



Canadian Institutes
of Health Research

Instituts de recherche
en santé du Canada

PROTECTED B WHEN COMPLETED

Routing Slip

This routing slip is to be included with your signature pages and is for CIHR's administrative use only.

Funding Opportunity

Operating Grant 2012-03-01

Application Number 275594

ResearchNet ID 160572

Suggested Peer Review Committees

1st: Cancer Biology & Therapeutics

2nd: Molecular & Cellular Biology of Cancer

Nominated Principal Applicant

Surname

GINGRAS

Given Names

Anne-Claude

PIN

157721

Project Title

A systems approach towards the therapeutic modulation of the acetylome

Relevant Research Area:

Title of Priority Announcement/Funding Pools:

Linked Programs:

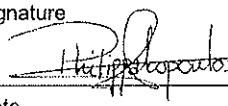
**Applicant Signatures**

The applicants are in the following order: Principal Applicant(s) and Co-Applicants. It is agreed that the general conditions governing grants, as well as the statement "Meaning of Signatures on Application Forms" as outlined in the CIHR Grants and Awards Guide, apply to any grant made pursuant to this application and hereby accepted by the applicant(s).

Consent to Disclosure of Personal Information

I understand that maintaining public trust in the integrity of researchers is fundamental to building a knowledge-based society. By submitting this application or by accepting funding from CIHR, NSERC and/or SSHRC, I affirm that I have read and I agree to respect all the policies of these Agencies that are relevant to my research, including the *Tri-Agency Framework: Responsible Conduct of Research* (<http://www.rcr.ethics.gc.ca/eng/policy-politique/framework-cadre/>). In cases of a serious breach of Agency policy, the Agency may publicly disclose my name, the nature of the breach, the institution where I was employed at the time of the breach and the institution where I am currently employed.

I accept this as a condition of applying for or receiving Agency funding and I consent to such disclosure.

Surname	Given Names	Role	Signature
Filippakopoulos	Panagis	Co-Applicant	X 
Institution	Faculty	Department	Date
Oxford University			
Surname	Given Names	Role	Signature
Pawson	Anthony	Co-Applicant	X 
Institution	Faculty	Department	Date
Samuel Lunenfeld Research Institute (Toronto)			
Surname	Given Names	Role	Signature
			X
Institution	Faculty	Department	Date
Surname	Given Names	Role	Signature
			X
Institution	Faculty	Department	Date
Surname	Given Names	Role	Signature
			X
Institution	Faculty	Department	Date



Canadian Institutes
of Health Research

Instituts de recherche
en santé du Canada

PROTECTED B WHEN COMPLETED

275594

GINGRAS, Anne-Claude
Mount Sinai Hospital (Toronto)

\$267244

Institution Signatures

It is agreed that the general conditions governing grants, as well as the statement "Meaning of Signatures on Application Forms" as outlined in the CIHR Grants and Awards Guide, apply to any grant made pursuant to this application and are hereby accepted by the applicant's and the applicant(s) employing Institution(s).

A signature is not required at institutions outside of Canada.

For institutions using the Electronic Approval Tool on ResearchNet, a signature is not required for block 1 if the Authorized Official can bind the institution to all obligations outlined in the "Meaning of Signatures on Application Forms". If the Authorized Official cannot bind the institution to all obligations in the "Meaning of Signatures on Application Forms", complete block 2.

1. Signature of Authorized Official: Mount Sinai Hospital (Toronto)

Print Name:

Date:

John Spreull, Senior VP, Finance

28/2/2012

Signature:

x

John Spreull

2. If the Authorized Official above cannot bind the institution to all obligations outlined in the "Meaning of Signatures on Application Forms", please provide additional signatures below as required.

Print Name:

Date:

DR. JAMES WOODGETT, DIRECTOR, RESEARCH

28/2/2012

Signature:

x

James Woodgett

Print Name:

Date:

Signature:

x



Canadian Institutes
of Health Research

Instituts de recherche
en santé du Canada

**PROTECTED WHEN
COMPLETED**

Appl. # 275594

Application Details

Funding Opportunity:

Operating Grant: 2011-2012 (2012-03-01)

Proposed Start Date:

Proposed End Date:

Nominated Principal Applicant/Candidate:

Surname GINGRAS

Given Names Anne-Claude

Institution

Faculty

Department

Samuel Lunenfeld Research Institute (Toronto)

Telephone

416-586-5027

Fax

E-mail

gingras@mshri.on.ca

Project Title:

A systems approach towards the therapeutic modulation of the acetylome

Primary location where research to be conducted: Samuel Lunenfeld Research Institute (Toronto)

Faculty:

Department:

Institution which will administer project funds (Institution Paid):

Mount Sinai Hospital (Toronto)

Location of proposed Activity:

Period of support requested: 4 Year(s) Month(s)

THE FOLLOWING SECTIONS ARE NOT APPLICABLE TO ALL PROGRAMS

Budget section - Amounts Requested from CIHR in the First Full Year:

Operating: 267244

Equipment: 0

Total Amount Requested: 267244

New

Renewal

Funding Reference Number (FRN):

Is this application a resubmission of a previously unsuccessful new application?

Yes No

Is this application a resubmission of a previously unsuccessful renewal application?

Yes No FRN #:

Have you applied to this program in the last two years?

Yes No

Is this a multi-center study?

Yes No



Appl. # 275594

Certification Requirements:

- Human subjects Human stem cells Animals Biohazards
- Environmental Impact Containment Level _____
- Clinical Trial
- Contains a randomized trial
- In order to carry out the proposed research in this application, an exemption from Health Canada under Section 56 of the Controlled Drugs and Substances Act is required. Should my application be approved, I understand that I will need to seek an exemption from Health Canada and provide this exemption to CIHR before any funding will be released for this application.

Other Project Information:

- For statistical purposes, does this application propose research involving Aboriginal people?
- Are sex (biological) considerations taken into account in this study?
- Are gender (socio-cultural) considerations taken into account in this study?

Please describe how sex and/or gender considerations will be considered in your research proposal:
Basic research - molecular mechanisms apply to both sexes

**Other Applicants**

Surname	Given Names	Role
Filippakopoulos	Panagis	Co-Applicant
Institution Oxford University	Department	Faculty

Surname	Given Names	Role
Pawson	Anthony	Co-Applicant
Institution Samuel Lunenfeld Research Institute (Toronto)	Department	Faculty

Surname	Given Names	Role
Institution	Department	Faculty

Surname	Given Names	Role
Institution	Department	Faculty

Surname	Given Names	Role
Institution	Department	Faculty

Surname	Given Names	Role
Institution	Department	Faculty

Surname	Given Names	Role
Institution	Department	Faculty

Surname	Given Names	Role
Institution	Department	Faculty

**Lay title of the research****Abstract suitable for preparation of a press release**

Every cell in our bodies receives signals which tell it whether to grow and divide, to differentiate into a specific cell type (e.g. a muscle cell or a neuron) or even to commit suicide. These signals are transmitted by classes of proteins, known as enzymes, which modify other proteins. These modified proteins are in turn recognized by specialized proteins to dictate the proper cellular response. A key regulatory system in the cell utilizes a chemical group called "acetyl" for the modifications and the response to signals. In many cancers, the enzymes responsible for the addition of the acetyl modification and the proteins that recognize this modification are mutated. Drugs that target the process of acetyl modification and recognition are being developed and are showing great promise in the clinic to treat diverse cancers. Yet, despite this fact, there is very little known regarding how specificity is acquired in this acetyl modification pathway. For example, there are 40 distinct proteins that are able to recognize the acetyl modification. Similar to a lock and key mechanism, these proteins need to recognize the acetyl modification only in the context of specific proteins, but what these proteins are is unclear at the moment. The co-applicants have established a strong expertise in the use of an approach called mass spectrometry, which essentially identifies proteins using a unique "fingerprint" that each protein produces, and structural biology, which allows them to "see" how the lock and key mechanism functions. By combining these techniques with molecular and cell biology approaches, the applicants will determine how the acetyl modification system is organized. They will also test newly designed drugs that target the acetyl modification system to assess in a timely fashion which of the new drugs hold potential as therapeutic agents.

Project Descriptors *

Acetylome, Bromodomain, Epigenetics, protein-protein interactions, cancer therapeutics, mass spectrometry, systems biology, proteomics, protein fusions

Areas of Research *

Primary
GENOMICS, PROTEOMICS, AND BIOINFORMATICS

Secondary
CANCER

Classification Codes *

Primary
CANCER THERAPY GENERAL

Secondary
MULTIPLE DISEASE RELEVANCE

Themes *

1st Biomedical

2nd

3rd

4th

Categories *

1st Cancer Research

2nd Genetics

3rd

4th

**Suggested Peer Review Committees:**1st Cancer Biology & Therapeutics2nd Molecular & Cellular Biology of Cancer**Suggested External Referee(s)***

Name Washburn, Michael

E-mail mpw@stowers.org

Area of Expertise chromatin, interaction proteomics, mass spectrometry

Name

E-mail

Area of Expertise

Summary of Research Proposal/Résumé de la proposition de recherche

Lysine acetylation on histones is key to the epigenetic regulation of gene expression, and is mediated by the action of acetyltransferases (writers) and deacetylases (erasers). Not surprisingly, the enzymes implicated in such a critical function are often found altered in cancers; in this regard, histone deacetylase inhibitors now constitute a prime target for cancer therapies. It has been recently realized, however, that acetyltransferases and deacetylases modify not only histones, but also a wide variety of proteins throughout the cell. Like phosphorylation, which can be recognized by dedicated protein modules (or “readers”, e.g. the SH2 domain), acetylation is recognized by a specific protein domain, the bromodomain. Bromodomain-containing proteins are often mutated and/or amplified in cancers, and bromodomains, which have deep pockets to accommodate acetylated lysine residues, are also viable targets for small molecule inhibition (for example, the BRD4 inhibitor JQ1 is a potent antitumor agent). Unfortunately, our current knowledge of bromodomain specificity and function is limited, hampering the rational design of therapeutic agents.

The Gingras and Pawson laboratories have developed extensive expertise in interaction proteomics and in the identification of post-translational modifications. Within the context of this application, we have recently developed optimized methods to identify protein-protein interactions for proteins associated with chromatin and identify and accurately quantify the acetylated sites. We have already cloned and stably expressed in mammalian cell lines 55 of the 75 proteins known to participate in acetyl lysine signalling, and initiated a systematic interaction mapping effort. This work has been done in coordination with a systematic effort at the Structural Genomics Consortium (SGC; Oxford) by co-applicant Filippakopoulos to systematically determine the structures of all human bromodomains and begin to define their specificity for acetylated lysines *in vitro* (Appendix 1; Filippakopoulos et al., *Cell*, *in press*). Through the efforts at the SGC, molecular probes that target the interaction of the bromodomains with acetylated peptides have also been developed. In summary, in the past year we have built the technical know-how and a series of resources which put us in an excellent position to provide a new view of the acetylome.

Here, we propose to systematically define the specificity and function of human bromodomains through identification of their binding partners and acetylated targets. This knowledge will assist in the development of new inhibitors of bromodomains, accelerating the development of new therapeutics.

We propose the following specific aims:

- 1) Identify the interactome for all components of the human acetylation machinery, including translocation products and mutated proteins
- 2) Define the specificity for each bromodomain-containing protein towards acetylated lysines
- 3) Reconstitute the acetylome specificity map
- 4) Test new bromodomain inhibitors

This project will enable us to better understand the function for each bromodomain-containing protein, as well as the mode of recognition of acetylated lysine residues by the bromodomain. With the growing interest in developing specific bromodomain inhibitors, this knowledge and the reagents generated will serve to rapidly assess, in a cellular context, the consequences of treatment with bromodomain inhibitors, and accelerate their progression to preclinical studies.

GINGRAS, Anne-Claude

Operating Grant/Subvention de fonctionnement Application/Demande 2012-03-01

Lay abstract/Résumé non scientifique

Every cell in our bodies receives signals which tell it whether to grow and divide, to differentiate into a specific cell type (e.g. a muscle cell or a neuron) or even to commit suicide. These signals are transmitted by classes of proteins, known as enzymes, which modify other proteins. These modified proteins are in turn recognized by specialized proteins to dictate the proper cellular response. A key regulatory system in the cell utilizes a chemical group called "acetyl" for the modifications and the response to signals. In many cancers, the enzymes responsible for the addition of the acetyl modification and the proteins that recognize this modification are mutated. Drugs that target the process of acetyl modification and recognition are being developed and are showing great promise in the clinic to treat diverse cancers. Yet, despite this fact, there is very little known regarding how specificity is acquired in this acetyl modification pathway. For example, there are 40 distinct proteins that are able to recognize the acetyl modification. Similar to a lock and key mechanism, these proteins need to recognize the acetyl modification only in the context of specific proteins, but what these proteins are is unclear at the moment. The co-applicants have established a strong expertise in the use of an approach called mass spectrometry, which essentially identifies proteins using a unique "fingerprint" that each protein produces, and structural biology, which allows them to "see" how the lock and key mechanism functions. By combining these techniques with molecular and cell biology approaches, the applicants will determine how the acetyl modification system is organized. They will also test newly designed drugs that target the acetyl modification system to assess in a timely fashion which of the new drugs hold potential as therapeutic agents.

The proposal builds on concerted efforts from functional proteomics and experts at the Lunenfeld (Toronto; Gingras and Pawson) and a structural biologist at the Structural Genomics Consortium (Oxford; Filippakopoulos). The collaboration between the Pawson group and Dr. Filippakopoulos is longstanding, and has defined modes of molecular recognition (e.g. Filippakopoulos et al., *Cell* 2008); all co-applicants have a collaborative publication *in press* (Filippakopoulos et al., *Cell*, 2012; appended) that forms the basis of parts of Aims 2 and 4. By solving the structure of 29 bromodomains and characterizing their binding properties on peptides in isolation, this study began revealing important elements of bromodomain recognition. In particular, we found that the bromodomains of the BET family BRD4 exhibits a marked preference for distinct diacetylated peptides, suggesting a direct recognition of multiple marks by a single bromodomain containing protein.

As defined in the preliminary data, the Pawson and Gingras groups have established a strong and collaborative expertise in functional proteomics. We have defined efficient protocols for affinity purification coupled to mass spectrometry (AP-MS; e.g. Chen and Gingras, *Methods*, 2007; Dunham et al., *Proteomics*, 2011) that we have employed for the identification of interaction partners for multiple signaling proteins (e.g. Chen et al., *J Biol Chem*, 2008; Goudreault et al., *Mol Cell Proteomics*, 2009; Kean et al., *J Biol Chem*, 2011). We have also developed bioinformatics tools enabling the tracking of biological material (OpenFreezer, Olkovsky et al., *Nat Methods*, 2011) and mass spectrometry data (ProHits, Liu et al., *Nat Biotech*, 2010; Liu et al., *submitted*). Since identifying true interaction partners from AP-MS data is challenging because of non-specific background interactors, we have developed with collaborators statistical tools enabling the identification of true interaction partners (SAINT; Breitkreutz et al., *Science*, 2010; Choi et al., *Nat Methods*, 2011; Choi et al., *submitted*). All these resources are freely distributed to the scientific community and are subjected to active development.

Despite these advances, the protocols that we had developed for AP-MS analysis of soluble proteins were not appropriate for the analysis of chromatin-associated proteins (under standard cell lysis conditions, most of the chromatin-associated proteins precipitate as an insoluble pellet). Fortunately, the Pawson and Gingras lab recruited a joint postdoctoral fellow in Fall 2010 who had developed during his PhD a method he called mChIP for the solubilization of chromatin coupled to AP-MS (Lambert et al., *Mol Cell Proteomics*, 2009; Lambert et al., *Mol Sys Biol*, 2010). Since joining the SLRI, Jean-Philippe Lambert adapted the mChIP approach to mammalian cell AP-MS data, and showed that it enabled recovery of interactions which were otherwise lost. Jean-Philippe also developed methods to enrich acetylated peptides from affinity-purified samples; this method is an essential component of the current proposal. As described in the “preliminary data” section of the proposal, we have already cloned, expressed, purified, and analyzed by AP-MS the majority of the proteins composing the acetylome system using this protocol. In the past year, our laboratories have also made significant progress towards robust mass spectrometry-based quantification of AP-MS data, using intensity-based quantification (e.g. Kean et al., *J Biol Chem* 2010), SRM-based quantification (e.g. Bisson et al., *Nature Biotech*, 2011; Yong et al., under review at *Science*), SWATH quantification (Lambert et al., *in prep*), and isotope-based quantification (e.g. Jorgensen et al., *Science*, 2009). We recently used intensity-based quantification to measure the affinity of isolated bromodomains derived from BRD4 towards acetylated histone peptides (Filippakopoulos et al., *Cell*, *in press*), demonstrating the feasibility of our aims.

Lastly, we have engaged in key collaborations, which position us very well for the successful completion of our aims. We have access to a growing number of phage affinity reagents to target bromodomains through a collaboration with Sachdev Sidhu (U Toronto), and are working with the lead expert on NUT midline carcinoma, Christopher French (Harvard). Collaborations with statisticians, mass spectrometry vendors and computational biologists are also in place to ensure a timely generation and analysis of our data within the short period of this grant.

Overview:

The cellular machinery responsible for the acetylation of lysines is critically implicated in cancer biology. In particular the only protein domain that recognizes acetylated lysines, the bromodomain, is present in multiple proteins implicated in chromatin biology and regulation of gene expression, and many of these proteins are deregulated in human tumors. Structurally (see Appendix 1), the recognition of acetylated peptides by the bromodomain makes it a druggable target. Through the work of one of the co-applicants (PF), the first anti-cancer bromodomain inhibitor (JQ1; specific for the BET family of bromodomains) was developed and shows promise against a number of tumor types. However, this work also highlighted the need for more information regarding the cellular context in which bromodomain-containing proteins and acetylation enzymes act, and what are the cellular targets for bromodomain recognition of acetylated peptides. Here, we propose to systematically define the specificity and function of human bromodomains through identification of their binding partners and acetylated targets. This project builds on the strengths and the synergy of the research team, and will accelerate the selection and development of new chemical probes to target this family.

Introduction

The acetylation machinery. Lysine acetylation is the process by which an acetyl group is transferred (from Acetyl Coenzyme A) to the epsilon amine of a lysine residue, a modification that is catalyzed by a family of lysine acetyltransferases (KATs; 18 genes in human). Similar to protein phosphorylation in which kinases and phosphatases oppose each other for the addition (“writing”) and removal (“erasing”) of the phosphate group, acetyl groups (KAc) added by KATs can be removed by lysine deacetylases (KDACS, histone-specific KDACS are known as HDACs, histone deacetylases)^{1,2}. KDACS are represented by two main classes which differ in their cofactor dependence, the KDACS proper, 11 genes, and the sirtuins, 7 genes. KAc marks are recognized by a single dedicated protein domain, the bromodomain, which acts as a “reader” of the acetylation mark and provides context-specific recognition (Fig 1). As detailed below, bromodomains are encoded in 44 genes (some proteins contain more than one bromodomain). Most of the bromodomain-containing proteins are comprised of several modular domains³, suggesting roles as molecular scaffolds (Fig 1c). Though not all bromodomain-containing proteins have been characterized in detail, the majority of those studied have roles in chromatin biology and regulation of gene expression (Table 1; please see⁴ for a recent review).

Lysine acetylation first came to the forefront as a histone modification⁵. Histone tails are modified by a number of post-translational modifications (PTMs) that present different cues to the cellular machinery (the so-called histone code⁶). In general, acetylation of histone tails acts to open the chromatin structure, activating transcription, although certain acetylation marks are associated instead with chromatin compaction and with other processes as well, such as metabolism and DNA repair⁷. Regulation of gene expression by histone acetylation status has been the driving force behind the development of KDAC inhibitors⁸, as detailed below. However, while first studied intensively in the context of the histone code, lysine acetylation is now recognized as a widespread PTM occurring on a large proportion of the proteome^{8,9}. So far, 18330 sites on 6870 proteins have been reported in total in human cells in a PTM repository (<http://www.phosphosite.org>), which is likely still an underestimation (as detailed below, we consistently detect new sites by mass spectrometry). The prevalence of acetyl lysine on thousands of cellular proteins underlies the vital importance of this modification and also highlights the lack of information regarding which KATs and KDACS are responsible for these acetylation events, and, just as importantly, which sites on which proteins are “read” by the different bromodomains.

The acetylome in disease. Tight control of the acetylation system (here referred to as the “acetylome”) is clearly critical to cellular homeostasis. This is evidenced by the fact that many of the components of the acetylome have been associated with disease, particularly cancer (Table 1). Acetylome alterations in disease include expression changes, but also a number of genetic alterations, most notably deletions,

point mutations and translocations. For example, the bromodomain-containing acetyltransferase KAT3A (also known as CREBBP, CREB binding protein) is a transcriptional co-activator mutated in Rubinstein-Taybi syndrome, a multiple congenital anomaly syndrome characterized by mental retardation, dysmorphic facial features, postnatal growth deficiency and predisposition to cancer. KAT3A is also translocated, to a variety of fusion partners (including other components of the acetylation machinery such as MLL, KAT6A, KAT6B; Fig 2), in acute monocytic leukemia (AML)¹⁰⁻¹². Mutations in the bromodomain-containing protein BRWD3 are associated with X-linked mental retardation. BRWD3 is also disrupted by translocation in B-cell chronic lymphocytic leukemia¹³. The overexpression of multiple KDACs in cancers prompted the testing of KDAC inhibitors as anticancer therapeutics. To date, two KDAC (HDAC) inhibitors have been approved for clinical use. Vorinostat, a broad specificity small molecule inhibitor targeting KDACs 1-11, was approved for the treatment of persistent or recurrent cutaneous T-cell lymphoma (CTCL)¹⁴. Recently, the cyclic peptide Romidepsin, targeting KDAC1 and KDAC2, was approved for treatment of CTCL¹⁵. Though these inhibitors show clinical promise, their cellular mode of action is still not clear. These compounds are also bound to be somewhat non-specific: besides the fact that most of the currently used inhibitors can target more than one KDAC, the number of direct acetylation targets for each of the KDACs (let alone their identity) is still largely unknown. Furthermore, most KDAC inhibitors have been shown in clinical trials to be ineffective as monotherapies⁸, reinforcing the need for improved therapeutic agents targeting the acetylome.

BRD3, BRD4 and NUT midline carcinomas. The bromodomain-containing protein BRD4 (and more rarely BRD3) is fused to the NUT (Nuclear protein in testis) protein (gene name C15ORF55) in a rare and aggressive form of squamous cell carcinoma called NUT midline carcinoma (NMC)¹⁶. So far, less than 100 NMC cases have been reported¹⁷. However, this may be due in part to misdiagnosis associated – until recently – with a lack of diagnostic tools and registry, according to our collaborator and leading NMC expert, Dr. C French (letter attached). The cell of origin is unknown for NMCs, and NMCs occur in a variety of sites, most commonly in the upper digestive track and the mediastinum. NMCs have a devastating clinical course, and though initial response to chemotherapy is observed, recurrence is rapid and insensitivity to subsequent treatment results in an average survival time of less than 10 months¹⁸. By contrast to most carcinomas, which have very complex karyotypes, NMCs usually harbor a single reciprocal [t(15;19)(q14;p13.1)] chromosomal translocation which results in fusing the promoter and most of the BRD4 coding sequence (including all modular domains) to the nearly complete coding sequence of NUT (BRD4-NUT; Fig 2)¹⁷. NUT expression is normally restricted to testis: consistent with this, the reciprocal translocation product (NUT-BRD4) has not been detected in NMC. The role of the NUT protein in testis is not completely clear, though it was reported to possess an acidic domain enabling interaction with KAT3B (EP300)¹⁹. On the basis of this information, it was suggested that in spermatids, NUT binds and activates KAT3B, leading to global increases in histone acetylation.

The biological function of BRD4 (and of the rarer NUT fusion partner BRD3) is better understood than that of NUT. These proteins belong to a subgroup of bromodomain-containing proteins known as the BET family (for bromodomain and extraterminal) that share the same structure: two bromodomains, followed by the extraterminal (ET) domain which may be involved in protein/DNA recognition. BET proteins regulate transcriptional elongation by providing a docking site for the P-TEFb complex. The kinase component of P-TEFb, CDK9, phosphorylates the C-terminal tail of RNA polymerase II, resulting in transcriptional upregulation²⁰. BRD4 is implicated in post-mitotic gene transcription, and recruits P-TEFb to mitotic chromosomes resulting in the expression of growth-promoting genes; it is expected to contribute to epigenetic memory²¹. Two major BRD4 transcript variants are annotated that encode “long” and “short” isoforms differing in their C-termini. Both of these variants, as well as the BRD4 fragment fused to NUT, contain the tandem bromodomains and the ET domain; the longest form appears to have an extended interactome, which include proteins from viral pathogens such as papillomaviruses²². The link to viral oncogenic proteins, as well as the control of Myc expression^{23, 24},

implicates BRD4 in other cancer-related events outside of the NMCs, and BRD4 was recently found in an RNAi screen to be a therapeutic target in AML²⁵.

The driving role of the BRD4-NUT or BRD3-NUT fusions in NMCs is well established: in cell lines derived from patients with NMCs, knockdown of BRD4-NUT or BRD3-NUT induces rapid terminal squamous differentiation and arrests proliferation²⁶. This also strongly indicates that a block in differentiation drives NMCs. How exactly the fusion proteins drive NMC is not completely clear, but different lines of evidence have suggested that they may sequester KAT3B from differentiation specific genes¹⁹.

Recognition of acetylated lysines by bromodomains. Despite variation in primary amino acid sequence, all bromodomains (a domain of ~120 aa) share a common fold, and bind to acetylated lysines in a similar manner²⁷. The bromodomain consists of a left-handed bundle of four alpha helices that are linked by more variable loop insert regions²⁸. The acetylated lysine is recognized in a central deep and largely hydrophobic cavity, as determined by co-crystal structures of bromodomains complexed with acetylated lysine-containing peptides²⁹. Anchoring of the acetylated lysine (or lysines, see below) is provided by hydrogen bonding to a conserved asparagine residue (Fig 3; other hydrogen bond donors such as the side chains of threonine or tyrosine can replace the asparagine). How bromodomains specifically recognize distinct acetylated peptides is beginning to be understood, in large part due to our group's efforts to systematically determine the structures of multiple bromodomains (Appendix 1; Filippakopoulos et al., *Cell*, *in press*). For example, while the structural fold of all bromodomains is similar, the electrostatic potential of the surfaces surrounding the KAc binding site is diverse (Appendix 1), suggesting that they recognize different sequences (e.g. positively charged histones may not be favored by bromodomains that have a positive charge, such as that of the first bromodomain of PBRM1). Note that differences in surface charge have been observed even within the same protein, including the first and second domains of BRD4 (Fig 3), suggesting that different bromodomains within a protein may target different sites, an hypothesis that we addressed directly for BRD4, as described below. Additional selectivity elements are provided by the diverse loop regions that distinguish the different bromodomains.

Targeting bromodomains for therapies. Disrupting the interaction between a bromodomain and its substrates could in theory provide a much more specific control over the disease-associated acetylome than is currently afforded by KDAC inhibitors. Structurally, the deep and relatively hydrophobic cavity for binding acetylated lysine residues offers a good target for the development of small molecule inhibitors of the bromodomain-KAc interaction⁴. Given that the surface and loops of the bromodomains are diverse, it is expected that highly specific inhibitors can be designed (Table 2). We (PF) developed with our collaborators the first anti-cancer bromodomain inhibitor, (+)-JQ1, based on a triazolodiazepine scaffold, which binds to and inhibits all BET-family bromodomains (Fig 4)³⁰. (+)-JQ1, but not its stereoisomer, binds in the KAc pocket of the bromodomain. The fact that BRD4-NUT and BRD3-NUT fusions are driving oncogenes in NMCs provided an excellent test case for the development of bromodomain inhibitors in an attempt to treat these rare and aggressive tumors. In cell culture, the cell permeable (+)-JQ1 inhibitor displaces BRD4-NUT from chromatin, induces squamous differentiation, and reduces proliferation³⁰. Importantly, in xenograft models of NMCs, the well-tolerated (+)-JQ1 inhibitor led to tumor regression and improved mouse survival³⁰. This demonstrates not only the feasibility of developing specific inhibitors for bromodomain-KAc binding, but also that these inhibitors are of potential clinical value. While initially tested on NMCs, which offered a clean genetic model for BRD4-driven cancers, bromodomain-specific inhibitors have therapeutic value in a wider range of cancers. For example, after identifying BRD4 as a potential therapeutic target in AML by RNA interference (RNAi), (+)-JQ1 was shown to exhibit robust anti-leukemic effects both *in vitro* and *in vivo*, and these were accompanied with terminal differentiation and elimination of the leukemic stem cells²⁵. This therapeutic effect is also evident in multiple myelomas, and occurs through regulation of the Myc oncogenic driver in these tumors²³. In addition, bromodomain inhibitors also have potential as anti-

inflammatories, e.g. the BRD4 specific I-BET inhibitor (structurally related to JQ1) reduces expression of pro-inflammatory genes in activated macrophages³¹. Inhibitors to bromodomains outside of the BET family are also being developed, for example against KAT3B (CREBBP^{32, 33}) and PCAF³⁴, and this will remain a very active area of research, given the involvement of the acetylome in disease, and the striking results obtained with the first generation bromodomain inhibitors.

How is the acetylome organized? There is a pressing need to characterize systematically and in an unbiased manner the function of each of the components of the acetylation machinery and the relationships between the writer-reader-eraser modules. As mentioned above, structurally, bromodomains offer attractive targets for therapies and are the topic of an intense drug discovery program, yet very little is known regarding the cellular context in which many of the bromodomain-containing proteins reside or their specificity for their targets *in vivo*. As shown in Appendix 1, we have used synthetic libraries of histone-derived peptides to provide information regarding the specificity of isolated bromodomains *in vitro*. This expanded the number of bromodomain substrates, but also led to the realization that flanking PTMs (especially phosphorylation and acetylation) have a strong influence on the recognition of marks, indicating that bromodomains often recognize combinations of marks rather than isolated acetylated lysines. This observation was previously made for BRDT which requires the presence of several acetylation sites for high-affinity binding to histone tails³⁵. Structural determination of BRD4 with different diacetylated peptides derived from histone H4 showed that both KAc groups bind within the same pocket, and that the different sequences exhibit the same mode of binding (Appendix 1). However, our studies have also outlined the need to use alternative approaches to the synthetic libraries to identify specific sequences to which each of the bromodomains are associated in a cellular context. This is important especially in light of the therapeutic design of inhibitors: the inhibitors should be able to displace relevant (and presumably higher affinity) KAc peptides from the bromodomain. As described in this proposal, we have assembled a team of co-applicants and collaborators to provide *in vivo* context for the acetylome and accelerate development and testing of bromodomain-specific inhibitors.

Preliminary data:

Towards mapping the interactome for the acetylation machinery

Our research team has a long-standing interest in understanding how PTMs affect protein-protein interactions³⁶⁻⁴⁶, and we (TP, ACG) have developed robust and sensitive experimental approaches to systematically identify interactions using affinity purification coupled to mass spectrometry⁴⁷⁻⁵⁷ (e.g. Appendix 2). An important factor for success of the AP-MS approach is the expression of the epitope-tagged protein at near endogenous levels. While several approaches can be employed to ensure appropriate expression levels, a system that we have found particularly useful is the Flp-In T-REx system (Invitrogen) in which isogenic stable integrants are generated in a HEK293 derivative, and recombinant protein expression is under the control of a tetracycline inducible promoter^{58, 59}. We have demonstrated that this system is ideal for the expression of allelic series (e.g. wild type proteins and proteins in which the bromodomain has been mutated; e.g. Fig 17). This expression system was selected for the generation of the interactome for the acetylation system (Fig 5). We have also elected to use proteins tagged at their N-termini with a 3xFLAG tag and a magnetic bead purification strategy, based on our previous work using different tagging approaches.

The standard affinity purification protocol developed in our labs for different projects was developed for soluble proteins; however, multiple components of the acetylation machinery are known to be associated with chromatin. In addition, key substrates, including histones, are essentially insoluble under standard protocols, and normally not recovered in AP-MS approaches⁶⁰. To circumvent this problem, postdoctoral fellow Jean-Philippe Lambert developed during his PhD an approach for AP-MS of chromatin associated proteins in yeast, which he termed mChIP (modified chromatin

immunoprecipitation^{61, 62}), which he has now optimized for mammalian AP-MS (Fig 6; reviewed in Appendix 3). He demonstrated that this approach enabled recovery of multiple additional interactors for chromatin-associated proteins. In some cases, this is linked to a better solubilization of the bait protein (e.g. BAZ2A), while in other cases (e.g. BRD2), this is most likely due to better solubilization of the interactors (Fig 7). He also demonstrated that for this type of data, the sensitivity of the mass spectrometer is important: using new generation instruments (here an Orbitrap Velos) results in the identification of significantly more interactions than was previously possible.

The mChIP approach works because it solubilizes chromatin-associating proteins by shearing and cleaving the DNA using a combination of sonication and nuclease treatment. At the same time, this solubilization – especially in the context of single step purification and very sensitive mass spectrometers – creates a significant level of background (proteins which either interact with the affinity matrix or with the epitope tag independently of the bait). As in our previous studies, however, we found that the background was quantitatively reproducible and that the quantitative information embedded in each mass spectrometry experiment could be used to provide a confidence value to each potential protein-protein interaction detected here. To provide meaningful scoring of the interactions, we developed a series of software tools that we are distributing freely to the scientific community. ProHits tracks all mass spectrometry data, and functions as an important tool to annotate, visualize and interpret mass spectrometry data⁶³. Recently, we fully integrated within the ProHits system the SAINT statistical tool. SAINT (Significance Analysis of INTERactome) uses quantitative information (e.g. spectral counts, which is how often peptides for a given protein have been detected in the mass spectrometer) for the recovery of each potential interactor across a series of negative control purifications as compared to its recovery in association with the different baits in the dataset^{48, 58, 64} (Fig 8; Appendix 4). We have found that a particularly successful procedure, ensuring the identification of true interactors, was to perform each experiment in at least two biological replicates, to score each replicate independently with SAINT (SAINT scores are between 0 and 1) and then average the independent scores (AvgP). By selecting AvgP ≥ 0.8 , all interactions included are high-confidence, and have been detected in each of the biological replicates (this is the cutoff used for the data presented here).

Using the approaches described above, the components of the acetylation system (wt proteins) are each being expressed stably in the Flp-In T-REx cell lines, and AP-MS is being performed in biological duplicates. Together, the team (ACG, TP, PF) has already assembled a collection containing all lysine acetyltransferases and deacetylases, and 41 of the bromodomain-containing proteins. We are well on our way to transferring all of these cDNAs to the expression vector, generating stable cell lines, and performing AP-MS (the status of this project is represented on Fig 9). So far, 57 bait proteins have been analyzed by mass spectrometry in at least biological duplicates. This has led to the recovery of ~23000 pairs of unfiltered interactions, and ~1500 interactions amongst ~450 proteins at a SAINT AvgP ≥ 0.8 (Fig 10). As expected, we recover many published interactions, though the vast majority of the interactions detected here are novel. Which interactions are mediated by the bromodomain-acetyl-lysine interaction will be addressed in the proposal using a collection of point mutants of the critical asparagine residue (Fig 3) in the background of the full-length proteins (mutants already available are depicted in Fig 9). While we are still in the data acquisition phase of the project, we have begun analyzing in more depth some of the therapeutically-relevant interactions, such as those established by the BET family of bromodomain-containing proteins (Fig 11) and KAT3A/KAT3B (CREBBP; Fig 12). We have also detected a previously unknown interaction between the poorly characterized HDAC11 and the DICER1 protein (implicated in miRNA processing⁶⁵), which we validated (Fig 13).

Determining specificity of bromodomains for acetyl-lysine residues

Our interactome map is critical to understand the processes in which bromodomain-containing proteins and other acetylome components are acting; however, how the recognition of the acetylated lysine by the bromodomain is related to the interactome is unclear at this point. To begin to characterize interactome specificity, we are proposing a multi-pronged approach to identify acetylation-dependent

interactions, both in the context of recombinantly-expressed isolated bromodomains (tested both on synthetic peptides and on cell extracts), but also within the context of the full length proteins expressed in human cells. A key element of our strategy is to be able to identify acetylation sites directly from immunoprecipitates of bromodomain-containing proteins. To start, we have modified our mChIP strategy to add, after the tryptic digest, an additional affinity step consisting of an anti-acetyl lysine pull-down at the peptide level and analysis on a high mass resolution mass spectrometer (Fig 14; we validated the specificity of the anti-KAc antibody on acetylated BSA). We first tested this approach on the bromodomain-containing protein BRPF3, a component of the MOZ/MORF histone acetyl-transferase complex^{66, 67}. In addition to the known BRPF3 interaction partners, our interaction proteomics approach (see above) identified new interactors for this protein. After the acetyl-lysine enrichment approach, we also found acetylation sites on BRPF3 itself, and on many of its interactors (Fig 15A, B). While several of these sites were present in a public repository (PhosphoSitePlus.org), 15 are new (and have been manually validated), highlighting the sensitivity of our approach. We also optimized the mass spectrometric identification of acetylated sites by analyzing fractions of the same purification on different mass spectrometers (and in the case of the Orbitrap Velos by employing different modes of fragmentation). Importantly, we show that – in comparison to a proteome-wide study of acetylation⁶⁸ – our targeted study of the acetylome facilitates the identification of polyacetylated peptides (Fig 15C, D). This is an important result, especially in light of our recent study where we demonstrated that at least some of the bromodomains exhibit a preference for polyacetylated sites, both *in vitro* and *in vivo*, and that this preference may contribute to the recognition of histone code marks (Fig 16). It is also noteworthy that many of the lysine acetyltransferases (especially those of the MYST family) require auto-acetylation for activity. As such, being able to identify and study these critical residues is important. As detailed in the proposal, and highlighted in Appendix 1 and Fig 16, we have also developed approaches to determine the specificity of recombinant isolated bromodomains against synthetic peptides and histone preparations. The list of the recombinant bromodomains currently available for our studies is in Fig 9.

Development of new methods for accurate mass spectrometry-based quantification

An important aspect of the proposal is our ability to accurately quantify interactions regulated by the recognition of the acetylation site (e.g. from a parallel purification of a wild-type bromodomain-containing protein and a mutant in which the acetyl-lysine recognition is prevented). Furthermore, we wish to monitor the enrichment of acetylated peptides and quantify their relative abundance. While we and others have used several different quantitative proteomics approaches, including isotopic labeling approaches such as SILAC⁶⁹, and label-free approaches such as spectral counting^{48, 70, 71} and intensity measurement of the precursor ion in the mass spectrometer (so-called MS1 quantitation⁷²), here we propose to use quantification methods based on the intensity of the product ions in the MS/MS spectrum. We have previously used single reaction monitoring (SRM) to monitor changes in interactomes accurately and with high sensitivity⁵² (Appendix 5). Since SRM method building (i.e. the selection of the best precursor ions and daughter ions to monitor) is time-consuming, we have also developed (in collaboration industrial partner AB SCIEX, letter appended) a modified approach called MRM^{HR}, which accelerates the development of quantification assays while providing the same accuracy on quantification (Zheng et al., submitted). Lastly, we have implemented a data-independent quantification approach called SWATH⁷³ which enables us to determine – again with high accuracy – the relative abundance of every identified species in a given dataset (Fig 17). Since the SWATH approach is very recent⁷³, we benchmarked it here to determine its efficiency in measuring interactome differences on cancer-associated mutations⁷⁴; this type of sample is very similar to what we are proposing to monitor in this proposal. As shown in Fig. 17, AP-MS with SWATH quantification reveals differences between the interactions established by wild type CDK4 (cyclin dependent kinase 4) and two point mutants identified in melanoma that act in a dominant manner^{74, 75}. As expected, we could readily detect the disruption of the interaction between the mutants and the cyclin dependent inhibitors of the

INK family. Interestingly, however, we could also quantify increased interactions between the mutants and components of the HSP90 chaperone complex (suggesting folding issues with the mutants), and with another class of CDK inhibitor. These novel changes were validated by immunoblotting, confirming the usefulness of the approach. Importantly, quantification is accurate at the peptide level, which will enable us to monitor changes in acetylation. Further, the instrument on which we are performing SWATH measurements is the same as that used for acetylation site mapping, the AB-SCIEX TripleTOF 5600, ensuring the compatibility of our platform.

Here, we propose to systematically define the specificity and function of human bromodomains through identification of their binding partners and acetylated targets. This knowledge will assist in the development of new inhibitors of bromodomains, in order to accelerate the development of new therapeutics.

We propose the following specific aims:

- 1) Identify the interactome for all components of the human acetylation machinery, including translocation products and mutated proteins
- 2) Define the specificity for each bromodomain-containing protein towards acetylated lysines
- 3) Reconstitute the acetylome specificity map
- 4) Test new bromodomain inhibitors

Aim 1: Identify the interactome for all components of the human acetylation machinery, including translocation products and mutated proteins

As described in the preliminary data (and figures 9-13) we are well on our way to creating a map of the steady-state interactome for all full-length components of the acetylation machinery. Here, we will complete and validate this map, and expand it to the study of fusion proteins and point mutants involving the components of the acetylation machinery. Note that, while we strive to provide a global map of acetylome specificity, our team and our collaborators have a particular interest in the BET family of bromodomains, and we will therefore start all aims by studying this important group of proteins.

1.1 Complete the interactome map for wt proteins

We will continue generating stable cell lines, and performing two biological replicate AP-MS analyses for each of the baits detailed in Figure 9; note that, based on the comparison of the instrument types displayed in Fig. 7, we will generate a complete dataset using more sensitive mass spectrometers (Orbitrap-Velos or 5600 TripleTOF). We are making a deliberate decision not to artificially increase the acetylation level in these cells via the addition of KDAC inhibitors; see Aim 2.1 for experiments in which these inhibitors will be used. When the entire dataset is generated, the data will be analyzed using our bioinformatics pipeline (with ProHits and SAINT) as detailed in the preliminary data section. While some of the cDNAs have been difficult to obtain due to their length, we (PF, collaboration with Stefan Knapp) have initiated a gene synthesis program to complete the dataset, and we therefore do not expect any major difficulty with this sub-aim. In parallel to the generation of the interaction data, we will be determining the subcellular localization in our stable cells for each of the proteins, as we are doing for our other projects; this is important, as many of the acetylome components have been completely uncharacterized so far. All data will be further analyzed for functional enrichment⁷⁶⁻⁷⁸ (with Dr. Gary Bader, letter attached) to provide additional information regarding the biological roles of the acetylation system (see Appendix 2 for an example of the type of analysis we are proposing to do). From the preliminary data that we have obtained thus far, we have detected interesting and previously unsuspected links between acetylome components and other cellular proteins. For example, HDAC11 interacts specifically with DICER1, raising the interesting hypothesis that acetylation may be implicated in the regulation of miRNA biology; as detailed in Aim 2.2, experiments are in progress to map potential acetylation sites on DICER1 and components of the miRNA pathway. It is intriguing that DICER1 has previously been linked to acetylation regulation but that the mechanisms are unknown⁷⁹. In addition to generating a large interactome map in exponentially growing cells, we will, when appropriate,

investigate condition-specific interactions. In the case of the BET family of bromodomain-containing proteins that are recruited to chromatin during mitosis^{21, 80}, we will also perform comparative analysis of the interactions in asynchronous cells, and in cells arrested with paclitaxel (Taxol; 5ug/ml for 16 hours) or nocodazole (100ng/ml for 16 hours); this may assist in determining regulated interactions and provide a more dynamic view of the protein-protein interactions mediated by acetylation changes (these experiments will be performed using quantitative proteomics, as described in Fig 17).

1.2 Validate interactions and define protein complexes interacting with acetylome components

While we have now demonstrated that our experimental and bioinformatics pipeline enables us to accurately identify bona fide interaction partners, as opposed to contaminants, in AP-MS data (we are initially using especially stringent filtering criteria for the acetylome project presented here to ensure that follow-up experiments are performed on valid interactions), we are interested in defining how many independent protein complexes can associate with the interactome. We have previously determined^{49, 51, 70} that iterative AP-MS (in which a prey identified in a round of AP-MS is in turn cloned and used as a bait in the next round of AP-MS) is an efficient method to define the modularity of the interactome, and identify mutually exclusive and cooperative interactions⁸¹. Here, up to 50 hits from the first round of AP-MS (selected based on the biological functions and the complexity of the interactomes) will be used for this purpose, to further define the assembly of the acetylation system. Additional interactions of biological interest (see, e.g. Fig. 12) will be further validated by IP-western, and co-localization in human cells, as we have done extensively in the past. Biological follow-up and structure/function analysis of new protein complexes will be performed as we recently did for a kinase/phosphatase complex linked to Cerebral Cavernous Formations, STRIPAK^{70, 72, 82}.

1.3 Determine the consequences of genetic aberration of the acetylome on the recruitment of interaction partners.

We have established a quantitative proteomics approach (which we refer to as AP-SWATH) that is particularly suited for unbiased interactome analysis of wt and mutant proteins. AP-SWATH was developed to speed-up the study of the functional consequences of a mutation on the interactome (Lambert et al., *in prep*; Fig 17). Our central hypothesis in developing this project is that the mutant or fusion protein may either acquire or lose interactions which were detected with the wild type protein, and that these interactions will help to understand the function of the protein in healthy and diseased cells (such mutations have been referred to as “edgetic” mutations⁷⁴). Here, AP-SWATH will be applied to compare the interactome of the acetylome components, of their point mutants, of fusion products involving the acetylome components, and of their translocation partners. We have already begun generating stable lines of BRD4, the BRD4-NUT fusion protein, and its fusion partner NUT, a poorly characterized protein (implicated in aggressive carcinoma; collaboration with C.A. French). This will be expanded to other known translocation products within the acetylome (Fig 2; Table 1), such as the KAT6B-KAT3A fusion⁴ which our team is committed to clone, as we have done with the full length, wild-type, proteins. A caveat to the use of the fusion proteins is that their expression levels may differ from that of the wild-type protein; in fact, we have already observed this phenomenon with the BRD4-NUT fusions (not shown). To prevent inducing or losing interactions based on expression alone, we will perform parallel AP-SWATH experiments in which the induction of protein expression with tetracycline will be performed using different concentrations of tetracycline or different expression times⁵⁹.

Robustness of the interactome identified after driving different expression levels will be monitored for those fusion proteins (and their wt counterparts) for which expression level differences are more dramatic. We will continue collaborating with Alexey Nesvizhskii (letter attached) on scoring these potentially more difficult interactions. The biological consequences of interaction with translocation products or mutant proteins will be investigated in appropriate systems. For example, for fusion proteins such as BRD4-NUT which is essential for transformation in midline carcinoma, we will test whether the interactions are essential for the transformation phenotype by performing transient or stable knockdowns of the interaction partners in NMC cells (we have access through CA French to several patient-derived

lines: 8645⁸³, 11060³⁰, 10326 (a BRD3-NUT line⁸⁴), 14169, and 879).

1.4 Map the interactome for wild-type and cancer-associated mutant bromodomain-containing proteins expressed at endogenous levels.

While our growing experience indicates that HEK293s constitute an excellent system for interaction mapping, they may not express all proteins expressed in different types of cancers. Also, while the interactions detected by AP-MS in this cell model are usually robust, it is important to be able to validate that the interactions do occur amongst endogenous proteins. We have therefore established a collaboration with Sachdev Sidhu (letter attached) to generate phage affinity reagents to different biotinylated and soluble bromodomains (we have successfully employed such reagents for AP-MS). Currently, 59 high-affinity Fabs have been generated against 14 bromodomains (Fig 9), and this collection is growing. After determining which of these Fabs are inhibitory to the bromodomain-KAc interaction and which are not, non-inhibitory sequences will be transferred into a full length IgG scaffold to use as affinity reagents. Ideally, we would want to test in parallel several affinity reagents per target, first on the stable cell lines, each expressing a full length tagged bromodomain-containing protein (obtaining multiple reagents is important for discerning the background from the true interaction partners in AP-MS experiments; inhibitory Fabs will be used in Aim 4). Once the affinity reagents are validated, they will be tested for AP-MS on endogenous protein interactions, first in an HEK293 cell lysate, next on other cellular sources, including cancer cell lines (including the NMC cells described above) which our groups (TP, ACG) use routinely, as appropriate. Other antibodies to endogenous proteins are available, including to the NUT fusion partner (different antibodies are available, both commercially, and through CA French). As we have done extensively in the past, we will also systematically validate new and interesting interactions by performing immunoprecipitation followed by immunoblotting with endogenous proteins, using either the phage affinity reagents described here, or standard antibodies.

Aim 2: Define the specificity for each bromodomain-containing protein toward acetylated lysines

Bromodomains are the only known domain that recognize the KAc group, and are expected to add specificity to the acetylation system, yet, little is known regarding their individual specificity for acetylated peptides. As detailed in Appendix 1, we have used pull-downs with biotinylated recombinant bromodomains to define the specificity for acetylated peptides on spot arrays, in ITC measurements, and after enrichment from histone preps (in combination with mass spectrometric analysis). We have also optimized an approach to combine affinity purification with acetyl-lysine enrichment and mass spectrometry (AP-KAc-MS; Fig 14). Here, we will use a multi-pronged approach to define the specificity of each bromodomain-containing protein for specific acetylated lysine peptides.

2.1 Define which interactions are mediated by the bromodomains

While Aim 1 will provide a global map of the interactome for bromodomain-containing proteins, some or many of the interactions we are detecting may not be mediated by a bromodomain-KAc interaction. Indeed, as shown in Fig 2, many of the bromodomain-containing proteins contain several additional protein interaction domains; furthermore, some of them contain more than a single bromodomain. As shown on Fig 9, we have generated stable cell lines for the expression of point mutants within bromodomains which disrupt the interaction with the acetylated target proteins. Guided by the results of Aim 1 (using the bromodomain-containing proteins that recover rich interactomes in our cell system), we will generate additional mutants and cells lines. As detailed in Aim 1.3, we will perform AP-SWATH on pairs of wt and mutant bromodomains (all within the context of the full-length proteins). In order to potentially increase the number of KAc-mediated interactions, the steady-state level of KAc in the cells will be increased in some of the experiments by the addition of Trichostatin A (TSA; 200nM for 6 hours); parallel purifications of the treated and untreated cells both in the context of the wt and the bromodomain mutant protein will be performed. This will enable the definition of the bromodomain-associated interactome, i.e. those interactions that are mediated by the direct interaction of a bromodomain with an acetyl-lysine substrate.

2.2 Identify acetylation sites on acetylome-associated proteins.

As shown in Fig 15 (with BRPF3), we have been successful at developing an approach to systematically map acetylated sites on binding partners of KATs, KDACs and bromodomain-containing proteins. We are now expanding this analysis to each member of the BET family (BRD2, BRD3, BRD4, BRDT; each of these proteins contains two bromodomains). Initially, the samples will be lysed in the presence of sodium butyrate, a generic deacetylase inhibitor to stabilize KAc sites prior to their purification (again, in some experiments, acetylation will be enhanced by treatment of the cells with KDAC inhibitors). Since the BET-bromodomain specific inhibitor JQ1 efficiently displaces those interactions that depend on the KAc-bromodomain interaction, parallel experiments will be performed on cells treated with the (+)-JQ1 and the inactive (-)-JQ1 compounds. Proteins detected in the absence of JQ1, but not in its presence will be deemed likely bromodomain-KAc-mediated interactors; their KAc sites will be inspected closely, and consensus sites will be determined (see 2.3). In parallel to the purification of the BET-associated KAc peptides, we will determine the acetylation sites in purifications of other acetylome components. As detailed above, we are particularly interested in determining whether DICER1 itself is acetylated (based on the recovery of DICER1 in the purifications of HDAC11); this will initially be monitored by immunoblotting of affinity-purified DICER1 using anti-acetyl-lysine antibodies. If we indeed do detect reactivity with the KAc antibody, the role of HDAC11 in mediating deacetylation will be tested by silencing HDAC11 (and other HDACs as controls) and monitoring DICER acetylation level. Sites will be mapped using the method defined above.

2.3 Define the specificity of bromodomains in isolation for peptides and proteins.

As shown in Fig 15 and Appendix 1, we have used biotinylated bromodomains to define KAc specificity preference for BRD4 bromodomain 1 and BRD4 bromodomain 2. This approach will be repeated, starting from purified histone preparations, first on each of the bromodomains from the BET family. In addition to using purified histones as starting material, we will employ tryptic peptides generated from a mixture of cell lines (e.g. K562, HeLa, U2OS and MCF-7), to insure a diverse source of starting material. We expect that if we perform the pull-down at the peptide level (rather than at the protein level like we have done so far) we may not need to couple our approach with anti-KAc enrichment; if this is not the case, we will perform a dual purification). To help identify relevant targets, we will incorporate a SILAC isotopic-labeling approach to generate pools of heavy and light lysates which will be used for these experiments. As controls, we will use validated inhibitors when available (e.g. the JQ1 inhibitor which we have used for BRD4; also see Aim 3), bromodomain inhibitory Fabs (if available) (Aim 4) and mutants in the bromodomains themselves. This is important, as peptide interactions can likely occur with surfaces outside of the KAc-binding pocket. Resulting enriched peptide sequences for each of the isolated bromodomains will be analyzed for the presence of consensus motifs, in collaboration with Gary Bader⁸⁵. We are initially focusing on the BET family of bromodomains since these are the targets of our intensive structural analysis and drug development (PF, collaboration S Knapp). Note that our mass spectrometric experiments enable us to look at combinatorial modifications on the KAc-peptides, providing that they are in close proximity; all the mass spectrometry data will be searched for potential multi-modifications (KAc, Lysine methylation, phosphoS/T/Y, etc.) and the motif enrichment analysis will also take combinatorial modifications into account. This will likely be important for the other bromodomains that exhibit preference for peptides that are multiply modified (Appendix 1). If we detect a bias for peptides harboring a different modification (e.g. phosphorylation), we could combine the approach with another affinity step for the particular modification (e.g. IMAC enrichment for phosphorylation⁸⁶). Once the data analysis on the BET family is completed, we will systematically apply the approach to all other bromodomains, again, prioritizing the analysis to targets of biological interest in cancer (clear candidates are the KAT3A/KAT3B bromodomains). Importantly, this work will be done in the context of a parallel effort (PF) to characterize specificity based on *in vitro* spot assays and isothermal calorimetry measurements (see Appendix 1). Advantages of the pull-down from lysates include the possibility of detection of combinatorially modified peptides, and detection of peptides from

real proteins, accelerating the study of the functional roles of the bromodomains. Peptides corresponding to the new sequences determined by mass spectrometry will be synthesized (PF) for SPOT assays followed by biophysical measurements and structural analysis, as appropriate. For bromodomains for which the substrate is completely unknown, this approach will enable the team (PF) to set up primary screens for small molecule binding (by ALPHA screen or surface plasmon resonance) utilizing the peptides identified and displacing them using small molecule scaffolds³⁰.

Aim 3. Reconstitute a specificity map for the acetylome

In Aim 2, we focused specifically on the reader component of the acetylation machinery, the bromodomain, as these proteins are expected to be key to the specificity of the acetylome, and are the target of our small molecule generation efforts. However, we recognize that to fully define the acetylome, the specificity of each of the KATs and KDACs for their substrates will need to be better defined; this is particularly relevant as we and others have identified several protein-protein interactions amongst the writer-eraser-reader components of the system (Fig 18), suggesting an interplay between these modules, and/or a coordination of the acetylation system (interestingly, we previously noted such an enrichment of kinase-kinase interactions in a systematic analysis of the phosphorylation system in yeast – Appendix 2). Here we will attempt to further refine the relationships between the acetylome components.

3.1. Identification of targets of KATs and KDACs. Whenever possible, we will use literature-curated data as they become available to identify the targets of KATs and KDACs (this is an active area of research, and we expect to see several reports on this throughout the duration of the grant). However, it is possible that for one or multiple KATs or KDACs of interest, no data will be available. If this were the case, we would employ the same approach employed recently for finding SIRT1 substrates⁸⁷, and identify the target proteins by quantitative mass spectrometry. Briefly, cell extracts will be treated with N-succinimidyl N-methylcarbamate to cap free lysines, the sample split into two equal aliquots and recombinant FLAG-tagged KDAC from human cells added to one aliquot. Newly exposed lysines will then be labeled with Sulfo-NHS-SS-Biotin, the peptides purified with streptavidin beads and the precise acetylation sites detected by mass spectrometry. To confirm novel KAT substrates, we will employ *in vitro* KAT assays, in which FLAG-tagged recombinant KAT will be incubated with affinity-purified putative substrates and heavy labeled acetyl-CoA (containing two C₁₃ atoms), as was performed by Dr. Lambert during his PhD (unpublished work). Following mass spectrometry analysis, we will be able to identify both *in vivo* (light KAc) and *in vitro* (heavy KAc) acetylation sites on a given protein (a caveat is that if the sites are fully occupied, we may need to strip them first, e.g. by employing deacetylases). Finally the intersection between the list of KAT / KDAC substrates and bromodomain-specific interactions will be analyzed computationally to provide an overview of the acetylome.

3.2 Mapping of KDAC/bromodomain relationships. HDAC inhibitors are used in the clinic, but there still remains much to learn regarding their specificity and functional consequences on cells. As shown in Fig 18, we have identified several physical interactions between KDACs and bromodomain-containing proteins in our network. Here we propose to further study these relationships by investigating the consequence(s) of depletion of specific KDACs using RNAi, or their inhibition by specific KDAC/HDAC inhibitors, on the interactions established by individual bromodomains. For example, we detected a physical interaction between BRD3 and HDAC2; HDAC2 will be silenced by RNAi, or inhibited by romidepsin (a Class I HDAC inhibitor), and the interactions established by FLAG-BRD3 will be quantified by AP-SWATH. If the physical interaction between HDAC2 and BRD3 is functional, we expect depletion of HDAC2 to lead to an increased association of HDAC2 targets to BRD3. We will integrate the data from Aims 1-3 and the literature (as in^{39, 88}), and build functional association maps for the writer-eraser-reader system. This should provide a much improved view of the acetylation system in human cells, better setting the stage for target identification for future drug design.

Aim 4. Test new small molecule inhibitors of bromodomains

As shown in Fig 4, we (PF, collaboration with S Knapp) have already demonstrated that new inhibitors targeted to bromodomains can show selectivity and exquisite efficacy in a mouse tumor model. We have therefore embarked upon a systematic effort to generate new bromodomain inhibitors, guided in this by the crystal structures we have determined (Appendix 1). What these inhibitors have in common is that they target the KAc binding pocket, thereby disrupting interactions with bromodomain substrates. Here, we will capitalize on the team expertise to accelerate the testing of first generation molecules. Inhibitors targeting KAT3A (BDOIA298; 105 nM by SPR), BAZ2B (BDG00021335; 96nM IC50 on an ALPHA assay displacing a H3K14 peptide) and BRD9 (BDOTC000133a; 300nM by SPR) will be available for testing within the next 6 months (1 or 2 distinct scaffolds per protein), and the team aims to generate and test inhibitors for 10 targets during the duration of this grant. These small molecule chemical probes have *in vitro* affinity in the order of 100 nM and *in vivo* cell efficacy at 1-10 μ M and are being developed as part of the SGC's epigenetics probe program (http://www.thesgc.org/scientists/chemical_probes). The testing platform developed here will also be employed to monitor inhibitory Fabs as they become available.

4.1 *In vitro* testing of interactome disruption. As the specificity of each bromodomain for KAc-peptides is determined (Aim 2.3) and validated *in vitro*, synthetic peptides derived from these target sites will first be employed for affinity measurements and initial testing of the different drug-like molecules using our standard platforms (Appendix 1^{30, 89}). Specificity will be determined by employing soluble bromodomains from related and distant families. We (PF) have extensive expertise in this area and do not anticipate major difficulties. This initial screening will be followed by confirmatory pull-down experiments in a cell lysate (as in Aim 2.3) that interactions with all KAc-peptides is indeed abrogated (quantification by AP-SWATH will be performed). If the drug works as expected, there should be a marked reduction of KAc-peptide interaction, as we demonstrated for each of the isolated domains from BRD4 after addition of the JQ1 inhibitor (Fig 11, 16; Appendix 1). In addition to the small molecule inhibitors developed by the team, Fabs generated in collaboration with Sachdev Sidhu (letter appended) will be tested in the same system. Inhibitory Fabs will be further tested in Aim 4.3.

4.2 Disruption of the interactome within the context of full-length proteins. The SGC pipeline strives at making potent, selective and cell permeable chemical probes, and the compounds we will be testing here will have already been subjected to extensive testing in Oxford. If the inhibitors are membrane permeable (as is JQ1), they will be used to treat stable cell lines expressing the target bromodomain, and parallel purification and LC-MS/MS analysis of cells treated or not with the inhibitor will be performed (as in Fig 16). This is important to ensure that within the context of a living cell, the inhibitor is sufficient for dissociation. In the event that the first generation inhibitor is not cell permeable, we will test whether it can disrupt complex formation in digitonin-permeabilized cells⁹⁰. Alternatively, we will determine whether the inhibitor can disrupt interactions pre-established in a cell lysate. For cell permeable inhibitors, we will perform parallel measurements of proliferation rates with different concentrations of inhibitors and AP-MS or IP/western analysis. Taken together, these results should enable the team to refine the drug design and/or focus on the most promising chemical probes.

4.3 Intracellular scFvs as molecular probes. Fab sequences demonstrated to be potent for the disruption of KAc-bromodomain interactions and proven to be specific for a given bromodomain would constitute outstanding probes, if delivered within the cell, to accelerate the understanding of the molecular functions of the KAc-bromodomain interaction. Our collaborator Sachdev Sidhu will transfer the sequences we validate in Aim 4.1 into an expression system suitable for the expression of single-chain Fv molecules in mammalian cells; these ScFvs will be affixed to an epitope tag (we will use a dual epitope comprised of HA and StreptagII; SH tag), and transfected into our stable cell lines expressing the cognate FLAG-tagged bromodomain-containing protein (initially, we will perform transient transfection). We will be monitoring first that the SH-tagged scFv can precipitate the FLAG-tagged protein by IP/western. If this is the case, we will assess whether it exhibits specificity by comparing the interactions established by the SH-scFv to those of the FLAG-bromodomain-containing protein (they

should be the same). After this validation, parallel comparisons of the interactomes of the FLAG-tagged protein in the presence or absence of the SH-scFv protein will reveal which portion of the interactome is displaced by the scFv. The biological role of the scFv will then be directly tested by analyzing the consequences of its overexpression using appropriate readouts for the given bromodomain-KAc interactions being disrupted (e.g. changes in proliferation, etc.).

Perspectives

In summary, our project will provide researchers worldwide with several valuable datasets. First, we will be providing a comprehensive and high quality for all the components of the acetylation machinery. This interaction map will be unique: to the best of our knowledge, all other laboratories performing high-throughput AP-MS are focusing on the soluble fraction, and miss the interactions involving chromatin components, which are essential to understand the function of the acetylation machinery. Importantly, this map will also extend to fusion products mutated in cancers, which could help providing a better understanding of the mechanisms of cellular transformation by these oncogenes. Next, we will provide a map of the specificity for the recognition of acetylated lysine by bromodomains, which will complement the data we published from on peptide array efforts (Appendix 1). This is important, since many of the bromodomains do not currently have any known target. This information should enable us to uncover additional relationships between the components of the acetylation machinery (KAT, KDAC, bromodomain) and crosstalk to cellular signaling. Lastly, we will use our toolset to test new chemical (and Fab) bromodomain inhibitors. The strength of this proposal is that there is a strong synergy between team members and integration between the Toronto (proteomics) and Oxford (structural and chemical biology) groups: Essentially, new interactions or KAc recognition sites discovered here will fuel the determination of additional co-crystal structures, allowing for the development of new small molecule inhibitors. The reagents generated for the structural analysis (e.g. biotinylated recombinant proteins) will in turn be used for additional proteomics analysis here. Importantly, the new small molecule inhibitors of specific bromodomains will be rapidly tested in the cognate cell line / biotinylated protein assay in Toronto, accelerating the development of well-characterized molecular probes that can be used to perturb and better understand the biological systems in which bromodomains participate.

The team

The team is led by three dedicated researchers: at the SLRI (Toronto), AC Gingras (start of independent career in 2005) and Distinguished Investigator T Pawson specialize in functional proteomics and their application to the study of interaction specificity and signaling events. P Filippakopoulos (start of independent career in 2011) at the Structural Genomics Consortium in Oxford is an expert in structural biology and rational drug design (see letter of collaboration by the SGC group leader in structural biology, Stefan Knapp). PF has already received funding for his part of the project (structural biology and inhibitor development), and we are therefore only requesting funds for the Toronto component of the research project (proteomics). TP and ACG co-direct the SLRI mass spectrometry facility, and have built through capital equipment grants and an intensive collaboration with the mass spectrometry vendor AB-SCIEX the infrastructure necessary for the completion of this project. All co-applicants (including Dr. Lambert) are co-authors on the appended manuscript in *Cell*, which provides the foundation for much of Aims 2 and 3, and have a very active and open collaboration, which will be maintained via frequent webinars and yearly visits. In Toronto, the work will be carried out by a postdoc co-supervised by TP/ACG, JP Lambert who brings a strong expertise in proteomics of chromatin-associated complexes, by B Badillo (cloning and cell line generation), A Veri (endogenous interactions in cancer cells and pharmacology screens), B Larsen (SWATH mass spectrometry), M Tucholska (mass spectrometric analysis) and J Zhang (data analysis). The team will be rounded up by other SGC-Oxford team members, S Picaud (recombinant protein engineering) and I Felletar (protein production and biophysical assays). Lastly, we have enlisted the help of a number of colleagues to provide additional expertise to enable us to accomplish all the aims in the short time frame (4-years) of funding requested.

REFERENCES

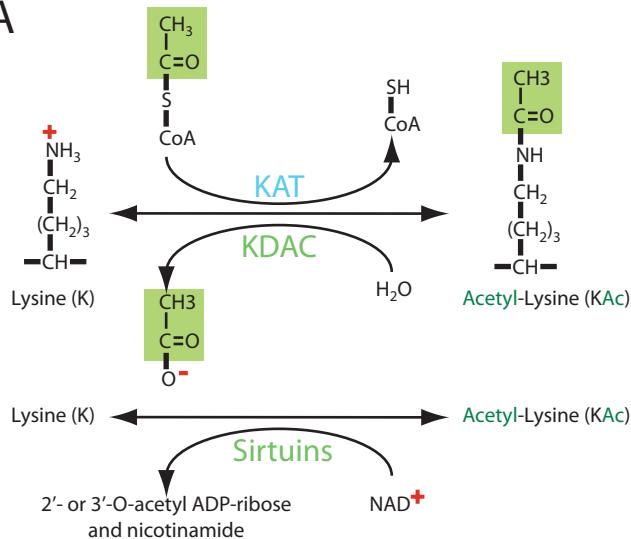
1. Kouzarides, T. Acetylation: a regulatory modification to rival phosphorylation? *Embo J* **19**, 1176-1179 (2000).
2. Lim, W.A. & Pawson, T. Phosphotyrosine signaling: evolving a new cellular communication system. *Cell* **142**, 661-667 (2010).
3. Pawson, T. & Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445-452 (2003).
4. Muller, S., Filippakopoulos, P. & Knapp, S. Bromodomains as therapeutic targets. *Expert Rev Mol Med* **13**, e29 (2011).
5. Allfrey, V.G., Faulkner, R. & Mirsky, A.E. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of RNA Synthesis. *Proc Natl Acad Sci U S A* **51**, 786-794 (1964).
6. Strahl, B.D. & Allis, C.D. The language of covalent histone modifications. *Nature* **403**, 41-45 (2000).
7. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
8. Thurn, K.T., Thomas, S., Moore, A. & Munster, P.N. Rational therapeutic combinations with histone deacetylase inhibitors for the treatment of cancer. *Future Oncol* **7**, 263-283 (2011).
9. Zhao, S. et al. Regulation of cellular metabolism by protein lysine acetylation. *Science* **327**, 1000-1004 (2010).
10. Sobulo, O.M. et al. MLL is fused to CBP, a histone acetyltransferase, in therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3). *Proc Natl Acad Sci U S A* **94**, 8732-8737 (1997).
11. Troke, P.J., Kindle, K.B., Collins, H.M. & Heery, D.M. MOZ fusion proteins in acute myeloid leukaemia. *Biochem Soc Symp*, 23-39 (2006).
12. Vizmanos, J.L. et al. t(10;16)(q22;p13) and MORF-CREBBP fusion is a recurrent event in acute myeloid leukemia. *Genes Chromosomes Cancer* **36**, 402-405 (2003).
13. Kalla, C. et al. Translocation t(X;11)(q13;q23) in B-cell chronic lymphocytic leukemia disrupts two novel genes. *Genes Chromosomes Cancer* **42**, 128-143 (2005).
14. Mann, B.S., Johnson, J.R., Cohen, M.H., Justice, R. & Pazdur, R. FDA approval summary: vorinostat for treatment of advanced primary cutaneous T-cell lymphoma. *Oncologist* **12**, 1247-1252 (2007).
15. Piekarz, R.L. et al. Phase II multi-institutional trial of the histone deacetylase inhibitor romidepsin as monotherapy for patients with cutaneous T-cell lymphoma. *J Clin Oncol* **27**, 5410-5417 (2009).
16. French, C.A. et al. BRD4-NUT fusion oncogene: a novel mechanism in aggressive carcinoma. *Cancer Res* **63**, 304-307 (2003).
17. French, C.A. Pathogenesis of NUT Midline Carcinoma. *Annu Rev Pathol* **7**, 247-265 (2012).
18. French, C.A. Demystified molecular pathology of NUT midline carcinomas. *J Clin Pathol* **63**, 492-496 (2010).
19. Reynoird, N. et al. Oncogenesis by sequestration of CBP/p300 in transcriptionally inactive hyperacetylated chromatin domains. *Embo J* **29**, 2943-2952 (2010).
20. Wu, S.Y. & Chiang, C.M. The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. *J Biol Chem* **282**, 13141-13145 (2007).
21. Zhao, R., Nakamura, T., Fu, Y., Lazar, Z. & Spector, D.L. Gene bookmarking accelerates the kinetics of post-mitotic transcriptional re-activation. *Nat Cell Biol* **13**, 1295-1304 (2011).
22. Rahman, S. et al. The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol Cell Biol* **31**, 2641-2652 (2011).
23. Delmore, J.E. et al. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904-917 (2011).

24. Mertz, J.A. et al. Targeting MYC dependence in cancer by inhibiting BET bromodomains. *Proc Natl Acad Sci U S A* **108**, 16669-16674 (2011).
25. Zuber, J. et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478**, 524-528 (2011).
26. Yan, J., Diaz, J., Jiao, J., Wang, R. & You, J. Perturbation of BRD4 protein function by BRD4-NUT protein abrogates cellular differentiation in NUT midline carcinoma. *J Biol Chem* **286**, 27663-27675 (2011).
27. Mujtaba, S., Zeng, L. & Zhou, M.M. Structure and acetyl-lysine recognition of the bromodomain. *Oncogene* **26**, 5521-5527 (2007).
28. Dhalluin, C. et al. Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491-496 (1999).
29. Owen, D.J. et al. The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *Embo J* **19**, 6141-6149 (2000).
30. Filippakopoulos, P. et al. Selective inhibition of BET bromodomains. *Nature* **468**, 1067-1073 (2010).
31. Nicodeme, E. et al. Suppression of inflammation by a synthetic histone mimic. *Nature* **468**, 1119-1123 (2010).
32. Borah, J.C. et al. A small molecule binding to the coactivator CREB-binding protein blocks apoptosis in cardiomyocytes. *Chem Biol* **18**, 531-541 (2011).
33. Gerona-Navarro, G. et al. Rational design of cyclic peptide modulators of the transcriptional coactivator CBP: a new class of p53 inhibitors. *J Am Chem Soc* **133**, 2040-2043 (2011).
34. Zeng, L. et al. Selective small molecules blocking HIV-1 Tat and coactivator PCAF association. *J Am Chem Soc* **127**, 2376-2377 (2005).
35. Moriniere, J. et al. Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* **461**, 664-668 (2009).
36. Chen, C., Nott, T.J., Jin, J. & Pawson, T. Deciphering arginine methylation: Tudor tells the tale. *Nat Rev Mol Cell Biol* **12**, 629-642 (2011).
37. Pawson, T. & Kofler, M. Kinome signaling through regulated protein-protein interactions in normal and cancer cells. *Curr Opin Cell Biol* **21**, 147-153 (2009).
38. Filippakopoulos, P. et al. Structural coupling of SH2-kinase domains links Fes and Abl substrate recognition and kinase activation. *Cell* **134**, 793-803 (2008).
39. Linding, R. et al. NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res* **36**, D695-699 (2008).
40. Hardy, W.R. et al. Combinatorial ShcA docking interactions support diversity in tissue morphogenesis. *Science* **317**, 251-256 (2007).
41. Wells, C.D. et al. A Rich1/Amot complex regulates the Cdc42 GTPase and apical-polarity proteins in epithelial cells. *Cell* **125**, 535-548 (2006).
42. Jones, N. et al. Nck adaptor proteins link nephrin to the actin cytoskeleton of kidney podocytes. *Nature* **440**, 818-823 (2006).
43. Gingras, A.C. et al. Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes Dev* **13**, 1422-1437 (1999).
44. Gingras, A.C., Kennedy, S.G., O'Leary, M.A., Sonenberg, N. & Hay, N. 4E-BP1, a repressor of mRNA translation, is phosphorylated and inactivated by the Akt(PKB) signaling pathway. *Genes Dev* **12**, 502-513 (1998).
45. Gingras, A.C. et al. Hierarchical phosphorylation of the translation inhibitor 4E-BP1. *Genes Dev* **15**, 2852-2864 (2001).
46. Filippakopoulos, P., Muller, S. & Knapp, S. SH2 domains: modulators of nonreceptor tyrosine kinase activity. *Curr Opin Struct Biol* **19**, 643-649 (2009).

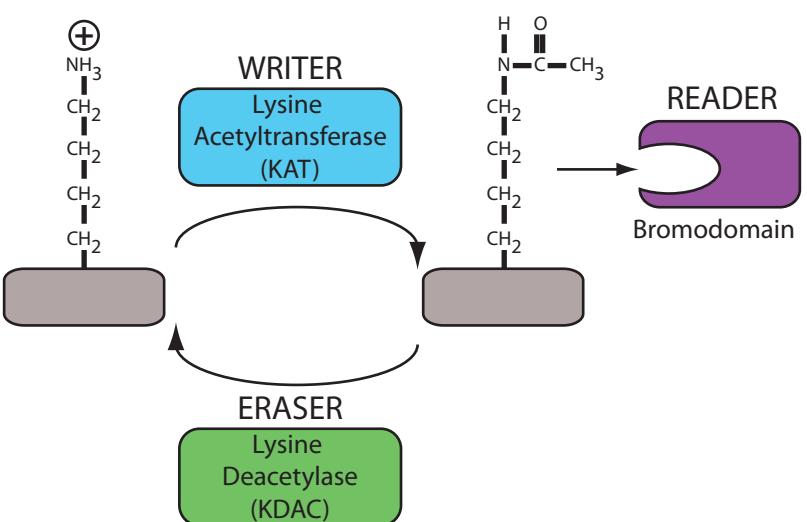
47. Dunham, W.H. et al. A cost-benefit analysis of multidimensional fractionation of affinity purification-mass spectrometry samples. *Proteomics* **11**, 2603-2612 (2011).
48. Breitkreutz, A. et al. A global protein kinase and phosphatase interaction network in yeast. *Science* **328**, 1043-1046 (2010).
49. Gingras, A.C., Gstaiger, M., Raught, B. & Aebersold, R. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* **8**, 645-654 (2007).
50. Chen, G.I. & Gingras, A.C. Affinity-purification mass spectrometry (AP-MS) of serine/threonine phosphatases. *Methods* **42**, 298-305 (2007).
51. Gingras, A.C. et al. A novel, evolutionarily conserved protein phosphatase complex involved in cisplatin sensitivity. *Mol Cell Proteomics* **4**, 1725-1740 (2005).
52. Bisson, N. et al. Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nat Biotechnol* **29**, 653-658 (2011).
53. Chen, C. et al. Mouse Piwi interactome identifies binding mechanism of Tdrkh Tudor domain to arginine methylated Miwi. *Proc Natl Acad Sci U S A* **106**, 20336-20341 (2009).
54. Ingham, R.J. et al. WW domains provide a platform for the assembly of multiprotein networks. *Mol Cell Biol* **25**, 7092-7106 (2005).
55. Smith, M.J., Kulkarni, S. & Pawson, T. FF domains of CA150 bind transcription and splicing factors through multiple weak interactions. *Mol Cell Biol* **24**, 9274-9285 (2004).
56. Jin, J. et al. Proteomic, functional, and domain-based analysis of in vivo 14-3-3 binding proteins involved in cytoskeletal regulation and cellular organization. *Curr Biol* **14**, 1436-1450 (2004).
57. Ho, Y. et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180-183 (2002).
58. Skarra, D.V. et al. Label-free quantitative proteomics and SAINT analysis enable interactome mapping for the human Ser/Thr protein phosphatase 5. *Proteomics* **11**, 1508-1516 (2011).
59. Glatter, T., Wepf, A., Aebersold, R. & Gstaiger, M. An integrated workflow for charting the human interaction proteome: insights into the PP2A system. *Mol Syst Biol* **5**, 237 (2009).
60. Lambert, J.P., Pawson, A.J. & Gingras, A.C. Mapping physical interactions within chromatin by proteomic approaches. *Proteomics* **in press** (2012).
61. Lambert, J.P. et al. Defining the budding yeast chromatin-associated interactome. *Mol Syst Biol* **6**, 448 (2010).
62. Lambert, J.P., Mitchell, L., Rudner, A., Baetz, K. & Figeys, D. A novel proteomics approach for the discovery of chromatin-associated protein networks. *Mol Cell Proteomics* **8**, 870-882 (2009).
63. Liu, G. et al. ProHits: integrated software for mass spectrometry-based interaction proteomics. *Nat Biotechnol* **28**, 1015-1017 (2010).
64. Choi, H. et al. SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat Methods* **8**, 70-73 (2011).
65. Ji, X. The mechanism of RNase III action: how dicer dices. *Curr Top Microbiol Immunol* **320**, 99-116 (2008).
66. Ullah, M. et al. Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes. *Mol Cell Biol* **28**, 6828-6843 (2008).
67. Doyon, Y. et al. ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol Cell* **21**, 51-64 (2006).
68. Choudhary, C. et al. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834-840 (2009).
69. Jorgensen, C. et al. Cell-specific information processing in segregating populations of Eph receptor ephrin-expressing cells. *Science* **326**, 1502-1509 (2009).
70. Goudreault, M. et al. A PP2A phosphatase high density interaction network identifies a novel striatin-interacting phosphatase and kinase complex linked to the cerebral cavernous malformation 3 (CCM3) protein. *Mol Cell Proteomics* **8**, 157-171 (2009).

71. Chen, G.I. et al. PP4R4/KIAA1622 forms a novel stable cytosolic complex with phosphoprotein phosphatase 4. *J Biol Chem* **283**, 29273-29284 (2008).
72. Kean, M.J. et al. Structure-function analysis of core STRIPAK Proteins: a signaling complex implicated in Golgi polarization. *J Biol Chem* **286**, 25065-25075 (2011).
73. Gillet, L.C. et al. Targeted data extraction of the MS/MS spectra generated by data independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* (2012).
74. Zhong, Q. et al. Edgetic perturbation models of human inherited disorders. *Mol Syst Biol* **5**, 321 (2009).
75. Zuo, L. et al. Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. *Nat Genet* **12**, 97-99 (1996).
76. Merico, D., Isserlin, R. & Bader, G.D. Visualizing gene-set enrichment results using the Cytoscape plug-in enrichment map. *Methods Mol Biol* **781**, 257-277 (2011).
77. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G.D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, e13984 (2010).
78. Warde-Farley, D. et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* **38**, W214-220 (2010).
79. Giles, K.E., Ghirlando, R. & Felsenfeld, G. Maintenance of a constitutive heterochromatin domain in vertebrates by a Dicer-dependent mechanism. *Nat Cell Biol* **12**, 94-99; sup pp 91-96 (2010).
80. Dey, A. et al. A bromodomain protein, MCAP, associates with mitotic chromosomes and affects G(2)-to-M transition. *Mol Cell Biol* **20**, 6537-6549 (2000).
81. Choi, H., Kim, S., Gingras, A.C. & Nesvizhskii, A.I. Analysis of protein complexes through model-based biclustering of label-free quantitative AP-MS data. *Mol Syst Biol* **6**, 385 (2010).
82. Ceccarelli, D.F. et al. CCM3/PDCD10 heterodimerizes with germinal center kinase III (GCKIII) proteins using a mechanism analogous to CCM3 homodimerization. *J Biol Chem* **286**, 25056-25064 (2011).
83. Schwartz, B.E. et al. Differentiation of NUT midline carcinoma by epigenomic reprogramming. *Cancer Res* **71**, 2686-2696 (2011).
84. French, C.A. et al. BRD-NUT oncproteins: a family of closely related nuclear proteins that block epithelial differentiation and maintain the growth of carcinoma cells. *Oncogene* **27**, 2237-2242 (2008).
85. Tonikian, R. et al. A specificity map for the PDZ domain family. *PLoS Biol* **6**, e239 (2008).
86. Bodenmiller, B. & Aebersold, R. Quantitative analysis of protein phosphorylation on a system-wide scale by mass spectrometry-based proteomics. *Methods Enzymol* **470**, 317-334 (2010).
87. Andersen, J.L. et al. A biotin switch-based proteomics approach identifies 14-3-3zeta as a target of Sirt1 in the metabolic regulation of caspase-2. *Mol Cell* **43**, 834-842 (2011).
88. Linding, R. et al. Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415-1426 (2007).
89. Filippakopoulos, P. et al. Benzodiazepines and benzotriazepines as protein interaction inhibitors targeting bromodomains of the BET family. *Bioorg Med Chem* (2011).
90. Chi, N.C., Adam, E.J., Visser, G.D. & Adam, S.A. RanBP1 stabilizes the interaction of Ran with p97 nuclear protein import. *J Cell Biol* **135**, 559-569 (1996).

A



B



C

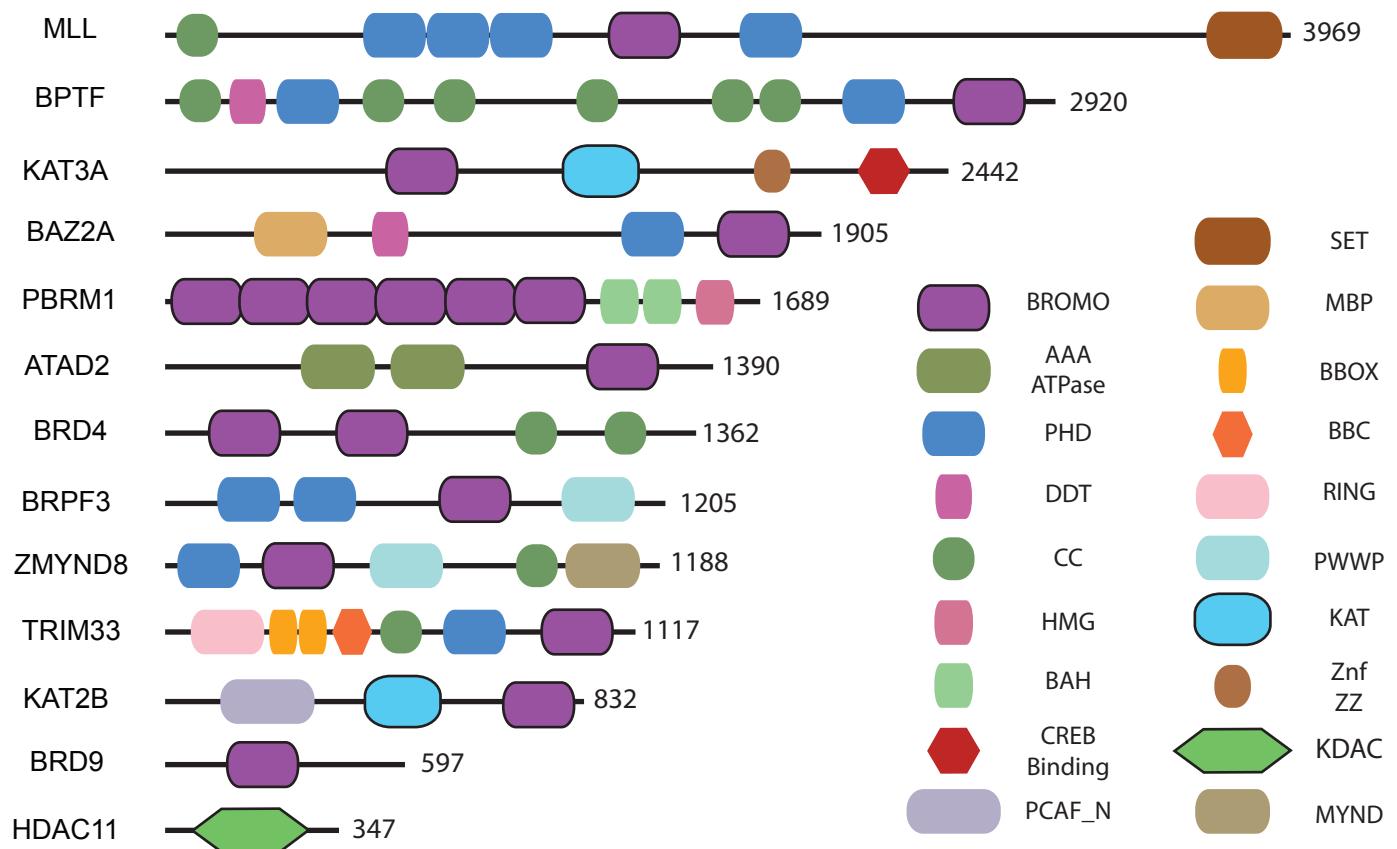


Figure 1. Lysine acetylation signaling system.

A) Lysine acetyltransferases (KATs; previously termed histone acetyltransferases, HATs) use acetyl CoA to transfer an acetyl group to the ϵ -amine of lysine. Two classes of enzymes can remove the acetyl group, the lysine deacetylases proper (KDACs; previously referred to also as histone deacetylases, HDACs), and the NAD^+ dependent sirtuins. **B)** The lysine acetylation system viewed as a writer (KAT), reader (bromodomain-containing proteins), eraser (KDAC) module. **C)** Overview of the architecture of the acetylome components (only selected proteins are shown); modular protein domains are depicted (see our website for description of protein domains; <http://pawsonlab.mshri.on.ca/>). Note that some proteins, e.g. KAT2B and KAT3A contain both a KAT module and a bromodomain.

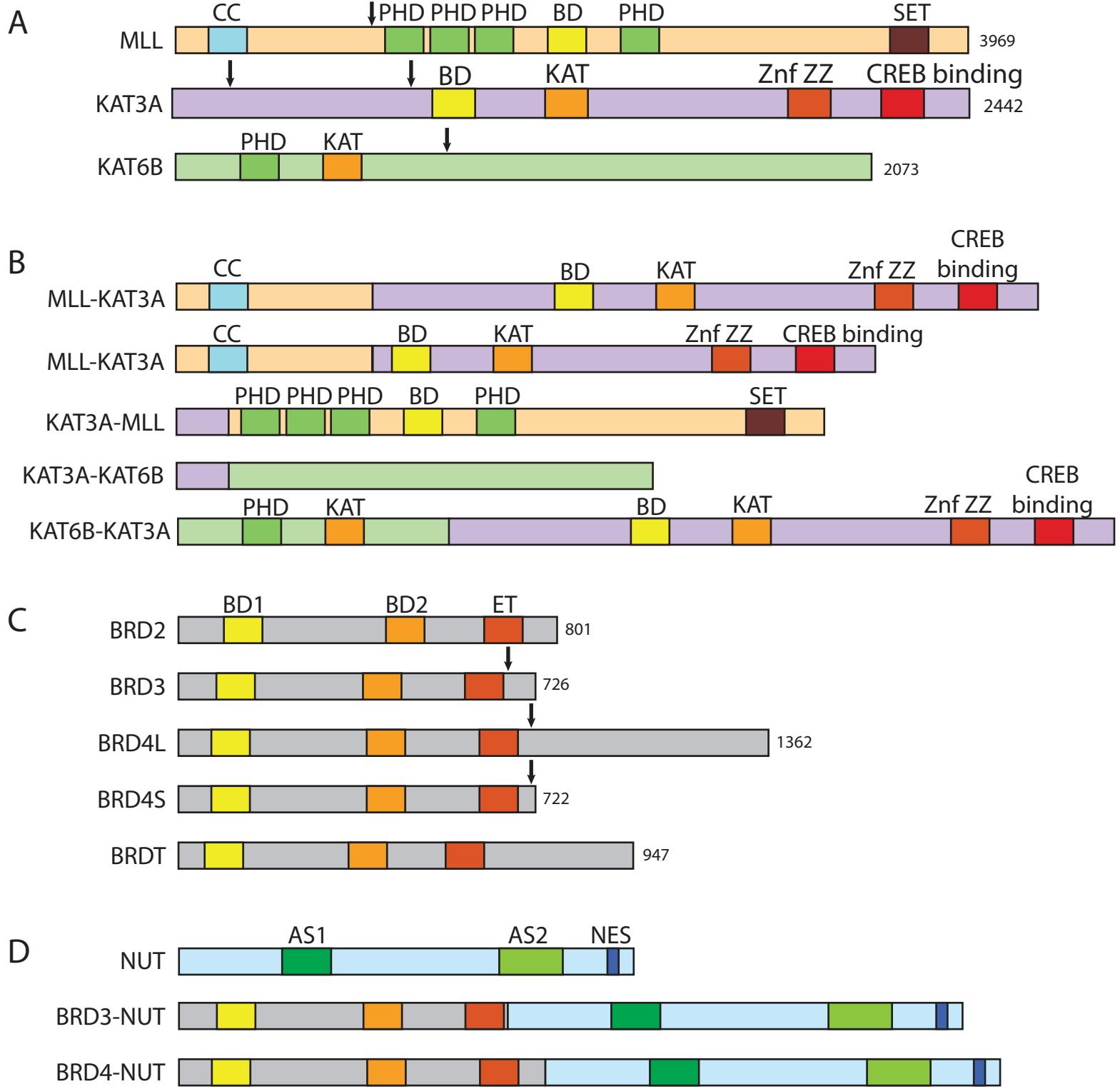
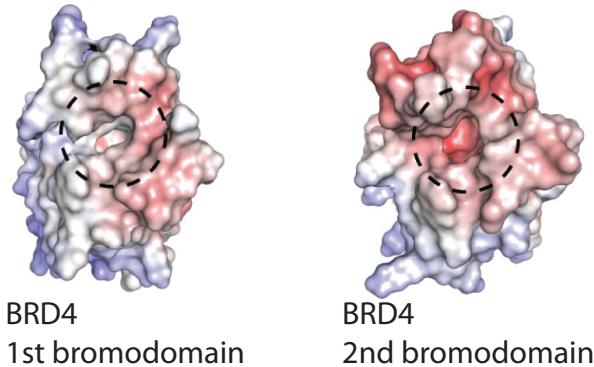


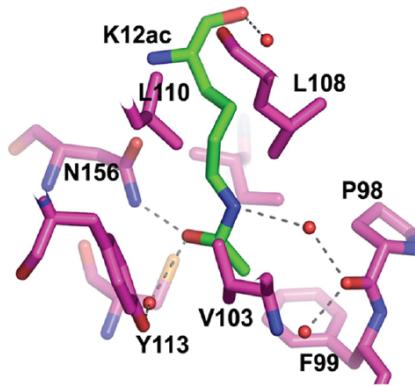
Figure 2. Selected translocations involving acetylome components.

A) MLL, KAT3A, KAT6B wild type domain organization (domain names are defined in Fig 1); common translocation sites are shown by arrows. **B)** Selected fusion proteins involving KAT3A (CREBBP). **C)** BET family of bromodomain-containing proteins. Four genes encode the BET family members in humans (BRD4 is represented by two major splice variants, the “long” BRD4L and the “short” BRD4S). BD1 and BD2 indicate the first and second bromodomains, respectively. ET is the extraterminal domain. **D)** NUT protein and BRD-NUT translocations. *Top*: Architecture of the NUT protein; though no structural domains have been identified in NUT, two acidic stretches (AS1 and AS2) are responsible for mediating protein-protein interactions. A nuclear export sequence (NES) is indicated; putative NLS sequences are detected in NUT, BRD3 and BRD4, but not displayed here. *Middle and bottom*: structural organization of the BRD4-NUT and BRD3-NUT fusion proteins which are causally implicated in NMC. Though these fusion stem from a reciprocal translocation, the reverse product (NUT-BRD3 or NUT-BRD4) is not expressed. Adapted from Muller et al., 2012.

A

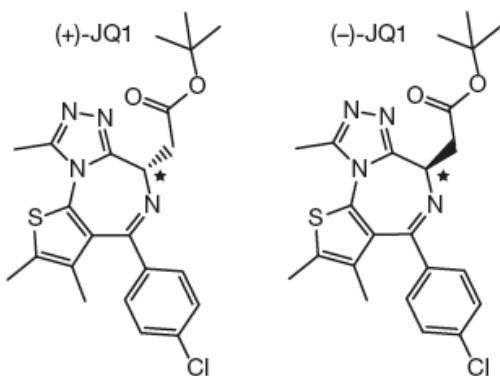


B

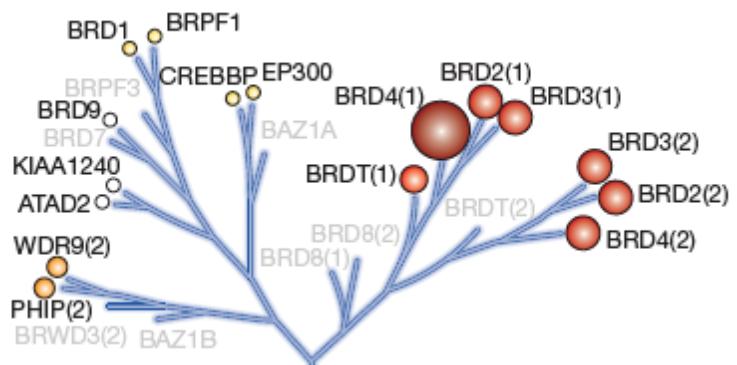
**Figure 3. Structural analysis of bromodomains.**

A) Structures of the first and second bromodomains from BRD4 (Filippakopoulos et al., *Cell*, in press; Appendix 1). Surface electrostatic potentials are displayed. The dashed circles indicates the location of the acetyl-lysine binding pocket. **B)** Stick diagram of the interactions established between a bromodomain (here BRD2) and an acetylated lysine (from histone H4); Asn156 interacts with the acetylated lysine, and its mutation to Phe or Ala drastically reduces the interaction (Umeshara et al.; *J Biol Chem*, 2010). Here, the bromodomain mutants will be generated by mutation of the conserved Asn.

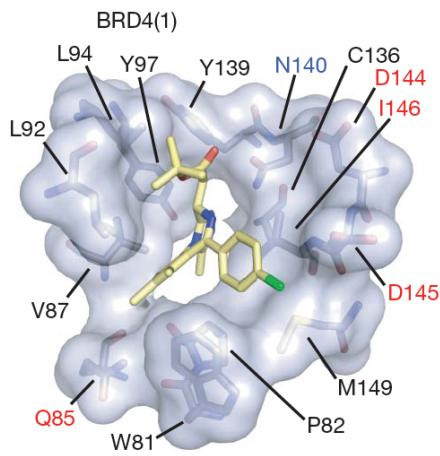
A



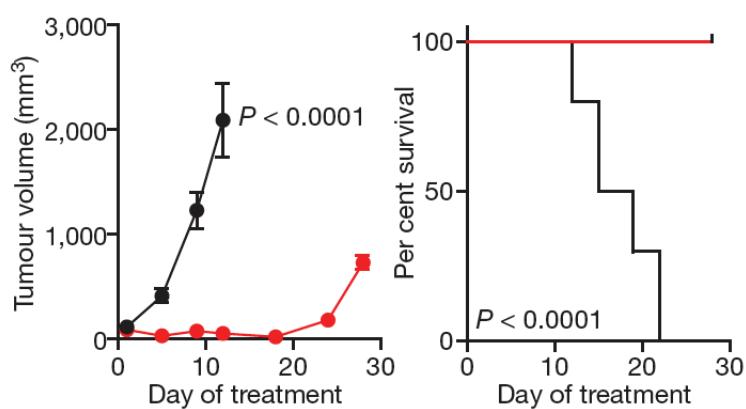
B



C



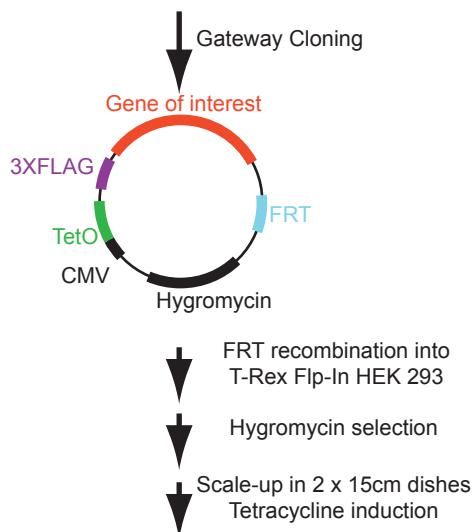
D

**Figure 4. The bromodomain as a therapeutic target: development of the JQ1 inhibitor.**

A) Structure of the active (+)-JQ1 inhibitor and its inactive (-)-JQ1 stereoisomer. **B)** Selectivity of JQ1 for the BET family of bromodomains as assessed by differential scanning fluorometry (the size and color of the circles is proportional to the temperature shifts; only a selected subset of the bromodomains tested are displayed here – the 21 other ones were not significantly affected). **C)** The JQ1 inhibitor (yellow) binds in the KAc pocket on the BET bromodomain (Here, the first bromodomain of BRD4; grey) structure. **D)** The JQ1 inhibitor (red circles and lines; 50mg/kg daily for 18 days) is effective at preventing tumor growth (left) and to prolong survival (right) of mice engrafted with primary cells derived from a patient with NUT midline carcinoma (NMC). Similar effects, including tumor regression, were observed with three independent NMC models. All panels are from Filippakopoulos et al., *Nature*, 2010.

A

44 Bromodomain containing proteins
18 Lysine acetyltransferases (KAT)
11 Lysine deacetylases (KDAC)
7 Sirtuins



B

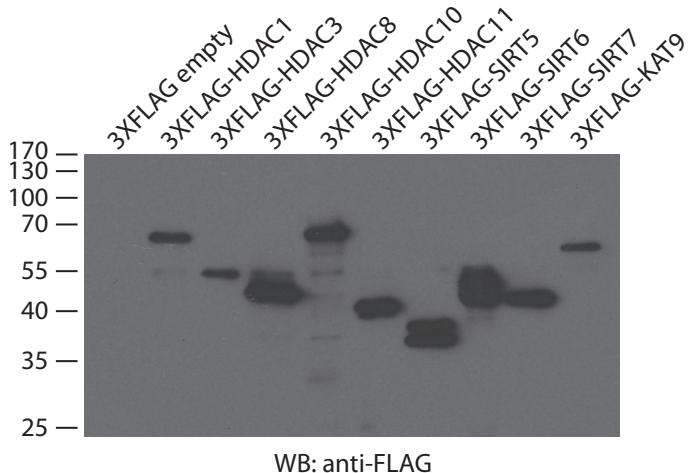


Figure 5. Strategy for the establishment of stable cell lines expressing acetylome components.

A) Each of the proteins implicated in the KAc mechanism is cloned in a Gateway-compatible vector downstream of a 3xFLAG epitope. The vector is derived from the Invitrogen Flp-In T-Rex system, enabling establishment of isogenic cell lines in which expression of the target protein is under the control of a tetracycline inducible promoter. For AP-MS analysis, 2 plates of cells (15cm) are used per analysis. **B**) Example of an immunoblot for different stable cell lines for this project. Note that – as we and others reported previously – the expression levels of the recombinant proteins is usually not excessive as compared to the endogenous proteins (see, e.g. Figure 17). This is important to prevent artefactual detection of interactions.

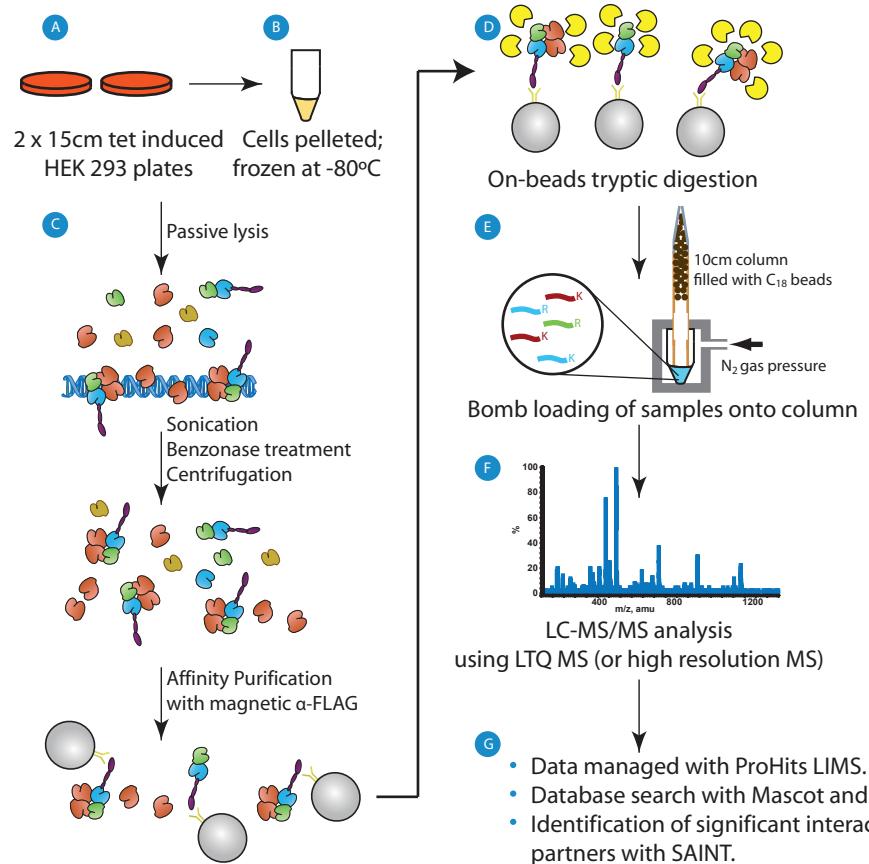


Figure 6. Schematic of the AP-MS protocol, mChIP, used for the interactome mapping.

Cells are generated as in Figure 5 (**A**), washed in PBS, and pelleted (**B**). Lysates are prepared by passive lysis in a buffer containing low amounts of detergents, followed by shearing of DNA by sonication and further nuclease digestion with benzonase; affinity purification is performed using magnetic beads coated with anti-FLAG antibody (**C**). Trypsin digestion is performed directly on the magnetic beads (**D**), and the peptides are acidified before being loaded using a pressure cell on home-packed reversed-phase capillary columns (**E**). The column is attached to an HPLC system and placed in line with a mass spectrometer (**F**) for mass spectrometry acquisition. Data analysis is performed to identify all proteins in the sample, and high confidence interactors (**G**; see Figure 8 for details).

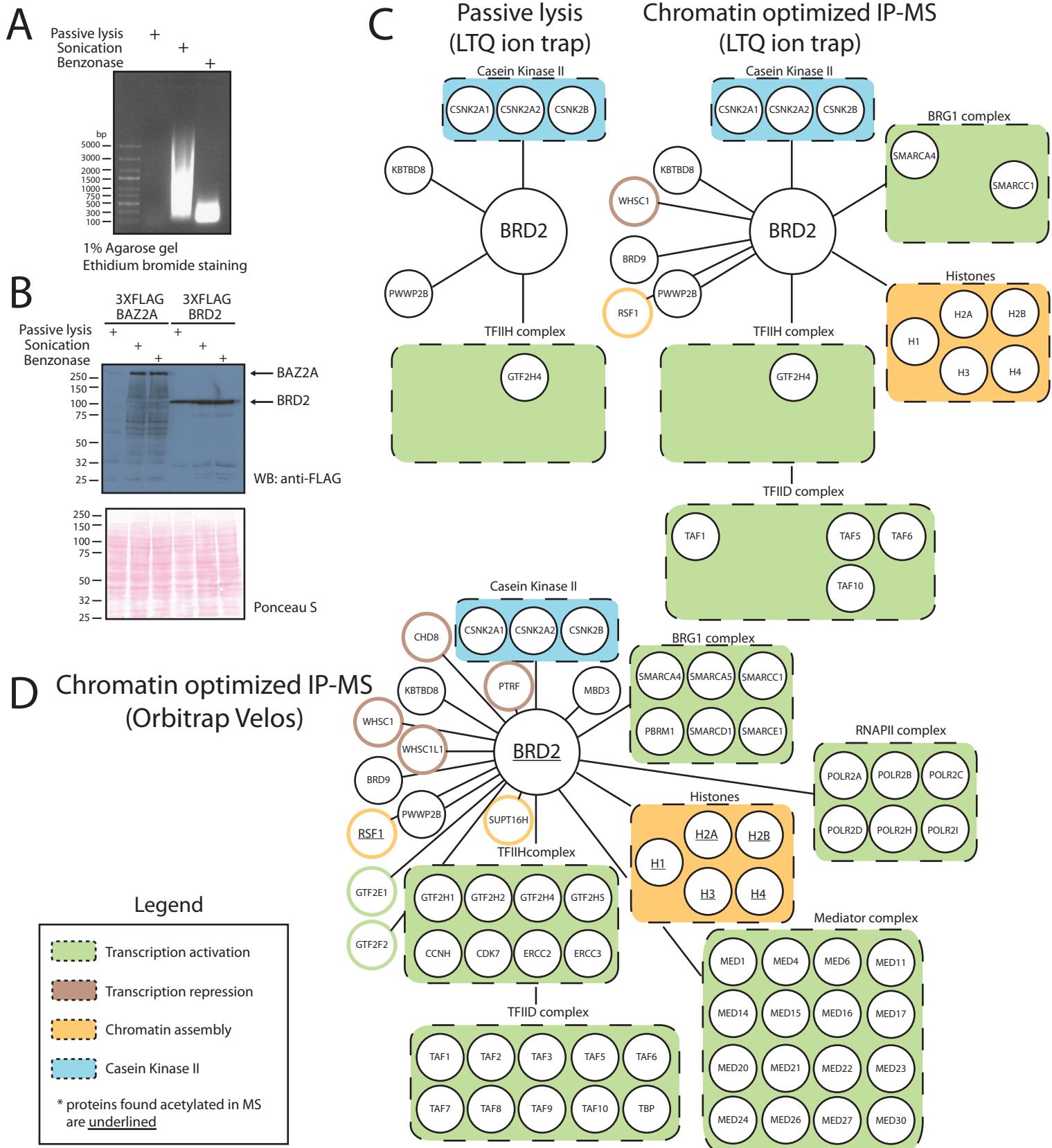
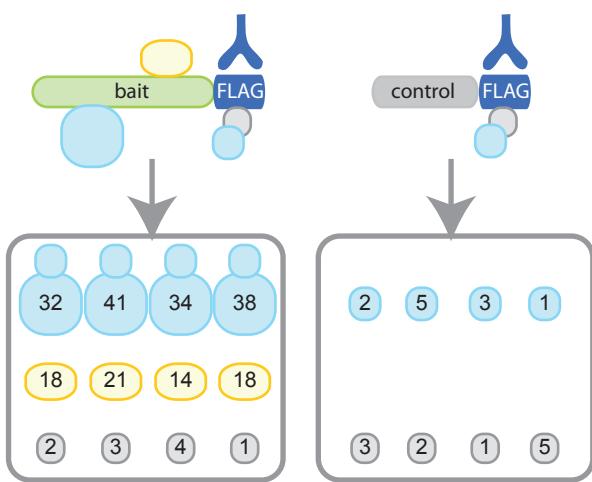


Figure 7. Use of mChIP for the acetylome project.

A) Sonication and benzonase treatment lead to fragmentation of the chromatin, detected here by DNA fragmentation; our optimized protocol employs both sonication and benzonase treatment to ensure reproducibility. **B**) Comparison of the solubility of two bromodomain-containing proteins, BAZ2A and BRD2 using the standard passive lysis protocol and the same protocol after addition of a sonication or benzonase step; note that BAZ2A is completely unsoluble unless chromatin is fragmented. BRD2 is partially soluble, but its recovery is increased by sonication or benzonase treatment. **C**) Comparison of the AP-MS results for BRD2 with the standard passive lysis protocol and with the mChIP protocol: note the detection of new interactors, including histones. **D**) Faster mass spectrometers (here the Orbitrap Velos) lead to the identification of more proteins from the sample shown in **C**. This is because mass spectrometers typically operate in a data-dependent manner in which only the most abundant co-eluting peptides are sequenced; a faster mass spectrometer means that more co-eluting peptides have the time to be sequenced, leading to a more sensitive detection.

A



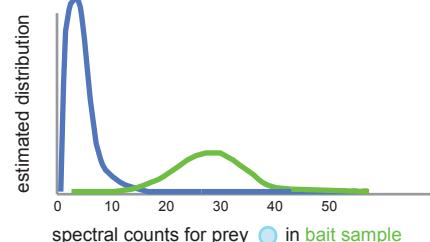
B

Simple frequency filter (or background removal)

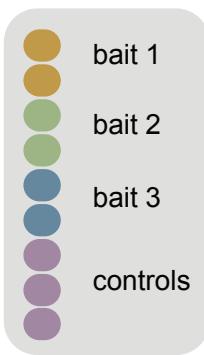
- correctly identified as a contaminant
- correctly identified as an interactor
- incorrectly filtered out

C

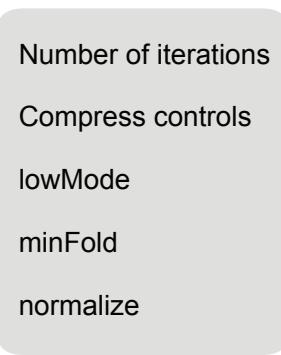
SAINT filtering



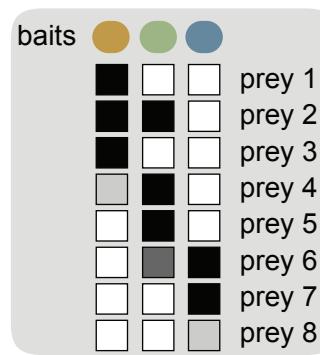
D



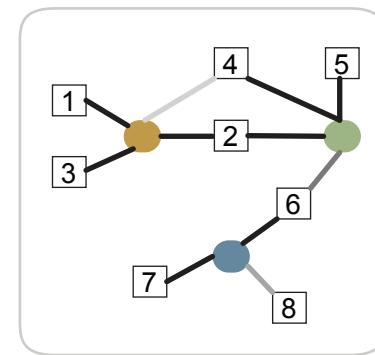
Select Samples



Specify SAINT options



Filter SAINT results

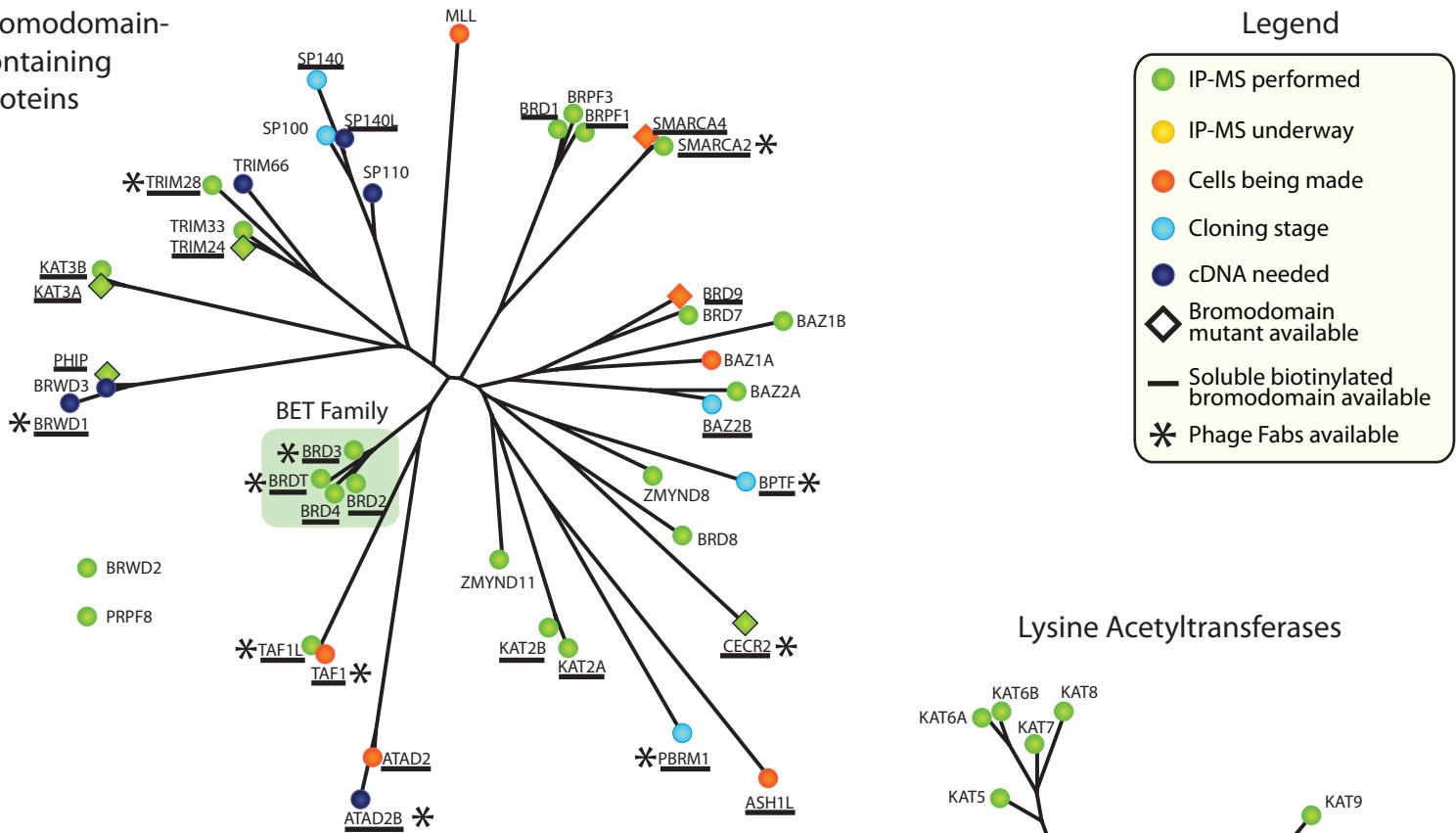


Export and visualize

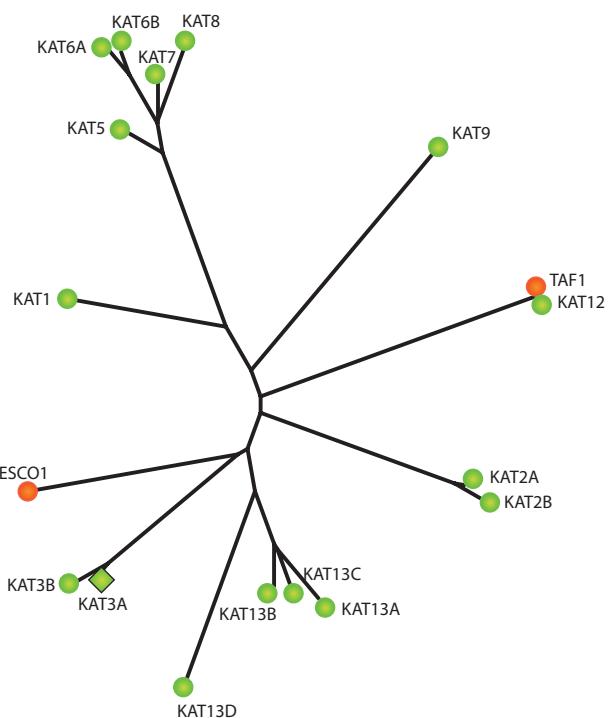
Figure 8. Scoring of high confidence interactions with SAINT.

A) Quantitative information is embedded within each mass spectrometry experiment. Here four independent purifications of a bait and four parallel purifications of a negative control protein are depicted. The yellow protein is only detected with the bait, and never with the control. The grey protein is detected with both the control and the bait, and its relative abundance (depicted here by the size of the circle and associated numbers) does not vary significantly between samples. The blue protein is different: it is found associated with the negative control, but is present in much larger amounts in each of the bait purifications. **B)** Often the only criterion to accept a protein as an interactor is that it is never seen in the controls; while this is logical, it leads to both underfiltering (typically when there are not enough controls in the analysis) and overfiltering (like with the blue protein here). **C)** SAINT uses the quantitative information in each of the mass spectrometry experiments to model the distribution of the spectral counts (semi-supervised algorithm) and provide a probability value for each of the interactions. **D)** SAINT has been fully implemented within our ProHits LIMS for interaction proteomics, enabling the user to rapidly select samples, specify appropriate SAINT scoring options based on the type of data at hand, and export and visualize the data. Panels A-C are adapted from Skarra et al., *Proteomics*, 2011; D is from a submitted manuscript by Liu et al.

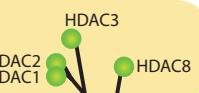
Bromodomain-containing proteins



Lysine Acetyltransferases



KDAC class I

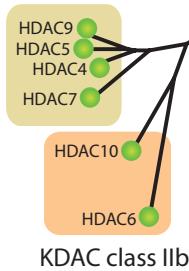


Lysine Deacetylases

KDAC class IV



KDAC class IIa



KDAC class III

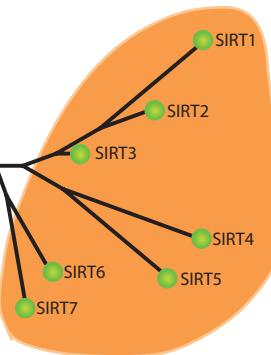
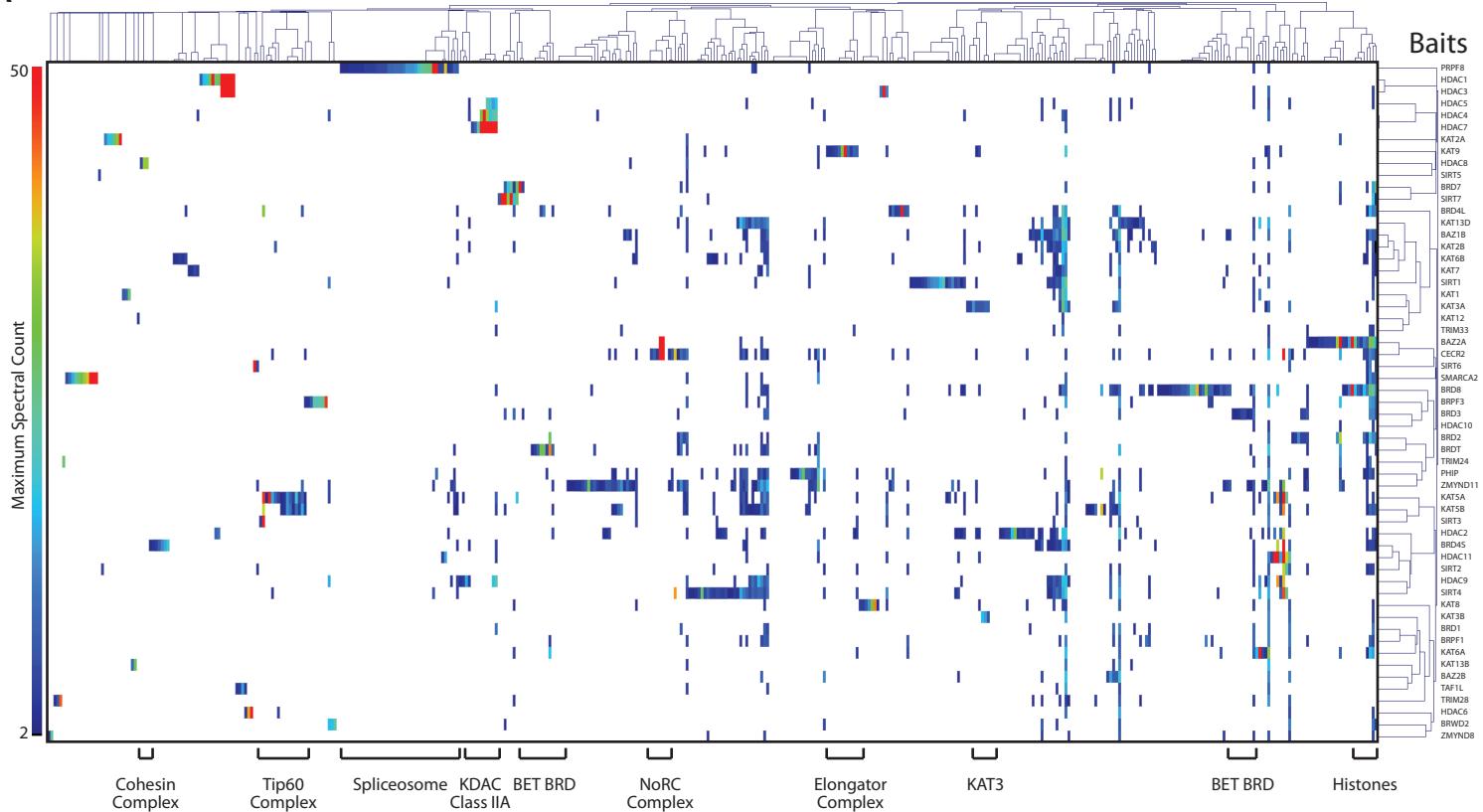


Figure 9. Current project status.

All bromodomain-containing proteins, KDACs and KATs are organized on sequence distance trees (based on the full-length proteins). Gene names are indicated, alongside with the progress of the cloning, expression and AP-MS analysis of the full-length, wild type proteins. Several point mutants in the bromodomains within the context of the full-length proteins have already been generated (diamond nodes), alongside numerous soluble and biotinylated isolated bromodomains which will be employed in pull-down analyses (underlined). Lastly, our collaborator Sachdev Sidhu has been systematically generating phage affinity reagents for bromodomains; the current status of all these reagents is indicated by asterisks. Note that PRPF8 and BRWD2 were previously reported to contain bromodomains, though this is controversial; they were included on this project for completeness.

A

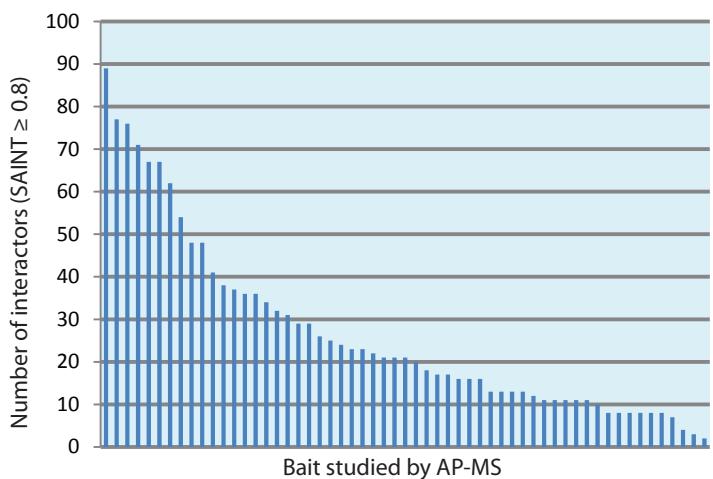
Associated Proteins



B

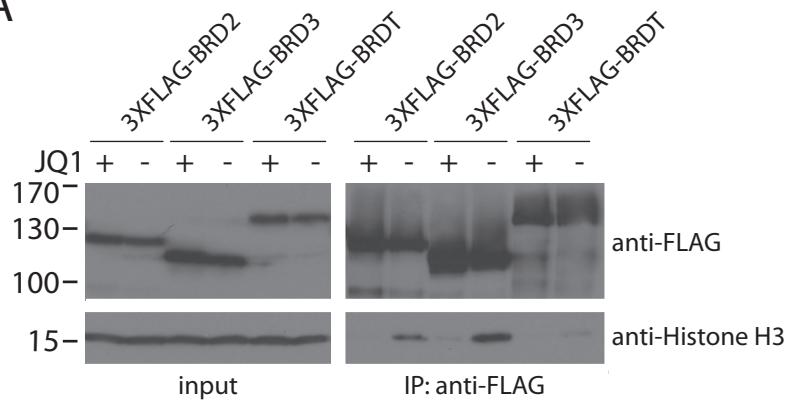
Number of baits	57
Number of LC-MS/MS runs	120
Number of negative controls	15
Unique interactions unfiltered	23298
Proteins unfiltered	2076
Interactions at SAINT ≥ 0.8	1521
Proteins at SAINT ≥ 0.8	446

C

**Figure 10. The interactome of the KAc machinery at a glance.**

A) Cluster analysis of the interactions detected with high-confidence SAINT score AvgP ≥ 0.8 ; baits are shown on the vertical and identified proteins on the horizontal axis. Selected complexes are labeled. **B)** Composition of the dataset so far. **C)** Number of interactors detected for each of the baits in the network. Most of the interactions were previously unreported.

A



B

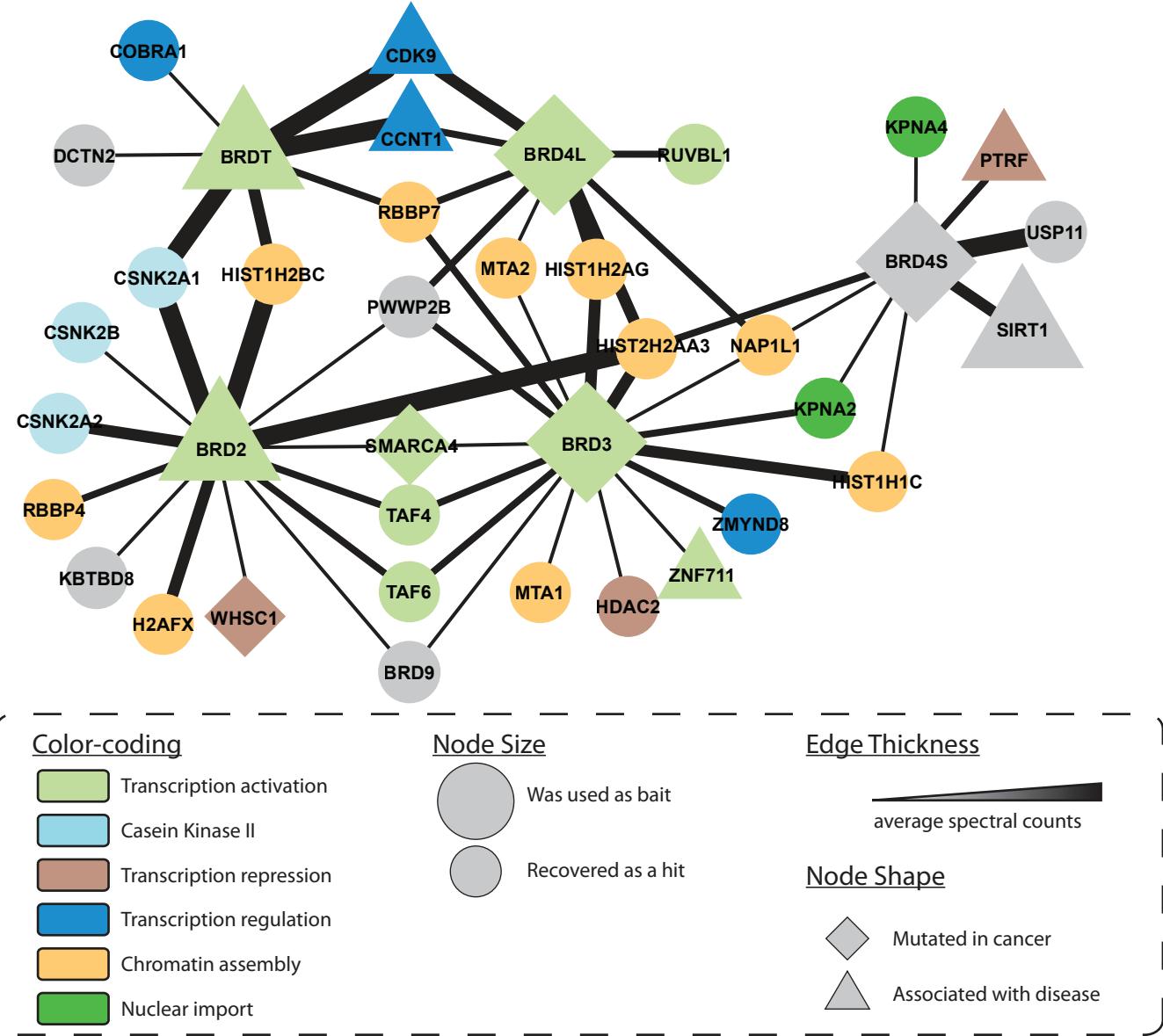


Figure 11. Interaction network for the BET family of bromodomain-containing proteins.

A) Members of the BET family of bromodomain-containing proteins BRD2, BRD3 and BRDT co-purify core histones (Histone H3 is shown here). Pretreatment with the active (+)-JQ1 BET inhibitor abolishes this interaction, demonstrating that the interaction depends on a KAc-bromodomain interaction. **B)** High confidence interaction network of the BET family in the absence of (+)-JQ1 inhibitor; see legend for details.

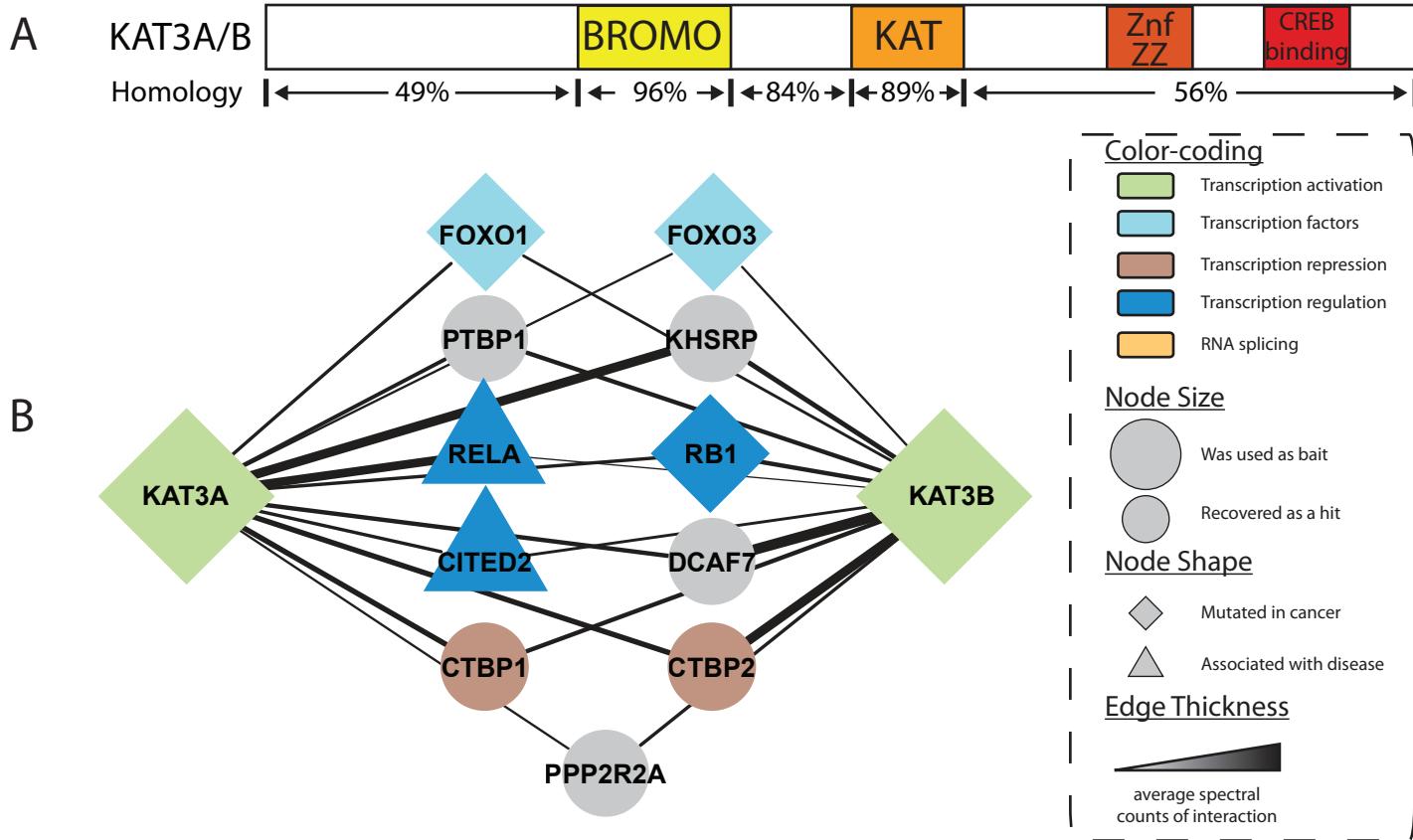


Figure 12. Interaction network of the close paralogues KAT3A / KAT3B .

A) Domain architecture of KAT3A / KAT3B. **B)** Interaction network for the KAT3A and KAT3B proteins reveals multiple key players in cancer.

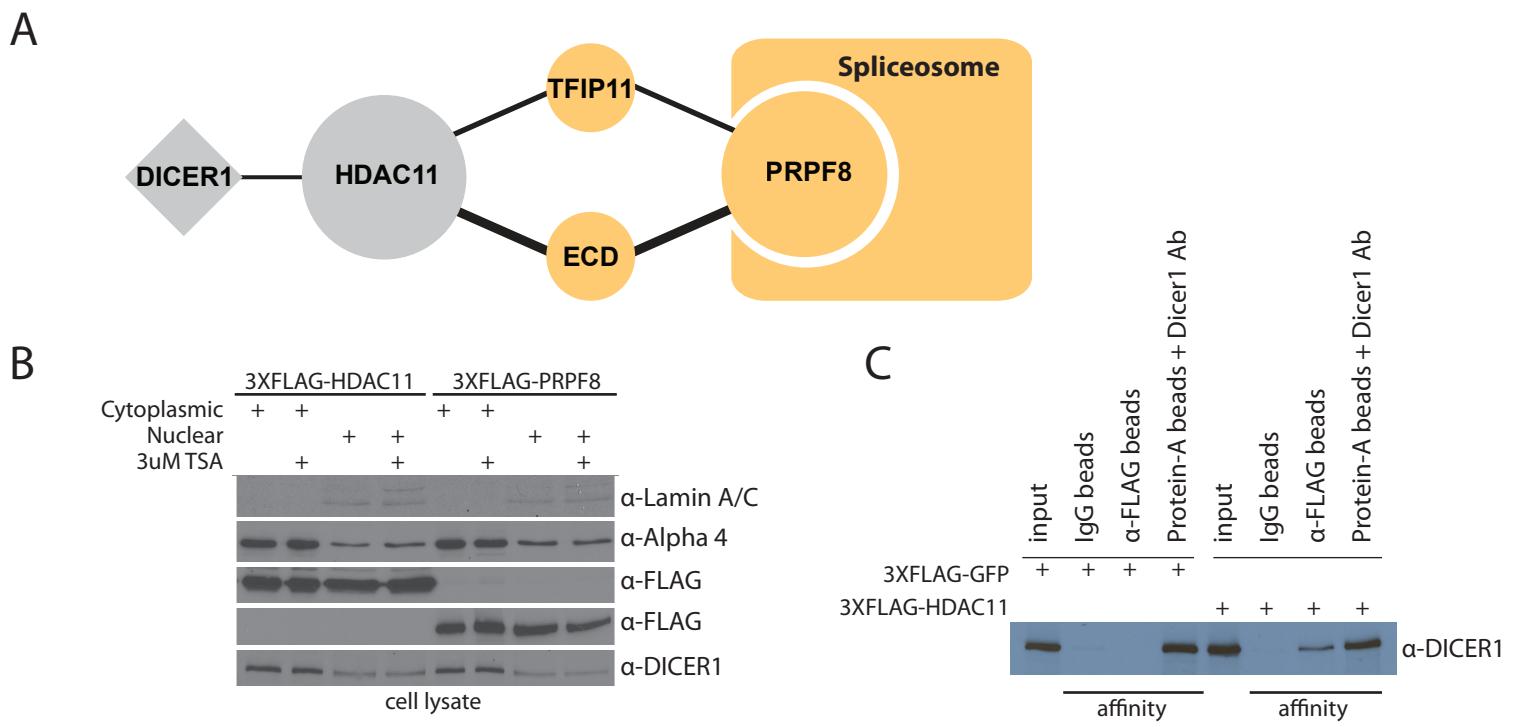
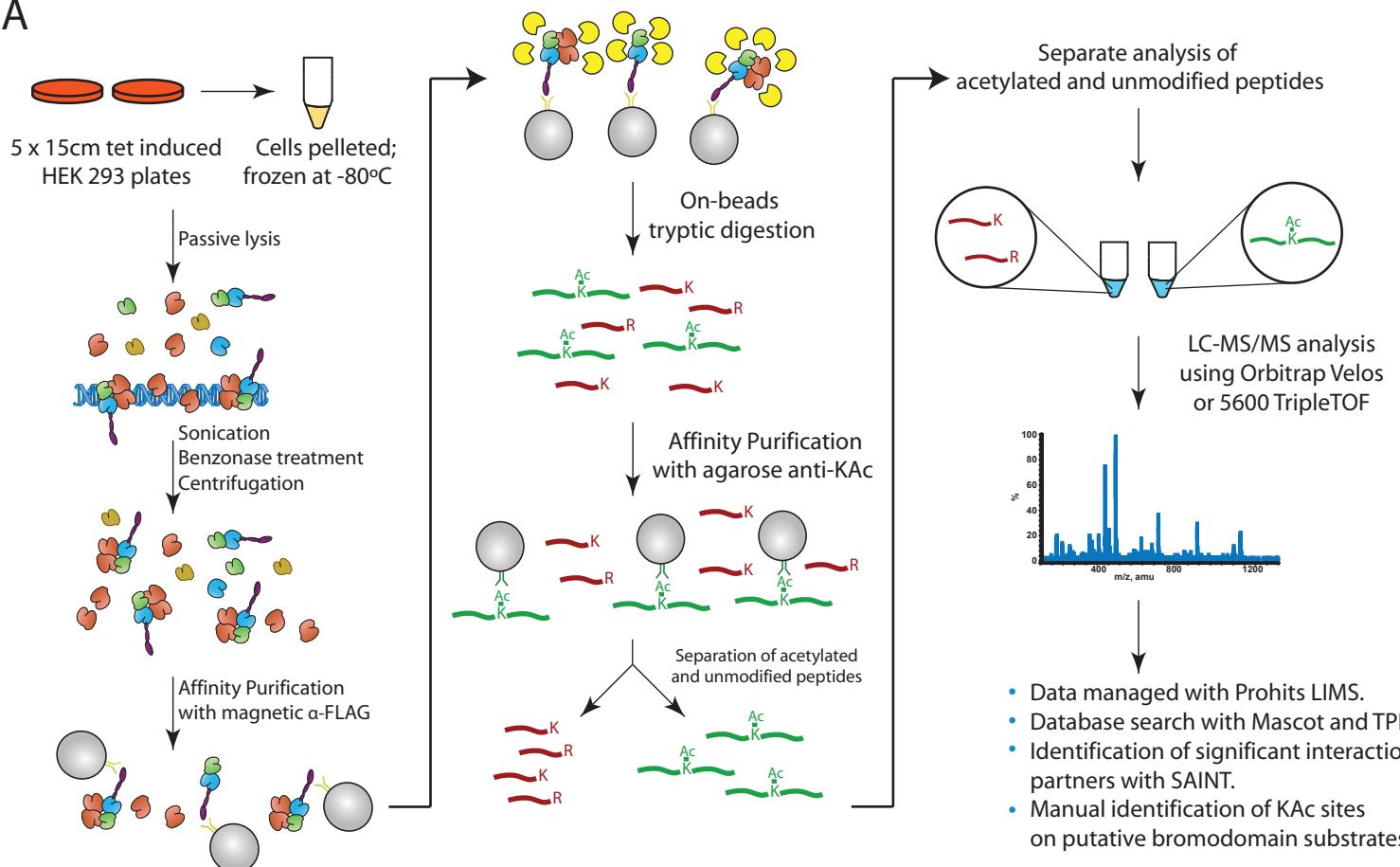


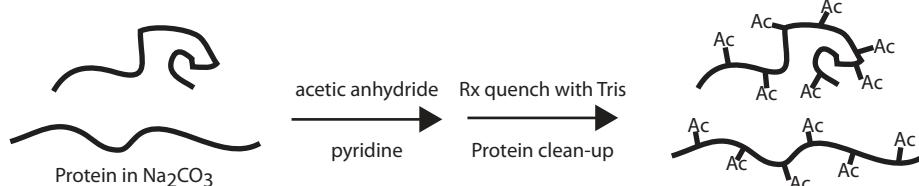
Figure 13. Interaction network of HDAC11.

A) Interaction network of HDAC11 and PRPF8 (see Fig 12 for legend). Affinity purification of HDAC11 revealed a physical association with DICER1 as well as with TFIP11 and ECD, two spliceosome components. This suggest a role for HDAC11 in RNA processing. **B)** 3XFLAG-HDAC11 and DICER1 were found to be present in the cytoplasm using a cell fractionation and immunoblotting approach. Lamin A/C and Alpha 4 are used as markers for nuclear and cytoplasmic fractions, respectively. The broad spectrum HDAC inhibitor trichostatin A (TSA) does not affect localization of DICER1 **C)** The physical interaction between HDAC11 and DICER1 was further validated by performing affinity purification and immunoblotting. The functional implication of the HDAC11-DICER1 interaction is being assessed.

A



B



C

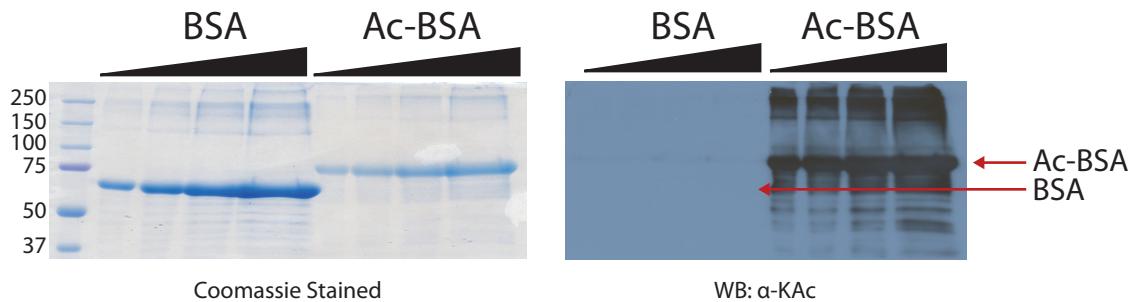
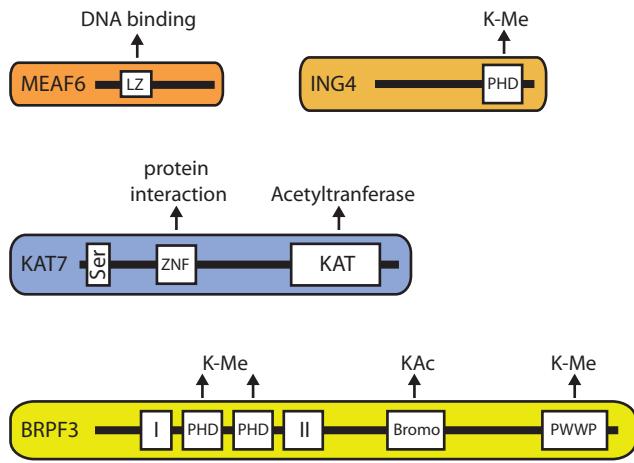


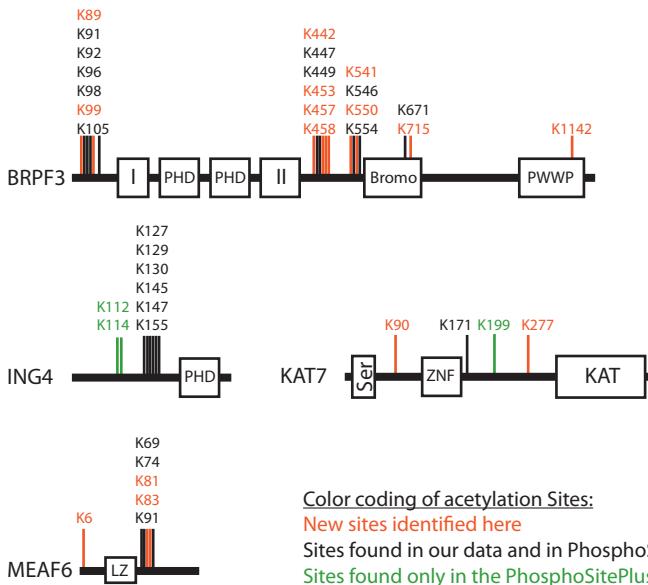
Figure 14. Schematic of the KAc enrichment approach used here and validation of commercial anti-KAc antibody.

A) The mChIP protocol is performed first, using more starting material than for simple protein identification (left). After tryptic digest, an additional immunoprecipitation of KAc peptides is performed using agarose anti-KAc affinity matrix (middle). Bound and unbound fractions are kept and analyzed separately on high mass resolution instruments to sensitively define the interaction partners of a given acetylome components and their acetylation status. To validate the anti-KAc antibodies, bovine serum albumin was acetylated *in vitro* (**B**) and utilized in Western blots (**C**). Note that the commercial antibody recognizes acetylated BSA effectively but not its unmodified form.

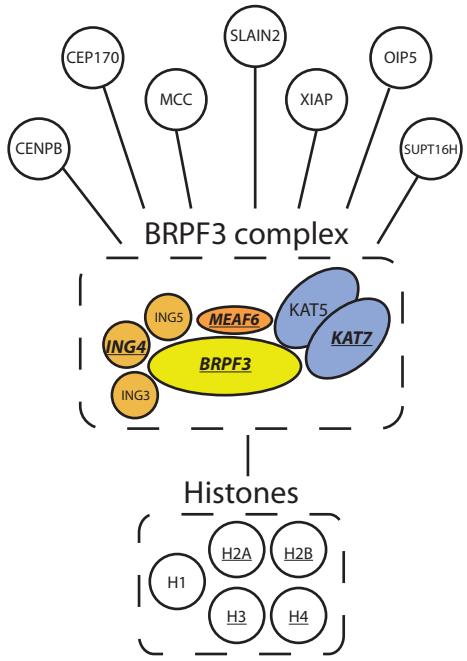
A



C



B



D

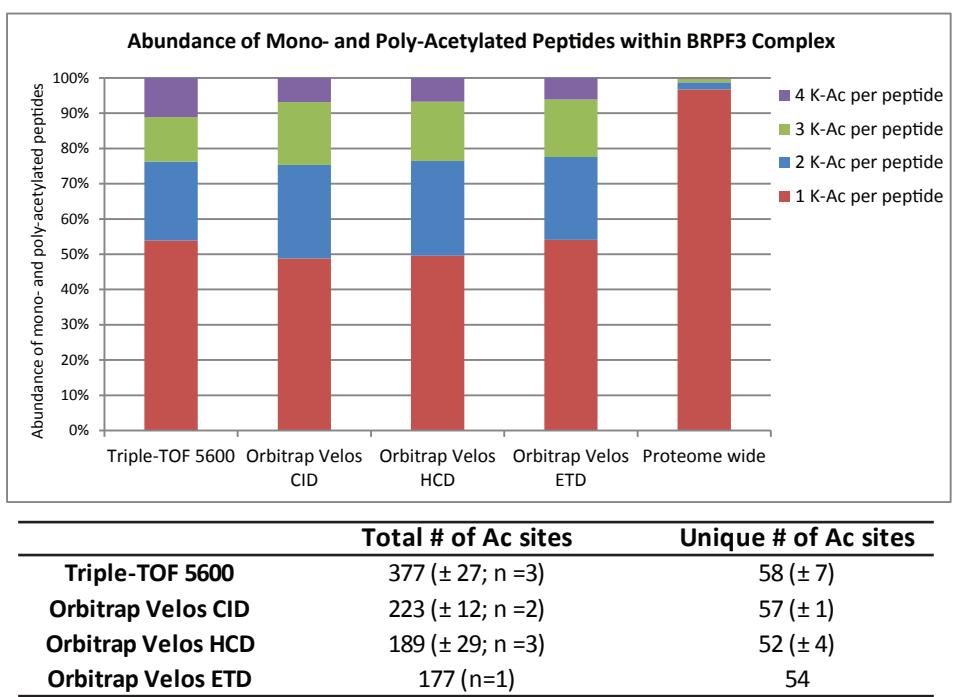


Figure 15. Benchmarking the KAc enrichment protocol on BRPF3.

A) Composition and structural organization of BRPF3 and its known interactors; note the presence of several modular domains that bind histone marks. **B)** BRPF3 complex and interactions as determined here; acetylated proteins (as determined by our mass spectrometry experiments) are underlined. ***Bold italics*** indicate the major components of the BRPF3 complex. **C)** Summary of the acetylated sites on the BRPF3 core components; new sites are shown in red. This study indicates that our approach is sensitive and enables the detection of multiple sites within a single AP-MS experiment. **D)** Comparison of different mass spectrometers and fragmentation types for the identification of KAc sites in the BRPF3 sample. The proportion of mono, di, tri and quadri-acetylated peptides detected is displayed (*top*), along with a summary of the quantitative data (*bottom*). Note that – as compared to a proteome wide study (Mann group) which predominantly identified monoacetylated sites – our approach enables efficient recovery and sequencing of multi-acetylated peptides. This is important since we have showed (e.g. Figure 16) that bromodomains effectively recognize combinations of marks. CID, collision induced dissociation; HCD, higher energy C-trap dissociation; ETD, electron transfer dissociation. The TripleTOF 5600 was selected for further studies.

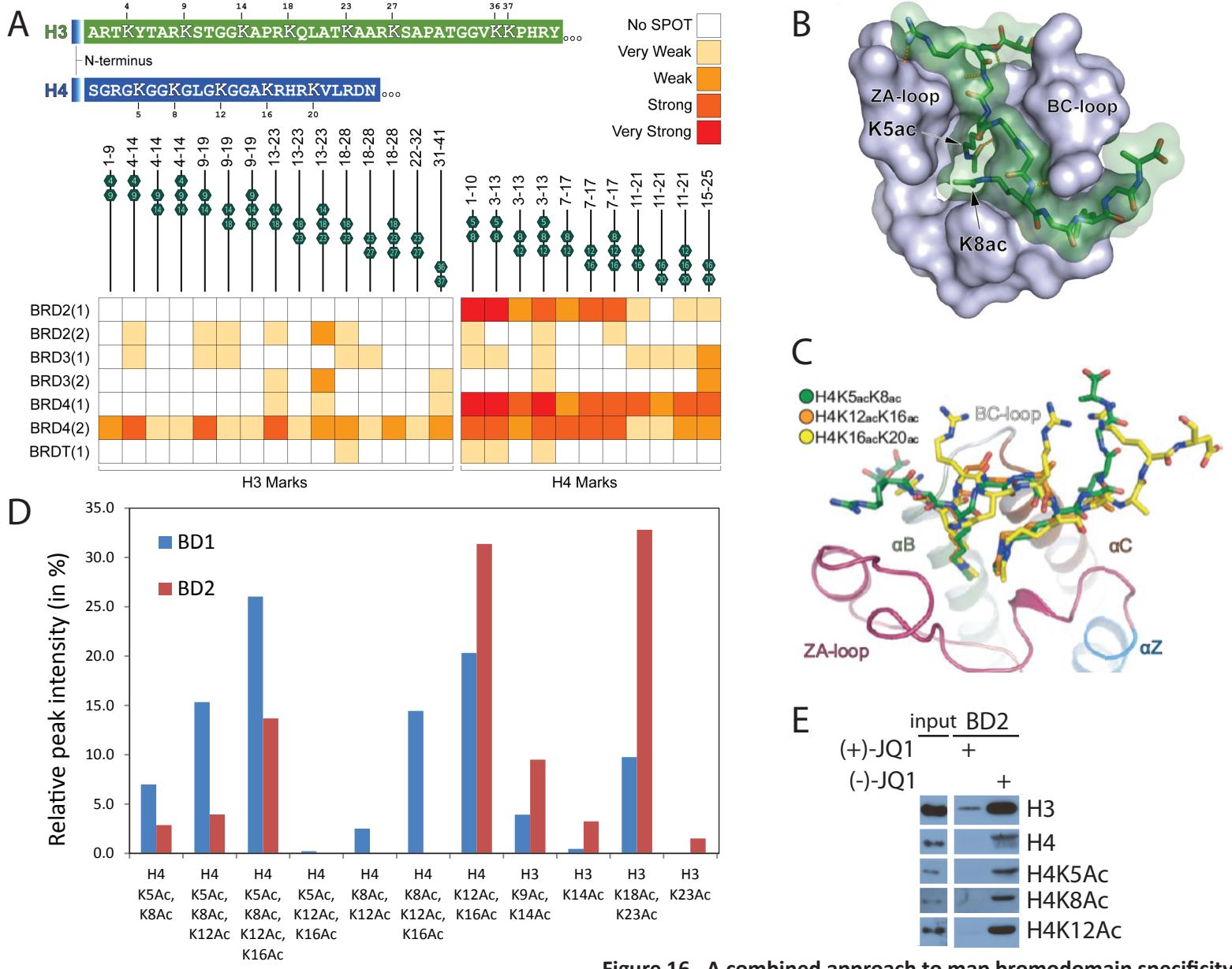
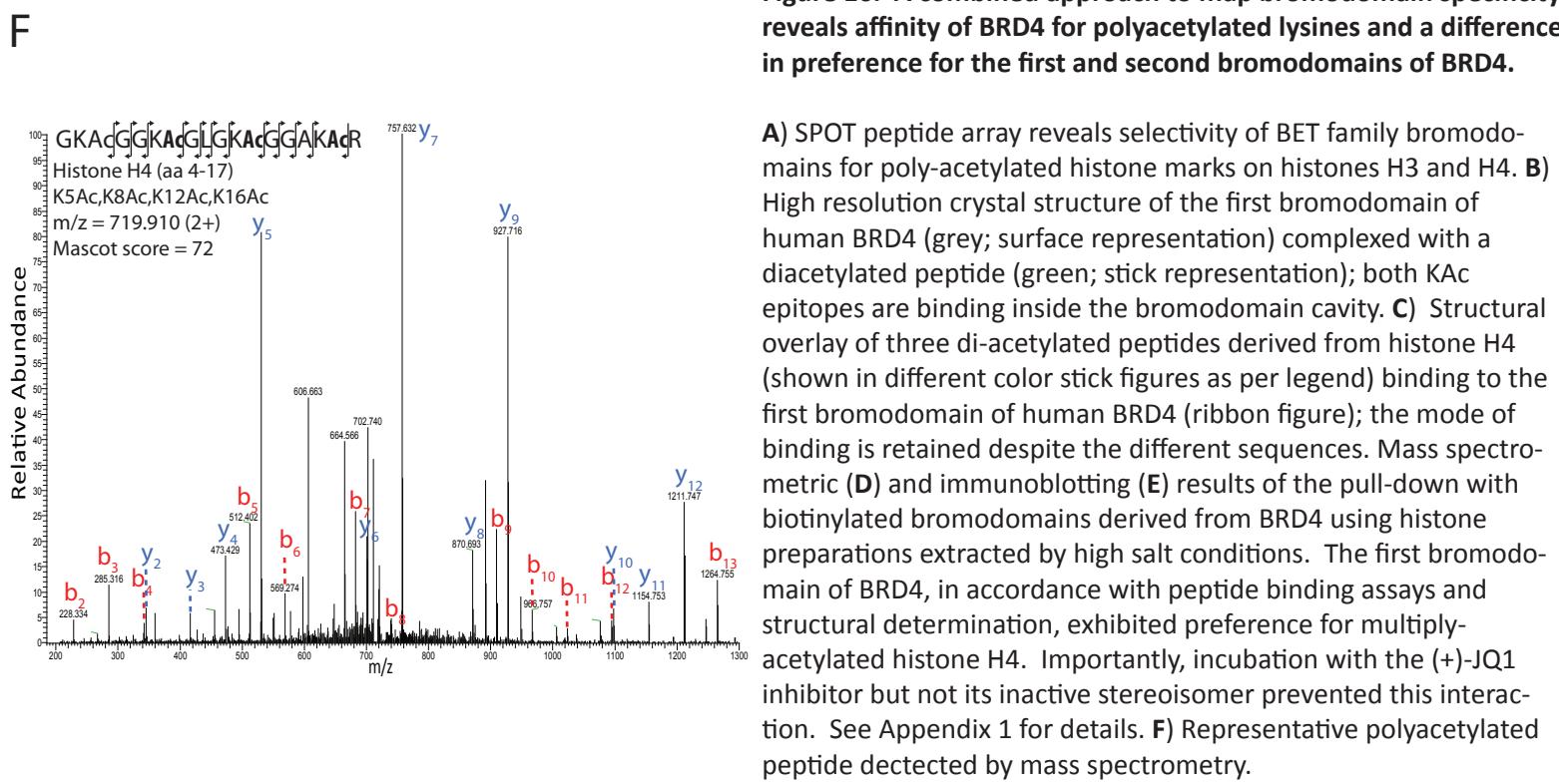
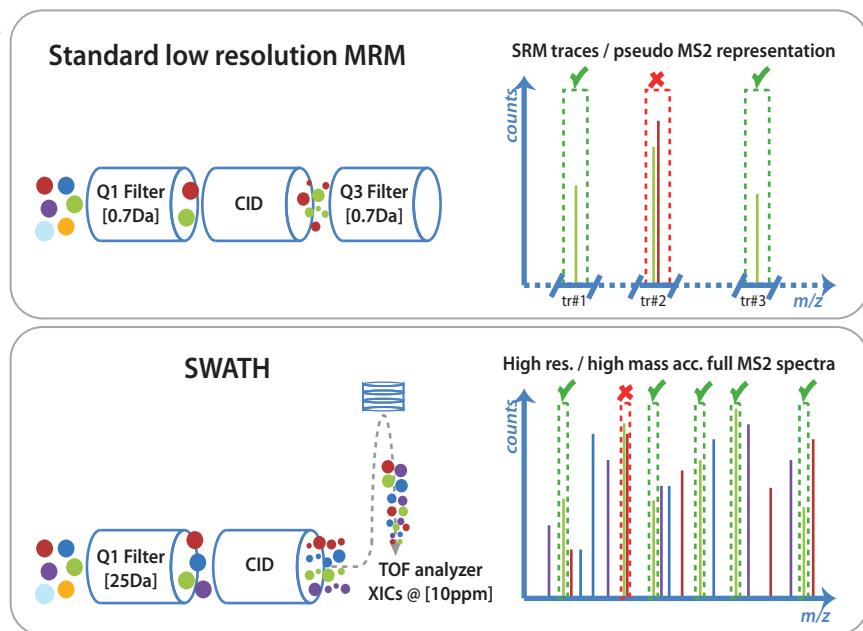


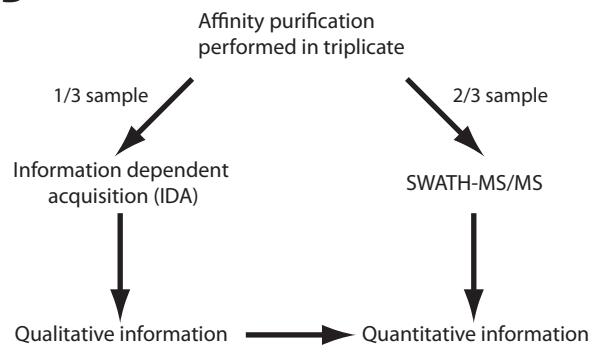
Figure 16. A combined approach to map bromodomain specificity reveals affinity of BRD4 for polyacetylated lysines and a difference in preference for the first and second bromodomains of BRD4.



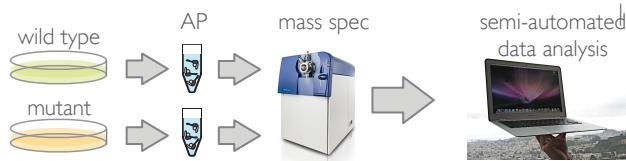
A



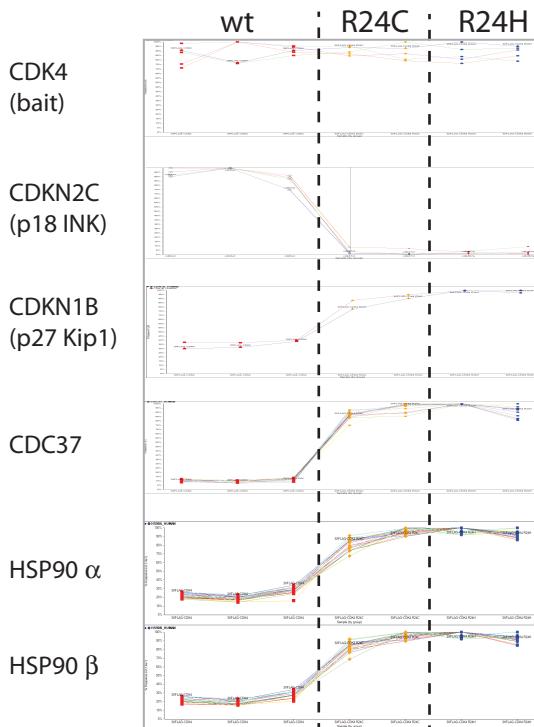
B



C



D



E

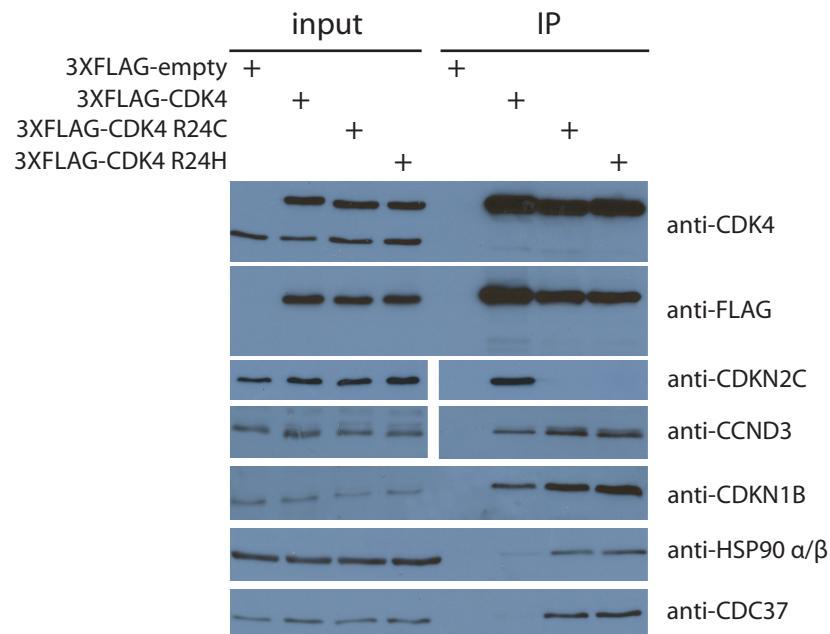


Figure 17. Quantitative mass spectrometry approaches using MS2 quantification for the analysis of affinity-purified samples.

A) The SWATH methodology is related to standard MRM (or SRM): In a first step, ions of a given mass are let through a filter, and then fragmented by CID. Masses of the fragment ions are recorded and quantification is performed. In standard MRM, the masses of the precursor ions to isolate are predetermined, and the isolation of the precursor and the detection of the product ions are at low resolution (~1 Da). By contrast, SWATH-MS does not require predetermination of the precursor masses: the entire m/z range is progressively fragmented in 25 amu "swaths"; the loss of accuracy in peptide isolation is compensated by ppm level precision in MS2.

B) Because the spectra associated with SWATH-MS data are most often mixed, a spectral library search strategy needs to be employed. For AP-MS samples, we generate the spectral library using standard data-dependent acquisition from the same samples that would be quantified by SWATH-MS and on the same instrument. C) Experimental design for AP-SWATH. Cell lines expressing similar levels of a wild-type and mutant bait proteins are processed in parallel by affinity purification and mass spectrometry; several replicates are analyzed for each sample. The data is analyzed using a semi-automated pipeline, as described in Bisson et al., Nature Biotech, 2011. D) Results of AP-SWATH for the bait CDK4 and two dominant point mutants implicated in melanoma; three replicates of the wild type and two replicates of each of the two point mutants were analyzed. Extracted peptide intensity for selected prey proteins shows a tight correlation (all intensities – on the vertical axis - are within a narrow range of each other). This indicates that the AP-SWATH approach can – like standard MRM/SRM – provide accurate peptide-level quantification. Normalization to the averaged bait levels in the wt runs was applied to all samples. E) Validation of the AP-SWATH quantification results by IP/Western. As previously reported, CDKN2C (also known as p18 INK) is unable to associate – and hence to repress – the CDK4 mutants. By contrast, other proteins, and in particular HSP90 and its kinase-specific co-chaperone CDC37 associate preferentially with the mutants.

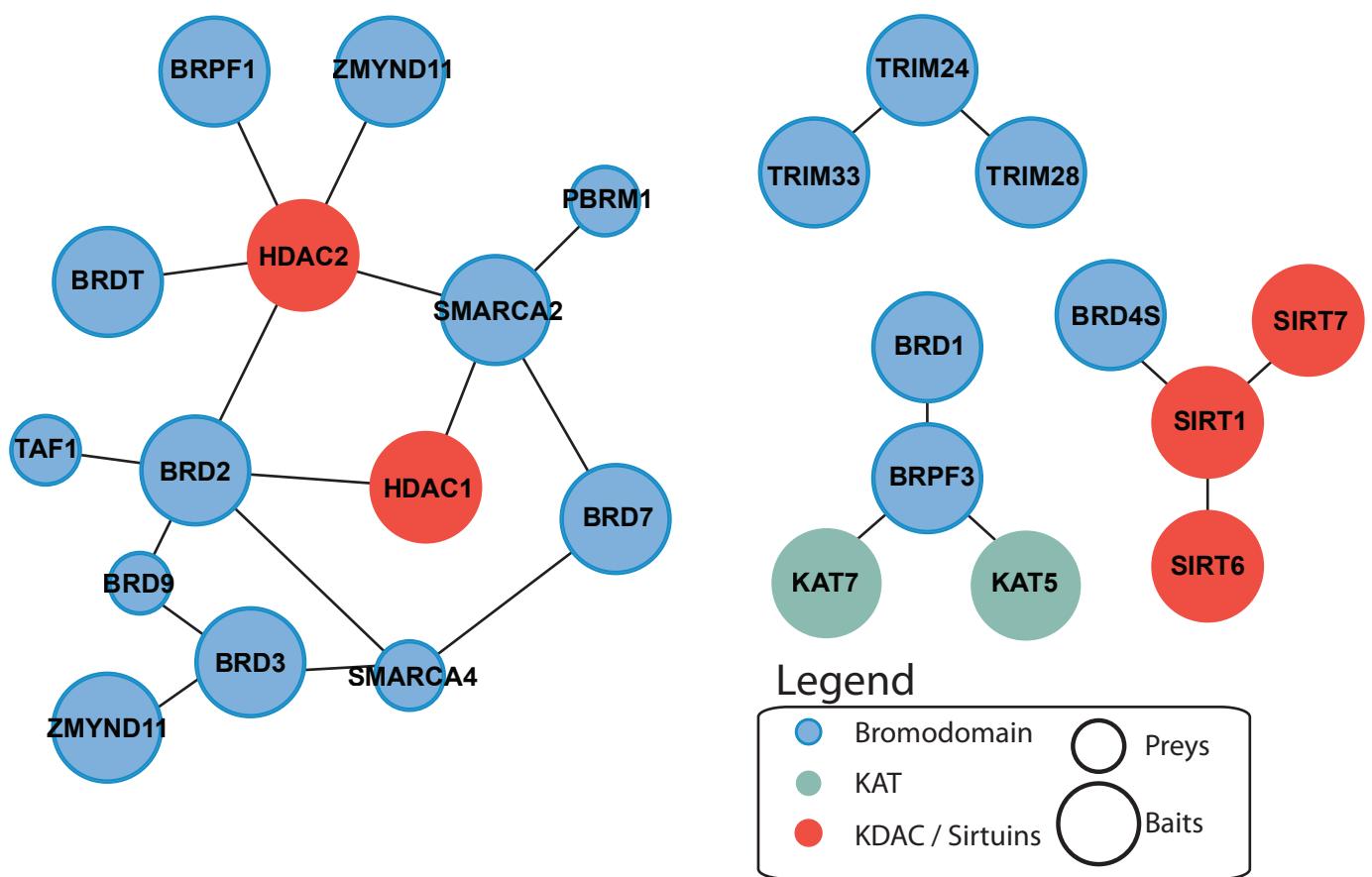
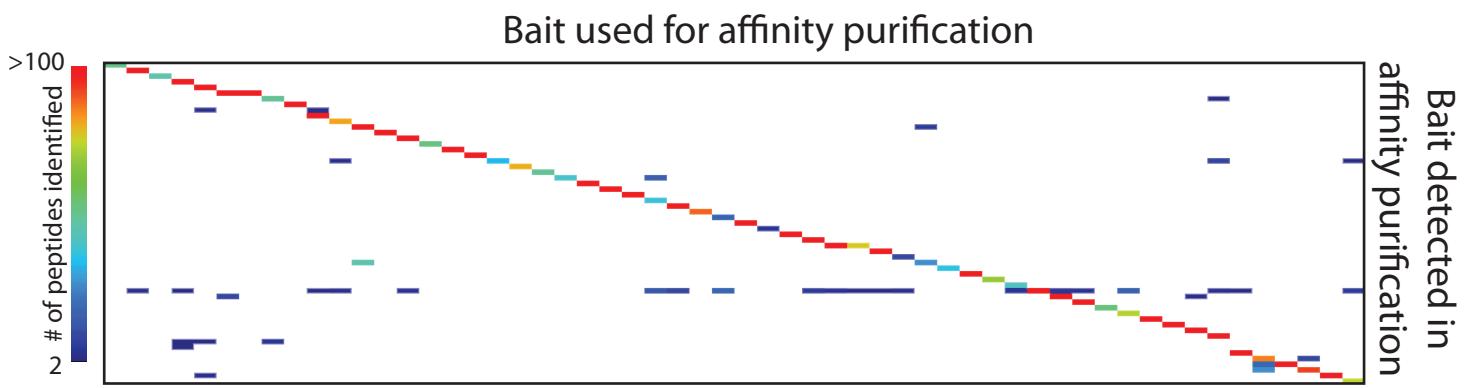


Figure 18. Interconnections between the components of the acetylome.

Several of the acetylome components were recovered as high confidence interactors of other acetylome proteins, suggesting possible crosstalks in the KAc modification mechanism (*top*). Connections detected so far are depicted in a network diagram (*bottom*).

Table I: Acetylome components with potential oncogenic functions

Gene	Other Aliases	# of BRD	Function	Cancer Type	Alteration in cancer
Bromodomain					
ASH1L	KMT2H	1	Methyltransferase		
ATAD2	ANCCA	1	ATPase, coactivator	breast	Overexpression
ATAD2B		1	ATPase, coactivator		
BAZ1A		1	Chromatin assembly and remodeling		
BAZ1B		1	Chromatin assembly and remodeling		
BAZ2A		1	unknown	lymphoblastic leukemia	Translocation
BAZ2B		1	unknown		
BPTF	FALZ	1	Chromatin remodeling	neuroblastoma, leukemia	Amplification
BRD1	BRPF2	1	subunit of MYST KAT containing complex		
BRD2		2	Transcription factor		
BRD3		2	Transcription factor	NUT Midline carcinoma	Translocation
BRD4		2	Transcription factor	NUT Midline carcinoma	Translocation
BRD7		1	Transcription repressor	breast	Deletion
BRD8		2	Tip60 subunit		
BRD9		1	unknown	NSCLC	Amplification
BRDT		2	Transcription factor		
BRPF1	BR140	1	subunit of MYST containing complex		
BRPF3		1	subunit of MYST containing complex		
BRWD1	WDR9	2	Chromatin remodeling		
BRWD2	WDR11	0*	unknown	glial	Insertion, deletion
BRWD3		1	JAK/STAT signaling	B-CLL	Translocation
CECR2		1	Chromatin remodeling		
KAT2A	GCN5L2	1	KAT		
KAT2B	p/CAF	1	KAT		
KAT3A	CREBBP CBP	1	KAT	ALL, AML, DLBCL, B-NHL	Translocation, overexpression, mutation
KAT3B	EP300 p300	1	KAT	colorectal, breast, pancreatic, AML, ALL, DLBCL	Mutation, overexpression
MLL	KMT2A	1	Methyltransferase	AML, ALL	Translocation, overexpression
PBRM1	PB1 BAF180	6	SWI/SNF PBAF subunit	clear cell renal carcinoma, breast, pancreatic	Mutation, deletion, overexpression
PHIP	WDR11	2	Insulin signaling		
PRPF8	PRP8	0*	Spliceosome subunit		
SMARCA2	BRM	1	SWI/SNF ATPase	prostatic	Reduced expression
SMARCA4	BRG1	1	SWI/SNF ATPase	NSCLC, pancreatic	Mutation
SP100		1	Transcriptional regulator	apparent tumor suppressor	

SP110	1	Transcriptional regulator			
SP140	1	Transcriptional regulator			
SP140L	1	Transcriptional regulator			
TAF1	KAT4	2	Transcription initiation		
TAF1L		2	Transcription initiation		
TRIM24	TIF1A	1	Transcriptional silencer	APL	Translocation
TRIM28	TIF1B KAP1	1	Transcriptional silencer	Gastric	Overexpression
TRIM33	TIF1G	1	Transcriptional silencer	papillary thyroid, CMML	Translocation, reduced expression
TRIM66	TIF1D	1	Transcriptional silencer		
ZMYND11	BS69	1	Corepressor		
ZMYND8	PRKCBP1	1	Transcriptional regulator		

Lysine Acetyltransferase

KAT1	HAT1		Chromatin assembly		
KAT2A	GCN5L2	1	Transcriptional co-activator		
KAT2B	p/CAF	1	Transcriptional co-activator	astrocytic tumor	
KAT3A	CREBBP CBP	1	Transcriptional co-activator	ALL, AML, DLBCL, B-NHL	Translocation, overexpression, mutation
KAT3B	EP300 p300	1	Transcriptional co-activator	colorectal, breast, pancreatic, AML, ALL, DLBCL	Mutation, overexpression
KAT4	TAF1	2	Transcription initiation		
KAT5	TIP60		Transcriptional co-regulator		
KAT6A	MOZ MYST3		Transcriptional co-activator	AML	Translocation
KAT6B	MORF MYST4		Transcriptional co-activator	AML	Translocation
KAT7	HBO1 MYST2		Transcriptional co-repressor		
KAT8	hMOF MYST1		Transcriptional co-activator	primary breast carcinoma, medulloblastoma	Downregulation
KAT9	ELP3		Transcription elongation		
KAT12	GTF3C4 TFIIE90		Transcription initiation		
KAT13A	NCOA1 SRC1		Transcriptional co-activator		
KAT13B	NCOA3 SRC3 ACTR		Transcriptional co-activator	breast	Overexpression
KAT13C	NCOA2 SRC2		Transcriptional co-activator		
KAT13D	CLOCK		Transcriptional co-activator		
ESCO1			Sister chromatid cohesion		

Lysine Deacetylase

HDAC1		Transcriptional regulator	PTCL, gastric, prostate, Hodgkin's lymphomas	Overexpression
HDAC2		Transcriptional regulator	PTCL, gastric, prostate, Hodgkin's lymphomas	Overexpression
HDAC3		Transcriptional regulator	Hodgkin's lymphomas	Overexpression

HDAC4	Transcriptional regulator		
HDAC5	Transcriptional regulator		
HDAC6	Transcriptional regulator	PTCL	Overexpression
HDAC7	Transcriptional regulator	PDAC	Overexpression
HDAC8	Transcriptional regulator	PDAC, neuroblastoma	Overexpression
HDAC9	Transcriptional regulator		
HDAC10	Transcriptional regulator		
HDAC11	unknown		
SIRT1	Transcriptional co-repressor	AML, colon, bladder, prostate, glioma, nonmalignant skin, ovarian	Overexpression
SIRT2	Transcriptional silencer	glioma	Overexpression
SIRT3	unknown	breast	Downregulation
SIRT4	unknown	pancreatic, breast	Overexpression
SIRT5	unknown	pancreatic, breast	Overexpression
SIRT6	DNA repair, telomere maintenance	colon, breast	Overexpression
SIRT7	Transcriptional regulator	breast	Overexpression

*BRWD2 and PRPF8 were included in the study of the acetylome since they have been previously proposed to contain bromodomains.

NSCLC, non-small cell lung cancer; B-CLL, B-cell chronic lymphocytic leukemia; ALL, Acute lymphoblastic leukemia; AML, Acute myeloid leukemia; DLBCL, Diffuse large B-cell lymphoma; B-NHL, B-cell non-Hodgkin lymphomas; APL, Acute promyelocytic leukemia; PTCL, peripheral T-cell lymphomas; PDAC, pancreatic ductal adenocarcinoma; CMML, Chronic myelomonocytic leukaemia.

Table 2: Known bromodomain inhibitors and their cellular effects

Inhibitors	Main targets	Binding K _D	Cellular effects	Reference
Bromodomain				
(+)-JQ1	BET BRD	50-90nM	Inhibition of BRD4/NUT translocation product	Nature. 2010;468(7327):1067-73.
			Inhibition of MYC oncogene	Proc Natl Acad Sci U S A. 2011;108(40):16669-74. Cell. 2011;146(6):904-17. Nature. 2011;478(7370):524-8.
I-BET	BET BRD	49nM	Anti-inflammatory action	Nature. 2010;468(7327):1119-23.
GW841819X	BET BRD	24nM	ApoA1 activity induction, Anti-inflammatory action inhibits p53 interaction	J. Med. Chem. 2011, 54, 3827–3838.
Ischemin	KAT3A	19uM	with CBP; prevents apoptosis in ischemic cardiomyocytes	Chem. Biol. 2011, 18, 531–541.
Cyclic Peptides	KAT3A	8-75uM	Modulate p53 activity and response to DNA damage Inhibit KAT2B	J. Am. Chem. Soc. 2011, 133, 2040–2043.
Unnamed compound	KAT2B	1-5uM	bromodomain binding to acetylated HIV-Tat	J Am Chem Soc. 2005;127(8):2376-7.

Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family

Panagis Filippakopoulos^{1‡}, Sarah Picaud¹, Maria Mangos³, Tracy Keates¹, Jean-Philippe Lambert^{6,7}, Dalia Barsyte-Lovejoy³, Ildiko Felletar¹, Rudolf Volkmer⁵, Susanne Müller¹, Tony Pawson^{6,7}, Anne-Claude Gingras^{6,7}, Cheryl H. Arrowsmith^{3,4}, Stefan Knapp^{1,2,8‡}.

¹Nuffield Department of Clinical Medicine, Structural Genomics Consortium, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford, OX3 7LD, UK.

²Department of Biochemistry and Molecular Biology, George Washington University, School of Medicine and Health Sciences, 2300 Eye Street, NW, Suite 530, Washington, DC 20037, USA.

³Structural Genomics Consortium, University of Toronto, Toronto, Ontario, M5G 1L7, Canada.

⁴Ontario Cancer Institute, Campbell Family Cancer Research Institute and Department of Medical Biophysics, University of Toronto, 101 College Street, Toronto, Ontario, M5G 1L7, Canada.

⁵Institut für Medizinische Immunologie, Charité - Universitätsmedizin Berlin, Hessische Str. 3-4, 10115 Berlin, Germany.

⁶Centre for Systems Biology, Samuel Lunenfeld Research Institute, 600 University Avenue, Toronto, Ontario, M5G 1X5, Canada.

⁷Department of Molecular Genetics, University of Toronto, 1 Kings College Circle, Toronto, Ontario, M5S 1A8, Canada.

⁸Department of Oncology, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford, OX3 7LD, UK.

Running title: Comprehensive analysis of human Bromodomains

Data deposition: The crystal structures reported in this paper have been deposited in the Protein Data Bank, www.pdb.org.

#Correspondence:

e-mail: stefan.knapp@sgc.ox.ac.uk, Phone: +44 1865 617584 Fax: +44 1865 617575 or

e-mail: panagis.filippakopoulos@sgc.ox.ac.uk, Phone: +44 1865 617576 Fax: +44 1865 617575

Abstract

Bromodomains (BRDs) are evolutionary conserved protein interaction modules that specifically recognize ε-N-lysine acetylation motifs, a key event in the reading process of epigenetic marks. We identified 61 BRDs in the human genome that cluster into 8 families based on sequence similarity. Crystallization trials led to 29 high-resolution crystal structures, covering all BRD families. Comprehensive cross-family structural analysis identified conserved and family specific structural features necessary for specific acetylation dependent substrate recognition. Screening of more than 30 representative BRDs against systematic peptide arrays covering all possible histone acetylation sites and arrays that combined acetylation sites with other histone marks, resulted in the identification of novel BRD substrates, many of which were subsequently validated in solution by isothermal titration calorimetry. These data revealed a strong influence of flanking post translational modifications (PTMs) such as acetylation and phosphorylation sites suggesting that BRDs recognize combinations of marks rather than singly acetylated sequences. We also uncovered a structural mechanism for the simultaneous binding and recognition of diverse di-acetyl containing peptides by BRD4. This structural and substrate specificity data provide a powerful resource for further studies and a foundation for structure based drug design of specific inhibitors for this emerging target family.

Highlights

- 29 high resolution crystal structures used to characterize the entire human BRD family
- Peptide arrays used to establish core histone binding preferences
- Interactions with histone Kac sites quantified by ITC
- Explored effect of flanking PTMs on Kac recognition by human BRDs

Introduction

ϵ -N-acetylation of lysine residues (K_{ac}) is one of the most frequently occurring post translational modifications (PTMs) in proteins (Choudhary et al., 2009). Acetylation has a profound effect on the physiochemical properties of modified lysine residues neutralizing the positive charge of the ϵ -amino group (Kouzarides, 2000). Lysine acetylation is particularly abundant in large macromolecular complexes that function in chromatin remodeling, DNA damage and cell cycle control (Choudhary et al., 2009) and in particular in histones. Cellular acetylation levels are stringently controlled by two enzyme families: the histone acetyltransferases (HATs) and histone deacetylases (HDACs) (Shahbazian and Grunstein, 2007). Histone acetylation has been associated with transcriptional activation but specific marks have also been linked to DNA repair (Kouzarides, 2007).

Bromodomains (BRDs) are protein interaction modules that exclusively recognize acetylation motifs. BRDs are evolutionary conserved and present in diverse nuclear proteins comprising HATs (GCN5, PCAF), ATP-dependent chromatin-remodeling complexes (BAZ1B), helicases (SMARCA), methyltransferases (MLL, ASH1L), transcriptional co-activators (TRIM/TIF1, TAFs) transcriptional mediators (TAF1), nuclear scaffolding proteins (PB1) and the BET family (Muller et al., 2011) (**Figure 1A, Supplemental Table S1**). Despite large sequence variations, all BRD modules share a conserved fold that comprises a left-handed bundle of four alpha helices (α_Z , α_A , α_B , α_C), linked by loop regions of variable length (ZA and BC loops), which line the K_{ac} binding site and determine binding specificity. Co-crystal structures with peptides have demonstrated that K_{ac} is recognized by a central deep hydrophobic cavity, where it is anchored by a hydrogen bond to an asparagine residue present in most BRDs (Owen et al., 2000).

Dysfunction of BRD proteins has been linked to development of several diseases. For instance, recurrent t(15;19) chromosomal translocations that results in a fusion protein that comprises both BRD4 or BRD3 bromodomains and the NUT (nuclear protein in testis), leads to an aggressive form of human squamous carcinoma (French, 2008; French et al., 2001). Deregulation of transcription as a consequence of altered protein acetylation patterns is a

hallmark of cancer, a mechanism that is currently targeted by HDAC inhibitors (Lane and Chabner, 2009). It is likely that selective inhibitors capable of targeting BRDs will find broad application in medicine and basic research as exemplified by the recent development of highly specific and potent acetyl-lysine competitive BET bromodomain inhibitors (Chung et al., 2011; Dawson et al., 2011; Delmore et al., 2011; Filippakopoulos et al., 2010; Mertz et al., 2011; Nicodeme et al., 2010).

To date only a small number of lysine acetylation marks have been identified to specifically interact with individual BRDs and the often weak affinities reported for bromodomain interactions with their potential target sites have been determined by a variety of different techniques making data comparison difficult (Muller et al., 2011). Reported affinities range from nano- to milli-molar dissociation constants raising the issue of which affinity window is relevant for specific BRD-peptide interactions. For instance, the bromodomains of BRD2 have been shown to bind histone 4 acetylated at lysine 12 (H4K12_{ac}) (Kanno et al., 2004) with dissociation constants (K_D) that range from 360 μ M for the di-acetylated peptide H4K5_{ac}K12_{ac} (Umeshara et al., 2010) to 2.9 mM for the monoacetylated H4K12_{ac} peptide (Huang et al., 2007). Closely spaced multiple K_{ac} sites have also been shown to significantly increase affinity of the histone H4 N-terminus for BRDT by simultaneous binding to the same BRD (Moriniere et al., 2009). Thus, the field would greatly benefit from a more systematic analysis of BRD structure and peptide binding properties in order to better understand acetylation-mediated signalling as interpreted through BRDs.

Here we present a comprehensive structural characterization of the human bromodomain family together with identified K_{ac} specific interaction sites of these essential protein recognition modules with their target sites in histones. Using available sequence databases we identified 61 BRDs in the human proteome that are present in 46 diverse proteins. High throughput cloning led to the establishment of 171 expression systems that yielded functional recombinant proteins. Using these reagents we crystallized and determined the structures of 29 BRDs, including 25 structures that had not been previously published. We performed a SPOT blot analysis that covered all possible K_{ac} sites of human histones (Nady et al., 2008) for 43 members of the bromodomain family. We identified 485 linear K_{ac} dependent BRD-binding motifs, and determined accurate binding affinities in solution for 81 known cellular histone marks by

isothermal titration calorimetry (ITC). Furthermore, we found that bromodomain peptide recognition is dependent on patterns of multiple modifications rather than on a single acetylation site. This study provides a comprehensive structural comparison of this protein family interpreted in the context of a large array of histone interaction data, establishing a powerful resource for future functional studies of this family of epigenetic reader domains.

Results

The Human bromodomain family

Analysis of sequence databases (NCBI, UniProt, PFAM) identified 46 diverse human proteins that contain a total of 61 diverse BRDs. BRD-containing proteins are large multidomain proteins associated with chromatin remodeling, transcriptional control, methyl or acetyl transferase activity or helicases (**Figure 1A, Supplemental Table S1**). The domain organization in BRD containing proteins is evolutionary highly conserved and the BRD motif is often flanked by other epigenetic reader domains. Most frequently observed combinations include the presence of a plant homeodomains (PHD) N-terminal to the BRD, multiple BRDs as well as various other domains that generally mediate protein interactions such as bromo-adjacent homology (BAH) domains (Goodwin and Nicolas, 2001).

Phylogenetic analysis of the BRD family outside the two central core helices was complicated by the low sequence homology and non-conserved insertions in BRD loop regions. We therefore used three dimensional structure based alignments including available NMR models together with secondary structure prediction (Jones, 1999) and manual curation of the aligned sequences to establish an alignment of all human BRDs (**Supplemental Figure S1A**). The derived phylogram clustered into 8 major BRD families designated by Roman numerals (I-VIII) (**Figure 1B**). An un-rooted phylogram of the human BRD family is shown in **Supplemental Figure S1B** and key references for all bromodomain containing proteins are included in **Supplemental Table S1**.

Multiple protein interaction modules can be tightly linked to form a single stable interaction domain or they can be connected by flexible linker sequences allowing conformational adaptation to diverse sequences motifs. An example of a tightly linked dual domain reader is the PHD-BRD domain of TRIM24, in which both domains interact through a large interface that orients both peptide binding cavities to the same side of the protein (Tsai et al., 2010). In contrast, the two BRDs in TAF1 are free to orient independently as shown by the different domain orientations in the dual domain structure determined here and a previously published

model (Jacobson et al., 2000) (**Figure 1C**). The frequent combination of multiple interaction modules in the same protein suggests that the epigenetic reading process involves concomitant recognition of several post translational modifications.

Structural analysis

In order to establish a platform of recombinant BRDs for functional and structural studies we sub-cloned all human BRDs into bacterial expression systems in frame with a cleavable N- (or C-) terminal His₆ tag. A total of 1031 constructs resulted in the identification of 171 expression systems covering 44 BRDs that yielded stable and soluble proteins. Details of the cloned constructs are summarized in **Supplemental Table S2** and descriptions of one representative expression system per BRD are summarized in **Supplemental Table S3**. The expressed proteins provide an excellent coverage of representative BRDs of all eight families.

A total of 133 recombinant BRD constructs covering 44 unique BRDs were expressed at levels sufficient for structural studies, resulting in the determination of a total of 33 crystal structures of apo-BRDs (**Table 1**), or BRDs in complex with acetylated peptides (**Table 2**). Together with previously published structural information, each BRD family is represented by at least one structural model, and families **I**, **II** and **VIII** are either completely or nearly completely covered (**Figure 1B**). All structures presented here were refined at high resolution. A summary of the crystallization conditions, data collection and refinement statistics is compiled in **Supplemental Table S4**.

Despite the low degree of overall sequence homology, all BRD domains shared a conserved overall fold comprising four alpha helices (α_Z , α_A , α_B , α_C) linked by highly variable loop regions (ZA and BC loops) that form the docking site for interacting recognition motifs (**Figure 1D**, **Supplemental Figure S3A**). The C- and N-termini are highly diverse and may comprise additional helices that extend the canonical BRD fold (e.g. the sixth BRD of PB1 has an additional C-terminal helix) or largely extended kinked helices that are present as C- or N-terminal extensions (for instance in TAF1L or ATAD2). The four helices form a deep cavity which is extended by the two loop regions (ZA- and BC- loops) creating a largely hydrophobic K_{ac} pocket. The most notable structural difference within the BRD core fold is a hairpin insertion

located between helix α_Z and the ZA loop that is present in all family **VIII** members. The proximity to the K_{ac} binding site suggests that this insert may play a role in recruitment of acetylated binding partners. Loop insertions are frequently found within the ZA loop resulting in substantial differences in the rim region of the binding pocket. Hydrophobic residues in the ZA loop may contribute to protein instability and the low crystallization success rate observed in our work for BRDs of families **VI** and **V**. Indeed, in the recently published structure of the MLL tandem PHD-BRD module a flexible insertion found in the MLL ZA loop was deleted in order to generate a more stable construct (Wang et al., 2010).

In stark contrast to the conserved fold of BRDs (**Supplemental Figure S2A**), their surface properties are highly diverse. The electrostatic potential of the surface area around the K_{ac} binding site ranges from highly positively to strongly negatively charged, suggesting that BRDs recognize largely different sequences (**Figure 2**). Based on their surface properties, interactions with highly basic histones are not likely for BRDs with highly positive surfaces, as observed for instance for the third BRD of PB1.

Structural superimposition of 33 BRD crystal structures and 4 NMR models revealed conserved motifs throughout the folded protein domain. To refer to specific sites we chose the first BRD of BRD4 as a reference sequence for numbering of residues (**Supplemental Figure S3A**). The N-terminal helix α_Z is highly diverse but it contains three conserved hydrophobic residues oriented towards the core of the helical bundle. This conserved motif follows the generic sequence $\phi_1x_1x_2(x_3)\phi_2x_3x_4x_5(x_6)\phi_3$ where ϕ_i are hydrophobic residues and x_j representing any amino-acid. The insertions at x_3 are present in the N-terminal domain of BET family members. Insertions x_6 are present in the C-terminal BRDs of TAF1 and TAF1L and possibly PRKCBP1 (**Supplemental Figure S3A, B**). Helix α_Z is flanked by a diverse sequence region and a beta hairpin insert present in all family **VIII** BRDs (**Supplemental Figure S1A, S2B**). These diverse loop inserts are typically followed by a short helical segment in the ZA loop. The C-terminus of the helical segment is stabilized by a highly conserved phenylalanine (F83 in BRD4(1)) that is deeply buried by hydrophobic residues present in helix α_C , bridging both sides of the helical bundle (**Supplemental Figure S3C**). The ZA loop harbours also three conserved proline residues in

addition to hydrophobic residues such as the conserved V87/Y97 pair that closely pack to hydrophobic residues present in α_C stabilizing the loop conformation. A conserved tyrosine (Y97) defines the N-terminus of the ZA loop helix present in all BRDs except TRIM28 and the 6th BRD of PB1 which have unusually short ZA loops that have lost this structural element (**Figure 3A** and **Supplemental Figure S3C**).

Helix α_A is preceded by a P ϕ_1 D motif (ϕ_1 is a hydrophobic residue). The conserved aspartate caps the helix α_A forming a hydrogen bond with a backbone amide. Also for this helix the main sites of conservation are hydrophobic residues that contribute to the stability of the core of the structure (**Figure 3B**). The loop region AB contains a highly conserved tyrosine (Y119) which hydrogen bonds to a conserved aspartate (D128) located in helix α_B , presumably stabilizing the loop-helix fold. The long helix α_B shows a conserved pattern of the sequence $\phi_{xx}D\phi_{xx}\phi\phi_xN\phi_{xx}Y/F$ (**Figure 3C**). A conserved asparagine (N135) hydrogen bonds with the ZA loop backbone linking to this α_B loop region which is additionally stabilized by a small hydrophobic core formed around the conserved aromatic residue (Y139) preceding the K_{ac} docking residue (N140). An asparagine residue that anchors K_{ac} by formation of a critical hydrogen bond initiates the BC loop. Structural comparison suggested that this asparagine can be replaced by other hydrogen bond donors, such as threonine or tyrosine side chains. In MLL however, an aspartate occupies this position suggesting that this domain either does not bind acetylated lysine residues, or has a significantly different mechanism to recognize its target sequence. The different variations of hydrogen bond interactions of K_{ac} with the BC loop are shown in **Supplemental Figure S3E**. Similar to helix α_Z , the C-terminal helix α_C exhibits little sequence conservation apart from a number of hydrophobic core residues (**Figure 3D**). In summary, we have identified several highly conserved sequence motifs in BRDs that serve to stabilize the structural fold and conformation of loop regions flanking the K_{ac} binding pocket. An overview of the sequence conservation is shown in **Supplemental Figure S3D**.

Interactions of BRDs with histone acetylation sites

Histone tails are hotspots of post-translational modifications which play key roles in regulation of transcription and all aspects of chromatin biology. However, to date no systematic study has

addressed binding specificity of reader domains. Here we used SPOT peptide arrays that cover all possible K_{ac} sites of the human histones (H1-4, H2A H2B, H3 and H4) in order to identify interaction sites for 33 representative BRDs. To distinguish between K_{ac}-dependent and independent binding we also included all corresponding unmodified peptides. In general, affinities of K_{ac} for BRDs are low suggesting that additional interaction domains may be required for higher affinity target-specific binding, *in vivo*. In some cases we observed K_{ac} independent interaction of BRDs with non-acetylated control peptides. To date, it is not clear if BRDs participate in K_{ac} independent protein interactions as it has been described for PHD domains that recognize a broad variety of differently methylated, acetylated and non-modified peptides (Lan et al., 2007; Org et al., 2008; Tsai et al., 2010).

We identified 485 interactions of BRDs to histone peptides that depend on the presence of a single K_{ac} site (**Figure 4**, **Supplemental Figure S4**, **Table 3**). The nuclear body protein SP140 as well as the related protein LOC39349 and PCAF showed non-specific binding to most peptides. In contrast, the second, fourth, fifth and sixth BRD of PB1, MLL, and TRIM28 interacted with only a few histone K_{ac}-peptides. Also a number of promiscuous sequences were identified, such as the H2AK36 and H2BK85 containing peptides which interacted with most BRDs. To validate the detected interactions and to obtain accurate binding constants in solution we synthesized 53 singly acetylated peptides and determined binding constants by isothermal titration calorimetry (ITC) (**Supplemental Table S5**). We included also 14 peptides that did not bind to BRDs in the SPOT array. As expected these peptides did not show measurable interactions by ITC suggesting that false negatives are not a major concern in the SPOT array study. Also in agreement with the array study, binding of 20 identified interacting peptides was confirmed by ITC experiments showing dissociation constants (K_D) between 3 and ~300 μM. However, 16 peptides that were selected based on published recognition sites did not give rise to detectable interactions in the SPOT array and still exhibited K_D values between 10 and 730 μM by ITC. The detection limit of SPOT arrays is about 500 μM but the data suggests that SPOT arrays do not detect all possible interacting motifs. Steric constraints of the immobilized peptides and potentially the lack of sufficient N- and C-terminal flanking regions are the most likely

reasons for the failure to detect BRD recognition motifs in SPOT arrays. In addition, 35 (12 %) of acetyl-lysine containing peptides were not recognized by acetyl-lysine specific antibodies. However, most of these peptides contained proline residues in close proximity of the K_{ac} site, a likely reason for the failure of the antibody to recognize these sites. Other peptides showed cross reactivity with the His₆ antibody and have been removed from the analysis.

The false negative rate was particularly high for BET family members. Recently, it was demonstrated that murine BRDT preferentially recognized di-acetylated motifs whereas most mono-acetylated peptides tested did not bind tightly to mBRDT bromodomains (Moriniere et al., 2009). This observation prompted us to design a systematic histone H3 array in which we explored combinations of acetylated and trimethylated (K_{me3}) lysines as well as phosphoserine/threonine (pS/pT) modifications around each acetylated lysine (**Figure 5, Supplemental Figure S5**).

Interactions previously reported for singly acetylated lysine sites were largely confirmed. Interestingly, most of the 43 BRDs tested were highly sensitive to modifications flanking the K_{ac} mark. For instance, BRD4(2) did not interact with H3 peptides singly acetylated on K4. In contrast, this domain showed strong interaction with di-acetylated H3 (H3K4_{ac}K9_{ac}) but not with the same peptide acetylated at K4 but tri-methylated at K9 (H3K4_{ac}K9_{me3}). The strongest interaction was observed using di-acetylated H3K4_{ac}K9_{ac} in combination with phosphorylation at T3. Similarly, the BRD of FALZ showed no interaction with non- or singly- acetylated K4 but interacted strongly with H3 pT3K4_{ac}K9_{ac}. Also WDR9(2) and EP300 exclusively interacted with the triply modified H3 pT3K4_{ac}K9_{ac} peptide. The WDR9(2) interaction with H3K14_{ac} showed strong dependence on S10 and T11 phosphorylation as well as acetylation at K18. Indeed, ITC experiments showed that the binding affinity of many BRDs was significantly increased for multiply modified peptides (**Supplemental Tables S5 and S6**). For example, to the K_D of CREBBP decreased from 733 μM for H3K14_{ac} to 131 μM for H3pS10K14_{ac}K18_{ac} suggesting that many BRDs recognize a pattern of modifications rather than a single K_{ac} mark.

To obtain better insight into BRD recognition of multiply acetylated histone tails we designed a systematic μ-spot array of peptide 11-mers that harboured multiple K_{ac} sites of the N-terminal

tails of histones H3 and H4 (**Figure 6A**). Screening against BRDs of the BET family showed that BRD4(2) interacted with most combinations of two and three acetylated lysines, whereas BRD4(1) seemed to specifically recognize multiple marks found on the H4 tail. A tetra-acetylated peptide that contained the H4 acetylation sites K5, K8, K12 and K16 bound with single digit μM K_D to the first BRDs of BRD2 and BRD4, increasing affinity at least 20-fold when compared to single marks. The second BET BRDs bound to tetra-acetyl H4 peptides with about 10-fold weaker affinities, suggesting that the first BRD in BET proteins recognizes the H4 tail (**Figure 6B**). Recently it was demonstrated that BRDT requires two K_{ac} residues for high affinity binding (Moriniere et al., 2009). Our peptide binding data suggest that the BET family and several other BRDs may also recognize multiply acetylated peptides. However, our binding data cannot discriminate between simultaneous recognition of two K_{ac} as opposed to increased avidity for a multiply modified peptide. In order to determine whether the diverse sequence and spacing of histone K_{ac} residues can be accommodated by a single BRD we systematically determined co-crystal structures of BRD4(1) with the di-acetylated peptides H4K5_{ac}K8_{ac}, H4K12_{ac}K16_{ac} and H4K16_{ac}K20_{ac}. In all cases the two acetylated lysines bound simultaneously and with identical conformations to the BRD4(1) K_{ac} binding site (**Figure 6C**). The N-terminal K_{ac} always formed the anchoring hydrogen bond with the conserved asparagines (N140). In the N-terminal region of H4, flexible glycine residues allow variable peptide conformations with two (H4₁₋₁₁K5_{ac}K8_{ac} - **Supplemental Figure S6A**) or three (H4₁₁₋₂₁K12_{ac}K16_{ac} - **Supplemental Figure S6B**) linking residues, whereas the large side-chains in H4₁₅₋₂₅K16_{ac}K20_{ac} (**Supplemental Figure S6C**) fit perfectly into surface grooves created by the ZA and BC loops. These structures explain the similar affinities observed for the various combinations of di- and tri-acetylated H4 peptides. They also suggest that the greater apparent affinity of BRD4(1) and BRD2(1) for tetra acetylated H4 peptides is an avidity effect. However, not all di-acetylated H4 sequences are compatible with this bidentate recognition process. The co-crystal structure of H4₇₋₁₇K8_{ac}K12_{ac} with BRD4(1) revealed a canonical mono-acetylated recognition mode (**Supplemental Figure S6D**) suggesting that the H4₉₋₁₁ linker sequence is not suitable for a simultaneous recognition of the two K_{ac} by a single BRD. Consistent with this notion, ITC experiments revealed a binding stoichiometry (N) of 0.5, indicating binding of two BRDs to the

H4K8_{ac}K12_{ac} peptide whereas only a single binding event with significantly increased affinity was observed for the H4K5_{ac}K8_{ac} peptide (**Supplemental Table S7**). A representative set of ITC data is shown in **Figure 6D**.

We were interested in the sequence requirements of the di-acetyl-lysine BET recognition and designed a systematic peptide array in which we modulated the spacer sequence and residue properties of residues located between the two K_{ac} binding sites (**Figure 7A**). For the first bromodomains of the BET family, a spacer of 2 glycine residues was optimal. However, BRD2(1) tolerated also longer linker sequences. For two-residue linkers, bulky amino acids in the first linker position were not tolerated but changes of residue properties in the second linker position did not strongly influence binding. Intriguingly, the wild type sequence “GG” of the H4 K5-K8 linker region seems highly optimized for interaction of the first bromodomain of BET family members. Binding of di-K_{ac} marks separated by three residue spacers as found in sequences linking the H4 K8-K16 and K16-K20 required a glycine or a hydrophobic residue in the first linker position for optimal binding to the first BET bromodomains. Acidic residues in any linker position led to loss of interaction with H4 histone tail peptides. In contrast, the second BRDs of BET bromodomains bound either weakly (BRD2), not at all (BRD3) or promiscuously (BRD4) to histone sequences and their variants present in this array. ITC data collected on the first BRD of BRD4 showed a 30 fold increase in affinity between the singly acetylated peptide H4K5_{ac} to the most optimal wild type peptide H4K5_{ac}K8_{ac}. In contrast, di-acetylation had only a modest effect on binding affinities of the second bromodomain of BRD4. As reported for interactions with single acetylation sites in the case of the BRDs of BRD2 (Umehara et al., 2010), alanine mutants of the conserved asparagine (N140 and N443 in the first and second BRDs in BRD4) did also abolish binding of di-acetylated peptides in both SPOT assays as well as in ITC (**Figure 7B, 7C and Supplemental Table S7**).

In order to address the question if full-length BRD4 also interacts with the identified K_{ac} sites in the context of intact nucleosomes we performed pull-down assays on nucleosomal preparations using Flag-tagged BRD4 and antibodies that specifically recognize K_{ac} sites. In agreement with our peptide array studies we identified histone interaction of BRD4 with the H4 sites K5_{ac}, K8_{ac}, K12_{ac}, K16_{ac} and H3 K14_{ac} (**Figure 7D**). Unfortunately, no antibodies are currently available that

specifically recognize di-acetylated marks in histones H3 and H4. We therefore analysed K_{ac} enriched tryptic digests prepared from pull-downs of salt extracted histone using C-terminally biotinylated BRD4(1) and BRD4(2) by mass spectroscopy (**Figure 7E**). We were able to detect numerous poly-acetylated histone peptides associated with BRD4 BRDs; incubation with the active BRD4 inhibitor (+)-JQ1 (Filippakopoulos et al., 2010), but not its inactive stereoisomer (-)-JQ1 abrogated interaction with acetylated histones, indicating that the purifications were specific. Importantly, we observed that the first bromodomain BRD4(1) interacted mostly with poly-acetylated histone H4 peptides and that the majority of peptides identified contained at least two K_{ac} sites. These results are in good agreement with the strong increase in binding affinity and the preference for BRD4(1) for histone di-acetyl marks observed in our *in vitro* binding studies.

Discussion

Recent developments in biotechnology and structural biology have facilitated rapid generation of structural data enabling determination of high resolution structures of most members of certain protein families within a short time frame (Barr et al., 2009). The study presented here represents a comprehensive structural description of the entire human bromodomain family with at least one representative structural model for each branch in the BRD phylogenetic tree. Structural coverage of families I, II and VIII is complete or nearly complete. These crystal structures enabled a detailed sequence comparison of this highly diverse domain family. Importantly, although the protein family database Pfam (Finn et al., 2010) extended the BRD fold from the initially predicted central helices α_A and α_B (Haynes et al., 1992), to a 110 residue motif (Jeanmougin et al., 1997), sequence based tools still fail to predict correct domain boundaries for BRDs that contain long ZA- and BC- loop insertions. The excellent structural coverage of the BRD family enabled the identification of BRD signature motifs and family specific secondary structure elements, such as the ZA loop helix α_{AZ} and the subfamily VIII specific beta hairpin insert.

Probing BRD histone recognition by peptide arrays

Peptide arrays offer a rapid technology for screening protein-peptide interactions. The technology was developed more than a decade ago (Reineke et al., 2001) and has recently been applied to study epigenetic methyl-lysine reader domain interactions with histone tails (Nady et al., 2011; Nady et al., 2008). Recent progress in array technology allows peptide densities of up to 40,000 spots per square centimetre of solid support, enabling in principle genome wide analysis of reader domains with peptidic recognition motifs (Beyer et al., 2009). To date, more than 100 histone PTMs that function as recruitment platforms for chromatin proteins have been described (Kouzarides, 2007). We chose therefore a systematic peptide array that covered all possible histone acetylation sites to characterize a representative set of BRD reader domains. However, many bromodomains may interact with acetylation sites present in non-

histone proteins. In fact, we did not observe interaction with histone peptides for a number of bromodomains.

Recognition sites for only a few BRDs have been previously characterized and reported substrate affinities range from the low μM to the mM K_D range. Specific recognition sites in histones identified by our SPOT arrays that contained only a single K_{ac} site per peptide had binding affinities between 3 μM to 350 μM , which fall into the affinity range that has been reported for other BRD K_{ac} interactions (Shen et al., 2007; Zeng et al., 2008). However, comparison of binding constants determined in solution by ITC with SPOT intensity did not always correlate, suggesting that peptides linked to cellulose supports used in this study did not allow quantification of binding affinities. However, a recent study found good correlation with SPOT intensities and substrate Km values for deacetylases indicating improved correlation for proteins with enzymatic activity (Smith et al., 2011).

Many bromodomains recognize patterns of post translational modifications

The weak contribution of the K_{ac} mark to the binding affinity of BRDs to their target sites makes bromodomain interactions particularly sensitive to changes in the environment of the K_{ac} site. The high density of PTMs in histones and other signalling molecules results in a large number of potential combinations of marks that regulate chromatin-templated recognition processes. In this study we selected a limited but systematic set of combinations that may be present in histone H3 and comprehensively profiled this array against the human BRD family. The observed strong influence of neighbouring PTMs, such as phosphorylation, on recognition of bromodomains and their target sites, suggests tight coupling of phosphorylation signalling with epigenetic mechanisms of regulation. Many examples of this coupling have been reported for chromatin modifying enzymes. For instance, H3S10 phosphorylation has been shown to be functionally linked to GCN5 mediated acetylation at H3K14 (Lo et al., 2001; Lo et al., 2000) and crosstalk of the three marks, H3K9_{ac}, H3pS10 and H4K16_{ac} regulates transcriptional elongation of certain genes by providing a nucleosome platform that recruits BRD4/PTEF-b (Zippo et al., 2009). Also H3pS10 is a prerequisite for H3K4 tri-methylation (Li et al., 2011) which in turn has been shown to prevent phosphorylation at H3T3 by haspin (Eswaran et al., 2009). These data strongly

suggest that combinatorial motifs rather than single PTMs determine the cellular outcome of processes regulated by epigenetic reader domains. This hypothesis would also explain the large amount of contradictory results in studies where single marks have been assigned specific function such as transcriptional activation or silencing. Thus, the reading process of the “histone code” is a sophisticated, nuanced chromatin language that recognizes combinations of marks rather than single PTMs (Berger, 2007).

Simultaneous binding of multiple acetyl lysines to a single BRD

Recent structural and biophysical studies demonstrated that murine BRDT requires at least two adjacent acetylation sites for tight interaction with the histone H4 tails (Moriniere et al., 2009). Our SPOT array and ITC data showed that multiple K_{ac} sites are generally required for specific recognition of the histone H4 tail by all human BET family members.

We were interested if interactions of di-acetyl-lysine also occur outside the BET bromodomain family. Using rigid docking of the H4K5_{ac}K8_{ac} peptide onto all available crystal structures revealed that a number of other bromodomains would have an acetyl-lysine binding site architecture that would be compatible with the binding of this di-acetylated peptide (**Supplemental Figure 6E**).

The presence of multiple reader modules in chromatin modification complexes led to the proposal that distinct epigenetic signatures are interpreted by a multivalent reading process that engages diverse binding modules (Ruthenburg et al., 2007). For instance, recently the dual reader module PHD-BRD in BPTF has been shown to specifically recognize a combination of H4K16_{ac} and H3K4_{me3} at the mono-nucleosome level (Ruthenburg et al., 2011). Similarly, the tandem Tudor domain of UHRF1 recognizes H3 when K9 is trimethylated, and K4 is unmodified - a histone modification state associated with heterochromatin (Nady et al., 2011). Combinations of PHD and BRDs are particularly frequent and are a hallmark of BRD proteins in families **V** and **VI** and the PHD-BRD structure showed that the two reader domains form a single, stable functional unit (Tsai et al., 2010). The work by Tsai et al. also suggests that the TRIM24 PHD-BRD di-domain binds two different histone tails in opposite orientations. Similarly our array and ITC studies on BRD4 showed that the first BRD of this protein has high affinity for the histone H4

tail while the second BRD most likely recognizes multiply acetylated marks in histone H3. This would be consistent with the notion that proteins that harbour multiple reader domains act as integration platforms for different chromatin proteins.

Epigenetic reader domains are promising drug targets

BRDs have recently emerged as promising targets for the development of protein interaction inhibitors (Chung et al., 2011; Filippakopoulos et al., 2010; Hewings et al., 2011; Nicodeme et al., 2010). The acetylation of lysine residues neutralizes the charge of the primary amine. As a consequence, BRD acetyl-lysine binding sites are deep and largely hydrophobic binding pockets that represent attractive targeting sites for the development of K_{ac} competitive inhibitors. Proteins containing epigenetic reader modules have been implicated in the development of many diseases (Baker et al., 2008; Muller et al., 2011; Reynoird et al., 2010).

The recent development of potent and highly specific K_{ac} competitive inhibitors for BET BRDs provides a compelling case for targeting these BRDs for the treatment of an extremely aggressive subtype of squamous cell carcinoma that is caused by chromosomal rearrangement of BRD3 or BRD4 with NUT (Filippakopoulos et al., 2010; French, 2010). Recent data strongly suggested that targeting BET bromodomains will be beneficial for many diverse cancer types due to down regulation of oncogenes such as c-Myc (Dawson et al., 2011; Delmore et al., 2011; Mertz et al., 2011). The structural data presented here provide the foundation for the rational design of selective BRD inhibitors which will be valuable tools for our understanding of the role of epigenetic reader modules in health and disease.

Experimental Procedures

Protein purification

BRD constructs were sub-cloned into pET28 derived expression vectors. All proteins were expressed as His6 tagged fusions and were purified using Ni-chelating affinity chromatography. Analytical details for construct design, protein expression and purification are given in the Supplemental information and in **Supplemental Tables S2 and S3**.

SPOT assays

Peptides were synthesized on cellulose membranes using a MultiPep SPOT peptide arrayer (Intavis). His6 tagged BRDs were added to a final concentration of 1 µM and blots were developed using an ECL kit (Thermo Scientific) following the manufacturer's protocol.

Isothermal Titration Calorimetry (ITC)

Experiments were carried out on a VP-ITC or an ITC200 microcalorimeter (MicroCal™, Northampton, MA). In most cases a single binding site model was employed, supplied with the MicroCal™ Origin software package.

Crystallization and Structure Determination

Individual proteins were crystallized at either 4 or 20 °C and X-ray diffraction data were collected on beam lines listed in **Supplemental Table S4**. Structures were solved by molecular replacement and were refined as described in detail in the supplemental method section. Crystallization conditions, data collection, refinement statistics and PDB accession codes are listed in **Table I and II** and in **Supplemental Table S4**.

Histone immunoprecipitation

HEK293 cells were grown in DMEM with 10 % FBS and transfected with Brd4-Flag (UniProt: O60885, residues 1-1362, cloned on pCDNA5) vector using GeneJuice (EMD) according to the manufacturer's instructions. Nucleosome isolation protocol was based on Ruthenburg and co-workers (Ruthenburg et al., 2011) with modification at the end of the nuclease treatment.

LC-MS/MS analysis and quantitation

HeLa cells were grown in DMEM supplemented with 10 % fetal bovine serum and antibiotics (Penicillin-Streptomycin). Histones were extracted with high salt (Shechter et al., 2007) and incubated with recombinant biotinylated BRDs in the presence of (+)-JQ1 or (-)-JQ1 prior to purification on streptactin sepharose beads. After elution of bound histones using Trifluoroacetic acid (TFA) and sample lyophilization, trypsin digestion was performed and trypsin was inhibited. Acetylated peptides were purified using anti-K_{ac} agarose beads (ImmuneChem Pharmaceuticals Inc.), and both the unbound fraction and the bound fraction (eluted in TFA) were prepared for mass spectrometry (MS). LC-MS/MS was performed using a NanoLC-Ultra 2D plus HPLC system (Eksigent) coupled to a LTQ-Orbitrap Velos (Thermo Electron) equipped with a nanoelectrospray ion source (Proxeon Biosystems). Spectra were assigned by Mascot (Matrix Science, v2.3) against the human RefSeq database (version 45). Relative quantitation of acetylated peptides was achieved with Proteome Discoverer 1.2 (Thermo Electron). The efficiency of purification with each BRD was monitored by analyzing the fraction unbound to anti-K_{ac}; specificity was ascertained by analyzing the samples incubated with the inhibitor (+)-JQ1.

Acknowledgements

The SGC is a registered charity (number 1097737) that receives funds from the Canadian Institutes for Health Research, the Canada Foundation for Innovation, Genome Canada, GlaxoSmithKline, Pfizer, Eli Lilly, the Novartis Research Foundation, the Ontario Ministry of Research and Innovation and the Wellcome Trust. T.P. and A.-C.G. are supported by a Global Leadership Round in Genomics and Life Sciences from the Ontario Ministry of Health and Innovation. J.-P.L. is supported by a fellowship from the Natural Sciences and Engineering Research Council of Canada. A.-C.G. and C.H.A. hold Canada Research Chairs in Functional Proteomics and Structural Proteomics, respectively.

References

- Akai, Y., Adachi, N., Hayashi, Y., Eitoku, M., Sano, N., Natsume, R., Kudo, N., Tanokura, M., Senda, T., and Horikoshi, M. (2010). Structure of the histone chaperone CIA/ASF1-double bromodomain complex linking histone modifications and site-specific histone eviction. *Proc Natl Acad Sci U S A* **107**, 8153-8158.
- Baker, L.A., Allis, C.D., and Wang, G.G. (2008). PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. *Mutat Res* **647**, 3-12.
- Barr, A.J., Ugochukwu, E., Lee, W.H., King, O.N., Filippakopoulos, P., Alfano, I., Savitsky, P., Burgess-Brown, N.A., Muller, S., and Knapp, S. (2009). Large-scale structural analysis of the classical human protein tyrosine phosphatome. *Cell* **136**, 352-363.
- Berger, S.L. (2007). The complex language of chromatin regulation during transcription. *Nature* **447**, 407-412.
- Beyer, M., Block, I., Konig, K., Nesterov, A., Fernandez, S., Felgenhauer, T., Schirwitz, C., Leibe, K., Bischoff, R.F., Breitling, F., et al. (2009). A novel combinatorial approach to high-density peptide arrays. *Methods Mol Biol* **570**, 309-316.
- Charlop-Powers, Z., Zeng, L., Zhang, Q., and Zhou, M.M. (2010). Structural insights into selective histone H3 recognition by the human Polybromo bromodomain 2. *Cell Res* **20**, 529-538.
- Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834-840.
- Chung, C.W., Coste, H., White, J.H., Mirguet, O., Wilde, J., Gosmini, R.L., Delves, C., Magny, S.M., Woodward, R., Hughes, S.A., et al. (2011). Discovery and characterization of small molecule inhibitors of the BET family bromodomains. *J Med Chem* **54**, 3827-3838.
- Dawson, M.A., Prinjha, R.K., Dittman, A., Girotopoulos, G., Bantscheff, M., Chan, W.I., Robson, S.C., Chung, C.W., Hopf, C., Savitski, M.M., et al. (2011). Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature*.
- Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastritis, E., Gilpatrick, T., Paranal, R.M., Qi, J., et al. (2011). BET Bromodomain Inhibition as a Therapeutic Strategy to Target c-Myc. *Cell* **146**, 904-917.
- Eswaran, J., Patnaik, D., Filippakopoulos, P., Wang, F., Stein, R.L., Murray, J.W., Higgins, J.M., and Knapp, S. (2009). Structure and functional characterization of the atypical human kinase haspin. *Proc Natl Acad Sci U S A* **106**, 20198-20203.
- Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., et al. (2010). Selective inhibition of BET bromodomains. *Nature* **468**, 1067-1073.
- Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., et al. (2010). The Pfam protein families database. *Nucleic Acids Res* **38**, D211-222.
- French, C.A. (2008). Molecular pathology of NUT midline carcinomas. *J Clin Pathol*.
- French, C.A. (2010). NUT midline carcinoma. *Cancer Genet Cytogenet* **203**, 16-20.
- French, C.A., Miyoshi, I., Aster, J.C., and et al. (2001). BRD4 bromodomain gene rearrangement in aggressive carcinoma with translocation t (15;19). *American Journal of Pathology* **159**, 1987-1992.

Goodwin, G.H., and Nicolas, R.H. (2001). The BAH domain, polybromo and the RSC chromatin remodelling complex. *Gene* 268, 1-7.

Haynes, S.R., Dollard, C., Winston, F., Beck, S., Trowsdale, J., and Dawid, I.B. (1992). The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. *Nucleic Acids Res* 20, 2603.

Hewings, D.S., Wang, M., Philpott, M., Fedorov, O., Uttarkar, S., Filippakopoulos, P., Picaud, S., Vuppusetty, C., Marsden, B., Knapp, S., *et al.* (2011). 3,5-Dimethylisoxazoles Act As Acetyl-lysine-mimetic Bromodomain Ligands. *J Med Chem* 54, 6761-6770.

Huang, H., Zhang, J., Shen, W., Wang, X., Wu, J., Wu, J., and Shi, Y. (2007). Solution structure of the second bromodomain of Brd2 and its specific interaction with acetylated histone tails. *BMC Struct Biol* 7, 57.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAF(II)250 double bromodomain module. *Science* 288, 1422-1425.

Jeanmougin, F., Wurtz, J.M., Le Douarin, B., Chambon, P., and Losson, R. (1997). The bromodomain revisited. *Trends Biochem Sci* 22, 151-153.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292, 195-202.

Kanno, T., Kanno, Y., Siegel, R.M., Jang, M.K., Lenardo, M.J., and Ozato, K. (2004). Selective recognition of acetylated histones by bromodomain proteins visualized in living cells. *Molecular Cell* 13, 33-43.

Kouzarides, T. (2000). Acetylation: a regulatory modification to rival phosphorylation? *EMBO J* 19, 1176-1179.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693-705.

Lan, F., Collins, R.E., De Cegli, R., Alpatov, R., Horton, J.R., Shi, X., Gozani, O., Cheng, X., and Shi, Y. (2007). Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* 448, 718-722.

Lane, A.A., and Chabner, B.A. (2009). Histone deacetylase inhibitors in cancer therapy. *Journal of Clinical Oncology* 27, 5459-5468.

Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D., and Patel, D.J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* 442, 91-95.

Li, Y., Sun, L., Zhang, Y., Wang, D., Wang, F., Liang, J., Gui, B., and Shang, Y. (2011). The histone modifications governing TFF1 transcription mediated by estrogen receptor. *J Biol Chem*.

Lo, W.S., Duggan, L., Emre, N.C., Belotserkovskya, R., Lane, W.S., Shiekhattar, R., and Berger, S.L. (2001). Snf1—a histone kinase that works in concert with the histone acetyltransferase Gcn5 to regulate transcription. *Science* 293, 1142-1146.

Lo, W.S., Trievel, R.C., Rojas, J.R., Duggan, L., Hsu, J.Y., Allis, C.D., Marmorstein, R., and Berger, S.L. (2000). Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. *Mol Cell* 5, 917-926.

Mertz, J.A., Conery, A.R., Bryant, B.M., Sandy, P., Balasubramanian, S., Mele, D.A., Bergeron, L., and Sims, R.J., 3rd (2011). Targeting MYC dependence in cancer by inhibiting BET bromodomains. *Proc Natl Acad Sci U S A* 108, 16669-16674.

Moriniere, J., Rousseaux, S., Steuerwald, U., Soler-Lopez, M., Curtet, S., Vitte, A.L., Govin, J., Gaucher, J., Sadoul, K., Hart, D.J., *et al.* (2009). Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* 461, 664-668.

Mujtaba, S., He, Y., Zeng, L., Farooq, A., Carlson, J.E., Ott, M., Verdin, E., and Zhou, M.M. (2002). Structural basis of lysine-acetylated HIV-1 Tat recognition by PCAF bromodomain. *Mol Cell* 9, 575-586.

- Mujtaba, S., He, Y., Zeng, L., Yan, S., Plotnikova, O., Sachchidanand, Sanchez, R., Zeleznik-Le, N.J., Ronai, Z., and Zhou, M.M. (2004). Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Mol Cell* 13, 251-263.
- Muller, S., Filippakopoulos, P., and Knapp, S. (2011). Bromodomains as therapeutic targets. *Expert Rev Mol Med* 13, e29.
- Nady, N., Lemak, A., Walker, J.R., Avvakumov, G.V., Karetta, M.S., Achour, M., Xue, S., Duan, S., Allali-Hassani, A., Zuo, X., et al. (2011). Recognition of multivalent histone states associated with heterochromatin by UHRF1. *J Biol Chem*.
- Nady, N., Min, J., Karetta, M.S., Chedin, F., and Arrowsmith, C.H. (2008). A SPOT on the chromatin landscape? Histone peptide arrays as a tool for epigenetic research. *Trends Biochem Sci* 33, 305-313.
- Nicodeme, E., Jeffrey, K.L., Schaefer, U., Beinke, S., Dewell, S., Chung, C.W., Chandwani, R., Marazzi, I., Wilson, P., Coste, H., et al. (2010). Suppression of inflammation by a synthetic histone mimic. *Nature* 468, 1119-1123.
- Org, T., Chignola, F., Hetenyi, C., Gaetani, M., Rebane, A., Liiv, I., Maran, U., Mollica, L., Bottomley, M.J., Musco, G., et al. (2008). The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO Rep* 9, 370-376.
- Owen, D.J., Ornaghi, P., Yang, J.C., Lowe, N., Evans, P.R., Ballario, P., Neuhaus, D., Filetici, P., and Travers, A.A. (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *Embo J* 19, 6141-6149.
- Reineke, U., Volkmer-Engert, R., and Schneider-Mergener, J. (2001). Applications of peptide arrays prepared by the SPOT-technology. *Curr Opin Biotechnol* 12, 59-64.
- Reynoird, N., Schwartz, B.E., Delvecchio, M., Sadoul, K., Meyers, D., Mukherjee, C., Caron, C., Kimura, H., Rousseaux, S., Cole, P.A., et al. (2010). Oncogenesis by sequestration of CBP/p300 in transcriptionally inactive hyperacetylated chromatin domains. *EMBO J* 29, 2943-2952.
- Ruthenburg, A.J., Li, H., Milne, T.A., Dewell, S., McGinty, R.K., Yuen, M., Ueberheide, B., Dou, Y., Muir, T.W., Patel, D.J., et al. (2011). Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell* 145, 692-706.
- Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, C.D. (2007). Multivalent engagement of chromatin modifications by linked binding modules. *Nat Rev Mol Cell Biol* 8, 983-994.
- Shahbazian, M.D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. *Annual Review of Biochemistry* 76, 75-100.
- Shechter, D., Dormann, H.L., Allis, C.D., and Hake, S.B. (2007). Extraction, purification and analysis of histones. *Nat Protoc* 2, 1445-1457.
- Shen, W.Q., Xu, C., Huang, W., Zhang, J.H., Carlson, J.E., Tu, X.M., Wu, J.H., and Shi, Y.Y. (2007). Solution structure of human Brg1 bromodomain and its specific binding to acetylated histone tails. *Biochemistry* 46, 2100-2110.
- Singh, M., Popowicz, G.M., Krajewski, M., and Holak, T.A. (2007). Structural ramification for acetyl-lysine recognition by the bromodomain of human BRG1 protein, a central ATPase of the SWI/SNF remodeling complex. *Chembiochem* 8, 1308-1316.
- Smith, B.C., Settles, B., Hallows, W.C., Craven, M.W., and Denu, J.M. (2011). SIRT3 substrate specificity determined by peptide arrays and machine learning. *ACS Chem Biol* 6, 146-157.
- Sun, H., Liu, J., Zhang, J., Shen, W., Huang, H., Xu, C., Dai, H., Wu, J., and Shi, Y. (2007). Solution structure of BRD7 bromodomain and its interaction with acetylated peptides from histone H3 and H4. *Biochem Biophys Res Commun* 358, 435-441.
- Tsai, W.W., Wang, Z., Yiu, T.T., Akdemir, K.C., Xia, W., Winter, S., Tsai, C.Y., Shi, X., Schwarzer, D., Plunkett, W., et al. (2010). TRIM24 links a non-canonical histone signature to breast cancer. *Nature* 468, 927-932.

- Umehara, T., Nakamura, Y., Jang, M.K., Nakano, K., Tanaka, A., Ozato, K., Padmanabhan, B., and Yokoyama, S. (2010). Structural basis for acetylated histone H4 recognition by the human BRD2 bromodomain. *J Biol Chem* 285, 7610-7618.
- Vollmuth, F., and Geyer, M. (2010). Interaction of propionylated and butyrylated histone H3 lysine marks with Brd4 bromodomains. *Angew Chem Int Ed Engl* 49, 6768-6772.
- Wang, Z., Song, J., Milne, T.A., Wang, G.G., Li, H., Allis, C.D., and Patel, D.J. (2010). Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. *Cell* 141, 1183-1194.
- Zeng, L., Zhang, Q., Gerona-Navarro, G., Moshkina, N., and Zhou, M.M. (2008). Structural basis of site-specific histone recognition by the bromodomains of human coactivators PCAF and CBP/p300. *Structure* 16, 643-652.
- Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., and Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* 138, 1122-1136.

Figure Legends

Figure 1: Domain organization, phylogenetic tree and overall fold of bromodomains. **A:** Domain organization of representative proteins that contain BRDs. The name and the length of the selected proteins are shown on the bar chart in the left panel. The positions of the different domains are highlighted as shown by the legend on the right. **B:** Phylogenetic tree of the human bromodomain family. The different families are named by Roman numbers (I-VIII). Structures determined in this study, by NMR or by other groups are indicated by blue, red and green dots, respectively. **C:** Domain flexibility as seen in the tandem BRD modules of TAF1 di-domain structure (orange-PDB: 1EQF) and a new structure (green-PDB: 3UV5), highlighting the ability of BRDs to adopt different relative orientations which may influence the recognition of their target sequences. **D:** Overall structure of the BRD4(1) BRD. N-, C-termini and secondary structure elements are labelled.

Figure 2: Electrostatic surface potential of human bromodomains. The domains are grouped into the eight BRD families. Electrostatic surface potential is shown between -10kT/e (red) and +10kT/e (blue). The BRD names and structures (pdb accession code) are shown in the figure. All domains are shown in identical orientation with their acetyl-lysine binding site oriented towards the reader and highlighted with a dashed circle on the top-left structure (PCAF).

Figure 3: Conserved residues and sequence logos. Conserved residues are shown as sequence logos (lower panel in each figure) and their location is shown in the ribbon diagram above. Secondary structure elements and residue labels are coloured in all figures as follows: α_Z blue, α_A red, α_B green, α_C brown, ZA loop magenta. Motifs and location of residues are shown for **A:** helix α_A and AB loop, **B:** AB-loop and Helix α_B , **C:** helix α_C , **D:** ZA-loop.

Figure 4: Detected interactions of bromodomains with histones in SPOT arrays. 33 BRDs were screened against an array of singly acetylated peptides that cover all possible acetylation sites in histones H1-4 (right panel), H2A, H2B, H3 and H4 (left panel). Non K_{ac} specific interactions (corresponding non-acetylated) peptides are shaded in gray. Spots are shaded by different spot intensities as indicated in the figure.

Figure 5: Influence of neighbouring post translational modifications on bromodomain interaction with histone H3. Shown are interactions detected in SPOT arrays shaded by spot intensity as indicated in the legend at the top left corner of the figure. The influence of lysine trimethylation (K_{me3}), acetylation and phosphorylation (pT, pS) has been studied on binding to a central acetylated lysine epitope. The combination of the different modifications is indicated in the right panel. Non-modified peptides have been included as controls to identify lysine acetylation independent of interactions.

Figure 6: Binding of the N- and C- BRDs of BRD4 to multiply acetylated histone H3 and H4. **A:** Interactions detected in microSPOT arrays for BRD4(1) and BRD4(2) comprising multiply acetylated H3 (shown in green) or H4 (shown in blue) peptides. **B:** Peptide lengths are given together with the location of the K_{ac} marks (green hexagons). Binding of seven BRD members of the BET subfamily is summarized, highlighting the effect of multiple K_{ac} marks as well as neighbouring Ser or Thr phosphorylation. **C:** Structural overlay of di-acetylated peptides to the Kac binding site of BRD4(1). The binding mode is retained although the linker between marks and the flanking residues are not the same. **D:** Two representative ITC traces of poly-acetylated histone 4 peptide binding to BRD4(1). Peptide sequences are shown in the inset.

Figure 7: Effect of distance between acetylated lysines and *in-vivo* binding of BRD4. **A:** Effect of poly-Gly and linker sequence on binding of doubly-acetylated peptides to BET BRDs profiled by SPOT assays. Interactions are weaken or abolished when the docking asparagine is mutated to an Ala (N140A for the first and N433A for the second BRDs of BRD4). **B:** Effect of poly-Gly linker on the binding of H4K5_{ac}K8_{ac} to BRD4(1) evaluated by ITC, demonstrating that the natural recognition sequence has the optimal sequence for binding. Peptide sequences are given in the inset. **C:** The second BRD of BRD4 is more promiscuous as demonstrated by ITC, exhibiting weaker binding for all tested peptides. **D:** Immunoprecipitation of Flag tagged Brd4 from transfected cell nucleosome fraction and western blotting using anti acetylated histone antibodies. Input represents 1% of total input. IgG was used for control immunoprecipitations. **E:** Individual BRD4 bromodomains purify histones with distinct acetylation status from histone fractions. Acetylated histone peptides associated with biotinylated BRD4(1) or BRD4(2) were identified by LC-MS/MS and the relative peak intensity of individual peptides was expressed as a

ratio of peak area of the specific peptide to the sum of all peak areas for acetylated histone peptides in each sample.

Tables

Table 1: Available crystal structures of human bromodomains*

Protein	PDB Code	Resolution [Å]	Reference
ASH1L	3MQM	2.54	this study
ATAD2	3DAI	1.95	this study
BAZ2B	3G0L	2.03	this study
BRD1	3RCW	2.21	this study
BRD2(1)	2DVS	2.04	(Umehara et al., 2010)
BRD2(2)	2G4A	NMR	unpublished
BRD3(1)	2NXB	1.40	this study
BRD3(2)	2O01	1.70	this study
BRD4(1)	2OSS	1.35	this study
BRD4(2)	2OUO	1.89	this study
BRD7	2I7K	NMR	(Sun et al., 2007)
BRD9	3HME	2.23	this study
BRDT(1)	2RFJ	2.05	this study
CECR2	3NXB	1.83	this study
EP300	3I3J	2.33	this study
CREBBP	3DWY	1.98	this study
FALZ	2F6N, 3UV2	1.45, 1.58	(Li et al., 2006), this study
GCN5L2	3D7C	2.06	this study
KIAA1240	3LXJ	2.33	this study
MLL	3LQH	1.72	(Wang et al., 2010)
PB1(1)	3IU5	1.63	this study
PB1(2)	3HMF	1.63	this study
PB1(3)	3K2J	2.20	this study
PB1(4)	3TLP	2.13	this study
PB1(5)	3G0J	1.78	this study
PB1(6)	3IU6	1.79	this study
PCAF	3GG3	2.25	this study
PHIP(2)	3MB3	2.25	this study
SMARCA2	2DAT	NMR	unpublished
SMARCA4	2GRC, 3UVD	1.50	(Singh et al., 2007), this study
TAF1(2)	3AAD, 3UV4	3.30, 1.89	(Akai et al., 2010), this study
TAF1(1/2)	1EQF, 3UV5	2.10, 2.03	(Jacobson et al., 2000), this study
TAF1L(2)	3HMH	2.05	this study
TIF1	3O33	2.00	(Tsai et al., 2010)
TRIM28	2RO1	NMR	(Zeng et al., 2008)
WDR9(2)	3Q2E	1.74	this study

*Only the first structure deposited into the PDB has been included in the table. NMR structures are also available for several bromodomains. However, due to space limitations they have not been included unless no crystal structure is available for the target.

Table 2: Available structures of BRD peptide complexes

Protein	PDB Code	Reso. [Å]	Histone Mark	Peptide/Ligand	Reference
BRD2(1)	2DVQ	2.04	H4K12	SGRGKGGKGLG K_{ac} GGA	(Umehara et al., 2010)
BRD2(1)	2DVR	2.30	H4K12	GGKGLG K_{ac} GGA	(Umehara et al., 2010)
BRD2(2)	2E3K	2.30	H4K5/K12	SGRG K_{ac} GGKGLG K_{ac} GGA	unpublished
BRD4(1)	3MUK	1.75	H3K23prop	AT K_{prop} AARK	(Vollmuth and Geyer, 2010)
BRD4(1)	3MUL	1.65	H3K14buty	GK_{but}	(Vollmuth and Geyer, 2010)
BRD4(1)	3UVW	1.49	H4K5K8	SGRG K_{ac} GG K_{ac} GLGY	this study
BRD4(1)	3UVX	1.91	H4K12K16	GK_{ac} GGAK ac RHRKV	this study
BRD4(1)	3UVY	2.02	H4K16K20	AK_{ac} RHR K_{ac} VLRDN	this study
BRD4(1)	3UW9	2.37	H4K8K12	GK_{ac} GLG K_{ac} GGAKR	this study
FALZ	2RI7	1.45	H3K4	H3(1-9) K4_{me2}	(Li et al., 2006)
FALZ	2F6J	2.00	H3K4	H3(1-15) K4_{me3}	(Li et al., 2006)
FALZ	2FSA	1.90	H4K4	H3(1-15) K4_{me2}	(Li et al., 2006)
FALZ	3QZV	2.00	H4K12	H4(7-17) K12_{ac}	(Ruthenburg et al., 2011)
FALZ	3QZS	1.80	H4K16	H4(12-21) K16_{ac}	(Ruthenburg et al., 2011)
FALZ	3QZT	1.50	H4K20	H4(16-25) K20_{ac}	(Ruthenburg et al., 2011)
CREBBP	1JSP	NMR	p53K382	SHLKSKKGQSTSRR K_{ac} LMFK	(Mujtaba et al., 2004)
MLL	3LQI	1.92	H3K4	H3(1-9) K4_{me2}	(Wang et al., 2010)
MLL	3LQJ	1.90	H3K4	H3(1-9) K4_{me3}	(Wang et al., 2010)
PB1(2)	2KTB	NMR	H3K14	ARTKQTARKSTGG K_{ac} APRKQL	(Charlop-Powers et al., 2010)
PCAF	2RNW	NMR	H3K9	ARTKQTARK ac STGGKA	(Zeng et al., 2008)
PCAF	2RNX	NMR	H3K36	STGGV K_{ac} KPHRYKC	(Zeng et al., 2008)
PCAF	2RNY	NMR	H4K20	GGAKRHR K_{ac} VLRDNIQ	(Zeng et al., 2008)
PCAF	1JM4	NMR	HIV/Tat	SYGR K_{ac} KRRQR	(Mujtaba et al., 2002)
TIF1 α	3O34	1.90	H3K23	H3(13-32) K23_{ac}	(Tsai et al., 2010)
TIF1 α	3O35	1.76	H3K27	H3(23-31) K27_{ac}	(Tsai et al., 2010)
TIF1 α	3O36	1.70	H4K16	H4(14-19) K16_{ac}	(Tsai et al., 2010)
TIF1 α	3O37	2.00	H3K4	H3(1-10) K4_{ac}	(Tsai et al., 2010)

Table 4: BRD histone interactions

Acetyl-lysine specific histone interactions identified in the array of peptides with singly acetylated lysine residues. Strong specific spots are shown in bold and weak intensity spots are shown in smaller font size. Interactions confirmed by ITC are underlined, published quantified interactions are marked by a “*” and published but not quantified ones are marked by a “#”. Single acetylation marks identified in the H3 microSPOT array are marked by a “¥”.

BRD	Histone				
	H1-4	H2A	H2B	H3	H4
BRDT(1)	K33, K74, K80, K84, K89	K15, K36 , K74, K75 , K95, K99, K125, K129	K12, K30, K34, K116	K4, K9, K18, K23, K27, K36, K37, K56, K115, K122	K5, K8, K20, K37, K44, K91
BRD4(1)	K31, K33, K62, K63, K121¥	K74 , K75 , K99,	K24, K116, K120, K125,	K4, K5, K18, K23, K27, K36 , K37	K5, K20
BRD2(1)	K74 , K80, K84, K89	K95	K5, K34	K37, K56, K115	<u>K5</u> #, K44, K79
BRD3(1)		K127	K43, K120	K122, K27¥, K64¥	K79
BRD3(2)		K36	K85		
BRD4(2)	K84, K194	K5 , K9, K15, K36 , K74, K75	K43 , K46,	<u>K14</u> , K18 , K36 , K37 , K56 , K64, K115, K122	K5 *, K20 , K44 , K77, K79, K91
TAF1(1)	K74	K15 , K36		K56	
TAF1L(1)		K15, K36	K85		
TAF1(2)	K31, K62, K105	K15 , K75	K116, K120	K18, K56, K115, K64¥	K44 , K77 , K79
TAF1L(2)	K31 , K33, K84, K105, K118, K147, K148	K74, K75	K11, K12, K20, K23, K24, K116, K120, K125	K4, K9, K14, K23, K27,	K5, K8, K77 , K79, K91
CECR2		K15 , K36 , K75		K9¥, K14¥, K18 , K27¥, K36, K56, K64¥, K79¥, K115¥	
FALZ	K25, K31, K33, K62, K80, K84, K105, K185,		K11, K12, K30, K34, K43	K18, K23, K27	<u>K5</u> , K8, K20,
GCN5L2	K31, K33, K138, K139, K174, K175, K189, K191, K194 K196, K199, K211, K212, K216, K218	K5, K9, K125, K127, K129		K4, K9, K36, K37, K56¥	K20
PCAF	K25, K31, K33, K105, K108, K109, K151, K152, K155, K156, K158, K159	K99, K116, K119, K125, K127			K12, <u>K16</u> , <u>K20</u> *, K59
ASH1L	K74	K36	K30, K34, K33, K36, K85 , K120	K37 , K56 , K79, K122	K59 , K79
PB1(3)	K62, K74, K80, K84	K15	K23,	K18*, K36*, K56, K64¥	
PB1(1)	K33, K80, K84, K89	K15		K56, K64, K122	K44, K91
PB1(4)				K37**, K122¥	
PB1(2)		K36	K85	K14*	

PB1(5)			K85		
SMARCA2	K33, K74, K170, K182	K15, K95, K107	K5, K34	<u>K14</u> [#] , K18, K23, K27 [*] , K36, K37 , K56	K5, K8 [#] ,
SMARCA4	K80, K84, K105	K75,	K20, K23, K24	K18 , K23, K27, K36	
PB1(6)	K84		K43	K56, K115 [*]	
MLL		K15			
LOC39249	K22, K128, K135, K136, K138, K139, K158, K159, K170, K173, K174, K176, K182, K185, K211, K212, K216, K217, K218	K119, K125, K127, K129	K27, K30, K34		K5, K8 , K12, K16, K20
SP140				K14 [*] , K23 [*] , K36 [*]	
TRIM28				K37 [*]	
TIF1	K31, K33, K62, K63, K84, K116, K147, K148, K182	K15, K36 , K74, K75, K99,	K12, K20	K4, K9*, K14*, K18, K23*, K27, K36, K37, K56 , K64, K115	K5, K44 , K31, K77, K79, K91
BAZ2B	K89, K98	K36, K74, K129	K5, K11, K12, K34, K85, K108, K116	K14 , K36, K37, K79, K115, K122	K31, K77, K79
PHIP(2)	K31, K33, K62 , K80, K84, K89			K9, K18, K36, K37	K91
ATAD2	K74	K36	K85	K56	
KIAA1240	K74	K5, K15, K36	K43 , K46,	K9, K56 , K115, K122	<u>K5</u> , K59 , K77, K79
BRPF1	K74	K15		K56 , K115, K122	K44 , K77, K79, K91
BRD9				K18 [*] , K14 [*] , K27 [*] , K36 [*] , K56 [*] , K115 [*]	
CREBBP	K74	K15		K34, K43, <u>K36</u> [*] , <u>K37</u> , K56 , K64, K122	<u>K44</u> , K59, K77, K79, K91
EP300	K31, K33, K51, K63, K74 , K116	K5, K15, K36 , K74, K75,	K23, K34, K43 , K46 , K108, K116, K116, K120, K125	K4, K14, K18, K23, <u>K36</u> , K37, K56 , K64, K79 , K115 , K122 ,	K5 , K8, K12, K44 , K77, K79, K91

FIGURE 1

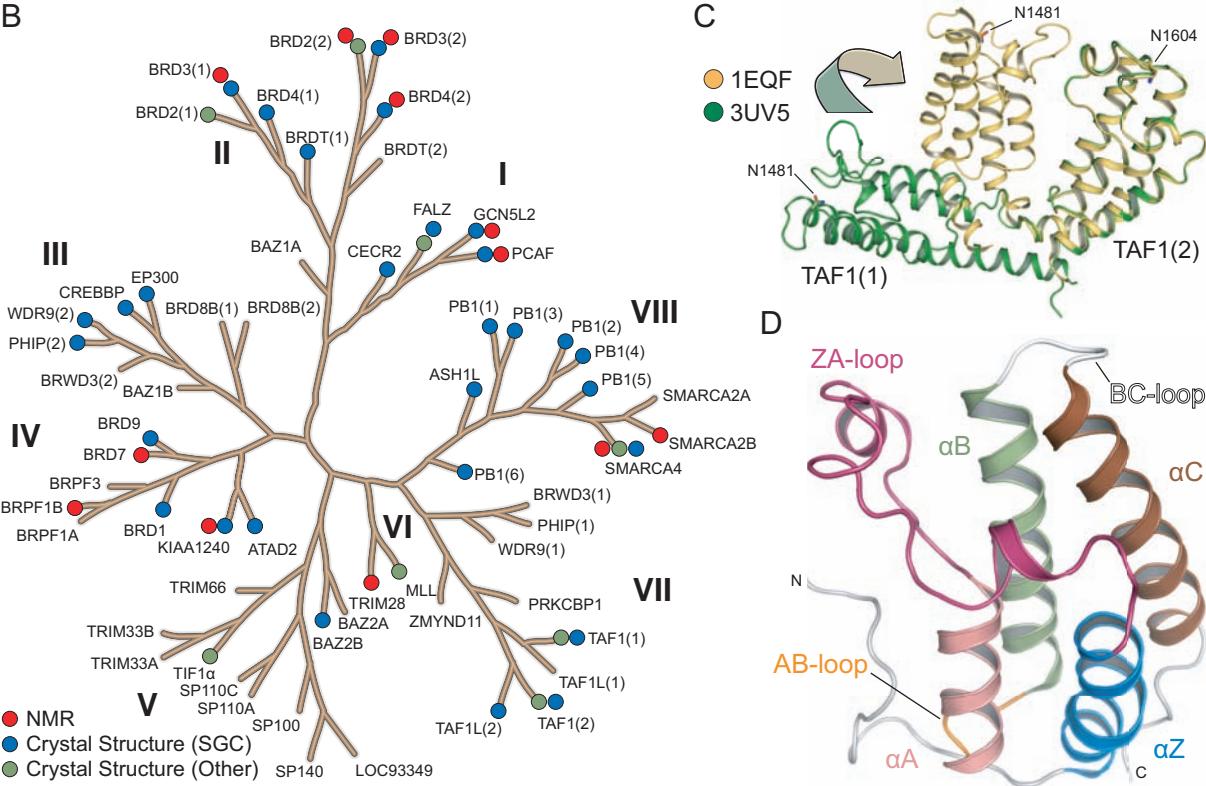
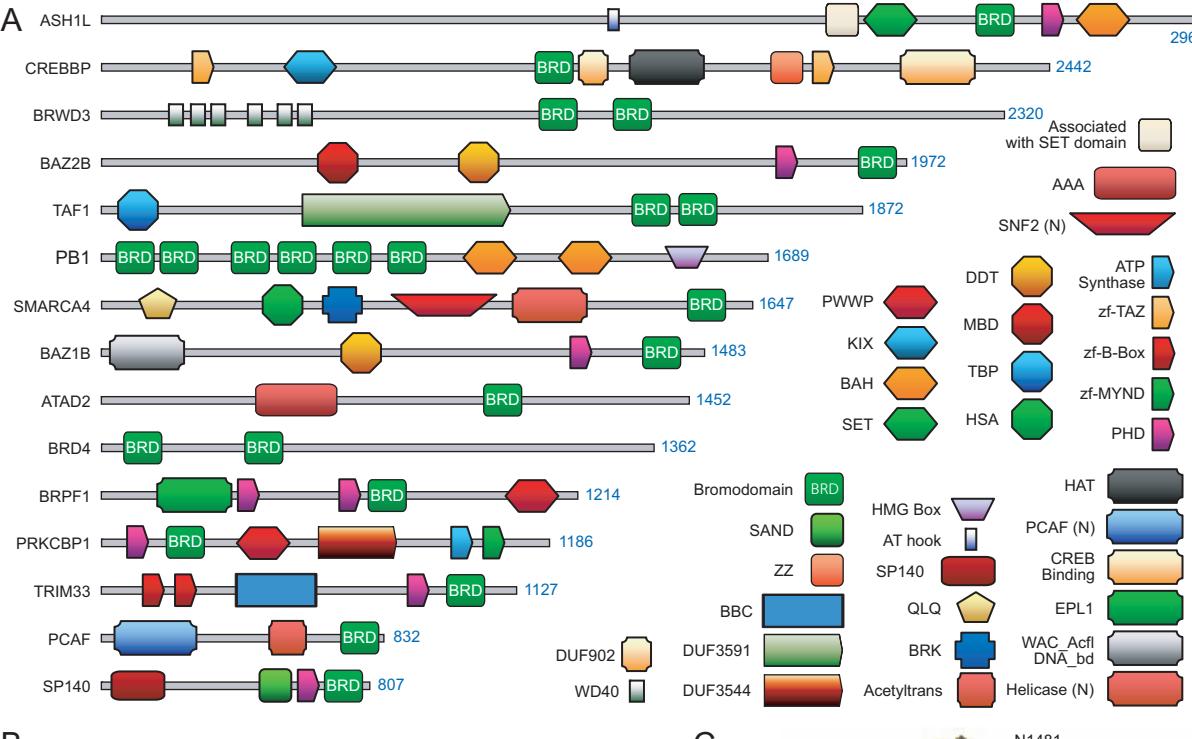


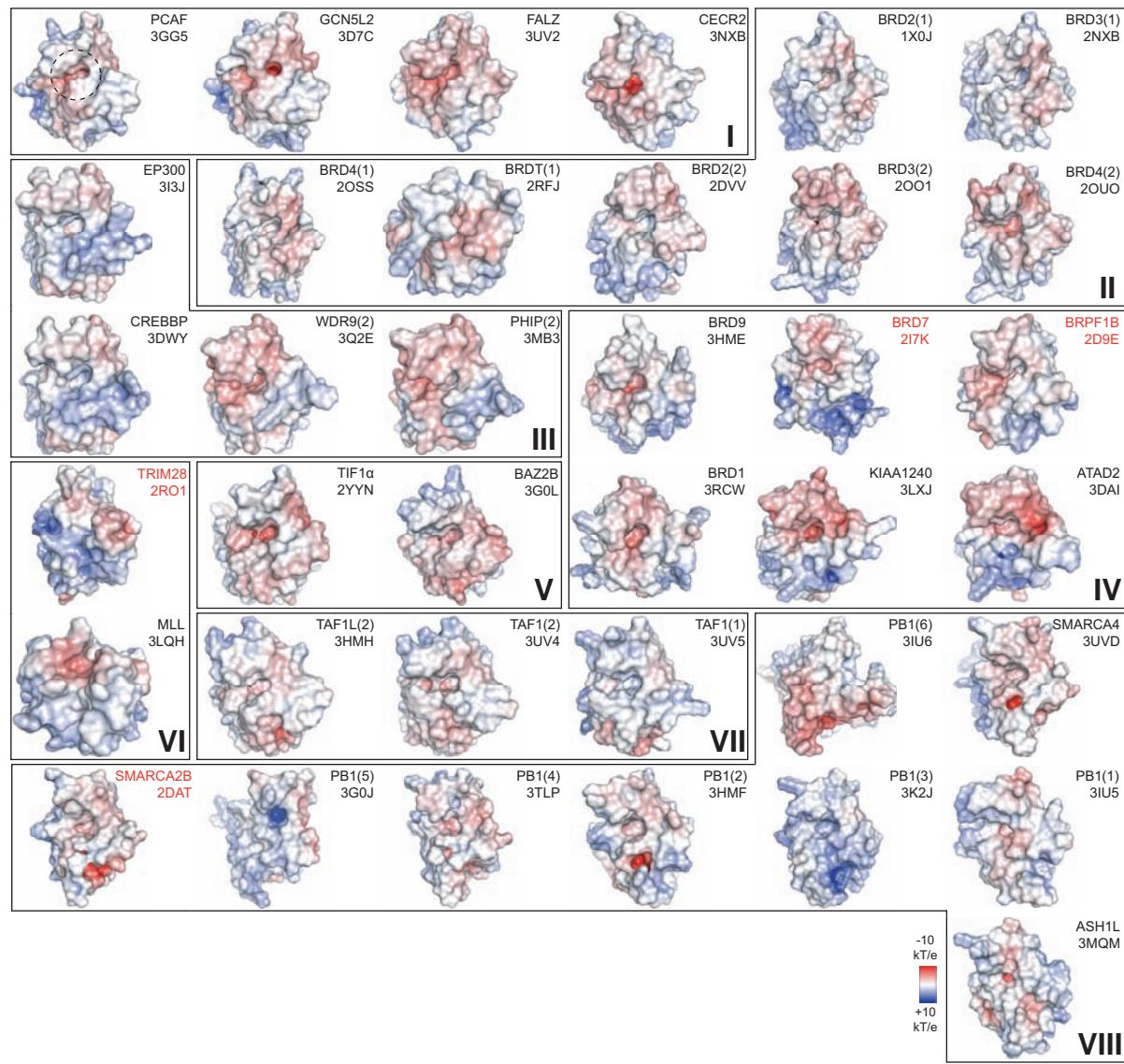
FIGURE 2

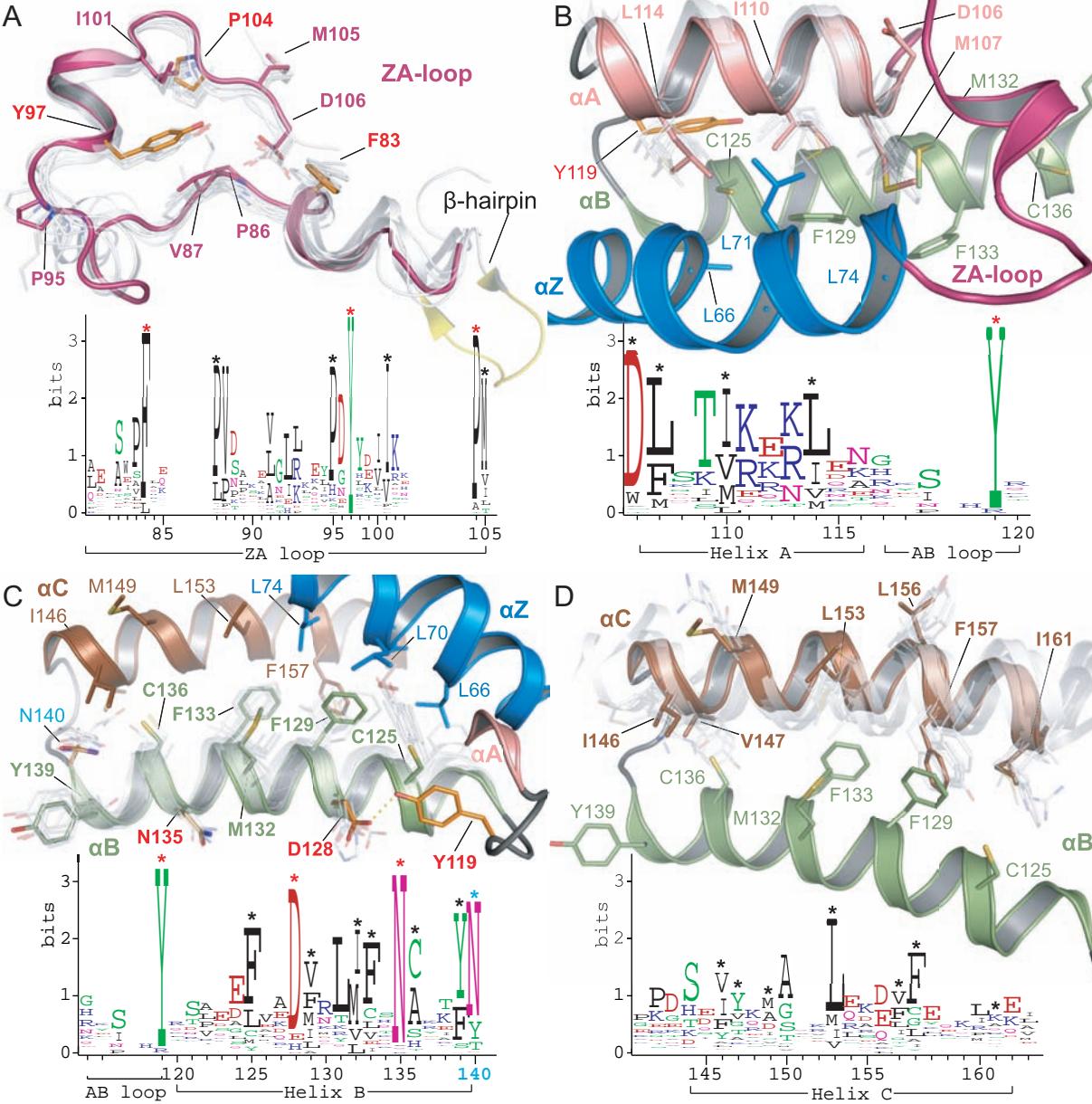
FIGURE 3

FIGURE 4



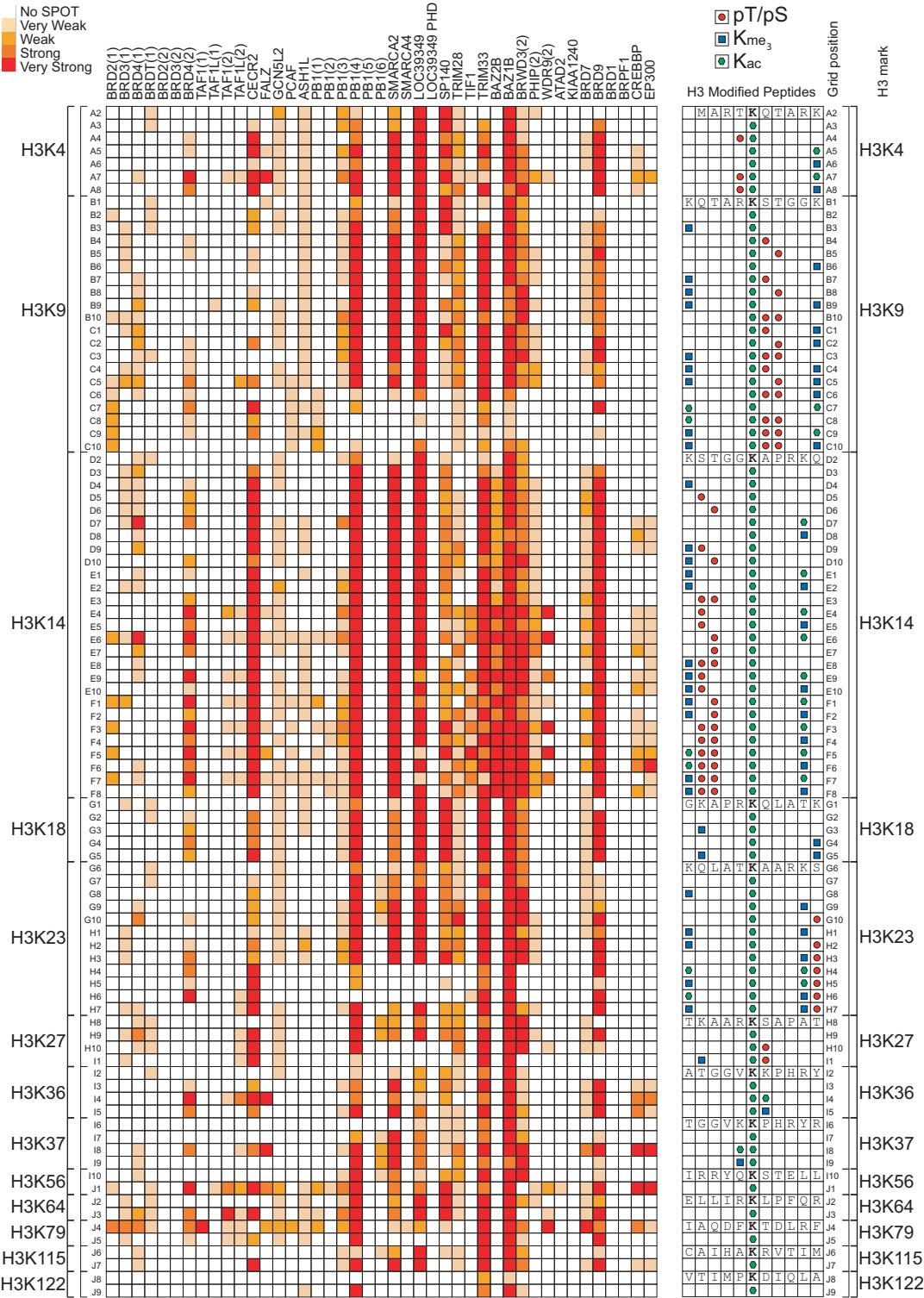
FIGURE 5

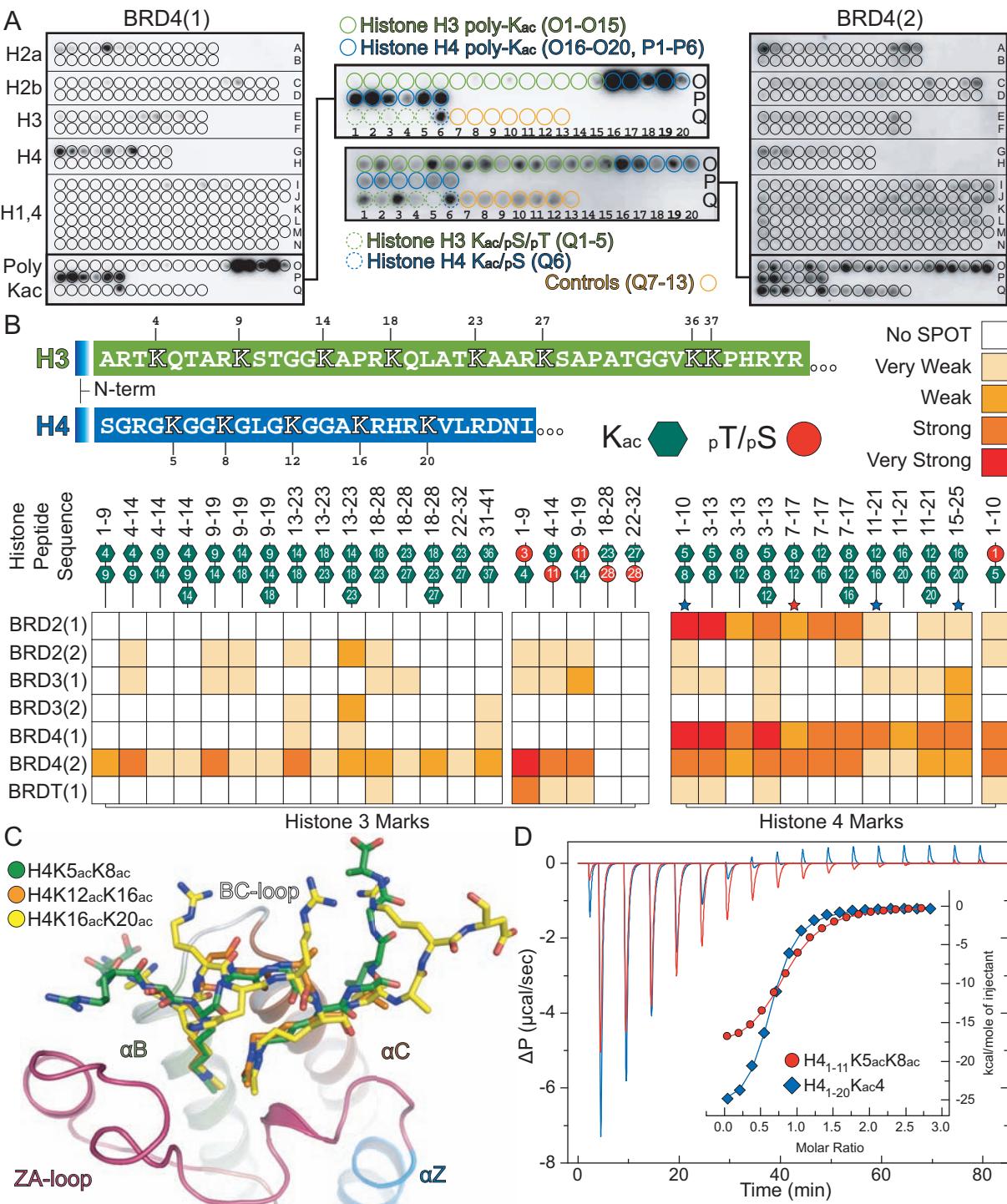
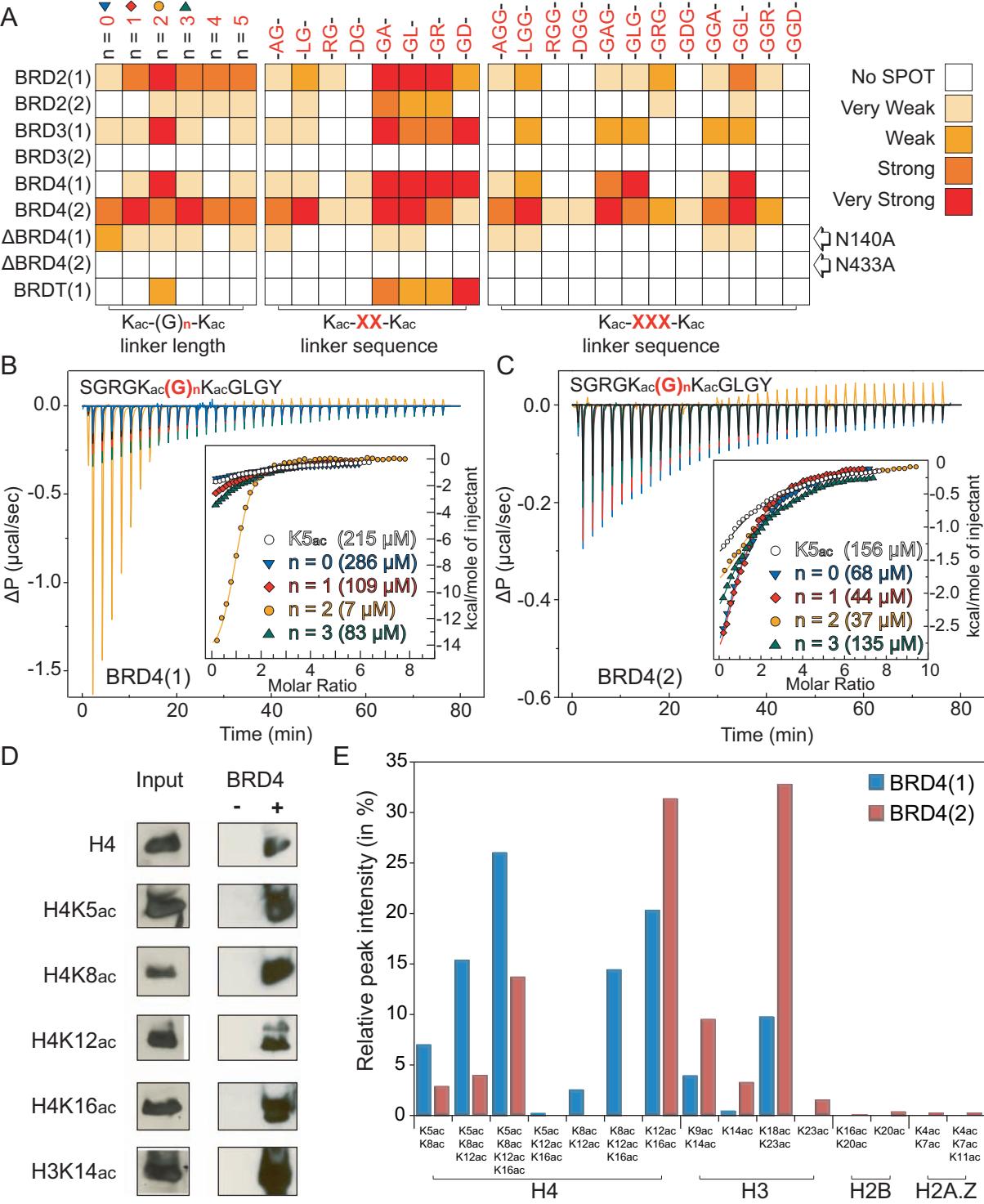
FIGURE 6

FIGURE 7

Supplemental Experimental Procedures

Cloning

cDNA encoding human BRD containing proteins were obtained from different sources. Most of them were synthesized (Genscript) for codon optimization, some were provided by the MGC collection, the IMAGE collection or from commercial sources. The obtained cDNA sequences were used as templates to amplify BRD regions employing the Polymerase Chain Reaction (PCR) in the presence of Platinum® Pfx DNA polymerase (Invitrogen™, UK). All relevant details are listed in **Supplemental Table S2**. PCR products were purified (QIAquick PCR Purification Kit, Qiagen Ltd. UK) and further sub-cloned into pET28 derived expression vectors, pNIC28-Bsa4 (gi|124015065) or pNIC-CTHF (gi|124015079), using ligation independent cloning (Stols et al., 2002). Constructs were transformed into competent Mach1™ cells (Invitrogen™, UK) to yield the final plasmid DNA and were verified by sequencing.

Protein expression and purification for biophysical characterization

Constructs were transformed into competent BL21 (DE3) cells (Invitrogen™) or into BL21 (DE3)-R3-pRARE2 cells (phage-resistant derivative with a pRARE plasmid encoding rare codon tRNAs). Cells were grown at 37 °C either in Luria-Bertani medium (LB-broth, Merck) or in Terrific Broth (Merck) from overnight cultures. Protein expression was induced overnight with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 18 °C at an OD₆₀₀ nm of 0.9 or 3.0 respectively. Cultures were harvested by centrifugation (8,700 x g for 15 min at 4 °C) on a Beckman Coulter Avanti J-20 XP centrifuge, and then re-suspended in lysis buffer (50 mM HEPES, pH 7.5 at 20 °C, 500 mM NaCl, 5 mM Imidazole, 5 % glycerol and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) in the presence of 1:200 (v/v) Protease Inhibitor Cocktail III (Calbiochem). Cells were lysed at 4 °C using an EmulsiFlex-C5 high pressure homogenizer (Avestin - Mannheim, Germany) and the DNA was removed by precipitation on ice for 30 min with 0.15 % (v/v) of PEI (Polyethyleneimine). Lysates were cleared by centrifugation (16,000 x g for 1h at 4 °C, JA 25.50

rotor, on a Beckman Coulter Avanti J-20 XP centrifuge) and were applied to a Nickel affinity column (nickel nitrilotriacetic acid (Ni-NTA) resin - Ni-NTA, Qiagen Ltd., 5 ml, equilibrated with 20 ml lysis buffer). Columns were washed once with 30 ml of lysis buffer then twice with 10 ml of lysis buffer containing 30 mM Imidazole. Proteins were eluted using a step elution of imidazole in lysis buffer (50, 100, 150, 2 x 250 mM Imidazole). All fractions were collected and monitored by SDS-polyacrylamide gel electrophoresis (Bio-Rad Criterion™ Precast Gels, 4-12 % Bis-Tris, 1.0 mm, from Bio-Rad, CA. Gel run conditions: 180 V, 400 mA, 55 min in XT MES buffer). The eluted proteins were treated overnight at 4 °C with TEV (Tobacco Etch Virus) protease to remove the hexa-histidine expression tag and were further purified by size exclusion chromatography on Superdex 75 16/60 HiLoad gel filtration columns (GE/Amersham Biosciences) on ÄktaPrime™ plus systems (GE/Amersham Biosciences). Proteins in 10 mM HEPES, 500 mM NaCl and 5 % glycerol were concentrated to 10 mg/ml with a Amicon® Ultra (MILLIPORE) concentrators employing 10 kDa cut-offs (10 MWCO), flash frozen in liquid nitrogen and stored at -80 °C. Protein concentration was estimated using a NanoDrop ND-1000 spectrophotometer (also see **Supplemental Table S3**).

Protein expression and purification for spot assays

Soluble BRD constructs were expressed in LB media (5 x 50 ml) in the presence of 50 µg/ml kanamycin. Cell growth was allowed at 37 °C to an optical density of about 0.5 (OD_{600nm}). Protein expression was induced by 1 mM IPTG, overnight, at 18 °C. Cells were harvested by centrifugation at 4000 x g for 15 min at 4 °C and re-suspended in 4 ml of binding buffer (50 mM HEPES pH7.5 at 20°C; 500 mM NaCl; 5 % glycerol, 5 mM Imidazole) complemented with 0.5 mM TCEP and 1:200 (v/v) Protease Inhibitor Cocktail III (Calbiochem). Cells were disrupted with an ultrasonic processor (SONICS Vibra-Cell™, amplitude 60, 10 sec ON, 10 sec OFF, for 2 min). Lysates were cleared by centrifugation and were applied to Ni-NTA columns (Qiagen Ltd., 2 ml, equilibrated with 20 ml lysis buffer). Columns were washed once with 30 ml of lysis buffer then twice with 10 ml of the same buffer containing 30 mM Imidazole. Proteins were eluted using a step elution of imidazole in lysis buffer (50, 100, 150, 2 x 250 mM Imidazole). All fractions were

collected and monitored by SDS-polyacrylamide gel electrophoresis. Fractions containing the recombinant proteins were pooled and concentrated using Amicon® Ultra (MILLIPORE) concentrators (10 MWCO) to a final volume of 1 ml before being loaded on NAP™-10 column (GE-Healthcare) in order to exchange the buffer to 25 mM HEPES (pH 7.5 at 20 °C), 150 mM NaCl and 5 % glycerol. Samples were flash frozen in liquid nitrogen and stored at -80 °C until used (**Supplemental Table S3**).

SPOT assays

Three different types of SPOT membrane were utilized in order to probe preferences of BRD recognition sites: the first array contained peptides from all four core histones (H2A, H2B, H3 and H4) covering all possible lysine acetylation sites, with peptide sizes ranging from 10 to 14 amino acids. The second array contained short peptides (11 residues) from human histone 3 exploiting possible post-translational modifications around each acetyl-lysine epitope, including phosphorylation (on serine and threonine residues), acetylation (on lysine residues) and trimethylation (on lysine residues). The third array covered all possible lysine acetylation sites found on the four core human histones, as well as peptides containing multiple adjacent acetyl-lysine epitopes, employing a microSPOT (μ SPOT) technology using a smaller foot-print.

Membranes synthesis

Peptides for the first two arrays were synthesized directly on cellulose membranes (Intavis) using a MultiPep SPOT peptide arrayer (Intavis) and commercially available standard (Intavis) or modified (Bachem) L-amino acid precursors as previously described (Nady et al., 2008). The quality of the synthesized array was evaluated as follows: i) immune reactive peptides were identified by incubation of the membrane with an anti-His antibody in the absence of any His6-tagged BRD. 8 peptides were recognized by this antibody and were excluded from the analysis. In all cases a general monoclonal primary antibody against acetylated lysine (#9681, Cell Signaling Technology) was used to probe the proper incorporation of acetylated lysine on each membrane, using the protocol provided by the manufacturer for Western blotting.

Phosphorylation was probed using a primary antibody against phosphorylated Ser10 (ab47297, Abcam) and Thr11 (ab5168, Abcam), employing an anti-rabbit HRP fragment secondary antibody (Amersham Biosciences).

Protein-peptide interaction assay

Membranes were washed 3 x 5 min with PBST (3.2 mM Na₂HPO₄, 0.5 mM KH₂PO₄, 1.3 mM KCl, 135 mM NaCl and 0.1 % Tween 20, pH 7.4) and were subsequently blocked with 5 % milk in PBST overnight at 4 °C in order to minimize non-specific binding of the proteins to the membranes. After 2 washes with PBST (5 min each) followed by a single wash with PBS (3.2 mM Na₂HPO₄, 0.5 mM KH₂PO₄, 1.3 mM KCl, 135 mM NaCl pH 7.4) for 5 min, his-6 tagged BRD proteins were added to a final concentration of 1 µM and the membranes were incubated over night at 4 °C in PBS. Each membrane was washed 3 times in PBST, blocked for 1 h with 10 % milk in PBST, and washed again 3 x 5 min with PBST. HPR-conjugated anti-His-tag antibodies (Novagen #71841) were added in 5 % milk/PBST solution at a dilution of 1:2000. After 1 h incubation, membranes were washed 3 x 20 min in PBST. The assay was developed with an ECL kit (Pierce ECL Western Blotting substrate, Thermo Scientific) following the manufacturer's protocol. Chem-illuminescence was detected with an Image reader (Fujifilm LAS-4000 ver.2.0) with an incremental exposure time of 5 min for a total of 80 min. Intensities of the resulting spots were quantified with the Kodak 1D ver.3.6.2 Scientific Imaging System. All experiments were performed at room temperature.

Membrane striping

Membranes were washed 3 x 10 min in water, 2 x 30 min in stripping buffer A (6 M Guanidinium HCl, 1 % Triton X-100) followed by an overnight incubation with Stripping Buffer A and Talon Beads at room temperature. The next day each membrane was washed twice with Stripping Buffer B (500 mM Imidazole, 500 mM NaCl, 20mM TRIS-HCl, pH 7.5) for 30 min each, followed by a series of washes with deionized and distilled water (ddH₂O) at room temperature and at 60 °C. Finally, a series of washes was performed, 10 min each with ddH₂O, 10 % TFA

(trifluoroacetic acid), ddH₂O, 20 % EtOH, 50 % EtOH and 95 % EtOH. Membrane were dried overnight and stored for extended periods of time at -20 °C until they were re-used.

Crystallization and Structure Determination

Individual proteins were crystallized in sitting drops at either 4 °C or 20 °C. Crystals were cryo-protected, flash frozen and X-ray diffraction data were collected at 100 K on beam lines X10SA at the Swiss Light Source (SLS), at Diamond (beam lines I02, I03, I04, I04.1), or at a Rigaku FRe Superbright home source. Diffraction images were indexed, and integrated using MOSFLM (Leslie and Powell, 2007), HKL2000 or XDS (Kabsch, 2010b) and data were scaled using SCALA (Evans, 2007), SCALEPACK (Otwinowski and Minor, 1997), or XSCALE (Kabsch, 2010a), respectively. Structures were solved by molecular replacement using PHASER (McCoy et al., 2005) and were refined against maximum likelihood targets using REFMAC (Murshudov et al., 1997). Iterative rounds of refinement were interspersed with manual rebuilding in COOT (Emsley and Cowtan, 2004). Thermal motions were analyzed using TLSMD (Painter and Merritt, 2006) and hydrogen atoms were included in late refinement cycles. Crystallization conditions, data collection and refinement statistics, PDB accession codes are compiled in **Supplemental Table S4**.

Structure-based alignment and sequence analysis

Multiple sequence/structural alignments were carried out using STRAP (Gille and Frommel, 2001) and ICM Pro (MolSoft LLC version 3.7-2c) (Abagyan et al., 1994) and were further manually edited (**Supplemental Figure S1A**). In the absence of an X-Ray or NMR structure model the PFAM boundaries for each BRD were extended using PSIPRED (version 2) (Jones, 1999) for secondary structure prediction and the secondary structure elements were further used to guide the STRAP/ICM alignment. A phylogenetic tree of the resulting structure based alignment was generated using the ClustalW2 program (Larkin et al., 2007) and is given in **Figure 1B** and **Supplemental Figure S1B**. Sequence conservations were visualized using the WebLogo (Crooks et al., 2004) online web server.

Isothermal Titration Calorimetry (ITC)

Experiments were carried out on a VP-ITC microcalorimeter or an ITC200 (MicroCal™, LLC Northampton, MA). All experiments were performed at 10 or 15 °C in 50 mM HEPES pH 7.5, 150 mM NaCl. All titrations were conducted using an initial injection of 2 µl followed by 29 identical injections of 8 µl (VP-ITC) or 0.3 µl followed by identical injections of 1 µl (ITC200). The dilution heats were measured on separate experiments and were subtracted from the titration data. Thermodynamic parameters were calculated using $\Delta G = \Delta H - T\Delta S = -RT\ln K_B$, where ΔG , ΔH and ΔS are the changes in free energy, enthalpy and entropy of binding respectively. In most cases a single binding site model was employed, supplied with the MicroCal™ Origin software package. Multiple binding events were also confirmed with the software package SEDPHAT (Houtman et al., 2007). Binding constants and thermodynamic parameters are given in **Supplemental Tables S5** (Single K_{ac} marks), **S6** (multiple marks) and **S7** (linker sequences).

Histone immunoprecipitation

HEK293 cells were grown in DMEM with 10 % FBS and transfected with Brd4-Flag (UniProt: O60885, residues 1-1362, cloned on pCDNA5) vector using GeneJuice (EMD) according to the manufacturer's instructions. The transfected cells were treated with 500 nM (+)-JQ1 or (-)-JQ1 (Filippakopoulos et al., 2010) for 16 h and cells collected by scraping. Nucleosome isolation protocol was based on Ruthenburg and co-workers (Ruthenburg et al., 2011) with modification at the end of the nuclease treatment. Briefly the cells were washed in buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM MgCl₂, 340 mM sucrose, 10 % (v/v) glycerol, protease inhibitors (Roche) TSA 1 µg/ml, beta-mercaptoethanol 5 mM. The cell pellet was resuspended in buffer A with 0.1 % Triton x-100 and incubated on ice for 10 min. The nuclei were pelleted, washed twice with buffer A and resuspended in buffer A to nucleic acid concentration of 1.2 µg/µl measured as previously described (Brand et al., 2008). CaCl₂ was added to 2 mM, followed by the micrococcal nuclease (Worthington) (1 U/50 µg DNA) and the reaction incubated at 37 °C for 10 min. The reaction stopped adding 4 mM EGTA on ice. To facilitate the release of digested nucleosomes, NaCl was added to final concentrations of 200 mM and reactions spun down at

13,000 rpm for 5 min. The soluble nucleosomes in the supernatant were collected and diluted (1:5) in the IP buffer (100 mM KCl 5 % glycerol, 10 mM TRIS-HCl pH 8.0, 10 mg/ml BSA, protease inhibitors, 1 µg/ml TSA, beta-mercaptoethanol 5 mM). Anti-Flag M2 antibody (Sigma) or normal mouse IgG (Abcam) and protein G Dynabeads (Invitrogen) were added and incubated at 4 °C overnight. The reactions were washed 5 times in the IP buffer and eluted with 100 mM TRIS-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 5 mM DTT, 1 % Triton, 3 % SDS 15 mM beta-mercaptoethanol. The immunoprecipitated proteins were run on Tris-Bis PAGE (Invitrogen) and transferred to nitrocellulose membranes (Pall) that were probed with the following antibodies: H4K5_{ac} (Abcam 61236), H4K12_{ac} (Upstate 07-595), H4 (Abcam 7311), H3K9_{ac} (Abcam 12179), H4K16_{ac} (Active Motif 39167), H4K18_{ac} (Cell Signaling Technology 2594) and secondary HRP conjugated antibodies (Cell Signaling Technology).

Preparation of asynchronous and M phase HeLa cells

HeLa cells were grown in DMEM supplemented with 10 % fetal bovine serum and antibiotics (Penicillin-Streptomycin cocktail). In order to generate cells arrested in M phase, cells at ~50 % confluence were first treated with 2 mM thymidine for 24 h; the media was then removed, the cells washed with 1X PBS, and fresh media added for three hours to allow release from the thymidine block. Nocodazole (100 ng/ml) was then added for 16 h to arrest cells in M phase. The cells were then harvested by “mitotic shake-off” to harvest rounded mitotic cells which were then washed with cold PBS and frozen as a dry pellet at -80 °C.

High salt extraction of histones

Histones were extracted from asynchronous or M phase cells using high-salt as previously described (Shechter et al., 2007) with minor modifications. Briefly, frozen cell pellets from 20 x 15 cm plates of HeLa cells (corresponding to approximately 4 x 10⁸ cells) were resuspended in 10 ml of cold extraction buffer (10 mM HEPES-NaOH pH 8.0, 10 mM KCl, 1.5 mM MgCl₂, 0.34 M sucrose, 10 % glycerol, 0.1 mM PMSF, 10 mM sodium butyrate) with 0.2 % NP-40 and incubated on ice for 10 min with occasional mixing. Samples were then centrifuged (6,500 x g, 5

min at 4 °C) and the supernatant discarded. An additional 10 ml of cold extraction buffer without NP-40 was used to resuspend the pellet, and the mixture was centrifuged (6,500 x g, 5 min at 4 °C). The supernatant was completely removed and the pellet resuspended in 10 ml of no-salt buffer (3 mM EDTA, 0.2 mM EGTA), vortexed intermittently for 2 min and further incubated at 4 °C on a nutator for 30 min. The samples were then centrifuged (6,500 x g, 5 min at 4 °C) and the supernatant discarded. Histones were extracted by vortexing the pellet in 10 ml of high-salt solubilisation buffer (50 mM TRIS-HCl pH 8.0, 2.5 M NaCl, 0.5 % NP-40) for 2 min, followed by incubation at 4 °C on a nutator for 30 min. DNA was pelleted by centrifugation (16,000 x g, 10 min at 4 °C) and the supernatant containing histones transferred to a fresh tube. Three buffer exchanges were performed with a spin filter device of 5000 Da molecular weight cut-off to reduce salt concentration before storing at -80 °C until further use.

Biotin tagged bromodomains affinity purification

500 µg of salt-extracted histones from asynchronous or M phase HeLa cells were mixed with 50 µg of recombinant biotinylated BRD4(1) or BRD4(2) and incubated at 4 °C on a nutator for 60 min in the presence of 1 µM (+)-JQ1 or 1 µM (-)-JQ1. Samples were transferred to a fresh tube containing 30 µl of streptactin sepharose beads (IBA BioTAGnology) and incubated for 60 min at 4 °C on a nutator. The beads were then washed five times with 1 ml of wash buffer (50 mM HEPES-NaOH pH 8.0, 500 mM KCl, 2 mM EDTA, 0.1 % NP-40, 10 % glycerol) and two times with 1 ml of no-salt wash buffer (20 mM TRIS-HCl pH 8.0, 2 mM CaCl₂). Bound histones were eluted by incubating the beads in 1 ml of 0.5 % Trifluoroacetic acid (TFA) at 4 °C on a nutator. The supernatants were transferred to fresh tubes and then evaporated to dryness and stored at -80 °C.

Protein digestion and acetylated peptide purification

Dried sample were resuspended in 100 µl of 20 mM TRIS-HCl pH 8.0, and 1 µg of trypsin (Sigma-Aldrich; Singles) was added to each sample. Samples were incubated overnight at 37 °C with agitation, and supplemented with an extra 0.5 µg of trypsin before another incubation of 4 h.

The samples were then boiled for 10 min to inhibit trypsin activity and subsequently let to cool down to room temperature for approximately 15 min. Samples were then diluted to 400 μ l with peptide wash buffer solution (50 mM MOPS pH 7.2, 10 mM NaPO₄, 50 mM NaCl) and incubated with 30 μ l of anti-acetyl lysine agarose beads (ImmuneChem Pharmaceuticals Inc.) overnight at 4 °C. The next morning, the beads were collect by gentle centrifugation and the supernatant (unbound fraction) transferred to a fresh tube for subsequent analysis by mass spectrometry. The beads were washed once with 1 ml of peptide wash buffer solution and once with 1 ml of no-salt wash buffer (20 mM TRIS-HCl pH 8.0, 2 mM CaCl₂). Peptides were eluted by incubating the beads with 1 ml of 0.5 % TFA for 30 min at 4 °C and then transferred to a fresh tube before being evaporated to dryness.

LC-MS/MS analysis and quantitation

Dried peptides were dissolved in 5 % formic acid and analyzed by LC-MS/MS using a NanoLC-Ultra 2D plus HPLC system (Eksigent, Dublin, USA) coupled to a LTQ-Orbitrap Velos (Thermo Electron, Bremen, Germany) equipped with a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark). The LTQ-Orbitrap Velos instrument under Xcalibur 2.0 was operated in the data dependent mode to automatically switch between MS and up to 10 subsequent MS/MS acquisition. Raw MS and MS/MS spectra were processed using Prohits (Liu et al., 2010). Peptides and proteins were identified using the Mascot software (Matrix Science, London, UK) and the human RefSeq database (version 45, released on February 2nd 2010, containing 34604 sequences). Mass tolerance of 7 ppm in MS mode and 0.6 Da in MS/MS mode with trypsin specificity were used, and 4 missed cleavage sites were allowed. No fixed modification was selected, but N-acetyl protein, N-pyroglutamine, oxidized methionine, acetylation of lysine and phosphorylation of serine, threonine and tyrosine were searched as variable modifications. Relative quantitation of acetylated peptides using MS spectra was achieved with Proteome Discoverer 1.2 (Thermo Electron, Bremen, Germany). The peak area for acetylated peptides co-purifying with BRD4(1) or BRD4(2) were calculated first by summing the areas under the curve for a given acetylated peptide in the asynchronous or nocodazole-

arrested cells (for a given bromodomain), then by computing the area of all acetylated peptides associated with either bromodomain. The relative peak intensity of each acetylated peptide was then determined by expressing (in percent) the ratio of the area under the peak of a given peptide peak over the total area under the peak of all peptides. The efficiency of purification with each bromodomain was monitored by analyzing the fraction unbound to the anti-acetyl lysine beads; specificity in the interaction was ascertained by analyzing by quantitative MS each of the samples incubated with the inhibitors (-)-JQ1 and (+)-JQ1.

Supplemental Figure Legends

Supplemental Figure S1A: Structure guided sequence alignment of the Human bromodomain family.

Sequences are clustered by BRD families, highlighted by different colours corresponding to the 8 families (Roman numbers I-VIII). The sequence region used for the generation of the phylogenetic tree is indicated by arrows. Location of bromodomain structural elements are shown and named on the top of the figure. Helices are indicated with solid black boxes (extracted from crystal structures) or red dashed boxes (predicted using PSIPRED) (Jones, 1999). Available representative structures (pdb-accession codes) are shown on the right. The first bromodomain of BRD4 (bold and highlighted by an arrow) has been chosen as a reference sequence and the corresponding numbering is shown on top of the alignment.

Supplemental figure S1B: Phylogram of the human bromodomain family.

The 8 families are distinguished by colours and have been labelled by Roman numbers I-VIII. The phylogram was generated using the structural alignment (**Supplemental Figure S1A**), secondary structure prediction and manual curation as well as the ClustalW2 program (Larkin et al., 2007).

Supplemental Figure S2: Overall structure of human bromodomains.

A: Ribbon diagrams of all publicly available crystal (black labels) and NMR (red labels) structures. **B:** Structural Overlay of ASH1L and the reference structure of BRD4(1). The hairpin insert, a hallmark of bromodomains of family VIII is highlighted and details are shown in the expanded view on the right panel.

Supplemental Figure S3: Structural Conservation in helix α_Z and the ZA loop region.

A: Structural overview of the BRD4(1) bromodomain. Secondary structure elements and residue labels are coloured as follows: α_Z blue, α_A salmon, α_B green, α_C brown, ZA loop magenta. **B:** Sequence conservation in helix α_Z . **C:** Anchoring of the ZA loop to the bromodomain core.

Shown are two orientations (left and right panel) of the ZA loop and its interaction with the core structure (shown as surface). The surface is coloured according to the structural elements that harbour the depicted surface residues as described in panel A. **D:** Sequence conservation of for the entire family of human BRDs. Regions of low or no conservation are annotated (in red), highlighting the advantage of employing structural and structure prediction data to define the structurally conserved BRD module.

Supplemental Figure S4: SPOT membrane layout of Figure 4 and representative membranes for each family of BRDs.

Top panel: Core (H2A, H2B, H3 and H4) and linker (H1-4) histone sequences used in the SPOT membranes of Figure 4 are shown with K_{ac} residues numbered (in the template) and highlighted (on the sequences) with a green hexagon. Boxes without a number represent control (non-acetylated) peptides for the histone marks following in the array. **Lower panel:** Representative membranes for each sub-family are given as follows: family I - FALZ, family II - BRDT(1), family III - EP300, family IV - BRPF1A, family V - LOC93349, family VI - MLL, family VII - TAF1L(2) and family VIII - SMARCA2A.

Supplemental Figure S5: microSPOT membrane layout of Figure 5 and representative membranes for each family of BRDs.

Top Panel: The N-terminal sequence of human histone 3 is shown with highlighted marks that were studied for cross-talk, including lysine acetylation and (tri-)methylation as well as threonine and serine phosphorylation. Peptide sequences are given in the inset of **Figure 5**. The membrane layout highlights the position of central epitopes (numbered). The blow-up shows as an example the arrangement of marks around the H3K14_{ac} epitope. **Lower panel:** Representative membranes for each sub-family are given as follows: family I - CECR2, family II - BRD4(2), family III - WDR9(2), family IV - BRD9, family V - BAZ2B, family VI - TRIM28, family VII - TAF1(1) and family VIII - PB1(4).

Supplemental Figure S6: Binding of BRDs to di-acetylated H4 peptides.

A: Detail from the crystal structure of H4₁₋₁₁K5_{ac}K8_{ac} binding to BRD4(1). The protein surface has been coloured according to its electrostatic properties and key residues are annotated. **B:** Detail of H4₁₁₋₂₁K12_{ac}K16_{ac} binding to the surface of BRD4(1). **C:** Detail of H4₁₅₋₂₅K16_{ac}K20_{ac} binding to the surface of BRD4(1). **D:** Binding of two BRD4(1) modules to the H4₇₋₁₇K8_{ac}K12_{ac} peptide. **E:** The H4₁₋₁₁K5_{ac}K8_{ac} peptide from the BRD4(1) complex shown in **A** is superimposed on each of the available BRD structures demonstrating the feasibility of these modules to bind to di-acetylated peptides with a -GG- linker between K_{ac} marks. Proteins are shown as van-der-Waal surfaces (white) with residues that clash with the peptide backbone coloured in red and are clustered in sub-families as shown in **Figure 1B**.

Supplemental Figure S7: SPOT membrane layout of Figure 7 and membranes for each BET bromodomain.

Membrane layout (top) of the peptides used to probe the effect on binding of a poly-glycine linker and of different residue properties between two acetyl lysine marks on histone 4. Stained membranes of BET family bromodomains and the two inactive mutants of BRD4 are given in the lower panel.

Supplemental Tables

Supplemental Table S1: Human Bromodomain family

Protein	Name	Alias	Protein Function	Sub-cellular Localization*	UniProt ID	Reference
ASH1L	ash1 (absent, small, or homeotic)-like	ASH1, KMT2H	Methyltransferase	N, C	Q9NR48	(Gregory et al., 2007)
ATAD2	Two AAA domain containing protein	ANCCA	Transcriptional regulator	N	Q6PL18	(Ciro et al., 2009)
BAZ1A	Bromodomain adjacent to zinc finger domain, 1A	ACF1, WALp1, WCRF180	Chromatin remodeling factor	N	Q9NRL2	(Racki et al., 2009)
BAZ1B	Bromodomain adjacent to zinc finger domain, 1B	WSTF, WBSCR9	Chromatin remodeling factor, Transcriptional regulator	N	Q9UIG0	(Cavellan et al., 2006)
BAZ2A	Bromodomain adjacent to zinc finger domain, 2A	TIP5, WALp3	Transcriptional repressor	N, C	Q9UIF9	(Zhou et al., 2009)
BAZ2B	Bromodomain adjacent to zinc finger domain, 2B	WALp4	Unknown	N, C	Q9UIF8	(Jones et al., 2000)
BRD1	Bromodomain containing protein 1	BRL, BRPF2	Transcriptional regulator, Scaffold protein	N, C	O95696	(Ullah et al., 2008)
BRD2	Bromodomain containing protein 2	FSH, RING3	Transcriptional regulator	N	P25440	(LeRoy et al., 2008)
BRD3	Bromodomain containing protein 3	ORFX, RING3L	Transcriptional regulator	N	Q15059	(LeRoy et al., 2008)
BRD4	Bromodomain containing protein 4	CAP, MCAP, HUNK1	Transcriptional regulator	N	O60885	(Yang et al., 2008)
BRD7	Bromodomain containing 7	BP75, NAG4, CELTIX1	Transcriptional regulator	N	Q9NPI1	(Kaeser et al., 2008)
BRD8B	Bromodomain containing 8 B	SMAP, SMAP2	Transcriptional regulator	N	Q9H0E9-2	(Cai et al., 2003)
BRD9	Bromodomain containing 9		Unknown	N, C	Q9H8M2	NA
BRDT	Bromodomain, testis-specific	BRD6	Chromatin remodeling factor	N	Q58F21	(Moriniere et al., 2009)
BRPF1	Bromodomain and PHD finger-containing protein 1A	BR140, Peregrin	Transcriptional activator	N, C	P55201-1	(Laue et al., 2008)
BRPF3	Bromodomain and PHD finger containing, 3		Unknown	N	Q9ULD4	NA
BRWD3	Bromo domain-containing protein disrupted in leukemia	BRODL	JAK-STAT signaling	N, C	Q6RI45	(Muller et al., 2005)
CECR2	Cat eye syndrome chromosome region		Chromatin remodeling factor	N	Q9BXF3	(Fairbridge et al., 2010)
CREBBP	CREB binding protein	CBP, KAT3A	Histone acetyl transferase	N	Q92793	(Kalkhoven, 2004)
EP300	E1A binding protein p300	p300, KAT3B	Histone acetyl transferase	N	Q09472	(Kalkhoven, 2004)
FALZ	Fetal Alzheimer antigen	BPTF, FAC1	Transcription factor	N	Q12830	(Li et al., 2006)
GCN5L2	General control of amino-acid synthesis 5-like 2	KAT2A, GCN5	Histone acetyl transferase	N	Q92830	(Yang et al., 1996)
KIAA1240	KIAA1240 protein	ATAD2B	Not known	N	Q9ULI0	NA
LOC93349	SP140 -like	SP140L	Unknown	U	Q13342	NA
MLL	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)	HRX, TRX1, CXXC7, ALL-1	Histone Methyl transferase	N	Q03164	(Dou et al., 2005)
PB1	Polybromo 1	PBRM1, BAF180	Chromatin remodeling factor	N, C	Q86U86	(Xue et al., 2000)

PCAF	P300/CBP-associated factor	KAT2B	Histone acetyl transferase	N	Q92831	(Dhalluin et al., 1999)
PHIP	Pleckstrin homology domain interacting protein	WDR11, ndrp	Insulin signalling	N	Q8WWQ0	(Podcheko et al., 2007)
PRKCBP1	Protein kinase C binding protein 1	ZMYND8, RACK7	Transcriptional regulator	N	Q9ULU4	(Fossey et al., 2000)
SMARCA2	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin a2	BRM, SNF2L2	Chromatin remodeling factor, Splicing regulator	N	P51531	(Harikrishnan et al., 2005)
SMARCA4	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin a4	BRG1, SNF2L4, SNF2LB	Chromatin remodeling factor	N	P51532	(Rada-Iglesias et al., 2011)
SP100	Nuclear antigen Sp100		Transcriptional regulator	N, C	P23497	(Yordy et al., 2005)
SP110	Nuclear antigen Sp110 A Nuclear antigen Sp110 C	IPR1	Transcriptional regulator	N	Q9HB58	(Bloch et al., 2000)
SP140	SP140 nuclear body protein	LYSP100	Transcriptional regulator	N, C	Q13342	(Zong et al., 2000)
TAF1	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor	TAFII250	Transcription initiation	N	P21675	(Wassarman and Sauer, 2001)
TAF1L	TAF1-like RNA polymerase II, TATA box binding protein (TBP)-associated factor	TAF(II)210	Transcription initiation	N	Q8IZX4	(Wang and Page, 2002)
TIF1 α	Transcriptional intermediary factor 1	TRIM24, PTC6, RNF82,	Transcriptional regulator	N, C	O15164	(Tsai et al., 2010)
TRIM28	Tripartite motif-containing 28	KAP1, RNF96, TIF1 β	Transcriptional regulator	N	Q13263	(Rowe et al., 2010)
TRIM33	Tripartite motif-containing 33 A	PTC7, RFG7, TIF1 γ	Control of Transcription elongation	N	Q9UPN9	(Bai et al., 2010)
TRIM66	Tripartite motif-containing 66	TIF1 δ	Transcriptional repressor	N	O15016	(Khetchoumian et al., 2004)
WDR9	WD repeat domain 9	BRWD1	Chromatin remodeling factor	N	Q9NSI6	(Huang et al., 2003)
ZMYND11	Zinc finger, MYND domain containing 11	BS69, BRAM1	Transcriptional repressor	N	Q15326	(Masselink and Bernards, 2000)

* Nuclear or Cytoplasmic

Supplemental Table S2: Cloning information of BRD constructs

(Separate EXCEL file)

Supplemental Table S3: Expression and purification details for studied human BRDs

(Separate EXCEL file)

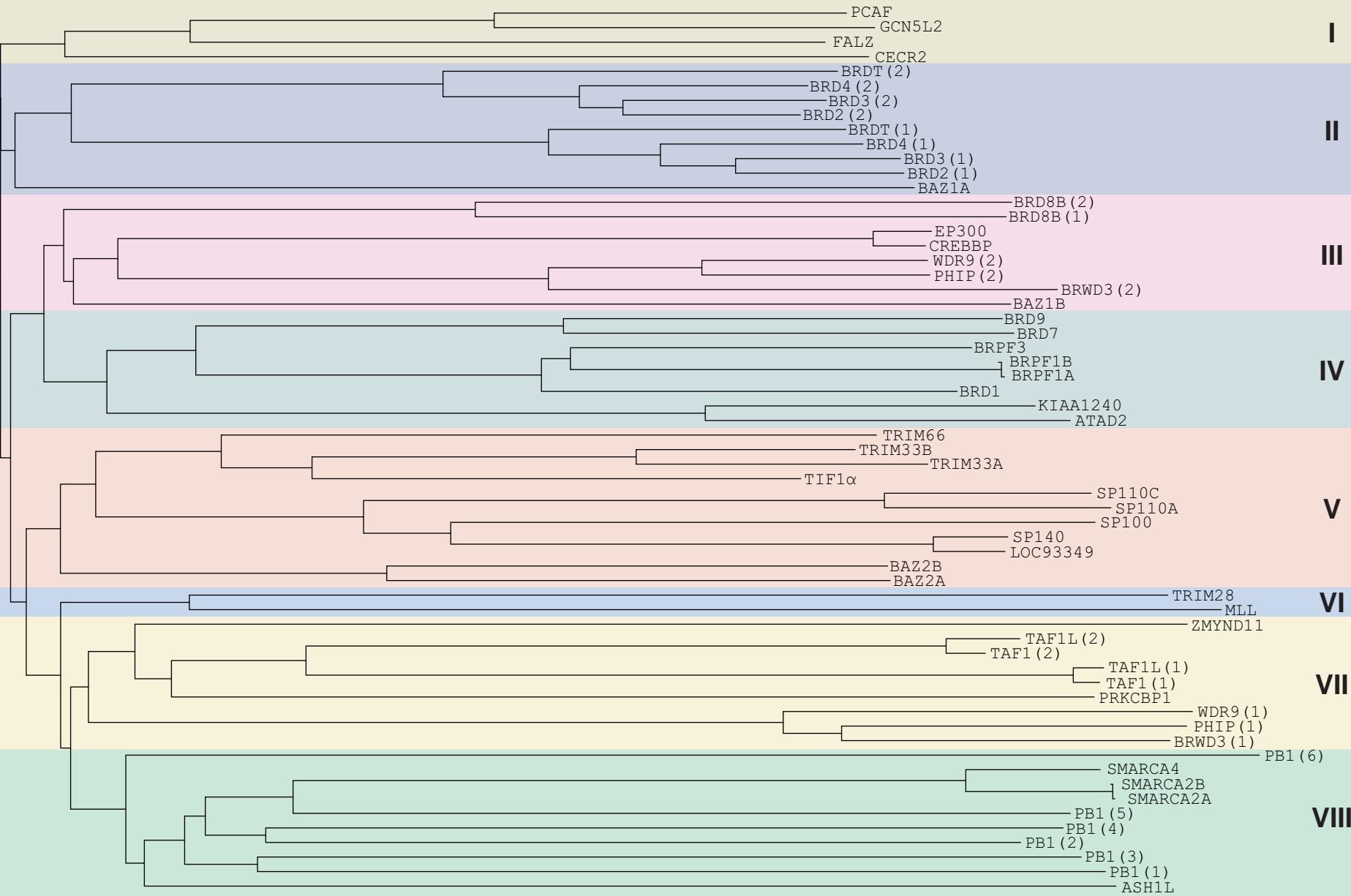
Supplemental Table S4: Diffraction and refinement statistics for crystallized human BRDs

(Separate EXCEL file)

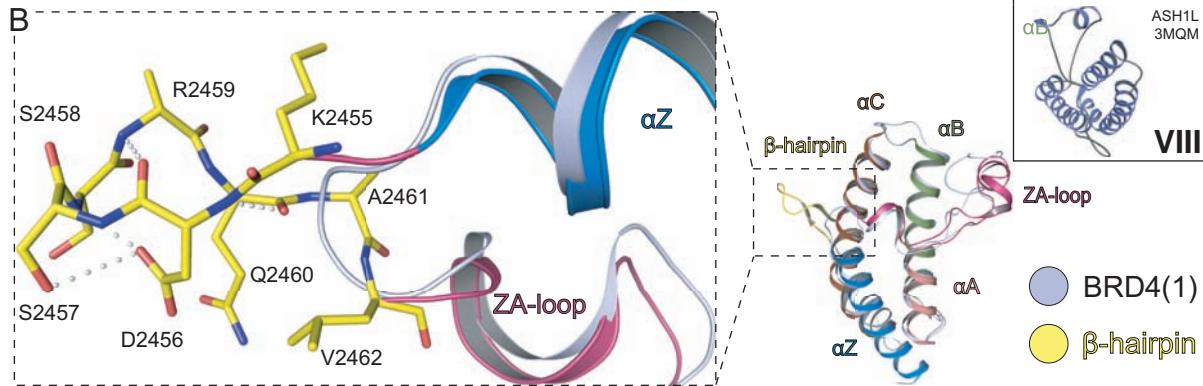
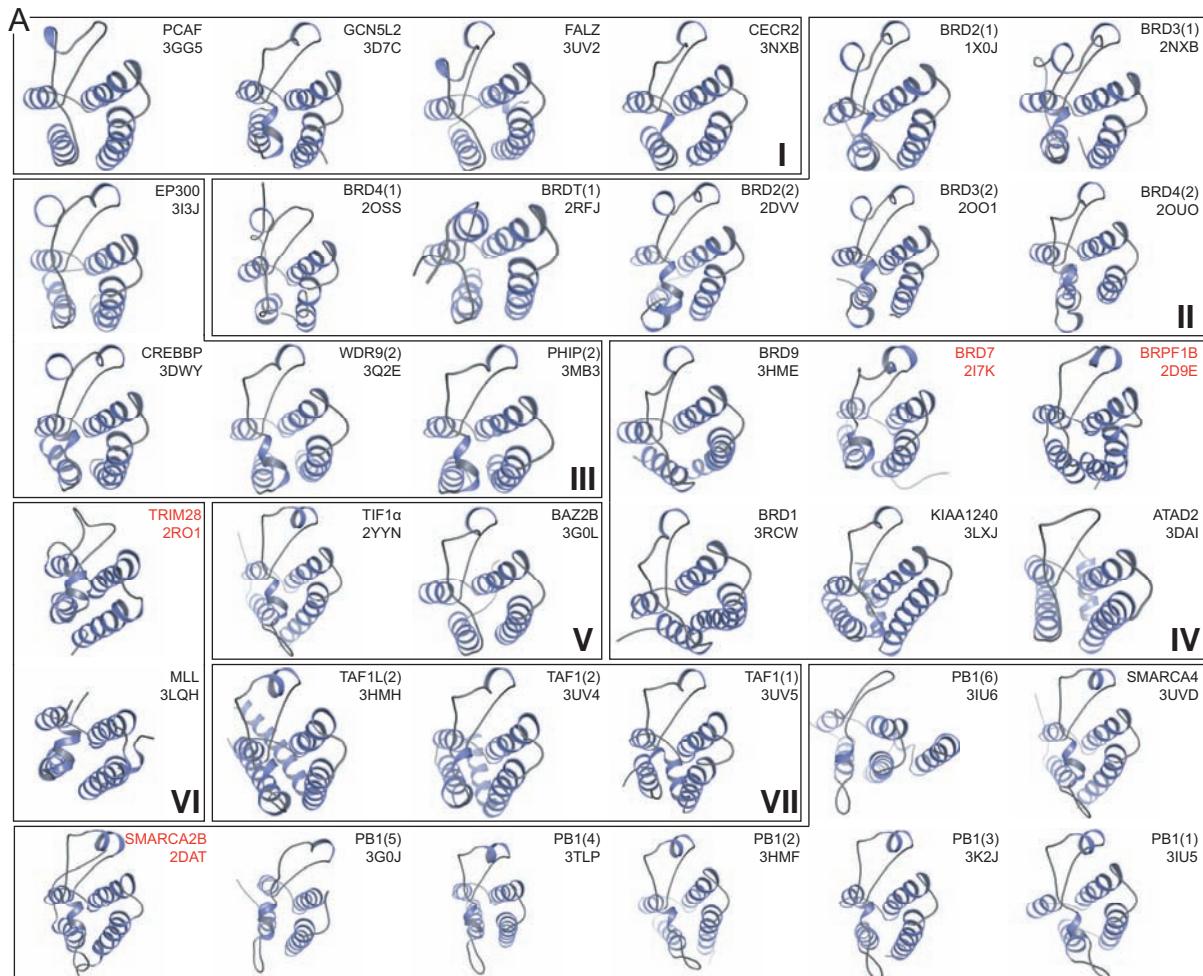
SUPPLEMENTAL FIGURE S1A



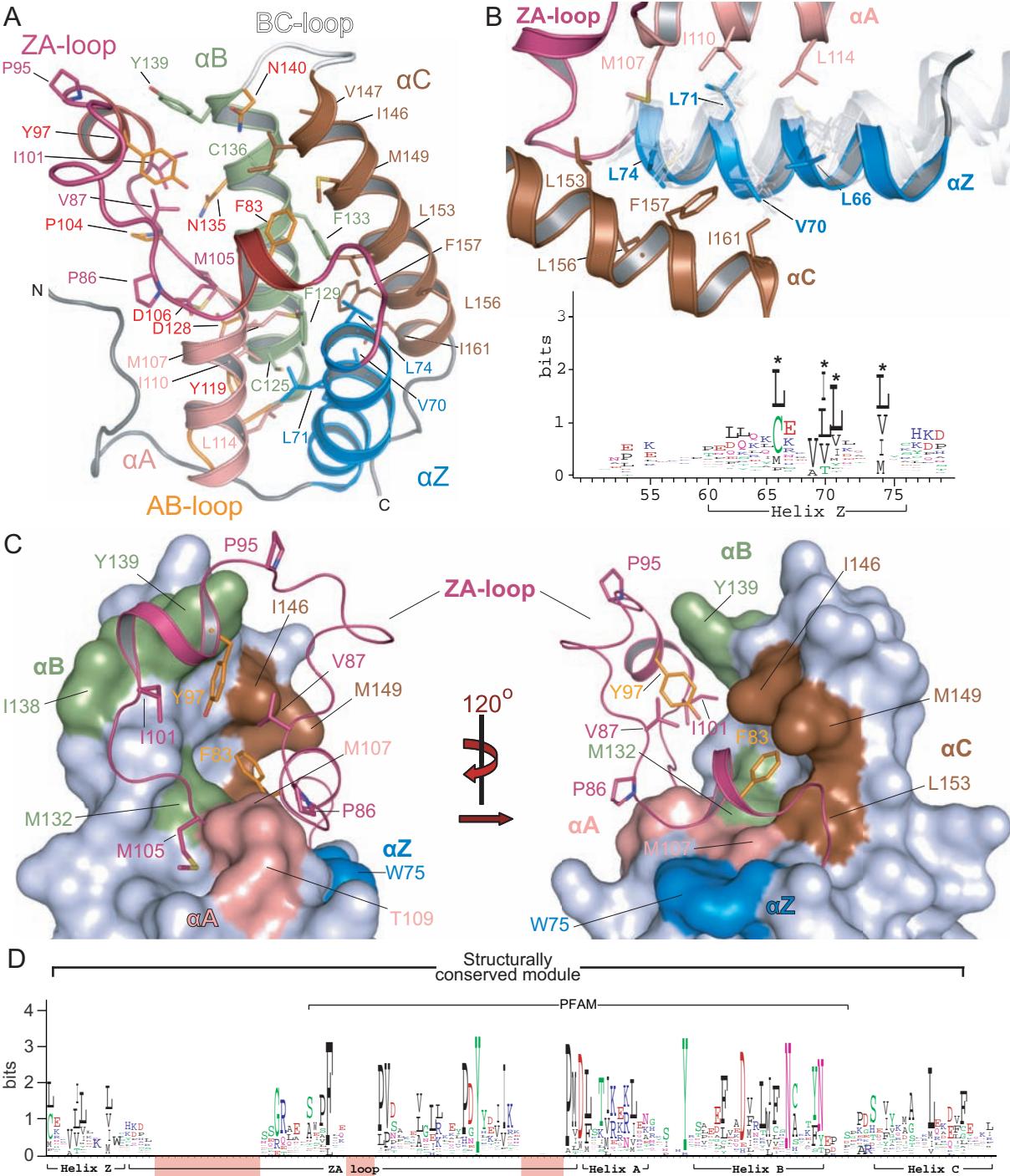
SUPPLEMENTAL FIGURE S1B



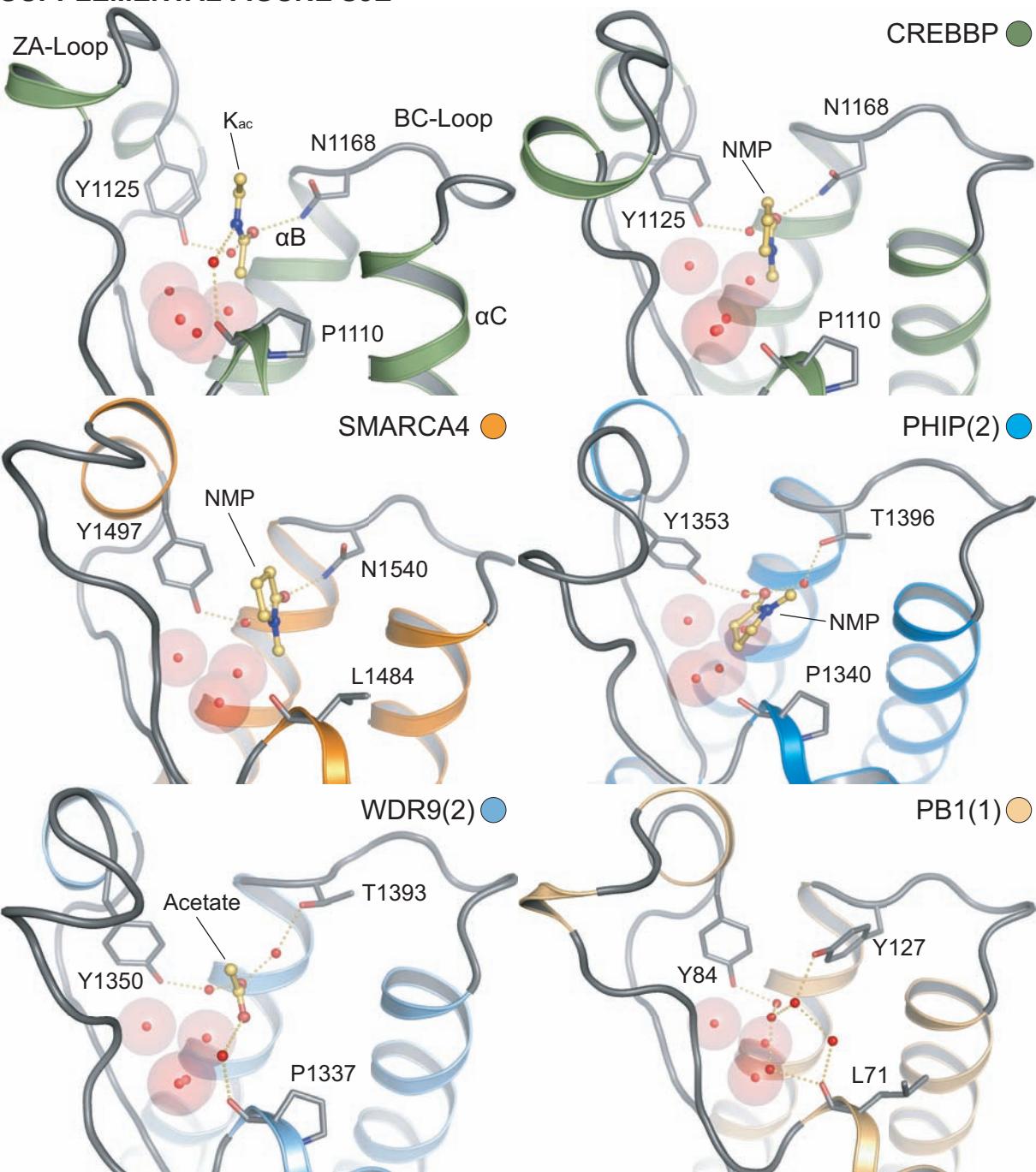
SUPPLEMENTAL FIGURE S2



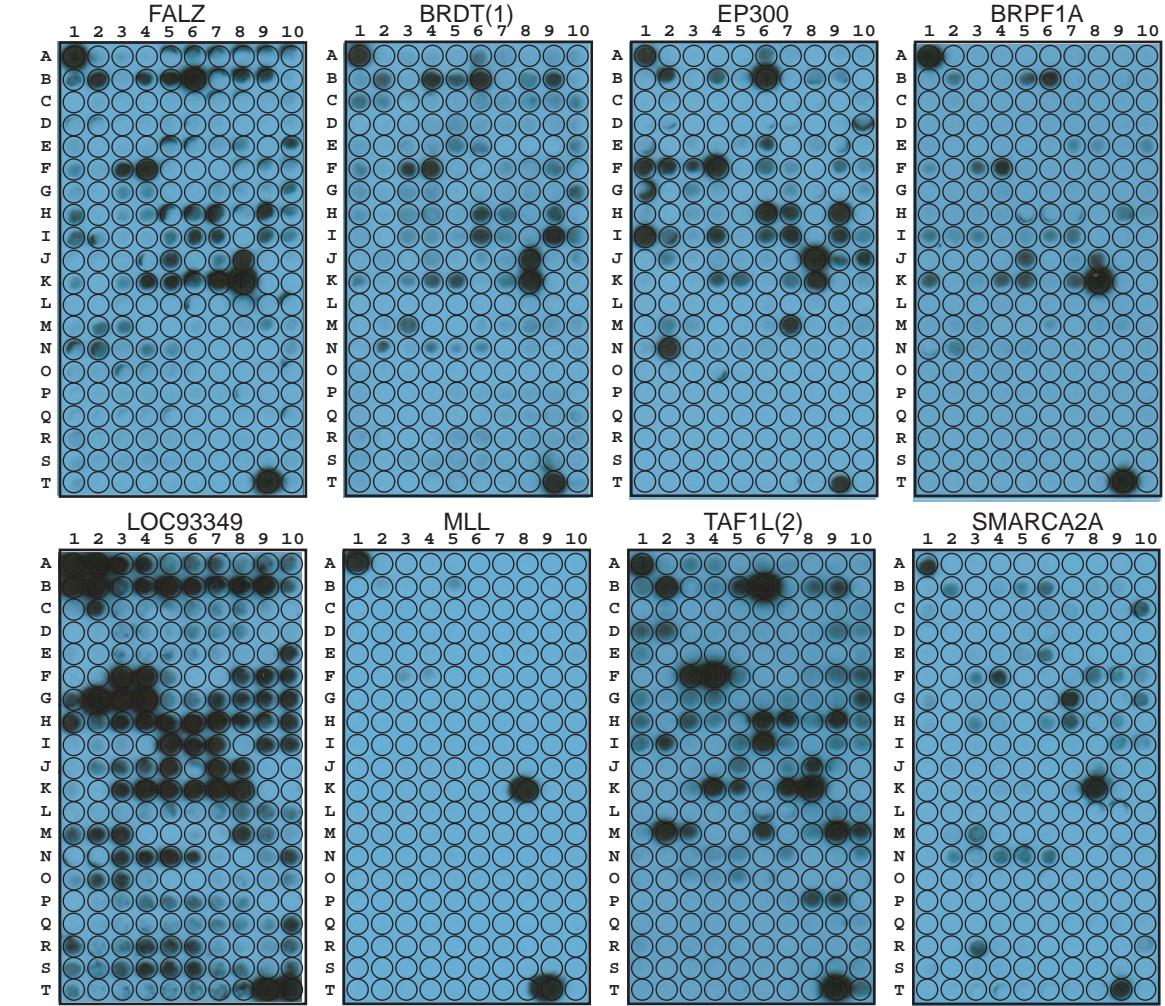
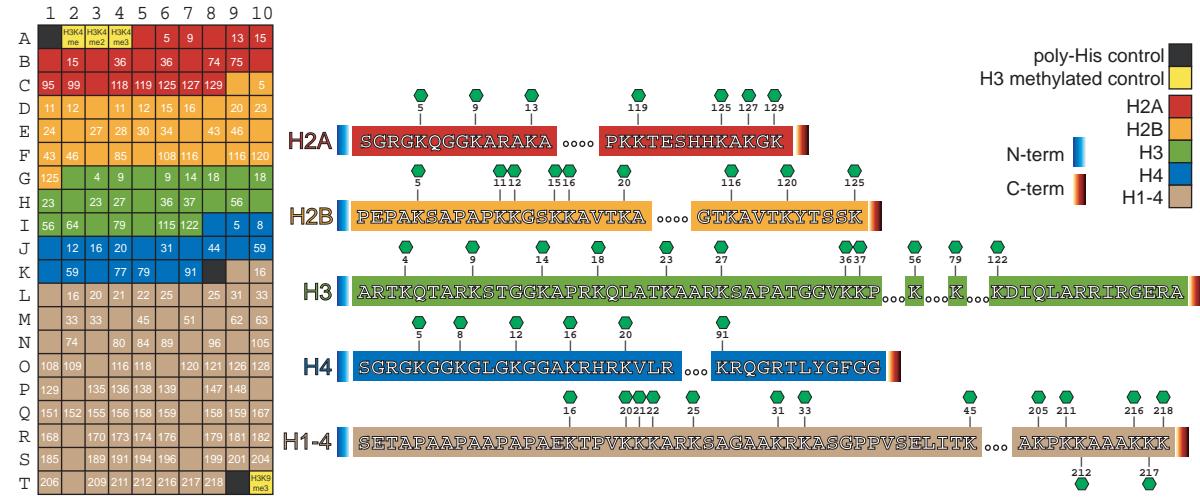
SUPPLEMENTAL FIGURE S3



SUPPLEMENTAL FIGURE S3E

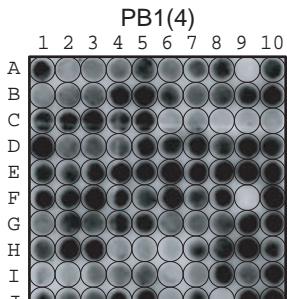
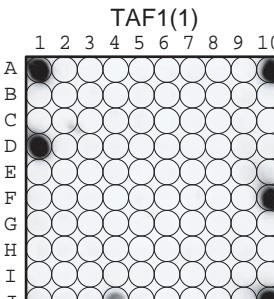
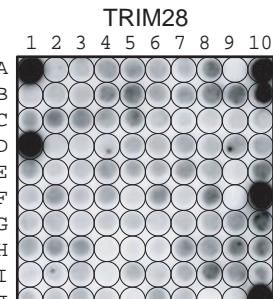
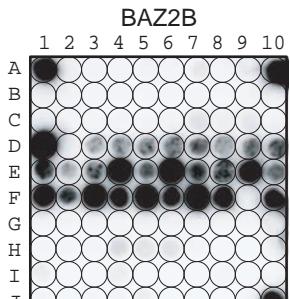
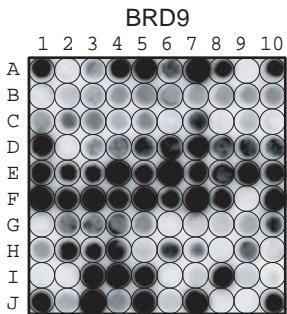
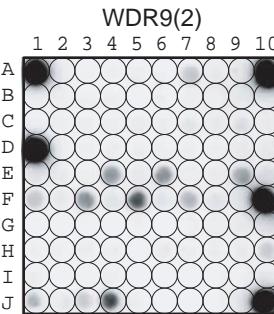
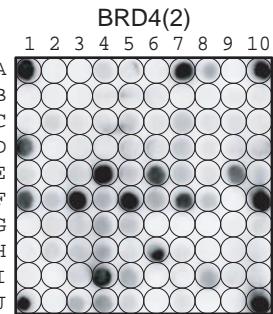
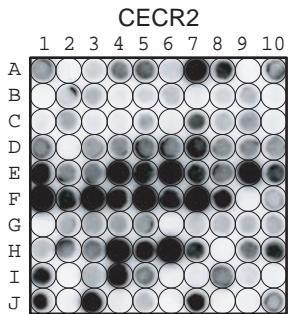
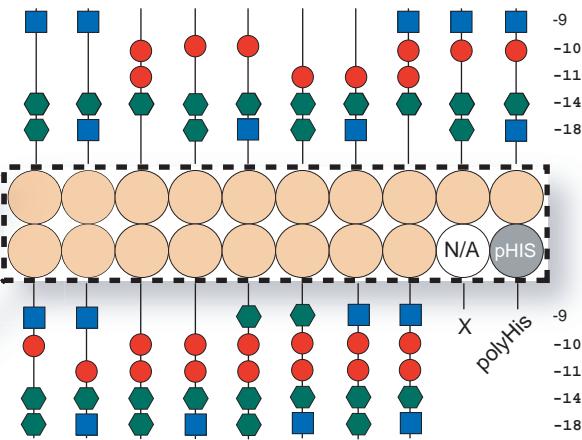
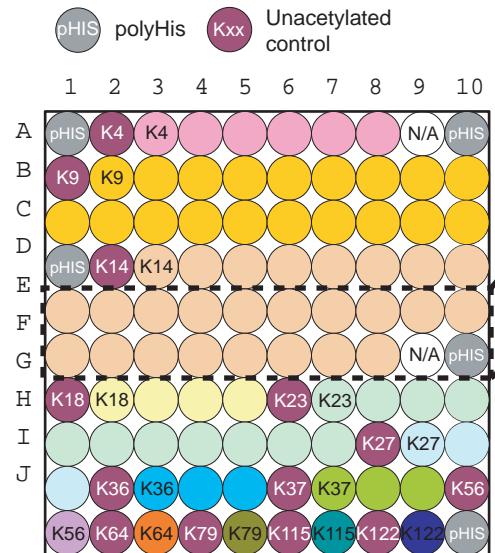
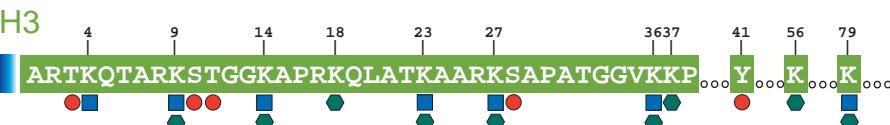


SUPPLEMENTAL FIGURE S4



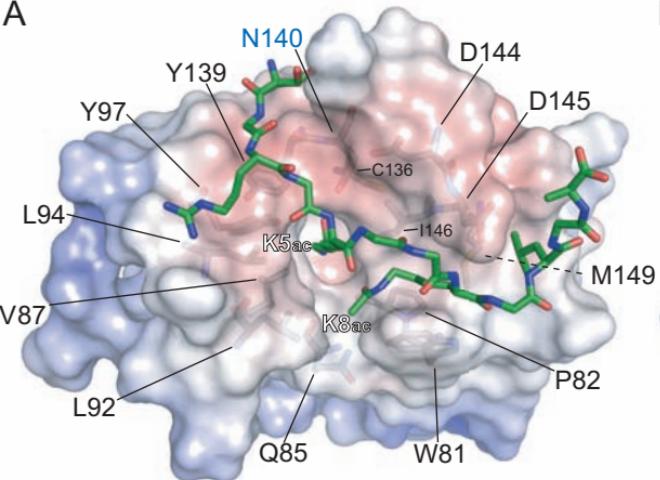
SUPPLEMENTAL FIGURE S5

- ◆ Acetylation
 - Methylation
 - Phosphorylation

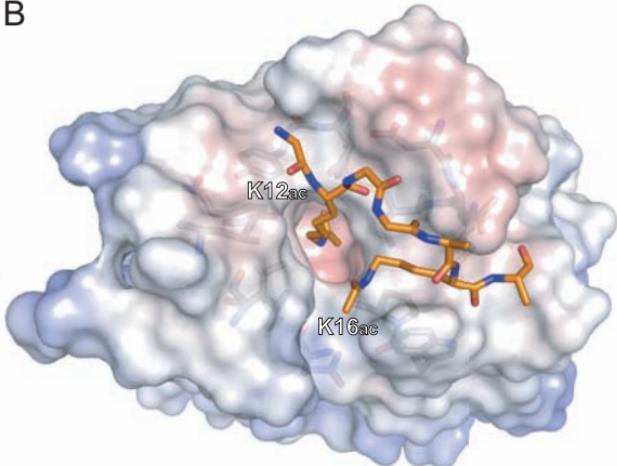


SUPPLEMENTAL FIGURE S6

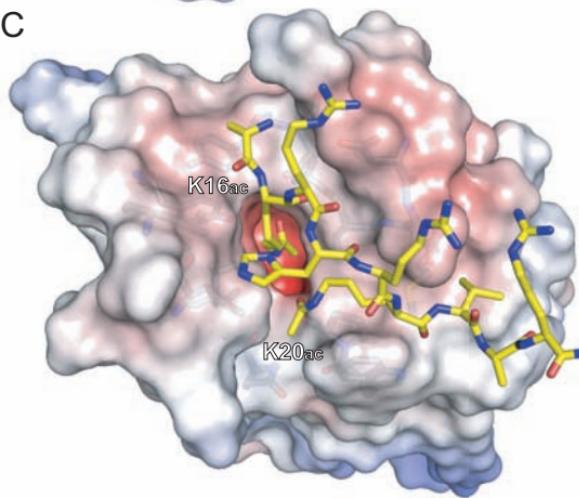
A



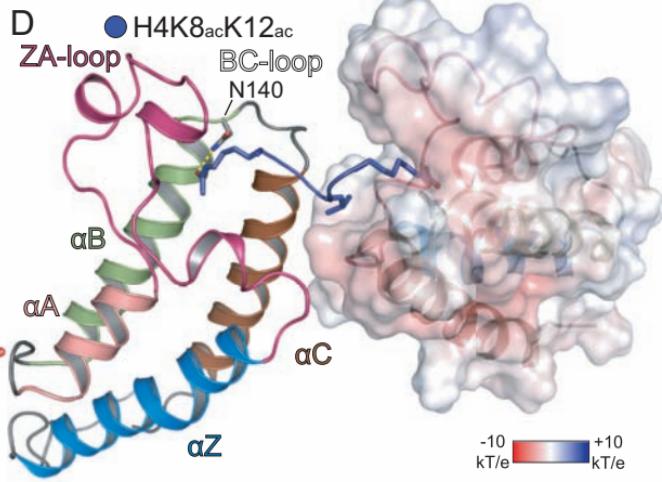
B



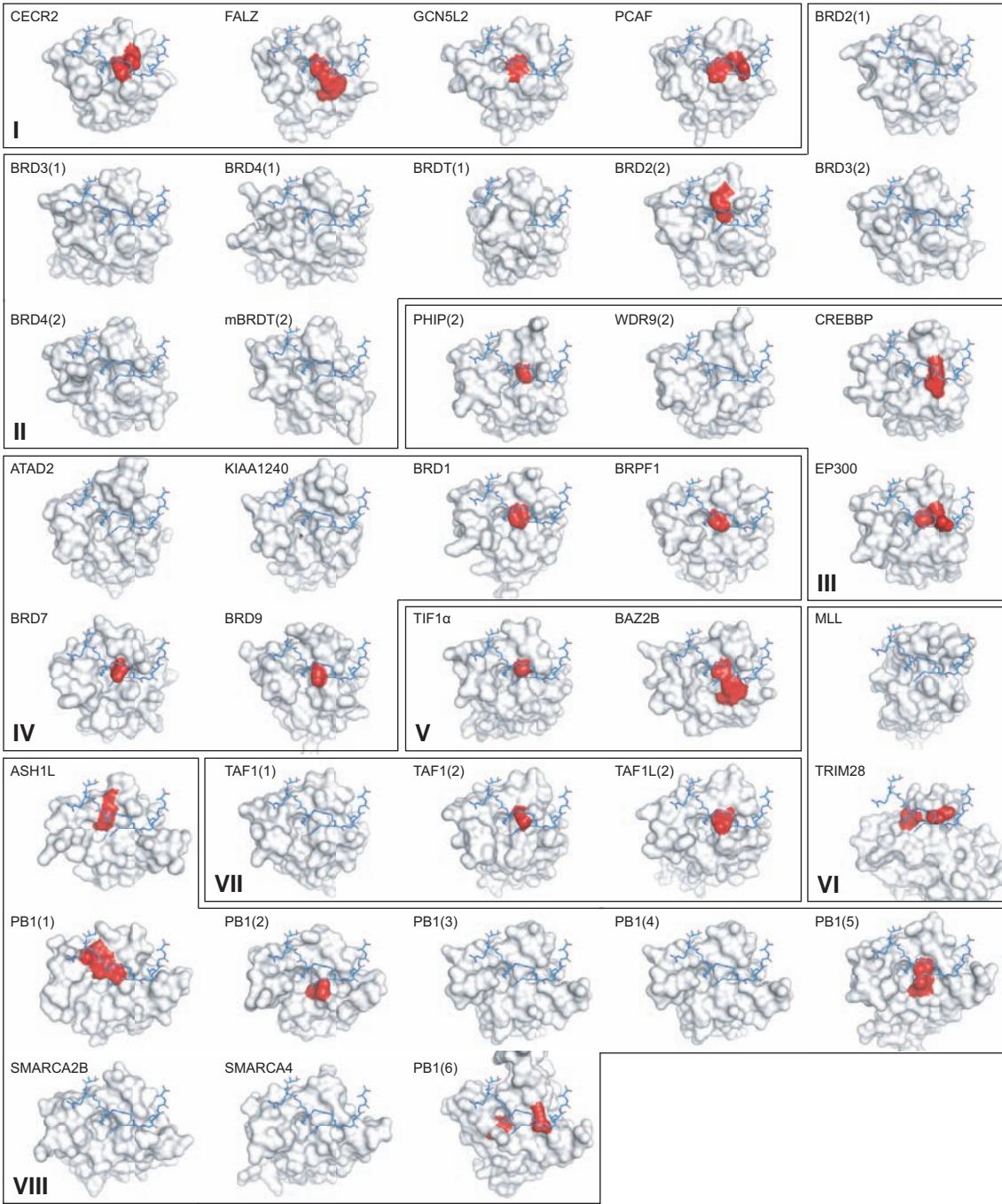
C



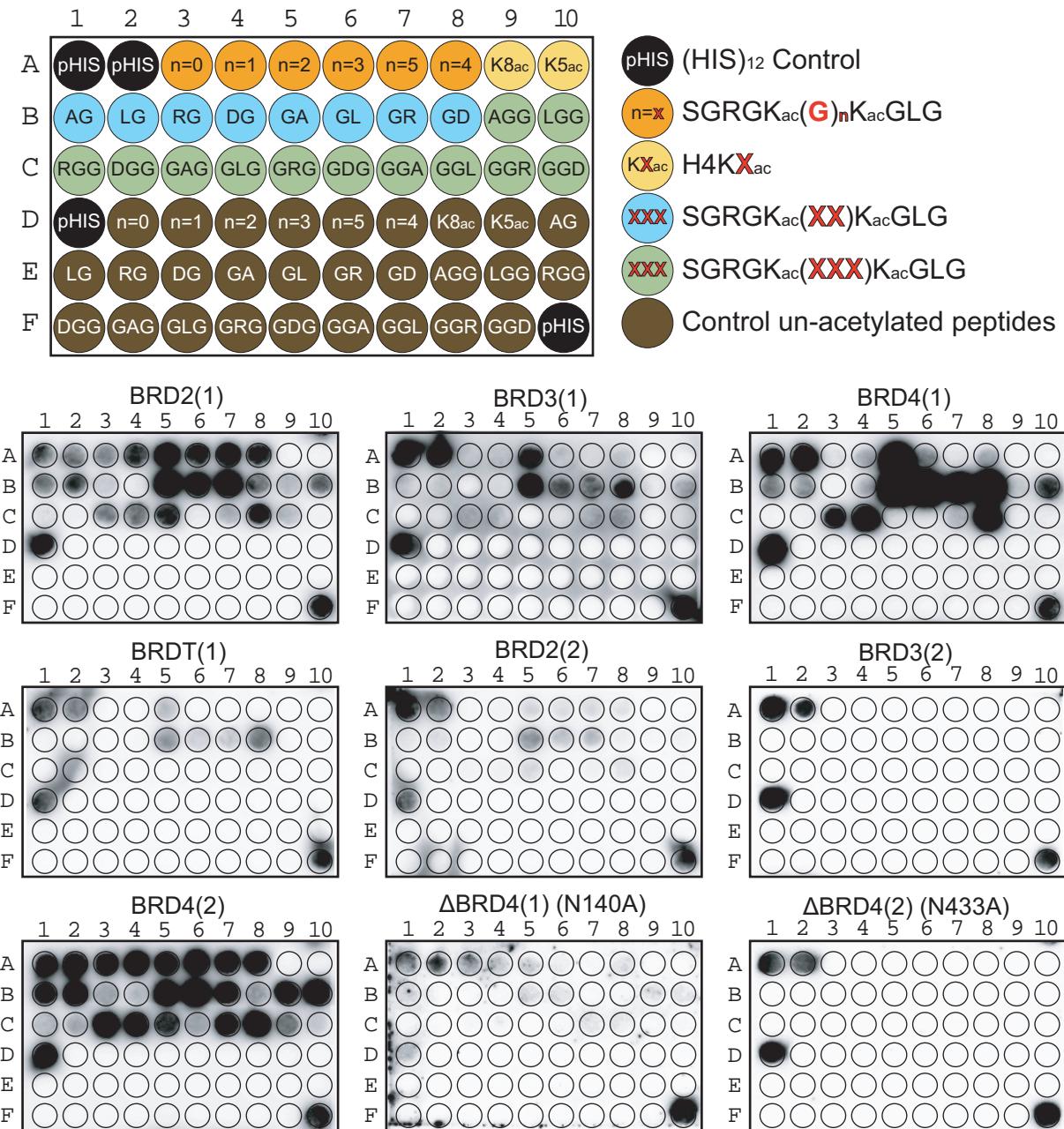
D



SUPPLEMENTAL FIGURE S6E



SUPPLEMENTAL FIGURE 7



A Global Protein Kinase and Phosphatase Interaction Network in Yeast

Ashton Breitkreutz, et al.
Science **328**, 1043 (2010);
DOI: 10.1126/science.1176495

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of February 27, 2012):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/328/5981/1043.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2010/05/20/328.5981.1043.DC1.html>

A list of selected additional articles on the Science Web sites related to this article can be found at:

<http://www.sciencemag.org/content/328/5981/1043.full.html#related>

This article cites 22 articles, 7 of which can be accessed free:

<http://www.sciencemag.org/content/328/5981/1043.full.html#ref-list-1>

This article has been cited by 6 article(s) on the ISI Web of Science

This article has been cited by 36 articles hosted by HighWire Press; see:
<http://www.sciencemag.org/content/328/5981/1043.full.html#related-urls>

This article appears in the following subject collections:

Cell Biology

http://www.sciencemag.org/cgi/collection/cell_biol

at pH 5.0. This complex is likely to interact with the cytoplasmic region of basal body of the secretion apparatus and to respond to an unidentified pH sensor. The sensor is unlikely to be part of the translocon because the translocon deletion mutant displayed wild-type levels of effector secretion upon pH upshift (fig. S7). The sensor might be the needle subunit itself, which has been implicated in signaling the translocator to effector switch in *Shigella* (11) and Yop secretion by *Yersinia* (12). Another possibility is that translocon pore assembly changes the pH gradient within the needle channel and that the sensor is located toward the base of the secretion apparatus. Changes in pH from mildly acidic to neutral can have dramatic effects on protein folding; for example, some bacterial toxins refold after their translocation from acidic endosomes to the host-cell cytosol in a partially un-

folded state (13). The SPI-2 T3SS pH sensor might thus undergo a conformational change on exposure to neutral pH and transduce a dissociation signal to the SsaL/SsaM/SpiC complex.

References and Notes

- J. E. Galán, H. Wolf-Watz, *Nature* **444**, 567 (2006).
- C. Rappl, J. Deiwick, M. Hensel, *FEMS Microbiol. Lett.* **226**, 363 (2003).
- D. Chakravortty, M. Rohde, L. Jäger, J. Deiwick, M. Hensel, *EMBO J.* **24**, 2043 (2005).
- X.-J. Yu, M. Liu, D. W. Holden, *Mol. Microbiol.* **54**, 604 (2004).
- M. J. Pallen, S. A. Beatson, C. M. Bailey, *BMC Microbiol.* **5**, 9 (2005).
- B. K. Coombes, N. F. Brown, Y. Valdez, J. H. Brumell, B. B. Finlay, *J. Biol. Chem.* **279**, 49804 (2004).
- Materials and methods are available as supporting material on *Science Online*.
- F. D. Schubot et al., *J. Mol. Biol.* **346**, 1147 (2005).
- D. Drecktrah, L. A. Knodler, D. Howe, O. Steele-Mortimer, *Traffic* **8**, 212 (2007).

- C. R. Beuzón, G. Banks, J. Deiwick, M. Hensel, D. W. Holden, *Mol. Microbiol.* **33**, 806 (1999).
- R. Kenjale et al., *J. Biol. Chem.* **280**, 42929 (2005).
- J. Torruellas, M. W. Jackson, J. W. Pennock, G. V. Plano, *Mol. Microbiol.* **57**, 1719 (2005).
- J. A. Young, R. J. Collier, *Annu. Rev. Biochem.* **76**, 243 (2007).

- We thank J. Mota, C. Tang, and members of the Holden laboratory for comments on the manuscript. This research was supported by grants G0800148 and 074553/Z/04/Z to D.W.H. from the Medical Research Council and Wellcome Trust.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1189000/DC1
Materials and Methods
Figs. S1 to S8
References

2 March 2010; accepted 25 March 2010
Published online 15 April 2010;
10.1126/science.1189000
Include this information when citing this paper.

A Global Protein Kinase and Phosphatase Interaction Network in Yeast

Ashton Breitkreutz,^{1,*} Hyungwon Choi,^{2*} Jeffrey R. Sharom,^{1,3*} Lorrie Boucher,^{1*} Victor Nedveda,^{4*} Brett Larsen,¹ Zhen-Yuan Lin,¹ Bobby-Joe Breitkreutz,¹ Chris Stark,¹ Guomin Liu,¹ Jessica Ahn,¹ Danielle Dewar-Darch,¹ Teresa Reguly,¹ Xiaojing Tang,¹ Ricardo Almeida,⁴ Zhaohui Steve Qin,⁵ Tony Pawson,^{1,3} Anne-Claude Gingras,^{1,3,†} Alexey I. Nesvizhskii,^{2,6,†} Mike Tyers^{1,3,4†}

The interactions of protein kinases and phosphatases with their regulatory subunits and substrates underpin cellular regulation. We identified a kinase and phosphatase interaction (KPI) network of 1844 interactions in budding yeast by mass spectrometric analysis of protein complexes. The KPI network contained many dense local regions of interactions that suggested new functions. Notably, the cell cycle phosphatase Cdc14 associated with multiple kinases that revealed roles for Cdc14 in mitogen-activated protein kinase signaling, the DNA damage response, and metabolism, whereas interactions of the target of rapamycin complex 1 (TORC1) uncovered new effector kinases in nitrogen and carbon metabolism. An extensive backbone of kinase-kinase interactions cross-connects the proteome and may serve to coordinate diverse cellular responses.

Protein phosphorylation mediates cellular responses to growth factors, environmental signals, and internal processes by the regulation of protein interactions, enzyme activity, or protein localization (1). However, the protein interactions of kinases, phosphatases, and their regulatory subunits and substrates remain sparse-

ly mapped, particularly in high-throughput (HTP) datasets [fig. S1 (2)]. To chart the budding yeast kinase and phosphatase interaction (KPI) network, we systematically characterized protein kinase and phosphatase complexes by rapid magnetic bead capture, on-bead protein digestion, and mass spectrometric identification of associated proteins, using different epitope tags and expression systems [fig. S2; (2)]. One hundred thirty protein kinases, 24 lipid and metabolic kinases, 47 kinase regulatory subunits, 38 protein phosphatases, 32 phosphatase regulatory subunits, and 5 metabolic phosphatases were analyzed (tables S1 and S2).

We eliminated nonspecific interactions using a statistical model called Significance Analysis of Interactome (SAINT). In contrast to simple threshold models, SAINT assigns the number of peptide identifications for each interactor to a probability distribution, which is then used to estimate the likelihood of a true interaction (2). We validated

SAINT on multiple independent purifications for several kinases and expression levels (fig. S3 and tables S3 to S5). A final KPI dataset of 1844 interactions between 887 protein partners was generated from more than 38,000 unfiltered identifications at a stringent SAINT threshold of $P > 0.85$ (fig. S4 and tables S1 and S2). High-confidence interactions were recovered for 120 protein kinases (fig. S5; see fig. S6 and table S6 for validation). For a number of kinases, we demonstrated that associated proteins were substrates in vitro (figs. S7 and S8 and table S7). Our dataset doubled the number of KPIs obtained in previous low-throughput (LTP) studies and performed as well as LTP data against an unbiased HTP high-confidence (HTP-HC) benchmark dataset [fig. S1 (2)]. Clustering of all kinases and phosphatases by their interaction profiles revealed locally dense regions in the KPI network (Fig. 1A and fig. S9).

The Cdc14 phosphatase formed one of the largest single hubs in the network with 53 interaction partners, including 23 kinases and 5 phosphatases (Fig. 1B, fig. S6, and table S6). Cdc14 antagonizes mitotic cyclin-dependent kinase (CDK) activity and is activated by the mitotic exit network (MEN) upon completion of anaphase (3). Many Cdc14 interactors were shared with its anchor protein Net1 and the nicotinamide adenine dinucleotide (NAD^+)-dependent histone deacetylase Sir2 that together with Cdc14 form the nucleolar RENT complex (4). New connections between Cdc14 and other mitotic regulators included the CDK-inhibitory kinase Swe1, the cytokinesis checkpoint protein Boi1 (5), and two activators of cytokinesis, Cbk1 and Ace2 (6). Cdc14, Net1, and Sir2 each interacted with the DNA damage checkpoint kinases Chk1 and Dun1. In support of a role for Cdc14 in the DNA damage response, we found that ectopic expression of Cdc14 caused sensitivity to the DNA-damaging agent methylmethane sulfonate (MMS), while a strain defective for Cdc14 function was sensitive to the ribonucleotide reductase inhibitor hydroxyurea (Fig. 1C). Interactions between the

¹Centre for Systems Biology, Samuel Lunenfeld Research Institute, 600 University Avenue, Toronto, Ontario, M5G 1X5, Canada. ²Department of Pathology, University of Michigan, Ann Arbor, MI 48109, USA. ³Department of Molecular Genetics, University of Toronto, 1 Kings College Circle, Toronto, Ontario, M5S 1A8, Canada. ⁴Wellcome Trust Centre for Cell Biology and School of Biological Sciences, University of Edinburgh, Mayfield Road, Edinburgh, EH9 3JR Scotland, UK.

⁵Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA. ⁶Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: gingras@lunenfeld.ca (A.C.G.), nesvi@med.umich.edu (A.I.N.), tyers@lunenfeld.ca, m.tyers@ed.ac.uk (M.T.).

RENT and the nutrient-sensing TOR complex 1 (TORC1) were supported by the finding that increased Cdc14 activity caused rapamycin sensitivity, whereas reduced Cdc14 function caused rapamycin resistance (Fig. 1D), suggesting that Cdc14 may antagonize TOR signaling. Cdc14 also interacted with the energy-sensing adenosine 5'-monophosphate (AMP)-activated kinase (AMPK) Snf1 and its upstream kinase Sak1; AMPK activates glucose-repressed genes in yeast and is an upstream inhibitor of TOR activity in metazoans (7). Deregulation of Cdc14 caused a severe defect in growth on glycerol medium and sensitivity to 2-deoxyglucose (Fig. 1D).

Cdc14 exhibited connections with three different mitogen-activated protein kinase (MAPK) modules. Interaction of the pheromone MAPK pathway kinases Fus3 and Ste7 with Cdc14 was supported by the finding that constitutive expres-

sion of Cdc14 caused partial pheromone resistance (fig. S10). Cdc14 interacted with the high osmolarity glycerol (HOG) pathway MAPK kinase Pbs2; consistently, constitutive expression of Cdc14 caused sensitivity to osmotic stress (Fig. 1E). The HOG pathway is also known to stimulate mitotic exit (8). The upstream cell wall integrity (CWI) MAPK kinase Bck1 interacted with Cdc14; a *cde14-3* strain was sensitive to the cell wall stress agent calcofluor white (Fig. 1E). These CWI interactions extended along each pathway because the conditional MEN alleles *mob1-77* and *cdc15-2* exhibited specific synthetic lethal interactions with either *slt2Δ* or *bck1Δ* mutations; this lethality was alleviated by growth on iso-osmotic medium but not by a catalytically inactive mutant of Slt2 (Fig. 1F and fig. S10). These data reveal Cdc14 as a nexus for cell cycle, checkpoint, metabolic, and stress signals (fig. S10).

The TORC1 and TORC2 kinase complexes are conserved from yeast to human and control macromolecular synthesis and polarized morphogenesis, respectively; TORC1 is sensitive to the macrolide rapamycin, whereas TORC2 is not (9). In the KPI dataset, TORC1 and TORC2 formed a highly connected subnetwork of 28 interaction partners, including 13 kinases and 4 phosphatases (figs. S6 and S11 and table S6). These connections established new links between TORC1 and the mitochondrial retrograde (RTG) signaling pathway (10), which induces genes required for glutamate production (fig. S12). Multiple TORC1 subunits exhibited previously undocumented interactions with the kinases Fmp48, Nnk1, Npr1, and Ksp1 (Fig. 2A and fig. S11).

Fmp48 is a kinase of unknown function that is associated with mitochondrial subcellular fractions (11). Consistent with interactions among

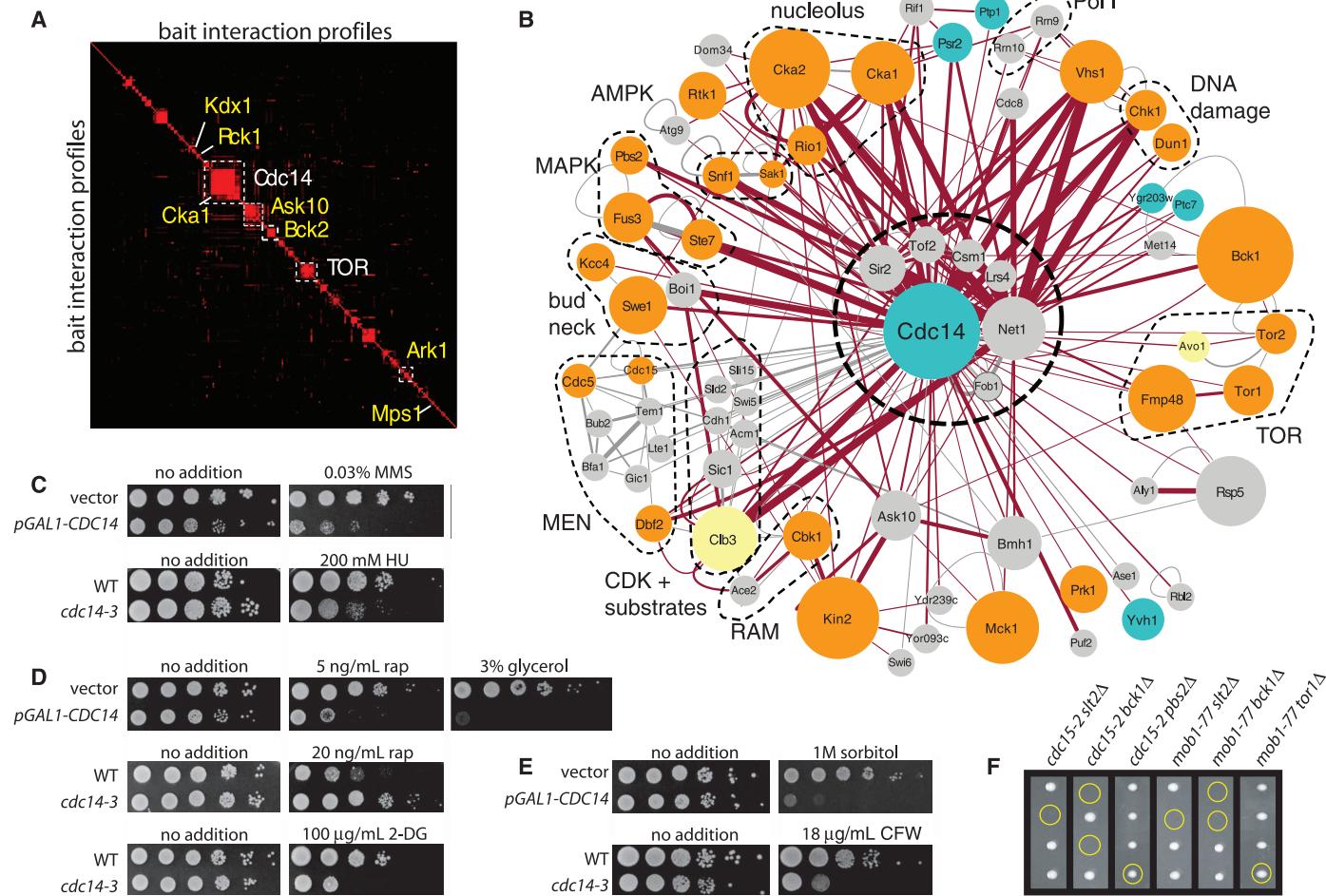


Fig. 1. Cdc14 phosphatase network. **(A)** Hierarchical two-dimensional clustering of bait interaction profiles in the KPI dataset. See fig. S9 for full clustergram. Networks for indicated clusters and other kinases are shown in fig. S19. **(B)** Cdc14-Net1-Sir2 (RENT) interaction network. Kinases are in orange, phosphatases in blue, kinase-associated proteins in yellow, and other proteins in gray. Red connecting lines indicate KPI interactions, gray lines LTP interactions, and gray dashed lines HTP-HC interactions. Line thickness indicates peptide count of interaction; node size is proportional to total number of interactions in the KPI dataset. Bold dashed circle indicates RENT complex and known associated proteins. RAM, regulation of Ace2p activity and cellular morphogenesis. **(C)** Sensitivity of a *GAL1-CDC14* strain to 0.03% methyl methanesulfonate (MMS) when induced by 0.02% galactose (see fig. S20 for expression titration) and a *cdc14-3* strain to 200 mM hydroxyurea (HU) at 33°C. **(D)** Sensitivity of a *GAL1-CDC14* strain to either rapamycin (5 ng/ml) or glycerol medium when induced by 0.05% galactose. Resistance of a *cdc14-3* strain to rapamycin (20 ng/ml) and sensitivity to 2-deoxyglucose (DG, 100 µg/ml) at 33°C. **(E)** Sensitivity of a *GAL1-CDC14* strain to 1 M sorbitol when induced by 0.05% galactose. Sensitivity of a *cdc14-3* strain to calcofluor white (CFW, 18 µg/ml) at 33°C. **(F)** Representative tetrads bearing combinations of *slt2Δ*, *bck1Δ*, *cdc15-2*, and *mob1-77* alleles. Double-mutant spore clones are circled in yellow; *pbs2Δ* and *tor1Δ* served as negative controls.

0.03% methyl methanesulfonate (MMS) when induced by 0.02% galactose (see fig. S20 for expression titration) and a *cdc14-3* strain to 200 mM hydroxyurea (HU) at 33°C. **(D)** Sensitivity of a *GAL1-CDC14* strain to either rapamycin (5 ng/ml) or glycerol medium when induced by 0.05% galactose. Resistance of a *cdc14-3* strain to rapamycin (20 ng/ml) and sensitivity to 2-deoxyglucose (DG, 100 µg/ml) at 33°C. **(E)** Sensitivity of a *GAL1-CDC14* strain to 1 M sorbitol when induced by 0.05% galactose. Sensitivity of a *cdc14-3* strain to calcofluor white (CFW, 18 µg/ml) at 33°C. **(F)** Representative tetrads bearing combinations of *slt2Δ*, *bck1Δ*, *cdc15-2*, and *mob1-77* alleles. Double-mutant spore clones are circled in yellow; *pbs2Δ* and *tor1Δ* served as negative controls.

Fmp48, TORC1, and the RTG inhibitor Mks1, elevated expression of *FMP48* caused a growth defect on nonfermentable glycerol medium and rapamycin resistance on a fermentable carbon source (Fig. 2B). Overexpression of *FMP48* caused abnormal mitochondrial morphology (Fig. 2C) and repression of genes encoding tricarboxylic acid cycle enzymes, electron transport chain components, and subunits of the adenosine 5'-triphosphate (ATP) synthase (Fig. 2D). Fmp48-associated kinase activity was specifically increased by rapamycin

treatment (Fig. 2E), suggesting that Fmp48 relays TORC1 signals to the RTG pathway and mitochondrial function.

The uncharacterized kinase Ykl171w, renamed Nnk1 for nitrogen network kinase, associated with all TORC1 subunits (fig. S11) and with Gdh2, the NAD⁺-dependent glutamate dehydrogenase that catalyzes deamination of glutamate to α-ketoglutarate and ammonia (12). Gdh2 was phosphorylated by Nnk1 complexes in vitro (Fig. 2F), and a *gdh2Δ* strain was resistant to rapamycin

when grown on glutamate as the sole nitrogen source (Fig. 2G), whereas overexpression of *NNK1* conferred hypersensitivity to rapamycin (Fig. 2H). Nnk1 also interacted with the TORC1 effector Ure2, which regulates the nitrogen catabolite response by sequestering the transcription factor Gln3 in the cytoplasm (12). Overexpression of *NNK1* induced rapid nuclear accumulation of Gln3 (Fig. 2I) and increased transcription of Gln3 target genes (Fig. 2J), suggesting that Nnk1 activity antagonizes the Ure2-Gln3 inter-

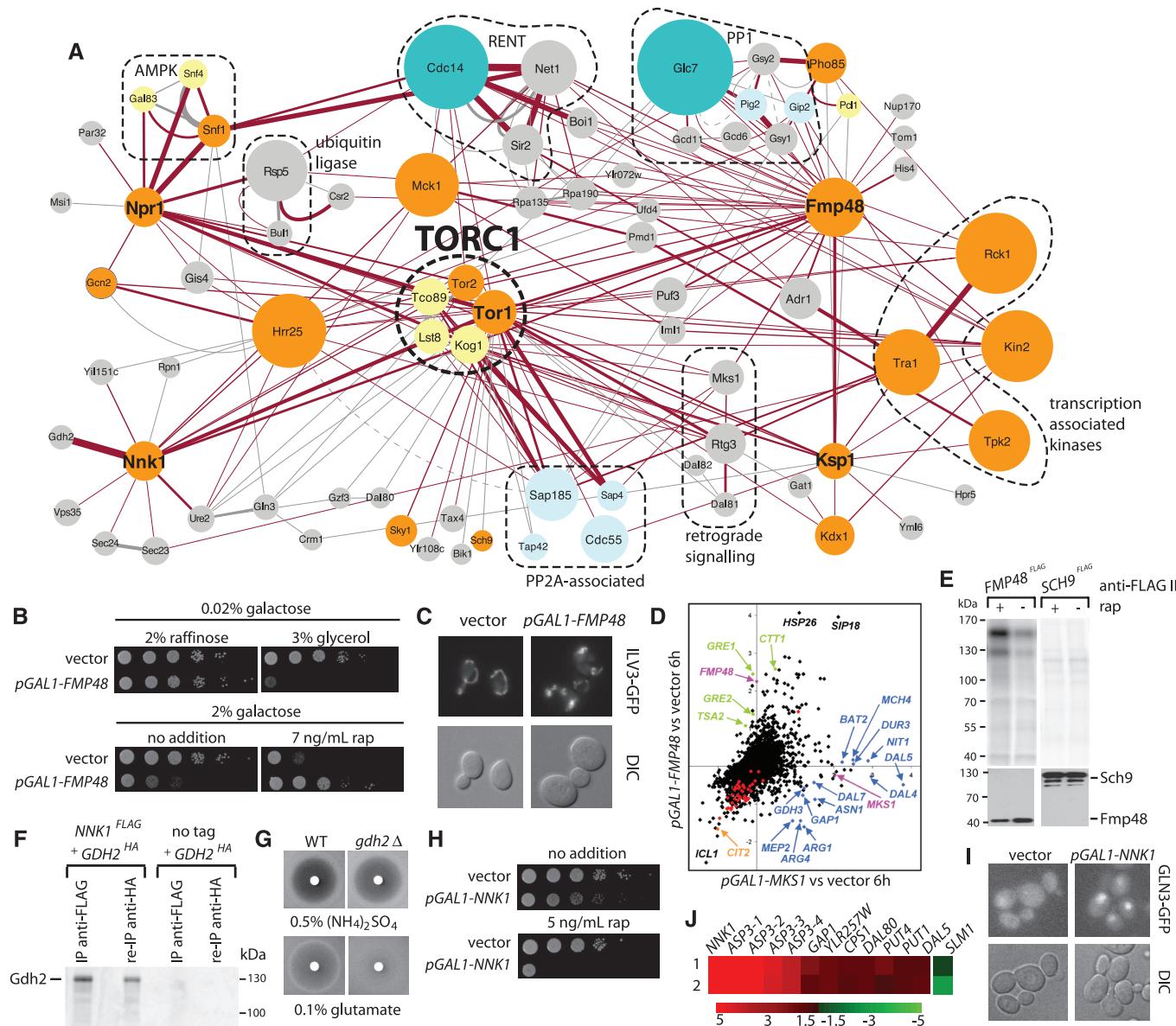


Fig. 2. TORC1 kinase network. **(A)** Partial network of new TORC1-associated kinases. **(B)** Overexpression of *GAL1-FMP48* inhibits growth on glycerol and confers rapamycin resistance. **(C)** Overexpression of *GAL1-FMP48* causes abnormal mitochondrial morphology as visualized with an *Ivy3*^{GFP} mitochondrial matrix fusion protein (GFP, green fluorescent protein). DIC, differential interference contrast. **(D)** Genome-wide expression profiles of *GAL1-FMP48* and *GAL1-MKS1* strains induced with 0.2% galactose. RTG-responsive (orange), mitochondrial (red), stress-responsive (green), and Gln3/Gcn4-responsive (blue) genes are marked. **(E)** Fmp48^{FLAG} or Sch9^{FLAG} complexes were immunopurified from cells grown in the presence or absence of rapamycin (200 ng/ml) for 30 min, then incubated with

[³³P]-γ-ATP, and radiolabeled species were resolved by SDS-polyacrylamide gel electrophoresis. Nonregulated Sch9-associated activity served as a negative control. **(F)** Immunopurified Nnk1^{FLAG} complexes were incubated with [³³P]-γ-ATP, then denatured, and radiolabeled Gdh2 species were repurified with antibody to hemagglutinin (HA). **(G)** A *gdh2Δ* strain is rapamycin resistant when glutamate is the sole nitrogen source. **(H)** Expression of *GAL1-NNK1* in 2% galactose confers sensitivity to rapamycin (5 ng/ml). **(I)** Expression of *GAL1-NNK1* in 2% galactose for 1 hour causes nuclear accumulation of Gln3^{GFP}. **(J)** Expression of *GAL1-NNK1* in 0.2% galactose for 1.5 hours specifically induces Gln3 target genes. Color bar indicates fold increase (red) or decrease (green) relative to empty vector control.

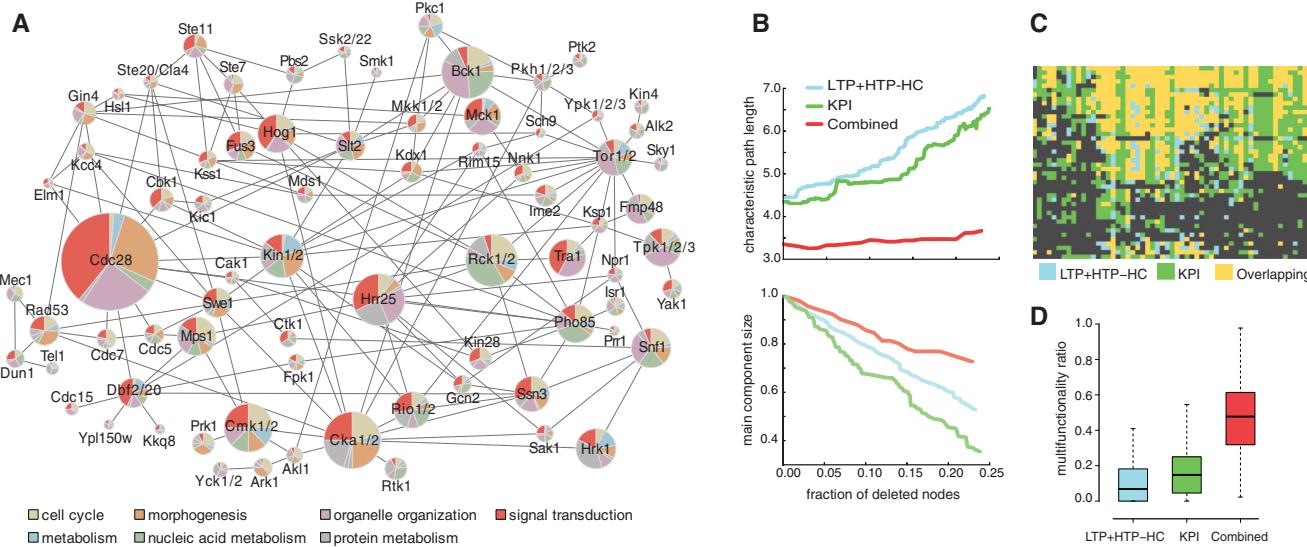


Fig. 3. A kinase-kinase (K-K) network connects the proteome. **(A)** Combined K-K interaction network derived from the KPI, LTP, and HTP-HC datasets. Interactions from known kinase regulatory subunits and paralogs were collapsed into single nodes (table S8). The reduced network contains 156 interactions between 75 kinases, 66 of which contain documented phosphorylation sites (table S9). Colors indicate fraction of GO Super-Slim biological processes assigned by interaction partners of each kinase (2). **(B)**

action. The expansive TORC1 network also included other nutrient-sensing kinases (Npr1, Snf1, Gcn2, and Ksp1; fig. S11) (13), transcription-associated kinases (Tra1 and Tpk2), MAPK module components (Bck1 and Kdx1), cell cycle kinases (Ime2, Mih1, and Clb2-Cdc28), an mRNA splicing kinase (Sky1), and a ribosome biogenesis kinase (Rio2). These findings underscore the central role of TOR in cell growth.

In a global protein interaction network constructed from the KPI, LTP, and HTP-HC datasets (2), kinase-kinase (K-K) interactions were significantly enriched compared to all other kinase interaction partners ($P < 3 \times 10^{-6}$) and collectively formed a highly interconnected K-K network (Fig. 3A, figs. S13 and S14, and table S8). Consistent with a trans-kinase phosphorylation network (14), we assigned 607 phosphorylation sites on 98 kinases (fig. S15 and table S9). This K-K network was extremely robust to fragmentation by hub deletion (Fig. 3B) and was far less modular than previous less-complete K-K networks [fig. S16 (2)]. Within the global network, kinases had a significantly higher centrality compared to nonkinase nodes [$P < 10^{-16}$ (2)], suggesting that kinases might unify cellular regulation. To test this idea, we identified dense clusters of interactions (cliques or complexes) in the global interaction network, then determined the extent of clique cross-connection by kinase interactions. More than 80% of the proteome was interlinked by kinases in this manner (fig. S17), a significantly larger fraction than in random networks [$P < 10^{-8}$ (2)]. The potential for kinases to co-regulate otherwise separate functions was further revealed by the diversity of Gene Ontology (GO) biological processes associated with kinase

interaction partners (Fig. 3C and fig. S18). The multifunctionality of kinases, as defined by associated GO terms, was markedly increased by the KPI dataset (Fig. 3D).

Cellular processes are controlled by a multitude of low-affinity interactions, as often mediated by short linear motifs embedded in disordered protein regions (15, 16). The KPI network is highly enriched for disordered regions as compared to the entire proteome [$P < 10^{-16}$ (2)]. This physical organization may allow the cell to overcome stochastic limitations in signal propagation, integration, and downstream responses (16). In human cells, kinase-mediated signaling can readily propagate across pathways (17) and may dictate complex decisions through a broadly distributed network of effectors (18, 19). Moreover, phosphorylation-based feedback loops often enable cooperative responses, tuning of network outputs, and entrained states (20–22). The densely connected and non-modular architecture of the KPI network suggests that the interaction of many such circuits will underpin cellular information flow (23).

References and Notes

1. T. Pawson, *Curr. Opin. Cell Biol.* **19**, 112 (2007).
2. Supporting material is available on *Science Online*.
3. F. Stegmeier, A. Amon, *Annu. Rev. Genet.* **38**, 203 (2004).
4. A. F. Straight et al., *Cell* **97**, 245 (1999).
5. M. Mendoza et al., *Nat. Cell Biol.* **11**, 477 (2009).
6. B. Nelson et al., *Mol. Biol. Cell* **14**, 3782 (2003).
7. D. G. Hardie, *Nat. Rev. Mol. Cell Biol.* **8**, 774 (2007).
8. V. Reiser, K. E. D'Aquino, L. S. Ee, A. Amon, *Mol. Biol. Cell* **17**, 3136 (2006).
9. S. Wullschleger, R. Loewith, M. N. Hall, *Cell* **124**, 471 (2006).
10. Z. Liu, R. A. Butow, *Annu. Rev. Genet.* **40**, 159 (2006).
11. J. Reinders, R. P. Zahedi, N. Pfanner, C. Meisinger, A. Sickmann, *J. Proteome Res.* **5**, 1543 (2006).
12. B. Magasanik, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16537 (2005).
13. A. Huber et al., *Genes Dev.* **23**, 1929 (2009).
14. A. Chi et al., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2193 (2007).
15. V. Neduvia et al., *PLoS Biol.* **3**, e405 (2005).
16. T. J. Gibson, *Trends Biochem. Sci.* **34**, 471 (2009).
17. W. M. Old et al., *Mol. Cell* **34**, 115 (2009).
18. H. Daub et al., *Mol. Cell* **31**, 438 (2008).
19. A. von Kriegsheim et al., *Nat. Cell Biol.* **11**, 1458 (2009).
20. J. E. Ferrell Jr., *Curr. Opin. Cell Biol.* **14**, 140 (2002).
21. F. Li, T. Long, Y. Lu, Q. Ouyang, C. Tang, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4781 (2004).
22. Q. A. Justman, Z. Serber, J. E. Ferrell Jr., H. El-Samad, K. M. Shokat, *Science* **324**, 509 (2009).
23. P. Nurse, *Nature* **454**, 424 (2008).
24. We thank B. Raught, A. Amon, L. Harrington, J. Bader, M. Costanzo, B. Andrews, C. Boone, R. Aebersold, B. Bodenmiller, I. Sadowski, and F. Sicheri for discussions; J. P. Zhang and D. Fermin for technical support; and M. Snyder, Y. Ohsumi, S. Piatti, S. Hahn, H. Reizman, S. Biggins, T. Petes, M. P. Longhese, and D. Mao for reagents. Supported by grants from the Canadian Institutes of Health Research to A.C.G. (MOP-84314), T.P. (MOP-57793), and M.T. (MOP-12246); the Ontario Research Fund to T.P. and A.C.G. (REO#-044); the National Institutes of Health to M.T. (R01RR024031 from the National Center for Research Resources) and A.I.N. (CA-126239); a Terry Fox Foundation Research Studentship from the National Cancer Institute of Canada to J.R.S.; Federation of European Biochemical Societies and Marie Curie Fellowships to V.N.; Canada Research Chairs in Functional Genomics and Bioinformatics (to M.T.) and in Functional Proteomics (to A.C.G.); the Lea Reichmann Chair in Cancer Proteomics to A.C.G.; and a Scottish Universities Life Sciences Alliance Research Professorship and a Royal Society Wolfson Research Merit Award to M.T.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5981/1043/DC1
Materials and Methods
Figs. S1 to S27
Tables S1 to S15
References
19 May 2009; accepted 7 April 2010
10.1126/science.1176495

PROTEOMICS**Mapping physical interactions within chromatin by proteomic approaches**

Journal:	<i>PROTEOMICS</i>
Manuscript ID:	pmic.201100547.R1
Wiley - Manuscript type:	Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Lambert, Jean-Philippe; Samuel Lunenfeld Research Institute, Pawson, Tony; Samuel Lunenfeld Research Institute, Gingras, Anne-Claude
Key Words:	chromatin, Histone, Affinity chromatography

SCHOLARONE™
Manuscripts

Review

1
2
3 **Title:** Mapping physical interactions within chromatin by proteomic approaches.
4
5

6 **Authors:** Jean-Philippe Lambert¹, Tony Pawson^{1,2}, Anne-Claude Gingras^{1,2}
7
8

9 ¹ Samuel Lunenfeld Research Institute at Mount Sinai Hospital, 600 University Ave, Toronto,
10 ON, M4M2Y8, Canada, and ² Department of Molecular Genetics, University of Toronto,
11 Toronto, ON, Canada.
12
13

14
15 **Abstract:**
16
17

18
19 Our ability to study protein-protein interactions has grown by leaps and bounds in recent years,
20 enabling numerous large-scale studies to be performed in a variety of organisms. Despite this
21 success, some classes of proteins, including those bound to chromatin, remain difficult to
22 characterize through proteomic approaches. Some of the problems faced by researchers
23 studying chromatin-bound proteins include low complex solubility, heterogeneous sample
24 composition, and numerous transient interactions which can be further complicated by the
25 presence of DNA itself. To tackle these issues, a number of innovative protocols have been
26 developed to better study the various facets of chromatin biology. In this review, we will
27 discuss novel approaches to study protein-DNA interactions as well as protein complexes
28 affecting chromatin.
29
30
31
32
33
34
35
36
37
38
39

1. Introduction

As Anuj Kumar and Michael Snyder stated "... no protein is an island entire of itself - or at least, very few proteins are" [1]. Whether proteins form homodimers, heterodimers, or are part of large multisubunit complexes, protein-protein interactions are critical for most protein functions [2]. As such, a significant fraction of the proteomic field has been devoted to the development and utilization of techniques for the identification of protein-protein interactions. The identification of protein-protein interactions has traditionally been performed through biochemical purification, which relies on the physical characteristics of the protein under study. The advent of efficient cloning techniques for protein tagging has permitted the emergence of efficient generic affinity purification reagents and epitope tags (reviewed by Dunham et al., this issue), which enabled the development of a variety of techniques to study protein-protein interactions. It was quickly realized that the mass spectrometer was a detector of choice to couple with affinity purification to enable both sensitive and unbiased protein detection [2]. Affinity purification coupled to mass spectrometry (AP-MS) has been used successfully in both small and large studies from a wide array of organisms including bacteria, *Saccharomyces cerevisiae* and mammalian cells and tissues [3]. Mapping protein-protein interactions by AP-MS is now a common tool in laboratories around the world.

Despite the success of the AP-MS approach, some classes of proteins remain notoriously difficult to study by AP-MS, including highly hydrophobic or membrane-embedded proteins, which are difficult to solubilize while preserving their interactions [4]. Poor protein complex solubilization is also a significant problem for the study of chromatin [5]. Chromatin is composed of DNA associated with a wide array of proteins involved in the maintenance and regulation of an organism's genomic material. The most fundamental unit of chromatin is the nucleosome, which consists of two copies of each core histone (histones H2A, H2B, H3 and H4) around which 147 base pairs of DNA is wrapped (Figure 1A). Histones are small basic proteins which exist as soluble dimers or tetramers prior to their assembly as octamers into chromatin (reviewed in [6]). The N-termini of histones, referred to as histone tails, do not participate in the histone-DNA interaction, but rather are exposed to solvent [7]. Histone tails are the target

of numerous post-translational modifications such as acetylation, methylation, phosphorylation and ubiquitination (Figure 1A). These modifications occur in a combinatorial manner referred to as the “histone code” to enable the recruitment of specific proteins to chromatin, resulting in distinct downstream events [8][9] (Figure 1B). Elucidating the biological consequences of histone modifications remains a challenging task which has greatly benefited from advances in the field of proteomics, as reviewed in the sections below. Beyond histones, numerous proteins are involved in chromatin functions, including transcription factors, nucleosome remodelers and structural components. The identification and characterization of the interplay between all these factors is necessary to obtain a holistic view of chromatin. Here we will focus on recent advances in proteomics that have improved our understanding of chromatin composition, interconnection and functions. We will cover approaches that successfully tackled chromatin from a DNA perspective (DNA-centric) and from the protein perspective (protein-centric).

2. Chromatin study using DNA-centric approaches

2.1 Global characterization of the protein composition of chromatin

Because of its very large size and charged nature, chromatin has a tendency to precipitate following cell lysis, a characteristic that has been harnessed to biochemically enrich it from cell extracts [10]. These chromatin-enriched protein fractions are typically generated by first isolating cell nuclei, which are then lysed in hypotonic buffer and subsequently gently centrifuged (5-15 minutes, <2000g). The soluble fraction (containing non-chromatin bound proteins) is removed, leaving a pellet which is greatly enriched for chromatin-associated proteins such as histones. Historically, these fractions were characterized as a whole by methods such as sedimentation analysis [11] or their constituents were visualized by gel electrophoresis [12-13]. The first reported quantitative proteomic characterization of chromatin-enriched fractions employing mass spectrometry was performed by Shiio *et al.* in 2003 [14] (Figure 2). To investigate the effects of c-Myc expression on chromatin composition, chromatin-enriched fractions from human B lymphocytes cells expressing the transcription

1
2
3 factor c-Myc, or not, were prepared, and the isolated proteins were labelled with isotopic
4 coded reagents (ICAT) and analyzed by MS [15]. This resulted in the detection and quantitation
5 of 282 distinct proteins. In particular, the expression levels of numerous transcription factors
6 were significantly altered following induction of c-Myc. As such, this study represented a very
7 successful first proteomic attempt at defining the complexity of chromatin, even though its
8 scope was restricted by the instrumentation available at the time.
9
10

11 Since this initial report, several other groups have also used quantitative mass
12 spectrometry for the characterization of chromatin-associated proteins. For example, Chou *et*
13 *al.* recently reported novel factors involved in DNA damage repair by comparing the
14 composition of chromatin-enriched fractions before and after DNA damage [16]. To do so,
15 stable isotope labelling by amino acids in cell culture (SILAC, see Trinkle-Mulcahy, this issue)
16 was employed to generate two populations of HeLa cells (light and heavy labelled), one of
17 which was exposed to UV light to induce DNA damage. The cell populations were then mixed,
18 the chromatin-enriched fractions prepared, and the proteins separated by SDS-PAGE and
19 analyzed by LC-MS/MS. This effort resulted in the detection of more than 1100 proteins, and
20 the identification of proteins quantitatively enriched in the chromatin fraction after DNA
21 damage. For instance, proteins known to be recruited to sites of DNA damage, such as 53BP1
22 and MDC1, were enriched in chromatin from the DNA damaged sample. The authors also
23 identified DNA-damage induced chromatin enrichment of three members of the polycomb
24 complexes PRC1 (CBX2) and PRC2 (EZH2 and SUZ12) and further showed that loss of the
25 polycomb complexes resulted in sensitivity to infrared radiation. Similarly, another protein
26 enriched in chromatin after DNA damage was found to be the poly (ADP-ribose) polymerase
27 (PARP), which was implicated in both DNA damage repair and in the recruitment of the
28 polycomb complex to sites of DNA damage.
29
30

31 Kubota *et al.* optimized protocols to generate chromatin-enriched fractions from
32 budding yeast to better characterize the role of Ctf18, a subunit of the replication factor C
33 complex, in maintaining genomic stability [17]. The protein composition of chromatin during S
34 phase in wild type or *ctf18Δ* yeast, challenged or not with the replication inhibitor hydroxyurea,
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

1
2
3 was analyzed. In cells lacking *ctf18*, replisome progression complex components were
4 increasingly associated with chromatin. Ctf18 was also shown to be necessary for activation of
5 the DNA replication checkpoint protein Rad53. In addition, the authors reported that deletion
6 of *elg1*, another replication factor present in budding yeast, resulted in the enrichment of
7 different chromatin-binding proteins following hydroxyurea treatment of *ctf18* cells, suggesting
8 that Elg1 and Ctf18 exhibit different functions *in vivo*. Based on these studies, it is apparent that
9 the analysis of chromatin-enriched fractions is a viable means of studying chromatin, especially
10 when combined with modern proteomics methodologies. However, the generation of
11 chromatin-enriched fractions is inherently biased towards the least soluble protein components
12 of chromatin, and as such is unlikely to allow for a complete protein catalogue of chromatin.
13
14
15
16
17
18
19
20
21

23 **2.1.1 The specific case of mitotic chromosomes**

24

25 As outlined above, chromatin composition undergoes profound changes following
26 various stimuli, DNA damage or progress through the cell-cycle. In particular, as cells approach
27 mitosis, chromosomes undergo significant condensation, resulting in the formation of mitotic
28 chromosomes. Mitotic chromosomes are extremely dense and are easily identified through
29 histological observations, but, critically, can also be biochemically isolated from the bulk of the
30 proteome. The classical approach is to arrest cells in mitosis with drugs (e.g. nocodazole) and
31 subsequently isolate the mitotic chromosomes by sedimentation through sequential sucrose
32 and Percoll gradients (Figure 3) [18]. Between 2002 and 2007, a number of groups performed
33 proteomic characterization of mitotic chromosomes, which resulted in the identification of 62
34 [19], 79 [20] and 240 [21] chromatin-associated proteins, respectively. While the goals of each
35 experiment were different, the protein fraction of the isolated mitotic chromosomes was
36 resolved by SDS-PAGE in each case, the protein bands excised and digested, and the resulting
37 peptide mixtures analyzed by mass spectrometry. This approach enabled Morrison *et al.* to
38 identify novel substrates for Aurora B kinase from isolated mitotic chromosomes [19] while
39 Grassmann *et al.* focused on the cell cycle-controlled association of the novel chromatin-
40 associated proteins [20]. On the other hand, Takata *et al.* identified differences in the protein
41 content of mitotic chromosomes from different cell lines [21]. While these studies
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

demonstrated that the characterization of mitotic chromosomes with proteomics tools was feasible, this remains a daunting task, with purifications often suffering from a very high level of non-chromatin contaminants. This is due in part to the fact that chromosomes possess a net negative charge which attracts a high abundance of cytosolic proteins (especially those positively charged) following nuclear envelope breakdown during mitosis [22].

In spite of the difficulties described above, Ohta *et al.* recently achieved the most exhaustive proteomic characterization of the protein content of mitotic chromosomes to date [23]. This *tour de force* was accomplished by combining classical mitotic chromosome purifications [18], as described above, with modern mass spectrometry techniques and, critically, with statistical tools. Briefly, following the isolation of mitotic chromosomes from nocodazole-arrested chicken DT40 cells, proteins were fractionated by SDS-PAGE, in-gel digested, and the resulting peptides further fractionated by ion exchange prior to mass spectrometric analysis. This effort resulted in the detection of 4029 distinct proteins associated with mitotic chromosomes. As in previous studies of mitotic chromosomes, a significant fraction of the proteins identified were clearly not functionally related to mitosis or to chromosomes, but rather appeared to be “hitchhikers”. To efficiently segregate the contaminants from true chromosome-binding proteins, the authors utilized machine learning approaches with different classifiers which included the protein enrichment in the chromosome fraction relative to cytosolic extract as well as ability of an individual protein to stably bind to chromosomes (when mixed with isotopically-labelled post-chromosomal extract). Used together in a random forest analysis, the selected classifiers demonstrated very high selectivity and specificity in classifying the identified proteins. To directly test the validity of their analysis, Ohta *et al.* tagged 50 previously uncharacterized proteins identified in their analysis with GFP and determined their localization in U2OS human cells. In confirmation of their statistical analysis, 44 of the 50 uncharacterized proteins (88%) localized within the nucleus as predicted. In addition, the authors further characterized the role of two novel kinetochore-associated proteins by showing drastic chromosomal segregation defects upon knockdown. Taken together, the wealth of information contained in this study provided the first detailed map of mitotic chromosomes, and will provide the foundation for numerous studies to come. However,

the method employed relies on the purification of mitotic chromosomes and cannot be extended to the study of chromatin at other cell-cycle stages or following various treatments or insults. As such, a universal isolation procedure enabling reproducible proteomic characterization of chromatin remains to be developed.

2.2 Targeted characterization of chromatin

2.2.1 Chromatin purification via exogenous DNA sequences

As we have seen above, the purification of endogenous proteins associated with chromatin is very challenging since chromatin is both complex and highly charged. In addition, the protein composition of chromatin varies according to the underlying DNA sequences, preventing a detailed understanding of protein-DNA association based solely on the global approaches described above. To better characterize DNA-specific protein associations, several groups have therefore employed immobilized DNA strands as affinity probes to specifically enrich proteins associated with specific DNA sequences (Figure 4A). For instance, Bürckstümmer *et al.* immobilized a well-defined 45 base pair double-stranded DNA sequence, called interferon stimulatory DNA (ISD), to characterize the proteins involved in the activation of the Interferon- β pathway, which is critical for the function of the innate immune system [24]. Following incubation of lysates (from various cell lines) with the ISD resin, the bound proteins were eluted by boiling in SDS-PAGE buffer, resolved on gel, digested in-gel with trypsin, and identified by mass spectrometry. This work resulted in the identification of over 1000 proteins putatively associated with the ISD resin, a clear indication that non-specific protein binding was also occurring. To circumvent this background issue, the authors focused only on those proteins demonstrated by RNA microarrays to also be regulated by interferon- β [24], reducing the number of putative Interferon- β pathway activators to seven, including AIM2 (absent in melanoma 2). The authors went on to demonstrate that AIM2 was a cytoplasmic DNA sensor for the inflammasome that associates with the adaptor ASC to mediate inflammasome activation (resulting, amongst other effects, in interleukin 1 beta maturation in monocytes).

Other groups have sought to overcome non-specific DNA binding by employing quantitative proteomic methods. For instance, Mittler *et al.* combined SILAC quantitation with immobilized DNA affinity purification to identify transcription factors which bind to methyl CpG islands [25]. CpG islands are clustered stretches of dinucleotide CG repeats in which the cytosine can become methylated, altering the recruitment of multiple transcription factors to specific genes. Mittler *et al.* generated DNA affinity probes containing the CpG island upstream of the MTA2 gene with and without cytosine methylation. The authors incubated metabolically labelled nuclear cell extracts with the methyl and unmethylated DNA probes, performed mild washes and eluted the DNA bound proteins by using restriction enzymes. The purified proteins from both samples were combined, resolved by SDS-PAGE, in-gel digested and identified by mass spectrometry. The authors were able to identify numerous proteins that were significantly enriched by the methylated probe, such as UHRF1, which had not been previously reported to bind to CpG islands. Rubio *et al.* employed a very similar method to determine which proteins were associated with the c-Myc insulator element containing wild-type or mutant CTCF binding sites [26]. As in the study by Mittler *et al.* [25], nuclear cell extracts were incubated with the immobilized DNA probes, the resin washed and the bound proteins eluted prior to protein labelling with ICAT reagents. The labelled proteins were combined, digested with trypsin, fractionated, and the peptides identified by mass spectrometry. As expected, the authors observed a strong enrichment of CTCF with the wild-type probes as compared to the mutant ones. Interestingly, the cohesin subunit SCC3 was also enriched with the wild type probes, pointing to a common localization of CTCF and cohesin at some genomic loci. ChIP-chip experiments confirmed that SCC3 colocalized with a subset of the CTCF sites, and further studies showed that recruitment of SCC3 was mediated by CTCF. This discovery is of clinical interest, as some patients suffering from Cornelia de Lange syndrome are known to carry mutations in cohesin subunits resulting in improper gene regulation. Interestingly, the improper recruitment of cohesin and CTCF to CpG islands could explain some of the abnormalities observed in patients suffering from Cornelia de Lange syndrome.

2.2.2 Chromatin purification via endogenous DNA sequences

The previous examples all relied on the introduction of exogenous DNA probes to perform affinity purification. The purification of endogenous protein-DNA complexes localized to particular genomic loci was achieved by Déjardin *et al.* who developed a novel method termed proteomics of isolated chromatin segments (PiCh) [27]. PiCh experiments are performed by first fixing cells with formaldehyde to stabilize protein-DNA interactions, followed by cell lysis and chromatin shearing before a specific desthiobiotin-labelled nucleic acid probe (locked nucleic acid) is hybridized to the chromatin segments of interest (Figure 4B). Streptavidin beads capture chromatin segments which are washed extensively and the proteins are eluted by heating the samples at 65°C in the presence of biotin. The purified proteins are then separated on gel and identified by mass spectrometry (Figure 4B). The authors used the PiCh approach to target telomeric sequences in three different cell types and compared these results to purifications using scrambled nucleic acid probes. This resulted in the detection of over 200 proteins reproducibly associated with telomeric sequences, including 33 previously characterized telomeric proteins. Immune-FISH confirmed telomere localization for seven of the eight tested new putative telomeric proteins. The major caveat of the PiCh approach remains its inability to study single copy genomic loci, as opposed to repetitive elements found at telomeres or at pericentric heterochromatic regions [27].

One possible solution to the study of single copy genomic loci by proteomics approaches is to increase the quantity of genomic material in the system under study. However, polyploidy results in severe phenotypes in most eukaryotes [28] and is not a practical solution. Or is it? The model organism *S. cerevisiae* can sustain extra chromosomes, a feat that has long been used to study chromosome segregation and stability [29]. Recently, Akiyoshi *et al.* utilized a small minichromosome which includes tandem repeats of lactose operators (*LacO*) and the centromere from Chromosome III (*CEN3*) to perform proteomic studies of kinetochore assembly *in vivo* [30]. Briefly, budding yeast cells harbouring a FLAG-tagged lac repressor (*lacI*) and the minichromosome described above, or a control minichromosome with a mutated *CEN3* region incapable of assembling kinetochores, were purified using magnetic anti-FLAG beads utilizing the affinity of the *lacI* protein for the *LacO* sequence. The proteins associated with the minichromosomes were digested with trypsin into peptides which were fractionated using ion

exchange prior to their analysis by mass spectrometry. After combining this purification strategy with quantitative proteomics, 518 proteins were identified, 44 of which were greatly enriched with the wild-type minichromosomes, including most known kinetochore components. The authors also observed that a previously unidentified kinetochore component, Fin1 (a regulatory subunit of the PP1 catalytic subunit Glc7), was strongly enriched (>13 fold) in their analysis. The authors went on to show that mislocalization of Fin1 results in severe mitotic phenotypes and that numerous control mechanisms exist to regulate Fin1 function. As such, the purification of minichromosomes was shown to be a powerful approach to characterize protein complexes associated with particular genomic loci in budding yeast.

As the examples discussed in the section above have shown, protein-DNA interactions can now be detected and studied through numerous proteomic approaches. Critically, these tools enable researchers to study chromatin constituents as a whole or to specifically target proteins associated with a particular genomic locus. Despite the progress and impressive results obtained from the studies reviewed in this section, sensitivity remains a problem for the characterization of protein-DNA interactions. This has been circumvented, in part, by increasing the amount of starting material. For instance Déjardin *et al.* [27] used approximately 3×10^9 HeLa cells per PICCh experiments while Ohta *et al.* [23] used 7.5×10^9 DT40 chicken cells for the characterization of mitotic chromosomes. Unfortunately, the need for a very large amount of starting material restricts these protocols to a handful of cell types. While improvements in mass spectrometers themselves have enhanced the sensitivity of numerous assays (such as the study of chromatin-enriched fractions), it is clear that other technical improvements are necessary to enable a wider use of the methods discussed in this section.

3. Chromatin study using protein-centric approaches

3.1.1 Purification of histone proteins and their associated proteins

Histones are master regulators of chromatin structure and function. From a structural point of view, as nucleosome components, histones participate not only in reducing the unruly

size of the genomic material, but also in reducing damage to it [31]. In addition, the positioning of histone proteins on DNA actively participates in regulating gene expression by either allowing or preventing access of the transcription machinery to the subjacent genomic material. Thus, regulation of histone synthesis, localization and post-translational modifications are critical to chromatin biology.

Early on, it was recognized that histone proteins are found in association with larger protein complexes, which regulate their biological functions [32]. To better understand the underlying mechanisms of histone regulation and deposition, a number of groups embarked on the identification of histone interaction partners. For instance, Tagami *et al.* identified interaction partners for histones H3.1 and H3.3, two variants of histone H3 with distinct cell-cycle expression profiles and functions [33]. Histone H3.1 and H3.3 were tagged at their C-termini with dual FLAG and HA tags, stably expressed in HeLa cells, and recovered from nuclear extracts by a dual purification, first with anti-FLAG beads and subsequently with anti-HA beads. The highly purified protein material was gently eluted from the affinity matrix with a competitive HA peptide and the recovered proteins resolved by SDS-PAGE, silver stained and the gel bands identified by mass spectrometry. Despite the high homology between H3.1 and H3.3, both common and distinct interaction partners were detected for each histone variant. For instance, the histone chaperones ASF1A and ASF1B and the lysine acetyltransferase HAT1 were common to both purifications, suggesting a common acetylation mechanism prior to histone deposition. On the other hand, histone H3.1 interacted strongly with the CAF-1 complex, a DNA replication-dependent histone chaperone, while histone H3.3 preferentially interacted with HIRA, a DNA replication-independent histone chaperone. A few years later, Lewis *et al.* [34] and Drané *et al.* [35] also utilized HeLa cells stably expressing Histone H3.1 and H3.3 with a C-terminal FLAG and HA tags to further characterize the interaction partners of H3.1 and H3.3. Using mass spectrometers with a greater sensitivity, both groups identified Daxx and ATRX to be specifically associated with histone H3.3, but not with H3.1, and observed that the Daxx-ATRX complex is required for deposition of H3.3 at telomeric regions [34] and at pericentric chromatin [35]. As such, a single histone variant has access to a diversity of interaction partners. An in-depth characterization of proteins involved in the synthesis, nuclear

import and deposition of H3.1 in chromatin was recently described by Campos and colleagues [36]. The authors initially employed the tandem FLAG and HA purification described above [33], but purified the bait from cell lysate fractionated into cytosolic, nuclear or chromatin-bound fractions. Using this approach, the authors identified a mixture of histone chaperones copurified with H3.1, such as tNASP, sNASP, RbAp46 and ASF1B, hinting at the presence of multiple distinct complexes within the affinity purified material. To distinguish these H3.1 subcomplexes, affinity purified H3.1 complexes were fractionated by gel filtration, individual fractions resolved by SDS-PAGE and silver stained. The additional fractionation enabled four distinct protein complexes containing H3.1 to be detected. The identification of these complexes, along with follow-up work, generated a clearer picture of the mechanism by which histone H3.1 is handled by chaperones from its synthesis all the way to nuclear import. This work shows the value of performing iterative biochemical purification to characterize histones and their associated proteins.

3.1.2 The impacts of AP protocol on the purification of chromatin associated proteins.

Depending on the question to be addressed (namely whether chromatin-associated or soluble interactions are desired), affinity purification can be coupled to protocols that solubilize chromatin or not. For instance, to identify proteins involved in the deposition of the centromeric histone H3 variant CENP-A, Foltz *et al.* performed affinity purification of dually tagged and stably expressed CENP-A (or histone H3.1 as a control) from chromatin-free cell extract [37] (Figure 5). These cell extracts were prepared by solubilizing cytoplasmic and nuclear membranes with polyamine and digitonin, followed by removal of chromatin by centrifugation. Tandem affinity purification of the histone variants was performed and the eluted proteins were digested with trypsin and identified by multidimensional chromatography coupled to mass spectrometry. The histone chaperone HJURP was found to be strongly associated with CENP-A, but not with the canonical histone H3.1. This novel interaction was validated through *in vitro* binding assays and found to be required for proper localization of CENP-A to centromeres. Note that no centromeric proteins were identified as interaction partners of CENP-A, since chromatin was removed from the cell extract during the sample

1
2
3 preparation. To characterize the chromatin-associated interaction partners of CENP-A, Foltz. *et*
4 *al.* modified their approach to ensure proper chromatin solubilization before affinity
5 purification [38] (Figure 5). Briefly, nuclear extracts were prepared from the same HeLa cell
6 lines expressing TAP-tagged versions of CENP-A and H3.1. The chromatin was then digested into
7 smaller soluble fragments with micrococcal nuclease, which digests exposed DNA between
8 nucleosomes. The resulting extract was used for tandem affinity purification and MS analysis as
9 described above. In contrast to the results obtained from the soluble extracts, CENP-A
10 purification from chromatin fractions contained multiple known centromeric proteins, as well
11 as three proteins not previously known to be localized to centromeric regions. The authors
12 demonstrated that these three proteins are localized to the centromere and that they were
13 required for CENP-A deposition at the centromere. The studies of CENP-A by Foltz *et al.* [37-38]
14 exemplified the impact that sample preparation prior to affinity purification can have on the
15 analysis of protein-protein interaction for baits associated with chromatin.
16
17

18 Over the years, a number of different approaches have been developed to solubilize
19 chromatin. As seen above, Foltz *et al.* utilized micrococcal nuclease to solubilize chromatin [38]
20 while others, like Du *et al.* utilized DNase I [39], which can completely digest DNA. Another
21 method commonly used to solubilize chromatin is sonication, which shears DNA by creating
22 small unstable cavities (i.e. bubbles) within the samples that release upon rupture a sufficient
23 amount of energy to fragment DNA. We recently described a method for affinity purification,
24 termed modified chromatin immunopurification or mChIP, which enables the purification of
25 chromatin-associated protein complexes in budding yeast [5]. mChIP relies on chromatin
26 solubilization, with a focus on sample preparation to minimize precipitation of chromatin
27 fragments [5]. Briefly, yeast cells expressing a TAP-tagged bait of interest under the control of
28 their native promoter are lysed, mildly sonicated, and clarified by centrifugation at low speed
29 before being subjected to a single affinity purification step. To date, mChIP purifications
30 coupled to mass spectrometry have been performed on more than 100 different chromatin-
31 related baits, including histones proteins and their chaperones [5, 40]. As part of the mChIP
32 development, we directly tested the impact of chromatin fragment size on the affinity
33 purification. To achieve this, cell extracts from TAP-tagged H2A.Z variant or untagged cells were
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 aliquoted in three fractions, and the chromatin solubilized using sonication, micrococcal
4 nuclease or DNase I prior to affinity purification [5] (Figure 6). In effect, this created a controlled
5 system in which chromatin fragments were large (sonication), medium (micrococcal nuclease)
6 or small/absent (DNase I). We observed that as the size of chromatin fragments decreased, the
7 quantity of proteins co-purifying with the tagged baits was also diminished. Furthermore, we
8 observed that when chromatin-associated baits were purified from extract containing large
9 DNA fragments, indirect interactions (i.e. via DNA) were observed which were minimized in the
10 preparations treated with nucleases. For instance, an interaction between H2AZ and the DNA
11 topoisomerase Top2 was observed from sonicated extract, but was reduced in the micrococcal
12 nuclease-treated sample and absent from the DNase I sample [5]. Conversely, the association
13 with proteins known to be physically associated with H2AZ, such as Nap1, was unaffected by
14 the size of chromatin fragments.
15
16

17 As the examples above illustrated, histones and other chromatin-associated proteins
18 can be effectively characterized by various AP-MS approaches. Still, it is critical to recognize
19 that every protocol has both benefits and limitations when it comes to defining the interactome
20 of chromatin-associated baits. Protocols that utilize high salt, high speed centrifugation or
21 multiple purification steps gain in specificity but often result in the loss of chromatin-dependent
22 interactions. On the other hand, protocols that solubilize chromatin effectively often result in
23 higher background contamination, rendering their analysis more difficult unless appropriate
24 controls are in place. It is thus critical to accept that no method is perfect or universal and that
25 sample preparation procedure will impact the protein-protein interactions detected for various
26 baits.
27
28

49 **3.2 Histone post-translational modifications and their impact on protein-protein interactions**

50
51

52 Histone proteins and their N-terminal tails are subject to a wide array of post-
53 translational modifications that have critical biological implications [41]. Identifying the
54 functions of individual histone modifications has been challenging since they can occur as
55 various states (e.g. lysine can be mono-, di- or trimethylated), on a single or multiple sites (in a
56
57
58
59
60

1
2
3 combinatorial manner), and often depend on precise cellular events. Great progress in defining
4 the genomic localization of particular histone PTMs was accomplished through the use of
5 chromatin immunoprecipitation coupled to gene array technology (ChIP-seq) which clearly
6 established that histone PTMs are not uniformly distributed [42-43]. Despite this success,
7 identifying which proteins are recruited to, or excluded from, chromatin by specific histone
8 marks remains difficult.
9
10
11
12
13
14
15
16
17

18 3.2.1 Purification via soluble histone tails 19

20 In recent years, various proteomic approaches have been utilized to untangle the
21 function of specific histone PTMs. As described below, the use of synthetic peptides as baits for
22 affinity purifications has been a particularly successful approach to study histone tail PTMs.
23 Synthetic peptides containing a precise combination of PTMs mimic well the N-terminal tail of
24 histones which are flexible and unfolded. Also, they enable researchers to study the
25 combinatorial effect of multiple PTMs on a single histone. Zegerman *et al.* first reported the use
26 of synthetic peptides to elucidate the roles of specific PTMs on histone tails [44]. The authors
27 used three histone H3 peptides trimethylated at either lysine 4 (H3K4Me3), 9 (H3K9Me3) or at
28 both residues along with two unmethylated control peptides corresponding to the tail of
29 histone H3 or H4. The peptides were immobilized on agarose beads, incubated with HeLa
30 nuclear extract, the proteins non-specifically bound to the beads washed away, and the protein
31 bound to the peptide eluted by the addition of excess soluble histone peptides. The purified
32 proteins were separated on gel and identified by mass spectrometry. Interestingly, a number of
33 proteins, and particularly the NuRD complex (for Nucleosome Remodeling and Deacetylase)
34 were found to be associated with the unmethylated histone H3 peptide and were lost upon
35 lysine 4 but not lysine 9 trimethylation.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The use of modified peptides as baits for affinity purification experiments was coupled
with quantitative mass spectrometry by Vermeulen *et al.* to improve the characterization of
proteins coupled to histone H3 tails trimethylated at lysine 4 [45]. Peptides corresponding to
the first 17 amino acids of histone H3, either unmethylated or trimethylated at lysine 4 were

synthesized with an additional biotinylated lysine for immobilization on streptavidin magnetic beads (Figure 7A). Nuclear extracts from metabolically-labelled HeLa S3 cells were prepared and individually incubated with the unmethylated or trimethylated peptide. The beads were then washed to remove non-specifically associated proteins, and then pooled. Associated proteins were eluted, resolved with SDS-PAGE gel, then the entire lane was in-gel digested and analyzed by LC-MS/MS. While the majority of proteins identified were not enriched with either peptide, proteins known to bind H3K4Me3 such as BPTF and CHD1, and the entire TFIID complex, were greatly enriched in the purification from the trimethylated peptide. *In vivo*, the chromatin association of TFIID was dependent on H3K4Me3, and was mediated by the plant homeodomain (PHD) of TAF3. Combinatorial PTMs on a single histone H3 tail were further analyzed in pulldown experiments using peptides harbouring either asymmetric dimethylation of arginine 2 (H3R2Me2a) or acetylation of lysine 9 and 14 (H3K9AcK14Ac). Specifically, nuclear extracts from HeLa cells labelled with three isotopic forms of SILAC were purified using peptides corresponding to H3/H3K4Me3/H3R2Me2aK4Me3 or H3/H3K4Me3/H3K4Me3K9AcK14Ac. Quantitative proteomics demonstrated that asymmetric dimethylation of arginine 2 reduced the binding of TFIID to histone H3 while acetylation of lysine 9 and 14 increased it.

Recently, Vermeulen *et al.* extended this work to all of the major trimethylation sites on lysines 4, 9, 27 and 36 of histone H3 and to lysine 20 of histone H4 [46]. The peptide pulldowns were performed as before [45] but a second set of purifications in which isotopic labels on the HeLa nuclear extract were inverted was also included. As such, proteins were listed as “specific” only if they were enriched in both the “forward” and “reverse” experiments. More than 600 distinct proteins were identified from individual purifications, with only a small subset (~10 to 60 proteins) showing enrichment with the modified histone peptides studied, including novel proteins. These previously uncharacterized proteins were next tagged with GFP, and their interaction partners were identified. For instance, C17orf49 protein was found to strongly bind H3K4Me3, but also the NuRF complex, a known binder for H3K4Me3. The combination of modified histone tail peptide pulldown with traditional affinity purification provided a detailed view of protein complexes associated with trimethylated lysine residues on histone H3 and H4.

The aforementioned examples clearly demonstrate the value of employing modified synthetic peptides to identify the “readers” of histone PTMs.

3.2.2 Purification via *in vitro*-assembled nucleosomes

Linear peptides are not perfect mimics of particular chromatin environments since the impact of the nucleosome structure and DNA sequence are not conserved. However, Shogreen-Knaak *et al.* have reported a protocol to generate *in vitro*-assembled nucleosome arrays containing a single or multiple histone PTMs [47]. These nucleosome arrays were shown to be useful in enzymatic assays [47] and for structural studies of nucleosomes [48]. These nucleosome arrays are also very attractive affinity baits and were utilized by Bartke *et al.* to study the impact of histone lysine trimethylation and DNA methylation on protein-protein interactions [49]. The method, termed SILAC nucleosome affinity purification (SNAP), functions as follows: first, chemically-modified histones are prepared, assembled into nucleosomes with biotinylated DNA [50] and finally immobilized on streptavidin beads (Figure 7B). Nuclear extracts from metabolically-labelled HeLa S3 cells are then incubated with recombinant nucleosomes containing methylated or unmethylated DNA, and subsequently, the beads are washed, combined and the bound proteins eluted in Laemmli sample buffer prior to SDS-PAGE, in gel digestion and mass spectrometry. Again, isotopic labels can be inverted to increase confidence in the association of a protein with a particular nucleosome. This protocol was employed to study protein binding to nucleosomes containing H3K4Me3, H3K9Me3, H3K27Me3 with or without methylated DNA. Agglomerative hierarchical clustering of the resulting data showed numerous interesting trends in the dataset. For instance, all purifications performed with nucleosomes containing methylated DNA showed a strong reduction in the binding of USF1 and USF2 proteins, but a strong increase in the binding of MeCP2, consistent with previous literature [51]. A similar situation was observed for the three lysine trimethylation sites studied on histone H3. A detailed characterization of the binding of KDM2A to H3K9Me3 (which is disrupted by DNA methylation) indicated that purified 3XFLAG-KDM2A could not bind to H3K9Me3-containing nucleosomes in an *in vitro* binding assay, and led to the discovery that

1
2
3 HP1 α was a necessary bridging factor, enabling proper genomic localization of KDM2A on
4 chromatin *in vivo*. Together this suggests a model where KDM2A recognizes H3K9Me3 through
5 its association with HP1 α , and DNA methylation through its DNA binding domain, concurrently
6 enabling an integration of these two modifications. This example illustrates the value of
7 studying histone PTMs in their native environment since the nucleosome structure, as well as
8 the associated DNA sequence, impact their functions.
9
10
11
12
13

14 15 16 17 18 **3.3 Functional proteomic studies of non-histone chromatin bound proteins** 19

20
21 Numerous non-histone proteins, such as chromatin remodelers, transcription factors
22 and histone modifiers are found in a network of protein complexes linking the multiple
23 functions of chromatin [52]. Deciphering the interplay between these chromatin-associated
24 complexes is critical to understanding chromatin itself. To date, numerous AP-MS studies
25 targeting specific chromatin-bound proteins have been performed and are beyond the scope of
26 this review. Most mammalian open reading frames (ORFs) have yet to be targeted by AP-MS
27 [53] however, preventing us from obtaining a clear picture of the myriad of subcomplexes
28 present in chromatin. Furthermore, even in budding yeast, where most ORFs have been
29 targeted by AP-MS experiments [54-55] numerous novel protein-protein interaction were
30 detected when AP-MS methodology specifically designed for chromatin bound proteins was
31 used [5, 40]. As such, a complete view of chromatin has yet to be achieved.
32
33

34 In the past few years, significant progress has been made, including the recent effort
35 spearheaded by the Nuclear Receptor Signalling Atlas consortium, in identifying endogenous
36 protein complexes implicated in gene expression modulation [56]. In total, Malovannaya *et al.*
37 utilized 1796 distinct antibodies in 3290 AP-MS experiments to characterize human nuclear
38 receptor coregulators, proteins working in concert with transcription factors to regulate gene
39 transcription. A stringent affinity purification procedure from HeLa S3 nuclear extract was
40 utilized, which included two ultracentrifugation steps for the removal of insoluble material and
41 of chromatin, limiting the background of the procedure [57]. Despite the scope of the work
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

presented by Malovannaya *et al.* (in which over 11 000 distinct gene products were identified by mass spectrometry) there is still much to be understood about the composition, localization and dynamic nature of protein complexes involved in chromatin biology.

4. Conclusion and future perspectives

Chromatin, with its diverse composition and structure, is involved in the regulation of a myriad of cellular functions. In recent years, technical improvements have rendered possible detailed characterization of chromatin, and of its protein constituents, with regards to its structure (by chromosome conformation capture [58]), protein localization (by ChIP coupled with sequencing [59]) and, as described in the previous sections, by proteomics studies. As we have seen, we are currently capable of identifying protein-protein and protein-DNA interconnections within chromatin. Furthermore, recent gains in sensitivity of the proteomic techniques have enabled the detection of subtle changes in the protein composition of chromatin following various stimuli. Thus, we have reached a point where we can generate detailed maps of chromatin which have already proven to be invaluable in understanding biological systems. Still as a result of its heterogeneous and dynamic nature, no study can yet claim to have "fully" characterized all physical interactions of proteins within chromatin. For instance, it remains impossible to effectively study protein complexes at a single genomic locus. Moreover, functional proteomic studies targeting chromatin components are providing an "average" view rather than loci-specific information. This problem is exacerbated in the study of histone PTMs in which modifications can act not only in a *cis* or *trans* manner, but also combinatorially, and as such act differently across the genome. Despite the challenges posed by chromatin to proteomic studies, protocols are emerging that enable a more holistic view of chromatin. As the proteomic "toolbox" keeps improving, the value of established protocols for the study of chromatin structure and function raises as well. These technical improvements, coupled to "tested and true" biochemical protocols, have already greatly improved our comprehension of chromatin. While some of the most fundamental aspects of chromatin biology remain to be elucidated, we

1
2
3 are hopeful that the near future will bring significant technical innovations in the study of
4 protein-protein interactions, resulting in a better characterization of chromatin.
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Figure 1: Overview of chromatin and of its constituents. (A) In cell, DNA is found to be wrapped around histone proteins generating a macromolecule termed nucleosome. The N-terminal tails of histone proteins protrude from nucleosomes and can be modified post-translationally by a wide array of enzymes (B). Histone PTMs function within a Writer-Reader-Eraser system in which PTMs enables the ordered recruitment of factors to chromatin. KAT, lysine acetyltransferase; KDAC, lysine deacetylase; BRCT, BRCA1 C-Terminus domain, a phosphoprotein binding domain.

Figure 2: Overview of the preparation and analysis of chromatin-enriched fractions. The low solubility of chromatin in hypotonic buffer is utilized to segregate away DNA associated proteins from the bulk of the proteome using centrifugation. Subsequently, both fractions are resolved on by SDS-PAGE gel and submitted to mass spectrometry analysis. Please note that in this and all other displayed figures, fractionation may involve other approaches, for example peptide-level separation by ion exchange. Furthermore, quantitative proteomics approaches can be incorporated in the purification schemes displayed.

Figure 3: Protocol for the isolation of mitotic chromosomes from nocodazole arrested cells and for their analysis by LC-MS/MS. Following lysis of mitotic cells, condensed chromosomes are isolated using sedimentation through both sucrose and percoll gradients prior to mass spectrometric analysis.

Figure 4: Isolation of chromatin-binding proteins using *in vitro* and *in vivo* approaches. (A) Immobilized DNA baits can be used to assemble protein-DNA complexes *in vitro* prior to identification through proteomic means. (B) The isolation of *in vivo* assembled protein-DNA complexes has been successfully performed from repetitive elements such as telomeres using the PICh approach.

1
2
3
4
5 Figure 5: Isolation of the centromeric histone variant CENP-A from the soluble and chromatin
6 bound fractions. By tailoring the sample preparation of tagged CENP-A prior to affinity
7 purification, Foltz *et al.* successfully , mapped both the soluble [37] and chromatin associated
8 interaction partners of CENP-A [38].
9
10
11
12
13
14
15

16 Figure 6: Chromatin fragments size influence protein-protein interactions detected by AP-MS.
17 The mChIP protocol was utilized to test the impact of chromatin fragment size on the
18 identification of interaction partners for Htz1-TAP. Reduction of chromatin size resulted in a
19 reduction of indirect (i.e. through chromatin) interaction but did not affect direct protein-
20 protein interaction.
21
22
23
24
25
26
27
28
29

30 Figure 7: *In vitro* identification of protein recruitment to histone PTMs. Immobilized histone tail
31 peptides and *in vitro* assembled nucleosomes can be used as baits for affinity purification,
32 enabling the identification of proteins binding to particular histone modifications.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Acknowledgement:
4
5

6 We would like to thank members of the Gingras lab for their critical comments and suggestions.
7 This work was supported by funding from the CIHR to A-C. G. (MOP-84314) and to T.P. (MOB-
8 6849) and to A.-C.G. and T.P. through the Ontario Research Fund GL2 program. A.-C.G. holds
9 the Canada Research Chair in Functional Proteomics and the Lea Reichmann Chair in Cancer
10 Proteomics. J.-P.L was supported by a post-doctoral fellowship from the Natural Sciences and
11 Engineering Research Council of Canada.
12
13

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1
2
3 **Table of abbreviations:**
4
5
6
7

8 AP-MS: affinity purification coupled to mass spectrometry
9
10

11 CEN3: centromere from chromosome III
12
13

14 ChIP: chromatin immunoprecipitation
15
16

17 ChIP-chip: chromatin immunoprecipitation coupled to gene array technology
18
19

20 H3K4Me3: histone H3 trimethylated at lysine 4
21
22

23 H3K9Me3: histone H3 trimethylated at lysine 9
24
25

26 H3K9AcK14Ac: histone H3 acetylated at lysine 9 and 14
27
28

29 ICAT: isotope-coded affinity tags
30
31

32 ISD: interferon stimulatory DNA
33
34

35 LacI: lac repressor
36
37

38 LacO: lactose operators
39
40

41 MNase: micrococcal nuclease
42
43

44 ORF: open reading frame
45
46

47 PARP: poly (ADP-ribose) polymerase
48
49

50 PHD: plant homeodomain
51
52

53 PICh: proteomics of isolated chromatin segments
54
55

56 PTM: post-translational modification
57
58

59 SILAC: stable isotope labelling by amino acids in cell culture
60
61

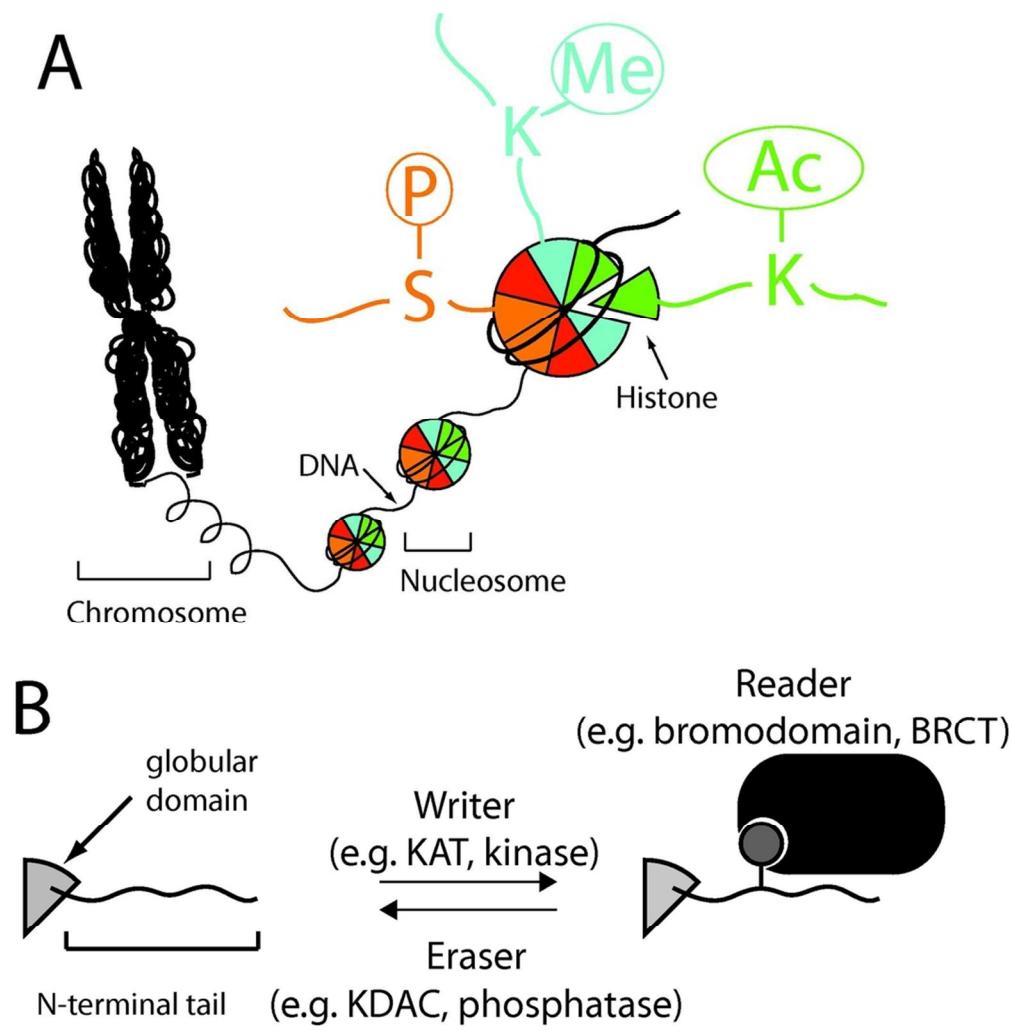
62 SNAP: SILAC nucleosome affinity purification
63
64

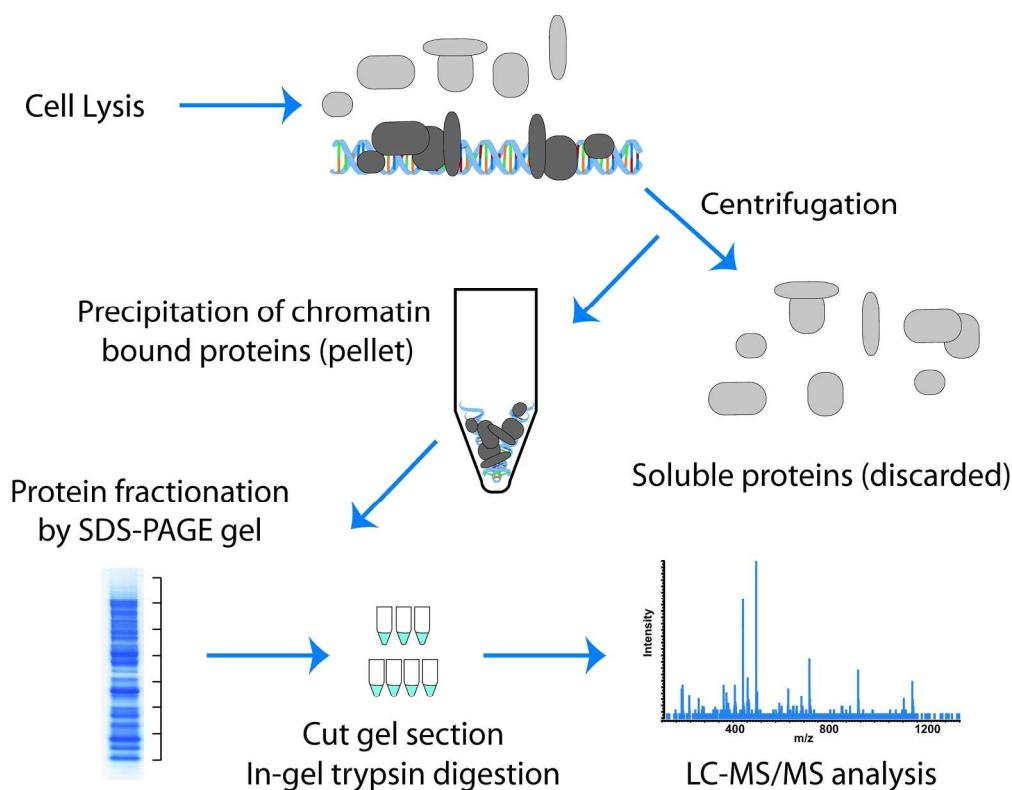
1
2
3 References:
4
5

- [1] Kumar, A., Snyder, M., Protein complexes take the bait. *Nature* 2002, **415**, 123-124.
- [2] Lamond, A. I., Mann, M., Cell biology and the genome projects a concerted strategy for characterizing multiprotein complexes by using mass spectrometry. *Trends Cell Biol* 1997, **7**, 139-142.
- [3] Gingras, A. C., Gstaiger, M., Raught, B., Aebersold, R., Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* 2007, **8**, 645-654.
- [4] Jensen, L. J., Bork, P., Biochemistry. Not comparable, but complementary. *Science* 2008, **322**, 56-57.
- [5] Lambert, J. P., Mitchell, L., Rudner, A., Baetz, K., Figeys, D., A novel proteomics approach for the discovery of chromatin-associated protein networks. *Mol Cell Proteomics* 2009, **8**, 870-882.
- [6] Campos, E. I., Reinberg, D., Histones: annotating chromatin. *Annu Rev Genet* 2009, **43**, 559-599.
- [7] Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., Richmond, T. J., Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 1997, **389**, 251-260.
- [8] Strahl, B. D., Allis, C. D., The language of covalent histone modifications. *Nature* 2000, **403**, 41-45.
- [9] Deribe, Y. L., Pawson, T., Dikic, I., Post-translational modifications in signal integration. *Nat Struct Mol Biol* 2010, **17**, 666-672.
- [10] Bonner, J., Dahmus, M. E., Fambrough, D., Huang, R. C., et al., The Biology of Isolated Chromatin: Chromosomes, biologically active in the test tube, provide a powerful tool for the study of gene action. *Science* 1968, **159**, 47-56.
- [11] Magee, B. B., Paoletti, J., Magee, P. T., Separation of lymphocyte chromatin into template-active fractions with specificity for eukaryotic RNA polymerase II or prokaryotic RNA polymerase. *Proc Natl Acad Sci U S A* 1975, **72**, 4830-4834.
- [12] Mendez, J., Stillman, B., Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol Cell Biol* 2000, **20**, 8602-8612.
- [13] Wysocka, J., Reilly, P. T., Herr, W., Loss of HCF-1-chromatin association precedes temperature-induced growth arrest of tsBN67 cells. *Mol Cell Biol* 2001, **21**, 3820-3829.
- [14] Shioi, Y., Eisenman, R. N., Yi, E. C., Donohoe, S., et al., Quantitative proteomic analysis of chromatin-associated factors. *J Am Soc Mass Spectrom* 2003, **14**, 696-703.
- [15] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., et al., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999, **17**, 994-999.
- [16] Chou, D. M., Adamson, B., Dephoure, N. E., Tan, X., et al., A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. *Proc Natl Acad Sci U S A* 2010, **107**, 18475-18480.
- [17] Kubota, T., Hiraga, S., Yamada, K., Lamond, A. I., Donaldson, A. D., Quantitative proteomic analysis of chromatin reveals that Ctf18 acts in the DNA replication checkpoint. *Mol Cell Proteomics* 2011, **10**, M110 005561.
- [18] Lewis, C. D., Laemmli, U. K., Higher order metaphase chromosome structure: evidence for metalloprotein interactions. *Cell* 1982, **29**, 171-181.
- [19] Morrison, C., Henzing, A. J., Jensen, O. N., Osheroff, N., et al., Proteomic analysis of human metaphase chromosomes reveals topoisomerase II alpha as an Aurora B substrate. *Nucleic Acids Res* 2002, **30**, 5318-5327.
- [20] Gassmann, R., Henzing, A. J., Earnshaw, W. C., Novel components of human mitotic chromosomes identified by proteomic analysis of the chromosome scaffold fraction. *Chromosoma* 2005, **113**, 385-397.
- [21] Takata, H., Uchiyama, S., Nakamura, N., Nakashima, S., et al., A comparative proteome analysis of human metaphase chromosomes isolated from two different cell lines reveals a set of conserved chromosome-associated proteins. *Genes Cells* 2007, **12**, 269-284.

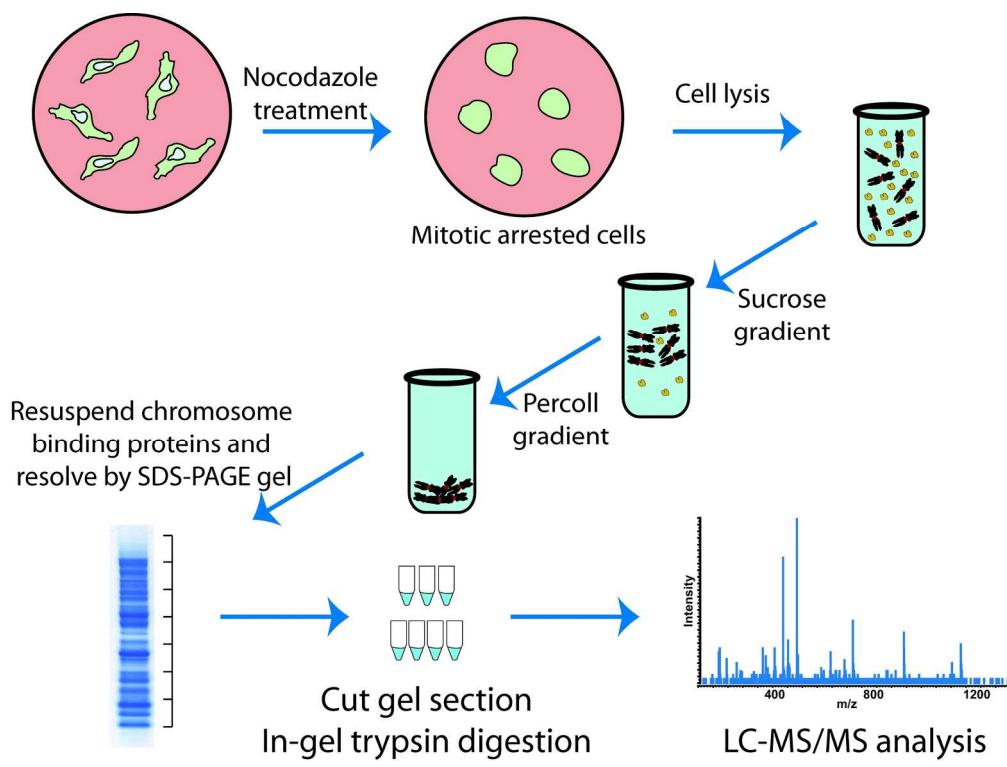
- [22] Rusk, N., Cell biology. Finding the trees in the forest. *Nat Methods* 2010, 7, 869.
- [23] Ohta, S., Bukowski-Wills, J. C., Sanchez-Pulido, L., Alves Fde, L., et al., The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. *Cell* 2010, 142, 810-821.
- [24] Burckstummer, T., Baumann, C., Bluml, S., Dixit, E., et al., An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome. *Nat Immunol* 2009, 10, 266-272.
- [25] Mittler, G., Butter, F., Mann, M., A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome Res* 2009, 19, 284-293.
- [26] Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., et al., CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A* 2008, 105, 8309-8314.
- [27] Dejardin, J., Kingston, R. E., Purification of proteins associated with specific genomic loci. *Cell* 2009, 136, 175-186.
- [28] Comai, L., The advantages and disadvantages of being polyploid. *Nat Rev Genet* 2005, 6, 836-846.
- [29] Clarke, L., Carbon, J., Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* 1980, 287, 504-509.
- [30] Akiyoshi, B., Nelson, C. R., Ranish, J. A., Biggins, S., Quantitative proteomic analysis of purified yeast kinetochores identifies a PP1 regulatory subunit. *Genes Dev* 2009, 23, 2887-2899.
- [31] Feser, J., Truong, D., Das, C., Carson, J. J., et al., Elevated histone expression promotes life span extension. *Mol Cell* 2010, 39, 724-735.
- [32] Eickbush, T. H., Moudrianakis, E. N., The histone core complex: an octamer assembled by two sets of protein-protein interactions. *Biochemistry* 1978, 17, 4955-4964.
- [33] Tagami, H., Ray-Gallet, D., Almouzni, G., Nakatani, Y., Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* 2004, 116, 51-61.
- [34] Lewis, P. W., Elsaesser, S. J., Noh, K. M., Stadler, S. C., Allis, C. D., Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci U S A* 2010, 107, 14075-14080.
- [35] Drane, P., Ouararhni, K., Depaux, A., Shuaib, M., Hamiche, A., The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev* 2010, 24, 1253-1265.
- [36] Campos, E. I., Fillingham, J., Li, G., Zheng, H., et al., The program for processing newly synthesized histones H3.1 and H4. *Nat Struct Mol Biol* 2010, 17, 1343-1351.
- [37] Foltz, D. R., Jansen, L. E., Bailey, A. O., Yates, J. R., 3rd, et al., Centromere-specific assembly of CENP-A nucleosomes is mediated by HJURP. *Cell* 2009, 137, 472-484.
- [38] Foltz, D. R., Jansen, L. E., Black, B. E., Bailey, A. O., et al., The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol* 2006, 8, 458-469.
- [39] Du, Y. C., Gu, S., Zhou, J., Wang, T., et al., The dynamic alterations of H2AX complex during DNA repair detected by a proteomic approach reveal the critical roles of Ca(2+)/calmodulin in the ionizing radiation-induced cell cycle arrest. *Mol Cell Proteomics* 2006, 5, 1033-1044.
- [40] Lambert, J. P., Fillingham, J., Siabazi, M., Greenblatt, J., et al., Defining the budding yeast chromatin-associated interactome. *Mol Syst Biol* 2010, 6, 448.
- [41] Bannister, A. J., Kouzarides, T., Regulation of chromatin by histone modifications. *Cell Res* 2011, 21, 381-395.
- [42] Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., et al., Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 2007, 448, 553-560.
- [43] Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., et al., High-resolution profiling of histone methylations in the human genome. *Cell* 2007, 129, 823-837.

- 1
2
3 [44] Zegerman, P., Canas, B., Pappin, D., Kouzarides, T., Histone H3 lysine 4 methylation disrupts binding
4 of nucleosome remodeling and deacetylase (NuRD) repressor complex. *J Biol Chem* 2002, 277, 11621-
5 11624.
6 [45] Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W., et al., Selective anchoring of TFIID to
7 nucleosomes by trimethylation of histone H3 lysine 4. *Cell* 2007, 131, 58-69.
8 [46] Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., et al., Quantitative interaction proteomics and
9 genome-wide profiling of epigenetic histone marks and their readers. *Cell* 2010, 142, 967-980.
10 [47] Shogren-Knaak, M. A., Fry, C. J., Peterson, C. L., A native peptide ligation strategy for deciphering
11 nucleosomal histone modifications. *J Biol Chem* 2003, 278, 15744-15748.
12 [48] Allahverdi, A., Yang, R., Korolev, N., Fan, Y., et al., The effects of histone H4 tail acetylations on
13 cation-induced chromatin folding and self-association. *Nucleic Acids Res* 2011, 39, 1680-1691.
14 [49] Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., et al., Nucleosome-interacting proteins
15 regulated by DNA and histone methylation. *Cell* 2010, 143, 470-484.
16 [50] Lowary, P. T., Widom, J., New DNA sequence rules for high affinity binding to histone octamer and
17 sequence-directed nucleosome positioning. *J Mol Biol* 1998, 276, 19-42.
18 [51] Adkins, N. L., Georgel, P. T., MeCP2: structure and function. *Biochem Cell Biol* 2011, 89, 1-11.
19 [52] Kadonaga, J. T., Eukaryotic transcription: an interlaced network of transcription factors and
20 chromatin-modifying machines. *Cell* 1998, 92, 307-313.
21 [53] Edwards, A. M., Isserlin, R., Bader, G. D., Frye, S. V., et al., Too many roads not taken. *Nature* 2011,
22 470, 163-165.
23 [54] Gavin, A. C., Aloy, P., Grandi, P., Krause, R., et al., Proteome survey reveals modularity of the yeast
24 cell machinery. *Nature* 2006, 440, 631-636.
25 [55] Krogan, N. J., Cagney, G., Yu, H., Zhong, G., et al., Global landscape of protein complexes in the
26 yeast *Saccharomyces cerevisiae*. *Nature* 2006, 440, 637-643.
27 [56] Malovannaya, A., Lanz, R. B., Jung, S. Y., Bulynko, Y., et al., Analysis of the human endogenous
28 coregulator complexome. *Cell* 2011, 145, 787-799.
29 [57] Malovannaya, A., Li, Y., Bulynko, Y., Jung, S. Y., et al., Streamlined analysis schema for high-
30 throughput identification of endogenous protein complexes. *Proc Natl Acad Sci U S A* 2010, 107, 2431-
31 2436.
32 [58] Dekker, J., The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat
33 Methods* 2006, 3, 17-21.
34 [59] Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., et al., Genome-wide profiles of STAT1 DNA
35 association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* 2007,
36 4, 651-657.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

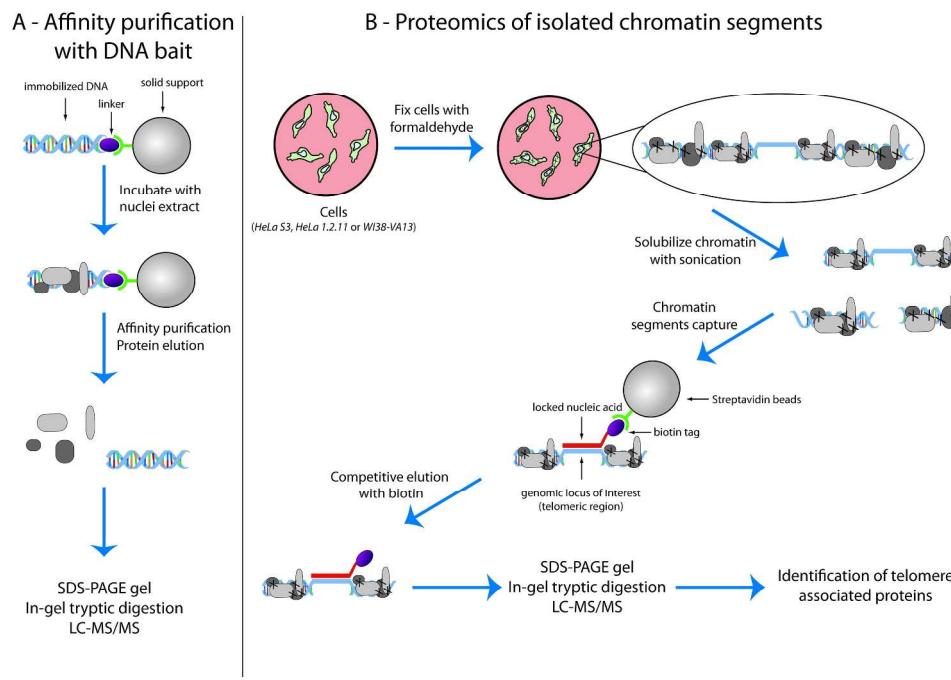




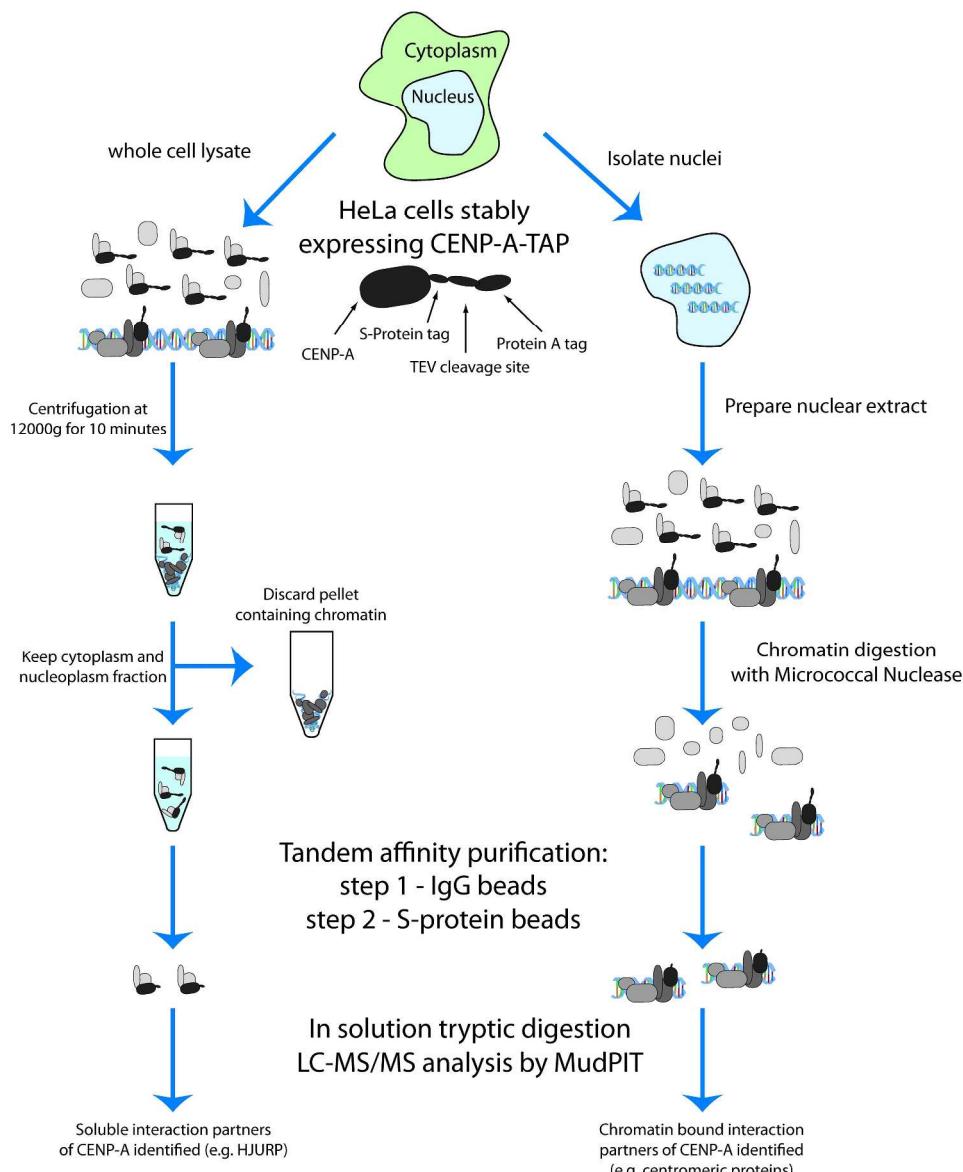
Overview of the preparation and analysis of chromatin-enriched fractions.
209x162mm (300 x 300 DPI)



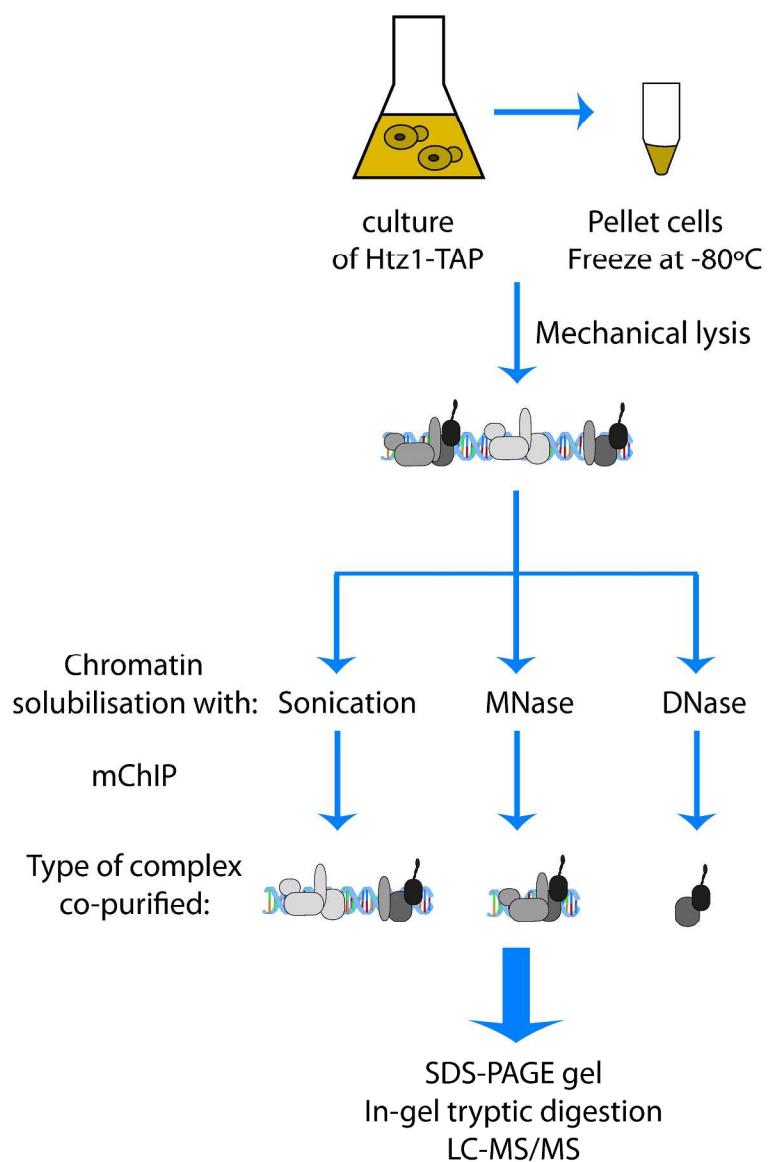
Protocol for the isolation of mitotic chromosomes from nocodazole arrested cells and for their analysis by LC-MS/MS.
192x143mm (300 x 300 DPI)



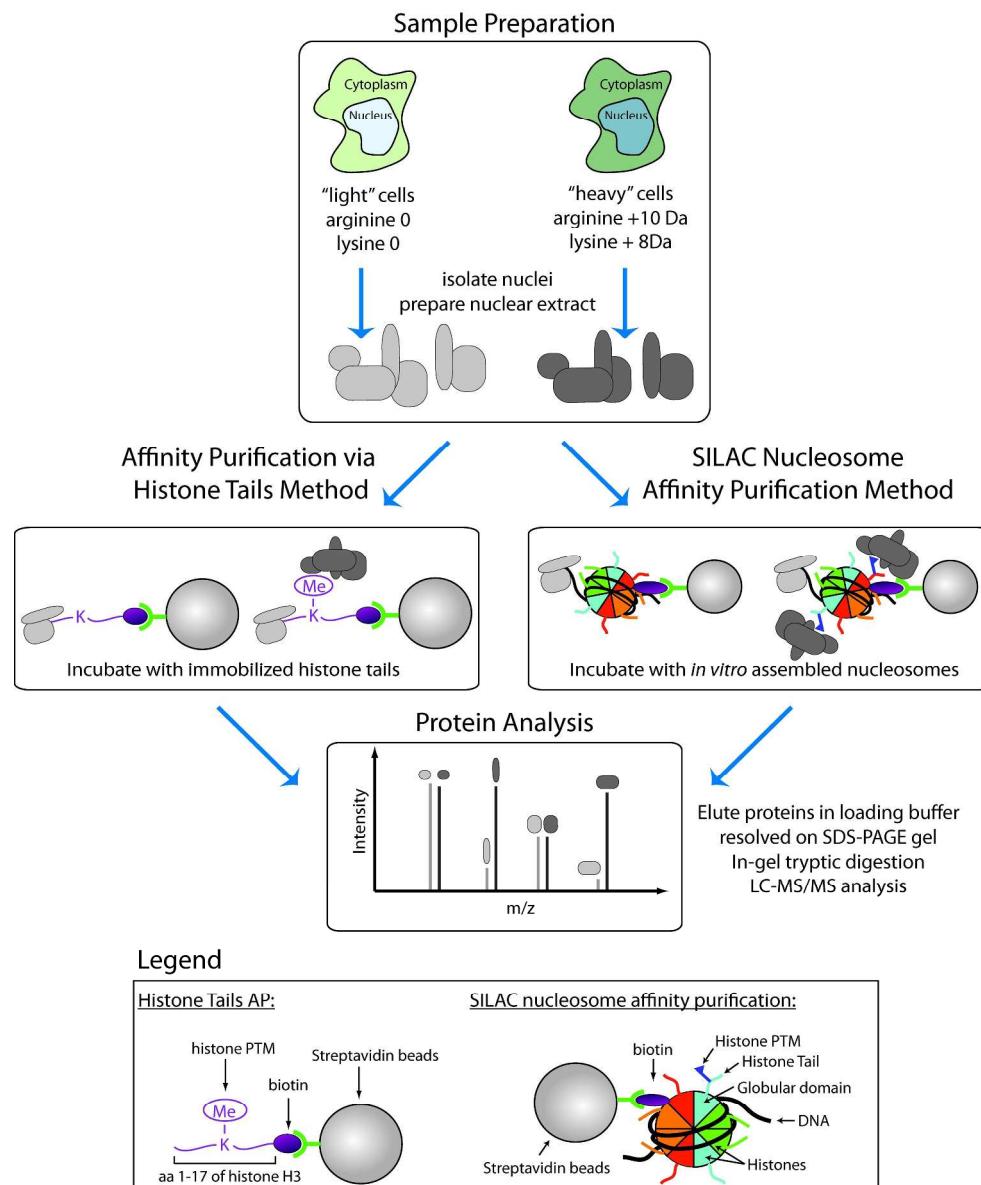
Isolation of chromatin-binding proteins using in vitro and in vivo approaches.
256x176mm (300 x 300 DPI)



Isolation of the centromeric histone variant CENP-A from the soluble and chromatin bound fractions.
302x370mm (300 x 300 DPI)



Chromatin fragments size influence protein-protein interactions detected by AP-MS.
235x351mm (300 x 300 DPI)



In vitro identification of protein recruitment to histone PTMs.
312x382mm (300 x 300 DPI)

SAINT: probabilistic scoring of affinity purification–mass spectrometry data

Hyungwon Choi¹, Brett Larsen², Zhen-Yuan Lin², Ashton Breitkreutz², Dattatreya Mellacheruvu¹, Damian Fermin¹, Zhaojun S Qin^{3,8}, Mike Tyers^{2,4–6}, Anne-Claude Gingras^{2,4} & Alexey I Nesvizhskii^{1,7}

We present ‘significance analysis of interactome’ (SAINT), a computational tool that assigns confidence scores to protein-protein interaction data generated using affinity purification–mass spectrometry (AP-MS). The method uses label-free quantitative data and constructs separate distributions for true and false interactions to derive the probability of a bona fide protein-protein interaction. We show that SAINT is applicable to data of different scales and protein connectivity and allows transparent analysis of AP-MS data.

The analysis of protein complexes and protein interaction networks is very important for biological research. A combination of affinity purification and mass spectrometry (AP-MS) has been increasingly used for both small-scale and large-scale analysis of protein complexes and interaction networks^{1–4}. However, the development of computational tools for the processing of AP-MS data has not kept pace with improvements in experimental approaches. In addition to the general challenge of false positive protein identifications in mass spectrometry-based proteomic data⁵, unfiltered AP-MS datasets contain many nonspecifically binding proteins; filtering these contaminants is the foremost computational challenge.

Whereas early methods filtered the noise using binary data (presence or absence of a protein), newer methods take into account quantitative information embedded in the mass spectrometric data (for example, label-free quantification, such as spectral counts). One such method converts the normalized spectral abundance factor (NSAF) into the posterior probability of a true interaction between a bait-prey pair using simple heuristics, which we term PP-NSAF hereafter⁶. Another method, named

CompPASS, computes scores that adjust observed spectral counts relative to the reproducibility of detection across biological replicates and to the frequency of observing prey proteins in purifications of different baits⁷. Although both approaches can effectively analyze the datasets for which they had been developed, these scores are an empirical transformation of spectral counts without a probability model that can be used to estimate the measurement errors in the data in a transparent manner.

We have recently introduced an advanced approach for statistical analysis of interaction data from AP-MS experiments using label-free quantification, which we termed significance analysis of interactome (SAINT)⁸. As PP-NSAF and CompPASS, we had designed our original SAINT approach to analyze a specific dataset, the yeast kinase and phosphatase interactome. Expanding on this method, here we present a generalized SAINT framework that can be used to compute interaction probabilities in a variety of datasets. The method incorporates negative controls that are commonly generated as a part of the experimental study but can also be applied to large datasets in the absence of such data. We illustrate the methodology and its advantages through the analysis of datasets of different sizes and network density: from a large, sparsely connected network involving human deubiquitinating enzymes to a smaller, highly interconnected network for chromatin remodeling proteins and even to the analysis of a single bait, the protein CDC23.

The aim of SAINT is to convert the label-free quantification (spectral count X_{ij}) for a prey protein i identified in a purification of bait j into the probability of a true interaction between the two proteins, $P(\text{true} | X_{ij})$. The spectral counts for each prey-bait pair are modeled with a mixture distribution of two components representing true and false interactions. Note that these distributions are specific to each bait-prey pair. The parameters for true and false distributions, $P(X_{ij} | \text{true})$ and $P(X_{ij} | \text{false})$, and the prior probability π_T of true interactions in the dataset, are inferred from the spectral counts for all interactions that involve prey i and bait j . SAINT normalizes spectral counts to the length of the proteins and to the total number of spectra in the purification.

In addition to the experimental data for bait proteins, AP-MS data often contain negative controls (Fig. 1a). When these are available, SAINT estimates the spectral count distribution for false interactions directly from the negative controls, which makes the modeling approach semisupervised (Online Methods). SAINT modeling can also be performed without negative control data, so long as a sufficient number of independent baits are profiled and provided that these baits are not densely interconnected. In this case (Fig. 1b), a prey detected in the purification of a bait is

¹Department of Pathology, University of Michigan, Ann Arbor, Michigan, USA. ²Centre for Systems Biology, Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada. ³Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA. ⁴Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. ⁵Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, UK. ⁶Centre for Systems Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK. ⁷Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA. ⁸Present address: Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, USA. Correspondence should be addressed to A.C.G. (gingras@lunenfeld.ca) or A.I.N. (nesvi@med.umich.edu).

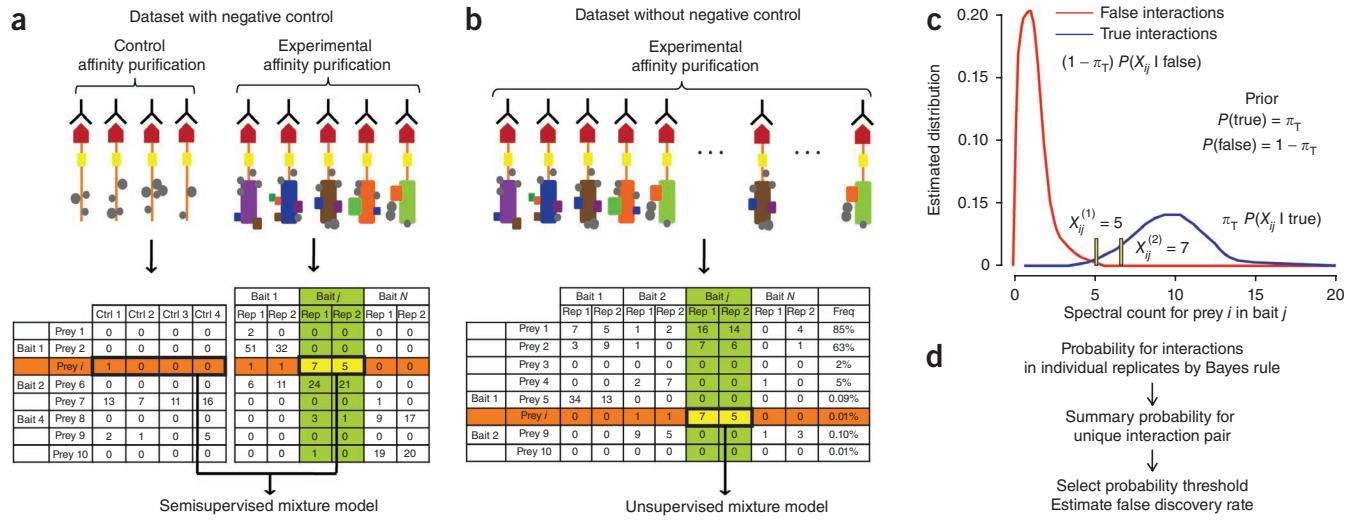


Figure 1 | Probability model in SAINT. **(a,b)** Interaction data in the presence (**a**) and absence (**b**) of control purifications. Schematic of the experimental AP-MS procedure is shown at the top and a spectral count interaction table is illustrated at the bottom. Ctrl, control; rep, replicate; freq, frequency. **(c)** Modeling spectral count distributions for true and false interactions. For the interaction between prey *i* and bait *j*, SAINT uses all relevant data for the two proteins, as shown in the column of the bait (green) and the data in the row of the prey (orange) in **a** and **b**. **(d)** Probability is calculated for each replicate by application of Bayes rule, and a summary probability is calculated for the interaction pair (*i,j*).

scored in reference to the quantitative information for the same prey across purifications of all other baits in the dataset. Although this is possible for large datasets such as the yeast kinase and phosphatase network⁸, and the human deubiquitinating (DUB) enzyme interaction network⁷ (which each contain more than 75 baits), this unsupervised approach involves additional assumptions and separate treatment of high- and low-frequency prey proteins (Online Methods).

One challenge in modeling AP-MS data is the limited number of replicates that are typically available for each bait. SAINT addresses this problem by inferring individual bait-prey interaction parameters through joint modeling of the entire bait-prey data. To this end, SAINT defines a protein-specific abundance parameter and establishes a multiplicative model in the mixture component distributions. In other words, if prey *i* and bait *j* interact, then the ‘interaction abundance’ (the spectral count of the prey *i* in purification with bait *j*) is assumed to be proportional to $\alpha_i \times \alpha_j$. Under this assumption, the protein-specific abundance parameters α_i and α_j can be learned not only from the interaction between the two proteins themselves but also from other bona fide interactions that involve either one of them. The same principle applies to false interactions. Hence, SAINT builds a large number of mixture distributions by pooling data (separate mixture distributions for individual prey-bait pairs), but all models are interconnected through the shared abundance parameters.

The probability distributions $P(X_{ij} | \text{true})$ and $P(X_{ij} | \text{false})$ are then used to calculate the posterior probability of true interaction $P(\text{true} | X_{ij})$ (Fig. 1c,d and Online Methods). For baits profiled in replicates, the next step involves the computation of a combined probability score from independent scoring of each replicate (Online Methods). Finally, SAINT probabilities can be used to estimate the false discovery rate (FDR). By ordering interactions in decreasing order of probability, a threshold can be selected that considers the average of the complement probabilities as the Bayesian FDR⁹. Although the accuracy of FDR estimates remains

to be validated, the availability of an objective reliability measure that has been widely used is an advantage over other methods.

We first tested performance of the generalized SAINT model using a human dataset⁶ centered around four key protein complexes that are involved in chromatin remodeling: prefoldin, hINO80, SRCAP and TRRAP or TIP60 (referred to as the TIP49 dataset). Although the original work focused the analysis on the interaction network between a core set of 65 proteins, here we analyzed the entire dataset provided by the authors of that study. The dataset consists of 27 baits (35 purifications) and 1,207 preys, and yielded 5,521 unfiltered interactions. The dataset also included 35 negative controls, which allows semisupervised modeling (Fig. 1a and Supplementary Table 1).

We applied SAINT to these data and compared the results to PP-NSAF⁶ and CompPASS Z and D^N scores^{7,10}, which we reimplemented in-house (Online Methods). We note that PP-NSAF⁶ removes all interactions involving prey proteins for which the sum of squared NSAF values across the negative control purifications is higher than that in the experiments that contain bait proteins. CompPASS is the only method that does not incorporate negative controls in scoring.

SAINT selected 1,375 interactions at the probability threshold 0.9, which was approximately equivalent to an estimated FDR of 2%. In PP-NSAF, as arbitrary cutoffs were set to define high, moderate and low probability interaction sets, the same number of top-scoring interactions was selected (corresponding to a PP-NSAF probability 0.2 or higher). In CompPASS, the same number of interactions corresponded to a D^N -score threshold of 1.48 (Supplementary Table 1).

We evaluated the performance of each algorithm first by benchmarking the selected interactions against two interaction databases named BioGRID¹¹ and iRefWeb¹² (Fig. 2a), and second by assessing the co-annotation rate of interaction partners to common Gene Ontology (GO) terms in ‘biological processes’ (Fig. 2b and Supplementary Table 1). SAINT-filtered interactions (with

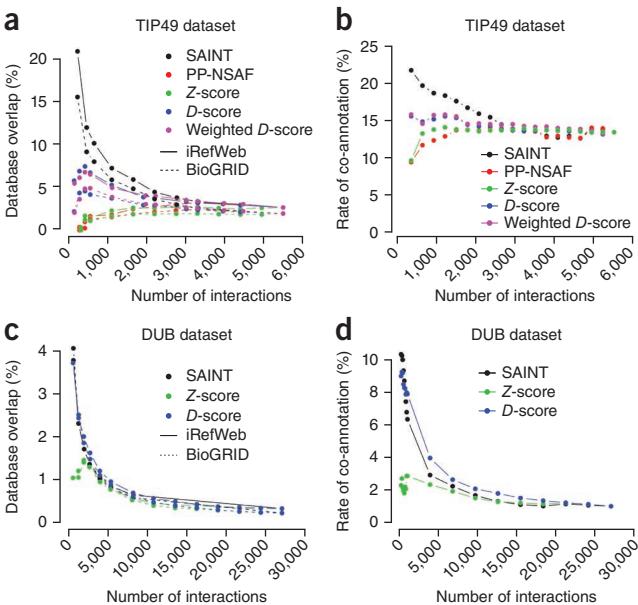


Figure 2 | Analysis of TIP49 and DUB datasets. **(a)** Benchmarking of filtered interactions in the TIP49 dataset by the overlap with interactions previously reported in BioGRID and iRefWeb databases. **(b)** Co-annotation of interaction partners to common GO terms in ‘biological processes’ in the TIP49 dataset. **(c)** Benchmarking against BioGRID and iRefWeb in the DUB dataset. **(d)** Co-annotation to GO terms in the DUB dataset.

controls) consistently showed the highest overlap with previously reported interactions and co-annotation rates to terms relevant to chromatin remodeling, including histone acetylation, protein amino acid acetylation, chromatin organization and modification, and cellular macromolecular complex assembly. Variation of the SAINT probability thresholds ($\sim 0.8\text{--}0.95$) did not qualitatively change this conclusion (data not shown). Note that omission of negative controls from SAINT modeling decreased the overlap with the interactions reported in BioGRID and iRefWeb (Supplementary Fig. 1). Explicit incorporation of the negative control data improved the robustness of modeling, especially in small or medium datasets.

We then tested the performance of SAINT for large-scale datasets without negative controls (Fig. 1b) on the human deubiquitinating enzymes (DUB) dataset⁷ (this dataset was used in the development of CompPASS). High confidence interactions from SAINT were compared to the high confidence set from CompPASS (Supplementary Table 2). Owing to the absence of negative controls, it was not possible to apply PP-NSAF to this dataset. SAINT probabilities and D^N scores were notably correlated (Pearson correlation, $r = 0.79$). At a probability threshold of 0.8, SAINT selected 1,300 interactions, whereas a threshold of CompPASS $D^N \geq 1$ (as used in ref. 7) reported 1,377 interactions. Of these, 1,051 interactions were identified by both methods. Reflecting the similarity of selected interactions, SAINT and CompPASS recovered previously reported interactions at comparable rates (Fig. 2c). In the top 1,000 interactions, SAINT showed higher overlap with published data. The co-annotation of interaction partners to the common GO terms also showed similar results between the two methods (Fig. 2d), including relevant terms such as positive and negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle, proteasome, and

so on (Supplementary Table 2). Although SAINT and CompPASS recovered largely overlapping interactions, SAINT removed the interactions identified with 1–2 spectral counts, which were still scored by CompPASS if they were specific to a single bait protein and detected in duplicates.

Another advantage of SAINT over other methods is that it is applicable to the analysis of small-scale datasets for which control purifications are available; this extends to the case of a single bait. We illustrate this by using a recent dataset¹³ that contains three experimental purifications of the bait CDC23 and three control purifications. In the original analysis, the authors of the study identified true interactions using ion intensity-based quantification followed by a simple *t*-test. We applied the SAINT approach to the same dataset by using spectral counts (the data were re-searched in-house; Online Methods). The results obtained by SAINT were nearly identical to those in the initial report (Supplementary Table 3), the sole exception being the single peptide hit C11orf51, which was reported as a new interactor in the original analysis¹³ but which was removed by SAINT.

The SAINT model presented here is based on label-free quantification using spectral counts, a parameter that is easily extracted from most AP-MS datasets. SAINT can also be extended to model other types of quantitative parameters such as peptide ion intensity¹⁴ or other continuous variables¹⁵, which can be accommodated by simply substituting the likelihood with an appropriate continuous distribution. SAINT is available as Supplementary Software.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturemethods/>.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

Supported by grants from the Canadian Institute of Health Research to A.-C.G. (MOP-84314) and M.T. (MOP-12246); the US National Institutes of Health to M.T. (5R01RR024031), A.I.N. and A.-C.G. (R01-GM094231), and A.I.N. (R01-CA126239); a Royal Society Wolfson Research Merit Award and a Scottish Universities Life Sciences Alliance Research Chair to M.T.; a Canada Research Chair in Functional Proteomics to A.-C.G.; and the Lea Reichmann Chair in Cancer Proteomics to A.-C.G. We thank M. Sardiu, M. Washburn and M. Sowa for providing additional information regarding the datasets used in this work, G. Bader for discussions and W. Dunham for reading the manuscript.

AUTHOR CONTRIBUTIONS

H.C. and A.I.N. developed, implemented and tested the SAINT method; H.C. wrote the software; B.L., A.B., Z.-Y.L., A.-C.G. and M.T. generated data for the initial SAINT modeling and provided feedback on the model performance; D.M. and D.F. assisted with data analysis and processing; Z.S.Q. contributed to statistical model development; H.C., A.-C.G. and A.I.N. wrote the manuscript; A.I.N. and A.-C.G. conceived the study; A.I.N. directed the project with input from A.-C.G.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/naturemethods/>. Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

1. Ewing, R.M. et al. *Mol. Syst. Biol.* **3**, 89 (2007).
2. Gavin, A.C. et al. *Nature* **440**, 631–636 (2006).
3. Jeronimo, C. et al. *Mol. Cell* **27**, 262–274 (2007).
4. Krogan, N.J. et al. *Nature* **440**, 637–643 (2006).
5. Nesvizhskii, A.I., Vitek, O. & Aebersold, R. *Nat. Methods* **4**, 787–797 (2007).
6. Sardiu, M.E. et al. *Proc. Natl. Acad. Sci. USA* **105**, 1454–1459 (2008).

7. Sowa, M.E., Bennett, E.J., Gygi, S.P. & Harper, J.W. *Cell* **138**, 389–403 (2009).
8. Breitkreutz, A. et al. *Science* **328**, 1043–1046 (2010).
9. Müller, P., Parmigiani, G & Rice, K. in *Bayesian Statistics Vol. 8* (eds., Bernardo, J.M. et al.) 349–370 (Oxford University Press, 2007).
10. Behrends, C., Sowa, M.E., Gygi, S.P. & Harper, J.W. *Nature* **466**, 68–76 (2010).
11. Breitkreutz, B.J. et al. *Nucleic Acids Res.* **36**, D637–D640 (2008).
12. Turner, B. et al. *Database (Oxford)* **2010**, baq023 (2010).
13. Hubner, N.C. et al. *J. Cell Biol.* **189**, 739–754 (2010).
14. Rinner, O. et al. *Nat. Biotechnol.* **25**, 345–352 (2007).
15. Griffin, N.M. et al. *Nat. Biotechnol.* **28**, 83–89 (2010).

ONLINE METHODS

Label-free quantification by spectral counting. Label-free quantification in this work was based on spectral counting. Spectral counts are the sum of every successful instance of sequencing a peptide from a particular protein by mass spectrometry, including redundant spectra. With proper normalization, spectral counts can be used as a quantitative measure of protein abundance in the sample. This method is conceptually similar to the approach of measuring gene expression using SAGE, EST or RNA-Seq fragment count data. For both DUB and TIP49 datasets, spectral count data were taken exactly as provided by the authors^{6,7}. Briefly, the DUB dataset and the TIP49 dataset were searched using SEQUEST¹⁶ using target-decoy database strategies against human databases; selected parameter sets were defined by the authors for filtering. The DUB dataset accepted peptides based on the following criteria. (i) High-stringency set: XCorr 2+ ≥ 2.5; 3+ ≥ 3.2; 4+ ≥ 3.5; +1 charge states were not collected. (ii) Complementary peptide set for proteins identified with high confidence: XCorr thresholds ≥ 1.0; ΔCn ≥ 0.05. The parameters selected by the authors of the TIP49 dataset were: XCorr 1+ ≥ 1.8 for 2+ ≥ 2.5, and 3+ ≥ 3.5 (fully tryptic peptides of at least seven amino acids long with max Sp score of 10). The reported spectral FDR for the entire TIP49 dataset was 0.065%; for the DUB data, a set FDR of 2% was selected to populate the interaction tables. No control data were used for the DUB dataset. In the case of the TIP49 dataset, 35 controls were provided alongside the experimental samples. These controls were generated from HeLa and HEK293 cell lines under nine different conditions. We merged the 35 measurements to 9 by taking the largest spectral count for each prey in each condition (**Supplementary Table 1**). For the analysis of the CDC23 data, the data were downloaded from Tranche (trancheproject.org), and re-searched in-house using X!Tandem/k-score against the RefSeq database using search parameters similar to those used in ref. 13. The search results were processed using PeptideProphet and ProteinProphet⁵, and filtered to achieve a protein-level FDR of less than 0.5%. The spectral counts were extracted using the in-house software Abacus (D.F. and A.I.N., unpublished data).

SAINT model. This section describes the generalized statistical modeling framework for the datasets with and without control purifications (**Fig. 1a,b**). In both cases, the spectral counts for prey *i* in purification with bait *j* are considered to be either from a Poisson distribution representing true interaction (with mean count λ_{ij}) or from a Poisson distribution representing false interaction (with mean count κ_{ij}). In the form of probability distribution, we write

$$P(X_{ij} | \bullet) = \pi_T P(X_{ij} | \lambda_{ij}) + (1 - \pi_T) P(X_{ij} | \kappa_{ij}) \quad (1)$$

where π_T is the proportion of true interactions in the data, and dot notation represents all relevant model parameters estimated from the data (here, specifically for the pair of prey *i* and bait *j*). The individual bait-prey interaction parameters λ_{ij} and κ_{ij} are estimated from joint modeling of the entire bait-prey association matrix, with the probability distribution (likelihood) of the form $P(X | \bullet) = \prod_{i,j} P(X_{ij} | \bullet)$. The proportion π_T is also estimated from the model, which relies on latent variables in the sampling algorithm (see below).

When at least three control purifications are available, and assuming that the control purifications provide a robust

representation of nonspecific interactors, the parameter κ_{ij} can be estimated from spectral counts for prey *i* observed in the negative controls. This is equivalent to assuming

$$\begin{aligned} P(X_{ij} | \bullet) &= \prod_{i,j: j \in E} (\pi_T P(X_{ij} | \lambda_{ij}) + (1 - \pi_T) P(X_{ij} | \kappa_{ij})) \times \\ &\quad \prod_{i,j: j \in C} (P(X_{ij} | \kappa_{ij})) \end{aligned} \quad (2)$$

where *E* and *C* denote the group of experimental purifications and the group of negative controls, respectively. This leads to a semisupervised mixture model in the sense that there is a fixed assignment to false interaction distribution for negative controls. As negative controls guarantee sufficient information for inferring model parameters for false interaction distributions, Bayesian nonparametric inference using Dirichlet process mixture priors can be used to derive the posterior distribution of protein-specific abundance parameters in the model. As a result, the mean parameters in the Poisson likelihood functions follow a nonparametric posterior distribution, allowing more flexible modeling at the proteome level. Under this setting, all model parameters are estimated from an efficient Markov chain Monte Carlo algorithm¹⁷.

To elaborate on the two distributions, the mean parameter for each distribution is assumed to have the following form. For false interactions, it is assumed that spectral counts follow a Poisson distribution with mean count

$$\log(\kappa_{ij}) = \log(l_i) + \log(c_j) + \gamma_0 + \mu_i \quad (3)$$

where l_i is the sequence length of prey *i*, and c_j is the bait coverage, the spectral count of the bait in its own purification experiment, γ_0 is the average abundance of all contaminants and μ_i is prey *i* specific mean difference from γ_0 . For true interactions, it is assumed that spectral counts follow a Poisson distribution with mean count

$$\log(\lambda_{ij}) = \log(l_i) + \log(c_j) + \beta_0 + \alpha_{bj} + \alpha_{pi} \quad (4)$$

where β_0 is the average abundance of prey proteins in those cases where they are true interactors of the bait, α_{bj} is bait *j* specific abundance factor and α_{pi} is prey *i* specific abundance factor. In other words, the mean spectral count for a prey protein in a true interaction is calculated using a multiplicative model combining bait- and prey-specific abundance parameters. This formulation substantially reduces the number of parameters in the model, avoiding the need to estimate every λ_{ij} separately.

For datasets without negative control purifications, the mixture component distributions for true and false interactions have to be identified solely from experimental (noncontrol) purifications. In this case, a user-specified threshold is applied to divide preys into high-frequency and low-frequency groups, denoted as $Y_i = 1$ or 0 if prey *i* belongs to the high- or low-frequency group, respectively. An arbitrary 20% threshold was applied in the case of the DUB dataset; however, the results were not very sensitive to the choice of the threshold. For preys in the high frequency group, the model considers spectral counts for the observed prey proteins (ignoring zero count data, which represent the absence of protein identification), as there are sufficient data to estimate distribution parameters. In the low-frequency group, nondetection of a prey is included to help the separation of high-count from low-count hits. The entire mixture model can then be expressed as

$$P(X_{ij} | \bullet) = \prod_{i,j} (\pi_T P(X_{ij} | \lambda_{ij}) + (1 - \pi_T) P(X_{ij} | \kappa_{ij}))^{Z_{ij}} \quad (5)$$

where $Z_{ij} = 1(Y_i=0) + 1(Y_i=1, X_{ij}>0)$ and the false and true interaction distributions are modeled by equations (3) and (4), respectively.

The posterior probability of a true interaction given the data is computed using Bayes rule

$$P(\text{true}|X_{ij}) = T_{ij} / (T_{ij} + F_{ij}) \quad (6)$$

where $T_{ij} = \pi_T P(X_{ij} | \lambda_{ij})$ and $F_{ij} = (1 - \pi_T) P(X_{ij} | \kappa_{ij})$. If there are replicate purifications for bait j , the final probability is computed as an average of individual probabilities over replicates. Note that one alternative approach is to compute the probability assuming conditional independence over replicates, that is, $\prod_{k \in j} P(X_{ijk} | \lambda_{ijk})$ and $\prod_{k \in j} P(X_{ijk} | \kappa_{ijk})$ for true and false interactions, with additional index k denoting replicates for bait j . Unlike average probability, this probability puts less emphasis on the degree of reproducibility, and thus may be more appropriate in datasets where replicate analysis of the same bait is performed using different experimental conditions (for example, purifications using different affinity tags) to increase the coverage of the interactome.

When probabilities have been calculated for all interaction partners, the Bayesian false discovery rate (FDR) can be estimated from the posterior probabilities as follows. For each probability threshold p^* , the Bayesian FDR is approximated by

$$\text{FDR}(p^*) = (\sum_k 1(p_k \geq p^*)(1 - p_k)) / (\sum_k 1(p_k \geq p^*)) \quad (7)$$

where p_k is the posterior probability of true interaction of protein pair k . The output from SAINT allows the user to select a probability threshold to filter the data to achieve the desired FDR.

Implementation of other scores. CompPASS^{7,10} calculates two different scores. First, Z score is constructed by mean centering and scale normalization in the conventional Z statistic, where

mean and s.d. are estimated from the data for each prey. D score is based on the spectral count adjusted by a scaling factor that reflects the reproducibility of prey detection over replicate purifications of the same bait. If X_{ij} denotes the spectral count between prey i and bait j , then $D_{ij} = ((k / f_i)^{p_{ij}} \cdot X_{ij})^{1/2}$, where k is the total number of baits profiled in the experiment, f_i is the number of experiments in which prey i was detected and p_{ij} is the number of replicate experiments of bait j in which prey i was detected. After computing the scores, a threshold D^T is selected from simulation data so that 95% of the simulated data falls below the chosen threshold. Note that CompPASS merges replicate data for bait j to produce a unique spectral count X_{ij} for a given pair. In doing so, it takes nonzero counts only when the prey is identified in a single replicate or otherwise averages counts over multiple replicates. In the analysis of TIP49 dataset, we used both the original D score and the more recently implemented ‘weighted D score’, which is designed for datasets with large protein complexes¹⁰. The weighted D scores are shown in **Figure 2** for the TIP49 dataset.

To replicate PP-NSAF⁶, we removed 330 contaminants from the dataset using the vector magnitude approach. After filtering, probabilities were computed using an in-house script following the method presented in ref. 6. Although our implementation did not reproduce exactly the same scores for the interactions reported in ref. 6, the scores computed by the in-house implementation showed a clear linear correspondence to the reported scores (Pearson correlation 0.89).

Software. The source C code and a user manual for the generalized SAINT model described in this work (SAINT 2.0) can be downloaded from <http://saint-apms.sourceforge.net/>, where updates will be distributed. The published version is also available as **Supplementary Software**.

16. Eng, J.K., McCormack, A.L. & Yates, J.R.I. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
17. Ishwaran, H. & James, L.F. *J. Am. Stat. Assoc.* **96**, 161–173 (2001).

Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor

Nicolas Bisson¹, D Andrew James^{1,4}, Gordana Ivosev², Stephen A Tate², Ron Bonner², Lorne Taylor¹ & Tony Pawson^{1,3}

Signaling pathways are commonly organized through inducible protein-protein interactions, mediated by adaptor proteins that link activated receptors to cytoplasmic effectors¹. However, we have little quantitative data regarding the kinetics with which such networks assemble and dissolve to generate specific cellular responses. To address this deficiency, we designed a mass spectrometry method, affinity purification-selected reaction monitoring (AP-SRM), which we used to comprehensively and quantitatively investigate changes in protein interactions with GRB2, an adaptor protein that participates in a remarkably diverse set of protein complexes involved in multiple aspects of cellular function. Our data reliably define context-specific and time-dependent networks that form around GRB2 after stimulation, and reveal core and growth factor-selective complexes comprising 90 proteins identified as interacting with GRB2 in HEK293T cells. Capturing a key hub protein and dissecting its interactions by SRM should be equally applicable to quantifying signaling dynamics for a range of hubs in protein interaction networks.

Signals from cell surface receptors are often relayed through adaptor proteins that couple activated receptors or scaffolds to multiple downstream targets². In the case of tyrosine kinase receptors (RTK), this is exemplified by adaptors composed exclusively of SH2 (Src Homology 2) and SH3 protein interaction domains^{3–5}, such as GRB2 (ref. 6). GRB2 has a single SH2 domain that binds preferentially to pY- ϕ -N- ϕ motifs⁷, as found on a number of activated RTKs, flanked by two SH3 domains that bind proline and arginine-rich motifs in downstream effectors⁸. GRB2 thereby links activated RTKs to targets such as the SOS guanine nucleotide exchange factor (GEF), and thus to stimulation of the Ras/MAPK (mitogen-activated protein kinase) pathway⁹. As GRB2 participates in a variety of RTK signaling pathways^{6,10,11}, we have investigated whether it might capture a substantial fraction of proteins involved in phosphotyrosine (pTyr) signaling, and so allow a broad investigation of dynamic protein-protein interactions induced by tyrosine phosphorylation.

Advances in technologies such as mass spectrometry (MS) and robotics have enhanced our understanding of signal transduction

at the systems level, in part by defining how signaling proteins interact with one another^{12–14}. Although these approaches have generated a great deal of information, they require precise quantification to address the dynamic nature of protein interaction networks, a property that underlies their biological regulation and output. Protein networks may vary from cell to cell, and also fluctuate in response to different cues and over time following stimulation. Understanding the dynamics of signaling networks therefore remains an important challenge.

The combination of AP and MS has been used successfully to decipher protein interaction networks in cells¹⁵, but very few studies have quantified the kinetics of such assemblies¹⁶. Current approaches to this problem involve metabolic labeling using stable isotope labeling of amino acids in cell culture (SILAC) and/or MS1 quantification or spectral counting¹⁷. However, the variability of measurements in successive experiments that results from stochastic sampling can hinder our ability to define the dynamics of protein interactions by AP-MS even when integrated approaches are used to circumvent this issue^{18–20}.

Recently, targeted proteomics techniques such as selected reaction monitoring (SRM) have allowed a substantial improvement in the reproducibility with which MS data can be acquired²¹. SRM relies on a function specific to triple quadrupole mass spectrometers, which permits selection of a pair of masses, corresponding to a precursor ion and one of its unique fragments (**Supplementary Fig. 1**). These precursor/fragment ion pairs, called transitions, are monitored over time to give a sensitive, reproducible and quantitative measurement of the corresponding peptide and protein across different samples²². SRM analysis has been extensively employed in the fields of pharmacology and toxicology²³, but its use in proteomics is fairly recent^{20,24,25}.

We hypothesized that GRB2 is an important hub in growth factor signaling. To first determine the complement of its associated proteins, we generated HEK293T cell lines that stably express a 3xFLAG-tagged GRB2 fusion protein. We used the phosphatase inhibitor pervanadate, which broadly elevates cellular tyrosine phosphorylation, to obtain a full spectrum of GRB2 interactors with which to develop an SRM assay. These proteins were isolated from cell lysates with anti-FLAG AP and analyzed by liquid chromatography-tandem MS (LC-MS/MS). We discovered 108 GRB2-associated proteins (**Supplementary Table 1**), many of which are known to be

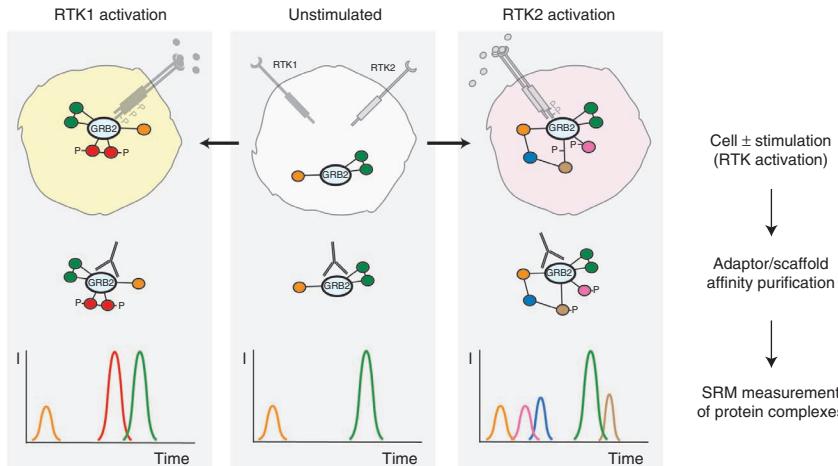
¹Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Joseph and Wolf Lebovic Health Complex, Toronto, Ontario, Canada. ²AB Sciex, Concord, Ontario, Canada. ³Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. ⁴Current address: Sanofi-Pasteur, Toronto, Ontario, Canada. Correspondence should be addressed to T.P. (pawson@lunenfeld.ca).

Received 27 April; accepted 3 June; published online 26 June 2011; doi:10.1038/nbt.1905

Figure 1 Schematic representation of the AP-SRM workflow. The strategy combines AP of a single adaptor/scaffold protein with SRM-MS to quantitatively and reproducibly measure changes in protein interaction networks using a single assay. In summary, cells are subjected to stimulation, leading to activation of phosphorylation-dependent signaling pathways downstream of a specific receptor (RTK1 or RTK2). An adaptor/scaffold protein (in this case GRB2), which may bind the activated RTK, is affinity-purified along with its interacting proteins. The isolated proteins are analyzed by SRM and compared to complexes enriched from unstimulated control cells. RTK, receptor tyrosine kinase; P, phosphorylation; I, intensity counts.

involved in signaling pathways downstream of growth factors, such as insulin or the epidermal growth factor (EGF). For example, we found 9 lipid or protein kinases, 10 phosphatases, 18 GTPase guanine nucleotide exchange factors and GTPase-activating proteins (GAPs) as well as 22 adaptor proteins or scaffolds. Searches of the Human Protein Reference Database (HPRD) and BioGRID database suggest that 73 of these proteins have not been previously reported to associate with GRB2 (**Supplementary Table 1**). We also found 62 phosphorylation sites on 26 of these proteins, including known or potential GRB2 SH2-binding sites (**Supplementary Table 2**).

We combined SRM with standard AP to devise a method (AP-SRM) to measure the proteins that constitute the extended GRB2 interaction



network in a single assay, and then to address the dynamics with which they assemble in response to pTyr signaling (**Fig. 1**). We reproducibly measured 90 of the 108 GRB2-associated proteins by SRM. This analysis used 326 peptides, and involved 1,157 transitions (**Supplementary Table 3**). Therefore, a simple GRB2 affinity purification can isolate a remarkable fraction of the pTyr signaling network for quantitative analysis using SRM.

To test the reliability of the AP-SRM assay, we quantified the extent to which association of each of the GRB2-interacting proteins depends on the integrity of the SH2 and SH3 domains in GRB2. To this end, we generated a cell line expressing a GRB2 mutant with substitutions that inactivate the conventional ligand-binding surface of the two SH3 domains, but that retains a wild-type SH2 domain (SH2⁺). Conversely, we expressed a mutant with functional SH3 domains, but

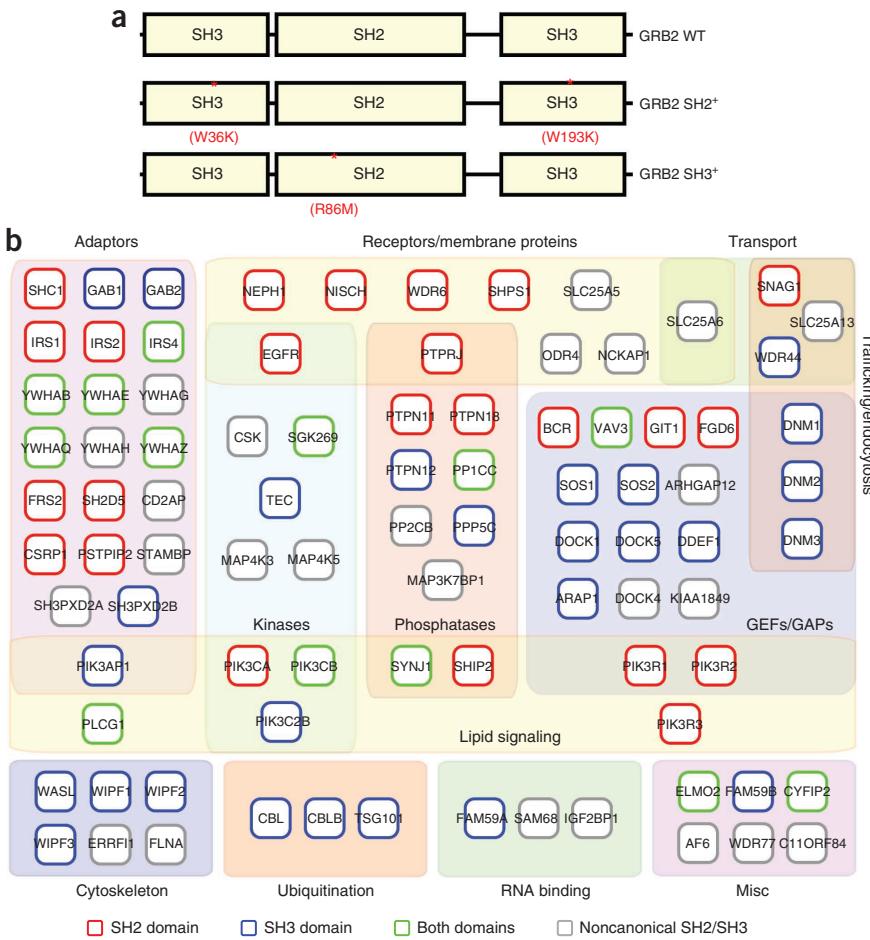


Figure 2 SH2/SH3 domain association of the 90 proteins measured in the GRB2 SRM assay. **(a)** Schematic representation of GRB2, depicting the structure of wild-type GRB2, as well as the SH2⁺ (W36K, W193K) and SH3⁺ (R86M) mutant forms. **(b)** The association of each protein analyzed ($n = 89$) with the SH2 or SH3 domains of GRB2 was measured and compared. Proteins associated primarily with the SH2 domain are represented in red (e.g., SHC1) and those with the SH3 domains in blue (e.g., CBL). Proteins that may associate with both the SH2 and SH3 domains are in green (e.g., IRS4). Twenty-five proteins remained associated with GRB2 after AP of the triple mutant (in gray). These proteins may bind either GRB2 through the inter-domain linker regions or SH2/SH3 domains in an atypical manner that is not abrogated through conventional mutations. For the purpose of this study, we focused solely on interactions abrogated with the triple mutation, although all data are provided in **Supplementary Data Sets 1** and **2**. We defined a protein as associating with the SH2 or SH3 domains if its binding was increased with high confidence by fivefold or more in the SH2⁺ or SH3⁺ mutants, respectively, relative to the triple mutant. Proteins are grouped according to their cellular function(s).

with a substitution that abolishes pTyr binding by the SH2 domain (SH3^+) (Supplementary Fig. 2). We first focused on a subset of 15 proteins analyzed by AP-SRM, which have been previously reported (Supplementary Table 4) to recognize specific GRB2 SH2 or SH3 domains. We measured the relative association of each protein with the SH2⁺ and SH3⁺ mutants in pervanadate-treated cells, and compared the patterns of association with those observed in the presence of a triply mutated form of GRB2 in which all domains are disrupted. We further developed a statistical analysis workflow to calculate a probability of occurrence (weight) of a given protein behavior (e.g., an increase in binding) when comparing two cell states (Supplementary Methods and Supplementary Figs. 3 and 4).

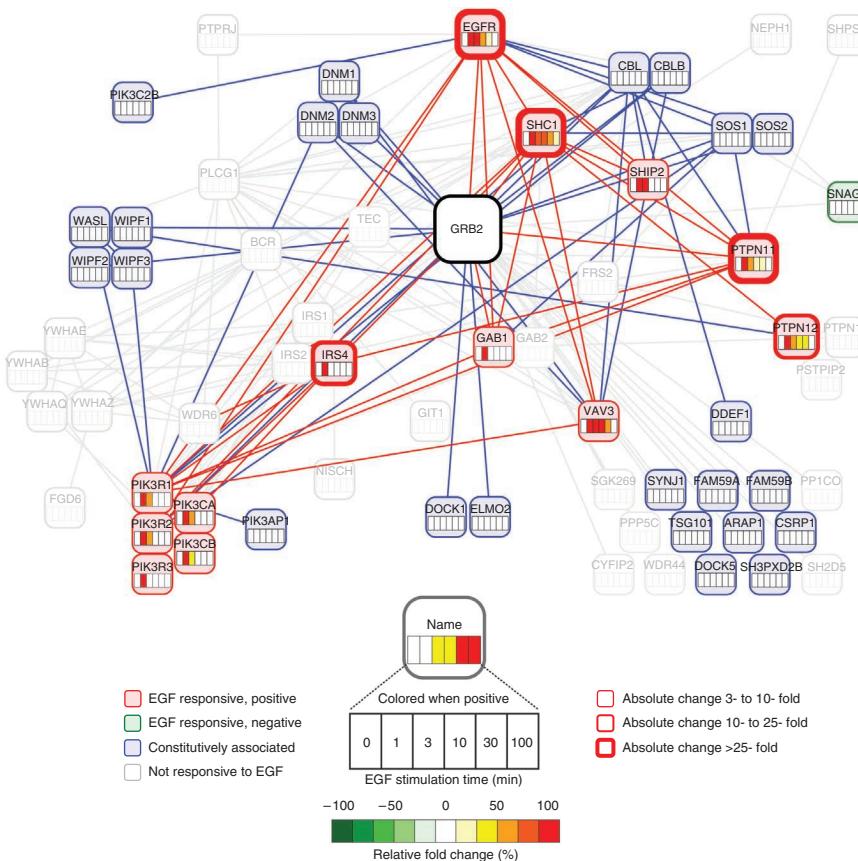
The AP-SRM strategy correctly assigned the GRB2 domain-binding properties of the 15 proteins in our data set that had been previously reported to associate with GRB2 using conventional techniques (Supplementary Table 4), arguing for the reliability and accuracy of the AP-SRM assay. Among the larger set of GRB2-associated proteins, we found 39 that specifically interact through the SH2 domain and 40 through one or both SH3 domains (Fig. 2). Surprisingly, 13 proteins (e.g., scaffolds such as IRS4) were present in both groups, suggesting that they can be recruited to GRB2 through different binding modes involving either the SH2 or SH3 domains. Experiments involving short interfering (si)RNA knockdown showed that a subset of these proteins may associate indirectly with GRB2, and through multiple nodes, thus enhancing network robustness (Supplementary Fig. 5). These experiments also confirmed the association with GRB2 SH2 or SH3 domains of 37 of the 73 putative new binders identified (Fig. 2 and Supplementary Table 1). Overall, these observations made using a single assay define the specific means by which all 90 proteins measured by AP-SRM are recruited to GRB2.

To investigate how the GRB2 protein interaction network fluctuates in response

Figure 3 Dynamic quantitative analysis of the GRB2 protein complexes upon stimulation of cells with epidermal growth factor (EGF). The GRB2-centered protein interaction network displays the association of time-dependent subnetworks following EGF stimulation. Nodes represent 65 proteins (including GRB2) found to be associated with an SH2 or SH3 domain (Fig. 2b). Red-shaded nodes represent proteins that are recruited to GRB2 complexes after EGF stimulation irrespective of time, green-shaded nodes those that are decreased, and blue-shaded nodes those present in GRB2 complexes in nonstimulated (control) cells. The thickness of the node line is proportional to the intensity of the change compared to control levels. Rectangles inside the nodes show the relative fold change for each time point (refer to graphic key). Interactions between proteins in the network are represented with edges if already reported in HPRD or BioGRID databases. Edges between red nodes are red and edges between proteins present in GRB2 complexes in nonstimulated cells are blue. The discriminating weight for data reported in these graphs ≥ 0.7 . Proteins were grouped manually after their homology, domain composition, cellular function and number of previously reported interactions. The 25 proteins that are associated with GRB2 independently of its SH2/SH3 domain were omitted but were measured; data may be found in Supplementary Data Sets 1 and 2.

to stimulation with a physiological growth factor, we stimulated HEK293T cells with EGF for varying times (0, 1, 3, 10, 30 and 100 min), and used AP-SRM to analyze the composition of GRB2 complexes at each time point. By this quantitative measure, 13 proteins showed an increased association with GRB2, albeit with varying levels and kinetics, for at least one of the time points. The composition of EGF-induced GRB2 complexes was maximal after 1 min of stimulation, when all of the inducible proteins were bound to GRB2 (Fig. 3 and Supplementary Figs. 6a and 7a). This correlates with maximal activation of extracellular signal-regulated kinase (ERK) activation (Supplementary Fig. 8c). In addition, we identified 23 constitutive GRB2 interactors involved in processes such as cytoskeletal organization and endocytosis. All of these proteins, detected independently of EGF stimulation, associate with GRB2 through its SH3 domains (Fig. 2b). These data (i) provide a quantitative view of the time-dependent protein interaction network that forms around GRB2 following stimulation with EGF, (ii) indicate that GRB2 rapidly engages a range of pathways affecting distinct aspects of signaling, metabolism and cytoskeletal architecture and (iii) provide the basis for a more detailed analysis, as discussed below.

The quantitative temporal analysis identified three distinct groups of proteins that vary in the kinetics with which they bind GRB2: an EGFR-like group (five members), a SHIP2-like group (five members) and an IRS4-like group (three members) (Supplementary Fig. 8a). Proteins in the EGFR-like group bind GRB2 rapidly, and then dissociate from GRB2 quickly, although remaining in GRB2 complexes for 10–30 min. Interestingly, three proteins from this group, EGFR⁶, SHC1 (ref. 11) and PTPN11 (ref. 26), associate directly with the SH2 domain of GRB2 and are involved in the activation of the Ras-MAPK pathway. We therefore examined the dynamics of their interaction



with GRB2 more closely. After 1 min of EGF stimulation, the increases in the levels of the EGFR, SHC1 and PTPN11 associated with GRB2 were 169-, 45- and 29-fold respectively (**Supplementary Fig. 6b**). The interaction of the EGFR with GRB2 was sustained at 3 min. After 30 min of EGF stimulation, only SHC1 still showed substantial

association with GRB2. The sustained SHC1-GRB2 complex may therefore permit prolonged signaling to the Ras-MAPK pathway. The results underscore the precision that can be achieved with SRM measurements in deciphering the dynamic temporal behavior of protein interaction networks, and reveal that different GRB2-binding

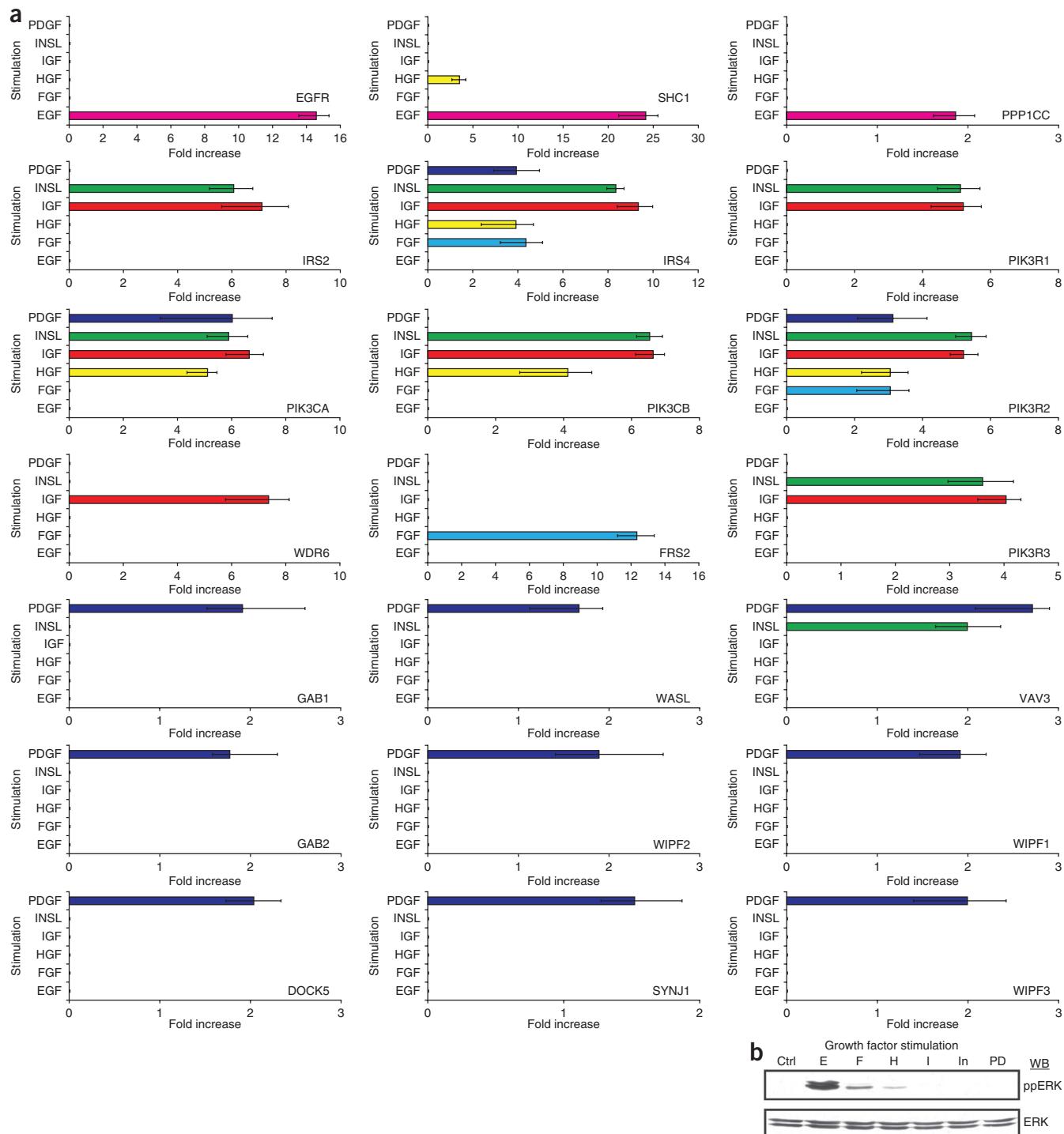


Figure 4 The composition of GRB2 complexes is generally dependent on the stimulation type. **(a)** Only proteins found at elevated levels were increased in GRB2 complexes upon stimulation with one or more of the growth factors tested are shown. Bar graphs display the increase of the levels of proteins associated with GRB2 following a stimulation compared to control. The discriminating weight for data reported in these graphs ≥ 0.7 . **(b)** Western blots displaying the levels of ppERK and ERK in cell lysates for each of the conditions tested are shown in the bottom right corner. E, EGF; F, FGF; H, HGF; I, IGF; In, Insulin; PD, PDGF.

partners have distinct kinetics of complex formation downstream of EGF stimulation. This in turn can potentially fine-tune the nature of the output from the activated receptor over time.

Having shown the temporal variation in GRB2 complexes induced by a single growth factor, we next sought to characterize GRB2 complexes formed in response to activation of distinct endogenous receptors. To this end, we stimulated HEK293T cells for 20 min with one of six different growth factors (GF) that signal through RTKs, namely EGF, fibroblast GF (FGF), hepatocyte GF (HGF), insulin-like GF-1 (IGF), insulin and platelet-derived GF-B-chain dimer (BB) (PDGF). We selected these growth factors based on the strong expression of their cognate receptors in HEK293T cells²⁷, and on a direct interaction of the receptor with the GRB2 SH2 domain in the case of receptors for EGF⁶, HGF¹⁰ and PDGF²⁸. The intensity of 21 proteins increased in GRB2 complexes for at least one of the six growth factors tested (Fig. 4 and Supplementary Fig. 7b). As expected, most of these proteins (16) bound the GRB2 SH2 domain (Fig. 2b). In addition, 23 constitutive interactors, all associated with GRB2 SH3 domains, were detected independently of any stimulation (Fig. 2b and Supplementary Fig. 9).

The above comparison led to the identification of growth factor-specific networks in stimulated cells. Some proteins were only present in GRB2 complexes in cells stimulated with one or two of the six growth factors tested, such as EGFR, FRS2 and GAB1. However, several proteins, for example IRS4, PIK3R2 and PIK3CA/B, showed increased association with GRB2 in the presence of the majority of the growth factors used for cell stimulation (Fig. 4). These proteins may therefore be part of a core GRB2 signaling machinery, whereas others are more context specific.

Some of the proteins demonstrated to exclusively bind the GRB2 SH3 domain(s) independently of any stimulation, including the cytoskeletal regulators WASL and WIPF1/2/3, also displayed increased interaction with GRB2 after treatment with PDGF. This could be due to novel indirect interactions formed between GRB2 and these proteins, allosteric changes that expose SH3-binding sites in the target proteins, or increased affinities for GRB2 SH3 domains when the SH2 is occupied, as previously demonstrated *in vitro*²⁹. Overall, these data suggest that growth factor-specific subcomplexes form around GRB2 in a given cell type, that could generate distinct targeted cell responses to different extracellular stimuli.

The AP-SRM approach was instrumental in our analysis of GRB2 signaling dynamics as it has several key advantages compared to other nontargeted MS-based quantitative methods. First, SRM is highly sensitive, with the lower limit of detection being in the subfemtomolar range^{20,22}. As a result, AP-SRM requires only minimal sample amounts (10–20% of what is normally used for other approaches). This may be a key consideration when working with expensive reagents (e.g., siRNA, drugs) or cells with limited availability (e.g., from primary tissues). Second, the quantitative measurements that are obtained are linear over a range of 4–5 orders of magnitude²⁴, which allows the comparison of samples containing proteins at different concentrations. Third, SRM permits the measurement of absolute protein abundances (not performed here). Fourth, AP-SRM does not require any *in vivo* labeling, which is expensive, limited to very few concomitant conditions (three for SILAC), cannot be applied to human tissues, and makes the quantification dependent on simultaneous peptide identification. Finally, as AP-SRM is a targeted method, not based on stochastic sampling, the reproducibility of measurements is extremely high²⁰. We calculated this reproducibility to be no less than 88.5% for three SRM measurements from each of three biological replicates (independent purifications) (Supplementary Figs. 10–13

and Supplementary Table 5). We have also found the variability to be greater between individual samples compared to successive analytical replicates of the same sample (Supplementary Figs. 10–14).

On the other hand, there are several limitations that need to be considered before adopting this type of method. First, AP-SRM does not allow for simultaneous *de novo* identification of proteins. Thus, the approach has to be hypothesis-driven and identification experiments may have to be performed before assay development. However, MS/MS information may be obtained using MIDAS (multiple RM/SRM-initiated detection and sequencing)³⁰ to ensure specific measurements (Supplementary Figs. 15 and 16). Second, SRM technology with time scheduling can currently accommodate 3,000 transitions (300–350 proteins) at the most. It is likely that this figure will increase as more sensitive technologies are developed. Third, time must be devoted to assay development, a step that is less demanding for most alternative protein quantification technologies. Finally, the triple quadrupole instruments required to perform SRM may not be part of the standard toolbox of proteomics facilities.

In summary, we have developed a new strategy in which affinity purification of a single adaptor or scaffold protein is combined with the sensitivity and reproducibility of SRM to quantitatively measure the fluctuations in protein interaction networks, in a single assay. We show that AP-SRM is a powerful tool to study network dynamics in cells, using GRB2 as a model. Our data not only show the remarkable connectivity and versatility of GRB2 in growth factor signaling, but also shed light on its involvement in the formation of stimulation-specific and time-dependent protein complexes. By focusing on key hub proteins, this method can be extended to obtain a truly quantitative view of dynamic intracellular protein interactions and network assembly activated by different classes of extracellular signals.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/nbt/index.html>.

Accession codes. The data associated with this manuscript may be downloaded from ProteomeCommons.org Tranche using the following hash:

MS/MS data. 7HWUuM9l4FjpxM392yIQ2ySUzcQkn4RbC+aq+uFB/CHZeRA05QLlqDtOfstQHezvfmN2U4hYHhcHc0rtMaNMWA+X2kAAAAAAAC1w==

PTMs data. RGFw3wr71szVRFFFVnw4PjjwteLjC0pYnqZP+yMRGkFD51FeUQVL1a56//7KC11+96fr1wfQFrLo7y3vsgog1NPwAAAAAAAC0A==

SRM data. pMTcB6xW0KKrrp0XvJAnO04jQC5x67aKnlb6SUCgBt2dh/M2ba0MEDxrtdJaMFxbEQL8AyKBj4fpPPd6qfJSB+VocdkAAAAAAAHMtA==

The protein interactions from this publication have been submitted to the IMEx (<http://imex.sf.net>) consortium through IntAct (PMID:19850723) and assigned the identifier IM-15417.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

The authors would like to thank K. Colwill and A.-C. Gingras for insightful comments on the manuscript, R. Bagshaw, G. Findlay, J. Gish, J. Jin, B. Larsen and O. Rocks for discussions and A. Yu Dai, B. Larsen and B. Liu for technical assistance. This project is supported through Ontario Research Fund grants from the Ontario Ministry of Research and Innovation and grants from Genome Canada through the Ontario Genomics Institute, the Canadian Institutes of Health Research (CIHR, grant MOP-6849), the Terry Fox Research Institute/Ontario Institute for Cancer Research and the Canadian Cancer Society

Research Institute. N.B. holds a research fellowship from the CIHR. T.P. holds the Apotex Chair in Molecular Oncology and is a Distinguished Investigator of the CIHR.

AUTHORS CONTRIBUTIONS

N.B., S.A.T., and T.P. conceived the project. N.B., D.A.J., S.A.T. and L.T. designed the experiments. N.B. and D.A.J. performed the experiments. N.B., G.I., S.A.T., R.B. and L.T. analyzed the data. N.B. and T.P. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/nbt/index.html>.

Published online at <http://www.nature.com/nbt/index.html>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Lemmon, M.A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
2. Pawson, T. Dynamic control of signaling by modular adaptor proteins. *Curr. Opin. Cell Biol.* **19**, 112–116 (2007).
3. Lim, W.A. & Pawson, T. Phosphotyrosine signaling: evolving a new cellular communication system. *Cell* **142**, 661–667 (2010).
4. Liu, B.A. *et al.* The human and mouse complement of SH2 domain proteins—establishing the boundaries of phosphotyrosine signaling. *Mol. Cell* **22**, 851–868 (2006).
5. Mayer, B.J. & Gupta, R. Functions of SH2 and SH3 domains. *Curr. Top. Microbiol. Immunol.* **228**, 1–22 (1998).
6. Lowenstein, E.J. *et al.* The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. *Cell* **70**, 431–442 (1992).
7. Songyang, Z. *et al.* Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. *Mol. Cell. Biol.* **14**, 2777–2785 (1994).
8. Sparks, A.B. *et al.* Distinct ligand preferences of Src homology 3 domains from Src, Yes, Ab1, Cortactin, p53bp2, PLCgamma, Crk, and Grb2. *Proc. Natl. Acad. Sci. USA* **93**, 1540–1544 (1996).
9. McCormick, F. Signal transduction. How receptors turn Ras on. *Nature* **363**, 15–16 (1993).
10. Ponzetto, C. *et al.* Specific uncoupling of GRB2 from the Met receptor. Differential effects on transformation and motility. *J. Biol. Chem.* **271**, 14119–14123 (1996).
11. Skolnik, E.Y. *et al.* The SH2/SH3 domain-containing protein GRB2 interacts with tyrosine-phosphorylated IRS1 and Shc: implications for insulin control of ras signaling. *EMBO J.* **12**, 1929–1936 (1993).
12. Barrios-Rodiles, M. *et al.* High-throughput mapping of a dynamic signaling network in mammalian cells. *Science* **307**, 1621–1625 (2005).
13. Breitkreutz, A. *et al.* A global protein kinase and phosphatase interaction network in yeast. *Science* **328**, 1043–1046 (2010).
14. Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008).
15. Gingras, A.C., Gstaiger, M., Raught, B. & Aebersold, R. Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* **8**, 645–654 (2007).
16. Kaake, R.M., Wang, X. & Huang, L. Profiling of protein interaction networks of protein complexes using affinity purification and quantitative mass spectrometry. *Mol. Cell. Proteomics* **9**, 1650–1665 (2010).
17. Choudhary, C. & Mann, M. Decoding signaling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol.* **11**, 427–439 (2010).
18. Hubner, N.C. *et al.* Quantitative proteomics combined with BAC TransgeneOmics reveals *in vivo* protein interactions. *J. Cell Biol.* **189**, 739–754 (2010).
19. Rinner, O. *et al.* An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. *Nat. Biotechnol.* **25**, 345–352 (2007).
20. Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D.A. & White, F.M. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci. USA* **104**, 5860–5865 (2007).
21. Addona, T.A. *et al.* Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* **27**, 633–641 (2009).
22. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4**, 222 (2008).
23. Ackermann, B.L., Berna, M.J., Eckstein, J.A., Ott, L.W. & Chaudhary, A.K. Current applications of liquid chromatography/mass spectrometry in pharmaceutical discovery after a decade of innovation. *Annu. Rev. Anal. Chem.* **1**, 357–396 (2008).
24. Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B. & Aebersold, R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* **138**, 795–806 (2009).
25. Traweger, A. *et al.* Protein phosphatase 1 regulates the phosphorylation state of the polarity scaffold Par-3. *Proc. Natl. Acad. Sci. USA* **105**, 10402–10407 (2008).
26. Shi, Z.Q., Yu, D.H., Park, M., Marshall, M. & Feng, G.S. Molecular mechanism for the Shp-2 tyrosine phosphatase function in promoting growth factor stimulation of Erk activity. *Mol. Cell. Biol.* **20**, 1526–1536 (2000).
27. Shaw, G., Morse, S., Ararat, M. & Graham, F.L. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB J.* **16**, 869–871 (2002).
28. Bazenet, C.E., Gelderloos, J.A. & Kazlauskas, A. Phosphorylation of tyrosine 720 in the platelet-derived growth factor alpha receptor is required for binding of Grb2 and SHP-2 but not for activation of Ras or cell proliferation. *Mol. Cell. Biol.* **16**, 6926–6936 (1996).
29. Chook, Y.M., Gish, G.D., Kay, C.M., Pai, E.F. & Pawson, T. The Grb2-mSos1 complex binds phosphopeptides with higher affinity than Grb2. *J. Biol. Chem.* **271**, 30472–30478 (1996).
30. Unwin, R.D., Griffiths, J.R. & Whetton, A.D. A sensitive mass spectrometric method for hypothesis-driven detection of peptide post-translational modifications: multiple reaction monitoring-initiated detection and sequencing (MIDAS). *Nat. Protoc.* **4**, 870–877 (2009).

ONLINE METHODS

Constructs. Human GRB2 (NCBI clone NM_002086) was subcloned into pMScVpuro (Clontech) with an N-terminal 3xFLAG epitope (Sigma). Point mutations were introduced using the Quickchange strategy (Stratagene). All inserts were fully sequenced and protein expression was verified.

Cell treatments. For pervanadate stimulations, a final concentration of 100 μM of Na₃VO₄ pre-activated with H₂O₂ was added to cell medium and was left for 20 min. For growth factor stimulations, a final concentration of 100 ng/ml of EGF, FGF, HGF, IGF or PDGF (Peprotech) or 200 μM of insulin (Novo Nordisk) was added to cell medium and was left for 20 min (unless otherwise stated).

Affinity purification (AP). Cleared whole cell extracts were incubated with 25 μl of packed M2 agarose beads (prewashed) (Sigma-Aldrich) for 2 h at 4 °C with slight rocking. Beads were washed three times with 1 ml of lysis buffer. Samples destined to MS or SRM were treated as described below. Alternatively, protein complexes on agarose beads were boiled in 2× SDS loading buffer (Laemmli) and separated by SDS-PAGE.

Antibodies. Antibodies used were as follows: anti-FLAG (M2) (Sigma-Aldrich), anti-GRB2 (BD Biosciences), anti-ppERK (Cell Signaling), anti-IRS4 (Santa Cruz Biotechnologies), anti EGFR (Cell Signalling Technologies), anti-PIK3R1 (Upstate), anti-SHC1 (BD Biosciences), anti-PTPN11 (Santa Cruz Biotechnologies), anti-GAB1 (Santa Cruz Biotechnologies) and anti-tubulin (Sigma-Aldrich). All original western blots are provided in **Supplementary Figures 17 and 18**.

Mass spectrometry characterization experiments. Initial characterization experiments of GRB2 APs were performed on a QSTAR Elite QqTOF mass spectrometer equipped with a nanospray III ion source (AB Sciex), and coupled to an Eksigent 1D+ Nano LC. Samples were injected directly onto home-built packed-tip columns. In brief, a 4 μm tip was pulled at the end of a 25 cm piece of 360 μm (external diameter) × 75 μm (internal diameter) fused silica capillary tubing using a Model 2000 Micropipette Laser Puller (Sutton Instrument). The tubing was packed with a methanol slurry of 3.5 μm Zorbax C-18 (Agilent) stationary phase using a pressure vessel set at 1,500 p.s.i.

Samples were run using a 90 min gradient from 5–30% solvent B (solvent A: 0.1% formic acid in water; solvent B: 0.1% formic acid in acetonitrile) at a flow rate of 250 nl/min. An information-dependent acquisition (IDA) method was set up with the MS survey mass range set between 400 amu and 1,600 amu, and six dependent MS/MS scans with a mass range set between 100 amu and 1,600 amu. The equivalent of 50% of a GRB2 AP was injected for each characterization experiment.

Tandem mass spectra were extracted, charge state deconvoluted and deisotoped in Analyst version 2.0. All MS/MS samples were analyzed using Mascot (Matrix Science). Mascot was set up to search the subset of all human proteins extracted from the ENSEMBL database (HS%50, 46,768 entries). Searches were performed with carbamidomethyl (C) set as a fixed modification and deamidated (NQ), pyro-Glu (N-term Q), pyro-Glu (N-term E), phospho Tyr, and phospho Ser/Thr set as variable modifications. Trypsin was selected for enzyme digestion with up to two missed cleavages. QqTOF type fragmentation was selected with peptide mass tolerance set to 80 p.p.m., and fragment mass tolerance set to 0.15 Da.

Scaffold (Proteome Software) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were generally accepted if they could be established at >95.0% probability as specified by the Peptide Prophet algorithm. Protein identifications were generally accepted if they could be established at >95.0% probability and contained at least one identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Phosphorylation sites that were identified were assigned ASCORE values (**Supplementary Data Set 3**).

Characterization experiments were performed through five independent 3xFLAG-GRB2 affinity purifications from three different stable 293T cell lines and two transient transfections (data deposited in ProteomeCommons.org

Tranche). Proteins identified in FLAG-GFP affinity purifications were manually removed (**Supplementary Data Set 4**). From this data set, a total of 109 proteins (including GRB2) were selected for SRM assay development (**Supplementary Table 1**).

SRM assay development. Characterization experiments were combined and searched in Protein Pilot yielding 219 proteins identified with a confidence >95% (false-discovery rate of 5%). Of these 219 protein hits, 108 were determined to be potential GRB2-associated interactors after background subtraction and were selected for SRM assay development. These proteins corresponded to an initial set of 420 peptides and 1,500 SRM transitions. Five unscheduled runs with 300 transitions each were required to characterize the selected SRMs. Peptides that were not detected in the first pass were collected into a second set of unscheduled SRM analysis runs. These unscheduled runs were used to determine which peptides would be confidently detected, and to assign a retention time for each of them. The MIDAS workflow was used to confirm peptide sequence where intensities were sufficient to generate high-confidence spectra. A subset of peptides did not trigger assignable and/or defendable MS/MS (~10%). In these cases, three factors were used to determine if transitions were reporting the appropriate peptide: the transitions had (i) to perfectly co-elute, (ii) to have the same relative ratio of intensities compared to fragments in the original discovery spectrum that were used to select the SRMs and (iii) the elution time had to match the original time from the discovery phase within a 10-min window. After the second run, only detected peptides/transitions were included in the assay. For the peptides that were not included in the assay, as many as three possible problems may have occurred: (i) a highly reliable MS/MS confirming the peptide identification was not obtained from the 5500 QTRAP through MIDAS, (ii) all selected transitions corresponding to the same peptide did not co-elute on the LC (which is essential to provide confident identification) and (iii) less than three transitions were detected for a given peptide. After optimization, the scheduled SRM (sSRM) assay contained 1,157 transitions corresponding to 90 proteins and 326 peptides (including 36 phospho-peptides from 19 different proteins). A total of 62 proteins were represented by 3 peptides, 11 proteins were represented by 2 peptides and 17 proteins by a single peptide.

Nevertheless, 19 proteins that were originally identified were not included in the final assay. Among these 19 proteins, 10 had only a single high-confidence peptide identified from only one of the five different samples used for the initial identification experiments. It may have been possible to use alternative ways (bioinformatic predictions and MS databases combined with synthetic peptides) to fully include all 109 proteins identified in characterization experiments and/or any other protein(s) of interest. However, there is no assurance that this approach would have been successful in detecting these peptides experimentally, despite the enhanced sensitivity of the targeted approach.

SRM workflow. The following workflow was used for running samples: (i) GRB2 affinity-purified sample, (ii) injection of 30 fmole of BSA (Micrhom)/60 fmole alpha-casein (digested in-house) as an LC/MS standard and (iii) a blank injection using the sSRM assay, to minimize carryover (**Supplementary Fig. 19**). For each sample (biological replicate), three technical/analytical replicates were run consecutively. The sample run order was completely randomized within each analytical replicate batch (independently to the previous/next analytical runs).

Trypsin digestions for samples from the same data set (e.g., all samples from an EGF time course stimulation experiment) were performed on the same day, but different samples/biological replicates were processed on different days. Different batches of trypsin may have been used.

The amount of sample injected for the sSRM assay work was equivalent to 10% of a GRB2 AP (from a single 15-cm plate). This amount of material minimized the potential for the GRB2 bait peptides to saturate the LC/MS system, yet still ensured sufficient sensitivity to detect all interacting proteins (as judged by detection in a positive control from pervanadate-stimulated cells). A second benefit to reducing the sample loading amount was that carryover from successive runs was minimal, as judged by the analysis of the blank injection that followed each experimental injection. Some carryover was observed for a small subset of peptides (<10%) in the assay and was

measured at 0.01–0.3% of intensity in the blank run, when compared to the previous sample run. One hydrophobic peptide gave higher carryover (about 3%) (**Supplementary Fig. 19**).

The BSA/casein control was used to track chromatographic reproducibility as well as instrument stability over the course of the experiments (**Supplementary Fig. 20**).

AP-MS identification experiments using growth factor-stimulated cells were run to assess whether new proteins were found compared to pervanadate-stimulated cells used in initial discovery experiments; no novel proteins were identified.

LC/MS setup for AP-SRM. The GRB2 SRM assay was developed and run on the cHiPLC nanoflex system (Eksigent) with a nanoLC Ultra 2D+ HPLC (3 µl sample loop on autosampler). Mobile phase buffers were: (A) 0.1% formic acid in water; (B) 0.1% formic acid in acetonitrile. Experiments were run on a 75 µm (i.d.) × 15 cm nano cHiPLC ChromXP C18 (3 µm, 120 Å) column. During assay development, 20% of the total affinity purification was injected per experiment. All experiments were run on a 5500 QTRAP (AB Sciex). Samples were run with the following LC method: 5% B buffer for 2 min followed by a linear gradient to 30% B buffer over 88 min. Column regeneration was performed through a 5 min wash in 90% B buffer followed by re-equilibration at 5% B buffer for 15 min. Highly reproducible chromatography was achieved on this system for the entire length of the study as the retention time for every peptide in the assay was consistent within ± 2.5 min for the entire time that data were acquired for this study (**Supplementary Fig. 20**).

SRM assay design, improvements and controls. Several strategies used in assay development and optimization contributed to increased specificity and elimination of false-positive detection, two problems that have been associated with SRM in the past¹⁷. First, we configured the assay from experimental data we acquired directly. This allows the simultaneous acquisition of distinct types of information essential for assay development, such as chromatographic retention times and fragmentation patterns. A specific retention time was thereby assigned to every peptide in the assay to provide a precise window for data acquisition. Second, peptide identity was continuously verified by MS/MS using the MIDAS workflow (**Supplementary Figs. 15 and 16**)³⁰. Third, the complexity of our samples was reduced by GRB2 enrichment, thus diminishing the risks of isobaric interference as well as false-positive rates. Finally, for each experimental condition (including controls), three independent samples were prepared from three different affinity purifications (referred to as ‘biological replicates’),

and each of them was analyzed three times (referred to as ‘technical/analytical replicates’). A positive control sample was run with each sample group.

Data analysis. The overall features of the data analysis workflow that was used are detailed in the **Supplementary Methods** and illustrated in **Supplementary Figure 3**. Briefly, as the response calculated for each transition gives independent evidence for the corresponding protein, weights that are propagated to the peptide and protein levels were assigned rather than eliminating variables that fail selection criteria. The initial weight values for the individual transitions are based on the measured signal-to-noise ratio but mapped to the range 0 to 1, so that weighted averages can be calculated. The transitions are treated independently.

To allow for variations in peptide ‘behavior’ (that is, changes in abundance in an experimental versus control condition), fold change and weight values for proteins were separated such that each protein had values for three possible behaviors: increased, decreased and unchanged. Individual peptides, however, may only change in one direction. In calculating peptide weight values, statistics that favor larger differences in expression level and smaller s.d. for the measurements were selected.

Normalization was based on a subset of GRB2 transitions that met certain criteria to improve the reliability. Because AP experiments were performed over a period of time, each sample was treated independently to allow for different experimental conditions, such as instrument sensitivity and biological variability that change the overall response. Although the expectation is that the sample replicates will have the same trend, the absolute change may be different, as can the variance because it depends on the square root of the intensity. This consideration precludes the use of statistical tests, such as the *t*-test, that assume variables are drawn from a population with the same variance.

After all processing, two features are available to help the user assess the results: (i) the fold change itself; and (ii) the weight of the fold change, which represents the variability across all transitions and replicates, and can be used as a discriminating factor. This can be reviewed at the peptide and protein level to indicate the contributions of the successive measurements (in triplicate) and individual AP samples (also in triplicate) to the overall variance.

To determine an appropriate cut-off value for the calculated weights, experimental data were simulated. Simulated weights and the corresponding $(1 - p)$ value that would have been obtained from a *t*-test were calculated. This strategy determined that a value of 0.7 corresponds roughly to $P = 0.05$ (**Supplementary Fig. 4**).



HARVARD MEDICAL SCHOOL BRIGHAM AND WOMEN'S HOSPITAL



Department of Pathology
75 Francis Street
Boston, Massachusetts 02115
Tel: 617 732-7675, Fax: 617 264-5169
E-mail: cfrench@partners.org

Christopher A. French, M.D.
Assistant Professor in Pathology

February 26, 2012

Anne-Claude Gingras, Ph.D.

Samuel Lunenfeld
Research Institute
Mount Sinai Hospital
992A-600 University Ave, Toronto, ON, M5G 1X5
CANADA

Anthony Pawson, Ph.D.

Samuel Lunenfeld
Research Institute
Mount Sinai Hospital
Joseph & Wolf Lebovic Health Complex
600 University Ave
Toronto Ontario M5G 1X5
CANADA

Panagis Filippakopoulos, Ph.D.

Structural Genomics Consortium
University of Oxford
Old Road Campus Research Building
Roosevelt Drive
Headington, Oxford, OX3 7DQ
UK

Stefan Knapp, Ph.D.

Structural Genomics Consortium
University of Oxford
Old Road Campus Research Building
Roosevelt Drive
Headington, Oxford, OX3 7DQ
UK

Dear Anne-Claude, Panagis, Stefan and Tony,

I want to offer my full and continued support to your important research program, and in particular to your CIHR proposal entitled "A systems approach towards the therapeutic modulation of the acetylome". I am particularly impressed by the synergy between the Toronto and Oxford teams, and the powerful techniques you have been implementing for the

study of the acetylome. As you know, my group is trying to fully understand the mechanisms behind the transformation mediated by the BRD4-NUT fusion, and to translate these findings to the clinic. Ultimately, we want to see a cure for this disease.

Parenthetically, I have diagnosed three cases of NUT midline carcinoma from Canada in the last year (Calgary, Toronto, and Quebec city). While a rare disease, it is relentlessly aggressive and we are seeing more and more cases every year as the oncology community become more aware.

I am therefore most enthusiastic about your unbiased proteomics approaches to define the interactome of BRD4-NUT (and eventually other NUT fusions) and the comparison you are proposing to undertake with the normal BRD4 and NUT proteins. I have already sent you full length BRD4-NUT fusion proteins for expression in mammalian, as well as various truncation products to express recombinantly, and will make additional constructs available to your groups as they are generated.

I am very eager to follow up on the new interactions you will be identifying for BRD4, BRD4-NUT and the NUT protein in isolation. As you are well aware, we are well experienced in cell culture-based and *in vivo* functional assays, and will be able to test the functional role(s) of any new interactor on growth, proliferation and tumorigenicity. I can make available to you for your primary testing of inhibitors for protein complex disruption and for the identification of protein interactors for the endogenous BRD4-NUT protein a number of cell lines derived from NUT midline Carcinoma (NMC) patients for which we have developed efficient protocols for RNA interference and recombinant protein expression. I would be curious to see if there are any interactors which are different in these cell line as compared to cells in which you force expression of FLAG-BRD4-NUT. I would also be happy to share my monoclonal antibody to NUT for your interaction proteomics analysis; I have used the anti-NUT monoclonal C52 successfully as well for immunoprecipitation, and this could be used in parallel in your studies to ensure that the interactions detected are specific.

Again, I wish you all the best with your grant and I am looking forward to start working on your interactors.

Sincerely,



Chris French

Chris French, M.D.



SGC



Prof. Stefan Knapp
Principal Investigator

Oxford January 12, 2012

Anne-Claude Gingras

Senior Investigator, Samuel Lunenfeld Research Institute

Tony Pawson

Distinguished Investigator, Samuel Lunenfeld Research Institute

Panagis Filippakopoulos

Principal Investigator, Structural Genomics Consortium, Oxford

Dear Anne-Claude, Tony and Panagis,

I am very happy to provide this letter in support of your grant to the CIHR entitled “A systems approach towards the therapeutic modulation of the acetylome”. I am excited by the data we have obtained as a group so far on this project for which the first publication is in press in *Cell*, and I am looking to our expanded collaboration.

My group at the Structural Genomics Consortium is committed to the realization of this project that capitalizes on the strengths developed in functional proteomics in Toronto, and our own expertise in structural and chemical biology. The committee should know that as group leader of the chemical biology team at the SGC Oxford, I will continue to closely interact with Panagis who was my postdoctoral fellow until very recently, and is now a group leader at the SGC. Panagis recently obtained a prestigious Wellcome Trust career transition award to continue working on the BET family of bromodomains, and he is group leader for the SGC bromodomain team. As such, he has access to all the resources of the SGC, and also has independent funding which will enable him to complete his part of the proposal. He is the ideal co-applicant for your group, for which I will continue to be a close collaborator.

I want to comment especially on the development of potent, cell permeable and selective protein inhibitors at the SGC. As you know very well, the SGC is partnering with different collaborators to systematically build such inhibitors (or chemical probes) for many proteins, including kinases and epigenetic proteins. We generate these inhibitors by structure-based screening strategies and by high-throughput screening (for example using alpha-screens). This is possible for us because we produce the target proteins in highly purified forms, and have in many cases also determined their structures. In the past years, we have developed a large program to generate chemical probes to disrupt the interaction between bromodomains and their acetylated-lysine targets. The interest for bromodomains as targets for chemical probes stems both from their key biological functions, and the fact that structurally they constitute excellent drug targets. We have indeed validated that the BET family inhibitor (+)-JQ1 which has already been used in Toronto, exhibits strong selectivity,

and efficiently competes out acetylated lysine peptides. This inhibitor is also potent to displace interactions in cells, and also prevents NUT midline carcinoma growth in a xenograft model, as was shown in Panagis' recent *Nature* paper. Recent work demonstrating that the inhibitor can be used against other types of tumors has further stimulated the interest of the SGC in developing additional chemical probes to other bromodomain-containing proteins. We already have another BET family inhibitor (PFI-1) that was released recently. (http://www.thesgc.org/scientists/chemical_probes/PFI-1/), and obtained recently probes against KAT3A/KAT3B (CREBB; EP300) and BAZ2B that we are making immediately available to you. As we continue screening for new lead compounds, we will continue sending them to Toronto as soon as they are available for testing using your proteomics tools.

Your project is extremely important to the understanding of the cellular function of bromodomain-containing proteins (many of which are poorly understood), and it also synergizes perfectly with our project, and especially with the rational drug design. As you are mapping preferred acetylated lysine peptide targets for each of the bromodomains using relevant cellular substrates, Panagis will be able to set-up affinity measurements assays, and select the best ligands for drug screening. This will be particularly important for poorly-characterized bromodomains, for which the targets may not have been present on the spot arrays we tested in the context of the *Cell* paper. I am therefore looking very much forward to our continued and already fruitful collaboration, and wish you all the best with your application.

Sincerely yours,



Prof. Stefan Knapp (Oxford University)

Alexey I. Nesvizhskii, Ph.D.
Associate Professor,
Department of Pathology,
Department of Computational Medicine and Bioinformatics
University of Michigan Medical School
Email: nesvi@umich.edu
Phone: 734-764-3616
Fax: 734- 936-7361



Feb 21, 2012

To: Anne-Claude Gingras
Senior Investigator, Samuel Lunenfeld Research Institute at Mount Sinai Hospital
Associate Professor, Department of Molecular Genetics, University of Toronto

Dear Anne-Claude,

I am writing this letter in full support of the CIHR grant entitled “A systems approach towards the therapeutic modulation of the acetylome” which you are writing with Drs Pawson, Knapp and Filippakopoulos.

There is no doubt in my mind that the project you are proposing will have strong relevance to human cancers. I am fully committed to continue working with you on the development of optimized analytical tools for mass spectrometry.

We have an excellent track-record at collaborating with each other to solve key problems in proteomics bioinformatics, particularly with regards to the development of tools for interaction proteomics. For example, we co-developed the Significance Analysis of INTeractome (SAINT) software tool which is critical in ensuring detection and reporting of only high quality protein interaction data. I am eager to continue developing these tools with you, and adapt them to new modes of quantification / new types of datasets as they become available.

My group is also developing additional tools for the analysis of proteomics data, which I am making fully available to you. For the interactomes from the modified chromatin IP (mChIP) protocols you have been generating, a possible difficulty in data analysis comes from the fact that multiple genes encoding histones share common peptides, making the identification of the exact histone more difficult. This issue of “shared peptides”, or “protein inference”, has been interesting me for many years, and was one of the reasons behind the development of the ProteinProphet tool which you have been using. In recent years, I have also expanded this type of logic to the distribution of spectral counts; this is implemented in the Abacus tool which I published recently. I would be very interested in helping you test Abacus for the acetylome data, and help you integrate it within your ProHits data management system.

As you know very well (since you are a co-author!) my lab has recently optimized an algorithm for the determination of the localization of a post-translational modification on a given peptide. Though we initially designed and tested this algorithm on phosphorylation data, this should also

work well for acetylation data, and we will fully test this with you. This tool will become important for you, as several of the peptides which you have been detecting in your acetylation site mapping contain more than one potential lysine modification sites (indeed, you already showed that in many cases, multiple acetylation is detected). I am also committed to work with you and your colleagues at AB SCIEX for scoring and quantifying SWATH data and have recruited a new student, Chih-Chiang Tsou who will embark on this project.

I am looking forward to our continued collaboration. Good luck with the application,

Best regards,

A handwritten signature in black ink, appearing to read "Nesvizhskii".

Alexey I. Nesvizhskii, Ph.D.



71 Four Valley Drive
Concord, Ontario L4K 4V8
Canada

February 27, 2012

Anne-Claude Gingras
Senior Investigator,
Samuel Lunenfeld Research Institute at Mount Sinai Hospital
600 University Avenue
Toronto, Ontario M5G 1X5

Dear Anne-Claude,

I am writing to offer my enthusiastic support for your CIHR application on the definition of the acetylome interactions in human cells that would accelerate the development of therapeutics for this cellular system critically implicated in human cancers.

AB Sciex has had a long-standing and fruitful collaboration with your group. We have developed improved approaches for Selected Reaction Monitoring on qTRAP instruments together, which have resulted, to date, in two publications and a submitted manuscript. Importantly, for our joint publication by Bisson et al. in *Nature Biotechnology*, we applied the skills of our mathematician Gordana Ivosev to build new analytical tools for analyzing this type of datasets. This is important, as these methods can be adapted relatively easily to the type of data you will be generating within the context of the proposed project.

We have also been working together on the testing of our newest generation mass spectrometer, the 5600 TripleTOF, for proteomics datasets and one of the first publications worldwide on this instrument came from our concerted efforts (Dunham et al., *Proteomics*, 2011). Since then, we have been actively working with your group to test different modes of quantification on this instrument. These include intensity-based measurements of the precursor ions (Kean et al., *J Biol Chem*, 2011), a modified SRM-like quantification that we call MRM^{HR} (Zheng et al., *under review at Science*), and the SWATH approach that you are proposing to use extensively here (Lambert et al., *in prep*). The data we have been obtaining has been very well received at the numerous scientific conferences where I, or some of the members of our group have been presenting.

We are particularly excited by the results we have been seeing with SWATH-based quantification, which accurately revealed changes in interactomes associated with mutations in the protein sequences for enzymes implicated in cancers. This approach should work very well for probing the effects of bromodomain mutations on the recovery of interactors. I am also excited by your proposed use of the approach to map interactions associated with the protein fusions of BRD4 and NUT and BRD3 and NUT, which are implicated in midline carcinomas.

With regard to the acetylation site mapping, the TripleTOF 5600 is also an ideal instrument, as you have demonstrated in your comparative analysis on the BRPF3 samples. In particular, the fact that this instrument enables detection of multiply-acetylated peptides may have strong implications for the specificity mapping that you are proposing here, given the fact that di- or multi-acetylated peptides appear to be preferred targets for at least some of the bromodomains.

AB Sciex is fully committed to continuing our work with you to facilitate seamless data analysis. As part of the collaboration we will continue to engage you in early access programs, helping us to define or refine new techniques and tools.

I am looking forward to our continued work together, and wish you all the best with your CIHR.

Sincerely,



Chris Lock, Ph. D.
Director, Mass Spectrometry Research
AB Sciex



71 Four Valley Drive
Concord, Ontario L4K 4V8
Canada

February 27, 2012

Tony Pawson
Distinguished Investigator,
Samuel Lunenfeld Research Institute at Mount Sinai Hospital
600 University Avenue
Toronto, Ontario M5G 1X5

Dear Tony,

I am writing to offer my enthusiastic support for your CIHR application on the definition of the acetylome interactions in human cells that would accelerate the development of therapeutics for this cellular system critically implicated in human cancers.

AB Sciex has had a long-standing and fruitful collaboration with your group. We have developed improved approaches for Selected Reaction Monitoring on qTRAP instruments together, which have resulted, to date, in two publications and a submitted manuscript. Importantly, for our joint publication by Bisson et al. in *Nature Biotechnology*, we applied the skills of our mathematician Gordana Ivosev to build new analytical tools for analyzing this type of datasets. This is important, as these methods can be adapted relatively easily to the type of data you will be generating within the context of the proposed project.

We have also been working together on the testing of our newest generation mass spectrometer, the 5600 TripleTOF, for proteomics datasets and one of the first publications worldwide on this instrument came from our concerted efforts (Dunham et al., *Proteomics*, 2011). Since then, we have been actively working with your group to test different modes of quantification on this instrument. These include intensity-based measurements of the precursor ions (Kean et al., *J Biol Chem*, 2011), a modified SRM-like quantification that we call MRM^{HR} (Zheng et al., *under review at Science*), and the SWATH approach that you are proposing to use extensively here (Lambert et al., *in prep*). The data we have been obtaining has been very well received at the numerous scientific conferences where I, or some of the members of our group have been presenting.

We are particularly excited by the results we have been seeing with SWATH-based quantification, which accurately revealed changes in interactomes associated with mutations in the protein sequences for enzymes implicated in cancers. This approach should work very well for probing the effects of bromodomain mutations on the recovery of interactors. I am also excited by your proposed use of the approach to map interactions associated with the protein fusions of BRD4 and NUT and BRD3 and NUT, which are implicated in midline carcinomas.

With regard to the acetylation site mapping, the TripleTOF 5600 is also an ideal instrument, as you have demonstrated in your comparative analysis on the BRPF3 samples. In particular, the fact that this instrument enables detection of multiply-acetylated peptides may have strong implications for the specificity mapping that you are proposing here, given the fact that di- or multi-acetylated peptides appear to be preferred targets for at least some of the bromodomains.

AB Sciex is fully committed to continuing our work with you to facilitate seamless data analysis. As part of the collaboration we will continue to engage you in early access programs, helping us to define or refine new techniques and tools.

I am looking forward to our continued work together, and wish you all the best with your CIHR.

Sincerely,



Chris Lock, Ph. D.
Director, Mass Spectrometry Research
AB Sciex



Sachdev Sidhu, PhD.

Associate Professor, The Donnelly Centre for Cellular + Biomolecular Research

Banting and Best Department of Medical Research

Ontario Institute for Cancer Research (Co-investigator) | University of Toronto

February 25, 2012

Anne-Claude Gingras,

Senior Investigator, Samuel Lunenfeld Research Institute at Mount Sinai Hospital

Associate Professor, Department of Molecular Genetics, University of Toronto

Tony Pawson,

Distinguished Investigator, Samuel Lunenfeld Research Institute at Mount Sinai Hospital

Professor, Department of Molecular Genetics, University of Toronto

Dear Anne-Claude and Tony,

I am very happy to provide this support letter for your CIHR grant entitled “A systems approach towards the therapeutic modulation of the acetylome”.

As you are well aware, my laboratory has been developing phage affinity reagents for isolated bromodomains, whose clones have been generated in collaboration with your co-applicant at the Structural Genomics Consortium. We have so far 22 domains for which Fabs (and derived IgGs) are available. Our goal is to soon obtain a panel of antibodies targeting the entire human bromo domain family. For example, we currently have seven unique Fabs targeting the first bromodomain of BRD3, which I know is of great interest to your labs. The Fabs are binding in the nanomolar range, and have been shown to immunoprecipitate proteins.

We expect that some of the Fabs bind at or near the KAc-binding pocket. As such, these could be candidate bromodomain-KAc inhibitors worth exploring in a cellular context. On the other hand, some Fabs may bind, for example, to the back surface of the bromodomain without interfering with KAc interactions; these would be good candidates to use as affinity handles for pulling down endogenous bromodomain-containing proteins or for detecting them by immunoblotting. Both types of Fabs are interesting, and we will aim to obtain both inhibitory and non-inhibitory reagents for each soluble bromodomain currently in our possession.

Besides the continued generation of Fabs and IgG for the individual bromodomains and their preliminary characterization by ELISA, we will be assessing the specificity of the interactions, by testing the closest relatives of the target bromodomains for reactivity. The next step will be to define the subset of the Fabs which act as inhibitors, and those that do not. To do so, we will test with your help whether the Fab and the acetylated peptide (the sequence of which will be determined by the beautiful study that your group has in press in Cell) can both bind to the

bromodomain simultaneously, and whether the Fab can prevent or disrupt the association of the KAc-peptide to the bromodomain.

From that point on, we will take one of two courses of actions: non-inhibitors will be prioritized as potential affinity reagents for immunoprecipitation, and will be reformatted into IgGs. We are very fortunate in this respect that you have the perfect test for these, as you can determine from parallel purifications of FLAG-tagged proteins with your FLAG antibodies and the new IgGs whether the interactomes are similar. If they are, these reagents could be employed to identify protein-protein interactions of endogenous proteins, including from cancer-derived cell extracts.

By contrast, the potential inhibitors will be prioritized to use as functional probes to define the biological roles for the bromodomain recognition of the acetylated peptides. Here again, I believe that your experimental design (e.g. pulldown of lysate-prepared KAc-peptides and MS analysis) will help accelerate the characterization of these inhibitors. Promising inhibitors will be cloned in a first step as single chain Fvs, and affixed with an epitope tag, and will be expressible in human cells.

I am very excited to continue this collaboration, and wish you all the best with your application.

Sincerely,



Sachedev Sidhu



Gary Bader, Ph.D.

Associate Professor, The Donnelly Centre for Cellular + Biomolecular Research
Department of Molecular Genetics + Computer Science | University of Toronto

February 29, 2012

Dear Anne-Claude, Tony and Panagis,

I am very enthusiastic about your project aiming to systematically map the specificity of bromodomains for acetyl-lysine peptides. I am very impressed by all the tools that your team has built, and the wealth of data that you have already accumulated together.

A major focus of my research is the specificity of protein domains for short peptide sequences. Given structural information and peptide specificity data, we have achieved a predictive model for how PDZ domains recognize their C-terminal hydrophobic ligands. We have now also expanded this research to WW domains and SH3 domains, both of which recognize proline-rich motifs (in collaboration with Tony and Anne-Claude – and Dev Sidhu). The data that your team will be obtaining in the context of the CIHR application will provide an outstanding dataset for the definition of specificity for all bromodomains. The fact that Panagis has already solved 30 structures for bromodomains, many of them with peptide ligands, provides an essential framework for this project. The combination of *in vitro* specificity mapping using synthetic peptides and peptides pools from cell lysates with direct mapping of acetylated sites on bromodomain-containing protein interactors will provide a dataset of unprecedented depth. This is extremely important in your case, as you have demonstrated in your beautiful paper in press in *Cell* that multiple bromodomains have a marked preference for multiply modified peptides. This is exciting, as it suggests cross-talk, e.g. to the phosphorylation machinery, and identifying these relationships if they exist would have strong repercussions on inhibitor development. Another important aspect of your research is that affinity measurements will be available for multiple sequences, enabling refinement of the computational models. Taken together with the solid bioinformatics expertise of my group, we will be able to provide a specificity map for all bromodomains, which will put you in an outstanding position to design better chemical probes.

I am also interested to help you reconstitute a relationship map between the writer-reader-eraser acetylation system components. I am particularly intrigued by Jean-Philippe's finding of multiple physical protein-protein interactions amongst the acetylases, deacetylases and bromodomain-containing proteins. This suggests – as you showed before for the phosphorylation machinery – coordinated regulation of this process.

I will also be able to help you with other aspects of your project: As you know very well, my group at the Donnelly center is developing computational methods for functional gene annotation, and we have released tools (such as GeneMANIA) that are very useful for at-a-glance views of the functional context of genes detected in large-scale data. Together with other tools (e.g. GSEA), which my group is very familiar with, I will be able to help you to functionally annotate your interactors and substrates to reveal cellular



Gary Bader, Ph.D.

Associate Professor, The Donnelly Centre for Cellular + Biomolecular Research
Department of Molecular Genetics + Computer Science | University of Toronto

pathways affected by the acetylation machinery.

I am looking forward to our continued collaboration, Tony and Anne-Claude, and am excited by the new collaboration with Panagis. I believe that your team can deliver on all the aims, and will be happy to contribute to your project.

Good luck with your CIHR,

Gary Bader
Associate Professor
University of Toronto

**Application for Funding – Budget****Funding Opportunity**

Operating Grant 2012-03-01

Nominated Principal Applicant/CandidateLast Name
GINGRASFirst Name
Anne-ClaudeInstitution
Samuel Lunenfeld Research Institute (Toronto)**Financial Assistance Required****Year 1**

Research Staff (excluding trainees)	No.	Salary	Benefits	CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Research Assistants	3.0	\$69,055	\$0	\$69,055	\$0	\$0	\$69,055
Technicians	1.0	\$43,045	\$11,794	\$54,839	\$0	\$0	\$54,839
Other personnel (as specified in Employment History)	0.0	\$0	\$0	\$0	\$0	\$0	\$0
Research Trainees	No.	Stipend	Benefits	CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Postdoctoral Fellows (post PHD, MD, etc.)	1.0	\$43,000	\$6,450	\$49,450	\$0	\$0	\$49,450
Graduate Students	1.0	\$24,500	\$0	\$24,500	\$0	\$0	\$24,500
Summer Students	0.0	\$0	\$0	\$0	\$0	\$0	\$0
Materials, Supplies and Services				CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Animals				\$0	\$0	\$0	\$0
Expendables				\$65,000	\$0	\$0	\$65,000
Services				\$0	\$0	\$0	\$0
Other (as specified in the Details of Financial Assistance Requested)				\$0	\$0	\$0	\$0
				CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Travel				\$4,400	\$0	\$0	\$4,400
Total Operating				\$267,244	\$0	\$0	\$267,244
Total Equipment				\$0	\$0	\$0	\$0
Total Request				\$267,244	\$0	\$0	\$267,244

**Application for Funding – Budget****Funding Opportunity**

Operating Grant 2012-03-01

Nominated Principal Applicant/CandidateLast Name
GINGRASFirst Name
Anne-ClaudeInstitution
Samuel Lunenfeld Research Institute (Toronto)**Financial Assistance Required****Year 2**

Research Staff (excluding trainees)	No.	Salary	Benefits	CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Research Assistants	3.0	\$69,055	\$0	\$69,055	\$0	\$0	\$69,055
Technicians	1.0	\$43,045	\$11,794	\$54,839	\$0	\$0	\$54,839
Other personnel (as specified in Employment History)	0.0	\$0	\$0	\$0	\$0	\$0	\$0
Research Trainees	No.	Stipend	Benefits	CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Postdoctoral Fellows (post PHD, MD, etc.)	1.0	\$43,000	\$6,450	\$49,450	\$0	\$0	\$49,450
Graduate Students	1.0	\$24,500	\$0	\$24,500	\$0	\$0	\$24,500
Summer Students	0.0	\$0	\$0	\$0	\$0	\$0	\$0
Materials, Supplies and Services				CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Animals				\$0	\$0	\$0	\$0
Expendables				\$65,000	\$0	\$0	\$65,000
Services				\$0	\$0	\$0	\$0
Other (as specified in the Details of Financial Assistance Requested)				\$0	\$0	\$0	\$0
				CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Travel				\$4,400	\$0	\$0	\$4,400
Total Operating				\$267,244	\$0	\$0	\$267,244
Total Equipment				\$0	\$0	\$0	\$0
Total Request				\$267,244	\$0	\$0	\$267,244

**Application for Funding – Budget****Funding Opportunity**

Operating Grant 2012-03-01

Nominated Principal Applicant/CandidateLast Name
GINGRASFirst Name
Anne-ClaudeInstitution
Samuel Lunenfeld Research Institute (Toronto)**Financial Assistance Required****Year 3**

Research Staff (excluding trainees)	No.	Salary	Benefits	CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Research Assistants	3.0	\$69,055	\$0	\$69,055	\$0	\$0	\$69,055
Technicians	1.0	\$43,045	\$11,794	\$54,839	\$0	\$0	\$54,839
Other personnel (as specified in Employment History)	0.0	\$0	\$0	\$0	\$0	\$0	\$0
Research Trainees	No.	Stipend	Benefits	CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Postdoctoral Fellows (post PHD, MD, etc.)	1.0	\$43,000	\$6,450	\$49,450	\$0	\$0	\$49,450
Graduate Students	1.0	\$24,500	\$0	\$24,500	\$0	\$0	\$24,500
Summer Students	0.0	\$0	\$0	\$0	\$0	\$0	\$0
Materials, Supplies and Services				CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Animals				\$0	\$0	\$0	\$0
Expendables				\$65,000	\$0	\$0	\$65,000
Services				\$0	\$0	\$0	\$0
Other (as specified in the Details of Financial Assistance Requested)				\$0	\$0	\$0	\$0
Travel				CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Total Operating				\$4,400	\$0	\$0	\$4,400
Total Equipment				\$267,244	\$0	\$0	\$267,244
Total Request				\$267,244	\$0	\$0	\$267,244

**Application for Funding – Budget****Funding Opportunity**

Operating Grant 2012-03-01

Nominated Principal Applicant/CandidateLast Name
GINGRASFirst Name
Anne-ClaudeInstitution
Samuel Lunenfeld Research Institute (Toronto)**Financial Assistance Required****Year 4**

Research Staff (excluding trainees)	No.	Salary	Benefits	CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Research Assistants	3.0	\$69,055	\$0	\$69,055	\$0	\$0	\$69,055
Technicians	1.0	\$43,045	\$11,794	\$54,839	\$0	\$0	\$54,839
Other personnel (as specified in Employment History)	0.0	\$0	\$0	\$0	\$0	\$0	\$0
Research Trainees	No.	Stipend	Benefits	CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Postdoctoral Fellows (post PHD, MD, etc.)	1.0	\$43,000	\$6,450	\$49,450	\$0	\$0	\$49,450
Graduate Students	1.0	\$24,500	\$0	\$24,500	\$0	\$0	\$24,500
Summer Students	0.0	\$0	\$0	\$0	\$0	\$0	\$0
Materials, Supplies and Services				CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Animals				\$0	\$0	\$0	\$0
Expendables				\$65,000	\$0	\$0	\$65,000
Services				\$0	\$0	\$0	\$0
Other (as specified in the Details of Financial Assistance Requested)				\$0	\$0	\$0	\$0
Travel				CIHR	Other Funding Sources		Total
					Cash*	In-Kind*	
Total Operating				\$4,400	\$0	\$0	\$4,400
Total Equipment				\$267,244	\$0	\$0	\$267,244
Total Request				\$267,244	\$0	\$0	\$267,244



Human Resources

**Employment History**

Name Larsen, Brett	Position Mass Spectrometry Spec	Current Salary Rate 126125	Current Source of Funding split(CIHR resrc
Name Veri, Amanda	Position Graduate student	Current Salary Rate 24500	Current Source of Funding N/A (start Sept 2
Name Zhang, JianPing	Position Programmer	Current Salary Rate 60576	Current Source of Funding Split on Gingras
Name Badillo, Beatriz	Position Technician, Level II	Current Salary Rate 54838	Current Source of Funding LLSC - Gingras +
Name Lambert, Jean-Philippe	Position Post-doctoral fellow	Current Salary Rate 49450	Current Source of Funding NSERC fellowsh
Name Tucholska, Monika	Position Research Technician, III	Current Salary Rate 63674	Current Source of Funding split(CIHR resrc
Name	Position	Current Salary Rate	Current Source of Funding
Name	Position	Current Salary Rate	Current Source of Funding
Name	Position	Current Salary Rate	Current Source of Funding
Name	Position	Current Salary Rate	Current Source of Funding
Name	Position	Current Salary Rate	Current Source of Funding
Name	Position	Current Salary Rate	Current Source of Funding
Name	Position	Current Salary Rate	Current Source of Funding
Name	Position	Current Salary Rate	Current Source of Funding
Name	Position	Current Salary Rate	Current Source of Funding
Name	Position	Current Salary Rate	Current Source of Funding
Name	Position	Current Salary Rate	Current Source of Funding

Detailed Budget Justification

All of the financial assistance requested will be used for salaries and laboratory expendables and services (no equipment) at the Toronto site only (Gingras and Pawson). Co-applicant Filippakopoulos is funded for his part of the project separately through European grants, and we are therefore not requesting funds for his part of the project. The highly technical nature of the project, with intensive cutting edge mass spectrometry and data analysis, involves several highly qualified personnel in the mass spectrometry facility, explaining the apparent disproportion between the consumables budget and the large budget we are requesting toward salaries.

To achieve the goals proposed we are requesting support for the following personnel and consumables:

1. RESEARCH STAFF AND TRAINEES

Jean-Philippe Lambert, post-doctoral fellow; 100% effort

Jean-Philippe is co-supervised by Tony Pawson and Anne-Claude Gingras, and has been the driving force behind the acetylome project. Jean-Philippe started at SLRI in October 2010, after a PhD at the University of Ottawa under the supervision of Daniel Figeys, a proteomics expert. During his PhD, Jean-Philippe developed the mChIP approach to identify protein-protein interactions amongst chromatin-associated proteins. Amongst his multiple publications (he already has co-authored 22 papers), the development of the mChIP approach (Lambert et al., Mol Cell Proteomics, 2009) and its application to the characterization of the budding yeast chromatin-associated interactome (Lambert et al., Mol Syst Biol, 2010) are particularly pertinent to the proposed studies. Since joining the Pawson and Gingras labs, Jean-Philippe has also written a review article, and contributed important experiments to the appended manuscript by Fillipakopoulos et al. (Cell, 2012, in press), which forms the basis of Aim 2. He has optimized the approach for the targeted study of the interactome which will be submitted shortly (to Mol Cell Proteomics) and has led the efforts on the validation of SWATH-based quantification (which will also be submitted to Mol Cell Proteomics). For this grant, he will be directly in charge of Aims 1 and 2, and coordinate the efforts of Amanda Veri and Beatriz Badillo. Jean-Philippe is currently supported by an NSERC which will expire in October 2012. The salary and benefits requested are in line with the SLRI salary scale for post-doctoral fellows.

\$ 43,000 p.a + 15% benefits (SLRI regulations) = \$49,450.

Amanda Veri, Graduate student –100% effort:

The minimal stipend for graduate students in the Department of Molecular Genetics at the University of Toronto is \$24,500. We request funding for Amanda Veri, who will join the Gingras lab as a graduate student in September 2012. Amanda was a summer student in the Gingras laboratory from May 2011 to December 2011, and immediately distinguished herself by her positive attitude, her hard work, and her ability to accomplish difficult tasks and troubleshoot problems. Within the context of this grant, she will be responsible for Aim 1.4 (map the endogenous interactions for selected bromodomain-containing proteins) and Aim 4 (testing the effects of the bromodomain inhibitors on acetyl-lysine-mediated interaction). She will work closely with post-doctoral fellow Jean-Philippe Lambert.

\$24,500 (no benefits) (U of Toronto guidelines)

Beatriz Badillo, Technician level II, 100% effort

Beatriz Badillo joined the Gingras lab in 2009, and has been involved in the improvements of the culture systems for interaction proteomics mapping; for her efforts, she was listed as a co-author on a manuscript on optimization of affinity purification coupled to mass spectrometry which we published last year (Dunham et al., Proteomics, 2011). She is also an author on a collaborative manuscript under

review at JBC. She is now fully dedicated to assisting Jean-Philippe Lambert with cloning and cell line generation. She has already cloned most of the wild type components of the acetylome. Within Aim 1, she will be in charge of generating all the DNA constructs to generate high quality interactome maps. Beatriz is an excellent molecular biologist and biochemist, and she will also be working with Jean-Philippe Lambert and Amanda Veri for the generation of the cDNA reagents and cell lines for the completion of all aims in this grant.

\$ 43,045 p.a + 27.4% benefits (SLRI): \$ 54,838

Brett Larsen, mass spectrometry specialist, 30% effort

Brett Larsen is our most senior mass spectrometry expert at the Lunenfeld, and is in charge of maintaining the 12 mass spectrometers currently in operation (he is assisted in this by excellent technicians), training the students and postdoctoral fellows and developing technically challenging protocols for the analysis of samples by mass spectrometry. He is co-supervised by Drs Pawson and Gingras. He is a major contributing author on seven publications with the Pawson and Gingras labs (not including submitted manuscripts), and is working closely with Jean-Philippe Lambert on all the technological / mass spectrometry aspects of this project. As such, the development of the SWATH quantification, improved acetylated lysine identification and other mass spectrometry related advances could not proceed without his help. Within the course of the project, he will continue working with Jean-Philippe Lambert to develop / improve mass spectrometry workflows. We expect that a significant part of his contribution – besides helping to run the samples – will be with data analysis. Brett's many years of experience in mass spectrometry data analysis, and in particular with regards to quantitative proteomics and analysis of post-translational modifications, put him in an exceptional position to ensure efficient and transparent analysis of the mass spectrometry data. Within the context of this project, Brett will also directly interact with JianPing Zhang to implement new tools specifically for the quantification and acetylation analysis proposed here. Because of his unique toolset and the competitive nature of the Toronto market, Brett commands a high salary (\$98,999.94/year). Until March 2012, a large portion of his salary was paid by a non-renewable CIHR Resource grant to Pawson and colleagues.

\$ 98,999.94 p.a. + 27.4% benefits = \$126,125.92;

Amount requested from the CIHR =30% = \$37837.78

We are only asking for the portion of Brett's salary that is directly related to this proposal; the acetylation project developed here is the most challenging of all mass spectrometry-related projects at the Samuel Lunenfeld Research Institute at the moment, and most of Brett's "research" time (as opposed to maintenance and training) will be spent on this project.

Monika Tucholska, Research Technician III – 30% effort:

Monika Tucholska was hired by the Pawson laboratory in January 2011. Prior to joining the Proteomics facility at SLRI, Monika obtained her Masters Degree in John Marshall's laboratory at Ryerson University and then went on to establish and maintain the mass spectrometry division of the Ryerson University Analytical Centre. Since arriving at the SLRI, Monika quickly established herself as a key personnel in the proteomics facility and has been of invaluable assistance in maintaining and performing quality control on the mass spectrometers that will be used in this research project and scheduling time on machines. For this project, Monika will assist Brett Larsen in maintenance of the mass spectrometers and will assist Brett and Jean-Philippe in sample acquisition and data analysis.

\$49,980.00 p.a. + 27.4% benefits = \$63,674.52

Amount requested from the CIHR = 30% = \$19,102.36

We are only asking for the portion of Monika's salary that is directly related to this proposal.

JianPing Zhang, programmer; 20% effort (no support for consumables requested)

JianPing is one of the two programmers behind the LIMS system for interaction proteomics, ProHits, and second author on the original ProHits manuscript (Liu et al., *Nature Biotech*) and on a new manuscript under review at *Proteomics*. JianPing is essential to the continued development of the software tools, including the implementation of quantification modules to incorporate MS/MS quantification which are proposed here. Importantly, JianPing will build new modules within our LIMS pipeline for the implementation of the acetylation analysis described in Aim 2. JianPing has been working under my supervision for 3 years (he was co-supervised by Mike Tyers until December 2010); he is paid as a contract employee, with a hourly salary of \$27.25 + 14% in lieu of benefits (\$31.05).

\$ 53,137 p.a + 14% in lieu of benefits (SLRI): \$ 60,576

Amount requested from CIHR = 20% = \$ 12,034

We are only asking for the portion of JianPing's salary that is directly related to this proposal.

TOTAL FOR RESEARCH STAFF AND TRAINEES: \$ 197,844.70

3A. EXPENDABLES

a. Reagents for mass spectrometry. The cost of materials for the proteomics experiments we are proposing is relatively high. All our experiments require sequencing-grade enzymes such as trypsin (~\$500). For the experiments using quantitative proteomics on the 5600 TripleTOF, we require RP-Chips for sample loading and separation (12 per year, at a cost of \$500/chip) Lastly, we require capillary tubing material, C18 reverse phase supports, HPLC plasticware, acetonitrile, HPLC-grade water and acids (formic acid, acetic acid, TFA) and bases (ammonium hydroxide), at an estimated cost of \$4,000.

\$11,000

b. Mammalian tissue culture. Last year, the cost of serum, selection medium and plates for the mammalian tissue culture experiments specifically for this project were ~\$ 4,000. For example, we have used ~6 cases of 150mm plates, 25 ml hygromycin, 4 bottles of FBS and 40 bottles of DMEM, not counting the transfection reagents and other supplies. SILAC and special growth media will be needed for some of the experiments (though we decrease the cost associated with SILAC reagents by buying reagents in bulk and making our own media, these costs are still high). Tissue culture will be expanded substantially for this project, as we need more material for the anti-acetyl lysine enrichment than we need for standard protein identification.

\$12,000

c. Affinity purification reagents. We require commercial antibodies and affinity reagents (e.g. large quantities of FLAG M2 manetic beads and anti-acetyl lysine antibodies) for our AP-MS experiments and custom antibodies for the other immunoprecipitation, immunoblotting and immunofluorescence experiments. This is a significant cost in our lab, and we spend nearly \$11,000 yearly on affinity reagents related to this project.

\$10,000

d. Molecular biology and standard reagents. Inhibitors (e.g. protease inhibitor cocktail, various deacetylase inhibitors) are estimated at \$2,500. Oligonucleotides and cloning enzymes (including Gateway clonase and PCR enzymes) are required for cloning and mutagenenesis (note that the project involves a large molecular biology component; \$8,000). Standard chemicals, molecular weight ladders and plasticware are also required for the project (\$15,000). Cost-recovery access to the Lunenfeld cDNA collection (\$10 per clone; estimated at \$500/year), and sequencing services are estimated at \$3,000.

\$ 29,000

e. Publication expenses: Preparation of figures for poster presentations and publication costs including page charges for one to two papers each year.

\$ 3,000

TOTAL FOR EXPENDABLES: \$65,000

4. TRAVEL

We request \$1,200 per year for each trainee (Lambert, Veri) paid on this grant to cover the cost of travel to scientific meetings to present their research findings on this project. While the vast majority of our meetings with co-applicant Filippakopoulos and other international collaborators are via webinars, we also request \$2000 each year for Jean-Philippe Lambert to travel to the Oxford site and learn new techniques there.

\$ 4,400

TOTAL TRAVEL: \$ 4,400

TOTAL PROJECT: \$ 267244.70

157721

CV Module

This page is for CIHR use only. It will not be included in the evaluation of your application for funding.

Family Name Gingras		Given Name Anne-Claude	Middle Initial(s)
Have you previously applied to CIHR for funding? Previous family name used Previous given name used		Yes <input checked="" type="checkbox"/> No <input type="checkbox"/> Title: Dr. <input checked="" type="checkbox"/> Mr. <input type="checkbox"/> Mrs. <input type="checkbox"/> Ms. <input type="checkbox"/> Prof. <input type="checkbox"/>	
Courier Address (If different from mailing address) Samuel Lunenfeld Research Institute Mt. Sinai Hospital Joseph and Wolf Lebovic Health Complex Samuel Lunenfeld Research Institute 600 University Avenue, Room 992A Toronto, Ontario CANADA (M5G 1X5)		Temporary Address Start Date _____ End Date _____	Primary Affiliation Name Samuel Lunenfeld Research Institute of Mount Sinai Hospital Start Date 07/2011 Primary Affiliation Address Samuel Lunenfeld Research Institute Mt. Sinai Hospital Joseph and Wolf Lebovic Health Complex Samuel Lunenfeld Research Institute 600 University Avenue, Room 992A Toronto, Ontario CANADA (M5G 1X5)
Contact numbers Phone Primary (416) 586-5027 Office Secondary (416) 586-4800 #8272 Laboratory Temporary Start Date _____ End Date _____		Fax Primary (416) 586-8869 Temporary Start Date _____ End Date _____	Electronic Addresses E-Mail gingras@lunenfeld.ca Web page address
Citizenship Canadian <input checked="" type="checkbox"/> Other <input type="checkbox"/> Other Country of Citizenship		Permanent Residence in Canada Permanent Resident <input type="checkbox"/> Date of permanent residency status DD/MM/YYYY Have you applied for permanent residency? Yes <input type="checkbox"/> No <input type="checkbox"/>	
Correspondence Language English <input checked="" type="checkbox"/> French <input type="checkbox"/>		Language English (Yes or No) French (Yes or No)	Read Write Speak Understand YES YES YES YES YES YES YES YES
Gender Male <input type="checkbox"/> Female <input checked="" type="checkbox"/>	Date of Birth (DD/MM/YYYY) 10/04/1972	Other Languages:	

Expertise

List up to ten (10) key words that best describe your expertise in research, instruments and technique.

Proteomics	Mass spectrometry
CCMs	Phosphorylation
Phosphatases	Affinity purification
Signal Transduction	GCKIII kinases
Protein interactions	angioma

Indicate and rank the disciplines that best correspond to your research interests. No additional pages may be added.

Discipline			Sub Discipline	
Rank	Code	Description	Code	Description
1.	11	MOLECULAR AND CELLULAR BIOLOGY		
2.	9	BIOCHEMISTRY		
3.	13	CANCER/ONCOLOGY		
4.				
5.				
6.				
7.				
8.				
9.				
10.				
11.				
12.				
13.				
14.				
15.				

Academic Background - One additional page may be added

Indicate all university degrees obtained and those in progress (where applicable) starting with the most recent. If you hold a co-degree from more than one institution (e.g. under the Soutien aux cotutelles de these de doctorat agreement between Quebec and France) enter each institution separately. Do not enter honorary degrees here, they should be listed in the Distinctions section.

Also indicate research training, such as postdoctoral or fellowship training. Trainees only: also list undergraduate and graduate research training experience.

Degree Type	Degree Name and Specialty	Institution/Organization and Country	Supervisor name	Start date (MM/YYYY)	Date received or expected (MM/YYYY)
Doctorate (PhD)	PhD Biochemistry	McGill University CANADA	Dr. Nahum Sonenberg	09/1994	09/2001
Bachelor's	Baccalaureat, Sciences Biochimie	Laval University CANADA	Dr. Andre Darveau	09/1991	05/1994

Work Experience

Starting with the most recent, indicate your current position, where applicable, and other academic and non-academic position(s) since the beginning of your university studies. For your current positions leave the end date blank. Additional pages will be accepted.

Position	Institution/Organization and Country	Department/Division and Faculty/School	Start Date (MM/YYYY)	End Date (MM/YYYY)
Associate Professor	University of Toronto CANADA	Department of Molecular Genetics	07/2011	
Senior Investigator	Samuel Lunenfeld Research Institute of Mount Sinai Hospital CANADA	Research N/A	07/2011	
Assistant Professor	University of Toronto CANADA	Department of Molecular Genetics	04/2006	07/2011
Investigator	Samuel Lunenfeld Research Institute of Mount Sinai Hospital CANADA	Research N/A	12/2005	07/2011
Postdoctoral fellow	Institute for Systems Biology, Seattle UNITED STATES	Proteomics N/A	01/2002	11/2005
PhD Student	McGill University CANADA	Biochemistry Faculty of Medicine, McGill U	05/1994	01/2001
Trainee/ research assistant	Centre de Recherche de la Croix-Rouge, Ste-Foy CANADA	Recherche Affiliated w/ Sciences et Genie, U Laval	09/1992	04/1994

Distinctions / Awards / Credentials

Starting with the most recent, indicate any recognitions received, including awards, fellowships, scholarships, licenses, qualifications, professional designation or credentials. Do not include Academic Appointments here, as they are detailed under Work Experience. Maximum 20 entries.

Name/Title and Type	Institution/Organization and Country	Effective Date (MM/YYYY)	End Date (MM/YYYY)	Specialty	Total Amount
2011 Canada's most powerful women: Top 100 Distinction	Women's Executive Network CANADA	12/2011			
Chair in Functional Proteomics (Renewed) Research award	Canada Research Chair Tier 2	07/2011	06/2016	Functional Proteomics	\$500,000
Chair in Cancer Proteomics Research award	Lea Reichmann CANADA	09/2008			
Early Researcher Award Research award	Ontario Ministry of Research and Innovation CANADA	2007			\$100,000
Chair in Functional Proteomics Research award	Canada Research Chair Tier 2	07/2006	06/2011	Functional Proteomics	\$500,000
Prix d'excellence (doctorat) Distinction	ADESAQ/FRSQ CANADA	2002			
Gordon A MacLachlan Doctoral Award Distinction	McGill University CANADA	05/2002			
Thomas Haliburton Henry Award Distinction	McGill University CANADA	05/2002			
Governor's General Gold Medal Distinction	McGill University	05/2002			
Prix d'Excellence (Sciences) Distinction	Academie des Grands Montrealais CANADA	04/2002			

Distinctions / Awards / Credentials

Starting with the most recent, indicate any recognitions received, including awards, fellowships, scholarships, licenses, qualifications, professional designation or credentials. Do not include Academic Appointments here, as they are detailed under Work Experience. Maximum 20 entries.

Name/Title and Type	Institution/Organization and Country	Effective Date (MM/YYYY)	End Date (MM/YYYY)	Specialty	Total Amount
Harold Weintraub Graduate Student Award Distinction	Fred Hutchinson Cancer Research Center UNITED STATES	02/2002			
Postdoctoral Fellowship Research award	CIHR CANADA	01/2002	12/2004		
Doctoral Award Research award	MRC CANADA	09/1998	09/2001		
Travel Grant Research award	National Cancer Institute of Canada CANADA	10/1997	11/1997		
Centennial (1967) Award Research award	NSERC CANADA	09/1994	08/1998		
Training Fellowship Research award	Fonds de la recherche en sante du Quebec CANADA	09/1993	05/1994		

Patents and Intellectual Property Rights

Record the total numbers of patents / copyrights in the following table.

OBTAINED			APPLICATIONS UNDER PROCESS			TOTAL PATENTS AND INTELLECTUAL PROPERTY RIGHTS
Total individual	Total collective	Sub-total	Total individual	Total collective	Sub-total	
0	0	0	0	0	0	0

PUBLICATIONS AND PRESENTATIONS

Give the number of publications and presentations in the course of your career. Detailed information should be attached as specified in the "Contributions - details" section.

Publications	Refereed Articles	Books and Monographs	Proceedings / Book Chapters / Contributions to a collective work	Abstracts / Notes	TOTALS
Already Published	70	0	17	0	87
Accepted or in the Press	2	0	3	0	5
					92

Invited presentations	65
-----------------------	----

LITERARY AND ARTISTIC WORKS

Provide the number of literary and artistic works created in the course of your career. Detailed information should be attached as specified in the "Contributions - details" section.

IN CIRCULATION			IN PROGRESS			TOTAL LITERARY AND ARTISTIC WORKS
Total individual	Total collective	Sub-total	Total individual	Total collective	Sub-total	
0	0	0	0	0	0	0

Supervisory Experience: To be completed by applicants requesting research trainees as part of their budget, salary support candidates and proposed supervisors of trainees.

Indicate the number of graduate students and postdoctoral fellows that you currently supervise or co-supervise. CIHR defines supervisory experience as the formal supervision or co-supervision of trainees. Enter zero (0) if not applicable.

Master 1Doctoral 1Post-Doctoral 5

Complete this form by listing the trainees that you have supervised/co-supervised (and are currently supervising/co-supervising) within the last five (5) years. Additional pages may be added if necessary.

* Flag those where you were/are the Primary Supervisor.

*	Name of Student	Program Type	Dates		Degree received or expected	Year Degree Rec'd (YYYY)	Research Project (Short title)	Current position and Institution
			Support Period From (MM/YY)	To (MM/YYYY)				
*	McBroom-Cerajewski, Linda	Postdoctoral Fellow, PhD	02/2012				Elements of specificity within the PP2A interaction network	
*	Knight, James	Postdoctoral Fellow, PhD	09/2011				Identification of substrates for GCKIII kinases	
*	Couzens, Amber	Postdoctoral Fellow, PhD	05/2011				Interplay between PP6 and Hippo signaling	
*	St-Denis, Nicole	Postdoctoral Fellow, PhD	12/2010				Cell-cycle interactome dynamics for Ser/Thr phosphatases	
*	Lambert, Jean-Phillipe	Postdoctoral Fellow, PhD	10/2010				Systematic proteomic characterization of human bromodomain specificity	
*	Dunham, Wade	Graduate Student	09/2009		Master's		Phosphorylation networks and modeling	U of Toronto
*	Kean, Michelle	Graduate Student	01/2008		Doctorate (PhD)		kinases and phosphatases in angioma	U. of Toronto
*	Chen, Ginny	Graduate Student	04/2006	12/2010	Doctorate (PhD)		PP4 Phosphatase interaction networks	New England Biolabs

Supervisory Experience: To be completed by applicants requesting research trainees as part of their budget, salary support candidates and proposed supervisors of trainees.

Indicate the number of graduate students and postdoctoral fellows that you currently supervise or co-supervise. CIHR defines supervisory experience as the formal supervision or co-supervision of trainees. Enter zero (0) if not applicable.

Master 1Doctoral 1Post-Doctoral 5

Complete this form by listing the trainees that you have supervised/co-supervised (and are currently supervising/co-supervising) within the last five (5) years. Additional pages may be added if necessary.

* Flag those where you were/are the Primary Supervisor.

*	Name of Student	Program Type	Dates		Degree received or expected	Year Degree Rec'd (YYYY)	Research Project (Short title)	Current position and Institution
			Support Period From (MM/YY)	To (MM/YYYY)				
*	D'Ambrosio, Lisa	Graduate Student	09/2007	12/2009	Master's	2009	dynein interactions with STRIPAK	Graduate student, U. of Toronto
*	Mullin, Michael	Postdoctoral Fellow, PhD	09/2007	12/2009	Postdoctorate		cell division and PP2A	Senior Scientist, Glaxo
*	Tisayakorn, Sally	Graduate Student	01/2006	12/2008	Master's	2009	Regulation of PP2A phosphatase assemblies	Quality Assurance, Allergan Inc.

Funds REQUESTED

List all sources of support applied for (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount requested (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal Molecular mechanisms of cerebral cavernous malformations		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Operating grant; with Derry
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$1,303,981	Support Period From (MM/YYYY) 10/2012	To (MM/YYYY) 09/2017
Title of Proposal A systems approach towards the therapeutic modulation of the acetylome		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Operating grant; with Pawson, Filippakopoulos
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$1,068,979	Support Period From (MM/YYYY) 10/2012	To (MM/YYYY) 09/2015
Title of Proposal Interplay between the PP6 phosphatase and Hippo signaling		
Funding Source Cancer Research Society (The)		Program Name
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$120,000	Support Period From (MM/YYYY) 07/2012	To (MM/YYYY) 06/2014
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)

Funds CURRENTLY HELD

List all sources of support currently held (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount awarded (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal Computational tools for mass spectrometry-based Interactome		
Funding Source National Institutes of Health (NIH) (USA)		Program Name RO1
Principal Applicant / Project Leader Nesvizhskii, A		Your Role Co-Applicant
Total Amount (CAN\$) \$60,000	Support Period From (MM/YYYY) 10/2010	To (MM/YYYY) 09/2015
Title of Proposal Global approaches to unravel PP2A function		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Operating Grant
Principal Applicant / Project Leader Gingras, Anne-Claude		Your Role Principal Applicant
Total Amount (CAN\$) \$960,385	Support Period From (MM/YYYY) 09/2010	To (MM/YYYY) 08/2015
Title of Proposal Understanding the assembly and function of dynamic signalling networks in complex diseases		
Funding Source Ontario Research Fund (ORF)		Program Name Global Leadership Round in Genomics
Principal Applicant / Project Leader Pawson, AJ		Your Role Co-Applicant
Total Amount (CAN\$) \$898,170	Support Period From (MM/YYYY) 07/2010	To (MM/YYYY) 06/2015
Title of Proposal Structure, function and regulation of PP4cs		
Funding Source Canadian Cancer Society Research Institute (CCSRI)		Program Name Operating Grant
Principal Applicant / Project Leader Gingras, Anne-Claude		Your Role Principal Applicant
Total Amount (CAN\$) \$688,605	Support Period From (MM/YYYY) 07/2009	To (MM/YYYY) 06/2014

Funds CURRENTLY HELD

List all sources of support currently held (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount awarded (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal Molecular Mechanisms of Cerebral Cavernous Malformations		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name CIHR Institute of Genetics – Bridge Funding; with Derry
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$100,000	Support Period From (MM/YYYY) 04/2012	To (MM/YYYY) 03/2013
Title of Proposal Core Proteomics Laboratory		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name CIHR Research Resource Grant - no direct funds to ACG
Principal Applicant / Project Leader Pawson, Anthony J.		Your Role Co-Applicant
Total Amount (CAN\$) \$497,150	Support Period From (MM/YYYY) 04/2008	To (MM/YYYY) 03/2012
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)

Funds HELD IN THE LAST FIVE YEARS

List all sources of support held in the last five years (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount awarded (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal Infrastructure for Functional Proteomics		
Funding Source Canada Foundation for Innovation (CFI)		Program Name Leaders Opportunity Fund (associated with CRC)
Principal Applicant / Project Leader Gingras, Anne-Claude		Your Role Principal Applicant
Total Amount (CAN\$) \$159,622	Support Period From (MM/YYYY) 06/2006	To (MM/YYYY) 07/2011
Title of Proposal SET and PP2A interactomes in leukaemia		
Funding Source Leukemia & Lymphoma Society of Canada (The) (LLSC)		Program Name Research Grant
Principal Applicant / Project Leader Gingras, Anne-Claude		Your Role Principal Applicant
Total Amount (CAN\$) \$116,000	Support Period From (MM/YYYY) 07/2009	To (MM/YYYY) 06/2011
Title of Proposal Ontario Proteomic Methods Centre (OMPC)		
Funding Source Ontario Research Fund (ORF)		Program Name Research Excellence Funding - 20K to ACG
Principal Applicant / Project Leader Pawson, Anthony		Your Role Co-Applicant
Total Amount (CAN\$) \$905,397	Support Period From (MM/YYYY) 10/2006	To (MM/YYYY) 03/2011
Title of Proposal Functional Proteomics of Serine/Threonine Phosphatases in the mTOR pathway		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Operating Grant
Principal Applicant / Project Leader Gingras, Anne-Claude		Your Role Principal Applicant
Total Amount (CAN\$) \$410,922	Support Period From (MM/YYYY) 08/2008	To (MM/YYYY) 07/2010

Funds HELD IN THE LAST FIVE YEARS

List all sources of support held in the last five years (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount awarded (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal Structure, function and regulation of PP4cs		
Funding Source National Cancer Institute of Canada (NCIC)		Program Name Terry Fox Foundation Research Grant for New Investigators
Principal Applicant / Project Leader Gingras, Anne-Claude		Your Role Principal Applicant
Total Amount (CAN\$) \$439,410	Support Period From (MM/YYYY) 07/2006	To (MM/YYYY) 06/2009
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)

Attachment Instructions

How to prepare and format all attachments:

Most Significant Contributions, Activities/Contributions, Interruptions/Delays, Patents/Copyrights (Part 2), and Publications (Part 2) details shall be contained in a CV attachment. Note: If you are using ResearchNet, you will need to provide each section identified as a separate PDF file.

The following format should be adhered to for this attachment.

- 8.5" X 11" (21.5 X 28.0 cm) white single-sided paper.
- Margins of $\frac{3}{4}$ " (2 cm).
- Minimum font size 12 point or 10 characters per inch.
- Six lines per inch, single-spaced, with no condensed type or spacing.
- Number pages consecutively after CV (If, for example, the print-out of the CV ends on page 8, the attachment would begin with page 9.).
- Each page header must contain the name and the sub-section header, e. g., Most Significant Contributions.

Most Significant Contributions

This section applies only to researchers, not to students. Identify a **maximum of five (5) contributions, with a maximum length of one page**, that best highlight your contribution or activities to research, defining the impact and relevance of each. (A contribution is understood to be a publication, literary or artistic work, conference, patent or copyright, contract or creative activity, commission, etc.) Your complete description may include the organization; position or activity type and description; from and to dates; and the basis on which this contribution is significant (i.e. relevance, target community and impact).

Activities / Contributions

The activities and contributions defined in this section should include both academic and non-academic achievements, and their impacts. **Limit the list to one page.**

Interruption(s) / Delays

Identify any administrative responsibilities, family or health reasons, or any other factors that might have delayed or interrupted any of the following: academia, career, scientific research, other research, dissemination of results, training, etc. Common examples of an interruption/delay might be a bereavement period following the death of a loved one, maternity/parental leave, or relocation of your research environment. **Limit the list to one page.**

Descriptions might include the start and end dates, the impact areas, and the reason(s) or a brief explanation of the absence.

Patents and Intellectual Property Rights

This section should include detail for patents and intellectual property rights for technology transfer, products, and services. Do not include Publications in this section. **Limit the list to one page.**

Descriptions for patents/intellectual property rights might include the title, patent/intellectual property rights number and date, country(ies) of issue, as well as the relevance or impact of this item and any inventor name(s) which pertain to it.

Publications List

List your most important publications and other research contributions over the past five years, according to the categories below. This is not necessarily a complete list, and is only intended to provide guidance. Categories can be added as needed. Use only items pertinent to the application. **There is no limit to the number of pages you can use.**

For Training or Salary Support Awards Candidates

- Candidates for training awards or New Investigator awards should list all publications, not just those of the last five years.
- All candidates for training or salary support awards must, for each multi-authored publication, define their role in the publication and indicate their percent contribution to the team effort.
- Candidates for training awards, with or without publications, are invited to comment on environmental factors that affected their capacity to publish.
- Candidates for salary support awards should, for multi-authored publications, underline the names of trainees whose work they supervised.

For Proposed Supervisors of Training Award Applicants

- Attach a maximum of two pages listing the titles and contributions over the past 5 years that will serve the application best.

MOST SIGNIFICANT CONTRIBUTIONS RELEVANT TO THIS APPLICATION

Though my contribution to the epigenetics field is fairly recent (Lambert et al., *Proteomics, in press*; Filippakopoulos et al., *Cell, in press*), several of my previous achievements and my track-record at successful collaborations place me in an outstanding position to accomplish the proposed project.

High-throughput interaction proteomics: “A global protein kinase and phosphatase network”; *Science*, 2010, 328:1043-6. This paper (a collaboration between the Gingras, Nesvizhskii and Tyers lab; Gingras is co-corresponding author) details a comprehensive survey of interaction partners for all yeast kinases and phosphatases. Importantly, this study revealed the unexpected finding that kinases physically associate with each other, which is likely to modulate the response to different signals, as well as the output of signaling. This study was profiled in *Science*, *Science Signalling* and in the *Nature Signalling Gateway*. In addition, this work enabled the generation of tools and protocols for interaction proteomics in *S. cerevisiae*, which have benefitted University of Toronto colleagues (e.g. Costanzo et al., *Science*, 2010; Baryshnikova et al., *Nature Methods*, 2010; Li et al., *Nat Biotech*, 2011; Sydorskyy et al., *Mol Cell Biol*). While this was the first large scale interactome project in my group, other interactome maps have since been completed, including a network centered on the human HSP90 system (collaboration with S. Lindquist, MIT; Taipale et al., *in prep*).

Optimized experimental approaches for mammalian interaction proteomics: “A novel, evolutionarily conserved phosphatase complex involved in cisplatin sensitivity”; Gingras et al., *Mol Cell Proteomics*, 2005; 4:1725-1740. In this manuscript, I optimized a mammalian TAP-tagging approach and used it to discover new regulatory subunits for protein phosphatase 4. Since starting my own group, we have developed more sensitive approaches, that have enabled us to provide new functional clues for signaling molecules (e.g. Chen et al., *J Biol Chem*, 2008; Goudreault et al., *Mol Cell Proteomics*, 2009; Dunham et al., *Proteomics*, 2011). We have also developed better quantification approaches for interaction proteomics, e.g. Kean et al., *J Biol Chem*, 2011, which we are coupling with RNA interference approaches and structural analysis to understand protein complex topology. More recent –and still unpublished – work capitalizes on a strong collaboration with mass spectrometry manufacturer AB SCIEX: we implemented a quantification method known as SWATH, and used it to accurately quantify parallel interactomes for wild type proteins and cancer-associated variants (Lambert et al., *in prep*). The mammalian protocols we have been developing are appropriate not only for phosphatases and kinase-associated complexes, but for a host of different proteins, as evidenced by collaborative work (e.g. Nakada et al., *Nature*, 2010; O’Donnell et al., *Mol Cell*, 2010; Lawo et al., *Curr Biol*, 2009).

Development of software tools for proteomics: “SAINT: probabilistic scoring of affinity purification - mass spectrometry data”; *Nature Methods*, 2011, 8:70-3. This manuscript – a collaboration between the Gingras and Nesvizhskii labs (Gingras is co-corresponding author) – describes the development of a new statistical tool for the analysis of interaction proteomics data. SAINT utilizes quantitative mass spectrometry information for given hits across the entire dataset (including, ideally, negative control runs) to calculate the probability of a bona fide protein-protein interaction. SAINT has been rapidly adopted by other research groups in Toronto and elsewhere (e.g. the Aebersold group at the ETH in Zurich) to remove contaminants from protein lists. SAINT even works on difficult cases, including a true positive interaction between the PP5 phosphatase and the chaperonin Hsp90, as we recently demonstrated (Skarra et al., *Proteomics*, 2011). We also developed (collaboration with Tyers, Gingras is senior author, a LIMS for interaction proteomics (Liu et al., *Nature Biotech*, 2010), which we recently linked to the SAINT tool (Liu et al., *submitted*; Choi et al., *submitted*). These and other bioinformatics tools enable our group, but also other researchers worldwide, to produce higher quality protein interaction maps.

PROFESSIONAL AFFILIATIONS AND ACTIVITIES

- 2011 Invitee committee member, Canadian Institutes of Health Research (CIHR), Genomics panel
- 2011 Co-editor (with Alexey Nesvizhskii) of a special issue of *Proteomics* on protein-protein interactions
- 2011 Consultant (category new investigators/ early career) for the International review of the Canadian Institutes of Health Research (CIHR)
- 2010-2011 Full member of the CCSRI (Canadian Cancer Society Research Institute) Panel F (Signaling)
- 2010- Director of the Research Training Center; Samuel Lunenfeld Research Institute
- 2010- Co-organizer (with Laurence Pelletier) of the Annual Samuel Lunenfeld Research Institute retreat
- 2010- Member of the Canadian Research Society Panel A
- 2010- Member of the Graduate student Recruitment committee, Department of Molecular Genetics, University of Toronto
- 2010- Director of the Bioinformatics core platform for the National Technology Platform (CFI; PI = Benoit Coulombe)
- 2009- Editorial Board, Molecular and Cellular Proteomics
- 2009-2010 Scientific Officer on the CCSRI (Canadian Cancer Society Research Institute) Panel F (Signaling)
- 2009 Co-organizer of the International Interactome Initiative (I3) Toronto Workshop (with Benoit Coulombe, IRCM and Tony Pawson, SLRI); co-leader of bioinformatics breakout session (with Pascal Braun, Dana Farber, Harvard Medical School).
- 2008-2011 Member of the Research highlight advisory panel for Nature Reviews Molecular Cell Biology
- 2008- Scientific Advisory Committee for the University of California in San Francisco (UCSF) mass spectrometry facility
- 2008- Member of the Canadian Institutes of Health Research (CIHR) Institute of Genetics New Principal Investigator Meeting Organizing Committee
- 2008- Member of the University of Toronto Molecular Genetics Scholarship Committee (Chair = Brigitte Lavoie 2008-2010; Lori Frappier 2011-)
- 2007-2008 Scientific Officer on the Virology and Structural Biology panel of National Cancer Institute of Canada (NCIC)
- 2007- Ad hoc reviewer: National Science and Engineering Research Council of Canada Discovery Grants; Swiss National Science Foundation; Austrian Science Fund; National Science Foundation.

INTERRUPTIONS/DELAYS

Gingras, Anne-Claude

NONE TO REPORT

PATENTS

Gingras, Anne-Claude

NONE TO REPORT

LIST OF PUBLICATIONS (2007-2012)**h-index = 45; cited 10088 times**

* indicates co-first authorship; ** indicates corresponding author; Gingras lab members underlined

CA = collaborator**PA = principal author****SRA = senior author****Peer reviewed primary articles (from a total of 72):**

1. Major, M.B., Camp, N.D., Berndt, J.D., Yi, XH, Goldenberg, S.J., Hubbert, C., Biechele, T.L., **Gingras, A.-C.**, Zheng, N., MacCoss, M.J., Angers, S. and Moon, R.T. (2007) Wilms Tumor Suppressor WTX negatively regulates WNT/b-catenin signaling. *Science* **316**:1043-1046. (**CA**)
2. Nakada, S., Chen, G.I., **Gingras, A.-C.**** and Durocher, D.** (2008) PP4 is a gH2AX phosphatase required for the recovery from the DNA damage checkpoint. *EMBO Reports*, **9**:1019-26. (**co-SRA**)
3. Rong, L., Livingstone, M., Sukarieh, R., Petroulakis, E., **Gingras, A.-C.**, Crosby, K., Smith, B., Polakiewicz, R.D., Pelletier, J., Ferraiuolo, M.A., and Sonenberg, N. (2008) Control of eIF4E cellular localization by eIF4E-binding proteins, 4E-BPs. *RNA* **14**:1318-27. (**CA**)
4. Chen, G.I., Tisayakorn, S., Jorgensen, C., D'Ambrosio, L.M., Goudreault, M. and **Gingras, A.-C.**** (2008) PP4R4/KIAA1622 forms a novel stable cytosolic complex with phosphoprotein phosphatase 4. *J Biol Chem*, **283**:29273-84. (**SRA**)
5. Goudreault, M., D'Ambrosio, L.M., Kean, M.J., Mullin, M., Larsen, B.G., Sanchez, A., Chaudhry, S., Chen, G.I., Sicheri, F., Nesvizhskii, A.I., Aebersold, R., Raught, B., and **Gingras, A.-C.**** (2009) A PP2A phosphatase high-density interaction network identifies a novel striatin-interacting phosphatase and kinase complex linked to the cerebral cavernous malformation 3 (CCM3) protein. *Mol Cell Proteomics*, **8**:157-71. (**SRA**)
6. Lawo, S., Bashkurov, M., Mullin, M., Gomez Ferreria, M., Kittler, R., Habermann, B., Tagliaferro, A., Poser, I., Hutchins, J., Buchholz, F., Peters, J-M., Hyman, A.A., **Gingras, A.-C.**, and Pelletier, L. (2009) The Augmin complex regulates centrosome and spindle integrity in mammalian cells. *Curr Biol*, **19**:816-26. (**CA**)
7. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin ZY, Liang W, Marback M, Paw J, San Luis BJ, Shuteriqi E, Tong AH, van Dyk N, Wallace IM, Whitney JA, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, **Gingras A.-C.**, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C. (2010) The genetic landscape of a cell. *Science*, **327**:425-21. (**CA**)
8. Bidinosti M., Ran I., Sanchez-Carbente M.R., Martineau Y., **Gingras A.-C.**, Gkogkas C., Raught B., Bramham C.R., Sossin W.S., Costa-Mattioli M., DesGroiseillers L., Lacaille J.C., Sonenberg N. (2010) Postnatal deamidation of 4E-BP2 in brain enhances association with raptor and alters kinetics of excitatory synaptic transmission. *Mol Cell*, **37**:797-808. (**CA**)
9. Breitkreutz A., Choi H., Sharon J., Boucher L., Neduvu V., Larsen B.G., Lin Z.-Y., Breitkreutz B.-J., Stark C., Liu G., Ahn, J., Dewar-Darch, D., Tang X., Almeida, V., Qin, Z.S., Pawson, T., **Gingras, A.-C.****, Nesvizhskii, A.**, Tyers, M.** (2010) A global protein kinase and phosphatase network. *Science*, **328**:1043-6. (**co-SRA**)
10. Mak, A.B., Ni, Z., Hewel, J., Chen, G.I., Zhong, G., Karamboulas, K., Blakely, K., Smiley, S., Marcon, E., Roudeva, D., Li, J., Olsen, J., Punna, T., Isserlin, R., Chetyrkin, S., **Gingras, A.-C.**, Emili, A., Greenblatt, J. and Moffat, J. (2010) A lentiviral-based functional proteomics approach identifies chromatin remodeling complexes important for the induction of pluripotency. *Mol Cell Proteomics*, **9**:811-23. (**CA**)

11. Choi, H., Kim, S., **Gingras, A.-C.**, Nesvizhskii, A.I. (2010). Analysis of protein complexes through model-based biclustering of label-free quantitative AP-MS data. *Mol Syst Biol*, **6**:385. (CA)
12. Sydorskyy, Y., Srikanth, T., Jeram, S.M., Wheaton, S., Vizeacoumar, F.J., Makhnevych, T., Chong, Y.T., **Gingras, A.-C.** and Raught, B. (2010). A novel mechanism for SUMO system control: regulated Ulp1 nucleolar sequestration. *Mol Cell Biol*, **30**:4452. (CA)
13. Nakada, S., Tai, I., Panier, S., Iemura, S.-I., Kumakubo, A., Munro, M., **Gingras, A.-C.**, Natsume, T., Suda, T. and Durocher, D. (2010) Non-canonical inhibition of DNA damage-dependent ubiquitylation by OTUB1. *Nature* 2010 **466**:941-6. (CA)
14. Liu, G., Zhang, J.P., Larsen, B., Stark, C., Breitkreutz, A., Lin, Z.-Y., Breitkreutz, B.-J., Ding, Y., Colwill, K., Pasculescu, A., Pawson, T., Wrana, J., Nesvizhskii, A.I., Raught, B., Tyers, M.**, and **Gingras, A.-C.**** (2010) ProHits: an integrated software platform for mass spectrometry-based interaction proteomics. *Nat Biotech*, **28**:1015-7. (co-SRA)
15. Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, Youn JY, Ou J, San Luis BJ, Bandyopadhyay S, Hibbs M, Hess D, **Gingras A.-C.**, Bader GD, Troyanskaya OG, Brown GW, Andrews B, Boone C, Myers CL. (2010) Quantitative analysis of fitness and geneic interactions on a genome scale. *Nat Methods*, **7**:1017-24 (CA)
16. O'Donnell L, Panier S, Wildenhain J, Tkach JM, Al-Hakim A, Landry MC, Escribano-Diaz C, Szilard RK, Young JT, Munro M, Canny MD, Kolas NK, Zhang W, Harding SM, Ylanko J, Mendez M, Mullin M, Sun T, Habermann B, Datti A, Bristow RG, **Gingras A.-C.**, Tyers M, Brown G and Durocher D. (2010) The MMS22L-TONSL complex mediates recovery from replication stress and homologous recombination. *Mol Cell*, **40**:619-31. (CA)
17. Choi, H., Larsen, B., Lin, Z.-Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., **Gingras, A.-C.**** and Nesvizhskii, A.I.** (2011) SAINT: probabilistic scoring of affinity purification - mass spectrometry data. *Nature Methods*, **8**:70-3. (co-SRA)
18. Templeton, G., Nimixk, M., Morrice, N.A., Campbell, D.G., Goudreault, M., **Gingras, A.-C.**, Takemiya, A., Shimazaki, K.I., Moorhead, G.B. (2011) Identification and characterization of Atl-2, an Arabidopsis homolog of an ancient protein phosphatase (PP1) regulatory subunit. *Biochem J*, **435**:73-83. (CA)
19. Skarra, D.V., Goudreault, M., Choi, H., Mullin, M., Nesvizhskii, A., **Gingras, A.-C.****, and Honkanen, R.** (2011) Label-free quantitative proteomics and SAINT analysis enable interactome mapping for the human Ser/Thr protein phosphatase 5. *Proteomics*, **11**:1508-16. (co-SRA)
20. Li, Z., Vizeacoumar, F., Bahr, S., Li, J., Warringer, J., Vizeacoumar, F., VanderSluis, B., Bellay, J., DeVit, M., Fleming, J., Stephens, A., Haase, J., Lin, Z.-Y., Baryshnikova, A., Min, R., Lu, H., Yan, Z., Jin, K., Barker, S., Datti, A., Giaever, G., Nislow, C., Bulawa, C., Costanzo, M., Myers, C., **Gingras, A.-C.**, Zhang, Z., Blomberg, A., Bloom, K., Andrews, B. and Boone, C. (2011) Systematic exploration of essential gene function with temperature-sensitive mutants. *Nature Biotech*, **29**:361-7. (CA)
21. Dunham, W., Larsen, B., Tate, S., Gonzalez Badillo, B., Tehami, Y., Kislinger, T., Goudreault M and **Gingras, A.-C.**** (2011) A cost-benefit analysis of multidimensional fractionation of affinity purification – mass spectrometry samples. *Proteomics*, **11**:2603-12. (SRA)
22. Kean, M.J., Ceccarelli, D., Goudreault, M., Tate, S., Larsen, B., Sanches, M., Gibson, L.C., Derry, W.B., Scott, I.C., Pelletier, L., Baillie, G.S., Sicheri, F., and **Gingras, A.-C.**** (2011) Structure-function analysis of core STRIPAK proteins: a signaling complex implicated in Golgi polarization. *J Biol Chem*, **15**:25065-75. (SRA)
23. Ceccarelli, D.F., Laister, R.C., Mulligan V.K., Kean, M.J., Goudreault, M., Scott, I.C., Derry, W.B., Chakrabartly A., **Gingras, A.-C.**** and Sicheri, F.** (2011) CCM3/PDCD10 Heterodimerizes with germinal center kinase III (GCKIII) proteins using a mechanism analogous to CCM3 homodimerization. *J Biol Chem*, **15**:25056-64. (co-SRA)

24. Olhovsky, M., Williton, K., Dai, A.Y., Pascalescu, A., Lee, J.P., Goudreault, M., Wells, C.D., Park, J.G., **Gingras, A.-C.**, Linding, R., Pawson, T., and Colwill, K. (2011) OpenFreezer: A reagent information management software system. *Nature Methods*, **28**:612-3. (**CA**)
25. Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.-P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., **Gingras, A-C.**, Arrowsmith, CH., and Knapp, S. (2012) Histone recognition and large structural analysis of the human bromodomain family. *In press, Cell*, CELL-S-11-01462 (**CA**)
26. Al-hakim, A.K., Bashkurov, M., **Gingras, A.-C.**, Durocher, D and Pelletier, L. (2012) Interaction proteomics identify NEURL4 and the HECT E3 ligase HERC2 as novel modulators of centrosome architecture. *Mol Cell Proteomics, Epub PMID:22261722* (**CA**)
27. Knight, J.R., Tian, R., Lee, R.E.C., Wang, F., Beauvais, A., Zou, H., Megeney, L.A., Gingras, A.-C., Pawson, T., Figeys, D. and Kothary, R. (2012) A novel whole-cell lysate kinase assay identifies substrates of the p38 MAPK in differentiating myoblasts. *In press, Skeletal Muscle*, MS4098815215883429. (**CA**)
28. Jovic, M., Kean, M.J., Szentpetery, Z., Polevoy, G., **Gingras, A.-C.**, Brill, J.A. and Balla, T. (2011) Two PI 4-kinases control lysosomal delivery of the Gaucher disease enzyme, β -glucocerebrosidase. *In press, Mol Biol Cell*, E11-06-0553. (**CA**)

Review articles and book chapters (from a total of 20):

1. Raught, B. and **Gingras, A.-C.**** (2007) Signaling to translation initiation. In *Translational control in biology and medicine*. M.B. Mathews, N. Sonenberg and J.W.B. Hershey, eds. Cold Spring Harbor Laboratory Press, Plainview, N.Y. 369-400. (**SRA**)
2. Chen, G.I. and **Gingras, A.-C.**** (2007) Affinity-purification mass spectrometry (AP-MS) of serine/threonine phosphatases. *Methods*, **42**:298-305. (**SRA**)
3. **Gingras, A.-C.****, Gstaiger, M., Raught, B and Aebersold, R. (2007) Analysis of protein complexes using mass spectrometry. *Nature Reviews Mol Cell Biol*, **8**:645-654. (**PA**)
4. **Gingras, A.-C.**** (2009) Journal Club: 35 later, mRNA caps still matter. *Nature Reviews Mol Cell Biol*, **10**:734. (**SRA**)
5. St-Denis, N. and **Gingras, A.-C.**** (2011) Mass spectrometric tools for systematic analysis of protein phosphorylation. Progress in molecular biology and translational sciences: Protein Phosphorylation in Health and Disease, Elsevier, edited by S. Shenolikar, *in press* (**SRA**).
6. **Gingras, A.-C.**** (2011) Protein Phosphatases, from Molecules to Networks. *EMBO Reports*, **12**:1211-13 (**PA, SRA**).
7. Lambert, J.-P., Pawson, T. and **Gingras, A.-C.**** (2011) Mapping physical interactions within chromatin by proteomic approaches. *In press, Proteomics, pmic.201100547* (**SRA**).
8. Dunham, W., Mullin, M and **Gingras, A.-C.**** (2011) Affinity-Purification coupled to Mass Spectrometry: Basic Principles and Strategies. *In press, Proteomics, pmic.201100523* (**SRA**).
9. Braun, P. and **Gingras, A.-C.** (2012) Protein interactions in the 20th century: from egg white to complex networks. *In press, Proteomics* (**CA**)

CV Module

This page is for CIHR use only. It will not be included in the evaluation of your application for funding.

Family Name Filippakopoulos		Given Name Panagis	Middle Initial(s)								
Have you previously applied to CIHR for funding? Previous family name used Previous given name used		Yes <input type="checkbox"/> No <input checked="" type="checkbox"/> Title: Dr. <input checked="" type="checkbox"/> Mr. <input type="checkbox"/> Mrs. <input type="checkbox"/> Ms. <input type="checkbox"/> Prof. <input type="checkbox"/>									
Courier Address (If different from mailing address) SGC - Old Road Campus Research Building Roosevelt Drive Oxford University Oxford, OX3 7DQ, UK UNITED KINGDOM		Temporary Address Start Date _____ End Date _____	Primary Affiliation Name University of Oxford Start Date 12/2011 Primary Affiliation Address SGC - Old Road Campus Research Building Roosevelt Drive Oxford University Oxford, OX3 7DQ, UK UNITED KINGDOM								
Contact numbers Phone Primary 44 (1865) 617576 Secondary Temporary Start Date _____ End Date _____		Fax Primary 44 (1865) 617575 Temporary Start Date _____ End Date _____	Electronic Addresses E-Mail panagis.filippakopoulos@sgc.ox.ac.uk Web page address								
Citizenship Canadian <input type="checkbox"/> Other <input checked="" type="checkbox"/> Other Country <u>GREECE</u> of Citizenship		Permanent Residence in Canada Permanent Resident <input type="checkbox"/> Date of permanent residency status DD/MM/YYYY Have you applied for permanent residency? Yes <input type="checkbox"/> No <input checked="" type="checkbox"/>									
Correspondence Language English <input checked="" type="checkbox"/> French <input type="checkbox"/>		Language English (Yes or No) French (Yes or No)	Read Write Speak Understand <table border="1"> <tr> <td>YES</td> <td>YES</td> <td>YES</td> <td>YES</td> </tr> <tr> <td>YES</td> <td>YES</td> <td>YES</td> <td>YES</td> </tr> </table>	YES							
YES	YES	YES	YES								
YES	YES	YES	YES								
Gender Male <input checked="" type="checkbox"/> Female <input type="checkbox"/>	Date of Birth (DD/MM/YYYY) 04/10/1973	Other Languages:									

Expertise

List up to ten (10) key words that best describe your expertise in research, instruments and technique.

crystallography	structural biology
bromodomains	peptide arrays
acetylation	isothermal calorimetry
drug development	cancer biology
protein interactions	protein domains

Indicate and rank the disciplines that best correspond to your research interests. No additional pages may be added.

Discipline			Sub Discipline	
Rank	Code	Description	Code	Description
1.	9	BIOCHEMISTRY	111	Protein/Amino Acid Biochemistry
2.	13	CANCER/ONCOLOGY	19	Anti-Cancer Agents
3.	13	CANCER/ONCOLOGY	1280	Cancer Therapy General
4.	9	BIOCHEMISTRY	1469	Proteomics
5.				
6.				
7.				
8.				
9.				
10.				
11.				
12.				
13.				
14.				
15.				

Academic Background - One additional page may be added

Indicate all university degrees obtained and those in progress (where applicable) starting with the most recent. If you hold a co-degree from more than one institution (e.g. under the Soutien aux cotutelles de these de doctorat agreement between Quebec and France) enter each institution separately. Do not enter honorary degrees here, they should be listed in the Distinctions section.

Also indicate research training, such as postdoctoral or fellowship training. Trainees only: also list undergraduate and graduate research training experience.

Degree Type	Degree Name and Specialty	Institution/Organization and Country	Supervisor name	Start date (MM/YYYY)	Date received or expected (MM/YYYY)
Doctorate (PhD)	PhD	University of Michigan UNITED STATES	Professor Dimitris Coucouvanis	10/1998	04/2004
Bachelor's	Baccalaureat, Sciences Chemistry	Aristotelian University of Thessaloniki GREECE	Professor Dimitris Kessissoglou	10/1992	06/1998

Work Experience

Starting with the most recent, indicate your current position, where applicable, and other academic and non-academic position(s) since the beginning of your university studies. For your current positions leave the end date blank. Additional pages will be accepted.

Position	Institution/Organization and Country	Department/Division and Faculty/School	Start Date (MM/YYYY)	End Date (MM/YYYY)
Principal Investigator & Wellcome Trust Career Development Fellow	University of Oxford UNITED KINGDOM	Nuffield Department of Medicine	12/2011	
Senior Scientist/Team Leader	University of Oxford UNITED KINGDOM	Structural Genomics Consortium	07/2007	11/2011
Postdoctoral Researcher	University of Oxford UNITED KINGDOM	Phosphorylation Dependant Signalling group	06/2004	06/2007

Distinctions / Awards / Credentials

Starting with the most recent, indicate any recognitions received, including awards, fellowships, scholarships, licenses, qualifications, professional designation or credentials. Do not include Academic Appointments here, as they are detailed under Work Experience. Maximum 20 entries.

Name/Title and Type	Institution/Organization and Country	Effective Date (MM/YYYY)	End Date (MM/YYYY)	Specialty	Total Amount
Medicine Merritt Award Research award	Oxford University Nuffield Department of Medicine Merritt Award	2009			

Patents and Intellectual Property Rights

Record the total numbers of patents / copyrights in the following table.

OBTAINED			APPLICATIONS UNDER PROCESS			TOTAL PATENTS AND INTELLECTUAL PROPERTY RIGHTS
Total individual	Total collective	Sub-total	Total individual	Total collective	Sub-total	
0	0	0	0	0	0	0

PUBLICATIONS AND PRESENTATIONS

Give the number of publications and presentations in the course of your career. Detailed information should be attached as specified in the "Contributions - details" section.

Publications	Refereed Articles	Books and Monographs	Proceedings / Book Chapters / Contributions to a collective work	Abstracts / Notes	TOTALS
Already Published	25	0	0	0	25
Accepted or in the Press	1	0	0	0	1
					26

Invited presentations	5
-----------------------	---

LITERARY AND ARTISTIC WORKS

Provide the number of literary and artistic works created in the course of your career. Detailed information should be attached as specified in the "Contributions - details" section.

IN CIRCULATION			IN PROGRESS			TOTAL LITERARY AND ARTISTIC WORKS
Total individual	Total collective	Sub-total	Total individual	Total collective	Sub-total	
0	0	0	0	0	0	0

Supervisory Experience: To be completed by applicants requesting research trainees as part of their budget, salary support candidates and proposed supervisors of trainees.

Indicate the number of graduate students and postdoctoral fellows that you currently supervise or co-supervise. CIHR defines supervisory experience as the formal supervision or co-supervision of trainees. Enter zero (0) if not applicable.

Master 0Doctoral 0Post-Doctoral 1

Complete this form by listing the trainees that you have supervised/co-supervised (and are currently supervising/co-supervising) within the last five (5) years. Additional pages may be added if necessary.

* Flag those where you were/are the Primary Supervisor.

*	Name of Student	Program Type	Dates		Degree received or expected	Year Degree Rec'd (YYYY)	Research Project (Short title)	Current position and Institution
			Support Period From (MM/YY)	To (MM/YYYY)				
*	Picaud, Sarah	Postdoctoral Fellow, PhD	02/2009	06/2013			Chemical probes for epigenetic readers of the Bromodomain Family	

Funds REQUESTED

List all sources of support applied for (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount requested (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal MYC onco-protein as a therapeutic target: Development of new molecules with anti-cancer potency		
Funding Source Other		Program Name Greek Ministry of Education grant
Principal Applicant / Project Leader Efstratiadis, Argiris		Your Role Co-Applicant
Total Amount (CAN\$) \$0	Support Period From (MM/YYYY) 01/2011	To (MM/YYYY) 01/2016
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)

Funds CURRENTLY HELD

List all sources of support currently held (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount awarded (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal Structural role of BET dromodomains in transcription		
Funding Source Ministry of Research and Innovation (MRI) (Ontario)		Program Name Research grant
Principal Applicant / Project Leader Pawson, Anthony Dr		Your Role Co-Applicant
Total Amount (CAN\$) \$1,625,700	Support Period From (MM/YYYY) 12/2011	To (MM/YYYY) 12/2016
Title of Proposal Chemical probes for interrogating protein-protein interactions in disease states		
Funding Source Wellcome Trust Career Development Program		Program Name Wellcome Trust Career Development Program
Principal Applicant / Project Leader Pawson, Anthony Dr		Your Role Principal Applicant
Total Amount (CAN\$) \$0	Support Period From (MM/YYYY) 10/2011	To (MM/YYYY) 09/2015
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)

Funds HELD IN THE LAST FIVE YEARS

List all sources of support held in the last five years (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount awarded (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)

Attachment Instructions

How to prepare and format all attachments:

Most Significant Contributions, Activities/Contributions, Interruptions/Delays, Patents/Copyrights (Part 2), and Publications (Part 2) details shall be contained in a CV attachment. Note: If you are using ResearchNet, you will need to provide each section identified as a separate PDF file.

The following format should be adhered to for this attachment.

- 8.5" X 11" (21.5 X 28.0 cm) white single-sided paper.
- Margins of $\frac{3}{4}$ " (2 cm).
- Minimum font size 12 point or 10 characters per inch.
- Six lines per inch, single-spaced, with no condensed type or spacing.
- Number pages consecutively after CV (If, for example, the print-out of the CV ends on page 8, the attachment would begin with page 9.).
- Each page header must contain the name and the sub-section header, e. g., Most Significant Contributions.

Most Significant Contributions

This section applies only to researchers, not to students. Identify a **maximum of five (5) contributions, with a maximum length of one page**, that best highlight your contribution or activities to research, defining the impact and relevance of each. (A contribution is understood to be a publication, literary or artistic work, conference, patent or copyright, contract or creative activity, commission, etc.) Your complete description may include the organization; position or activity type and description; from and to dates; and the basis on which this contribution is significant (i.e. relevance, target community and impact).

Activities / Contributions

The activities and contributions defined in this section should include both academic and non-academic achievements, and their impacts. **Limit the list to one page.**

Interruption(s) / Delays

Identify any administrative responsibilities, family or health reasons, or any other factors that might have delayed or interrupted any of the following: academia, career, scientific research, other research, dissemination of results, training, etc. Common examples of an interruption/delay might be a bereavement period following the death of a loved one, maternity/parental leave, or relocation of your research environment. **Limit the list to one page.**

Descriptions might include the start and end dates, the impact areas, and the reason(s) or a brief explanation of the absence.

Patents and Intellectual Property Rights

This section should include detail for patents and intellectual property rights for technology transfer, products, and services. Do not include Publications in this section. **Limit the list to one page.**

Descriptions for patents/intellectual property rights might include the title, patent/intellectual property rights number and date, country(ies) of issue, as well as the relevance or impact of this item and any inventor name(s) which pertain to it.

Publications List

List your most important publications and other research contributions over the past five years, according to the categories below. This is not necessarily a complete list, and is only intended to provide guidance. Categories can be added as needed. Use only items pertinent to the application. **There is no limit to the number of pages you can use.**

For Training or Salary Support Awards Candidates

- Candidates for training awards or New Investigator awards should list all publications, not just those of the last five years.
- All candidates for training or salary support awards must, for each multi-authored publication, define their role in the publication and indicate their percent contribution to the team effort.
- Candidates for training awards, with or without publications, are invited to comment on environmental factors that affected their capacity to publish.
- Candidates for salary support awards should, for multi-authored publications, underline the names of trainees whose work they supervised.

For Proposed Supervisors of Training Award Applicants

- Attach a maximum of two pages listing the titles and contributions over the past 5 years that will serve the application best.

MOST SIGNIFICANT CONTRIBUTIONS

I recently became a group leader (Oxford, UK) after completing post-doctoral training at the Structural Genomics Consortium in Oxford. Throughout my career, I have contributed several key studies by combining structural biology and drug design with collaborative biology studies. In my own group, I am focusing on the characterization of the functional roles of BET-family bromodomains and I continue to be implicated in determining the molecular mechanisms for bromodomain recognition of acetylated peptides. The following exemplify my most significant research contributions:

SH2 domain structures and tyrosine kinase activation: Regulation of phosphorylation is a key component of cellular signalling, often controlled by inter- and intra-molecular interactions of proteins with modular domains, such as the Src homology 2 (SH2) and 3 (SH3) domains. Tyrosine kinases in particular are medicinally interesting targets due to their oncogenic or tumour suppressor function depending on cellular context. Since their discovery, SH2 domains have been associated with negative regulation of kinase activity. I solved the three dimensional crystal structure of the human oncogene Fps/Fes and in collaboration with the laboratories of Tony Pawson (University of Toronto) and Giulio Superti-Furga (CeMM, Austria) unravelled a novel mechanism of tyrosine kinase activation, whereby the SH2 module acts as a stabilizer of the kinase active state. This work was a significant breakthrough as it explained data accumulated over 25 years of research that contradicted the structural information available at the time, highlighting the importance of modular domains in the regulation of protein activity (see Filippakopoulos, P. *et al.* Cell, 2008 **134** p.793 and Filippakopoulos, P. *et al.* Curr Opin Struct Biol, 2009 **19** p.643).

Protein phosphatase structural biology: Protein phosphatases are also key components of cellular signalling, working in concert with protein kinases, maintaining a fine balance of phosphorylation in cells. Dysregulation of their function has been associated with many diseases and members of the protein tyrosine phosphatase (PTP) family have been recognized as potential targets for development of therapeutics. Within the team I participated in the large scale structural and biophysical analysis of 24 PTP family members of this important family of proteins, generating a valuable resource which gave insight into intra-family diversity, catalytic activity, substrate recognition, and auto-regulatory self-association. This work constituted the first family-wide structural comparison of a large human protein family (see Barr, A.J. *et al.* Cell, 2008 **136** p.352).

Epigenetics and bromodomain specificity: Epigenetics describe changes in gene expression that are stable between cell divisions, and sometimes between generations, without involving changes in the underlying DNA sequence of an organism. Epigenetic events have been recognized for the potential to generate new therapeutics; however cross-talk with cellular signalling remains largely unknown. Signal transduction to the transcription machinery involves among other processes the ‘reading’ of the modification states of chromatin, the combination of DNA and histones. As a team leader at the SGC, and in collaboration with James Bradner, Christopher French and Andrew Kung (Harvard University) we identified and fully characterized a small molecule that specifically inhibited the readout of a modification state of chromatin. Using this molecule we demonstrated that it can treat in animals a severe and incurable form of cancer (NUT Midline Carcinoma) leading to survival, highlighting the tremendous potential of the unprecedented therapeutic value of inhibiting readers of epigenetic modifications, which will lead to new medicines (see Filippakopoulos, P. *et al.* Nature, 2010 **468** p.1067). Leading the same team of scientists within the SGC and in collaboration with the groups of Cheryl Arrowsmith, Anne-Claude Gingras and Tony Pawson, we established a large scale structural and functional analysis of human bromodomain proteins leading to the publication of a high quality resource article which defines bromodomain specificity (see Filippakopoulos, P. *et al.* Cell, 2012 **In Press**).

PROFESSIONAL AFFILIATIONS AND ACTIVITIES**TEACHING EXPERIENCE**

2003 – 2004	University of Michigan, Chemistry Department Graduate Student Mentor
1998 – 2004	University of Michigan, Chemistry Department Graduate Student Instructor
2000 – 2001	Advanced Inorganic Chemistry
1999 – 2001	Intermediate Inorganic Chemistry
1998 – 2004	General Chemistry

ADMINISTRATIVE EXPERIENCE

2007 – 2011	Team Leader - I lead a team of two scientists, a post doctoral fellow and a technical assistance, planning and supervising research within the Structural Genomics Consortium.
-------------	---

INVITED PRESENTATIONS (PEER REVIEWED)

2011	Protein Modules and Networks in Health and Disease (FEBS Workshop) - Seefeld, Austria - “ <i>Inhibiting Bromodomains? All BETs are OFF!</i> ”
2010	14 th Signal Transduction Society (STS) meeting - Weimar, Germany - “ <i>Inhibiting Bromodomains? All BETs are OFF!</i> ”
2009	Protein Modules and Networks in Health and Disease (FEBS Workshop) - Seefeld, Austria - “ <i>Chasing Histone Tails: The BROMO-domain Module</i> ”
2008	21 st Congress and Assembly of the International Union of Crystallography (IUCR) - Osaka, Japan - “ <i>Fes kinase structure reveals cooperative interactions between the SH2-kinase domains and substrate</i> ”
2007	18 th West Coast Protein Crystallography Workshop (WCPCW) - Asilomar, USA - “ <i>Human Adenylate Kinases – Plasticity and Function</i> ”

INTERRUPTIONS/DELAYS

Filippakopoulos, Panagis

NONE TO REPORT

PATENTS

Filippakopoulos, Panagis

NONE TO REPORT

List of Publications**Total: 25, Number of citations: 405**

- 1) Filippakopoulos, P.*, Picaud, S., Mangos, M., Keates, T., Lambert, J.P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., Gingras, A.C., Arrowsmith, C.H., Knapp, S.* (2012). Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. **Cell In-Press**
[* corresponding author]
- 2) Filippakopoulos, P., Picaud, S., Fedorov, O., Keller, M., Wrobel, M., Morgenstern, O., Bracher, F., Knapp, S. (2011). Benzodiazepines and benzotriazepines as protein interaction inhibitors targeting bromodomains of the BET family. **Bioorg Med Chem**, Nov 4.
- 3) Huber, K., Brault, L., Fedorov, O., Gasser, C., Filippakopoulos, P., Bullock, A.N., Fabbro, D., Trappe, J., Schwaller, J., Knapp, S., Bracher, F. (2012). 7,8-Dichloro-1-oxo- β -carbolines as a versatile scaffold for the development of potent and selective kinase inhibitors with unusual binding modes. **J Med Chem**, **55**:403-13.
- 4) Muller, S., Filippakopoulos, P., Knapp, S. (2011). Bromodomains as therapeutic targets. **Expert Rev Mol Med** **13**, e29.
- 5) Hewings, D.S., Wang, M., Philpott, M., Fedorov, O., Uttarkar, S., Filippakopoulos, P., Picaud, S., Vuppusetty, C., Marsden, B., Knapp, S., Conway, S.J., Heightman, T.D. (2011). 3,5-dimethylisoxazoles act as acetyl-lysine-mimetic bromodomain ligands. **J Med Chem** **54**, 6761-70
- 6) Philpott, M., Yang, J., Tumber, T., Fedorov, O., Uttarkar, S., Filippakopoulos, P., Picaud, S., Keates, T., Felletar, I., Ciulli, A., Knapp, S., Heightman, T.D. (2011). Bromodomain-peptide displacement assays for interactome mapping and inhibitor discovery. **Mol Biosyst** **7**, 2899-908.
- 7) Miduturu, C.V., Deng, X., Kwiatkowski, N., Yang, W., Brault, L., Filippakopoulos, P., Chung, E., Yang, Q., Schwaller, J., Knapp, S., King, R.W., Lee, J.D., Herrgard, S., Zarrinkar, P., Gray, N.S. (2011). High-throughput kinase profiling: a more efficient approach toward the discovery of new kinase inhibitors. **Chem Biol** **18**, 868-79.
- 8) Debdab, M., Carreaux, F., Renault, S., Soundararajan, M., Fedorov, O., Filippakopoulos, P., Lozach, O., Babault, L., Tahtouh, T., Baratte, B., Ogawa, Y., Hagiwara, M., Eisenreich, A., Rauch, U., Knapp, S., Meijer, L., Bazureau, J.P. (2011). Leucettines, a class of potent inhibitors of cdc2-like kinases and dual specificity, tyrosine phosphorylation regulated kinases derived from the marine sponge leucettamine B: modulation of alternative pre-RNA splicing. **J Med Chem** **54**, 4172-86.
- 9) Feng, L., Geisselbrecht, Y., Blanck, S., Wilbuer, A., Atilla-Gokcumen, G.E., Filippakopoulos, P., Kraling, K., Celik, M.A., Harms, K., Maksimoska, J., Marmorstein, R., Frenking, G., Knapp, S., Essen, L.O., Meggers, E. (2011). Structurally Sophisticated Octahedral Metal Complexes as Highly Selective Protein Kinase Inhibitors. **J Am Chem Soc** **133**, 5976-86.
- 10) Fedorov, O., Huber, K., Eisenreich, A., Filippakopoulos, P., King, O., Bullock, A.N., Szklarczyk, D., Jensen, L.J., Fabbro, D., Trappe, J., Rauch, U., Bracher, F., Knapp, S. (2011). Specific CLK inhibitors from a novel chemotype for regulation of alternative splicing. **Chem Biol** **18**, 67-76.

- 11) Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., Philpott, M., Munro, S., McKeown, M., Wang, Y., Christie, A.L., West, N., Cameron, M.J., Schwartz, B., Heightman, T.D., La Thangue, N., Kung, A.L., French, C.A., Wiest, O., Knapp, S., Bradner, J.E. (2010). Selective inhibition of BET bromodomains. **Nature** **468**, 1067-1073.
- 12) Johansson, C., Roos, A. K., Montano, S. J., Sengupta, R., Filippakopoulos, P., Guo, K., von Delft, F., Holmgren, A., Oppermann, U., Kavanagh, K. L. (2010). The Crystal structure of human GLRX5: iron sulphur cluster coordination, tetrameric assembly and monomer activity. **Biochem J** **433**, 303-311.
- 13) Filippakopoulos, P., Low, A., Sharpe, T.D., Uppenberg, J., Yao, S., Kuang, Z., Savitsky, P., Lewis, R.S., Nicholson, S.E., Norton, R.S., Bullock, A.N. (2010). Structural basis for Par-4 recognition by the SPRY domain- and SOCS box-containing proteins SPSB1, SPSB2, and SPSB4. **J Mol Biol** **401**, 389-402.
- 14) López-Ramos, M., Prudent, R., Moucadel, V., Sautel, C.F., Barette, C., Lafanechère, L., Mouawad, L., Grierson, D., Schmidt, F., Florent, J.C., Filippakopoulos, P., Bullock, A.N., Knapp, S., Reiser, J.B., Cochet, C. (2010) New potent dual inhibitors of CK2 and Pim kinases: discovery and structural insights. **FASEB J** **24**, 3171-3185.
- 15) Kwiatkowski, N., Jelluma, N., Filippakopoulos, P., Soundararajan, M., Manak, M.S., Kwon, M., Choi, H.G., Sim, T., Deveraux, Q.L., Rottmann, S., Pellman, P., Shah, J.V., Kops, G.J.P.L., Knapp, S., Gray, N.S. (2009). Small Molecule Kinase Inhibitors Provide Insight into Mps1 Cell Cycle Function. **Nat Chem Biol** **6**, 359-368.
- 16) Filippakopoulos, P., Muller, S., and Knapp, S. (2009). SH2 domains: modulators of nonreceptor tyrosine kinase activity. **Curr Opin Struct Biol** **19**, 643-649.
- 17) Eswaran, J.*, Patnaik, D.*, Filippakopoulos, P.*[†], Wang, F., Stein, R.L., Murray, J.W., Higgins, J.M., and Knapp, S. (2009). Structure and functional characterization of the atypical human kinase haspin. **Proc Natl Acad Sci U S A** **106**, 20198-20203.
[* equal contribution]
- 18) Barr, A.J., Ugochukwu, E., Lee, W.H., King, O.N., Filippakopoulos, P., Alfano, I., Savitsky, P., Burgess-Brown, N.A., Muller, S., and Knapp, S. (2009). Large-scale structural analysis of the classical human protein tyrosine phosphatome. **Cell** **136**, 352-363.
- 19) Akue-Gedu, R., Rossignol, E., Azzaro, S., Knapp, S., Filippakopoulos, P., Bullock, A.N., Bain, J., Cohen, P., Prudhomme, M., Anizon, F., et al. (2009). Synthesis, kinase inhibitory potencies, and in vitro antiproliferative evaluation of new Pim kinase inhibitors. **J Med Chem** **52**, 6369-6381.
- 20) Filippakopoulos, P., Kofler, M., Hantschel, O., Gish, G.D., Grebien, F., Salah, E., Neudecker, P., Kay, L.E., Turk, B.E., Superti-Furga, G., et al. (2008). Structural coupling of SH2-kinase domains links Fes and Abl substrate recognition and kinase activation. **Cell** **134**, 793-803.
- 21) Maksimoska, J., Williams, D.S., Atilla-Gokcumen, G.E., Smalley, K.S., Carroll, P.J., Webster, R.D., Filippakopoulos, P., Knapp, S., Herlyn, M., and Meggers, E. (2008). Similar biological activities of two isostructural ruthenium and osmium complexes. **Chemistry** **14**, 4816-4822.

- 22) Atilla-Gokcumen, G.E., Pagano, N., Streu, C., Maksimoska, J., Filippakopoulos, P., Knapp, S., and Meggers, E. (2008). Extremely tight binding of a ruthenium complex to glycogen synthase kinase 3. **Chembiochem** **9**, 2933-2936.
- 23) Pogacic, V., Bullock, A.N., Fedorov, O., Filippakopoulos, P., Gasser, C., Biondi, A., Meyer-Monard, S., Knapp, S., and Schwaller, J. (2007). Structural analysis identifies imidazo[1,2-b]Pyridazines as PIM kinase inhibitors with In vitro antileukemic activity. **Cancer Research** **67**, 6916-6924.
- 24) Eswaran, J., Lee, W.H., Debreczeni, J.E., Filippakopoulos, P., Turnbull, A., Fedorov, O., Deacon, S.W., Peterson, J.R., and Knapp, S. (2007). Crystal structures of the p21-activated kinases PAK4, PAK5, and PAK6 reveal catalytic domain plasticity of active group IIPAKs. **Structure** **15**, 201-213.
- 25) Bunkoczi, G., Salah, E., Filippakopoulos, P., Fedorov, O., Muller, S., Sobott, F., Parker, S.A., Zhang, H., Min, W., Turk, B.E., et al. (2007). Structural and functional characterization of the human protein kinase ASK1. **Structure** **15**, 1215-1226.

CV Module

This page is for CIHR use only. It will not be included in the evaluation of your application for funding.

Family Name Pawson	Given Name Anthony	Middle Initial(s) J
Have you previously applied to CIHR for funding? Previous family name used Previous given name used	Yes <input checked="" type="checkbox"/> No <input type="checkbox"/>	Title: Dr. <input checked="" type="checkbox"/> Mr. <input type="checkbox"/> Mrs. <input type="checkbox"/> Ms. <input type="checkbox"/> Prof. <input type="checkbox"/>

Courier Address (If different from mailing address) Samuel Lunenfeld Research Institute Mount Sinai Hospital 600 University Ave., Rm. 1084 Toronto, Ontario CANADA (M5G 1X5)	Temporary Address Start Date _____ End Date _____	Primary Affiliation Name Samuel Lunenfeld Research Institute of Mount Sinai Hospital Start Date 07/1985
		Primary Affiliation Address Samuel Lunenfeld Research Institute Mount Sinai Hospital 600 University Ave., Rm. 1084 Toronto, Ontario CANADA (M5G 1X5)

Contact numbers Phone Primary (416) 586-8262 office Secondary (416) 586-4800 #4524 lab Temporary Start Date _____ End Date _____	Fax Primary (416) 586-8869 Temporary Start Date _____ End Date _____	Electronic Addresses E-Mail pawson@lunenfeld.ca Web page address http://pawson.mshri.on.ca
--	--	---

Citizenship Canadian <input checked="" type="checkbox"/> Other <input type="checkbox"/> Other Country of Citizenship	Permanent Residence in Canada Permanent Resident <input type="checkbox"/> Date of permanent residency status DD/MM/YYYY Have you applied for permanent residency? Yes <input type="checkbox"/> No <input checked="" type="checkbox"/>
--	---

Correspondence Language English <input checked="" type="checkbox"/> French <input type="checkbox"/>	Language Read Write Speak Understand English (Yes or No) YES YES YES YES French (Yes or No) YES NO NO YES Other Languages: _____
Gender Male <input checked="" type="checkbox"/> Female <input type="checkbox"/> Date of Birth (DD/MM/YYYY) 18/10/1952	

Expertise

List up to ten (10) key words that best describe your expertise in research, instruments and technique.

signal transduction	protein structural anal.
cell transformation	biochemistry
protein kinases	axon guidance
protein-protein interact.	genetics
interaction domains	developmental biology

Indicate and rank the disciplines that best correspond to your research interests. No additional pages may be added.

Discipline			Sub Discipline	
Rank	Code	Description	Code	Description
1.	11	MOLECULAR AND CELLULAR BIOLOGY	1130	Second Messengers/Signal Transduction
2.	9	BIOCHEMISTRY	1469	Proteomics
3.	37	GENETICS	482	Genetic Engineering
4.	13	CANCER/ONCOLOGY	832	Oncogenes
5.	65	NEUROSCIENCES	797	Neural Growth/Development/Degeneration
6.	64	MICROBIOLOGY, VIROLOGY, AND PARASITOLOGY	668	Viral
7.	55	IMMUNOLOGY	592	Immunobiology
8.	52	HEMATOLOGY	123	Developmental Cell Biology/Differentiation
9.				
10.				
11.				
12.				
13.				
14.				
15.				

Academic Background - One additional page may be added

Indicate all university degrees obtained and those in progress (where applicable) starting with the most recent. If you hold a co-degree from more than one institution (e.g. under the Soutien aux cotutelles de these de doctorat agreement between Quebec and France) enter each institution separately. Do not enter honorary degrees here, they should be listed in the Distinctions section.

Also indicate research training, such as postdoctoral or fellowship training. Trainees only: also list undergraduate and graduate research training experience.

Degree Type	Degree Name and Specialty	Institution/Organization and Country	Supervisor name	Start date (MM/YYYY)	Date received or expected (MM/YYYY)
Doctorate (PhD)	Doctor of Philosophy Molecular Biology	King's College, London University UNITED KINGDOM	Dr. Alan Smith	07/1973	09/1976
Bachelor's	Bachelor Biochemistry	University of Cambridge UNITED KINGDOM	Dr. Tim Hunt	07/1970	07/1973

Work Experience

Starting with the most recent, indicate your current position, where applicable, and other academic and non-academic position(s) since the beginning of your university studies. For your current positions leave the end date blank. Additional pages will be accepted.

Position	Institution/Organization and Country	Department/Division and Faculty/School	Start Date (MM/YYYY)	End Date (MM/YYYY)
Distinguished Scientist	Samuel Lunenfeld Research Institute of Mount Sinai Hospital CANADA	Centre for Systems Biology	01/2006	
Senior Fellow	Massey College, University of Toronto CANADA	Research	01/2003	
Full Professor	University of Toronto CANADA	Medical Genetics & Microbiology	07/1989	
Senior Scientist	Samuel Lunenfeld Research Institute of Mount Sinai Hospital CANADA	Programme in Molecular Biology and Cancer	07/1985	
Director	Samuel Lunenfeld Research Institute of Mount Sinai Hospital CANADA		06/2002	12/2005
Head	Samuel Lunenfeld Research Institute of Mount Sinai Hospital CANADA	Programme in Molecular Biology and Cancer	07/1994	12/2005
Acting Director	Samuel Lunenfeld Research Institute of Mount Sinai Hospital CANADA		06/2000	06/2002
Associate Professor	University of Toronto CANADA	Medical Genetics & Microbiology	07/1985	07/1989
Assistant Professor	University of British Columbia CANADA	Microbiology	03/1981	03/1985
Postdoctoral Fellow	University of California, Berkeley UNITED STATES	Zoology	07/1977	07/1980

Work Experience

Starting with the most recent, indicate your current position, where applicable, and other academic and non-academic position(s) since the beginning of your university studies. For your current positions leave the end date blank. Additional pages will be accepted.

Position	Institution/Organization and Country	Department/Division and Faculty/School	Start Date (MM/YYYY)	End Date (MM/YYYY)
Postdoctoral Fellow	University of California, Berkeley UNITED STATES	Molecular Biology	07/1976	07/1977

Distinctions / Awards / Credentials

Starting with the most recent, indicate any recognitions received, including awards, fellowships, scholarships, licenses, qualifications, professional designation or credentials. Do not include Academic Appointments here, as they are detailed under Work Experience. Maximum 20 entries.

Name/Title and Type	Institution/Organization and Country	Effective Date (MM/YYYY)	End Date (MM/YYYY)	Specialty	Total Amount
Award for Outstanding Achievements in Cancer Res. Research award	Canadian Cancer Research Alliance CANADA	11/2011			
Inaugural Honorary Medal of the STS Distinction	Signal Transduction Society GERMANY	11/2010			
Honorary Doctorate of Science Degree Distinction	McMaster University CANADA	06/2010			
Doctor of Medicine Distinction	Karolinska Institutet SWEDEN	05/2010			
Named as "Nation builder of the decade" Distinction	The Globe & Mail newspaper CANADA	01/2010			
Fritz Lipmann Award Lecture Distinction	Gesellschaft fur Biochemie und Molekularbiologie GERMANY	09/2009			
Honorary Doctorate of Science Degree Distinction	University of Alberta CANADA	06/2008			
The Kyoto Prize in Basic Sciences Distinction	Inamori Foundation JAPAN	06/2008			
Diamond Jubilee Award. Distinction	The National Cancer Institute of Canada CANADA	11/2007			
Premier's Summit Award for Medical Research Research award	Government of Ontario CANADA	04/2007	04/2012		\$2,500,000

Distinctions / Awards / Credentials

Starting with the most recent, indicate any recognitions received, including awards, fellowships, scholarships, licenses, qualifications, professional designation or credentials. Do not include Academic Appointments here, as they are detailed under Work Experience. Maximum 20 entries.

Name/Title and Type	Institution/Organization and Country	Effective Date (MM/YYYY)	End Date (MM/YYYY)	Specialty	Total Amount
The Jubilee Lecture Award Distinction	The Biochemical Society, U.K. UNITED KINGDOM	01/2007			
Howard Taylor Ricketts Award Distinction	University of Chicago UNITED STATES	01/2007			
Appointed to The Order of Companions of Honour Distinction	Her Majesty Queen Elizabeth II UNITED KINGDOM	06/2006			
Daniel Nathans Memorial Award Distinction	Van Andel Research Institute UNITED STATES	12/2005			
Honorary Member, Japanese Biochemical Society Distinction	Japanese Biochemical Society JAPAN	10/2005			
Royal Medal Distinction	The Royal Society of London UNITED KINGDOM	06/2005			
Wolf Prize in Medicine Distinction	Wolf Foundation Council of Israel ISRAEL	05/2005			
Special Achievement Award Distinction	University of Miami UNITED STATES	02/2005			
Louisa Gross Horwitz Prize Distinction	Columbia University UNITED STATES	11/2004			
Ernst W. Bertner Memorial Award Distinction	MD Anderson Cancer Center UNITED STATES	10/2004			

Patents and Intellectual Property Rights

Record the total numbers of patents / copyrights in the following table.

OBTAINED			APPLICATIONS UNDER PROCESS			TOTAL PATENTS AND INTELLECTUAL PROPERTY RIGHTS
Total individual	Total collective	Sub-total	Total individual	Total collective	Sub-total	
1	9	10	0	0	0	10

PUBLICATIONS AND PRESENTATIONS

Give the number of publications and presentations in the course of your career. Detailed information should be attached as specified in the "Contributions - details" section.

Publications	Refereed Articles	Books and Monographs	Proceedings / Book Chapters / Contributions to a collective work	Abstracts / Notes	TOTALS
Already Published	381	0	36	559	976
Accepted or in the Press	2	0	0	0	2
					978

Invited presentations	587
-----------------------	-----

LITERARY AND ARTISTIC WORKS

Provide the number of literary and artistic works created in the course of your career. Detailed information should be attached as specified in the "Contributions - details" section.

IN CIRCULATION			IN PROGRESS			TOTAL LITERARY AND ARTISTIC WORKS
Total individual	Total collective	Sub-total	Total individual	Total collective	Sub-total	
0	0	0	0	0	0	0

Supervisory Experience: To be completed by applicants requesting research trainees as part of their budget, salary support candidates and proposed supervisors of trainees.

Indicate the number of graduate students and postdoctoral fellows that you currently supervise or co-supervise. CIHR defines supervisory experience as the formal supervision or co-supervision of trainees. Enter zero (0) if not applicable.

Master 0Doctoral 3Post-Doctoral 20

Complete this form by listing the trainees that you have supervised/co-supervised (and are currently supervising/co-supervising) within the last five (5) years. Additional pages may be added if necessary.

* Flag those where you were/are the Primary Supervisor.

*	Name of Student	Program Type	Dates		Degree received or expected	Year Degree Rec'd (YYYY)	Research Project (Short title)	Current position and Institution
			Support Period From (MM/YY)	To (MM/YYYY)				
	Kinght, James	Postdoctoral Fellow, PhD	10/2011		Doctorate (PhD)	2011	Identifying substrates for the GCKIII kinases	
*	van Eekelen, Mark	Postdoctoral Fellow, PhD	06/2011		Doctorate (PhD)	2011	Eph-ephrin signaling in cancer	
*	Lorenzen, Kristina	Postdoctoral Fellow, PhD	04/2011		Doctorate (PhD)	2008	Structural characterization of protein complexes using mass spectrometry	
*	Stacey, Melissa	Postdoctoral Fellow, PhD	01/2011		Doctorate (PhD)	2010	Phosphoprotein profile of Eph and ephrin expressing cells during tissue patterning	
*	Zhao, Yanling	Graduate Student	01/2011				Nuclear remodeling complexes, piRNA protein complexes	
*	Tian, Ruijun	Postdoctoral Fellow, PhD	10/2010		Doctorate (PhD)	2008	Mass spectrometry analysis of arginine methylation	
*	Wagner, Melany	Postdoctoral Fellow, PhD	10/2010		Doctorate (PhD)	2010	Signaling via the Shb scaffold	
	Lambert, Jean-Philippe	Postdoctoral Fellow, PhD	10/2010		Doctorate (PhD)	2010	Proteomic analysis of bromodomains	

Supervisory Experience: To be completed by applicants requesting research trainees as part of their budget, salary support candidates and proposed supervisors of trainees.

Indicate the number of graduate students and postdoctoral fellows that you currently supervise or co-supervise. CIHR defines supervisory experience as the formal supervision or co-supervision of trainees. Enter zero (0) if not applicable.

Master 0Doctoral 3Post-Doctoral 20

Complete this form by listing the trainees that you have supervised/co-supervised (and are currently supervising/co-supervising) within the last five (5) years. Additional pages may be added if necessary.

* Flag those where you were/are the Primary Supervisor.

*	Name of Student	Program Type	Dates		Degree received or expected	Year Degree Rec'd (YYYY)	Research Project (Short title)	Current position and Institution
			Support Period From (MM/YY)	To (MM/YYYY)				
*	Helbig, Andreas	Postdoctoral Fellow, PhD	09/2010		Doctorate (PhD)	2010	Phosphoproteomic analysis of complex systems	
*	Petsalakis, Evangelia	Postdoctoral Fellow, PhD	03/2010		Doctorate (PhD)	2009	Systematic discovery and analysis of pathways controlled by the Rho GTPases Rho, Rac1 and Cdc42	
*	Liu, Bernard	Postdoctoral Fellow, PhD	07/2009		Doctorate (PhD)	2009	Cellular rewiring of the polarity network using engineered PDZ domains	
*	Nott, Timothy	Postdoctoral Fellow, PhD	06/2009		Doctorate (PhD)	2009	Structure of Ras Gap	
*	Soliman, Mohamed	Graduate Student	04/2009				Re-wiring cell signaling using PH and death effector domains	
	Guettler, Sebastian	Postdoctoral Fellow, PhD	04/2008		Doctorate (PhD)	2007	Structural and functional analysis of the Hpo-Mats-Wts tumour suppressor network in human cells	
*	Kofler, Michael	Postdoctoral Fellow, PhD	08/2007		Doctorate (PhD)	2006	Regulation of the Fes tyrosine kinase by intra- and intermolecular interactions	
*	Findlay, Greg	Postdoctoral Fellow, PhD	05/2007		Doctorate (PhD)	2007	Investigation of Grb2 signaling networks	

Supervisory Experience: To be completed by applicants requesting research trainees as part of their budget, salary support candidates and proposed supervisors of trainees.

Indicate the number of graduate students and postdoctoral fellows that you currently supervise or co-supervise. CIHR defines supervisory experience as the formal supervision or co-supervision of trainees. Enter zero (0) if not applicable.

Master 0Doctoral 3Post-Doctoral 20

Complete this form by listing the trainees that you have supervised/co-supervised (and are currently supervising/co-supervising) within the last five (5) years. Additional pages may be added if necessary.

* Flag those where you were/are the Primary Supervisor.

*	Name of Student	Program Type	Dates		Degree received or expected	Year Degree Rec'd (YYYY)	Research Project (Short title)	Current position and Institution
			Support Period From (MM/YY)	To (MM/YYYY)				
*	So, Jonathan	Graduate Student	12/2006		Doctorate (PhD)		Synthetic biology, intracellular signaling	
*	Bagshaw, Richard	Postdoctoral Fellow, PhD	10/2006		Doctorate (PhD)	2006	Endocytic recycling in establishment of cell polarity	
*	Louria Hayon, Igal	Postdoctoral Fellow, PhD	06/2006		Doctorate (PhD)	2006	The role of Lnk family in adults and cancer stem cells	
*	Fellouse, Frederic	Postdoctoral Fellow, PhD	01/2006		Doctorate (PhD)	2006	Intrabodies as a tool for cell signaling	
*	Chen, Chen	Postdoctoral Fellow, PhD	09/2005		Doctorate (PhD)	2005	Nck function	
*	Zheng, Yong	Postdoctoral Fellow, PhD	09/2005		Doctorate (PhD)	2005	Phosphoproteomic analysis of growth factor receptor signaling	
*	Jing, Jin	Postdoctoral Fellow, PhD	03/2003		Doctorate (PhD)	2004	Proteomic analysis of intracellular signalling network	
*	Bisson, Nicolas	Postdoctoral Fellow, PhD	07/2007	12/2011	Doctorate (PhD)	2007	Eph and Nck signaling in cancer and development	Investigator, Laval University

Supervisory Experience: To be completed by applicants requesting research trainees as part of their budget, salary support candidates and proposed supervisors of trainees.

Indicate the number of graduate students and postdoctoral fellows that you currently supervise or co-supervise. CIHR defines supervisory experience as the formal supervision or co-supervision of trainees. Enter zero (0) if not applicable.

Master 0Doctoral 3Post-Doctoral 20

Complete this form by listing the trainees that you have supervised/co-supervised (and are currently supervising/co-supervising) within the last five (5) years. Additional pages may be added if necessary.

* Flag those where you were/are the Primary Supervisor.

*	Name of Student	Program Type	Dates		Degree received or expected	Year Degree Rec'd (YYYY)	Research Project (Short title)	Current position and Institution
			Support Period From (MM/YY)	To (MM/YYYY)				
*	Gray, Elizabeth	Graduate Student	12/2006	12/2011	Doctorate (PhD)	2011	Phospho-proteomics	Technology A.O., CCRM
*	Chris Tan	Graduate Student	01/2007	08/2011	Doctorate (PhD)		Evolution of phospho-regulation of cellular activities	Postdoctoral Fellow, Ce-M-M, Vienna
*	Rocks, Oliver	Postdoctoral Fellow, PhD	05/2007	03/2011	Doctorate (PhD)	2005	Spatio-temporal control of polarity protein activity	Group Leader, Max-Delbrück-Centrum
*	Holdorf, Amy	Postdoctoral Fellow, PhD	03/2002	12/2010	Doctorate (PhD)	2002	Adaptor proteins	Scientist, Broad Inst. of MIT
*	Vanderlaan, Rachel	Graduate Student	06/2006	10/2010	Doctorate (PhD)	2010	The role of ShcA in cardiovascular development and disease	Medical School, Univ. of Toronto
*	Pike, Kelly	Postdoctoral Fellow, PhD	01/2006	09/2010	Doctorate (PhD)	2005	Characterization of BLNK associated protein complexes: in resting and activated B cells	Research Associate, McGill University
*	Park, Jin Gyo	Postdoctoral Fellow, PhD	01/2005	05/2010	Doctorate (PhD)	2003	High-throughput analysis of protein domain functions	Scientist, Arizona State Univ.
*	Wegmeyer, Heike	Postdoctoral Fellow, PhD	10/2007	03/2010	Doctorate (PhD)	2007	Role of Shc scaffold proteins in regulating neuronal differentiation and survival in the mouse nervous system	Scientist, Roche Co., Germany

Supervisory Experience: To be completed by applicants requesting research trainees as part of their budget, salary support candidates and proposed supervisors of trainees.

Indicate the number of graduate students and postdoctoral fellows that you currently supervise or co-supervise. CIHR defines supervisory experience as the formal supervision or co-supervision of trainees. Enter zero (0) if not applicable.

Master 0Doctoral 3Post-Doctoral 20

Complete this form by listing the trainees that you have supervised/co-supervised (and are currently supervising/co-supervising) within the last five (5) years. Additional pages may be added if necessary.

* Flag those where you were/are the Primary Supervisor.

*	Name of Student	Program Type	Dates		Degree received or expected	Year Degree Rec'd (YYYY)	Research Project (Short title)	Current position and Institution
			Support Period From (MM/YY)	To (MM/YYYY)				
*	Jorgensen, Claus	Postdoctoral Fellow, PhD	05/2005	03/2010	Doctorate (PhD)	2005	Eph-ephrin signalling in normal and malignant cells	Team Leader, ICR, London, UK
*	Sherman, Andrew	Graduate Student	01/2006	01/2010	Master's		High throughput analysis of cell migration pathways	Thermo Fisher Co.
*	Scott, Rizaldy	Postdoctoral Fellow, PhD	09/2003	01/2010	Doctorate (PhD)	2002	Apoptosis-associated tyrosine kinase (AATYK) in cell cycle regulation and neuronal development	Research Associate, SLRI
*	Cohn, Jason	Graduate Student	01/2006	12/2009	Doctorate (PhD)		Characterization of SNAG1, a novel sorting nexin identified by gene trapping	
*	Wiggin, Giselle	Postdoctoral Fellow, PhD	09/2003	11/2009	Doctorate (PhD)	2003	The role of Lgl in mammalian cell polarity, cell proliferation and cancer	Scientist, Biotech Co., U.K.
Dar, Nisa	Graduate Student		01/2009	08/2009			Evolution of Wnt signalling pathway in eukaryotes	Graduate Student, Med. Gen., UofT
*	Hardy, William	Graduate Student	04/1999	04/2009	Doctorate (PhD)	2007	ShcA function in the mouse	Postdoctoral Fellow, Univ. of Mass.
*	Linding, Rune	Postdoctoral Fellow, PhD	01/2005	01/2009	Doctorate (PhD)	2004	Synthetic Biology	Team Leader, ICR, UK

Supervisory Experience: To be completed by applicants requesting research trainees as part of their budget, salary support candidates and proposed supervisors of trainees.

Indicate the number of graduate students and postdoctoral fellows that you currently supervise or co-supervise. CIHR defines supervisory experience as the formal supervision or co-supervision of trainees. Enter zero (0) if not applicable.

Master 0Doctoral 3Post-Doctoral 20

Complete this form by listing the trainees that you have supervised/co-supervised (and are currently supervising/co-supervising) within the last five (5) years. Additional pages may be added if necessary.

* Flag those where you were/are the Primary Supervisor.

*	Name of Student	Program Type	Dates		Degree received or expected	Year Degree Rec'd (YYYY)	Research Project (Short title)	Current position and Institution
			Support Period From (MM/YY)	To (MM/YYYY)				
*	Engelhardt, Mark	Postdoctoral Fellow, PhD	06/2006	12/2008	Doctorate (PhD)	2006	Synthetic signalling pathways	
*	Blasutig, Ivan	Graduate Student	12/2000	05/2008	Doctorate (PhD)	2008	Cytoskeletal regulation	Clinical Chemistry, Univ. of Toronto
*	Traweger, Andreas M.	Postdoctoral Fellow, PhD	11/2004	02/2008	Doctorate (PhD)	2004	Regulation of cell polarity	Scientist, Baxter AG, Austria
*	Lim, Caesar	Graduate Student	10/2003	12/2007	Master's	2007	Characterization of the KSHV protein K15	Medical School
*	Warner, Neil	Postdoctoral Fellow, PhD	09/2005	10/2007	Doctorate (PhD)	2005	Elucidation of the biological function of SAM	Postdoctoral Fellow, Univ. of Michigan
*	Smith, Matthew	Graduate Student	12/2000	08/2007	Doctorate (PhD)	2007	Characterization of FF domains	Postdoctoral Fellow, MaRS, Toronto

Funds REQUESTED

List all sources of support applied for (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount requested (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)

Funds CURRENTLY HELD

List all sources of support currently held (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount awarded (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal Mapping biological landscapes in four dimensions with ultra-high throughput and ultra-high-resolution cell biology		
Funding Source Ministry of Research and Innovation (MRI) (Ontario)		Program Name Research Fund
Principal Applicant / Project Leader Durocher, D.		Your Role Co-Applicant
Total Amount (CAN\$) \$214,447	Support Period From (MM/YYYY) 03/2012	To (MM/YYYY) 03/2016
Title of Proposal Chemical probes for interrogating protein-protein interactions in disease states		
Funding Source Ministry of Research and Innovation (MRI) (Ontario)		Program Name Research Grant
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$671,448	Support Period From (MM/YYYY) 10/2011	To (MM/YYYY) 09/2015
Title of Proposal Scaffolds and receptors in angiogenesis and tumorigenesis.		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Terry Fox Team Grant
Principal Applicant / Project Leader Nagy, Andras		Your Role Co-Applicant
Total Amount (CAN\$) \$1,012,110	Support Period From (MM/YYYY) 07/2010	To (MM/YYYY) 06/2015
Title of Proposal Synthetic antibody program: commercial reagents and novel therapy.		
Funding Source Genome Canada		Program Name 2010 Large-Scale Special Research Project
Principal Applicant / Project Leader Sidhu, Sachdev		Your Role Co-Applicant
Total Amount (CAN\$) \$393,555	Support Period From (MM/YYYY) 07/2011	To (MM/YYYY) 06/2014

Funds CURRENTLY HELD

List all sources of support currently held (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount awarded (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal Understanding the assembly and function of dynamic signaling networks.		
Funding Source Ministry of Research and Innovation (MRI) (Ontario)		Program Name Research Grant
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$2,092,800	Support Period From (MM/YYYY) 07/2010	To (MM/YYYY) 06/2014
Title of Proposal Bidirectional cell signaling networks that control cell movement and segregation. (Renewal of CIHR grant #MOP-13466)		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Operating Grant
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$878,400	Support Period From (MM/YYYY) 03/2009	To (MM/YYYY) 04/2014
Title of Proposal Protein interaction domains and adaptor proteins in cell signaling (renewal of CIHR grant #6849)		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Operating grant
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$1,302,700	Support Period From (MM/YYYY) 04/2009	To (MM/YYYY) 03/2014
Title of Proposal Signaling pathways controlling cell shape, movement and polarity in the kidney		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Operating Grant
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$677,875	Support Period From (MM/YYYY) 10/2007	To (MM/YYYY) 09/2012

Funds CURRENTLY HELD

List all sources of support currently held (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount awarded (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal The molecular dynamics and imaging of Eph receptor-guided cell positioning in tissue assembly		
Funding Source International Human Frontier Science Program Organization		Program Name Operating Grant
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$262,500	Support Period From (MM/YYYY) 09/2009	To (MM/YYYY) 08/2012
Title of Proposal Regulatory protein networks in cell polarity.		
Funding Source National Cancer Institute of Canada (NCIC)		Program Name Operating Grant
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$705,000	Support Period From (MM/YYYY) 07/2007	To (MM/YYYY) 06/2012
Title of Proposal Core Proteomics Laboratory		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Research Resource Grant
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$497,150	Support Period From (MM/YYYY) 04/2007	To (MM/YYYY) 03/2012
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)

Funds HELD IN THE LAST FIVE YEARS

List all sources of support held in the last five years (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount awarded (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal TFRI-OICR Selective Therapies Program		
Funding Source Ontario Institute for Cancer Research (OICR)		Program Name TFRI-OICR Selective Therapies Program
Principal Applicant / Project Leader Rottapel, Robert		Your Role Co-Applicant
Total Amount (CAN\$) \$305,087	Support Period From (MM/YYYY) 07/2009	To (MM/YYYY) 09/2011
Title of Proposal Ontario Proteomic Methods Centre (OPMC)		
Funding Source Ontario Research Fund (ORF)		Program Name Operating Grant
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$1,493,640	Support Period From (MM/YYYY) 10/2006	To (MM/YYYY) 03/2011
Title of Proposal Role of Nck adaptors in kidney and lower urinary tract development		
Funding Source Kidney Foundation of Canada (KFC)		Program Name Biomedical Research Grant
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$99,818	Support Period From (MM/YYYY) 07/2008	To (MM/YYYY) 06/2010
Title of Proposal The Dynactome: Mapping Spatio-Temporal Dynamic Systems in Humans		
Funding Source Ontario Genomics Institute (OGI)		Program Name Operating Grant
Principal Applicant / Project Leader		Your Role Principal Applicant
Total Amount (CAN\$) \$4,750,377	Support Period From (MM/YYYY) 01/2006	To (MM/YYYY) 06/2010

Funds HELD IN THE LAST FIVE YEARS

List all sources of support held in the last five years (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount awarded (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal Reciprocal cell signaling in angiogenesis and tumour formation.		
Funding Source National Cancer Institute of Canada (NCIC)		Program Name NCIC Program Project Grant
Principal Applicant / Project Leader Nagy, A.		Your Role Co-Applicant
Total Amount (CAN\$) \$1,211,420	Support Period From (MM/YYYY) 07/2005	To (MM/YYYY) 06/2010
Title of Proposal Signaling from mammalian guidance receptors.		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Operating Grant
Principal Applicant / Project Leader Pawson, Tony		Your Role Principal Applicant
Total Amount (CAN\$) \$1,148,845	Support Period From (MM/YYYY) 07/2004	To (MM/YYYY) 06/2009
Title of Proposal Protein interaction domains in cell signaling.		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Operating Grant
Principal Applicant / Project Leader Pawson, Tony		Your Role Principal Applicant
Total Amount (CAN\$) \$1,255,715	Support Period From (MM/YYYY) 04/2004	To (MM/YYYY) 03/2009
Title of Proposal Core Proteomics Laboratory		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Multi-User Equipment / Maintenance
Principal Applicant / Project Leader Pawson, Tony		Your Role Principal Applicant
Total Amount (CAN\$) \$165,000	Support Period From (MM/YYYY) 04/2004	To (MM/YYYY) 03/2007

Funds HELD IN THE LAST FIVE YEARS

List all sources of support held in the last five years (including CIHR) as an applicant or as a co-applicant. Include the principal applicant's name, title of the proposal, funding source, program name, total amount awarded (in Canadian dollars) and the period of the support. Indicate your role in the funding (principal applicant/project leader or co-applicant).

Title of Proposal Comparative and functional proteomics		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Collaborative Genomics Special Projects
Principal Applicant / Project Leader Pawson, Tony		Your Role Principal Applicant
Total Amount (CAN\$) \$1,000,000	Support Period From (MM/YYYY) 10/2002	To (MM/YYYY) 03/2007
Title of Proposal Regulation of signaling pathways by pathogenic viral and bacterial proteins.		
Funding Source Canadian Institutes of Health Research (CIHR)		Program Name Operating Grant
Principal Applicant / Project Leader Pawson, Tony		Your Role Principal Applicant
Total Amount (CAN\$) \$746,675	Support Period From (MM/YYYY) 04/2002	To (MM/YYYY) 03/2007
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)
Title of Proposal		
Funding Source		Program Name
Principal Applicant / Project Leader		Your Role
Total Amount (CAN\$)	Support Period From (MM/YYYY)	To (MM/YYYY)

Attachment Instructions

How to prepare and format all attachments:

Most Significant Contributions, Activities/Contributions, Interruptions/Delays, Patents/Copyrights (Part 2), and Publications (Part 2) details shall be contained in a CV attachment. Note: If you are using ResearchNet, you will need to provide each section identified as a separate PDF file.

The following format should be adhered to for this attachment.

- 8.5" X 11" (21.5 X 28.0 cm) white single-sided paper.
- Margins of $\frac{3}{4}$ " (2 cm).
- Minimum font size 12 point or 10 characters per inch.
- Six lines per inch, single-spaced, with no condensed type or spacing.
- Number pages consecutively after CV (If, for example, the print-out of the CV ends on page 8, the attachment would begin with page 9.).
- Each page header must contain the name and the sub-section header, e. g., Most Significant Contributions.

Most Significant Contributions

This section applies only to researchers, not to students. Identify a **maximum of five (5) contributions, with a maximum length of one page**, that best highlight your contribution or activities to research, defining the impact and relevance of each. (A contribution is understood to be a publication, literary or artistic work, conference, patent or copyright, contract or creative activity, commission, etc.) Your complete description may include the organization; position or activity type and description; from and to dates; and the basis on which this contribution is significant (i.e. relevance, target community and impact).

Activities / Contributions

The activities and contributions defined in this section should include both academic and non-academic achievements, and their impacts. **Limit the list to one page.**

Interruption(s) / Delays

Identify any administrative responsibilities, family or health reasons, or any other factors that might have delayed or interrupted any of the following: academia, career, scientific research, other research, dissemination of results, training, etc. Common examples of an interruption/delay might be a bereavement period following the death of a loved one, maternity/parental leave, or relocation of your research environment. **Limit the list to one page.**

Descriptions might include the start and end dates, the impact areas, and the reason(s) or a brief explanation of the absence.

Patents and Intellectual Property Rights

This section should include detail for patents and intellectual property rights for technology transfer, products, and services. Do not include Publications in this section. **Limit the list to one page.**

Descriptions for patents/intellectual property rights might include the title, patent/intellectual property rights number and date, country(ies) of issue, as well as the relevance or impact of this item and any inventor name(s) which pertain to it.

Publications List

List your most important publications and other research contributions over the past five years, according to the categories below. This is not necessarily a complete list, and is only intended to provide guidance. Categories can be added as needed. Use only items pertinent to the application. **There is no limit to the number of pages you can use.**

For Training or Salary Support Awards Candidates

- Candidates for training awards or New Investigator awards should list all publications, not just those of the last five years.
- All candidates for training or salary support awards must, for each multi-authored publication, define their role in the publication and indicate their percent contribution to the team effort.
- Candidates for training awards, with or without publications, are invited to comment on environmental factors that affected their capacity to publish.
- Candidates for salary support awards should, for multi-authored publications, underline the names of trainees whose work they supervised.

For Proposed Supervisors of Training Award Applicants

- Attach a maximum of two pages listing the titles and contributions over the past 5 years that will serve the application best.

Sadowski, I., J.C. Stone, and T. Pawson. 1986. A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. Mol. Cell. Biol. 6:4396-4408.

This paper provided the initial identification of the Src homology 2 (SH2) domain, and proposed that this element functions as a protein module involved in regulating the activity of tyrosine kinases and directing their interactions with downstream targets.

Anderson, A., C.A. Koch, L. Grey, C. Ellis, M.F. Moran, and T. Pawson. 1990. Binding of SH2 domains of phospholipase C(1, GAP and Src to activated growth factor receptors. Science 250:979-982.

This paper established that SH2 domains bind specifically to autophosphorylated receptor tyrosine kinases, and suggested that SH2 domains physically couple tyrosine kinases to their intracellular targets.

Olivier, J.P., T. Raabe, M. Henkemeyer, B. Dickson, G. Mbamalu, B. Margolis, J. Schlessinger, E. Hafen and T. Pawson. 1993. A Drosophila SH2/SH3 adaptor protein is implicated in coupling the Sevenless tyrosine kinase to an activator of Ras guanine nucleotide exchange, Sos. Cell 73:179-191.

This paper established using a combination of genetic and biochemical analysis that an SH2/SH3 adaptor couples tyrosine kinases to Sos, and thus to the Ras pathway.

Marengere, L.E.M., Z. Songyang, G.D. Gish, M.D. Schaller, T. Parsons, M.J. Stern, L.C. Cantley, and T. Pawson. 1994. SH2 domain specificity and activity modified by a single residue. Nature 369:502-505.

This paper demonstrated that SH2 domains specificity plays a key role in regulating signaling pathways in an intact organism.

Holland, S.J., N.W. Gale, G. Mbamalu, G.D. Yancopoulos, M. Henkemeyer and T. Pawson. 1996. Bi-directional signalling through the Eph family receptor Nuk and its transmembrane ligands. Nature. 383:722-725.

This paper suggested that Eph receptors and their transmembrane ligands might mediate a two-way signal between receptor - and ligand-expressing cells.

Activities/Contributions (current)

- Roche Oncology. Scientific Advisory Board.
- Genentech Scientific Review Board.
- Keystone Symposia Board of Directors.
- Protein Data Bank (PDB) Scientific Advisory Board.
- Co-President. Protein Modules Consortium.
- Structural Genomics Consortium. Scientific Advisory Board.
- Faculty 1000. Head of Faculty (Cell Biology).
- Cell Migration Consortium (NIGMS Large Scale Collaborative Grant). External Advisory Committee.
- MGH Cancer Center, Scientific Advisory Board.
- Gairdner Foundation Medical Advisory Board.
- A*STAR Biomedical Science, Singapore. International Advisory Board.
- Ontario Institute for Cancer Research, Scientific Advisory Board.
- Friedrich Miescher Institute, Basel, Scientific Advisory Board.
- Meeting on signal transduction, 2011 (Banff). Co-organizer.
- The evolution of protein phosphorylation. Keystone meeting 2011. Co-organizer.

Editorial Boards (past and present) – Trends in Genetics; Oncogene; Molecular and Cellular Biology; Chemistry and Biology; Cell Growth and Differentiation; Current Opinion in Cell Biology; Developmental Cell; Molecular Biology of the Cell; European Journal of Biochemistry; EMBO Journal, PNAS.

Distinguished/Keynote Lectures

- 2006 Euroconference on Protein Kinase Inhibitors. Paris, France. Keynote Speaker.
- 2006 Karolinska Research Lecture, Nobel Forum, Stockholm, Sweden.
- 2007 Distinguished Lecture, Beatson Research Institute, Glasgow, UK.
- 2007 Hadassah Horn Lecture, Weizmann Institute of Science, Israel.
- 2007 Special Conference on Signal Transduction. London Research Institute, Cancer Research UK. London, UK. Keynote Speaker
- 2007 Jean Shanks Lecture, Academy of Medical Sciences (UK), London, UK.
- 2007 International Conference on Anchored cAMP Signaling Mechanisms. Vollum Institute. Portland OR, USA. McAdams Wright Ragen Keynote Lecture.
- 2007 9th DGZ Young Scientists Meeting on “Signaling cascades in Development and Disease”. Munster, Germany. Keynote Talk.
- 2007 National Cancer Institute of Canada 60th Anniversary Conference. Toronto, ON.
- 2008 “What is Life?” Lecture Series. Nobel Forum, Karolinska Institute, Stockholm.
- 2008 B.C 2008 Cell Biology Retreat, Loon Lake, B.C. Keynote Speaker.
- 2008 HUPO 2008, Amsterdam. Opening Plenary Lecture.
- 2008 Network Biology, Hinxton, UK. Keynote Speaker.
- 2008 The 2008 Kyoto Prize Commemorative Lecture: Basic Sciences “Thinking about how living things work”, Kyoto, Japan.
- 2008 MD Anderson Cancer Center, Houston, TX. Blaffer Distinguished Scientist Lecture.
- 2009 FEBS Practical Course on Protein Interaction Modules. Split, Croatia. Keynote talk.
- 2009 Applied Medical Oncology Retreat. Toronto. Keynote talk.
- 2009 State-of-the Art Keynote Lecture, ASN Renal Week, San Diego, CA
- 2010 Harold Feinberg Memorial Lecture. Dept. Pharmacology, University of Illinois at Chicago.
- 2011 Director’s Lecture. Walter+Eliza Hall, Institute of Medical Research. Melbourne, Australia.

INTERRUPTIONS/DELAYS

Pawson, Anthony James

NONE TO REPORT

Patents and Copyrights

Pawson, Anthony James

- 1) Title:** "Method of Assaying for a Substance that Affects an SH2 Phosphorylated Ligand Regulatory System"

Inventor: Tony Pawson

- 2) Title:** "Novel Neural Receptor Tyrosine Kinase"

Inventors: Anthony Pawson, Mark Henkemeyer, Kenneth Letwin

- 3) Title:** "Method for Assaying for Substances Which Affect BCR-ABL Mediated Transformation"

Inventors: Lori Puil, Anthony Pawson, Ralph Arlinghaus, Gerald Gish, and Jiazin Liu

- 4) Title:** "Peptide Inhibitors of a Phosphotyrosine-binding Domain Containing Protein"

Inventors: Peter van der Geer, Sandra Wiley, Gerald D. Gish and Tony Pawson

- 5) Title:** "Method of Activating a Novel Regulatory Pathway"

Inventors: Anthony Pawson and Mark Henkemeyer

- 6) Title:** "Methods for Identifying Compounds that Inhibit or Enhanced Activation of a Transmembrane Ligand for Receptor Tyrosine Kinase"

Inventors: Sacha J. Holland, Geraldine Mbamalu, Anthony Pawson

- 7 Title:** "SHC Proteins"

Inventors: John P. O'Bryan, Channing Der, John P. O'Bryan and Tony Pawson

- 8) Title:** "A Method of Quantitating GTP and GDP Bound to a G Protein and Uses Thereof"

Inventors: Gerry R. Ross, Abhijit Guha, Jurgen S. Scheele, Anthony Pawson

- 9) Title:** "Ligands for Discoidin Domain Receptor Tyrosine Kinases"

Investors: Wolfgang Vogel and Anthony Pawson

- 10) Title:** "Neuronal Receptor Tyrosine Kinase"

Inventors: Anthony Pawson, Mark Henkemeyer and Kenneth Letwin

Refereed Papers – Published/In press 2007-2012

Seet, B., D. Berry, J.S. Maltzman, J. Shabason, M. Raina, G.A. Koretzky, C.J. McGlade and T. Pawson. 2007. Efficient T cell receptor signaling requires a high affinity interaction between the Gads C-SH3 domain and the SLP-76 RXXK motif. **EMBO J.** 26:678-689.

Pawson, T. and N. Warner. 2007. Oncogenic re-wiring of cellular signaling pathways. **Oncogene.** 26:1268-1275.

Matskova, L.V., C. Helmstetter, R.I. Ingham, G. Gish, C.K. Lindholm, I. Ernberg, T. Pawson and G. Winberg. 2007. The Shb signaling scaffold binds to and regulates constitutive signals from the Epstein-Barr virus LMP2A membrane protein. **Oncogene** 26:4908-4917.

Ceccarelli, D.F.J., I. Blasutig, M. Goudreault, J. Ruston, T. Pawson and F. Sicheri. 2007. Non-canonical interaction of phosphoinositides with the pleckstrin homology domains of Tiam1 and ArhGAP9. **J. Biol. Chem.** 282:13864-13874.

Jones, N., W.R. Hardy, M.B. Friese, C. Jorgensen, M.J. Smith, N.M. Woody, S.J. Burden and T. Pawson. 2007. Analysis of a Shc family adaptor protein, ShcD/Shc4, that associates with Muscle-Specific Kinase (MuSK). **Mol. Cell. Biol.** 27:4759-4773.

Borg, M., T. Mittag, T. Pawson, M. Tyers, J.D. Forman-Kay and H. S. Chan. 2007. Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity **Proc. Natl. Acad. Sci USA** 104:9650-9655.

Machida, K., C.M. Thompson, K. Diercj, K. Jablonowski, S. Karkkainen, B. Liu, H. Zhang, P.D. Nash, D.K., Newman, P. Nollau, T. Pawson, H. Renkema, K. Saksela, M.R. Schiller, D-G. Shin and B.J. Mayer. 2007. High throughput phosphotyrosine profiling using SH2 domains. **Mol. Cell.** 26:899-915.

Linding, R., L.J. Jensen, G.J. Osteheimer, M.A.T.M. van Vugt, C. Jorgensen, I. Miron, F. Diella, K. Colwill, L. Taylor, J. Jin, P. Metalnikov, V. Nguyen, A. Pascalescu, J.G. Park, L.D. Samson, R.B. Russell, J.R. Woodgett, P. Bork, M.B. Yaffe and T. Pawson. 2007. Systematic discovery of in vivo phosphorylation networks. **Cell.** 129:1415-1426.

Hardy, W.R., L. Li, Z. Wang, J. Sedy, J. Fawcett, E. Frank, J. Kucera and T. Pawson. 2007. Combinatorial ShcA docking interactions support diversity in tissue morphogenesis. **Science** 317:251-256.

Kim, J., A. Thanabalasuriar, T. Chaworth-Musters, C. Fromme, E.A. Frey, P. I. Lario, P. Metalnikov, K. Rizg, N.A. Thomas, P.R. Hardwidge, T. Pawson, N.C. Strynadka, B.B. Finlay, R. Schekman and S. Gruenheid. 2007. The bacterial virulence factor NleA binds and disrupts function of mammalian COPII. **Cell Host and Microbe.** 2:160-171.

Publications

Pawson, Anthony James

Lim, C., B.T. Seet, R.J. Ingham, G. Gish, L. Matkova, G. Winberg, I. Ernberg and T. Pawson. 2007. The K15 protein of Kaposi's sarcoma-associated herpesvirus recruits the endocytic regulator intersectin 2 through a selective SH3 domain interaction. **Biochemistry** 46:9874-9885.

Murphy, J.M., D.M. Korzhnev, D.F. Ceccareli, D.J. Briant, A. Zarrine-Afsar, F. Sicher, L.E. Kay and T. Pawson. 2007. Conformational instability of the MARK3 UBA domain compromises Ub recognition and promotes interaction with the adjacent kinase domain. **Proc. Natl. Acad. Sci. USA.** 104:14336-14341.

Linding, R., L. Jensen, A. Pascalescu, M. Olhovsky, K. Colwill, P. Bork, M.B. Yaffe, and T. Pawson. 2007. NetworKIN: a resource for exploring cellular phosphorylation networks. **Nucleic Acids Research** 36:D695-699.

Fawcett, J.P., J. Georgiou, J. Ruston, F. Bladt, A. Sherman, N. Warner, B. J. Saab, R. Scott, J. Roder and T. Pawson. 2007. Nck adaptor proteins control the organization of neuronal circuits important for walking. **Proc. Natl. Acad. Sci. USA.** 104:20973-20978.

Pawson, T. 2007. Dynamic control of signaling by modular adaptor proteins. **Curr Opin Cell Biol.** 19: 112-116.

Blasutig, I.M., L.A. New, A. Thanabalsuriar, T.K. Dayarathna, M. Goudreault, S.E. Quaggin, S. S.-C. Li, S. Gruenheid, N. Jones and T. Pawson. 2008. Phosphorylated YDXV motifs and Nck SH2/S3 adaptors act cooperatively to induce actin reorganization. **Mol. Cell Biol.** 28:2035-2046.

Ursini-Siegel, J., W.R. Hardy, D. Zuo, S.H.L. Lam, V. Sanguin-Gendreau, R.D. Cardiff, T. Pawson, and W.J. Muller. 2008. ShcA signaling is essential for breast cancer progression. **EMBO J.** 27:910-920.

Lundgren, T.K., A. Stenqvist, R.P. Scott, T. Pawson and P. Ernfors. 2008. Cell migration by a FRS2-adaptor dependent membrane relocation of ret receptors. **J Cell Biochem.** 104:879-894.

To, C., S. Kulkarni, T. Pawson, T. Honda, G.W. Gribble, M.B. Sporn, J.L. Wrana and G.M. Di Guglielmo. 2008. The synthetic triterpenoid CDDO-imidazolidine alters TGFbeta -dependent signaling and cell migration by affecting the cytoskeleton and the polarity complex. **J Biol Chem.** 283:11700-11713.

Warner, N., L.E. Wybenga-Groot and T. Pawson. 2008. Analysis of EphA4 receptor tyrosine kinase substrate specificity using peptide-based arrays. **FEBS J.** 275:2561-2573.

Stenqvist, A., T.K. Lundgren, M.J. Smith, O. Hermanson, G. Castelo-Branco, T. Pawson and P. Ernfors. 2008. Subcellular receptor redistribution and enhanced microspike formation by a Ret receptor preferentially recruiting Dok. **Neuroscience letters.** 435:11-16.

Publications

Pawson, Anthony James

Traweger, A., G. Wiggin, L. Taylor, S.A. Tate, P. Metalnikov and T. Pawson. 2008. Protein phosphatase 1 regulates the phosphorylation state of the polarity scaffold Par-3. **Proc. Natl. Acad. Sci. USA.** 105:10402-10407.

Sonnberg, S., B.T. Seet, T. Pawson, S.B. Fleming and A.A. Mercer. 2008. Poxvirus ankyrin repeat proteins are a unique class of F-box proteins that associate with cellular SCF1 ubiquitin ligase complexes. **Proc. Natl. Acad. Sci. USA.** 105:10955-10960

Findlay, G.M. and T. Pawson. 2008. How is Sos activated? Let us count the ways. **Nature Structural & Molecular Biology.** 15:538-540.

Lim, K.B., W. Bu, W.I. Goh, E. Koh, S.H. Ong, T. Pawson, T. Sudhaharan and S Ahmed. 2008. The Cdcd42 effector IRSp53 generates filopodia by coupling membrane protrusion with actin dynamics. **J. Biol. Chem.** 283:20454-20472.

Zogopoulos, G., C. Jorgensen, J. Bacani, A. Montpetit, P. Lepage, V. Ferretti, L. Chad, S. Selvarajah, B. Zanke, T.J. Hudson, T. Pawson and S. Gallinger. 2008. Germline EphB2 receptor variants in familial colorectal cancer. **PLoS ONE** 3:e2885.

Huang, H., L. Li, C. Wu, D. Schibli, K. Colwill, S. Ma, C. Li, P. Roy, K. Ho, Z. Songyang, T. Pawson, Y. Gao, S.S. Li. 2008. Defining the specificity space of the human SRC homology 2 domain. **Mol Cell. Proteomics** 7:768-784.

Filippakopoulos, P., M. Kofler, O. Hantschel, G.D. Gish, F. Grebien, E. Salah, P. Neudecker, L.E. Kay, B.E. Turk, G. Superti-Furga, T. Pawson and S. Knapp. 2008. Structural coupling of SH2-kinase domains links Fes and Abl substrate recognition and kinase activation. **Cell** 134:793-803.

Ruusala, A., T. Pawson, C.-H. Heldin, and P. Aspenström. 2008. Nck adapters are involved in formation of dorsal ruffles, cell migration and Rho signalling downstream of the PDGF β -receptor. **J. Biol. Chem.** 283:30034-30044.

Bakal, C., R. Linding, F. Llense, E. Heffern, E. Martin-Blanco, T. Pawson and N. Perrimon. 2008. Phosphorylation networks regulating JNK activity in diverse genetic backgrounds. **Science** 322:453-456.

Miller, M.L., L.J. Jensen, F. Diella, C. Jorgensen, M. Tinti, L. Li, M. Hsiung, S.A. Parker, J. Bordeaux, T. Sicheritz-Ponten, M. Olhovsky, A. Pascalescu, J. Alexander, S. Knapp, N. Blom, S. Li, G. Cesareni, T. Pawson, B.E. Turk, M.B. Yaffe, S. Brunak and R. Linding. 2008. Linear motif atlas for phosphorylation-dependent signaling. **Sci. Signal.** 1(35):ra2.

Chang, Q., C. Jorgensen, T. Pawson and D.W. Hedley. 2008. Effects of dasatinib on EphA2 receptor tyrosine kinase activity and downstream signalling in pancreatic cancer. **Br. J. Cancer.** 99:1074-1082.

Publications

Pawson, Anthony James

Pawson, T. and R. Linding. 2008. Network medicine. **FEBS Lett.** 582:1266-1270.

Pawson, T. and C. Jorgensen. 2008. Signal transduction by growth factor receptors. In “The Molecular Basis of Cancer” 3rd edition, Eds Mendelsohn, Howley, Israel, Gray and Thompson, Pub Saunders/Elsevier Inc. pp. 155-168.

Braun, P., M. Tasan, M. Dreze, M. Barrios-Rodiles, I. Lemmens, H. Yu, J.M. Sahalie, R.R. Murray, L. Roncari, A.S. de Smet, K. Venkatesan, J.F. Rual, J. Vandenhoute, M.E. Cusick, T. Pawson, D.E. Hill, J. Tavernier, J.L. Wrana, F.P. Roth and M. Vidal. 2009. An experimentally derived confidence score for binary protein-protein interactions. **Nat. Methods** 6:91-97.

Taylor, I.W., R. Linding, D. Warde-Farley, Y. Liu, C. Pesquita, D. Faria, S. Bull, T. Pawson, Q. Morris and J.L. Wrana. 2009. Dynamic modularity in protein interaction networks predicts breast cancer outcome. **Nature Biotechnology**. 27:199-204.

MacNeil, L.T., W.R. Hardy, T. Pawson, J.L. Wrana and J.G. Culotti. 2009. UNC-129 regulates the balance between UNC-40 dependent and independent UNC-5 signaling pathways. **Nature Neuroscience** 12:150-155

Pawson, T. and M. Kofler. 2009. Kinome signaling through regulated protein-protein interactions in normal and cancer cells. **Current Opinion in Cell Biology**. 21:147-153.

Ho, S.K., N. Kovačević, R.M. Henkelman, A. Boyd, T. Pawson and J.T. Henderson. 2009. EphB2 and EphA4 receptors regulate formation of the principal inter-hemispheric tracts of the mammalian forebrain. **Neuroscience** 160:784-795.

Pawson, T. and L. Taylor. 2009. Protein phosphorylation goes negative. **Molecular Cell**. 34:139-140.

Jones N., L.A. New, M.A. Fortino, V. Eremina, J. Ruston, I.M. Blasutig, L. Aoudjit, Y. Zou, X. Liu, G.L. Yu, T. Takano, S.E. Quaggin and T. Pawson. 2009. Nck proteins maintain the adult glomerular filtration barrier. **J Am Soc Nephrol.** 20:1533-1543.

Tan, C.S.H., A. Pascalescu, W.A. Lim, T. Pawson, G. Bader and R. Linding. 2009. Positive selection of tyrosine loss in metazoan evolution. **Science** 325:1686-1688.

Tan, C.S.H., B. Bodenmiller, A. Pascalescu, M. Jovanovic, M.O. Hengartner, C. Jorgensen, G.D. Bader, R. Aebersold, T. Pawson and R. Linding. 2009. Comparative analysis of protein phosphorylation reveals evolutionary conserved disease network. **Science Signaling**. 2(81): ra39.

Schmoranzer, J., J.P. Fawcett, M. Segura, S. Tan, R.B. Vallee, T. Pawson and G.G. Gundersen. 2009. Par3 and dynein associate to regulate local microtubule dynamics and centrosome organization during migration. **Current Biology** 19:1065-1074.

Murphy, J.M, D.F. Hansen, S. Wiesner, D.R. Muhandiram, M. Borg, M.J. Smith, F. Sicheri, L.E. Kay, J.D. Forman-Kay and T. Pawson. 2009. Structural studies of FF domains of the transcription factor, CA150, provide insights into the organization of FF domain tandem arrays. **J. Mol. Biol.** 393:409-424.

Jørgensen, C., A. Sherman, G.I. Chen, A. Pascalescu, A. Poliakov, M. Hsiung, B. Larsen, D.G. Wilkinson, R. Linding and T. Pawson. 2009. Cell-specific information processing in segregating populations of Eph receptor/ephrin-expressing cells. **Science** 326:1502-1509.

Chen, C., J. Jin, A. James, M.A. Adams-Cioaba, J.G. Park, Y. Guo, E. Tenaglia, C. Xu, G. Gish, J. Min and T. Pawson. 2009. Mouse Piwi Interactome Identifies Binding Mechanism of Tdrkh Tudor Domain to Arginine Methylated Miwi. **Proc. Natl. Acad. Sci. USA.** 106:20336-20341.

Scott, J.D. and T. Pawson. 2009. Signaling in space and time: Where proteins come together and when they're apart. **Science** 326:1220-1224.

Jin, J., X. Xie, C. Chen, J.G. Park, C. Stark, D.A. James, M. Olhovsky, R. Linding, Y. Mao and T. Pawson. 2009. Eukaryotic protein domains as functional units of cellular evolution. **Science Signaling** 2(98): ra76.

van Vugt, M.A.T.M., A.K. Gardino, R. Linding, G.J. Osthheimer, H.C. Reinhardt, S-E. Ong, C. S.A. Tan, H. Miao, S.M. Keezer, J.Li, T. Pawson, T.A. Lewis, S.A. Carr, S.J. Smerdon, T.R. Brummelkamp and M.B. Yaffe. 2009. A mitotic phosphorylation feedback network connects Cdk1, Plk1, 53BP1 and Chk2 to inactivate the G₂/M DNA damage checkpoint. **Plos Biol.** 8:e1000287.

Smith, M.J., W.R. Hardy, G.-Y. Li, M. Goudreault, S. Hersch, P. Metalnikov, A. Starostine, T. Pawson and M. Ikura. 2009. The PTB domain of ShcA couples receptor activation to the cytoskeletal regulator IQGAP1. **EMBO J.** 29:884-896.

Pawson T, Vanhook AM. 2009. Science Signaling Podcast. **Science Signaling**. 2(98):pc21.

Pershad, K., J.D. Pavlovic, S. Gräslund, P. Nilsson, K. Colwill, A. Karatt-Vellatt, D.J. Schofield., M.R. Dyson, T. Pawson, B.K. Kay and J. McCafferty. 2010. Generating a panel of highly specific antibodies to 20 human SH2 domains by phage display. **Protein Engineering, Design, and Selection.** pp. 1-10.

Nachman, J., R. Pomès, G. Gish, C. Virag, T. Pawson and E. Pai. 2010. Conformational determinants of phosphotyrosine peptides complexed with the Src SH2 domain. **PLoS ONE** 5:e11215.

Publications

Pawson, Anthony James

Richier, L., K. Williton, L. Clattenburg, K. Colwill, M. O'Brien, C. Tsang, A. Kolar, N. Zinck, P. Metalnikov, W.S. Trimble, S.B. Kruegger, T Pawson and J.P. Fawcett. 2010. NOS1AP associates with Scribble and regulates dendritic spine development. **J. Neuroscience** 30:4796-4805.

Breitkreutz. A., H. Choi, J. Sharom, L. Boucher, V. Neduvia, B. Larsen, Z-Y. Lin, B-J. Breitkreutz, C. Stark, G. Liu, J. Ahn, D. Dewar-Darch, T. Reguly, X. Tang, R. Almeida, Z.S. Qin, T. Pawson, A-C. Gingras, A.I. Nesvizhskii and M. Tyers. 2010. A Global Protein Kinase and Phosphatase Interaction Network in Yeast. **Science** 328:1043-1046.

Varelas, X., B.W. Miller, R. Sopko, S. Song, A. Gregorieff, F.A. Fellouse, R. Sakuma, T. Pawson, W. Hunziker, H. McNeill, J.L. Wrana and L. Attisano. 2010. The Hippo pathway regulates Wnt/beta-catenin signaling. **Dev. Cell** 18:579-591.

Deribe, Y.L., T. Pawson T and I. Dikic. 2010. Post-translational modifications in signal integration. **Nat Struct Mol Biol.** 17:666-672.

Lim, W.A. and T. Pawson. 2010. Phosphotyrosine signaling: Evolving a new cellular communication system. **Cell.** 142:661-667.

Roy, E., D. Togbe, A.D. Holdorf, D. Trubetskoy, S. Nabti, G. Küblbeck, A. Klevenz, A. Kopp-Schneider, F. Leithäuser, P. Möller, F. Bladt, G. Hämerling, B. Arnold, T. Pawson and A. Tafuri. 2010. The Nck adaptors are positive regulators of the size and sensitivity of the T cell repertoire. **Proc Natl Acad Sci USA.** 107:15529-15534

Liu, K., C. Chen, Y. Guo, R. Lam, C. Bian, C. Xu, D.Y. Zhao, J. Jin, F. MacKenzie, T. Pawson, and J. Min. 2010. Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. **Proc. Natl. Acad. Sci. USA.** 107:18398-18403.

Ursini-Siegel, J. S. Cory, D. Zuo, W.R. Hardy, E. Rexhepaj, S. Lam, B. Schade, K. Jirstrom, E. Bjur, C.A. Piccirillo, D. DeNardo, L.M. Coussens, D.J. Brennan, W.M. Gallagher, M. Park, T. Pawson, M. Hallett and W.J. Muller. 2010. Receptor tyrosine kinase signaling in breast cancer cells inhibits the adaptive immune response to favor a pro-tumorigenic state. **Cancer Research.** 70:7776-7787

Liu, G., J. Zhang, B. Larsen, C. Stark, A. Breitkreutz, Z.Y. Lin, B.J. Breitkreutz, Y. Ding, K. Colwill, A. Pascalescu, T. Pawson, J.L. Wrana, A.I. Nesvizhskii, B. Raught, M. Tyers and A.C. Gingras. 2010. ProHits: integrated software for mass spectrometry-based interaction proteomics. **Nat Biotechnol.** 28:1015-1017.

Publications

Pawson, Anthony James

Roy, E., D. Togbe, A. Holdorf, D. Trubetskoy, S. Nabti, G. Küblbeck, S. Schmitt, A. Kopp-Schneider, F. Leithäuser[†], P. Möller, F. Bladt, G. J. Hä默ling, B. Arnold, T. Pawson and A. Tafuri. 2010. Fine tuning of the threshold of T cell selection by the Nck adaptors. **J. Immunol.** 187:7518-7526

Smith, F.D., L.K. Langeberg, C. Cellurale, T. Pawson, D.K. Morrison, R.J. Davis, J.D. Scott. 2010. AKAP-Lbc enhances cyclic AMP control of the ERK1/2 cascade. **Nat Cell Biol.** 12:1242-1249.

Vanderlaan, R.D., W.R. Hardy, M.G. Kabir, A. Pascalescu, N. Jones, P.P. deTombe, P.H. Backx and T. Pawson. 2011. The ShcA phosphotyrosine docking protein uses distinct mechanisms to regulate myocyte and global heart function. **Circ. Res.** 108:184-193.

Pike, K., S. Kulkarni and T. Pawson. 2011. Immature T cell clustering and efficient differentiation require the polarity protein Scribble. **Proc Natl Acad Sci USA.** 108:1116-1121.

Colwill K, Persson H, Jarvik NE, Wyrzucki A, Wojcik J, Koide A, Kossiakoff AA, Koide S, Sidhu S, Dyson MR, Pershad K, Pavlovic JD, Karatt-Vellatt A, Schofield DJ, Kay BK, McCafferty J, Mersmann M, Meier D, Mersmann J, Helmsing S, Hust M, Dübel S, Berkowicz S, Freemantle A, Spiegel M, Sawyer A, Layton D, Nice E, Dai A, Rocks O, Williton K, Fellouse FA, Hersi K, Pawson T, Nilsson P, Sundberg M, Sjöberg R, Sivertsson A, Schwenk JM, Takanen JO, Hofer S, Uhlén M, Dahlgren LG, Flores A, Johansson I, Weigelt J, Crombet L, Loppnau P, Kozieradzki I, Cossar D, Arrowsmith CH, Edwards AM, Gräslund S. 2011. A roadmap to generate renewable protein binders to the human proteome. **Nat Methods** 8:551-558.

Tan,C.S.H., E.M. Schoof, P. Creixell, A. Pascalescu, W. A. Lim, T. Pawson, G. D. Bader, R. Linding. 2011. Response to Comment on “Positive Selection of Tyrosine Loss in Metazoan Evolution”. **Science.** 332, 917

Dyson, M.R., Y. Zheng, C. Zhang, K. Colwill, K. Pershad, B.K. Kay, T. Pawson, J. McCafferty. 2011. Mapping protein interactions by combining antibody affinity maturation and mass spectrometry. **Anal Biochem.** 417: 25-35.

Bisson, N., D.A. James, G. Ivosev, S.A. Tate, R. Bonner, L. Taylor and T. Pawson. 2011. Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. **Nat Biotech** 29:653-658.

Olhovsky, M., K. Williton, A.Y. Dai, A. Pascalescu, J.P. Lee, M. Goudreault, C.D. Wells, J.G. Park, A-C. Gingras, R. Linding, T. Pawson and K. Colwill. 2011. OpenFreezer: a reagent information management software system. (Correspondence) **Nat Methods** 8:612-613.

Publications

Pawson, Anthony James

Chen, C., T.J. Nott, J. Jin and T. Pawson. 2011. Deciphering arginine methylation: Tudor tells the tale. **Nature Reviews: Molecular Cell Biology**. 12: 1-14.

Kaniuk, N.A., V. Canadien, R.D. Bagshaw, M. Bakowski, V. Braun, M. Landekic, S. Mitra, J. Huang, W.D. Heo, T. Meye, L. Pelletier, H. Andrews-Polymenis, M. McClelland, T. Pawson, S. Grinstein and J.H Brumell. 2011. *Salmonella* exploits Arl8B- directed kinesin activity to promote endosome tubulation and cell-to-cell transfer. **Cell Microbiol.** 13:1812-1823.

Hu, S. T. Pawson, and R.M. Steven. 2011. UNC-73/Trio RhoGEF-2 activity modulates *Caenorhabditis elegans* motility through changes in neurotransmitter signaling upstream of the GSA-1/Gas Pathway. **Genetics** 189: 137-151

Guettler, S., J. LaRose, G. Gish, E. Petsalaki, A. Scotter, T. Pawson, R. Rottapel and F. Sicheri. 2011. Structural basis and sequence rules for substrate recognition by Tankyrase explain the basis for cherubism disease. **Cell** 147:1340-1354.

Grebien, F., O. Hantschel, J. Wojcik, I. Kaupe, B. Kovacic, A.M. Wyrzucki, G.D. Gish, S. Cerny-Reiterer, A. Koide, H. Beug, T. Pawson, P. Valent, S. Koide and G. Superti-Furga. 2011. Targeting the SH2-kinase interface in Bcr-Abl inhibits leukemogenesis. **Cell** 147:306-319.

Janes, P.W., B. Griesshaber, L. Atapattu, E. Nievergall, L.L. Hill, A. Mensinga, A. Chheang, B.W. Day, A.W. Boyd, P.I. Bastianes, C. Jorgensen, T. Pawson and M. Lackmann. Eph receptor function is modulated by heterooligomerization of A and B type Eph receptors. **J. Cell. Biol.** 195:1033-1045.

Tang, X., Orlicky, S., Mittag, T., Csizmok, V., Pawson, T., Forman-Kay, J.D., Sicheri, S., and Tyers, M. Composite low affinity interactions set a multisite phosphorylation threshold for recognition of Sic1 by the SCFCdc4 ubiquitin ligase. **Proc. Natl. Acad. Sci. USA.** Feb. 10, E-publ.

Lambert, J-P., T. Pawson and A-C. Gingras. 2012. Mapping physical interaction within chromatin by proteomic approaches. **Proteomics**. In press

Filippakopoulos, P., S. Picaud, M. Mangos, T. Keates, J.-P. Lambert, D. Barsyte-Lovejoy, I. Felletar, R. Volkmer, S. Müller, A.-C. Gingras, T. Pawson, C. Arrowsmith and S. Knapp. 2012. Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. **Cell**. In press.