

Cancer Genome (Data) Analysis in Personalised Medicine

Genome Variation | Data Formats | Resources | Sharing | Privacy

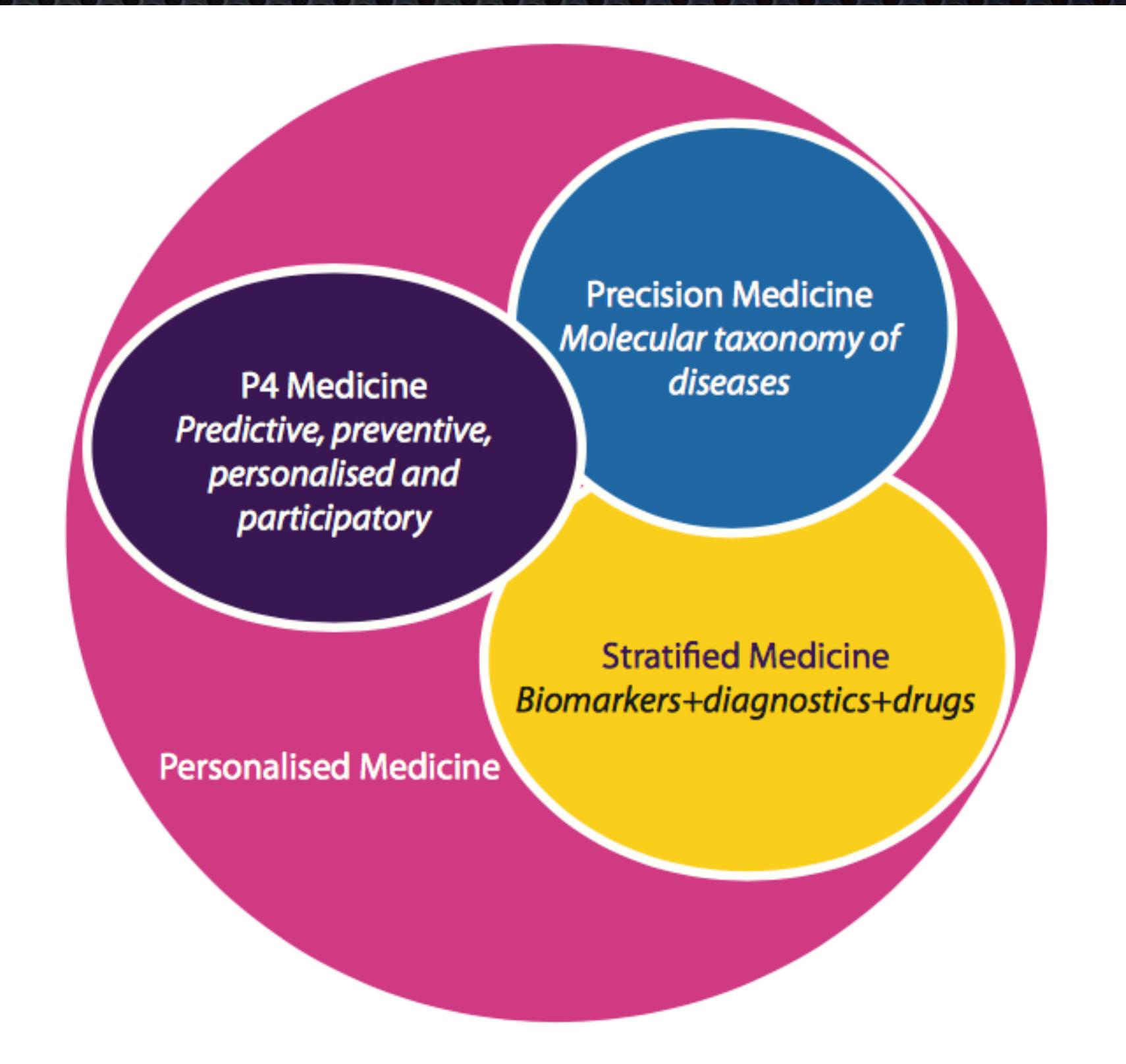
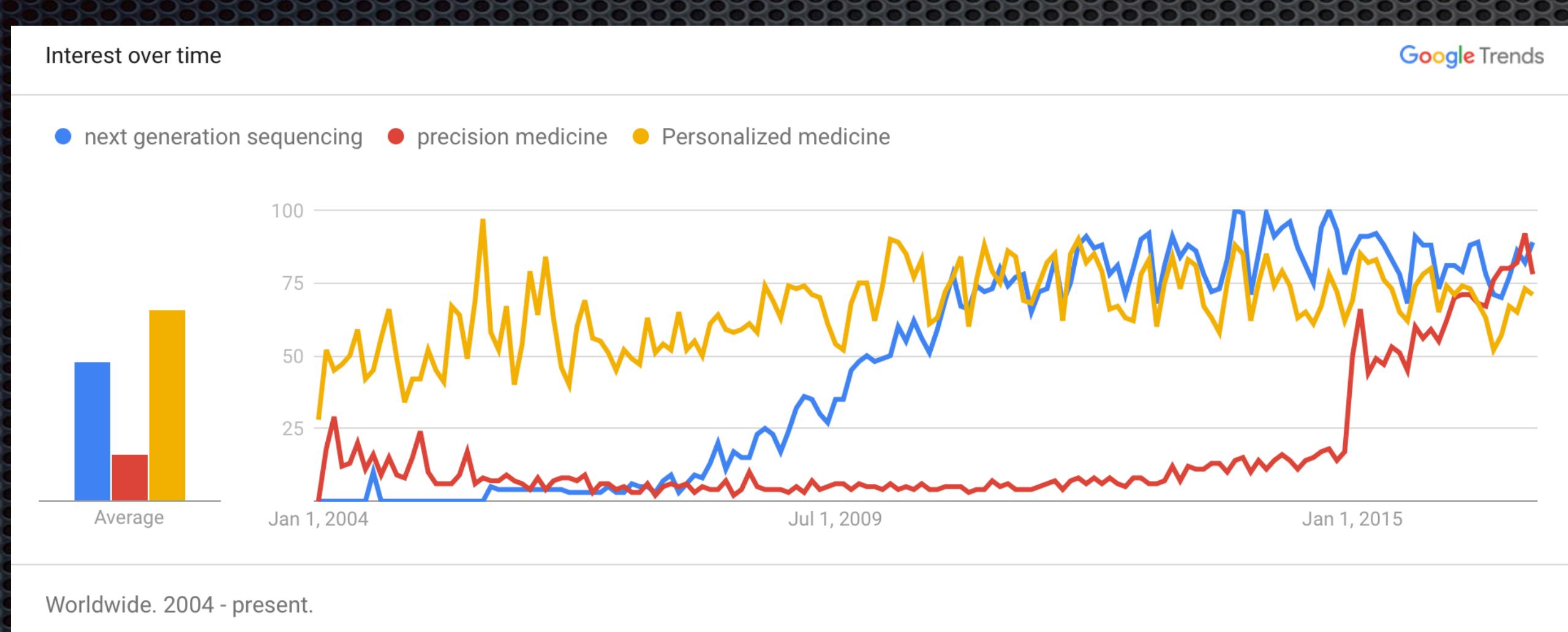
Michael Baudis **UZH SIB**
Computational Oncogenomics



University of
Zurich^{UZH}

Many names for one concept or many concepts in one name?

Stratified, personalised, precision, individualised, P4 medicine or personalised healthcare – all are terms in use to describe notions often referred to as the future of medicine and healthcare. But what exactly is it all about, and are we all talking about the same thing?



Source: PHG Foundation

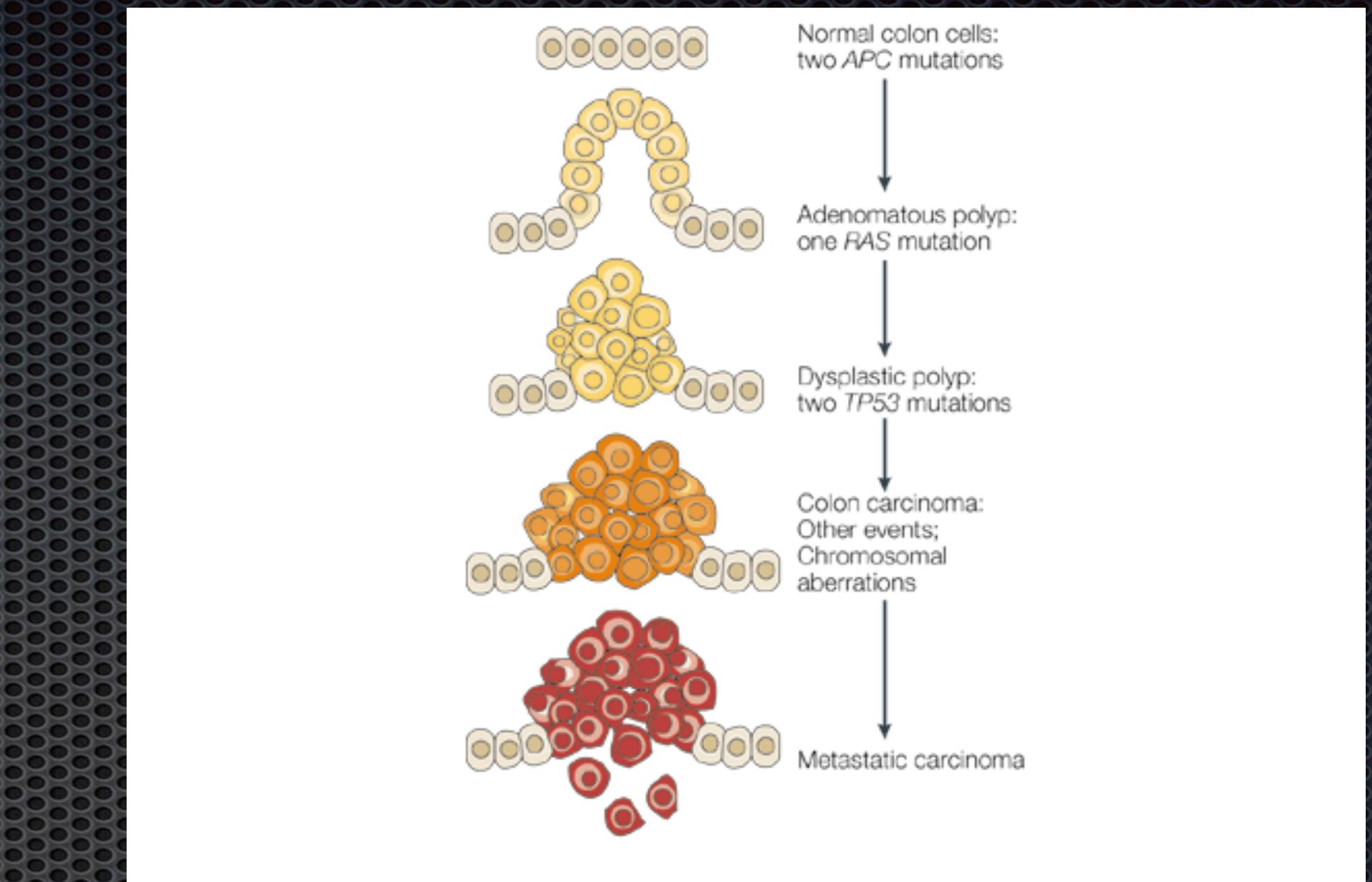
While medicine has always been "personal" and "precise" in the given context of available knowledge and technologies, the concept of "**Personalised Medicine**" describes the use of individual genome information, concept based metadata and individually targeted therapies.

Cancers are diseases of genomes

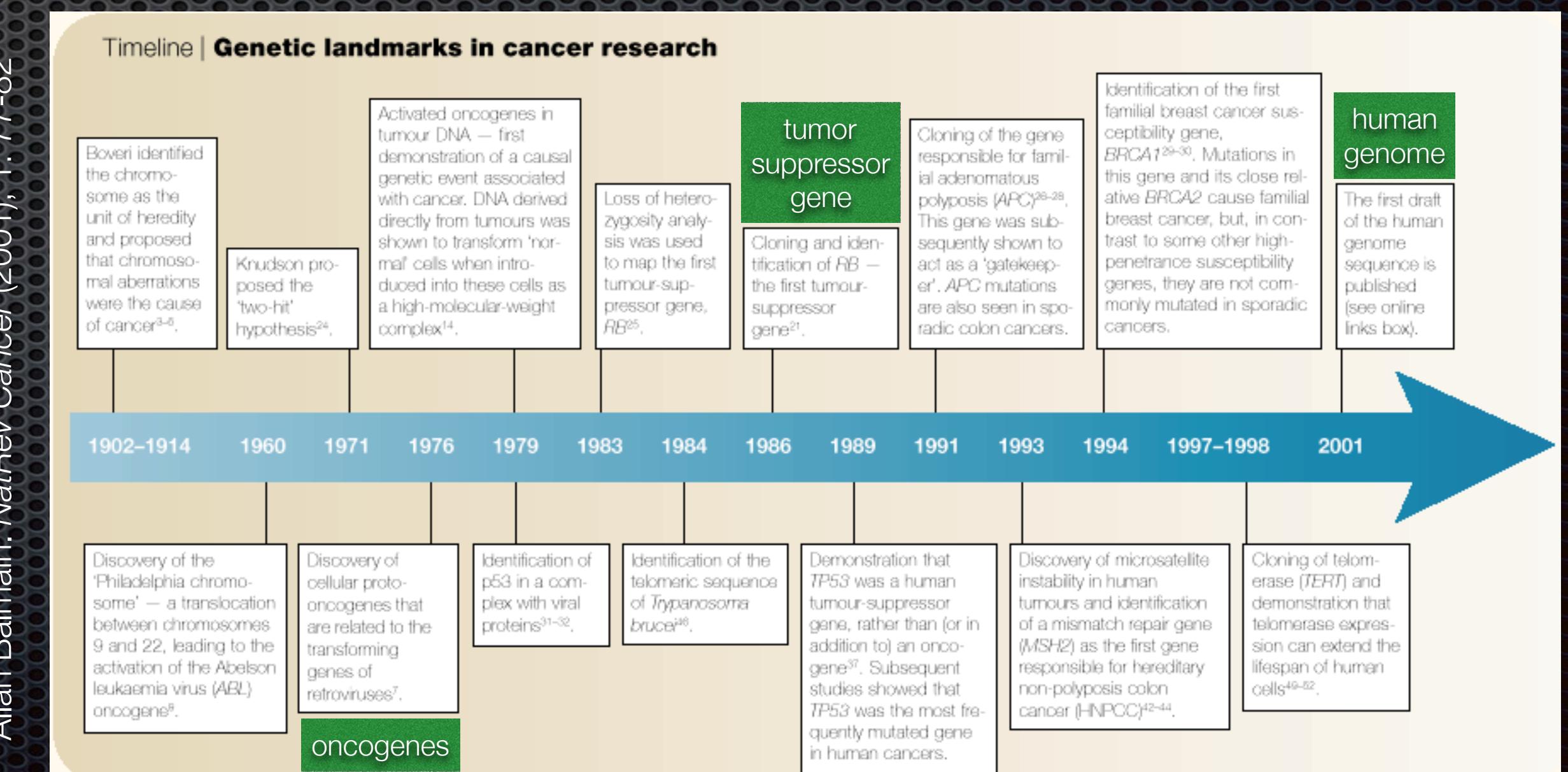
- every cancer's genome is different
- every individual's genome is different

Cancers arise from the clonal accumulation of **somatic genome mutations**, with varying but **limited** contribution of **inherited risk**

Allan Balmain. *NatRev Cancer* (2001); 1: 77-82



Knudson, A. G. (2001). Two genetic hits (more or less) to cancer. *Nature Reviews Cancer*, 1(2), 157–162.





Theodor Boveri (1914)

(based on observations in sea urchin eggs)

- **Cell-cycle checkpoints** (“Hemmungseinrichtung”)
- **Tumour-suppressor genes** (“Teilungshemmende Chromosomen”); can be eliminated during tumour progression
- amplified **Oncogenes** (“Teilungsfoerdernde Chromosomen” ... “im permanenten Übergewicht”)
- sequential **Progression** (benign to malignant)
- Cancer **predisposition** through inheritance of less able suppressor “chromosomes”
- high-penetrance cancer syndromes - (e.g. xeroderma pigmentosum) through **homozygosity**
- Clonal origin & Genetic mosaicism; wounding and inflammation in tumour promotion; loss of cell adhesion in metastasis; sensitivity of malignant cells to radiation therapy (based on Hertwig *et al.*)

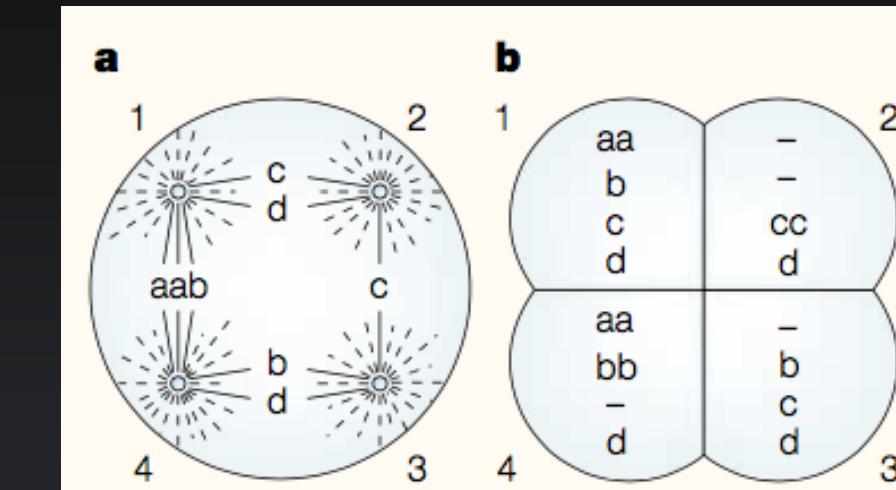
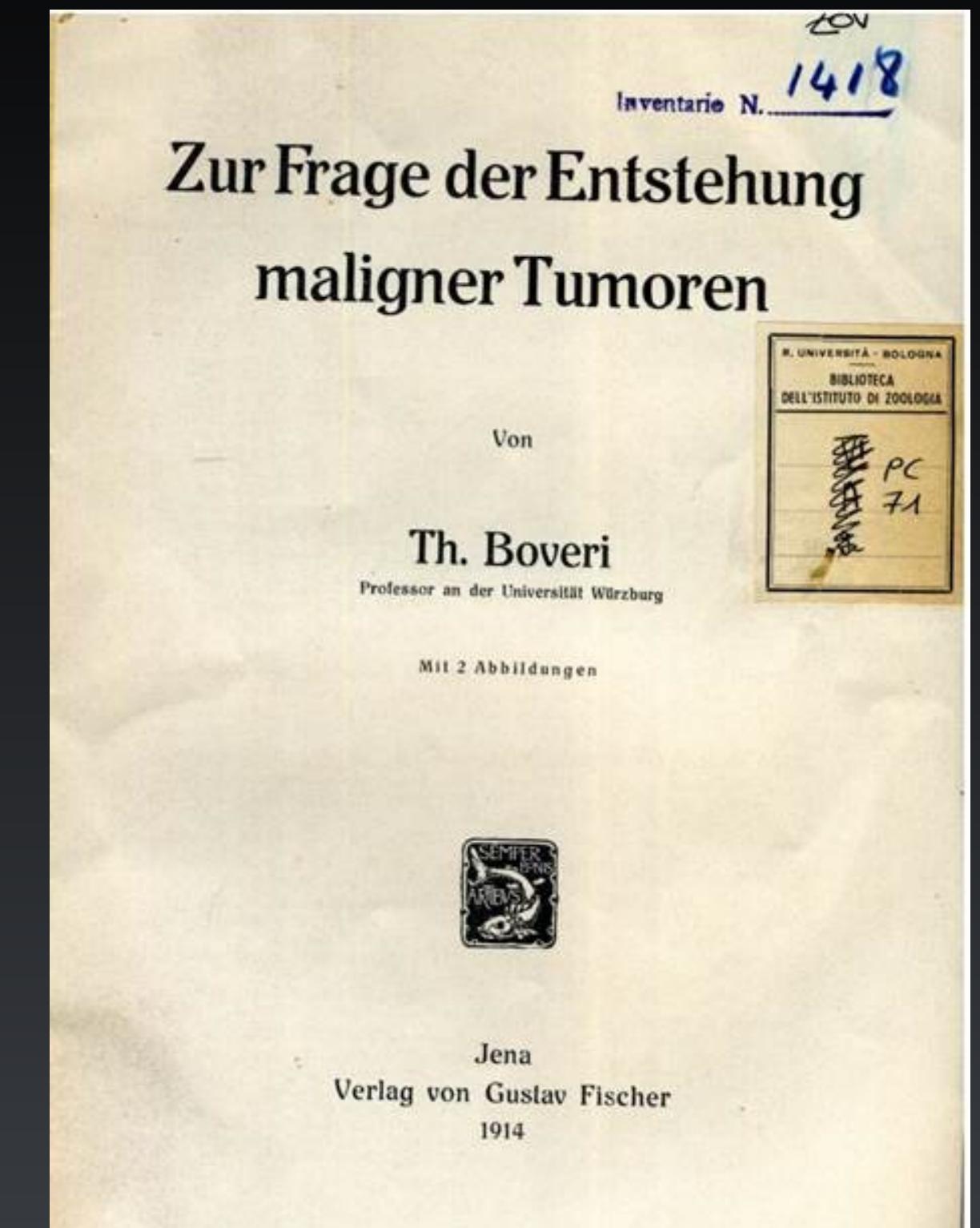


Figure 2 | Multiple cell poles cause unequal segregation of chromosomes. **a** | Boveri showed that fertilization of sea-urchin eggs by two sperm results in multiple cell poles. Individual chromosomes then attach to different combinations of poles — for example, one copy of chromosome c is attached to poles 1 and 2, and one copy is attached to poles 2 and 3. **b** | Chromosomes are segregated to the four poles at cell division, leaving some cells with too many copies of the chromosomes and some with too few — for example, cell 2 has two copies of chromosome c and cell 4 has none.



Allan Balmain
Cancer genetics: from Boveri and
Mendel to microarrays.
NatRev Cancer (2001); 1: 77-82

Anna Di Lonardo , Sergio Nasi , Simonetta Pulciani
Cancer: We Should Not Forget The Past
Journal of Cancer (2015), Vol. 6: 29-39
(for book cover & summary)



Janet Rowley (1972/73)

Chromosomal translocations in cancer

- Recurrent chromosomal translocations in leukemias and lymphomas
- "Philadelphia chromosome" in CML (Nowell & Hungerford, 1960) represents a reciprocal translocation between chromosomes 9 and 22
- 1972: t(8;21) ALL manuscript rejected by NEJM
- 1973: t(9;22) manuscript rejected by *Nature* "with some reasonable comments and some truly wrong"
- Clinical implications: **Tyrosine Kinase inhibitors** as standard first-line therapy in CML
 - first trials in 1998 (STI-571; Imatinib/Gleevec)
 - cf. Druker BJ, Lydon NB (2000). Lessons learned from the development of an Abl tyrosine kinase inhibitor... *J Clin Invest* 2000;105:3-7)

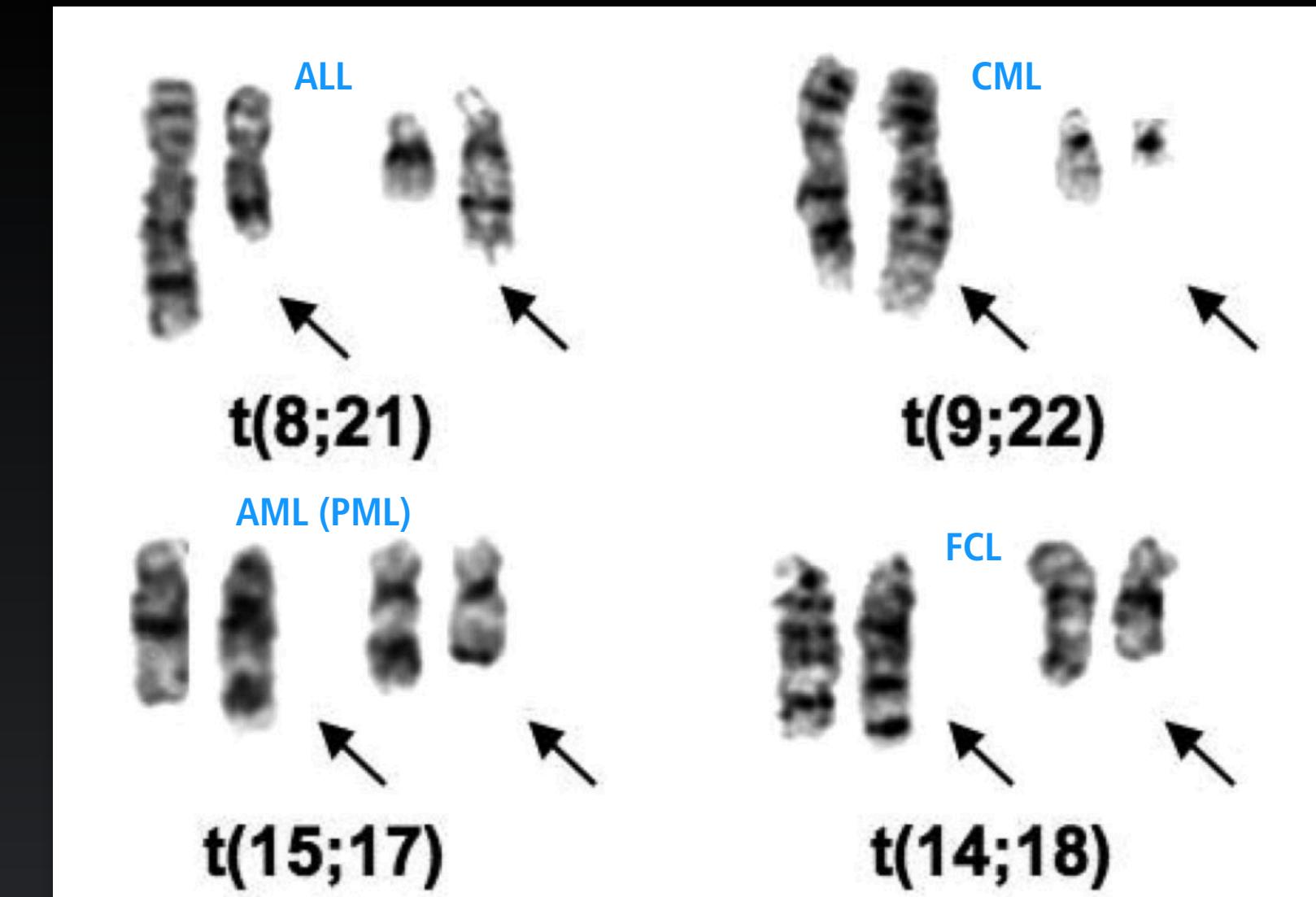
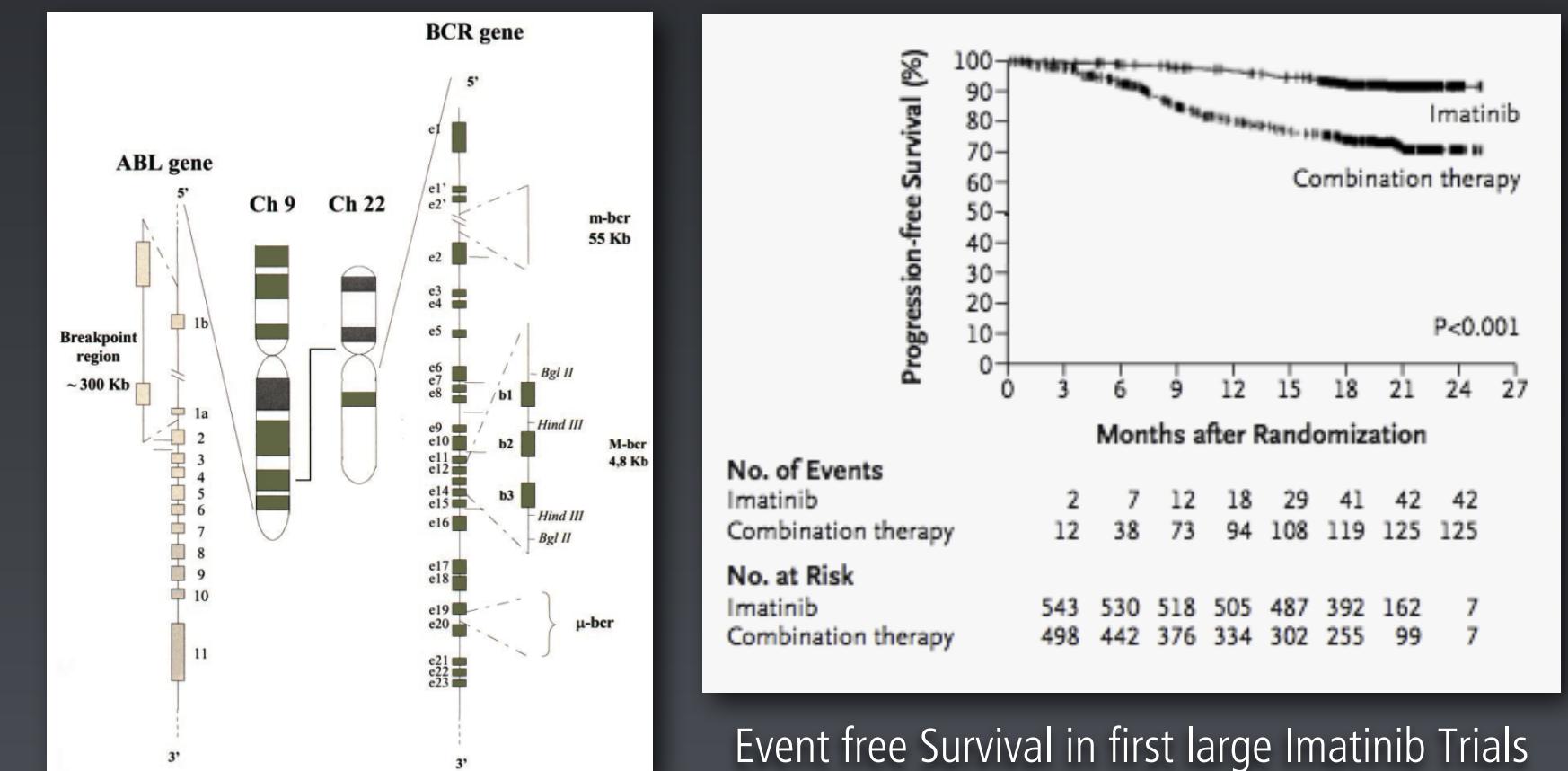


Figure 1. Partial karyotypes of common translocations discovered by Rowley.
The translocations appear in the order in which they were discovered.

Janet D Rowley. Chromosomal translocations: revisited yet again *Blood* (2008), 112(6)



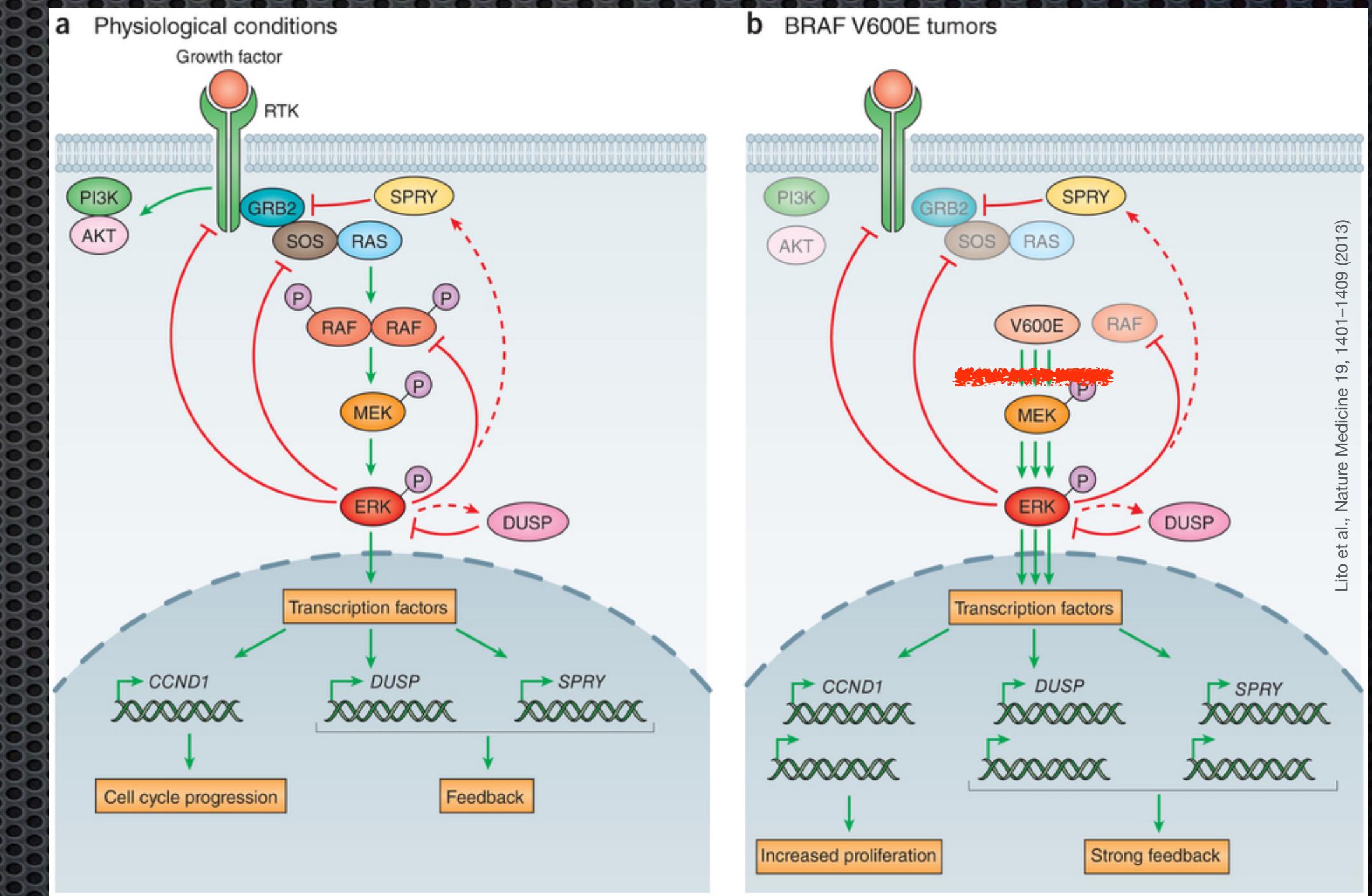
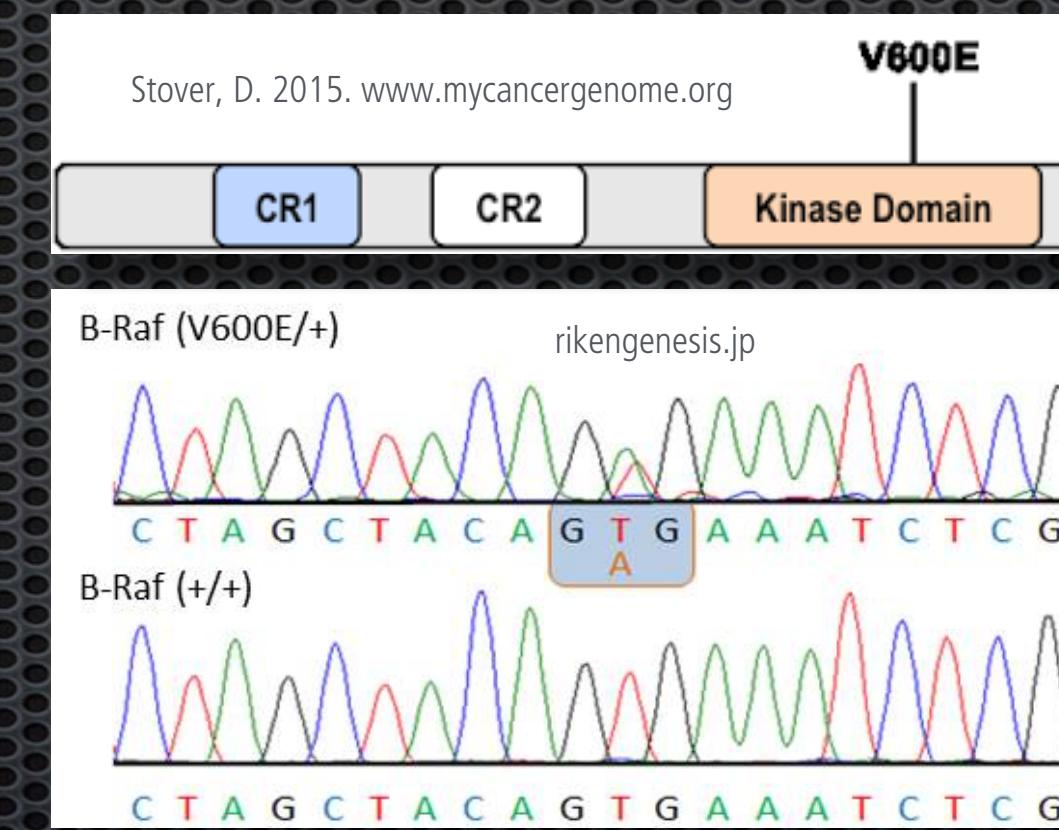
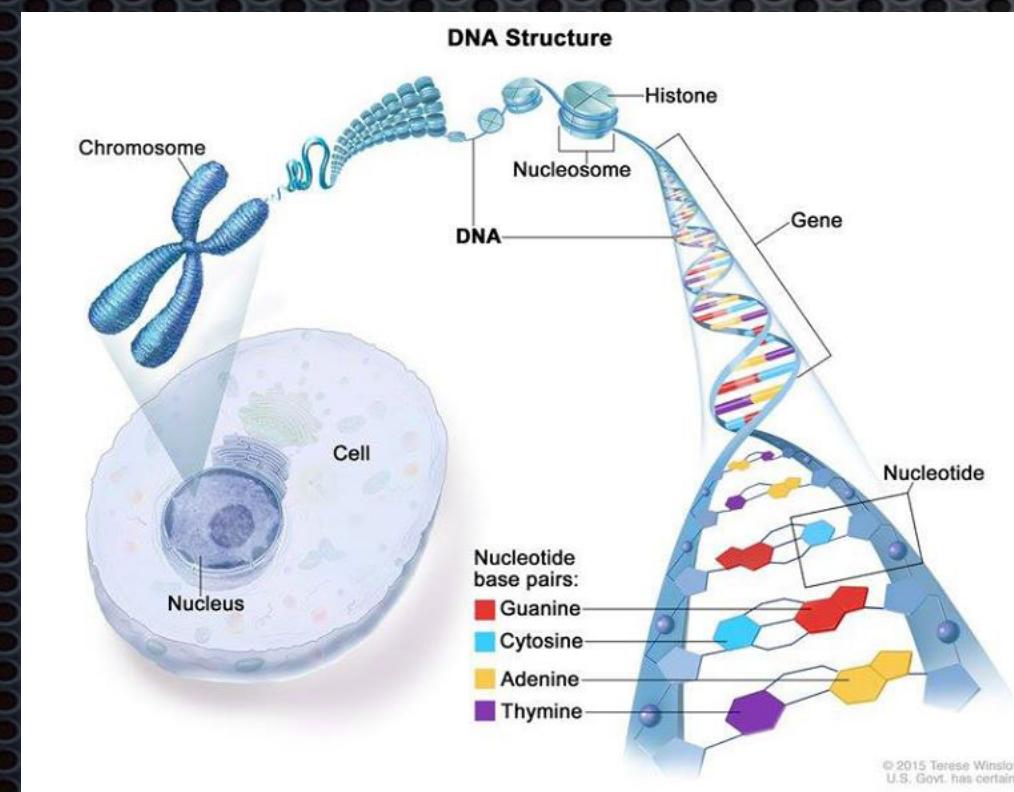
Event free Survival in first large Imatinib Trials

Pane et al. BCR/ABL genes
Oncogene (2002), 21 (56)

O'Brien et al. Imatinib compared with interferon and low-dose cytarabine... *NEJM* (2003) vol. 348 (11)

BRAF V600E (c.1799T>A) Mutation Oncogene Activation by Single Nucleotide Alteration

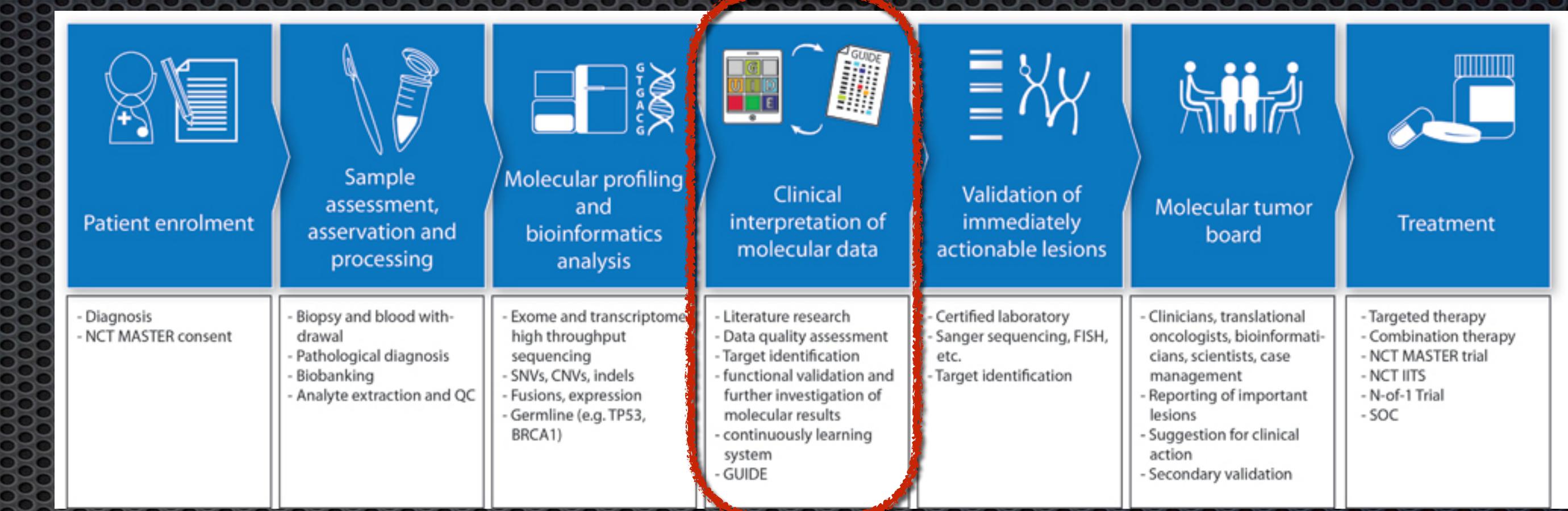
- a single nucleotide exchange Thymidine > Adenine leads to continuous RAF based activation of the MEK-ERK pathway
- BRAF V600E is a frequent mutation in >50% of malignant melanomas, but also CRC, lung ADC ...
- pharmacologic block of B-Raf (e.g. through **Vemurafenib**)



The BRAF V600E mutation leads to continuous phosphorylation of MEK, without the need for receptor based activation of the upstream pathway and loss of inhibitory feedback control.

Personalised Medicine in Cancer - A Genome Based Approach

- personalized cancer therapy uses information about the **individual genetic background** and **tumor sequence analysis** for the identification of somatic variants



Workflow of a cancer treatment protocol based on "personalized" assessment of actionable genomic lesions (source: NCT Heidelberg).

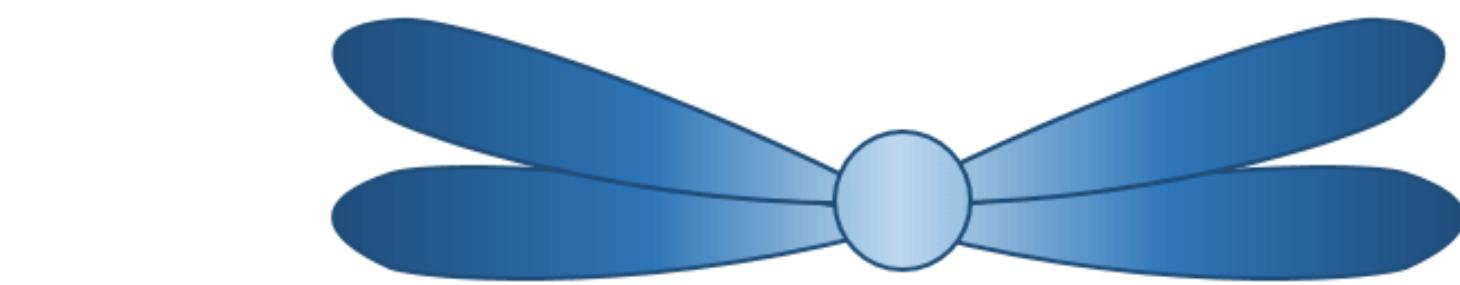
- currently mostly use of **targeted / panel sequencing** for identification of tens - hundreds of most common "actionable" mutations
- knowledge resources and literature search for interpretation of non-standard variants

Genome analyses at the core of Personalized Health™

- Genome analyses (including transcriptome, metagenomics) are the **core technologies** for Personalized Health™ applications
- In the context of **academic medicine**, this requires
 - standard sample acquisition procedures & central **biobanking**
 - **core sequencing facility** (large throughput, cost efficiency, uniform sample and data handling procedures)
- secure **computing/analysis** platform
- Standardized **data formats** and **sample identification** procedures
- Metadata rich, **reference variant resources** & expertise
- participation in reciprocal, international **data sharing** and **biocuration** efforts

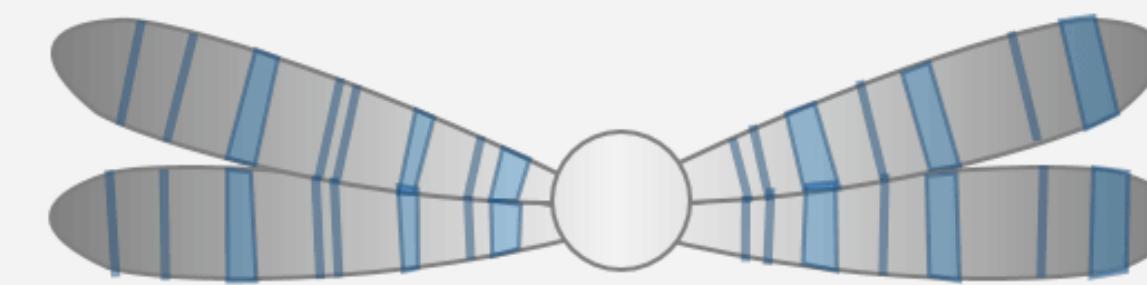
Genome Sequencing

whole genome sequencing (WGS)



100%
 all DNA
(3.1 billion base pairs)

exome sequencing



~1%
 protein-coding DNA only
(~31 million base pairs)

What does it cost to sequence a genome?

Human Genome

Project (HGP):

1991-2003

today:

2017

cost: \$2.7 billion

time: 12+ years

~\$1,500

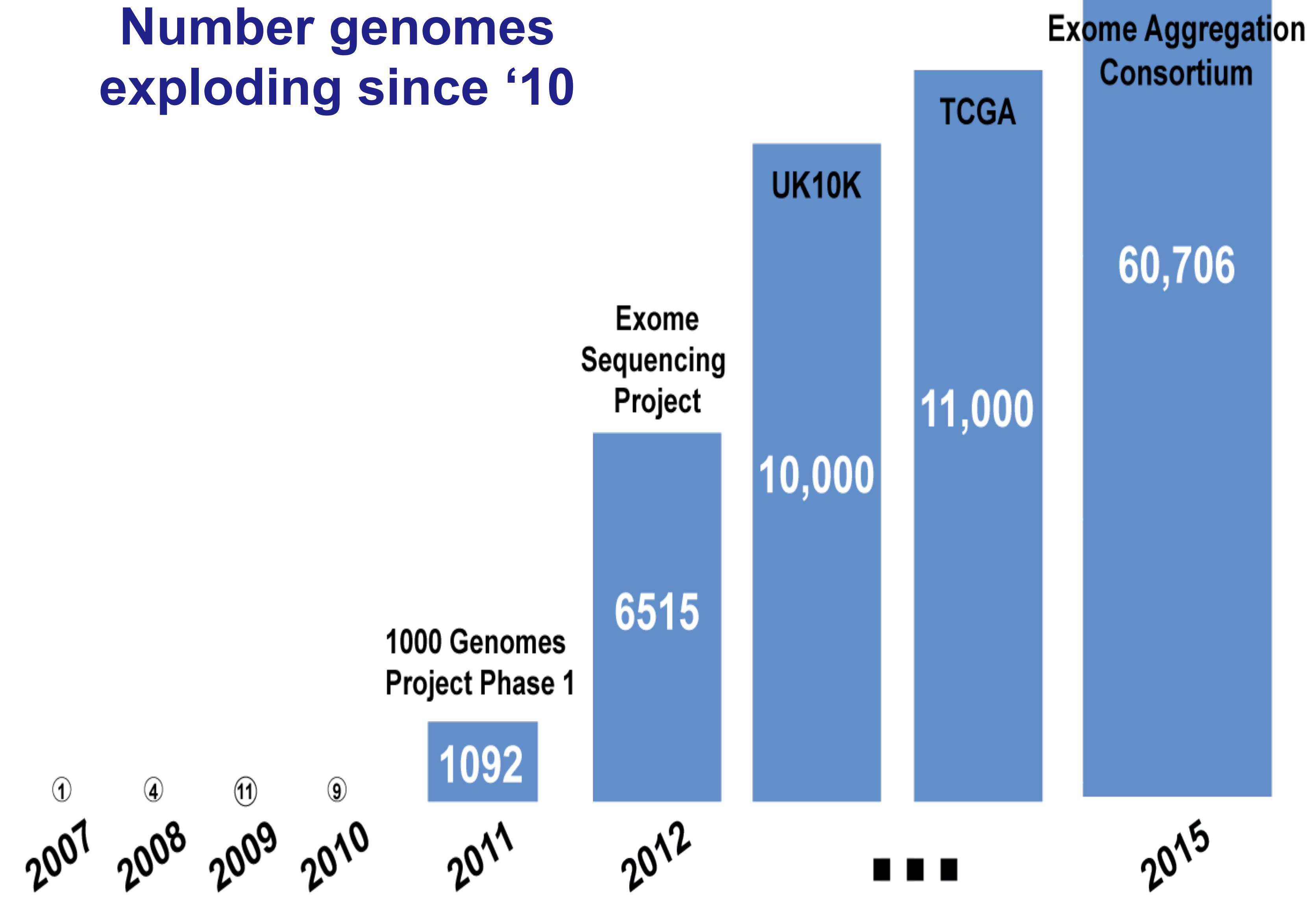
< 2 days

today:

2017

~\$530

~3 days



Curated Variant Data Resources as Backbone of Personalised Cancer Therapy

- cancer variant interpretation resources apply manual **data curation** and **bioinformatics** methods to provide information about putative targets and possible interventions

Database	Institute	Organized by
TARGET	BROAD	Gene
PCT	MD Anderson	Gene
cBioPortal / OncoKB	MSK	TCGA diseases
COSMIC	Sanger	Gene
IntOGen	University Pompeu Fabra	Gene
My Cancer Genome	Vanderbilt	Disease
ClViC	Washington University	Variant
DGIdb	Washington University	Drug/gene interaction

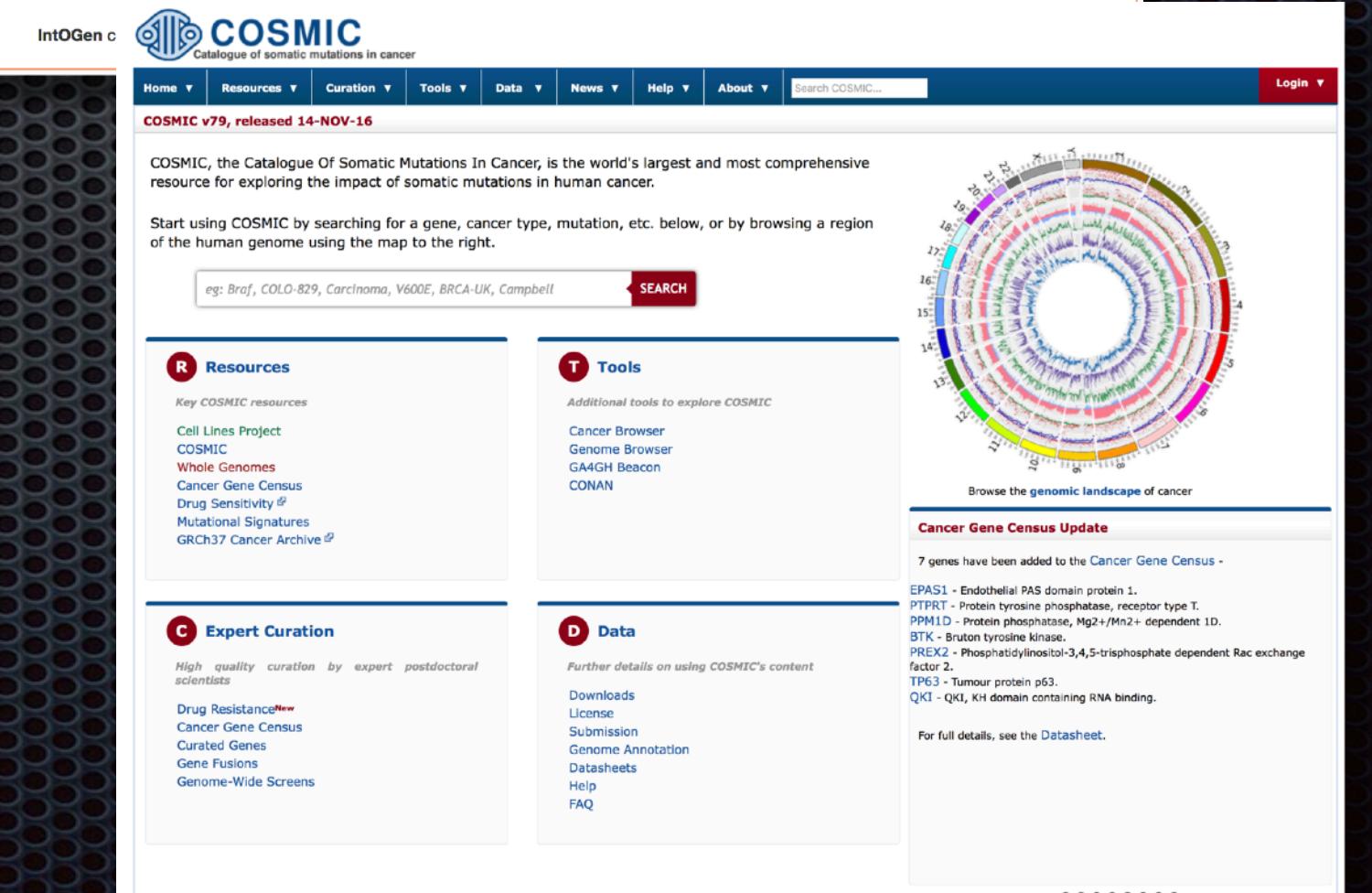
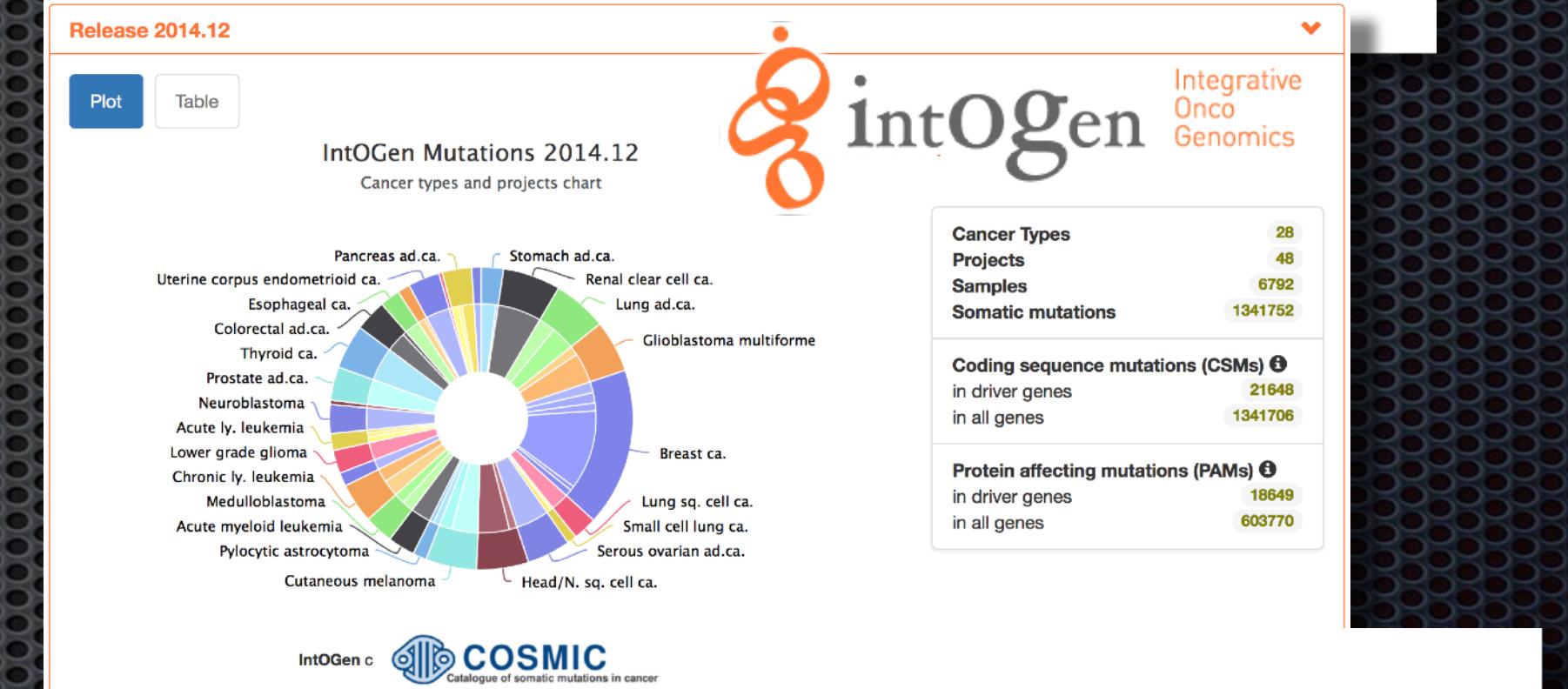
OncoKB Home About Team Levels of Evidence Actionable Genes Data Access News

OncoKB
Precision Oncology Knowledge Base
Annotation of Somatic Mutations in Cancer

418 Genes 3332 Variants 50 Tumor Types 71 Drugs

Search Gene

Level 1 FDA-approved Level 2 Standard-of-care Level 3 Clinical evidence



Genomes Everywhere

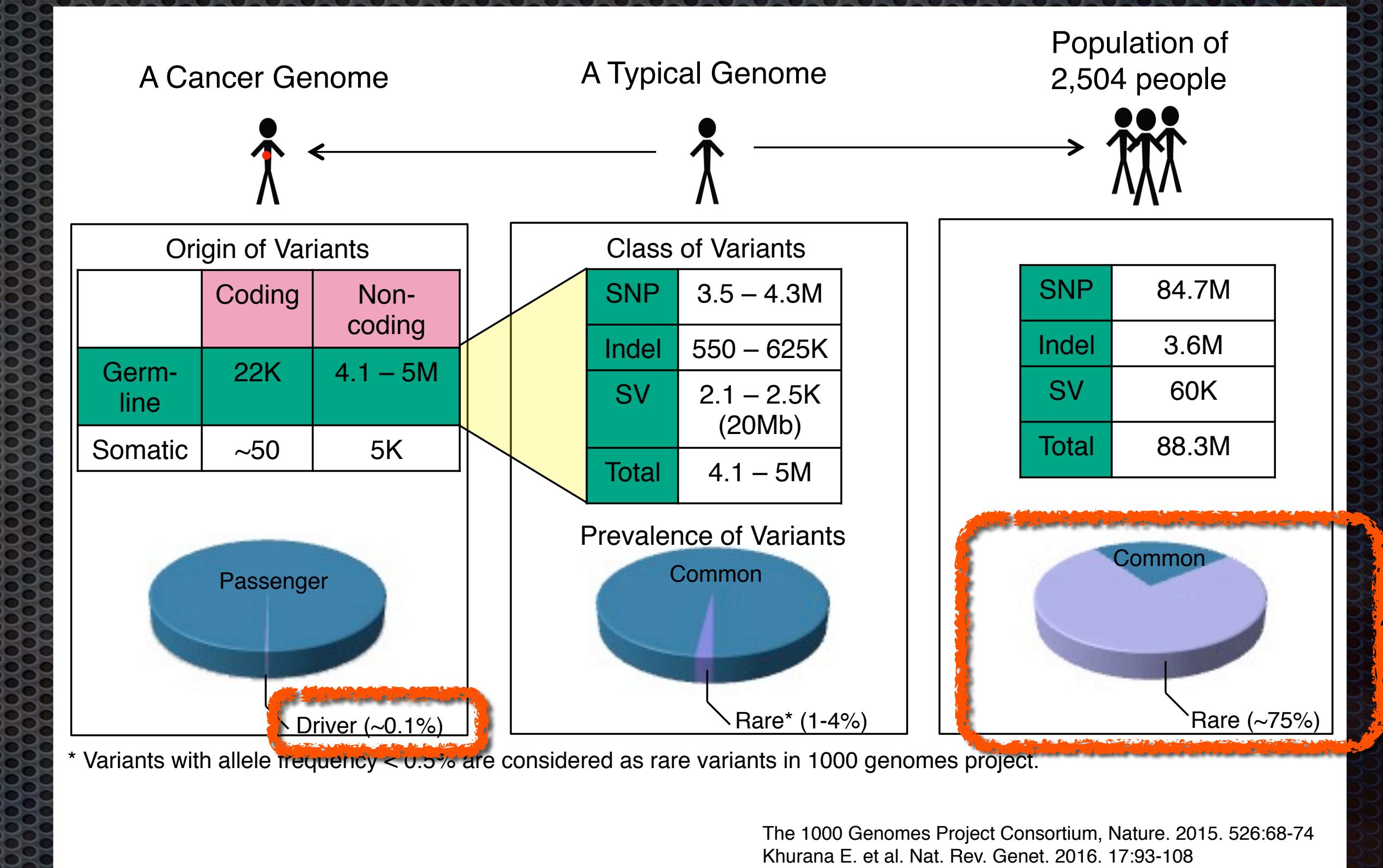
Organization / Initiative: Name	Organization / Initiative: Category	Cohort
100K Wellness Project	Research Project	107 unaffected individuals (scaling up to 100,000)
23andMe	Organization	>1 million customers (>80% consented to research)
Actionable Cancer Genome Initiative (ACGI)	Data-Sharing Project	Goal: 100,000 individuals
Ancestry.com	Organization	1.4 million customer DNA samples (what % consented to research?)
BioBank Japan	Repository	Specimens from >200,000 patients and unaffected controls
Cancer Moonshot2020	Consortium	Phase 1: 20,000 cancer patients
Children's Hospital of Philadelphia Biorepository	Repository	Capacity for 8.6 million samples
China Kadoorie Biobank	Repository	>512,000 participants (general population, China). Genotyping data available for ~100,000.
CIMBA	Consortium	>15,000 BRCA1 carriers, >8,000 BRCA2 carriers
Clinical Sequencing Exploratory Research (CSER)	Consortium	~4,000 patients and healthy controls
DECIPHER	Repository	19,014 patients (international)
deCode Genetics	Organization	500,000 participants (international)
East London Genes & Health	Research Project	100,000 unaffected individuals (East London, Pakistani or Bangladeshi heritage)
Electronic Medical Records and Genomics (eMERGE) Network	Repository, Consortium, Research Project	55,028 patients
European Network for Genetic and Genomic Epidemiology (ENGAGE)	Research Project	80,000 GWAS scans, and DNA and serum/plasma from >600,000 individuals
Exome Aggregation Consortium (ExAC)	Consortium	60,706 individuals
GENIE/AACR	Data-Sharing Project	>17,000 cancer patients (international)
Genome Asia 100K	Consortium	Goal: 100,000 individuals (Asia)
Genomics England	Organization	Goal: 100,000 genomes from 70,000 individuals (rare disease & cancer patients, and their relatives)
GoT2D	Consortium, Data-Sharing Project	Multiple case-control cohorts
International Cancer Genome Consortium (ICGC)	Consortium	currently data from >16'000 samples
International Genomics of Alzheimer's Project (IGAP)	Consortium	40,000 patients with Alzheimer's disease
International Multiple Sclerosis Genetics (IMSG) Consortium	Consortium	Goal: >50,000 patients with MS
Kaiser Permanente: Genes, Environment, and Health (RPGEH)	Repository, Research Project	200,000 DNA samples (scaling up to 500,000)
Leiden Open Variation Database (LOVD)	Repository	>170,000 individuals
Million Veteran Program	Research Project	Goal: 1 million individuals; first 200,000 is complete.
MyCode® Community Health Initiative	Repository, Research Project	Goal: >250,000 patients
Precision Medicine Initiative	Research Project	Goal: >1 million participants, starting in 2016 (US)
Psychiatric Genomics Consortium (PGC)	Consortium	>170,000 subjects
Resilience Project	Research Project	589,306 individuals
Saudi Human Genome Program	Research Project	Goal: ~100,000 patients and controls (Saudi Arabia)
Scottish Genomes Partnership (SGP)	Research Project	>3,000 individuals (Scotland)
T2D-GENES	Consortium, Data-Sharing Project	10,000 patients and controls (five ethnicities); 600 individuals (Mexican American)
TBResist	Consortium	>2,600 samples
UK Biobank	Repository, Consortium, Research Project	500,000 individuals (age 40-69 years; UK)
UK10K	Research Project	10,000 participants (6,000 patients and 4,000 controls)
Vanderbilt's BioVU	Repository	>215,000 samples

The trouble with human genome variation



Finding Somatic Mutations In Cancer: Many Needles in a Large Haystack

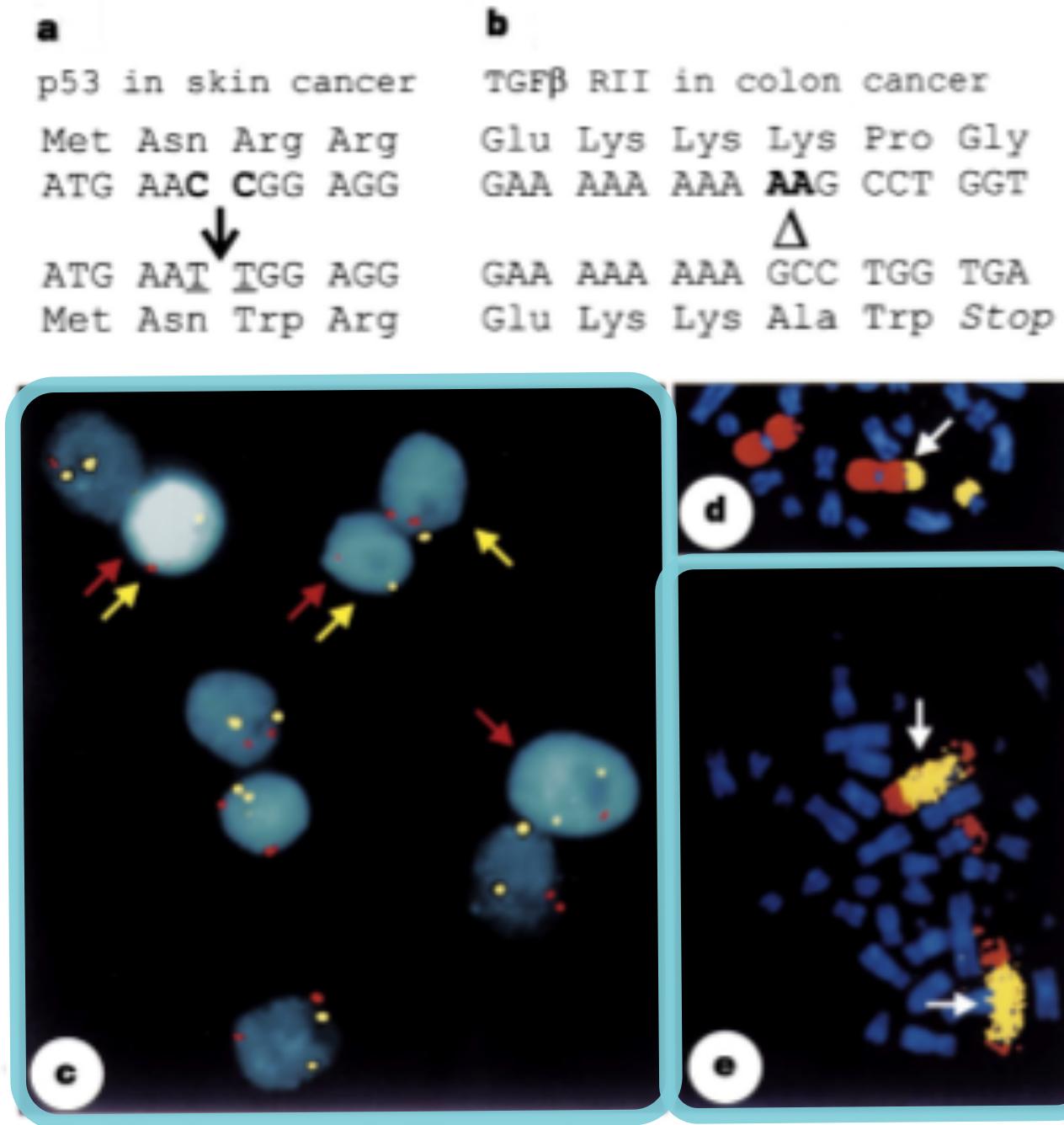
- a typical human genome (~3 billion base pairs) has ~5 million variants
- most of them are "**rare**"; i.e. can only be identified as recurring when sequencing thousands of people
- cancer cells accumulate additional variants, only **few** of which ("**drivers**") are relevant for the disease



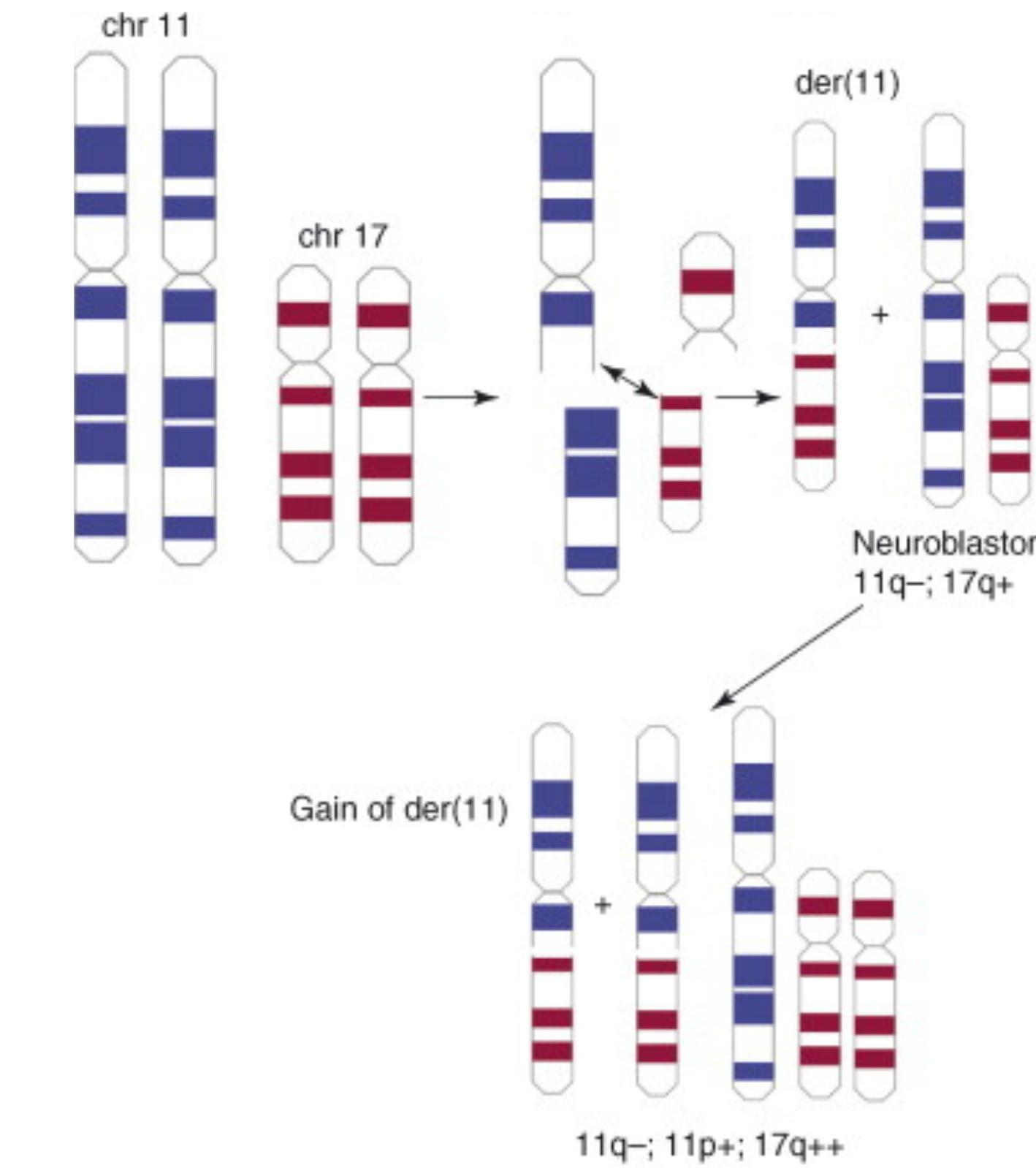
Graphic adapted from Mark Gerstein (GersteinLab.org; @markgerstein)

Mutations & genomic rearrangements in cancer

Lengauer et al, Genetic instabilities in human cancers. Nature (1998) vol. 396 (6712) pp. 643-9



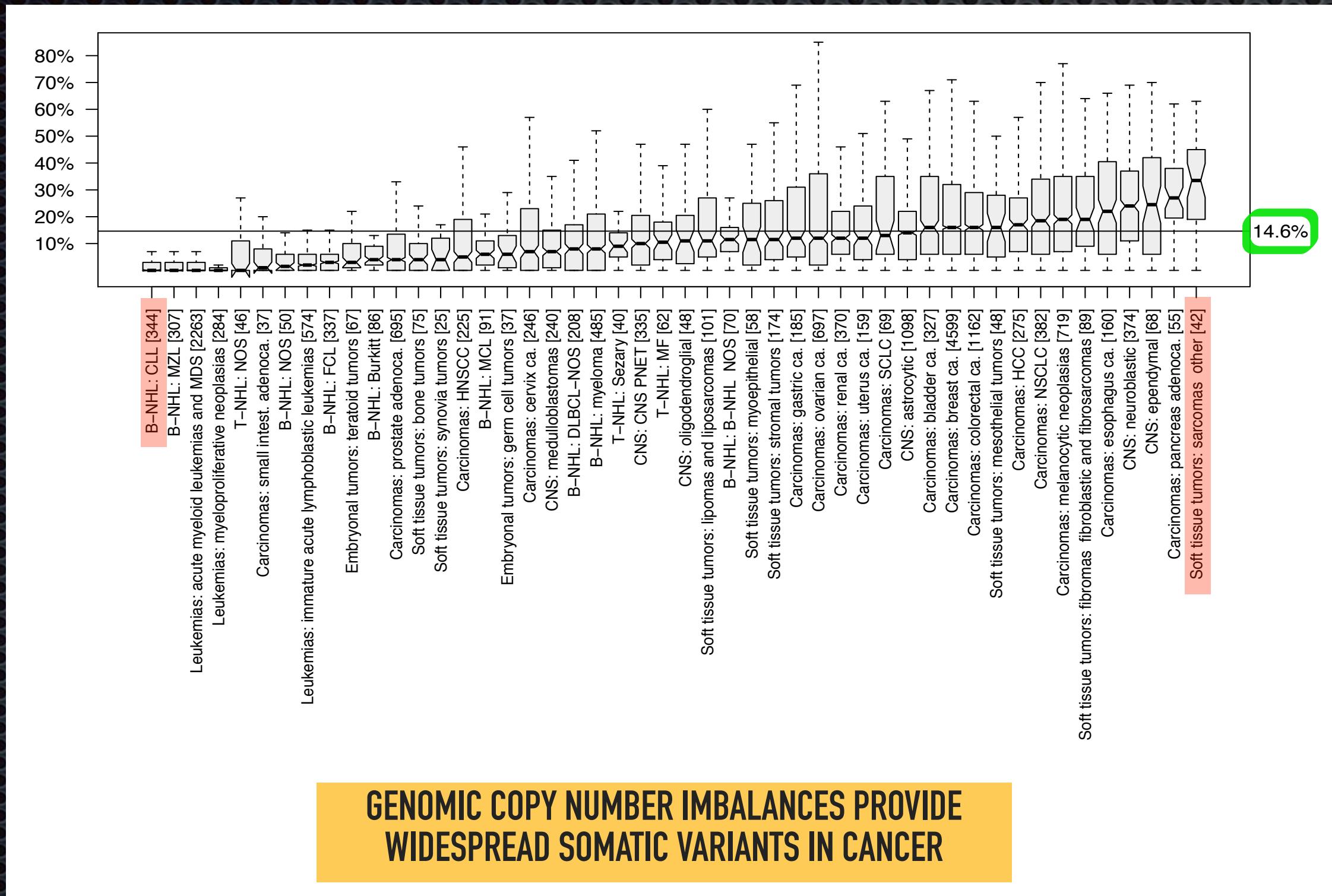
- a. small mutation (di-pyrimidine exchange at p53 in Xeroderma pigmentosum patient)
- b. two-base deletion in *TGFB* in a colorectal cancer patient with mismatch repair deficiency
- c. chromosomal losses (FISH; red=3, yellow=12) in CRC
- d. t(1;17) in neuroblastoma, whole-chromosomal painting
- e. *MYCN* gene amplification (multiple copies inserted into chromosome 1 derived marker)



Generation of copy number imbalances in cancer through imbalanced cytogenetic rearrangements - partial deletion of 11q, gain of 11pterq21 and 2 addl. copies of 17q

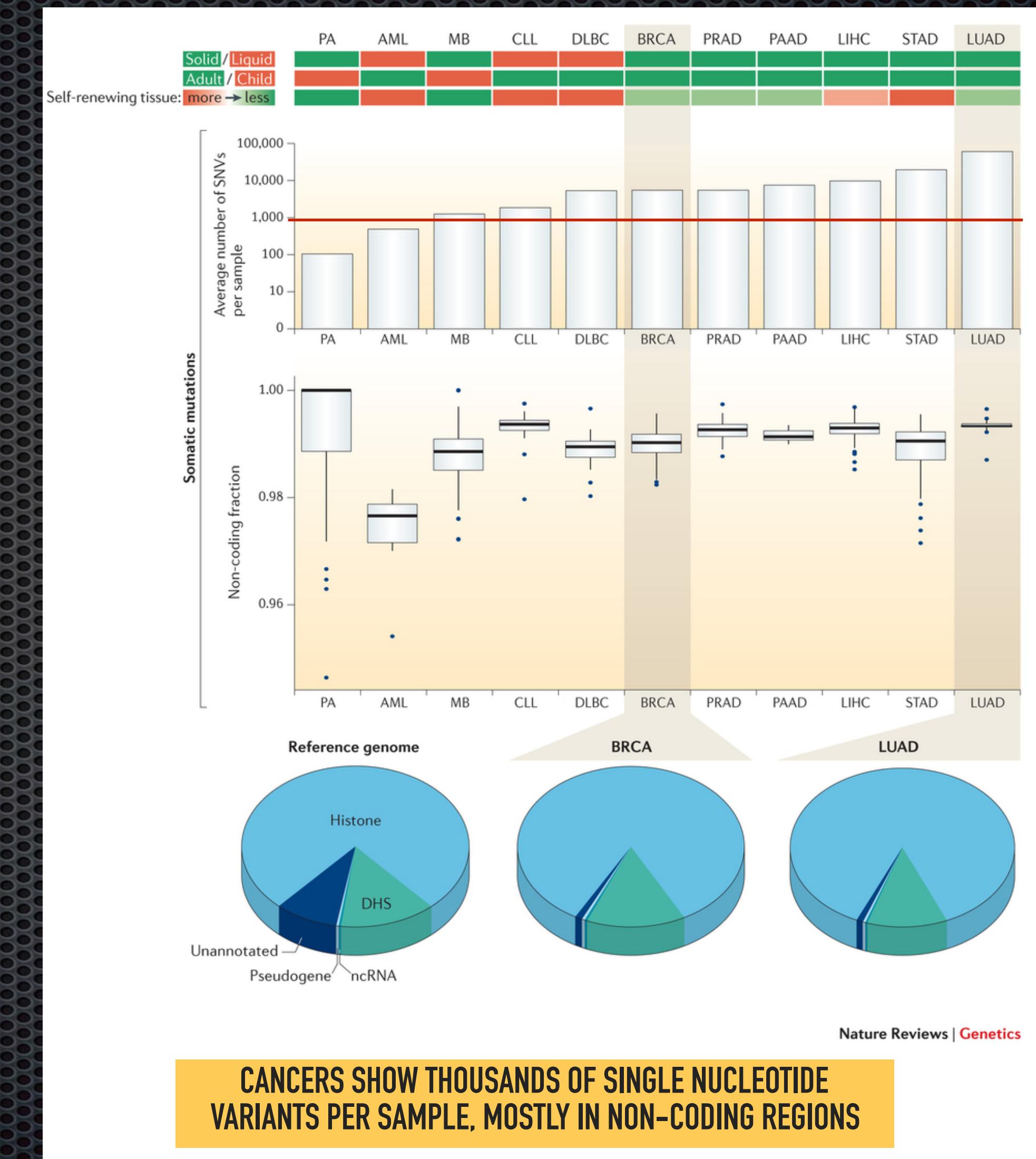
RL Stallings: Are chromosomal imbalances important in cancer? Volume 23, Issue 6, p278–283, 2007

Quantifying Somatic Mutations In Cancer



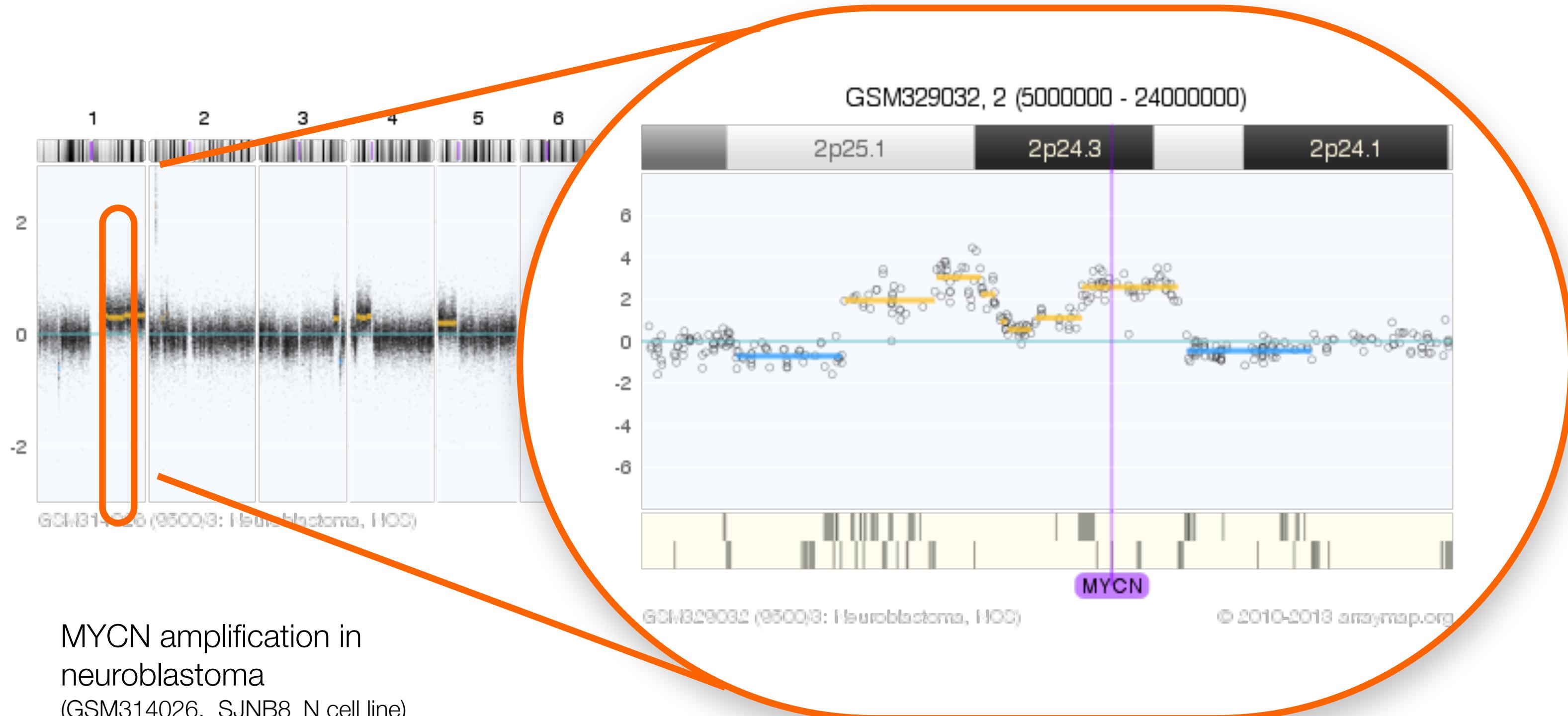
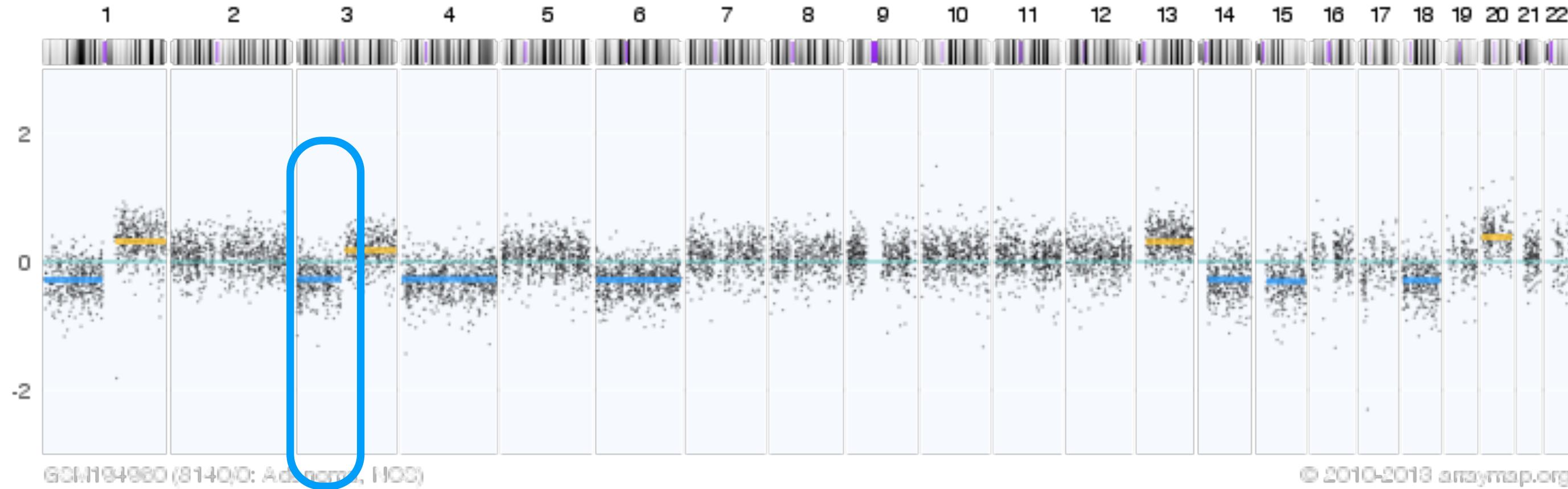
**GENOMIC COPY NUMBER IMBALANCES PROVIDE
WIDESPREAD SOMATIC VARIANTS IN CANCER**

On average ~15% of a cancer genome are in an imbalanced state (more/less than 2 alleles);
Original data based on >30'000 cancer genomes from arraymap.org



Pan-Cancer Analysis of Whole Genomes (PCAWG) data show widespread mutations in non-coding regions of cancer genomes (Khurana et al., Nat. Rev. Genet. (2016))

Genomic arrays: Many probes + bioinformatics determine copy number aberrations



low level/high level copy number alterations (CNAs)

arrayMap



Cancer genome data resources for research...

RESOURCES FOR CANCER GENOMICS

COSMIC
Catalogue of somatic mutations in cancer

Home ▾ Resources ▾ Curation ▾ Tools ▾ Data ▾ News ▾ Help ▾ About ▾ Search COSMIC... Login ▾

COSMIC v79, released 14-NOV-16

COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

Start using COSMIC by searching for a gene, cancer type, mutation, etc. below, or by browsing a region of the human genome using the map to the right.

eg: *Braf*, COLO-829, Carcinoma, V600E, BRCA-UK, Campbell **SEARCH**

R Resources

Key COSMIC resources

- Cell Lines Project
- COSMIC
- Whole Genomes
- Cancer Gene Census
- Drug Sensitivity
- Mutational Signatures
- GRCh37 Cancer Archive

T Tools

Additional tools to explore COSMIC

- Cancer Browser
- Genome Browser
- GA4GH Beacon
- CONAN

C Expert Curation

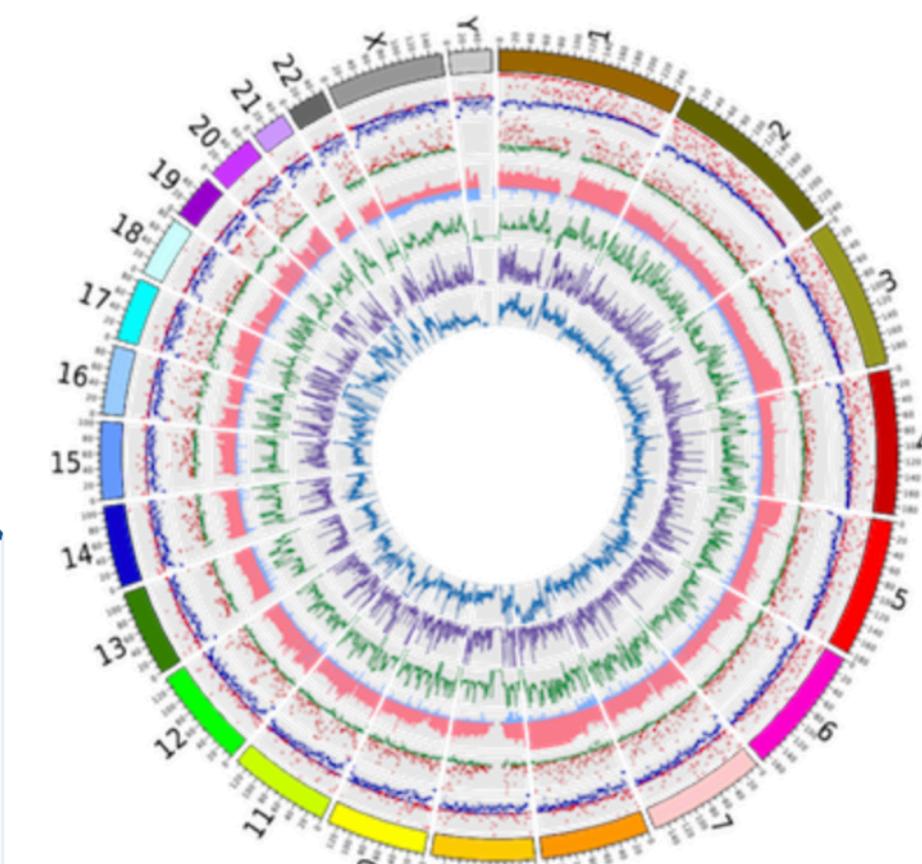
High quality curation by expert postdoctoral scientists

- Drug Resistance
- Cancer Gene Census
- Curated Genes
- Gene Fusions
- Genome-Wide Screens

D Data

Further details on using COSMIC's content

- Downloads
- License
- Submission
- Genome Annotation
- Datasheets
- Help
- FAQ



Browse the [genomic landscape of cancer](#)

Cancer Gene Census Update

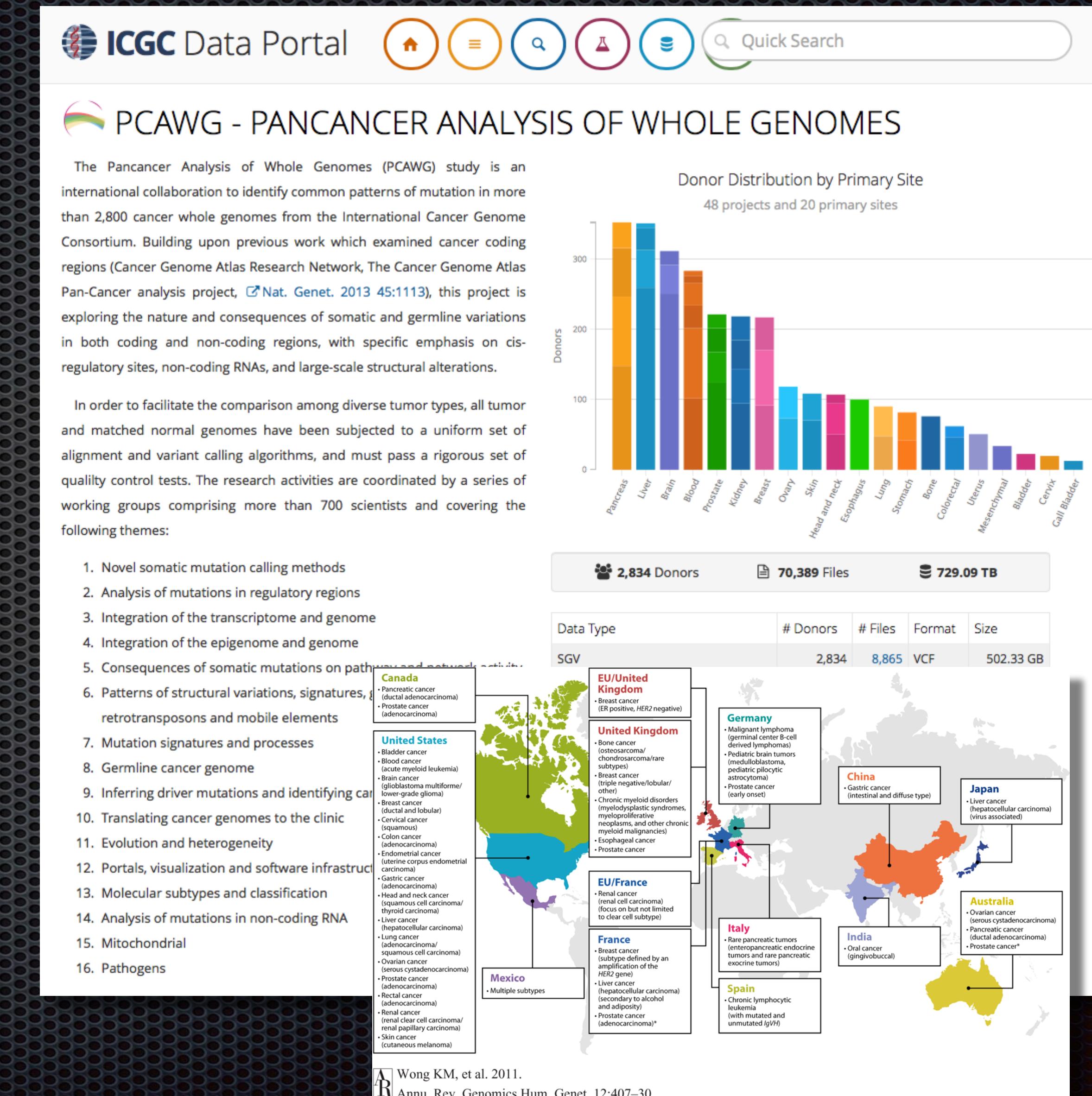
7 genes have been added to the [Cancer Gene Census](#) -

- EPAS1 - Endothelial PAS domain protein 1.
- PTPRT - Protein tyrosine phosphatase, receptor type T.
- PPM1D - Protein phosphatase, Mg²⁺/Mn²⁺ dependent 1D.
- BTK - Bruton tyrosine kinase.
- PREX2 - Phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 2.
- TP63 - Tumour protein p63.
- QKI - QKI, KH domain containing RNA binding.

For full details, see the [Datasheet](#).

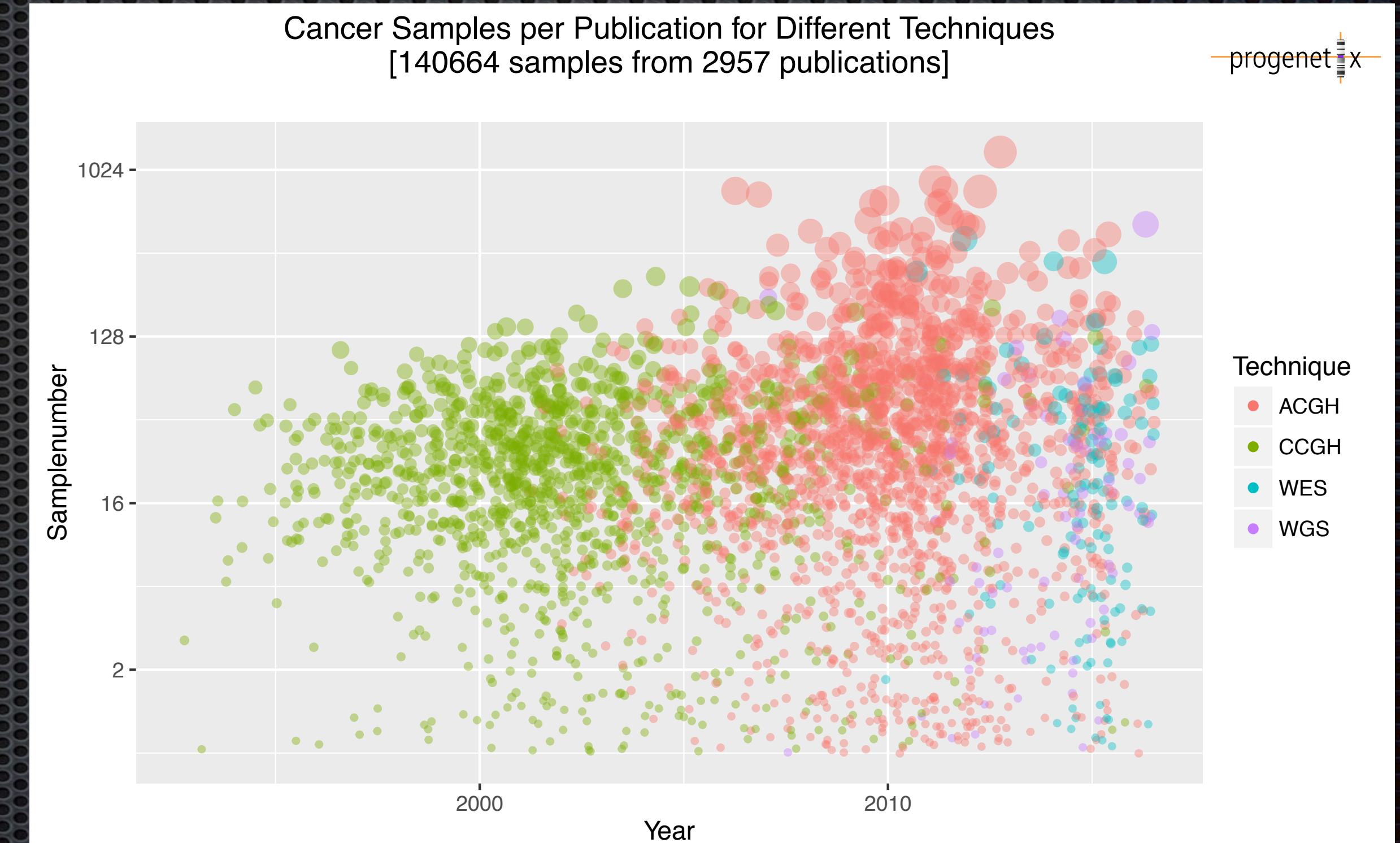
Genome-wide multi"omics" data generation for understanding tumor biology

- the International Cancer Genome Consortium (ICGC) as leading example of deep analysis of multiple cancer entities
- international collaboration of leading research centers for each of ~20 tumor types
- limitations:
 - focus on prominent cancer types w/ limited representation of rare entities
 - data access policies influenced by national regulations and legal frameworks
 - technical heterogeneity



Molecular Cytogenetic & Sequencing Studies for **Whole Genome Profiling** in Cancer

- genome screening to identify mutations in cancer samples
- for diagnostic purposes and therapeutic target identification
 - karyotyping (~1968)
 - Comparative Genomic Hybridization (1992)
 - genome **microarrays** (aCGH, SNP arrays ...; 1997)
 - Whole Exome Sequencing** (2010)
 - Whole Genome Sequencing** (2011)



Overview of publications reporting whole-genome screening analysis of cancer samples, by molecular-cytogenetic or genome sequencing methods. The data represents articles assessed for the progenetix.org cancer genome data resource (M. Baudis, 2001-2016)

Reference Resources for Cancer Genome Profiling

- continuously updated reference resources for cancer genome profiling data and related information
- basis for own research activities, collaborative projects and external use
- structured information serves for implementing GA4GH concepts



arrayMap



techniques	cCGH, aCGH, WES, WGS	aCGH (+?)
scope	sample (e.g. combination of several experiments)	experiment
content	>31000 samples	>60000 arrays
raw data presentation	no (link to sources if available)	yes (raw, log2, segmentation if available)
per sample re-analysis	no; supervised result (mostly as provided through publication)	yes (re-segmentation, thresholding, size filters ...)
final data	annotated/interpreted CN status for GP and cytogenetic regions	unsupervised CN status for GP and cytogenetic regions
main purposes	<ul style="list-style-type: none">Distribution of CNA target regions in most tumor types (>350 ICD-O)Cancer classification	<ul style="list-style-type: none">Gene specific hitsGenome feature correlation (fragile sites ...)

arrayMap

Resource for copy number variation data in cancer

arrayMap 

visualizing cancer genome array data @ arraymap.org

Search Samples
Search Publications
Gene CNA Frequencies
User Data
Array Visualization
Progenetix

 **University of Zurich** UZH

Citation
User Guide
Registration & Licensing
People
External Links ↗
FOLLOW US ON  [twitter](#)

 CC BY-SA

130.60.23.21

arrayMap is a curated reference database and bioinformatics resource targeting copy number profiling data in human cancer. The arrayMap database provides an entry point for meta-analysis and systems level data integration of high-resolution oncogenomic CNA data.

The current data reflects:

-  63060 genomic copy number arrays
-  763 experimental series
-  145 array platforms
-  141 ICD-O cancer entities
-  554 publications (Pubmed entries)



Genomic copy number imbalances on chromosome 9 in a case of Glioblastoma ([GSM491153](#)), indicating, among others, a homozygous deletion involving CDKN2A/B.

For the majority of the samples, probe level visualization as well as customized data representation facilitate gene level and genome wide data review. Results from multi-case selections can be connected to downstream data analysis and visualization tools, as provided through our Progenetix project.

arrayMap is developed by the group "Theoretical Cytogenetics and Oncogenomics" at the Institute of Molecular Life Sciences of University of Zurich.

BRAIN TUMOURS	5653 samples ↗
BREAST CANCER	8329 samples ↗
COLORECTAL CANCER	3238 samples ↗
PROSTATE CANCER	991 samples ↗
STOMACH CANCER	1062 samples ↗

ARRAYMAP NEWS

- 2016-08-03: SVG graphics**
- 2016-05-17: Transitioning to Europe PMC**
- [More news ...](#)

Feel free to use the data and tools for academic research projects and other applications. If more support and/or custom analysis is needed, please contact Michael Baudis regarding a collaborative project or a special license.

© 2000 - 2016 Michael Baudis, refreshed Mon, 19 Sep 2016 10:20:09 GMT in 6.87s on server 130.60.240.68. No responsibility is taken for the correctness of the data presented nor the results achieved with the Progenetix tools.

ta in cancer

ICD-O

Locus

HG18

HG19

FIND CNAS BY GENE OR REGION [ERBB2] 17:35097862-35138441:1 [?]

REGION SIZE | MAX COVERAGE (KB) 5000 250000 kb [?]

CLINICAL DATA

CITY [?]

Query Database

1949 of 65042 cases matched the selection criteria.

SUBSET	PERCENT IN SUBSET
8507/3: Invasive micropapillary carcinoma (13/39)	33.3
C692: retina (14/82)	17.1
8260/3: Papillary adenocarcinoma, NOS (11/65)	16.9
8500/3: Invasive carcinoma of no special type (1201/8188)	14.7
8560/3: Adenosquamous carcinoma (3/21)	14.3
Carcinomas: breast ca. (1254/6837)	14.2
C50: breast (1254/8929)	14.0
8500/2: Ductal carcinoma in situ, NOS (25/225)	11.1
C32: larynx (3/29)	10.3
8010/2: Carcinoma in situ, NOS (2/20)	10.0
C187: sigmoid incl. rectosigmoid junction (13/140)	9.3
8480/3: Mucinous adenocarcinoma (12/132)	9.1
8522/3: Infiltrating duct and lobular carcinoma (4/44)	9.1
8460/3: Micropapillary serous carcinoma [C56.9] (32/513)	6.2
8130/1: Urothelial papilloma, NOS (11/184)	6.0
C680: other urinary organs (11/184)	6.0
C54: corpus uteri (19/330)	5.8
8441/3: Serous adenocarcinoma, NOS (31/542)	5.7
Carcinomas: esophagus ca. (32/571)	5.6
Carcinomas: gastro ca. (80/1492)	5.4

9470/3: Medulloblastoma, NOS (M-94703)

9470/3

© 2010-2016 arraymap.org

9470/3: Medulloblastoma, NOS (M-94703)

PTCH1

© 2010-2016 arraymap.org

UID SERIESID PMID ICDMORPHOLOGYCODE ICDTOPOGRAPHYCODE

GSM1000061	GSE36942	23457519	8070/3	C10
GSM1000062	GSE36942	23457519	8070/3	C10
GSM1001316	GSE40777	23571474	8070/3	C53
GSM1001317	GSE40777	23571474	8010/3	C34
GSM1001318	GSE40777	23571474	8070/3	C09
GSM1001319	GSE40777	23571474	8010/3	C34
GSM1002668	GSE40834	24047479	9823/3	C42
GSM1002669	GSE40834	24047479	9823/3	C42
GSM1002670	GSE40834	24047479	9823/3	C42
GSM1002671	GSE40834	24047479	9823/3	C42
GSM1002672	GSE40834	24047479	9823/3	C42
GSM1002673	GSE40834	24047479	9823/3	C42
GSM1002674	GSE40834	24047479	9823/3	C42
GSM1002675	GSE40834	24047479	9823/3	C42
GSM1002676	GSE40834	24047479	9823/3	C42
GSM1002677	GSE40834	24047479	9823/3	C42
GSM1002678	GSE40834	24047479	9823/3	C42
GSM1002679	GSE40834	24047479	9823/3	C42
GSM1002680	GSE40834	24047479	9823/3	C42

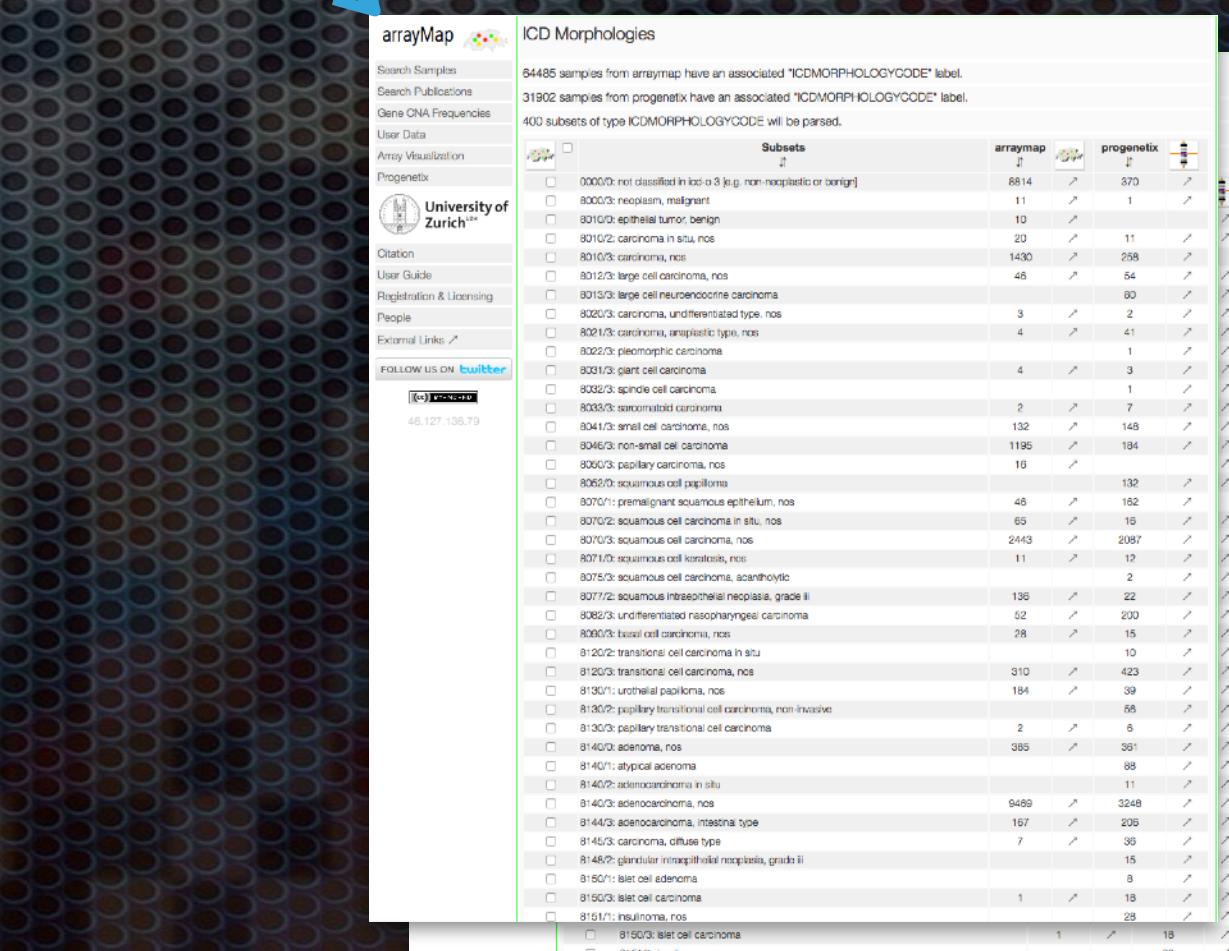
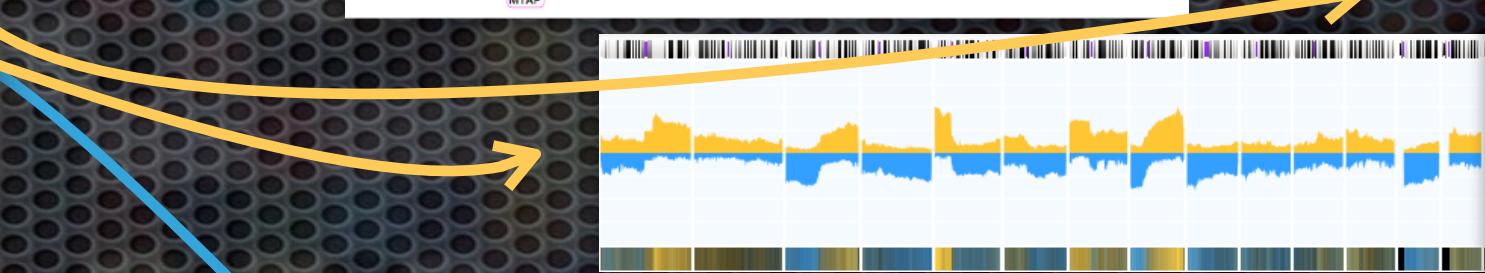
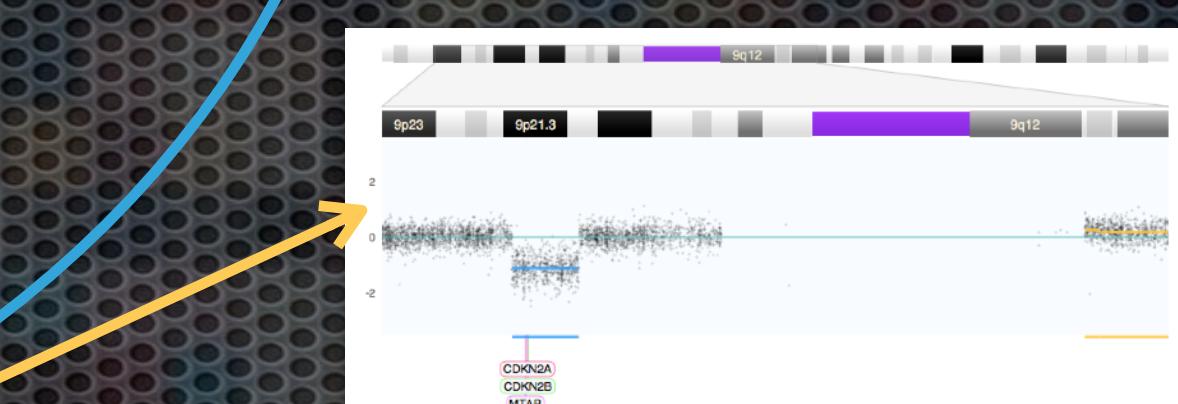
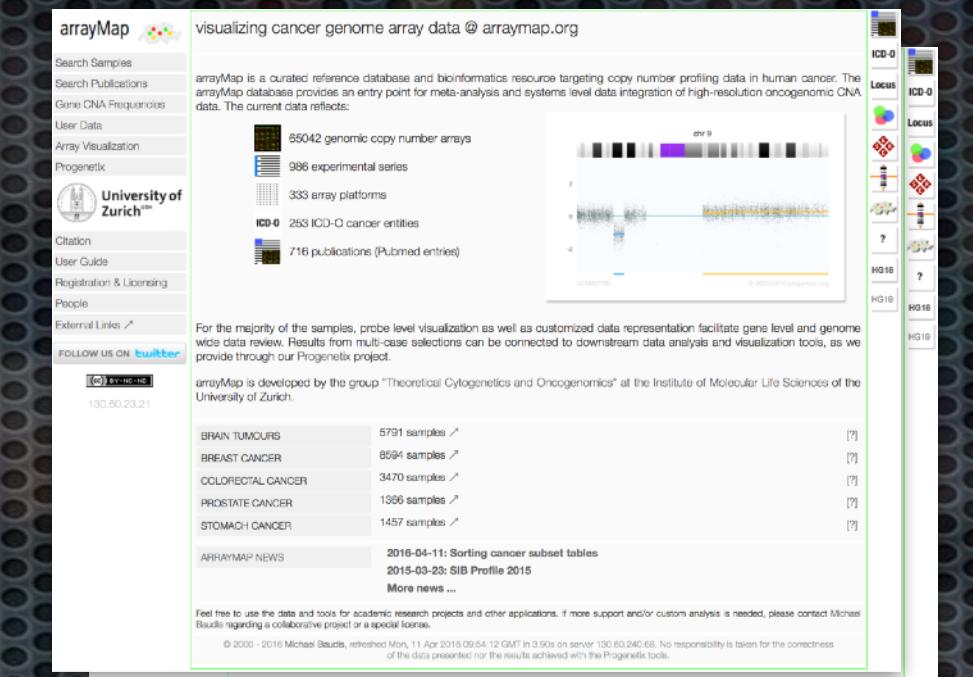
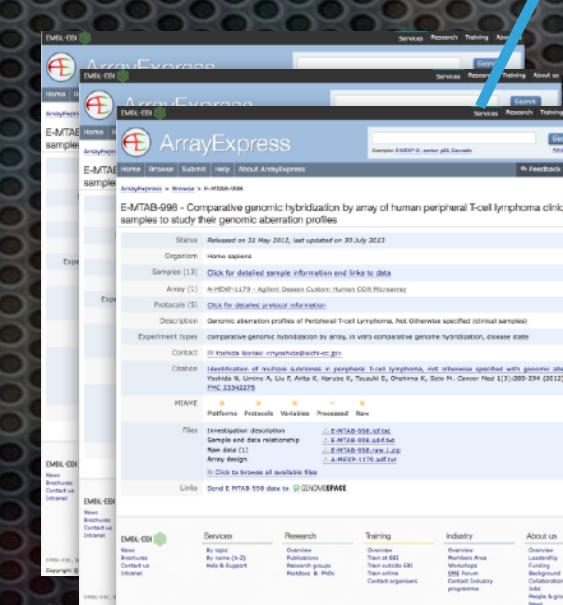
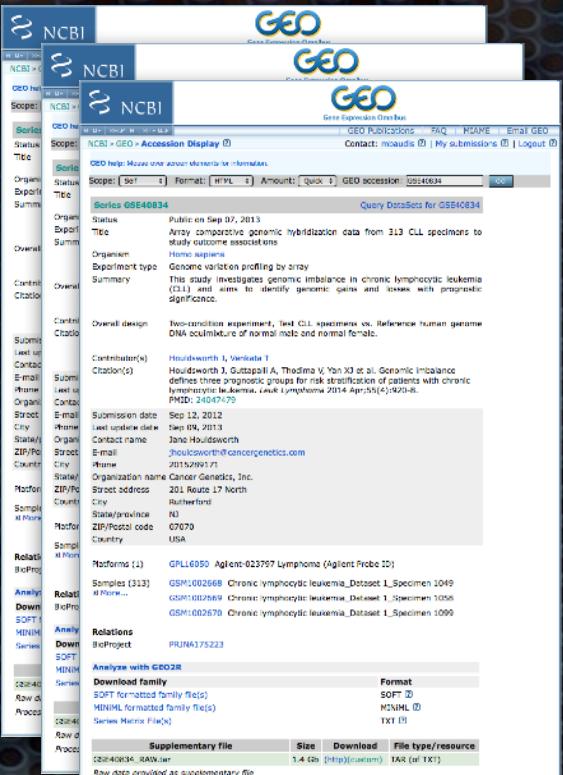
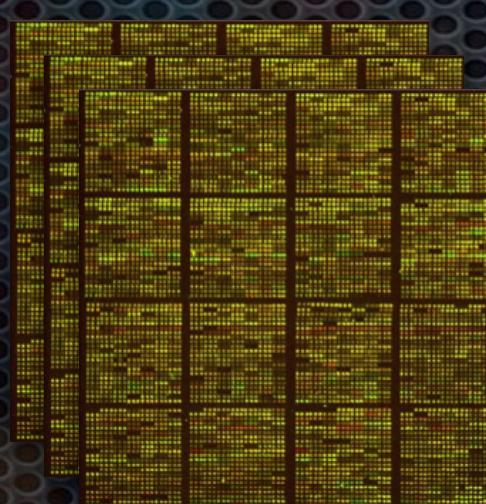
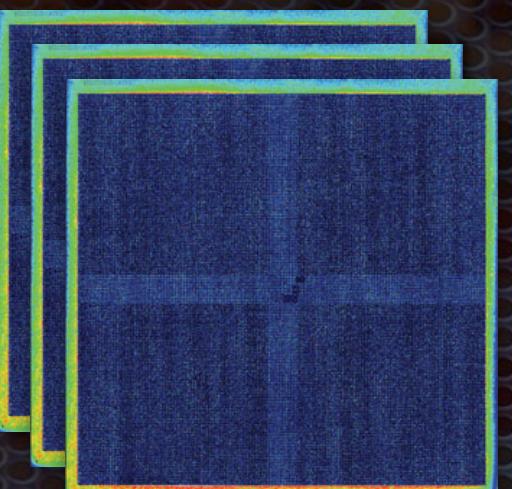
ARRAYMAP DATA PIPELINE

BIOCURATION

BIOINFORMATICS

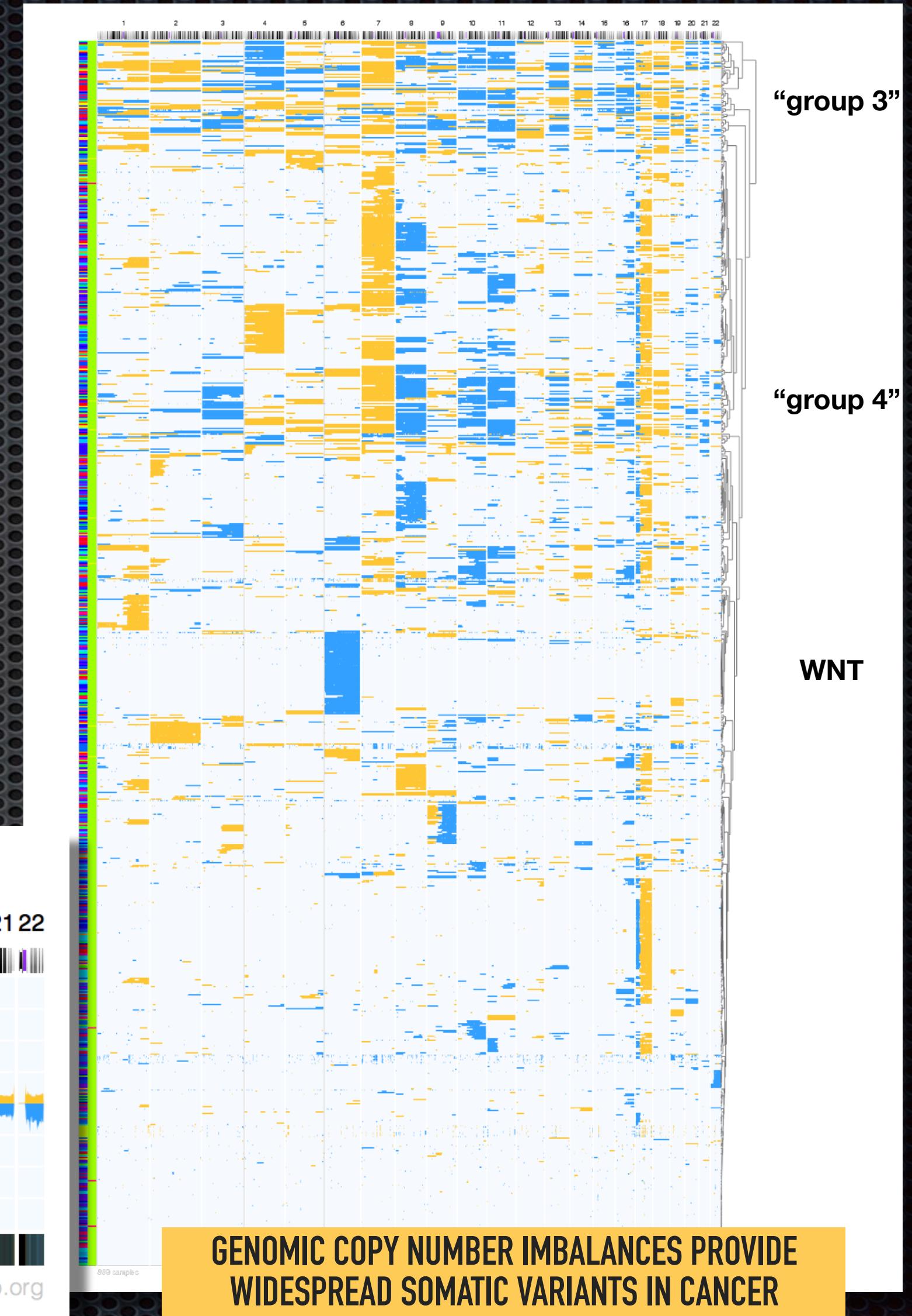
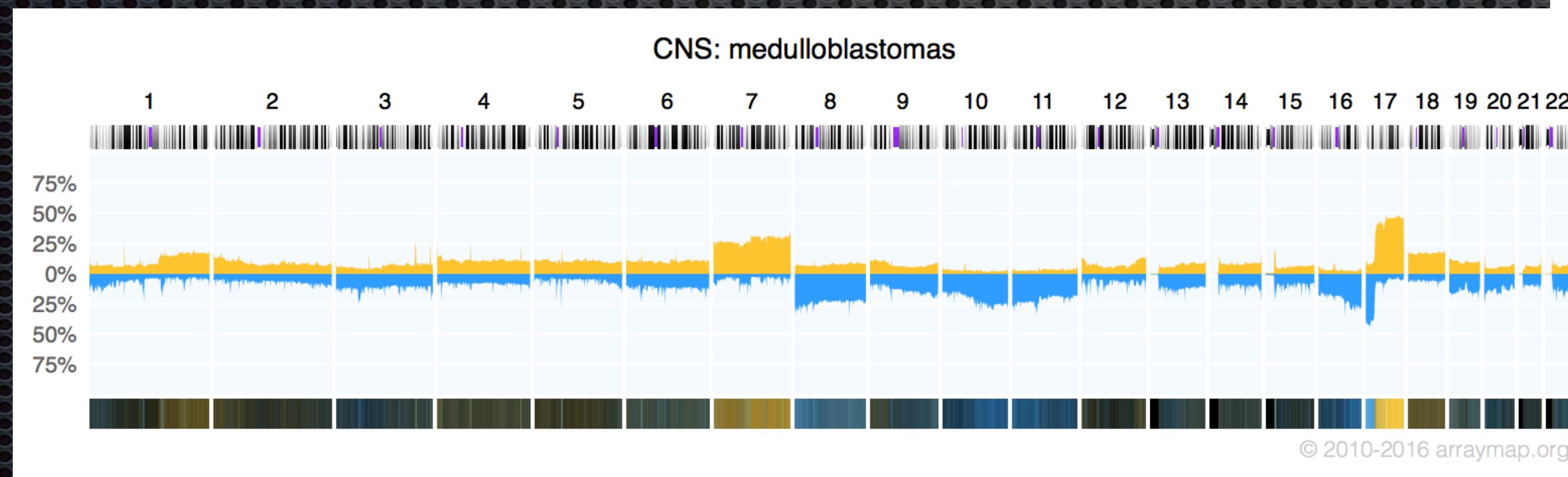


arrayMap



Somatic Mutations In Cancer: Patterns

- many tumor types express **recurrent mutation patterns**
- How can** those patterns be used for classification and determination of biological mechanisms?

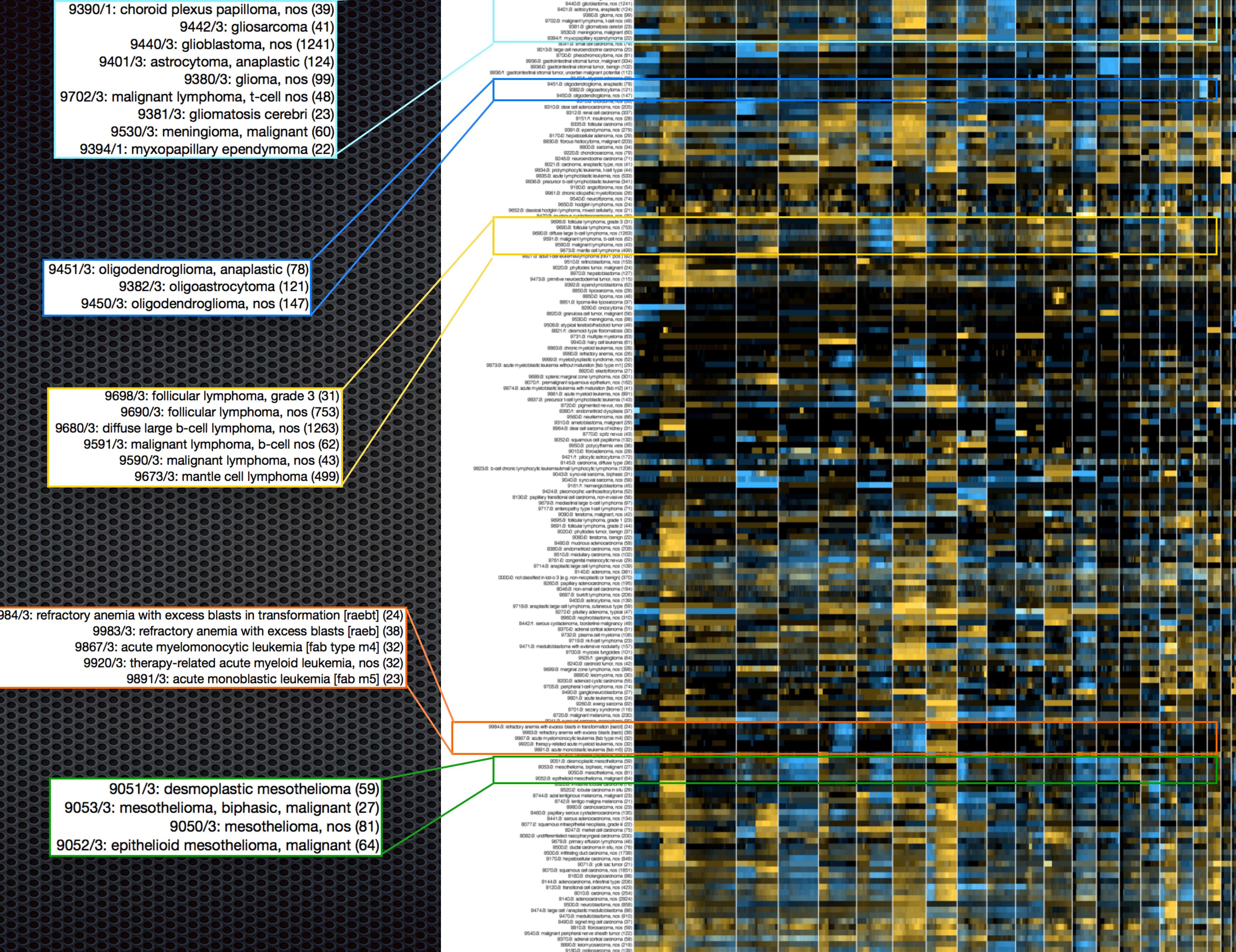


A genomic copy number histogram for malignant medulloblastomas, the most frequent type of pediatric brain tumors, displaying regions of genomic duplications and deletions. These can be decomposed into individual tumor profiles which segregate into several clusters of related mutation patterns with functional relevance and clinical correlation. From arraymap.org

Somatic Mutations In Cancer: Patterns III

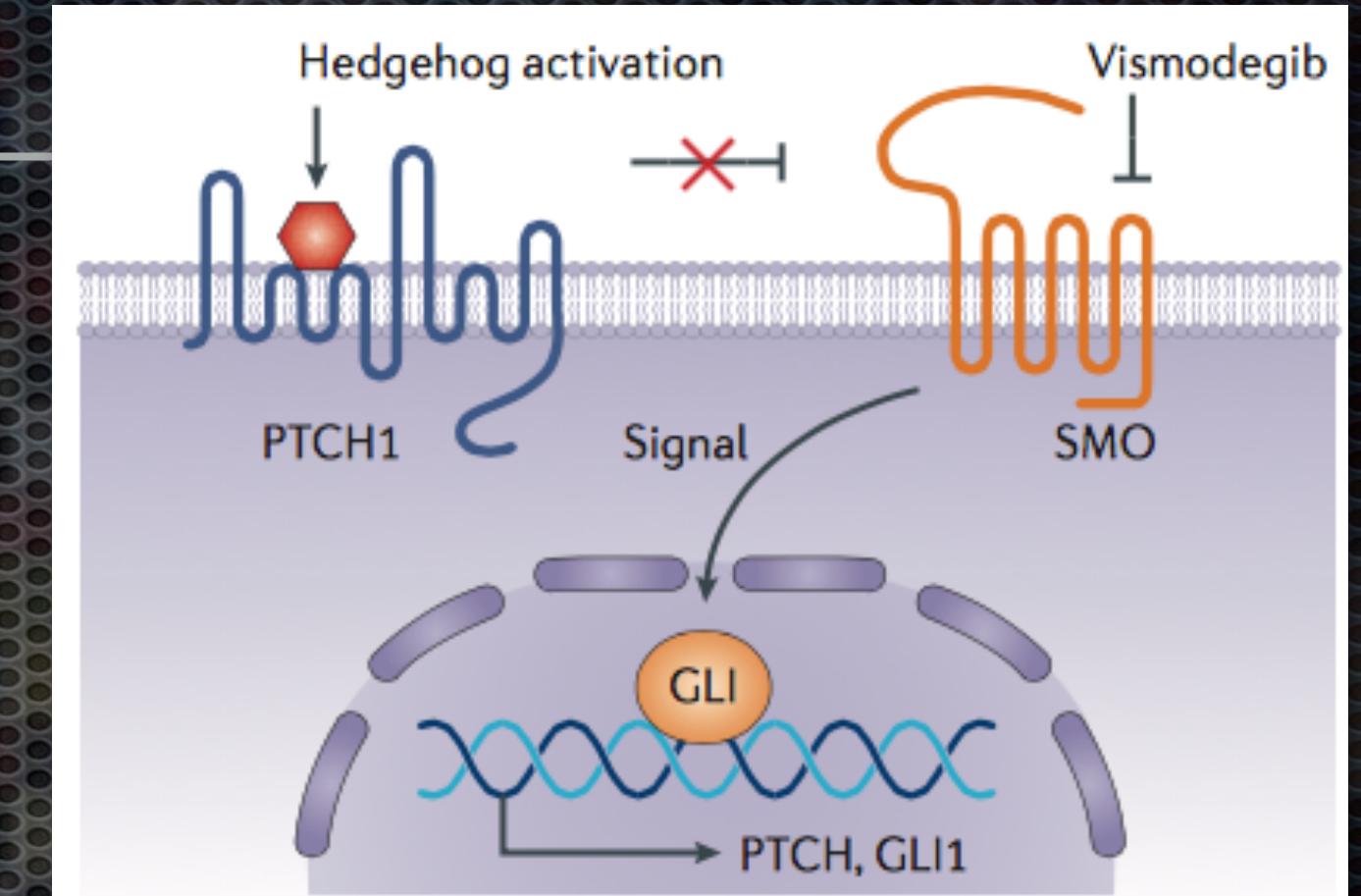
Making the case for genomic classifications

Some related cancer entities show similar copy number profiles

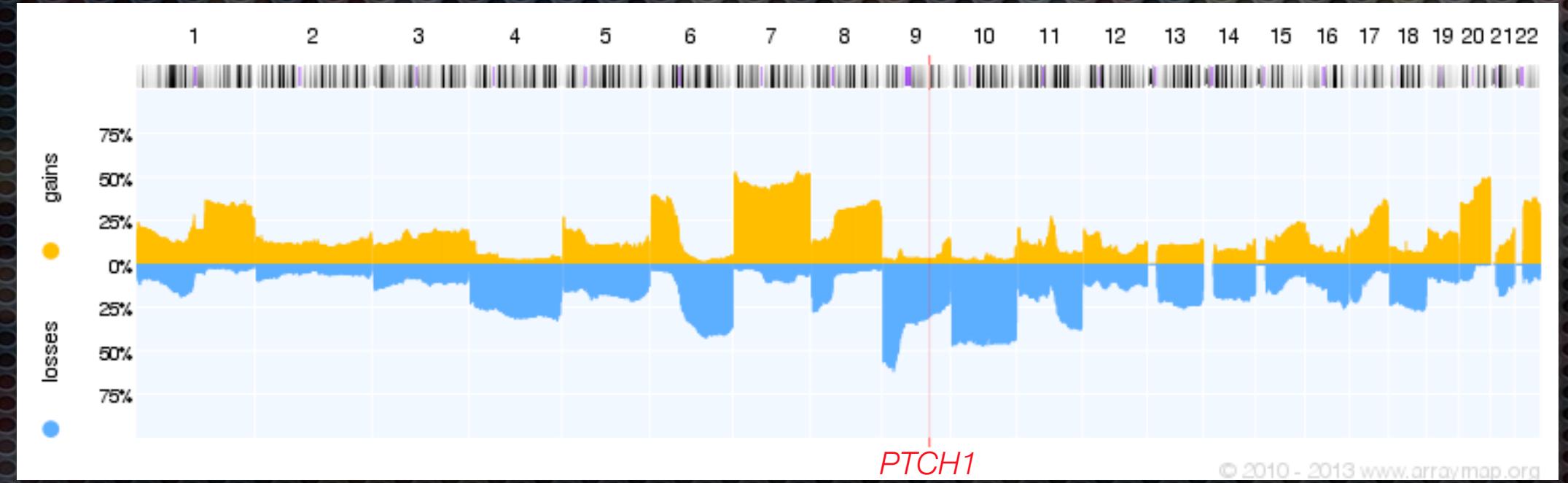


Large datasets for rare cancer gene events

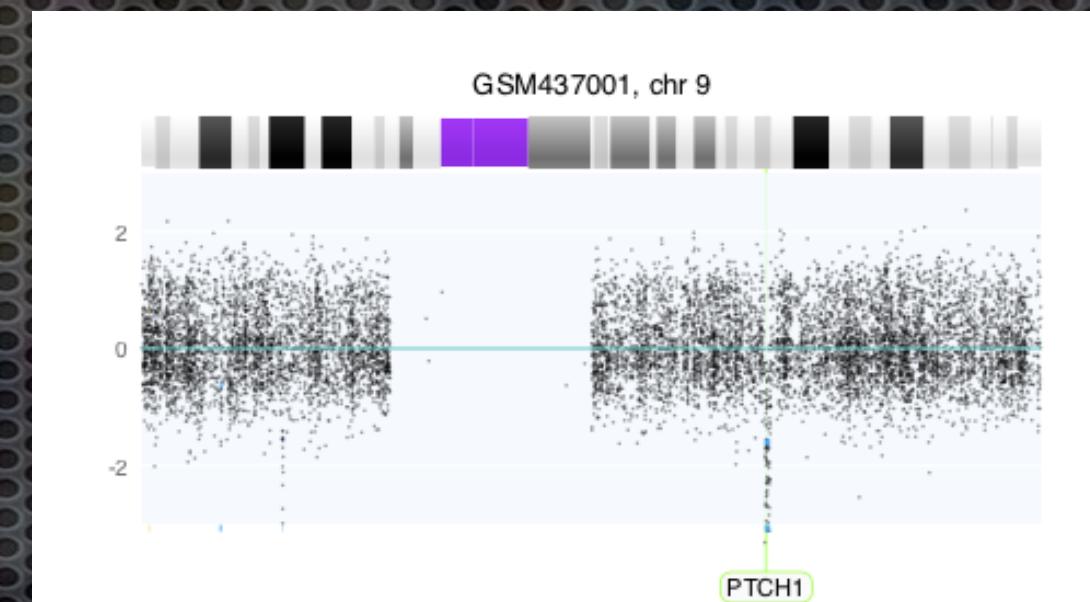
- The Sonic Hedgehog (SHH) pathway has become a “druggable” target in the therapy of syndromic and/or advanced basalomas (e.g. in Gorlin syndrome).
- In the pathway, PTCH1 acts as “tumor suppressor” counteracting SMO=>GLI mediated transcriptional activation.
- We were interested if the gene also could be involved in subsets of malignant melanomas ...



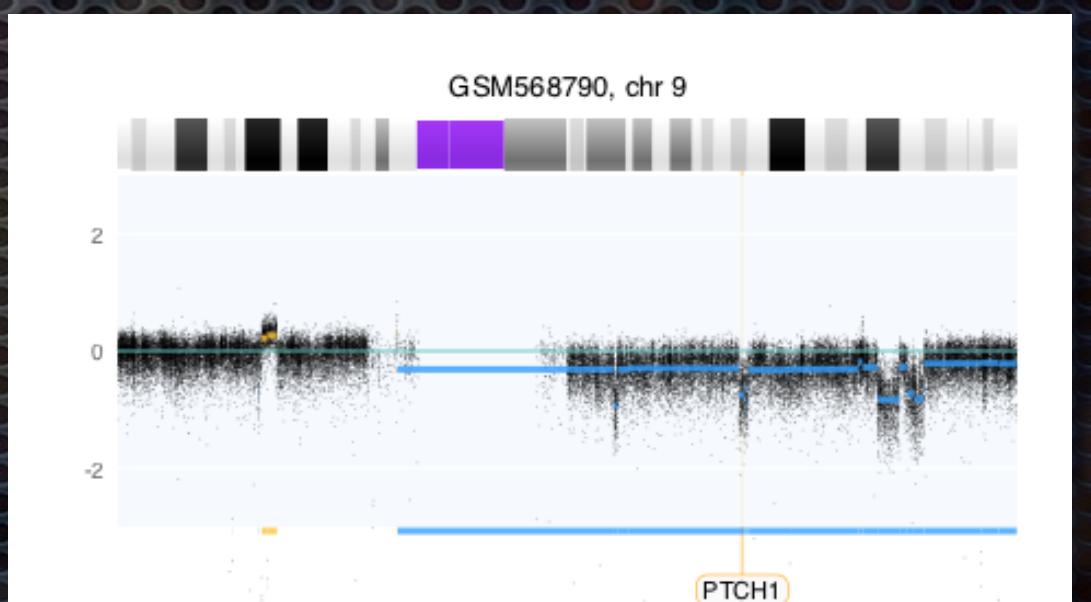
Dlugosz, A., Agrawal, S., & Kirkpatrick, P. (2012, June). Nature Reviews Drug Discovery, pp. 437–438



no “hot spot” (but 30% deletions)



probably pathogenic homozygous deletions in few cases
(3/~700): large datasets needed



Progenetix: Cancer Genome Profiles, Article Metrics, Epistemology, Resource Hub

cancer genome data @ progenetix.org

The Progenetix database provides an overview of copy number abnormalities in human cancer from currently **32317** array and chromosomal Comparative Genomic Hybridization (CGH) experiments, as well as Whole Genome or Whole Exome Sequencing (WGS, WES) studies. The data presented through Progenetix represents **364** different cancer types, according to the International classification of Diseases in Oncology (ICD-O).

Additionally, the website attempts to identify and present all publications (currently **2965** articles), referring to cancer genome profiling experiments. The database & software are developed by the group of Michael Baudis at the University of Zurich.

Publication data
2947 publications have been found.

New Search ...

1-250	251-500	501-750	751-1000	1001-1250	1251-1500	1501-1750	2751-2947	all
<input type="checkbox"/>								
CGH	aCGH	WES	WGS					

Technique

- ACGH
- CCGH
- WES
- WGS

PROGENETIX NEWS

- 2016-08-03: SVG graphics
- 2016-05-17: Transitioning to Eu...
- More news ...

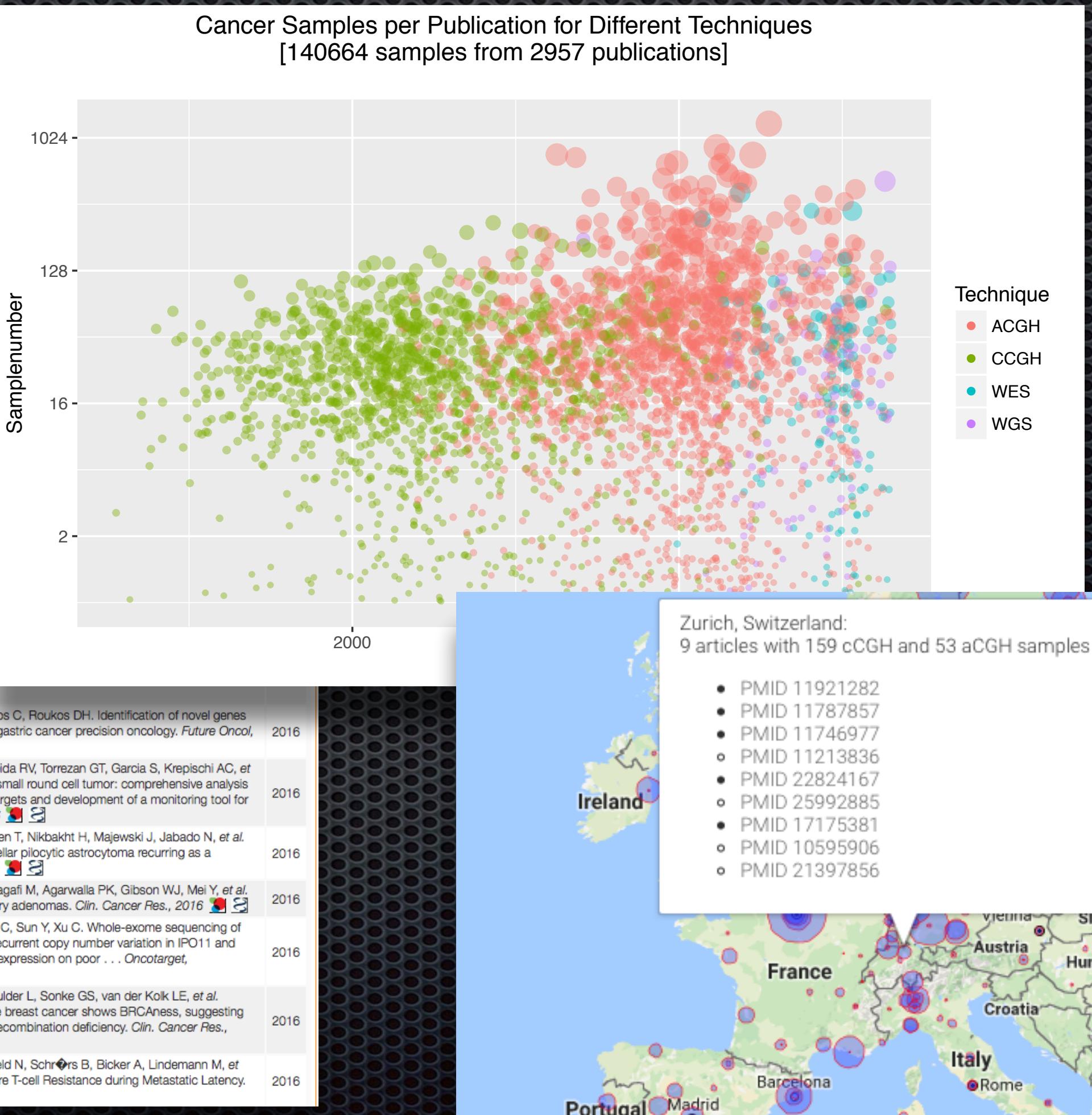
RELATED PUBLICATIONS

Google Scholar BETA

Search

Feel free to use the data and tools for academic research projects and other applications. Baudis regarding a collaborative project or a special license.

© 2000 - 2016 Michael Baudis, refreshed Sat, 17 Dec 2016 16:53:31 GMT in 5.63 of the data presented nor the results achieved.



Published online 12 November 2013 Nucleic Acids Research, 2014, Vol. 42, Database issue D105-D1062 doi:10.1093/nar/gkt510

Progenetix: 12 years of oncogenomic data curation

Hayoung Kim¹, Nitin Kumar^{1,2}, Ni Ai^{1,2}, Saumya Gupta^{1,2}, Priscilla Rata^{1,2} and Michael Baudis^{1,2*}

¹Institute of Molecular Life Sciences, University of Zurich, CH-8057 Zurich, Switzerland; ²Sisyphe Institute of Bioinformatics, University of Zurich, CH-8057 Zurich, Switzerland and ³Swiss Institute of Bioinformatics, University of Zurich, CH-8057 Zurich, Switzerland

Received August 21, 2013; Revised and Accepted October 21, 2013

ABSTRACT
DNA copy number aberrations (CNAs) can be found in the majority of cancer genomes and are crucial for understanding the molecular mechanisms of tumor formation and progression. Since the first release in 2001, the Progenetix project (<http://www.progenetix.org>) has been dedicated to provide the most comprehensive collection of CNAs from cancer genome projects. In this article, we present the latest version of Progenetix, which over the past 12 years our data curation efforts have resulted in the largest collection of cancer genome data presented through Progenetix. In addition, new features have been added. In particular, the gene set enrichment analysis has been extended to include various data representation options for providing users with improved search and reporting features. We report recent improvements of the database in terms of data completeness and other tools.

INTRODUCTION
DNA copy number aberrations (CNAs) are a form of genomic mutations found in the majority of individual cancer genomes. CNAs are often used to reveal both shared and distinct evolutionary processes in the same cancer type [1]. Understanding the role and importance of CNAs in the development of cancer is one of the main goals of oncogenomics [2,3].

* To whom correspondence should be addressed. E-mail: mbaudis@molbio.uzh.ch

© The Author(s) 2013. Published by Oxford University Press, *Nucleic Acids Research* 2014, 42, D105–D1062. This work is licensed under a Creative Commons Attribution Non-Commercial-ShareAlike 4.0 International License.

RESEARCH ARTICLE
DCCOA: A statistical method to define complexity dependence of co-occurring chromosomal aberrations

Nitin Kumar¹, Hubert Rehrauer², Hayoung Kim¹, Michael Baudis¹*

Abstract
Background: DNA copy number aberrations (CNAs) play a key role in cancer development and progression. Since most genes are located on chromosomes, it is important to study CNAs in the context of their chromosomal location. Testing methods for co-occurrence evaluate if two loci are not considered for each association due to the high genetic instability of many samples.

Method: We hypothesized that in cancer some linkage-independent CNAs may display a co-occurrence hypothesis. We developed a statistical method called Complexity Dependence of Co-occurring Chromosomal Aberrations (DCCOA) to test this hypothesis with a simulation-based algorithm.

Results: Application of DCCOA to example sets identified co-occurring CNA pairs from the complex background of cancer genomes. The identified pairs of co-occurring genes in their co-occurring changes can provide insights into cancer genome evolution.

Conclusion: We have developed a method to detect associations of regional copy number abnormalities in cancer genomes. Our results show that co-occurring CNAs are not randomly distributed but specifically co-occur in regions of regional CNAs, which may have negative impact on cancer development.

Background
Genetic alterations are an absolute requirement for malignant transformation. Both single and multiple genetic alterations and order of occurrence are important to understand cancer development. Recurrent CNAs have been identified in cancer genomes. Comparative Genomic Hybridization (CGH) [1,2] is a genome wide CGA approach that can detect changes in the genome throughout the last two decades. Building on the reverse engineering of CGH, array CGH (aCGH) [3] and whole genome sequencing (WGS) [4,5] have revolutionized the field of cancer genome analysis. The combination of CGH and WGS, genomic microarray technology (aCGH, CGH) [6,7], and short DNA sequences to derive regional copy number variations (RCNVs) [8,9] have been developed to analyze the complex changes in DNA copy number, which may have negative impact on cancer development.

Comments
Received: 10 June 2013; revised: 10 September 2013; accepted: 10 October 2013
Published online 12 November 2013 in *Nucleic Acids Research*. DOI: 10.1093/nar/gkt510
© 2013 The Author(s). *Nucleic Acids Research* published by Oxford University Press.
This is an open-access article distributed under the terms of the Creative Commons
Attribution Non-Commercial-ShareAlike 4.0 International License.

Background
Genetic alterations are an absolute requirement for malignant transformation. Both single and multiple genetic alterations and order of occurrence are important to understand cancer development. Recurrent CNAs have been identified in cancer genomes. Comparative Genomic Hybridization (CGH) [1,2] is a genome wide CGA approach that can detect changes in the genome throughout the last two decades. Building on the reverse engineering of CGH, array CGH (aCGH) [3] and whole genome sequencing (WGS) [4,5] have revolutionized the field of cancer genome analysis. The combination of CGH and WGS, genomic microarray technology (aCGH, CGH) [6,7], and short DNA sequences to derive regional copy number variations (RCNVs) [8,9] have been developed to analyze the complex changes in DNA copy number, which may have negative impact on cancer development.

Comments
Received: 10 June 2013; revised: 10 September 2013; accepted: 10 October 2013
Published online 12 November 2013 in *Nucleic Acids Research*. DOI: 10.1093/nar/gkt510
© 2013 The Author(s). *Nucleic Acids Research* published by Oxford University Press.
This is an open-access article distributed under the terms of the Creative Commons
Attribution Non-Commercial-ShareAlike 4.0 International License.

through the formation of oncogenic fusion genes, and directly promote cancer growth and metastasis [9–11]. In addition, CNAs are also involved in the formation of cancer subtypes and are crucial for understanding the molecular mechanisms of tumor formation and progression. Since the first release in 2001, the Progenetix project (<http://www.progenetix.org>) has been dedicated to provide the most comprehensive collection of CNAs from cancer genome projects. In this article, we present the latest version of Progenetix, which over the past 12 years our data curation efforts have resulted in the largest collection of cancer genome data presented through Progenetix. In addition, new features have been added. In particular, the gene set enrichment analysis has been extended to include various data representation options for providing users with improved search and reporting features. We report recent improvements of the database in terms of data completeness and other tools.

INTRODUCTION
DNA copy number aberrations (CNAs) are a form of genomic mutations found in the majority of individual cancer genomes. CNAs are often used to reveal both shared and distinct evolutionary processes in the same cancer type [1]. Understanding the role and importance of CNAs in the development of cancer is one of the main goals of oncogenomics [2,3].

* To whom correspondence should be addressed. E-mail: mbaudis@molbio.uzh.ch

© The Author(s) 2013. Published by Oxford University Press, *Nucleic Acids Research* 2014, 42, D105–D1062. This work is licensed under a Creative Commons Attribution Non-Commercial-ShareAlike 4.0 International License.

RESEARCH ARTICLE
DCCOA: A statistical method to define complexity dependence of co-occurring chromosomal aberrations

Nitin Kumar¹, Hubert Rehrauer², Hayoung Kim¹, Michael Baudis¹*

Abstract
Background: DNA copy number aberrations (CNAs) play a key role in cancer development and progression. Since most genes are located on chromosomes, it is important to study CNAs in the context of their chromosomal location. Testing methods for co-occurrence evaluate if two loci are not considered for each association due to the high genetic instability of many samples.

Method: We hypothesized that in cancer some linkage-independent CNAs may display a co-occurrence hypothesis. We developed a statistical method called Complexity Dependence of Co-occurring Chromosomal Aberrations (DCCOA) to test this hypothesis with a simulation-based algorithm.

Results: Application of DCCOA to example sets identified co-occurring CNA pairs from the complex background of cancer genomes. The identified pairs of co-occurring genes in their co-occurring changes can provide insights into cancer genome evolution.

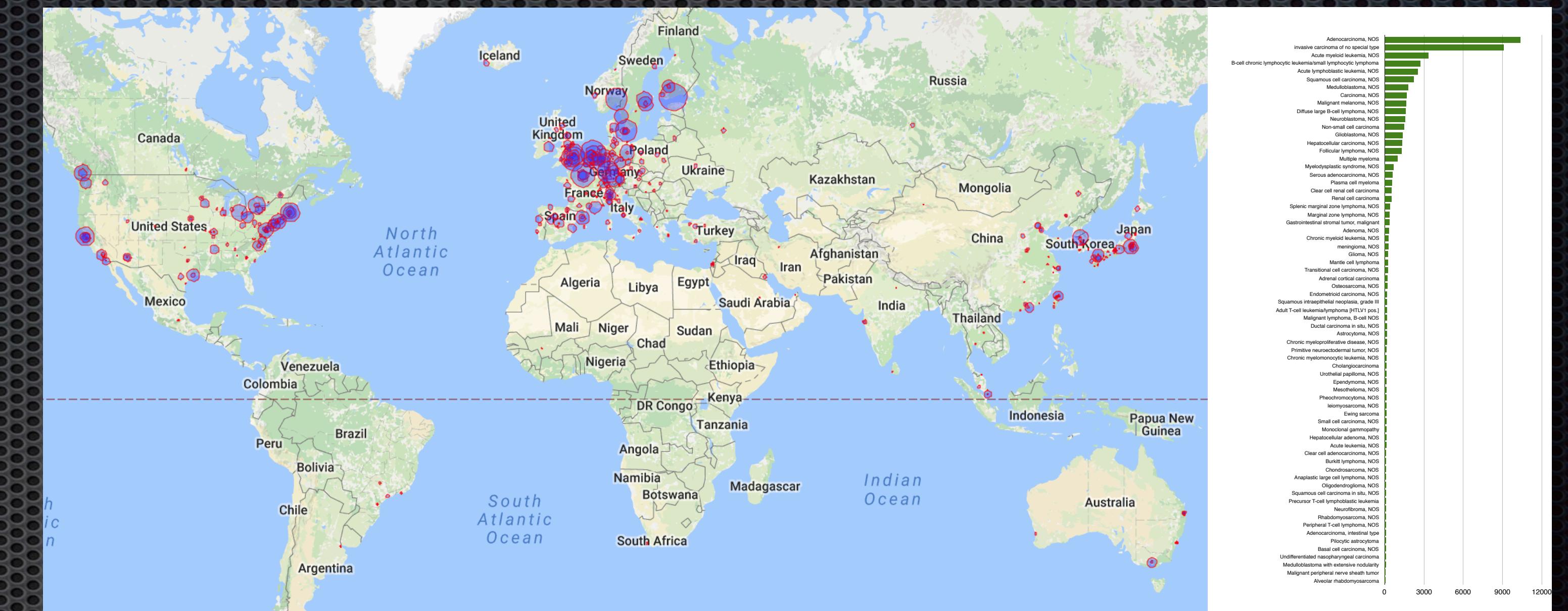
Conclusion: We have developed a method to detect associations of regional copy number abnormalities in cancer genomes. Our results show that co-occurring CNAs are not randomly distributed but specifically co-occur in regions of regional CNAs, which may have negative impact on cancer development.

Background
Genetic alterations are an absolute requirement for malignant transformation. Both single and multiple genetic alterations and order of occurrence are important to understand cancer development. Recurrent CNAs have been identified in cancer genomes. Comparative Genomic Hybridization (CGH) [1,2] is a genome wide CGA approach that can detect changes in the genome throughout the last two decades. Building on the reverse engineering of CGH, array CGH (aCGH) [3] and whole genome sequencing (WGS) [4,5] have revolutionized the field of cancer genome analysis. The combination of CGH and WGS, genomic microarray technology (aCGH, CGH) [6,7], and short DNA sequences to derive regional copy number variations (RCNVs) [8,9] have been developed to analyze the complex changes in DNA copy number, which may have negative impact on cancer development.

Comments
Received: 10 June 2013; revised: 10 September 2013; accepted: 10 October 2013
Published online 12 November 2013 in *Nucleic Acids Research*. DOI: 10.1093/nar/gkt510
© 2013 The Author(s). *Nucleic Acids Research* published by Oxford University Press.
This is an open-access article distributed under the terms of the Creative Commons
Attribution Non-Commercial-ShareAlike 4.0 International License.

Bias in Ascertainment / Background / Environment in Cancer Genome Studies

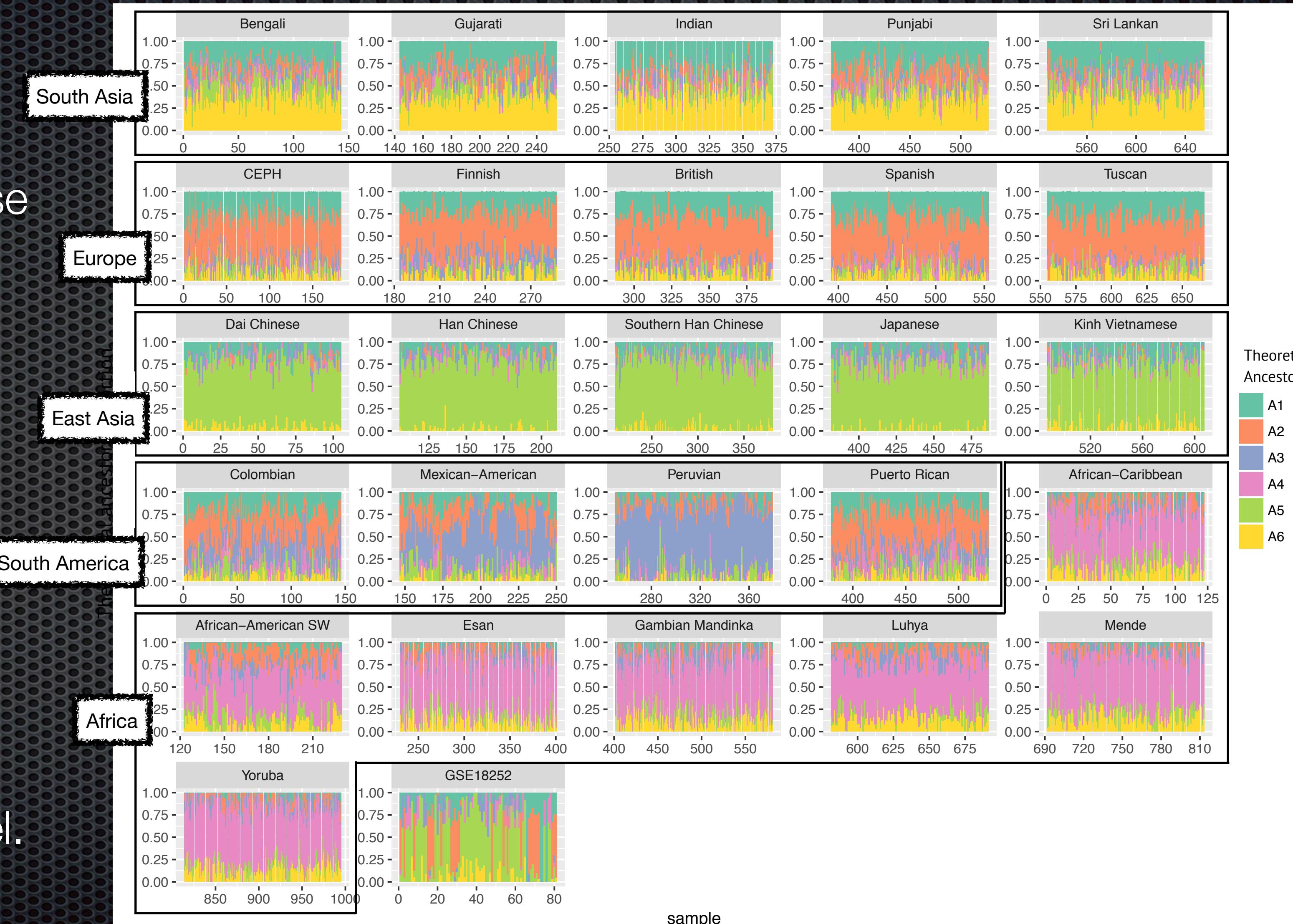
- the frequency of many genome variants depends on the genetic background
- cancer incidence & type can correlate to environmental factors
- geographic analysis can support interpretation and point to knowledge gaps



Geographic distribution of >140'000 cancer genome profiles reported in the literature. The numbers are derived from the 2947 publications registered in the Progenetix database.

Population stratification in cancer samples based on SNP array data

- 2504 genome profiles from 1000 Genome project phase 1 as reference
- 5 superpopulations: South Asia, Europe, South America, East Asia and Africa.
- SNP positions used in 9 Affymetrix SNP arrays are extracted to train a population admixture model.



GA4GH to solve accessibility...



Enabling genomic data sharing for the benefit of human health

The Global Alliance for Genomics and Health (GA4GH) is a policy-framing and technical standards-setting organization, seeking to enable responsible genomic data sharing within a **human rights framework**



**Genomic Data
Toolkit**



**Regulatory & Ethics
Toolkit**



**Data Security
Toolkit**



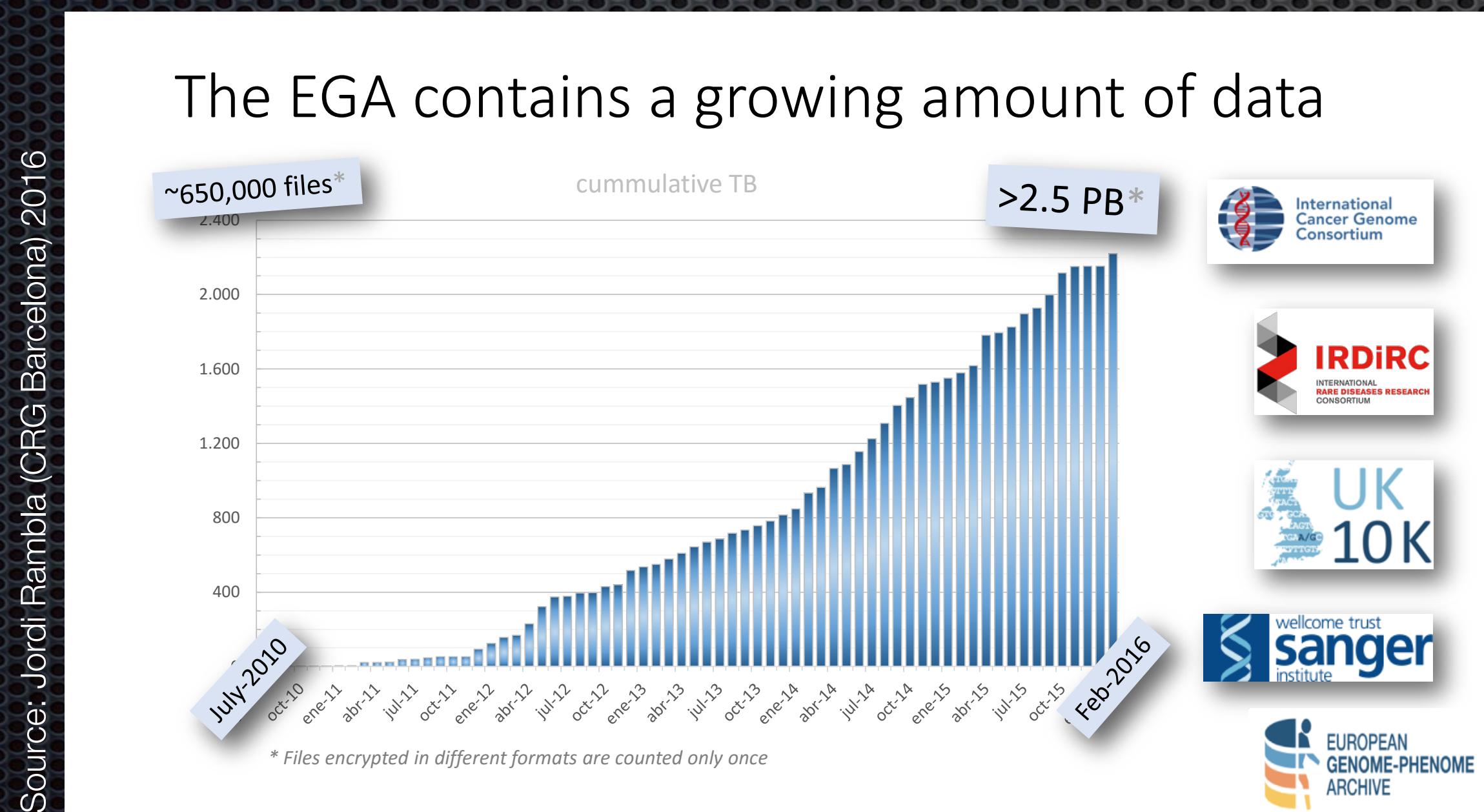
[VIEW OUR LEADERSHIP](#)

[MORE ABOUT US](#)

[BECOME A MEMBER](#)

Genome Datasets: Rapid Growth, Limited Access

population based and cancer research studies produce a rapidly increasing amount of genome sequence data



genome data is stored in an increasing number of institutional and core repositories, with **incompatible data** structures and **access** policies

GA4GH API promotes sharing

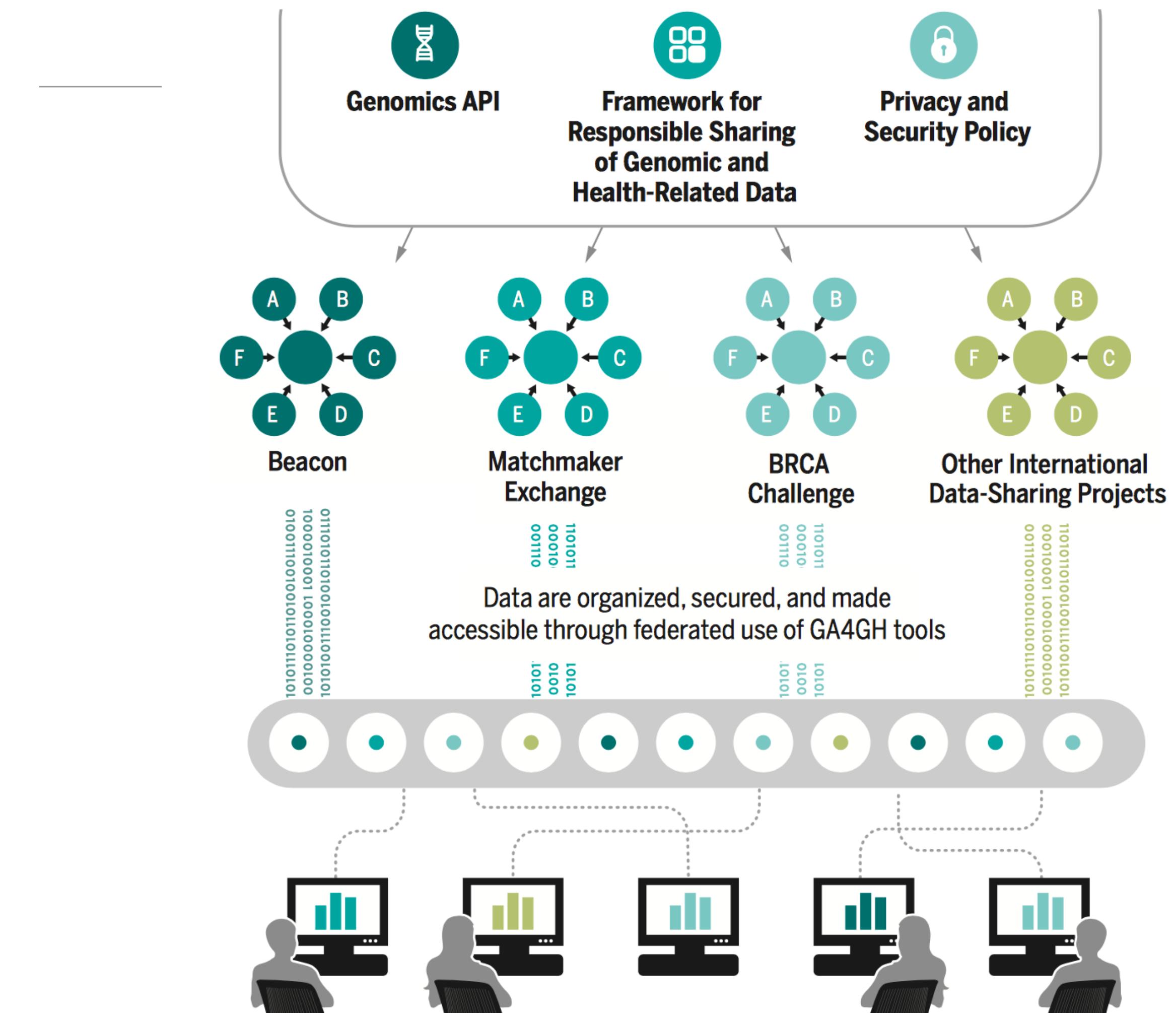
A federated data ecosystem. To share genomic data globally, this approach furthers medical research without requiring compatible data sets or compromising patient identity.



GENOMICS

A federated ecosystem for sharing genomic, clinical data

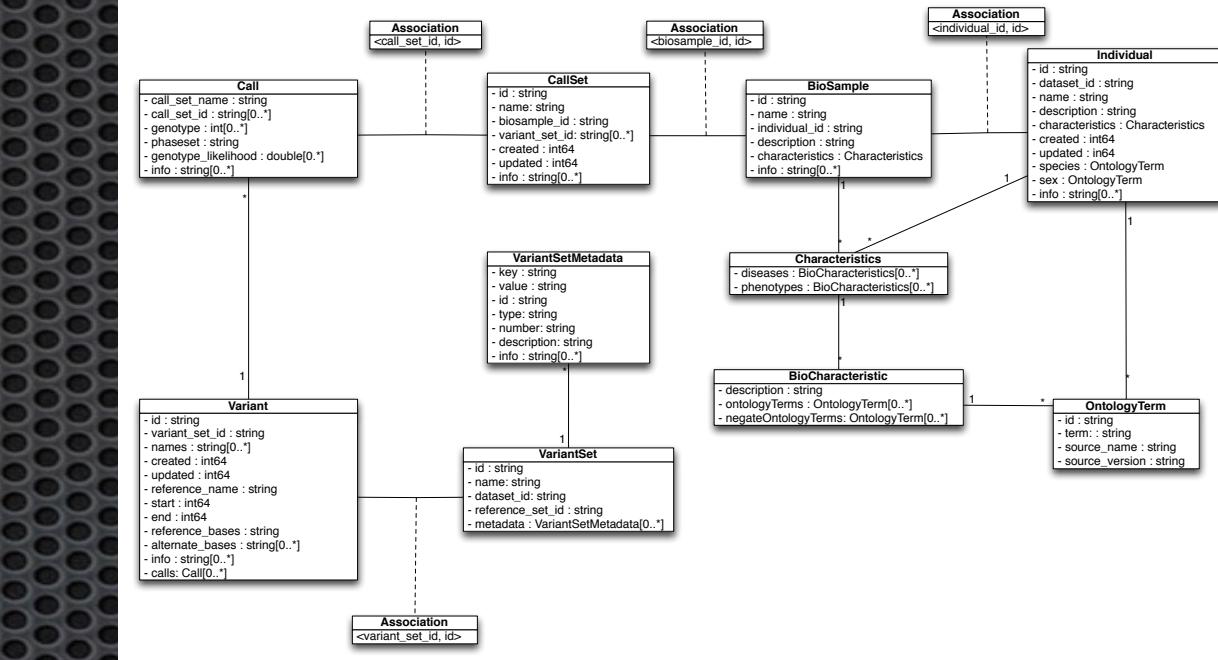
Silos of genome data collection are being transformed into seamlessly connected, independent systems



Developing the GA4GH Metadata Schema

▶ arrayMap for GA4GH

- metadata schema development through implementation of arrayMap resource data
- OntologyTerm objects for biodata
- implementation w/ ontology services



```

{
  "_id" : ObjectId("58297ca32ca4591e5a0df054"),
  "id" : "AM_V_1778741",
  "variant_set_id" : "AM_VS_HG18",
  "reference_name" : "10",
  "start" : 579049,
  "end" : 17236099,
  "alternate_bases" : "DUP",
  "reference_bases" : ".",
  "info" : {
    "svlen":16657050,
    "cipos": [
      -1000,
      1000
    ],
    "ciend": [
      -1000,
      1000
    ]
  },
  "calls" : [
    {
      "genotype" : [
        ".",
        "."
      ],
      "call_set_id" : "AM_CS_TCGA-61-1917-01A-01D-0648-01",
      "info" : {
        "segvalue" : 0.5491
      }
    }
  ],
  "created" : ISODate("2016-11-14T08:33:58.202Z"),
  "updated" : ISODate("2016-11-14T08:33:58.202Z"),
}
  
```

Driving Beacon Development

▶ Beacon*

- CNV/CNA as first type of structural variants
- disease specific queries
- quantitative reporting

Beacon Project

An open web service that tests the willingness of international sites to share genetic data.



Beacon Network

Search Beacons

A global search engine for genetic mutations.

GRCh37 ▾ e.g. 1: 100,000 A>C Search

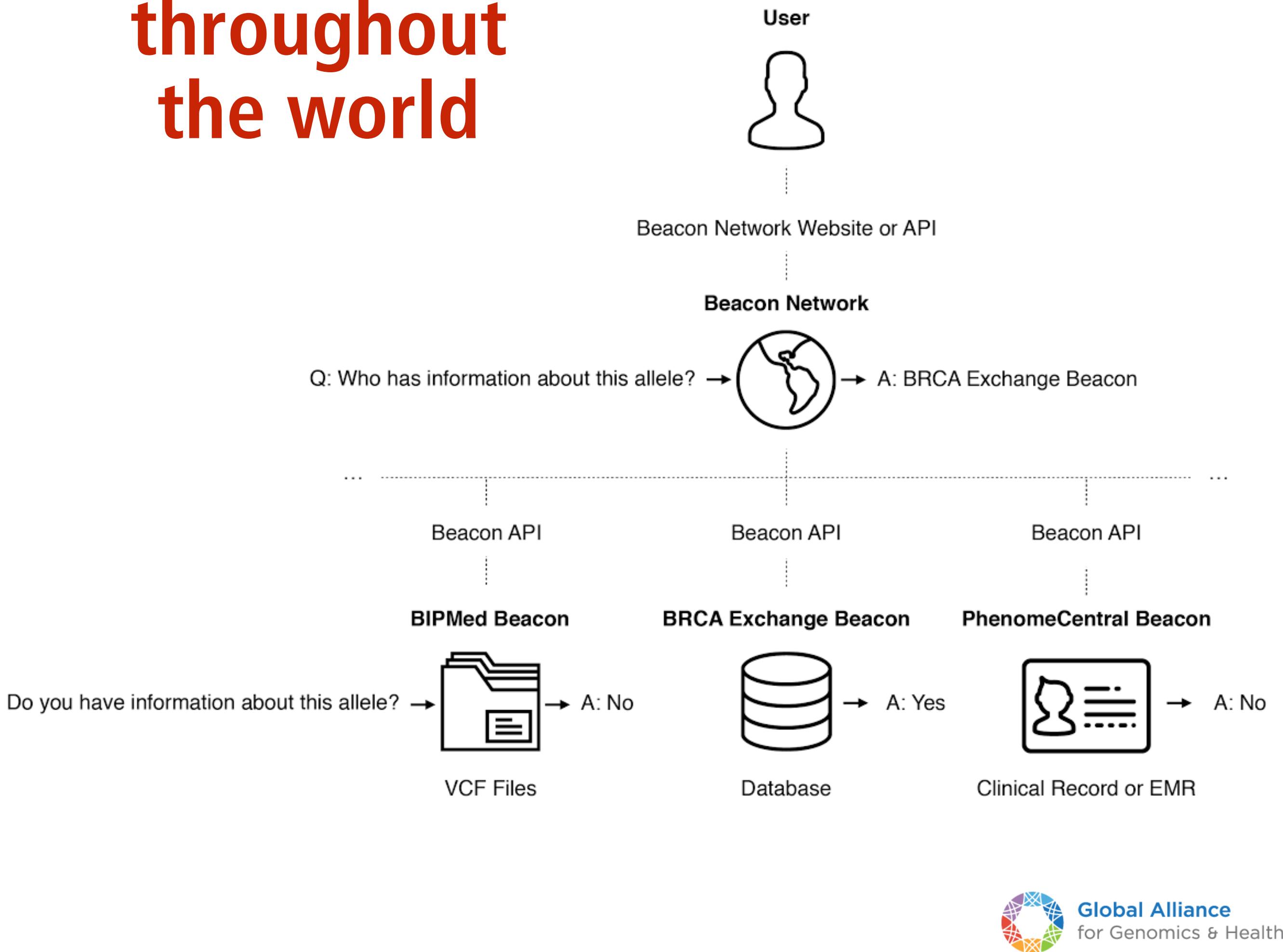
Quickstart: Search for a BRCA2 variant

Find genetic mutations shared by these organizations

- Global Gene Corp
- BRCA EXCHANGE
- Google
- BIPMed Beacon
- PC
- PhenomeCentral Beacon
- Clinical Record or EMR

Browse Beacons »

> 50 Beacons throughout the world



Beacon+ Concept

- Implementation of cancer beacon prototype, backed by arrayMap and DIPG data set
(MacKay *et al.*, Cancer Cell 2017, in print)
- structural variations (DUP, DEL) in addition to SNV
- diagnosis queries using ontology codes (NCIT, ICD-O)
- quantitative responses
- current version uses **GA4GH schema compatible** database

Beacon+

This forward looking Beacon interface implements additional, planned features beyond the current GA4GH specifications. [Info](#)

Query

Dataset: DIPG (CNV + selected SNV)

Reference name*: 17

Genome Assembly*: GRCh36 / hg18

Variant type*: SNV / indel

Position*: 7577121

Ref. Base(s)*: G

Alt. Base(s)*: A

Bio-ontology: pgx:icdom:9380_3

[Beacon Query](#)

Response

Dataset	Chro.	Assembly	Var. Type	Start Min	Start Max	End Min	End Max	Pos.	Ref.	Alt.	Bio Query	Call Count	Samples	f	Query
arraymap	9	GRCh36	DEL	19000000	21984490	21900000	25000000				pgx:icdom:8140_3	3781	403	0.0065	show JSON
dipg	17	GRCh36	SNV			7577121		G	A	pgx:icdom:9380_3	21	20	0.0187	show JSON	

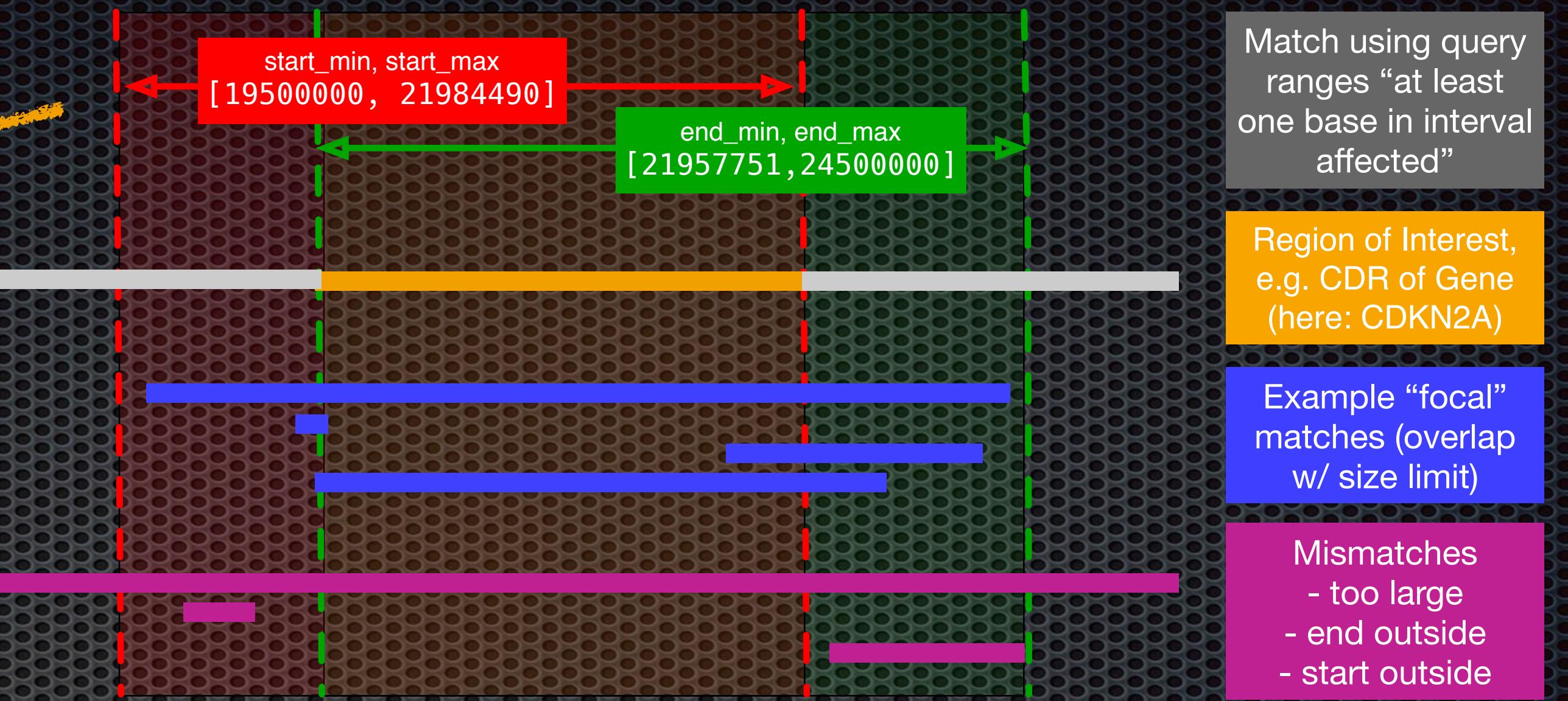
arrayMap  University of Zurich UZH  This Beacon implementation is developed by the Computational Oncogenomics Group at the University of Zurich, with support from the SIB Technology group and ELIXIR.   

```

{
  "allele_request" : {
    "$and": [
      { "reference_name" : "9" },
      { "variant_type" : "DEL" },
      { "start" : { "$gte" : 19500000 } },
      { "start" : { "$lte" : 21984490 } },
      { "end" : { "$gte" : 21957751 } },
      { "end" : { "$lte" : 24500000 } }
    ]
  },
  "api_version" : "0.4",
  "beacon_id" : "org.progenetix:progenetix-beacon",
  "exists" : true,
  "info" : {
    "url" : "http://progenetix.org/beacon/info/",
    "dataset_allele_responses" : [
      {
        "dataset_id" : "arraymap",
        "error" : null,
        "exists" : true,
        "external_url" : "http://arraymap.org",
        "sample_count" : 584,
        "call_count" : 3781,
        "variant_count" : 3244,
        "frequency" : 0.0094,
        "info" : {
          "description" : "The query was against database \\\"arraymap_ga4gh\\\", variant collection \\\"variants_cnv_grch36\\\". 3781 / 59428 matched callsets for 3602919 variants. Out of 62105 biosamples in the database, 2047 matched the biosample query; of those, 584 had the variant."
        },
        "ontology_ids" : [
          "ncit:C3058",
          "pgx:icdom:9440_3",
          "pgx:icdot:C71.9",
          "pgx:icdot:C71.0"
        ]
      }
    ],
  }
}

```

Metadata



- Beacon+**range queries** allow the definition of a genome region of interest, containing a specified variant (or other mappable feature)
- “fuzzy” matching of region ends is essential for features without base specific positions
- current Beacon implementation addresses CNV (<DUP>,), as are specified in VCF && GA4GH variant schema



Bioinformatics: **Ontologies**

- ontologies in information sciences describe concrete and abstract **objects**, there precisely defined **hierarchies** and **relationships**
- ontologies in bioinformatics support the move from a descriptive towards an **analytical science** in describing biological data and relations among it

"The widest use of ontologies within biology is for conceptual annotation – a representation of stored knowledge more computationally amenable than natural language."*

- Gene ontology (GO)
- NCI Neoplasm Core
- Uberon anatomical structures
- Experimental Factor Ontology (EFO)
- Disease Ontology (DO)



```
id: GO:0000118
name: histone deacetylase complex
namespace: cellular_component
def: "A protein complex that possesses histone deacetylase activity." [GOC:mah]
comment: Note that this term represents a location, not a function; the activity possessed by this complex is mentioned in the definition for the purpose of describing and distinguishing the complex. The function of this complex is represented by the molecular function term 'histone deacetylase activity'.
synonym: "HDAC complex" EXACT [
is_a: GO:0044451 ! nucleoplasm ]
is_a: GO:1902494 ! catalytic complex
```

- □ Neoplasm by Morphology
 - □ Epithelial Neoplasm [C3709](#)
 - □ Germ Cell Tumor [C3708](#)
 - □ Giant Cell Neoplasm [C7069](#)
 - □ Hematopoietic and Lymphoid Cell Neoplasm [C27134](#)
 - □ Melanocytic Neoplasm [C7058](#)
 - □ Benign Melanocytic Skin Nevus [C7571](#)
 - □ Dysplastic Nevus [C3694](#)
 - □ Melanoma [C3224](#)
 - □ Amelanotic Melanoma [C3802](#)
 - □ Cutaneous Melanoma [C3510](#)
 - □ Epithelioid Cell Melanoma [C4236](#)
 - □ Mixed Epithelioid and Spindle Cell Melanoma [C66756](#)
 - □ Non-Cutaneous Melanoma [C8711](#)
 - □ Spindle Cell Melanoma [C4237](#)
 - □ Meningothelial Cell Neoplasm [C6971](#)

Beacon+ Concept

- Implementation of cancer beacon prototype, backed by arrayMap and DIPG data set
(MacKay *et al.*, Cancer Cell 2017, in print)
- structural variations (DUP, DFI) in addition to SNV
- diagnosis queries using ontology codes (NCIT, ICD-O)
- quantitative responses
- current version uses **GA4GH schema compatible** database

Beacon+

This forward looking Beacon interface implements additional, planned features beyond the current GA4GH specifications. [Info](#)

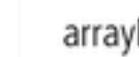
Query

Dataset: DIPG (CNV + selected SNV)
Reference name*: 17
Genome Assembly*: GRCh36 / hg18
Variant type*: SNV / indel
Position*: 7577121
Ref. Base(s)*: G
Alt. Base(s)*: A
Bio-ontology: pgx:icdom:9380_3

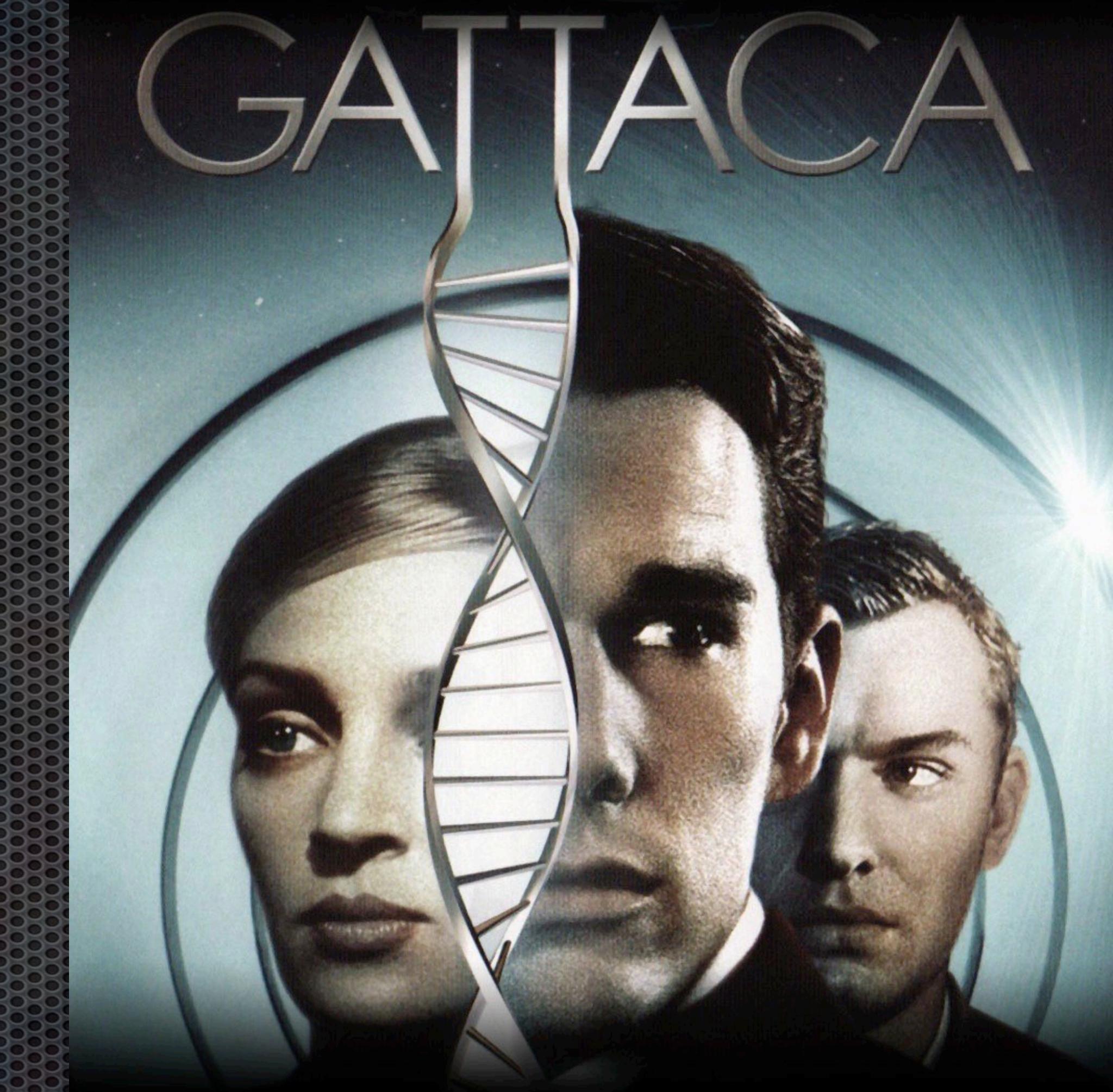
Beacon Query

Response

Dataset	Chro.	Assembly	Var. Type	Start Min	Start Max	End Min	End Max	Pos.	Ref.	Alt.	Bio Query	Call Count	Samples	f	Query
arraymap	9	GRCh36	DEL	19000000	21984490	21900000	25000000				pgx:icdom:8140_3	3781	403	0.0065	show JSON
dipg	17	GRCh36	SNV			7577121		G	A	pgx:icdom:9380_3	21	20	0.0187	show JSON	

arrayMap  University of Zurich UZH  This Beacon implementation is developed by the Computational Oncogenomics Group at the University of Zurich, with support from the SIB Technology group and ELIXIR.   

Genomes & Privacy



Generalkonsent

PRIVACY

HACKERS

Health
Insurance
Portability and
Accountability
Act

BENEFIT

CONSENT

LAWS

SAFETY

BLOCKCHAIN

ACCESS

SECURITY

Right to Research

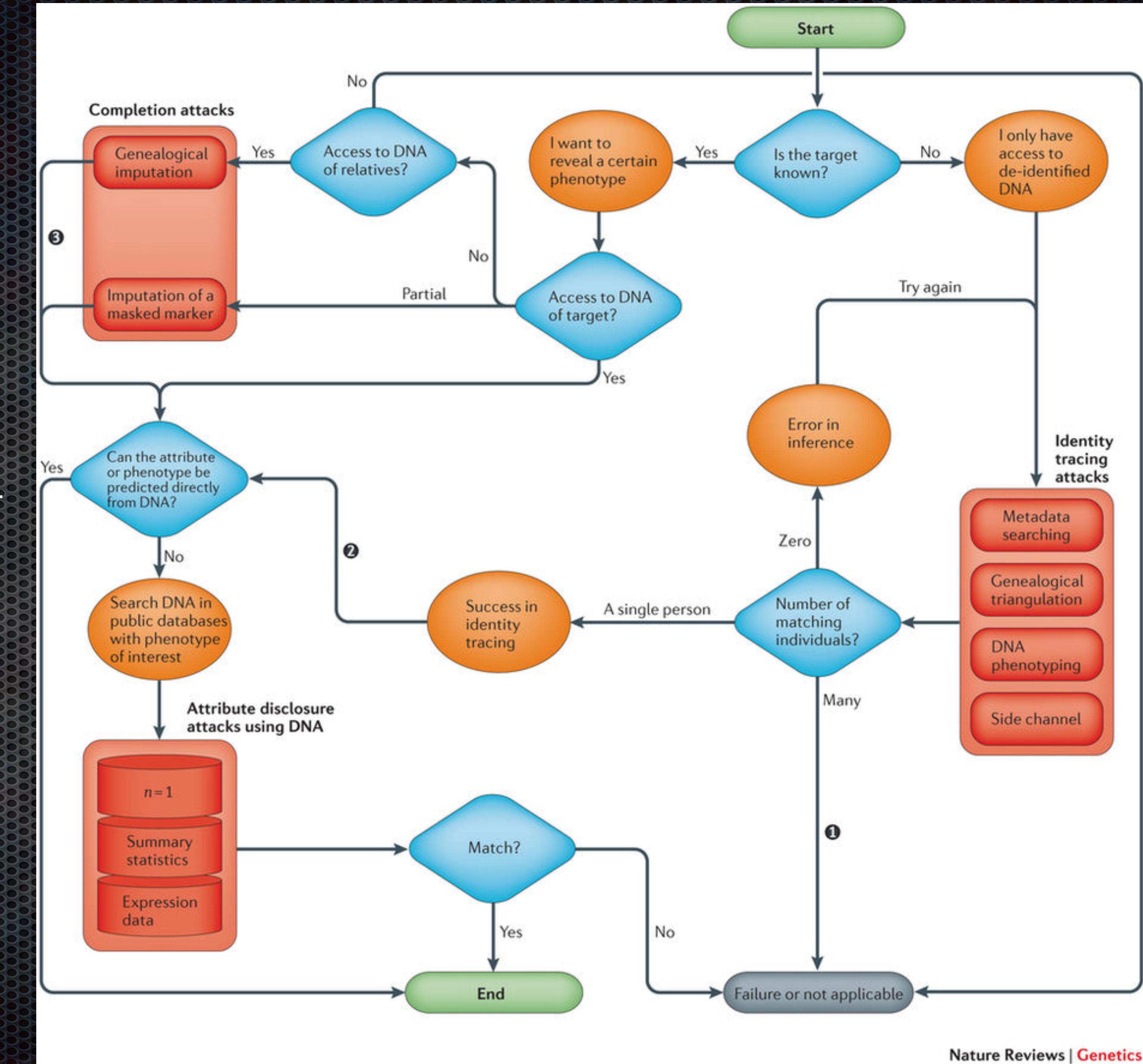
Genetic
Information
Nondiscrimination
Act

CRYPTOGRAPHY

Routes for breaching and protecting genetic privacy

The map contrasts different scenarios, such as identifying de-identified genetic data sets, revealing an attribute from genetic data and unmasking of data. It also shows the interdependencies between the techniques and suggests potential routes to exploit further information after the completion of one attack. There are several simplifying assumptions (black circles). In certain scenarios (such as insurance decisions), uncertainty about the target's identity within a small group of people could still be considered a success (assumption 1). For certain privacy harms (such as surveillance), identity tracing can be considered a success and the end point of the process (assumption 2). The complete DNA sequence is not always necessary (assumption 3).

Yaniv Erlich & Arvind Narayanan. *Nature Reviews Genetics* 15, 409–421 (2014)





Genome *Beacons* Compromise Security?

Querying for thousands of specific SNV occurrences in a genomic data pool can identify individuals

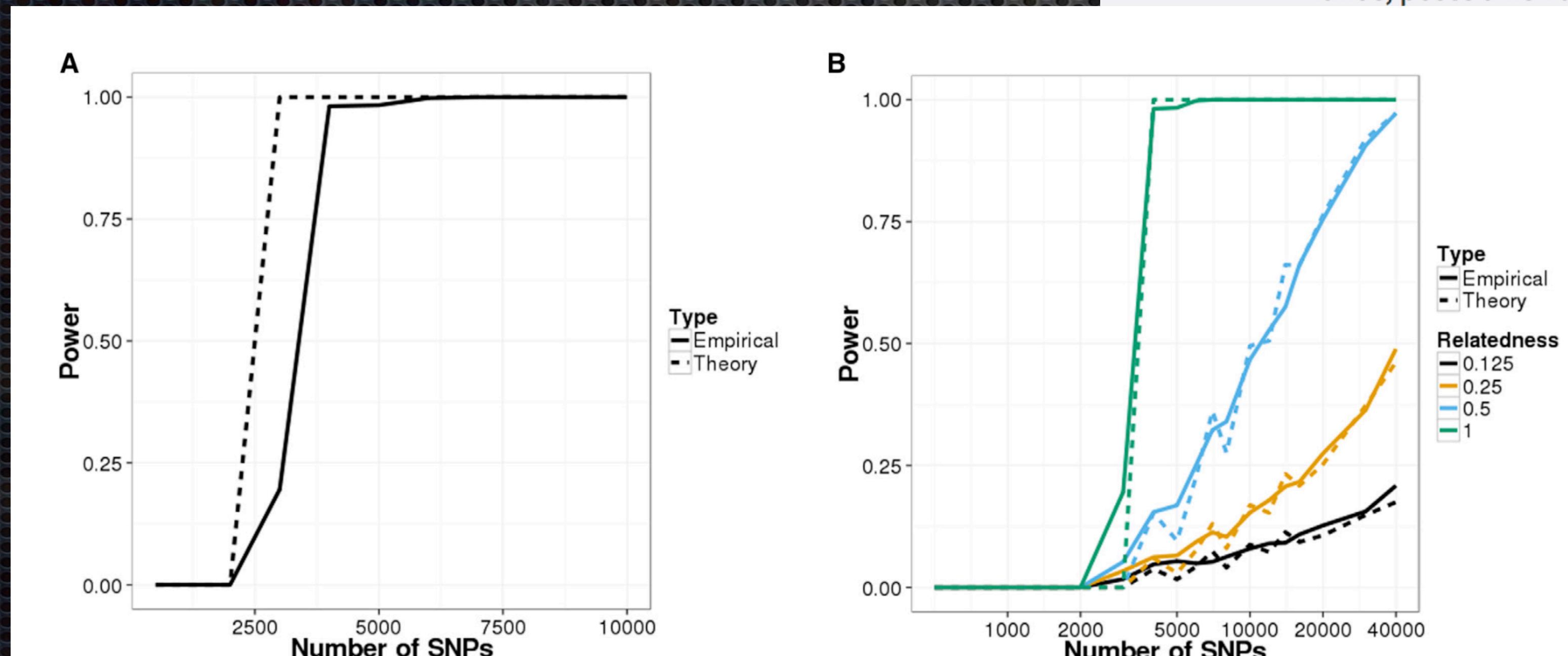


Figure 1. Power of Re-identification Attacks on Beacons Constructed with Simulated Data

Power curves for the likelihood-ratio test (LRT) on (A) a simulated beacon with 1,000 individuals and (B) detecting relatives in the simulated beacon. The false-positive rate was set to 0.05 for all scenarios.

Stanford researchers identify potential security hole in genomic data-sharing network

Hackers with access to a person's genome might find out if that genome is in an international network of disease databases.

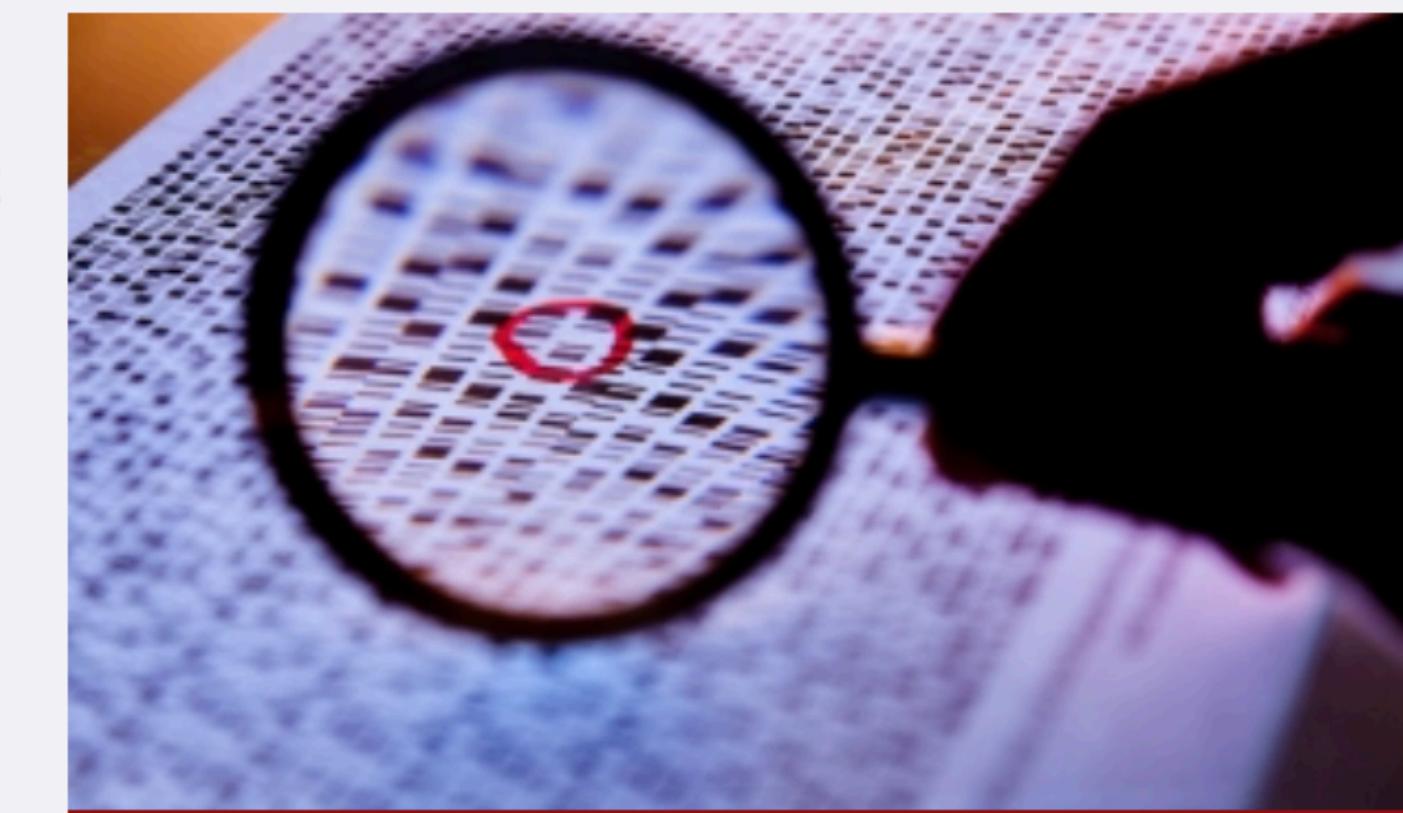
OCT 29
2015

Sharing genomic information among researchers is critical to the advance of biomedical research. Yet genomic data contains identifiable information and, in the wrong hands, poses a risk to individual

our genome
or saliva or
omic
eck to see if
with certain
sease, lung

Stanford
s that genomic
ture, PhD, a

and Carlos Bustamante, PhD, a professor of genetics, have
ng a network of global genomic databases and how to prevent it.
from the Global Alliance for Genomics and Health on implementing



Stanford researchers are working with the Global Alliance for Genomics and Health to make genomic information in the Beacon Project more secure.
Science photo/Shutterstock

American Journal of Human Genetics, also bears importantly on the mixtures of genomes, such as those from different people at a crime

The Right to Scientific Knowledge

In 1948, the General assembly of the United nations adopted the Universal Declaration of Human Rights (UDHR) to guarantee the rights of every individual in the world. Included were twin rights “to share in scientific advancement and its benefits” and “to the protection of the moral and material interests resulting from any scientific...production of which [a person] is the author” (art. 27, United nations 1948).

from Knoppers et al, 2014

A human rights approach to an international code of conduct for genomic and clinical data sharing

Bartha M. Knoppers · Jennifer R. Harris ·
Isabelle Budin-Ljøsne · Edward S. Dove

Received: 9 December 2013 / Accepted: 16 February 2014 / Published online: 27 February 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract Fostering data sharing is a scientific and ethical imperative. Health gains can be achieved more comprehensively and quickly by combining large, information-rich datasets from across conventionally siloed disciplines and geographic areas. While collaboration for data sharing is increasingly embraced by policymakers and the international biomedical community, we lack a common ethical and legal framework to connect regulators, funders, consortia, and research projects so as to facilitate genomic and clinical data linkage, global science collaboration, and responsible research conduct. Governance tools can be used to responsibly steer the sharing of data for proper stewardship of research discovery, genomics research resources, and their clinical applications. In this article, we propose that an international code of conduct be designed to enable global genomic and clinical data sharing for biomedical research. To give this proposed code universal application and accountability, however, we propose to position it within a human rights framework. This proposition is not without precedent: international treaties have long recognized that everyone has a right to the benefits of scientific

progress and its applications, and a right to the protection of the moral and material interests resulting from scientific productions. It is time to apply these twin rights to internationally collaborative genomic and clinical data sharing.

Introduction

In 1948, the General Assembly of the United Nations adopted the *Universal Declaration of Human Rights* (UDHR) to guarantee the rights of every individual in the world. Included were twin rights “to share in scientific advancement and its benefits” and “to the protection of the moral and material interests resulting from any scientific...production of which [a person] is the author” (Art. 27, United Nations 1948). In the 21st century, where are we in realizing the sharing of scientific advancement and its benefits, and the importance of protecting a scientific producer’s moral and material interests? In this article, we argue that these little-developed twin rights, what we call the right “to benefit from” and “to be recognized for”, have direct application to internationally collaborative genomic and clinical data sharing, and can be activated through an international code of conduct.

Sharing genomic and clinical data is critical to achieve precision medicine (National Research Council 2011), that is, more accurate disease classification based on molecular profiles to enable tailored effective treatments, interventions, and models for prevention. Better communication flow across borders and research teams, encompassing data from clinical and population research, enables researchers to connect the diverse types of datasets and expertise needed to elucidate the genomic basis and complexities of disease etiology. Such data integration can make it possible to reveal the genetic basis of cancer, inherited diseases,

B. M. Knoppers (✉) · E. S. Dove
Centre of Genomics and Policy, McGill University, 740 Dr.
Penfield Avenue, Suite 5200, Montreal H3A 0G1, Canada
e-mail: bartha.knoppers@mcgill.ca

E. S. Dove
e-mail: edward.dove@mcgill.ca

J. R. Harris · I. Budin-Ljøsne
Division of Epidemiology, Department of Genes
and Environment, Norwegian Institute of Public Health,
PO Box 4404, Nydalen 0403, Oslo, Norway
e-mail: Jennifer.Harris@fhi.no

I. Budin-Ljøsne
e-mail: Isabelle.Budin.Ljosne@fhi.no

Modernizing Patient Consent

- forward looking, transparent and technically feasible regulations for enabling access to research material and data while empowering patients

Generalkonsent: Eine einheitliche Vorlage soll schweizweite Forschung erleichtern

Art des Forschungs-materials	Biologisches Material und genetische Daten	Nicht-genetische Daten
Personenbezug		
Unverschlüsselt (identifizierend)	Information + Einwilligung in jedes einzelne Forschungsprojekt	Information über Weiterverwendung für zukünftige noch unbestimmte Forschungsprojekte + Generalkonsent für Forschungszwecke
Verschlüsselt	Information über Weiterverwendung für zukünftige noch unbestimmte Forschungsprojekte + Generalkonsent für Forschungszwecke	Information über Weiterverwendung für zukünftige noch unbestimmte Forschungsprojekte + Generalkonsent für Forschungszwecke + über Möglichkeit Weiterverwendung abzulehnen > Widerspruchsrecht
Anonymisiert	Genetische Daten: Information über Weiterverwendung für zukünftige noch unbestimmte Forschungszwecke + über Möglichkeit Weiterverwendung abzulehnen > Widerspruchsrecht Proben: Information zur Anonymisierung > Widerspruchsrecht	Ausserhalb des Geltungsbereichs des HFG

Switzerland: Definition of a unified "Generalkonsent", to provide a single framework to manage permissions for access to patient derived material and related data

Consent Codes: Upholding Standard Data Use Conditions

Stephanie O. M. Dyke^{1*}, Anthony A. Philippakis², Jordi Rambla De Argila^{3,4}, Dina N. Paltoo⁵, Erin S. Luetkemeier⁵, Bartha M. Knoppers¹, Anthony J. Brookes⁶, J. Dylan Spalding⁷, Mark Thompson⁸, Marco Roos⁸, Kym M. Boycott⁹, Michael Brudno^{10,11}, Matthew Hurles¹², Heidi L. Rehm^{2,13}, Andreas Matern¹⁴, Marc Fiume¹⁵, Stephen T. Sherry¹⁶



Consent Codes		
Name	Abbreviation	Description
Primary Categories (I^{IV})		
no restrictions	NRES	No restrictions on data use.
general research use and clinical care	GRU(CC)	For health/medical/biomedical purposes and other biological research, including the study of population origins or ancestry.
health/medical/biomedical research and clinical care	HMB(CC)	Use of the data is limited to health/medical/biomedical purposes, does not include the study of population origins or ancestry.
disease-specific research and clinical care	DS-[XX](CC)	Use of the data must be related to [disease].
population origins/ancestry research	POA	Use of the data is limited to the study of population origins or ancestry.
Secondary Categories (II^{IV}) (can be one or more extra conditions, in addition to I ^{IV} category)		
other research-specific restrictions	RS-[XX]	Use of the data is limited to studies of [research type] (e.g., pediatric research).
research use only	RUO	Use of data is limited to research purposes (e.g., does not include its use in clinical care).
no “general methods” research	NMDS	Use of the data includes methods development research (e.g., development of software or algorithms) ONLY within the bounds of other data use limitations.
genetic studies only	GSO	Use of the data is limited to genetic studies only (i.e., no research using only the phenotype data).
Requirements		
not-for-profit use only	NPU	Use of the data is limited to not-for-profit organizations.
publication required	PUB	Requestor agrees to make results of studies using the data available to the larger scientific community.
collaboration required	COL-[XX]	Requestor must agree to collaboration with the primary study investigator(s).
return data to database/resource	RTN	Requestor must return derived/enriched data to the database/resource.
ethics approval required	IRB	Requestor must provide documentation of local IRB/REC approval.
geographical restrictions	GS-[XX]	Use of the data is limited to within [geographic region].
publication moratorium/embargo	MOR-[XX]	Requestor agrees not to publish results of studies until [date].
time limits on use	TS-[XX]	Use of data is approved for [x months].
user-specific restrictions	US	Use of data is limited to use by approved users.
project-specific restrictions	PS	Use of data is limited to use within an approved project.
institution-specific restrictions	IS	Use of data is limited to use within an approved institution.

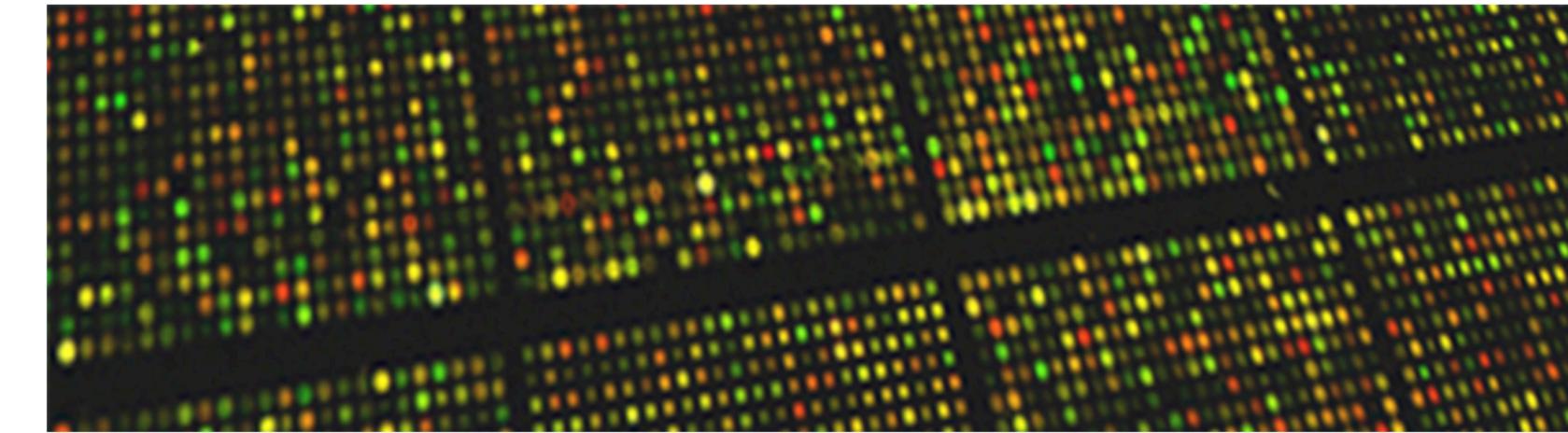
SOM Dyke, et al. Consent Codes: Upholding Standard Data Use Conditions. *PLoS Genetics* 12(1): e1005772.
<http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005772>

Contact: Dr. Stephanie Dyke (stephanie.dyke@mcgill.ca)

SHARE YOUR GENOME DATA?

- ▶ depositing genome data has the inherent risk of being identified and linked to your person
- ▶ What are the Risks?
- ▶ Would you contribute e.g. to OpenSNP?
- ▶ Discuss!

Welcome to *openSNP*



openSNP lets customers of direct-to-customer genetic tests publish their test results, find others with similar genetic variations, learn more about their results by getting the latest primary literature on their variations, and helps scientists find new associations.

[Sign Up!](#)[Download the data!](#)

For Genotyping Users For Scientists

Upload Your Genotyping File



Upload your raw genotyping or exome data from [23andMe](#), [deCODEme](#) or [FamilyTreeDNA](#) to the *openSNP* database to make it available for everybody.

Share Your Phenotypes & Traits



Phenotypes are the observable characteristics of your body, such as height, eye color or preference for coffee. Share your phenotype with other *openSNP* users, and find others with similar characteristics and traits. Your data may help scientists discover new genetic associations!

Share your stories on variations & phenotypes



With *openSNP* you can share stories about your genetic variations and phenotypes, and discover the stories of other users.

Find literature on genetic variation

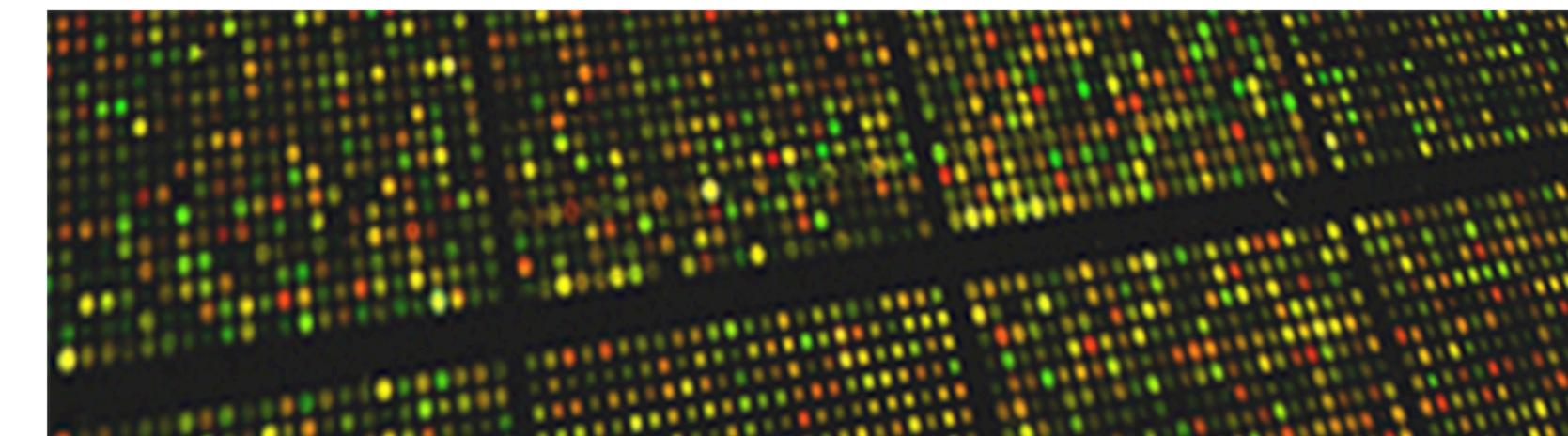


openSNP gets the latest open access journal articles on genetic variations from the [Public Library of Science](#). Popular articles are indexed via the social reference manager [Mendeley](#), and summaries are provided by [SNPedia](#).

SHARE YOUR GENOME DATA?



Welcome to *openSNP*



openSNP lets customers of direct-to-customer genetic tests publish their test results, find others with similar genetic variations, learn more about their results by getting the latest primary literature on their variations, and helps scientists find new associations.

[Sign Up!](#)[Download the data!](#)

For Genotyping Users For Scientists

Upload Your Genotyping File



Upload your raw genotyping or exome data from *23andMe*, *deCODEme* or *FamilyTreeDNA* to the *openSNP* database to make it available for everybody.

Share Your Phenotypes & Traits



Phenotypes are the observable characteristics of your body, such as height, eye color or preference for coffee. Share your phenotype with other *openSNP* users, and find others with similar characteristics and traits. Your data may help scientists discover new genetic associations!

Share your stories on variations & phenotypes

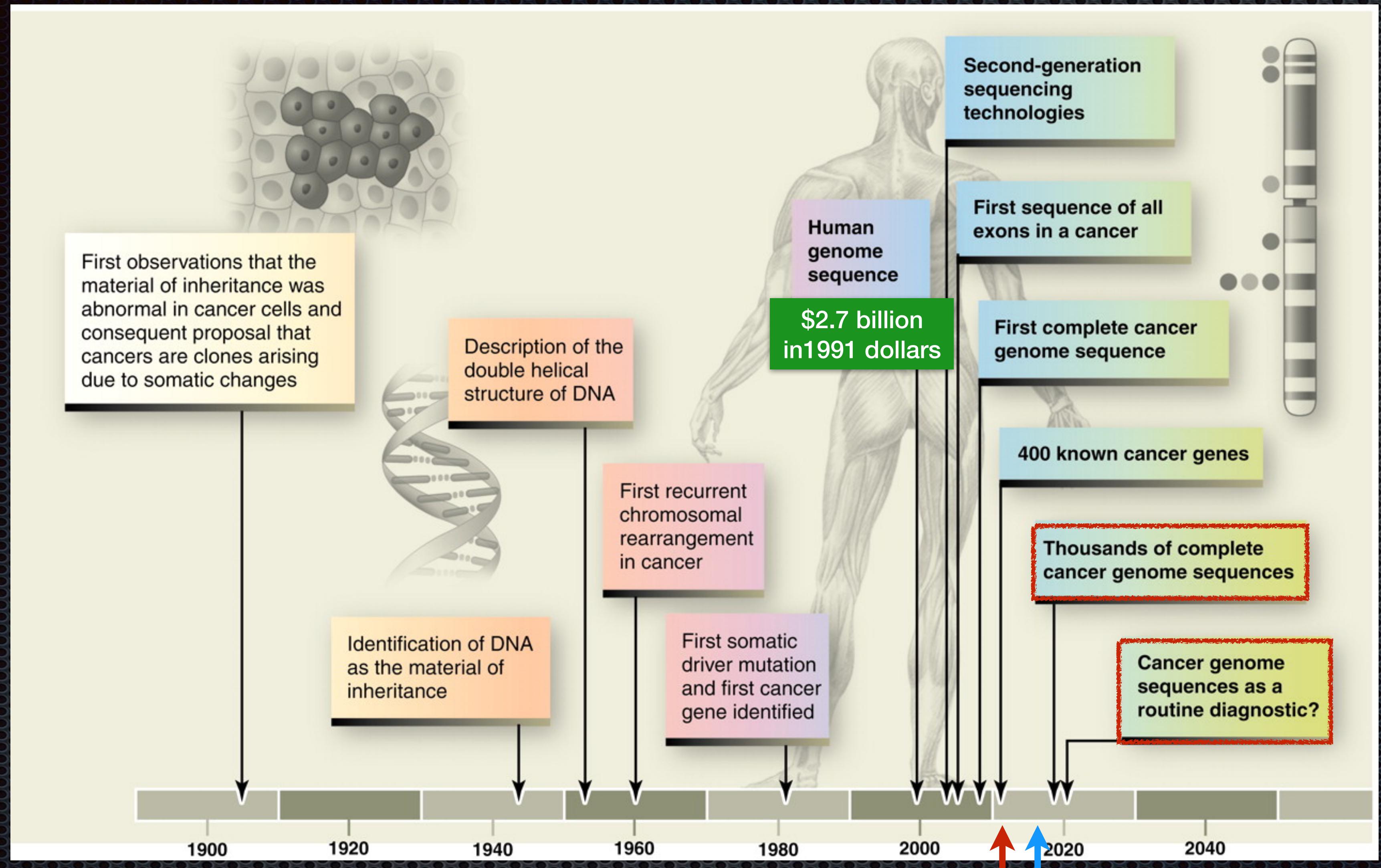


With *openSNP* you can share stories about your genetic variations and phenotypes, and discover the stories of other users.

Find literature on genetic variation



openSNP gets the latest open access journal articles on genetic variations from the *Public Library of Science*. Popular articles are indexed via the social reference manager *Mendeley*, and summaries are provided by *SNPedia*.



Th. Boveri

Nowell/Hungerford
J. Rowley

2016



Not yet practical
for full human
genomes, but real
product!
(OxfordNanopore)

Michael R. Stratton.
Exploring the Genomes
of Cancer Cells:
Progress and Promise.
Science (2011)

BAUDISGROUP @ UZH

NI AI
MICHAEL BAUDIS
(HAOYANG CAI)
PAULA CARRIO CORDO
QINGYAO HUANG
BO GAO
(LINDA GROB)
SAUMYA GUPTA
(ROMAN HILLJE)
(NITIN KUMAR)
(ALESSIO MILANESE)

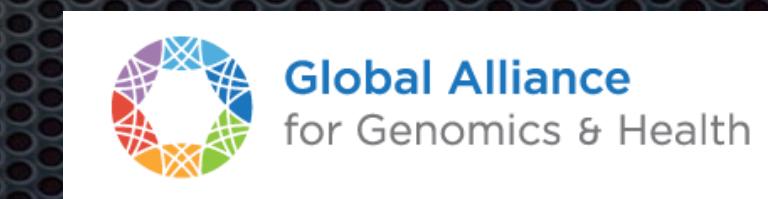
SIB

HEINZ STOCKINGER
SÉVERINE DUVAUD
DANIEL TEIXEIRA

ROSA NOGUERA
CAIUS SOLOVAN
ELENA CHITICARIU
GEORGIANA GUG



University of
Zurich UZH



GA4GH DWG + CWG

JACQUI BECKMANN
ANTHONY BROOKES
MELANIE COURTOT
MARK DIEKHANS
MELISSA HAENDEL
DAVID HAUSSLER
SARAH HUNT
STEPHEN KEENAN
SUZY LEWIS
DAVID LLOYD
MICHAEL MILLER
HELEN PARKINSON
GUNNAR RÄTSCH
ELEANOR STANLEY
DAVID STEINBERG
JULIA WILSON

ELIXIR & CRG

JORDI RAMBLA DE ARGILA
S. DE LA TORRE PERNAS
SUSANNA REPO
SERENA SCOLLEN

Prof. Dr. Michael Baudis
Institute of Molecular Life Sciences
University of Zurich
SIB | Swiss Institute of Bioinformatics
Winterthurerstrasse 190
CH-8057 Zurich
Switzerland

arraymap.org
progenetix.org
sib.swiss/baudis-michael
imls.uzh.ch/en/research/baudis