



NGS applications, data challenges and solutions

Guillaume Bourque

Dept. Human Genetics, McGill University
McGill University and Genome Quebec Innovation Center
Canadian Center for Computational Genomics (C3G)



@guilbourque

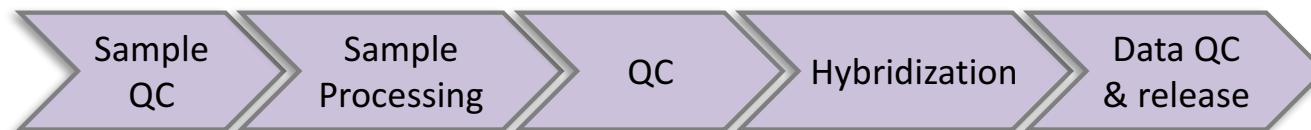
Dec 5th 2017

McGill Genome Center



- Genomics research centre created in 2002
- Partnership involving McGill University & Genome Quebec
- Uninterrupted operational funding support from Genome Canada
- Under scientific leadership of Dr. Mark Lathrop

Array technologies



DNA

- Genotyping
- CNV
- Methylation

Up to 75,000
samples per
year

RNA

- Gene expression
- Transcriptomics

Up to 2,000
samples per
year

Array technologies



GeneChip



GeneTitan

Affymetrix



HiScan



iScan

Illumina



Agena
MassARRAY

Illumina NGS applications



Genomic DNA

- “Shotgun” with PCR
- PCR-free
- WGBS
- 10X Genomics linked reads - **NEW**

Targeted DNA

- Whole exome
- Custom capture
- Methyl capture
- ChIP-seq
- Metagenomics
- Amplicons

RNA

- polyA+ RNA
- rRNA-depleted RNA
- Small RNAs
- RNA capture
- 10X Genomics single cell - **NEW**

Illumina sequencer fleet



3 MiSeq



1 HiSeq 4000



1 HiSeq X5



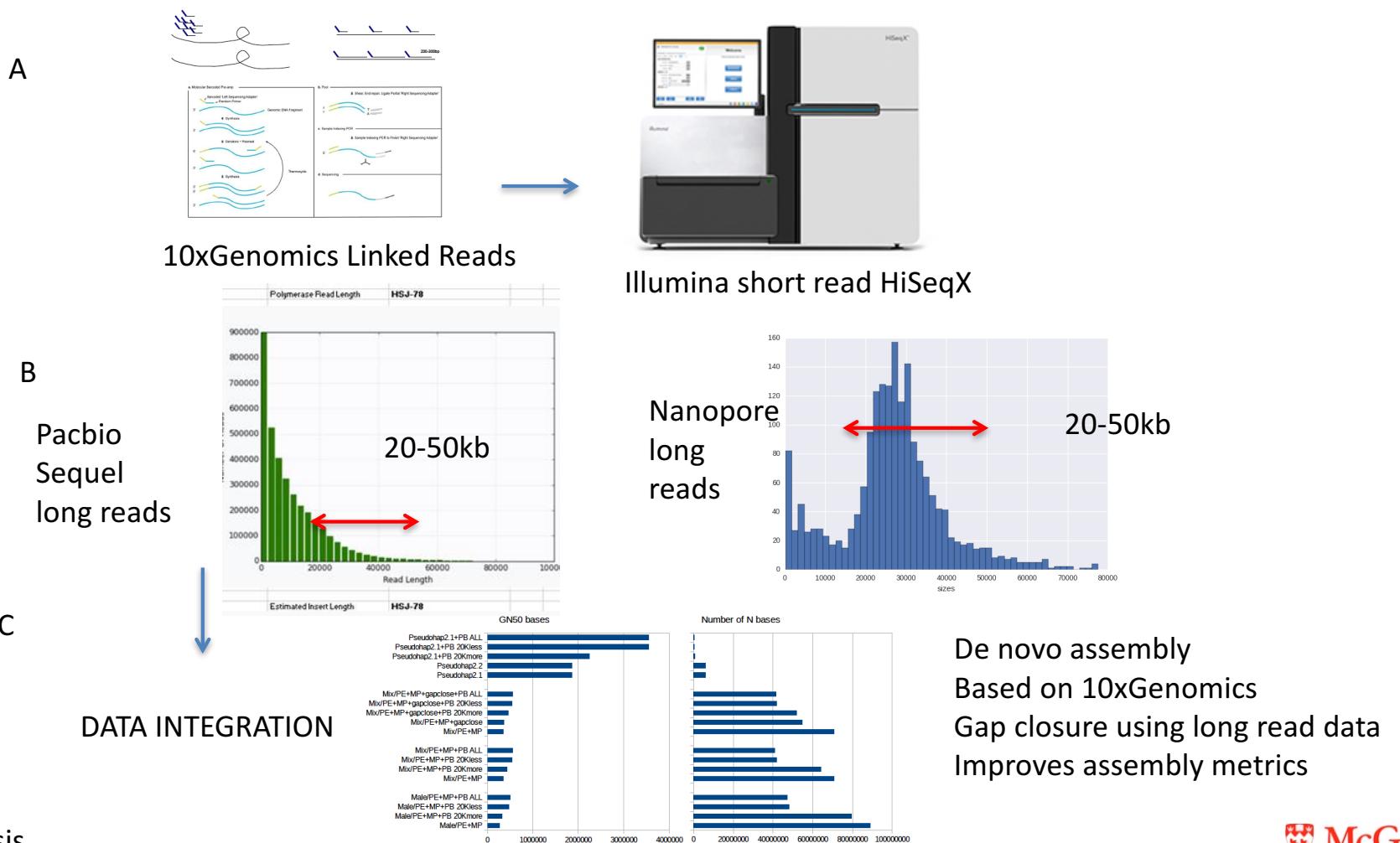
1 NovaSeq

3rd generation sequencing

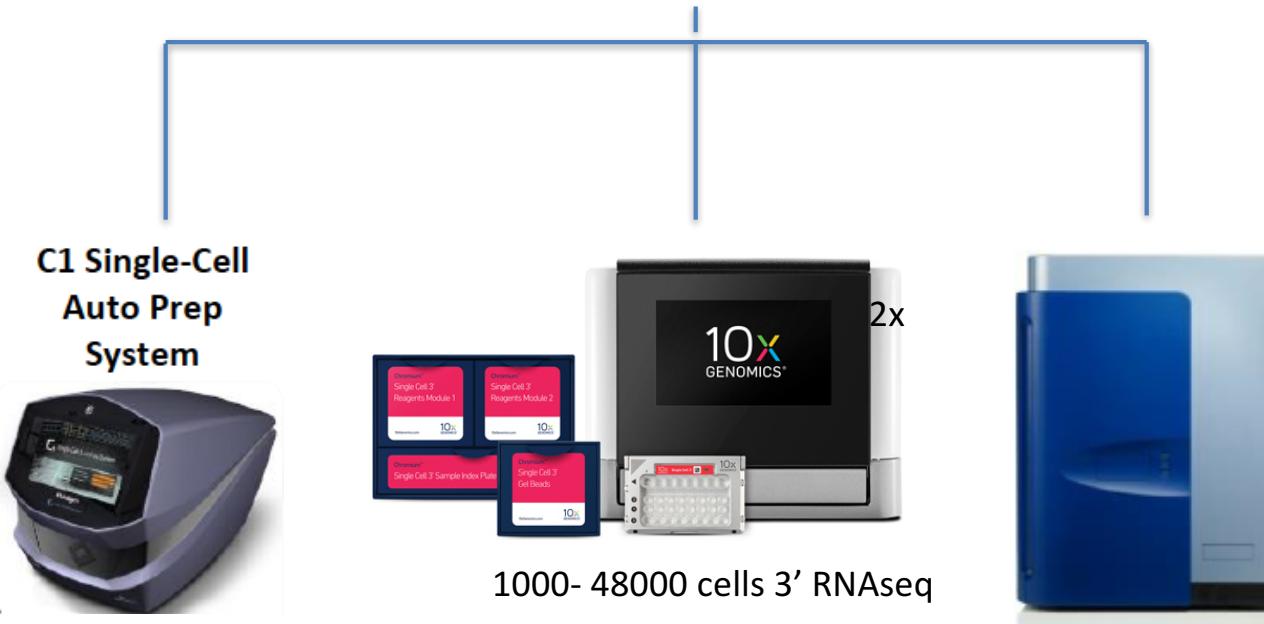
- Pacific Biosciences RSII & Sequel
- >10kb average read length
- Eliminates PCR bias/artefacts
- Applications:
 - Small- to mid-size genome assembly
 - Larger genomes in combination with short reads
 - Full-length transcripts/isoform
 - Haplotype and complex region resolution
 - Base modification
- Oxford Nanopore



CFI SEQUENCING TECHNOLOGY INTEGRATION at MUGQIC



McGill Single Cell Genomics



10s-800 cells, Cell visualization 70% of input cells captured

Single Cell mRNASeq:

Full transcript and 3' end counting

Single Cell DNASeq:

WGA, WES, Targeted DNASeq

Biomark HD
(for Bulk and Single-Cell qPCR)

Single Cell qPCR / STA:
up to 96 assays per cell

Publications:

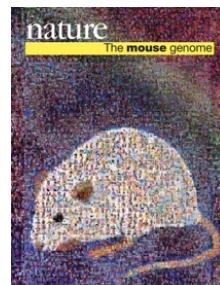
Binan et al Nature Communications 2016

Savage et al Cell Reports 2017

Boufaied et al Scientific Reports 2017

Applications (I)

- *De novo* sequencing
 - From the human genome... To all model organisms... To all relevant organisms (e.g. extreme genomes)... To “all” organisms?



Applications (II)

- Genome re-sequencing
 - Map genomic structural variations across individuals (to understand genetic disorders and also susceptibility factors)
 - Cancer genome sequencing
 - Agricultural crops



1000 Genomes Project



The Cancer Genome Atlas



Complete Resequencing of 40 Genomes Reveals Domestication Events and Genes in Silkworm (*Bombyx*)

Qingyou Xia,^{1,2*} Yiran Guo,^{3*} Ze Zhang,^{1,2*} Dong Li,^{1,2,4} Zhaoling Xuan,^{3*} Zhu Li,^{3*} Fangyan Dai,⁴ Yingran Li,³ Daojun Cheng,³ Ruiqiang Li,³ Tingcai Cheng,^{3,5} Tan Jiang,³ Celine Beiques,^{3,7} Xun Xu,² Chun Liu,² Xinghu Zha,³ Wei Fan,² Ying Lin,³ Yihong Shen,¹ Lan Jiang,³ Jef Huygen,⁶ Ines Heidmann,⁶ Mingming Li,³ Ping Wu,³ Xu Chen,³ Cheng Yu,³ Guofang Wang,³ Jun Li,³ Shiqing Liu,³ Ningning Li,³ Yan Zhou,³ Haizhi Liu,³ Jing Zhao,³ Chen Ye,² Zhouhe Du,³ Guojing Pan,¹ Aichun Zhao,³ Huijing Shi,^{3,7} Wei Zeng,³ Ping Wu,³ Chunfeng Li,³ Minhuai Pan,³ Jingling Lu,³ Xuyang Yin,³ Dawei Li,³ Juan Wang,³ Huihong Zheng,³ Wen Wang,³ Xueying Zhang,³ Songgang Li,³ Huanning Yang,³ Cheng Lu,³ Rasmus Nielsen,^{4,5} Zeyang Zhou,^{3,6} Jian Wang,³ Zhonghua Xiang,^{1,3†} Jun Wang^{3,4,‡}

Exome sequencing for Mendelian disease

REVIEWS

TRANSLATIONAL GENETICS

Exome sequencing as a tool for Mendelian disease gene discovery

Michael J. Bamshad^{*†}, Sarah B. Ng[‡], Abigail W. Bigham^{*§}, Holly K. Tabor^{*||}, Mary J. Emond[¶], Deborah A. Nickerson[†] and Jay Shendure[†]

Abstract | Exome sequencing — the targeted sequencing of the subset of the human genome that is protein coding — is a powerful and cost-effective new tool for dissecting the genetic basis of diseases and traits that have proved to be intractable to conventional gene-discovery strategies. Over the past 2 years, experimental and analytical approaches relating to exome sequencing have established a rich framework for discovering the genes underlying unsolved Mendelian disorders. Additionally, exome sequencing is being adapted to explore the extent to which rare alleles explain the heritability of complex diseases and health-related traits. These advances also set the stage for applying exome and whole-genome sequencing to facilitate clinical diagnosis and personalized disease-risk profiling.

“... about one-half to one-third (~3,000) of all known or suspected Mendelian disorders (for example, cystic fibrosis and sickle cell anaemia) have been discovered. However, there is a substantial gap in our knowledge about the genes that cause many rare Mendelian phenotypes.”

“Accordingly, we can realistically look towards a future in which the genetic basis of all Mendelian traits is known, ...”

Exome sequencing for Mendelian disease

BRIEF REPORT

Human Mutation



Mutations in *NOTCH2* in Families with Hajdu-Cheney Syndrome

Jacek Majewski,¹ Jeremy A. Schwartzentruber,¹ Aurélie Caqueret,² Lysanne Patry,² Janet M Kym M. Boycott,³ Louis-Georges Ste-Marie,⁴ Fergus E. McKenna,⁵ Ivo Marić,⁶ Hilde Van Es,⁷ FORGE Canada Consortium,¹ Jacques L. Michaud,² and Mark E. Samuels^{2,*}

¹Department of Human Genetics, McGill University and Genome Quebec Innovation Centre, Canada; ²Centre Ste-Justine, 3175, Côte Ste-Catherine, Montréal, Canada; ³Department of Genetics, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada; ⁴Université de Montréal, Centre de Recherche du CHUM, Hôpital Saint-Luc, 264 René Lévesque boulevard de la Côte-Saint-Luc, Montréal, Québec, Canada; ⁵Marshfield Clinic, Marshfield, Wisconsin; ⁶Ambulatory Centre for Defects of Locomotor Apparatus, University Hospitals Leuven, Leuven, Belgium; ⁷Department of Medicine, University of Montreal, Montréal, Québec, Canada

Exomes



SHORT REPORT

Combined malonic and methylmalonic aciduria: exome sequencing reveals mutations in the *ACSF3* gene in patients with a non-classic phenotype

Exomes

ORIGINAL ARTICLE

Novel inborn error of folate metabolism: identification by exome capture and sequencing of mutations in the *MTHFD1* gene in a single proband

David Watkins,¹ Jeremy A Schwartzentruber,² Jaya Ganesh,³ Bernard S Kaplan,⁵ Laura Dempsey Nunez,¹ Jacek Majewski,¹ David S Rosenblatt^{1,6,7}

Exomes

COMMUNICATIONS

A new ocular phenotype associated with an unexpected but known systemic disorder and mutation: novel use of genomic diagnostics and exome sequencing

Jacek Majewski,¹ Zibo Wang,¹ Irma Lopez,² Sulaiman Al Humaid,² Huanan Ren,¹ Julie Racine,² Alex Bazinet,² Grant Mitchel,¹ Nancy Braverman,¹ Robert K Koenekoop²



Cancer genome sequencing

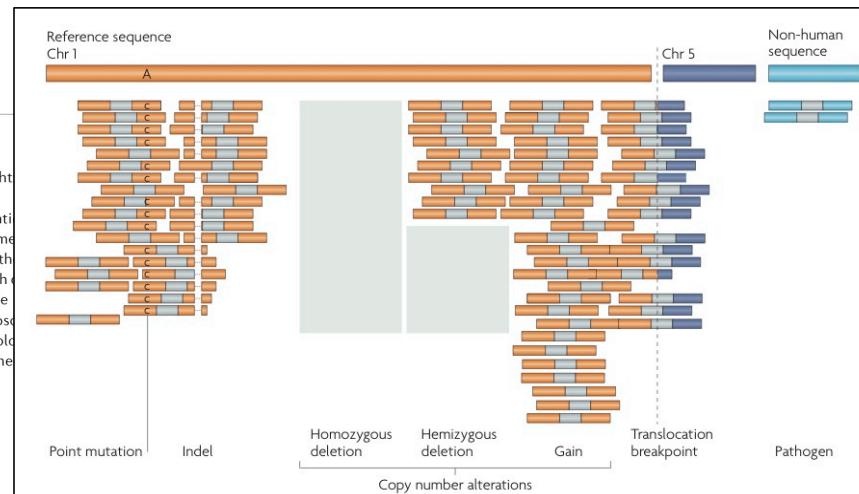
REVIEWS

APPLICATIONS OF NEXT-GENERATION SEQUENCING

Advances in understanding cancer genomes through second-generation sequencing

Matthew Meyerson, Stacey Gabriel and Gad Getz

Abstract | Cancers are caused by the accumulation of genomic alterations. Therefore, analyses of cancer genome sequences and structures provide insight for understanding cancer biology, diagnosis and therapy. The application of second-generation DNA sequencing technologies (also known as next-generation sequencing) — through whole-genome, whole-exome and whole-transcriptome approaches — is allowing substantial advances in cancer genomics. These methods are facilitating an increase in the efficiency and resolution of detection of each of the principal types of somatic cancer genome alterations, including nucleotide substitutions, small insertions and deletions, copy number alterations, chromosomal rearrangements and microbial infections. This Review focuses on the methodology, considerations for characterizing somatic genome alterations in cancer and the prospects for these approaches.

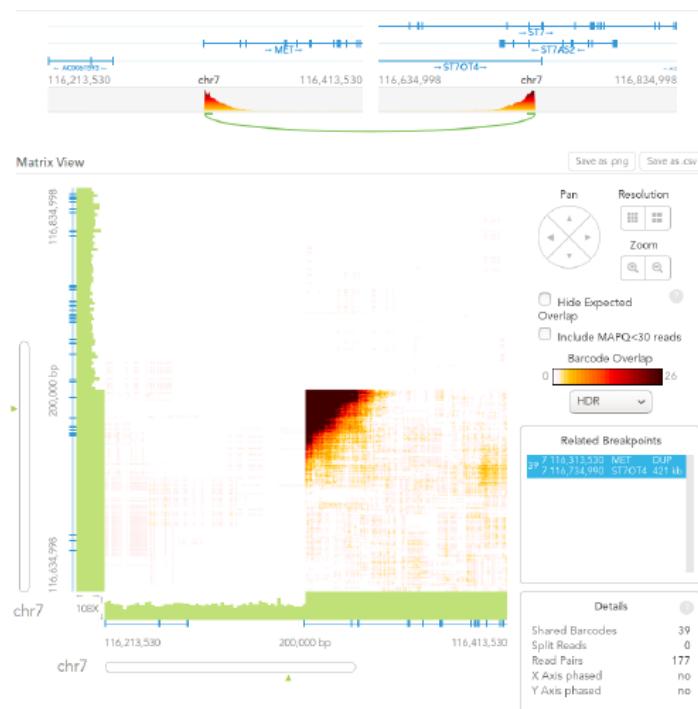


Can obtain a full catalogue of mutations

Analysis of tumor DNA using 10x Genomics technology

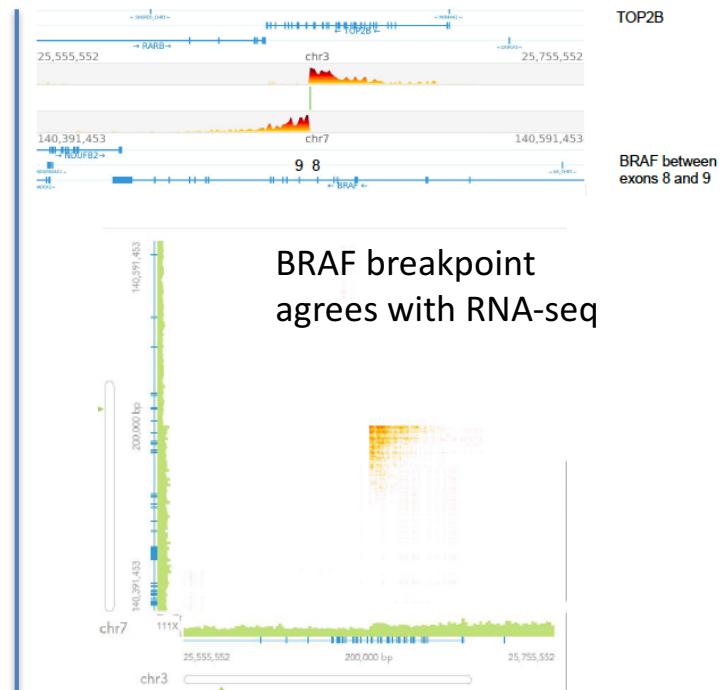
Example results

MET amplification and fusion of *MET* and *ST7*



Matrix view showing linked reads mapping to two different regions of chromosome 7 (y-axis) vs itself (x-axis) and signs of amplification

BRAF-TOP2B gene fusion



BRAF breakpoint agrees with RNA-seq

Ioannis Ragoussis

Mutations in paediatric glioblastoma

LETTER

doi:10.1038/nature10833

Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma

Jeremy Schwartzentruber^{1*}, Andrey Korshunov^{2*}, Xiao-Yang Liu^{3*}, David T. W. Jones⁴, Elke Pfaff⁴, Karine Jacob³, Dominik Sturm⁴, Adam M. Fontebasso³, Dong-Anh Khuong Quang³, Martje Tönjes⁵, Volker Hovestadt⁵, Steffen Albrecht⁶, Marcel Kool⁴, Andre Nantel⁷, Carolin Konermann⁸, Anders Lindroth⁸, Natalie Jäger⁹, Tobias Rausch¹⁰, Marina Ryzhova¹¹, Jan O. Korbel¹⁰, Thomas Hielscher¹², Peter Hauser¹³, Miklos Garami¹³, Almos Klekner¹⁴, Laszlo Bognar¹⁴, Martin Ebinger¹⁵, Martin U. Schuhmann¹⁶, Wolfram Scheurlen¹⁷, Arnulf Pekrun¹⁸, Michael C. Frühwald¹⁹, Wolfgang Roggendorf²⁰, Christoph Kramm²¹, Matthias Dürken²², Jeffrey Atkinson²³, Pierre Lepage¹, Alexandre Montpetit¹, Magdalena Zakrzewska²⁴, Krzysztof Zakrzewski²⁵, Paweł P. Liberski²⁴, Zhifeng Dong²⁶, Peter Siegel²⁶, Andreas E. Kulozik²⁷, Marc Zapata⁵, Abhijit Guha²⁸, David Malkin²⁹, Jörg Felsberg³⁰, Guido Reifenberger³⁰, Andreas von Deimling^{2,31}, Koichi Ichimura³², V. Peter Collins³², Hendrik Witt^{4,27}, Till Milde^{27,33}, Olaf Witt^{27,33}, Cindy Zhang²⁸, Pedro Castelo-Branco²⁸, Peter Lichter⁵, Damien Faury³, Uri Tabori^{28,29}, Christoph Plass⁸, Jacek Majewski³, Stefan M. Pfister^{4,27} & Nada Jabado^{3,34}

Jabado, Pfister and Majewski



Mutations in paediatric glioblastoma

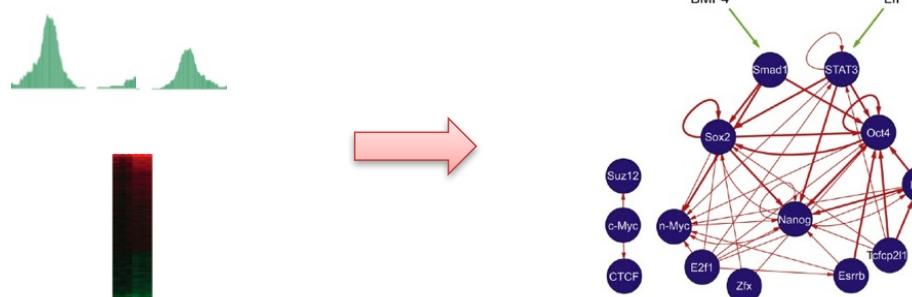
Sample ID	H3F3A	ATRX/DAXX	TP53	IDH1	NF1	PDGFRA
PGBM1	K27M	C1122fs	P152fs;R306X	WT	WT	WT
PGBM2	K27M	WT	R213X*	WT	WT	K385I
PGBM3	K27M	WT	N131del	WT	WT	WT
PGBM4	K27M	K1057fs*	G262fs*	WT	WT	WT
PGBM5	K27M	WT	WT	WT	Y2264fs	WT
PGBM6	K27M	M1800T*	WT	WT	F1247fs;V2230del	WT
PGBM8	K27M	WT	R273C	WT	WT	WT
PGBM9	K27M	WT	R273P*	WT	WT	WT
PGBM10	K27M	WT	WT	T990fs	WT	WT
PGBM11	G34R	S1394fs*	Y163C*	WT	WT	WT
PGBM12	G34R	E1727fs	R342X;R175H	WT	WT	Y849D*
PGBM13	G34R	R1739X*	T256fs*	WT	WT	WT
PGBM14	G34R	E1757X*	R273C;R248Q	WT	WT	WT
PGBM15	G34R	H2254R*	S51delins	WT	WT	WT
PGBM16	G34V	R2111X*	R342X*	WT	WT	K385M
PGBM17	WT	G1589V*	Y220C	R132H	WT	WT
PGBM18	WT	R1426X*	R273C;R196X	R132H	C622X;L1489fs	WT
PGBM19	WT	K1584fs†	R267Q;T230I	WT	R440X	WT
PGBM20	WT	N2443D*	R248W*	WT	WT	WT
PGBM21	WT	R238X (DAXX)	R267W;P152L	WT	R1947X	WT
PGBM22	WT	R1302fs;K1584fs†	R337C;R175H	WT	R2616X;R461Xs‡	WT
PGBM23	WT	WT	I254S	R132H	WT	WT
PGBM24	WT	WT	R196X*	WT	WT	WT
PGBM25	WT	WT	R342X*	WT	G1526fs*	WT
PGBM26	WT	WT	R175H*	WT	Y2264fs*	WT
PGBM27	WT	WT	I251L*	WT	WT	WT
PGBM28	WT	WT	R273H	WT	T676fs	WT
PGBM29	WT	WT	V10G	R132H	WT	WT
PGBM30	WT	WT	G245S*	WT	WT	WT
PGBM31	WT	WT	WT	WT	WT	WT
PGBM32	WT	WT	WT	WT	T1627S	WT
PGBM33	WT	WT	WT	WT	Splicing	WT
PGBM34	WT	WT	WT	WT	D842_I853delinsV	WT
PGBM35	WT	WT	WT	WT	WT	WT
...
PGBM49	WT	WT	WT	WT	WT	WT

Sequenced the exomes of 48 paediatric GBM samples, found:

- Somatic mutations in the H3.3-ATRX-DAXX chromatin remodelling pathway in 44% of tumours
- Recurrent mutations in H3F3A, which encodes the replication-independent histone 3 variant H3.3 in 31% of tumours

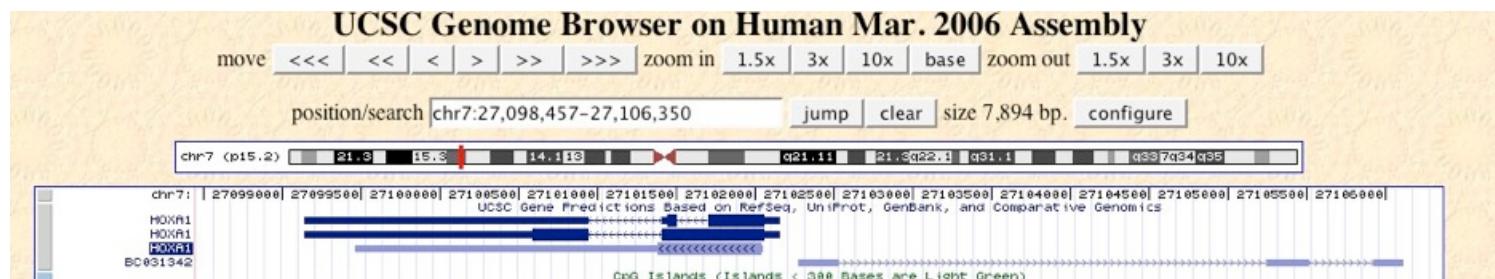
Applications (III)

- Quantitative biology of complex systems
 - New high-throughput technologies in functional genomics: ChIP-Seq, RNA-Seq, ChIA-PET, RIP-Seq, ...
 - From single-gene measurements, to thousands of probes on arrays, to profiles covering all 3B bases of the genome

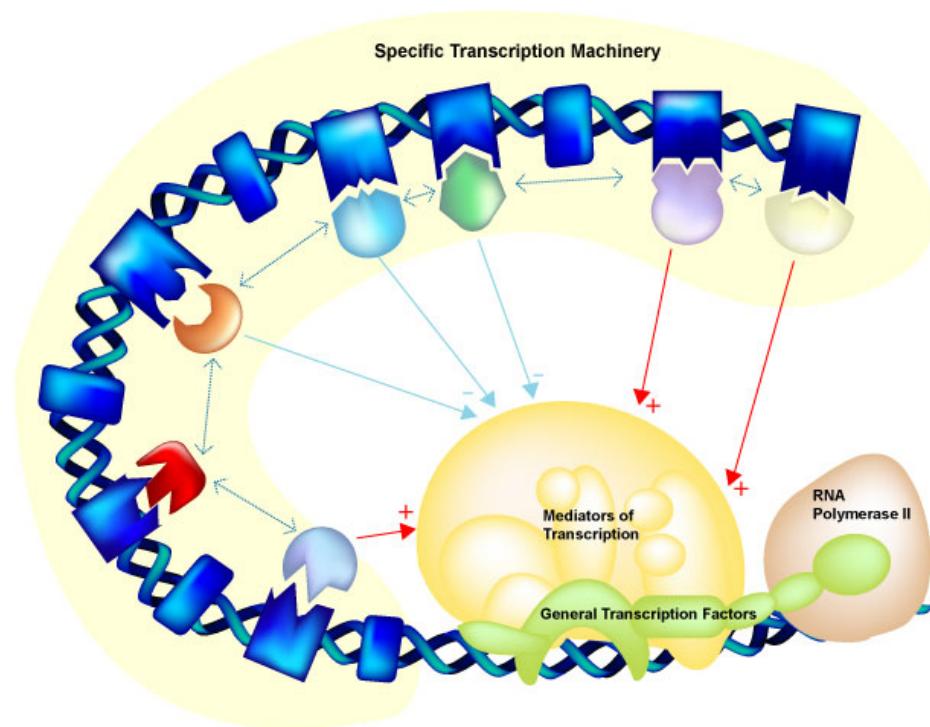


Functional genomics

The next step is to locate all of the genes and regulatory regions, describe their functions, and identify how they differ between different groups (i.e. “disease” vs “healthy”)...

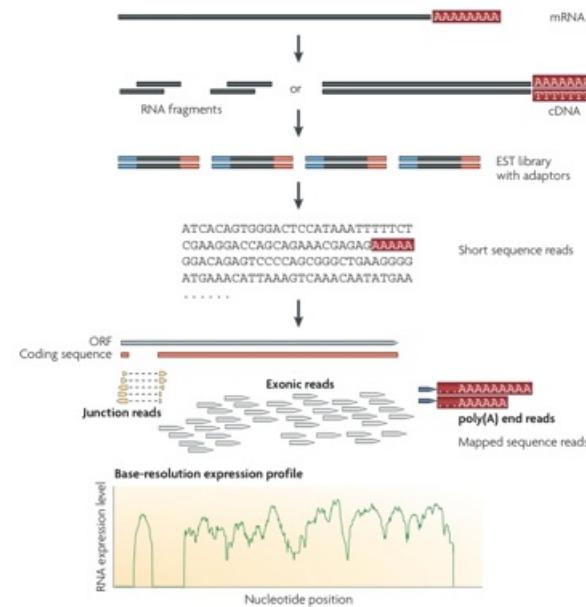


Transcription regulation



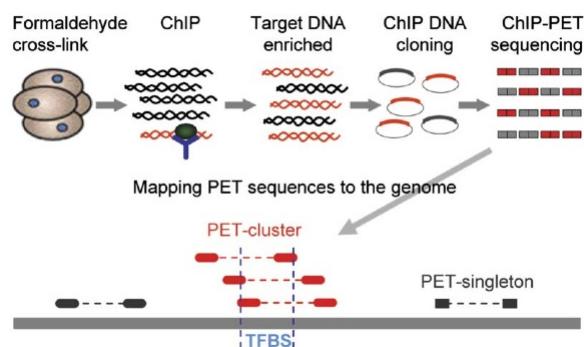
www.genwaybio.com

RNA-Seq: digital expression and much more

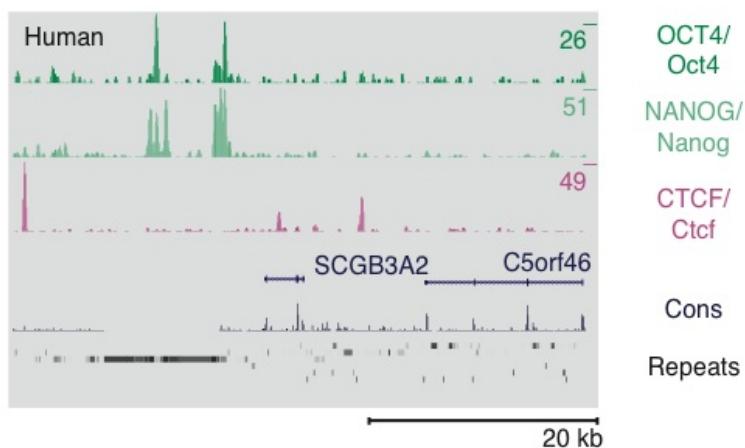


Wang et al. *Nat. Rev. Genet.*, 2009

ChIP-Seq : Genome-wide binding profiles

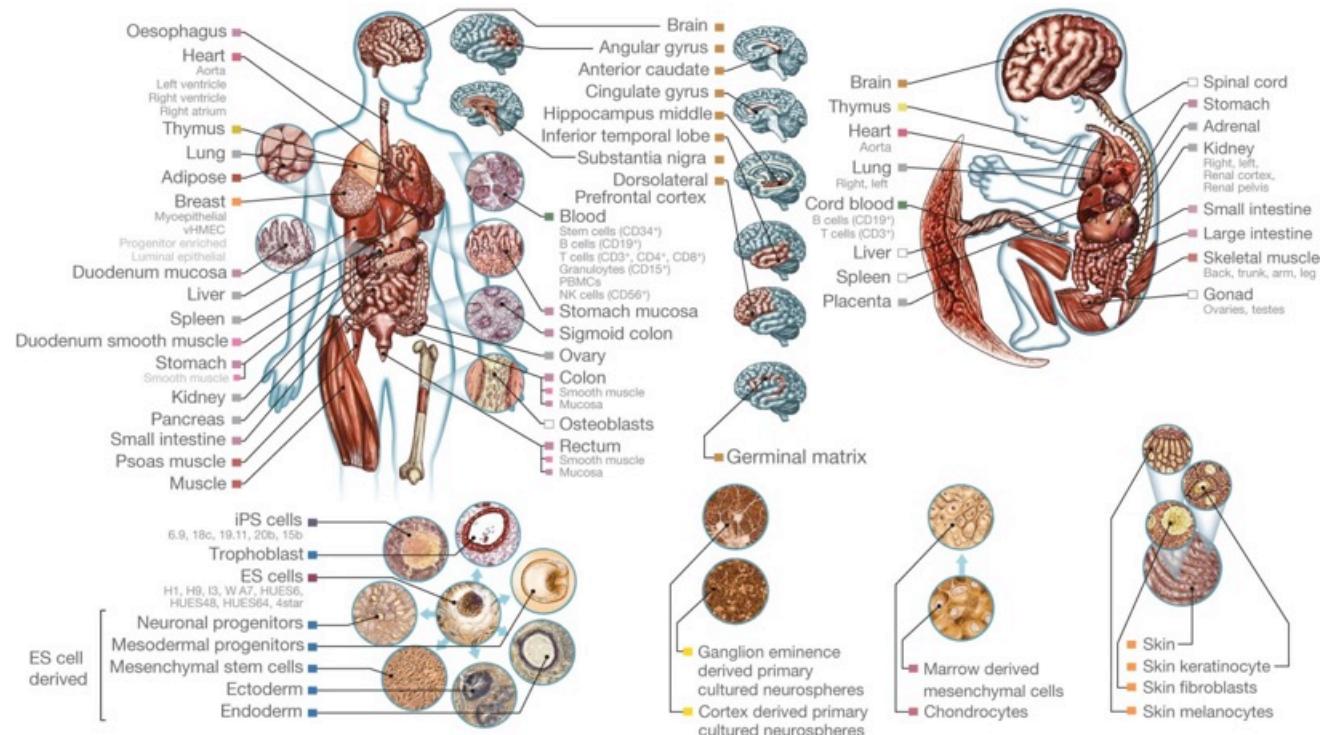


Wei et al., *Cell*, 2006

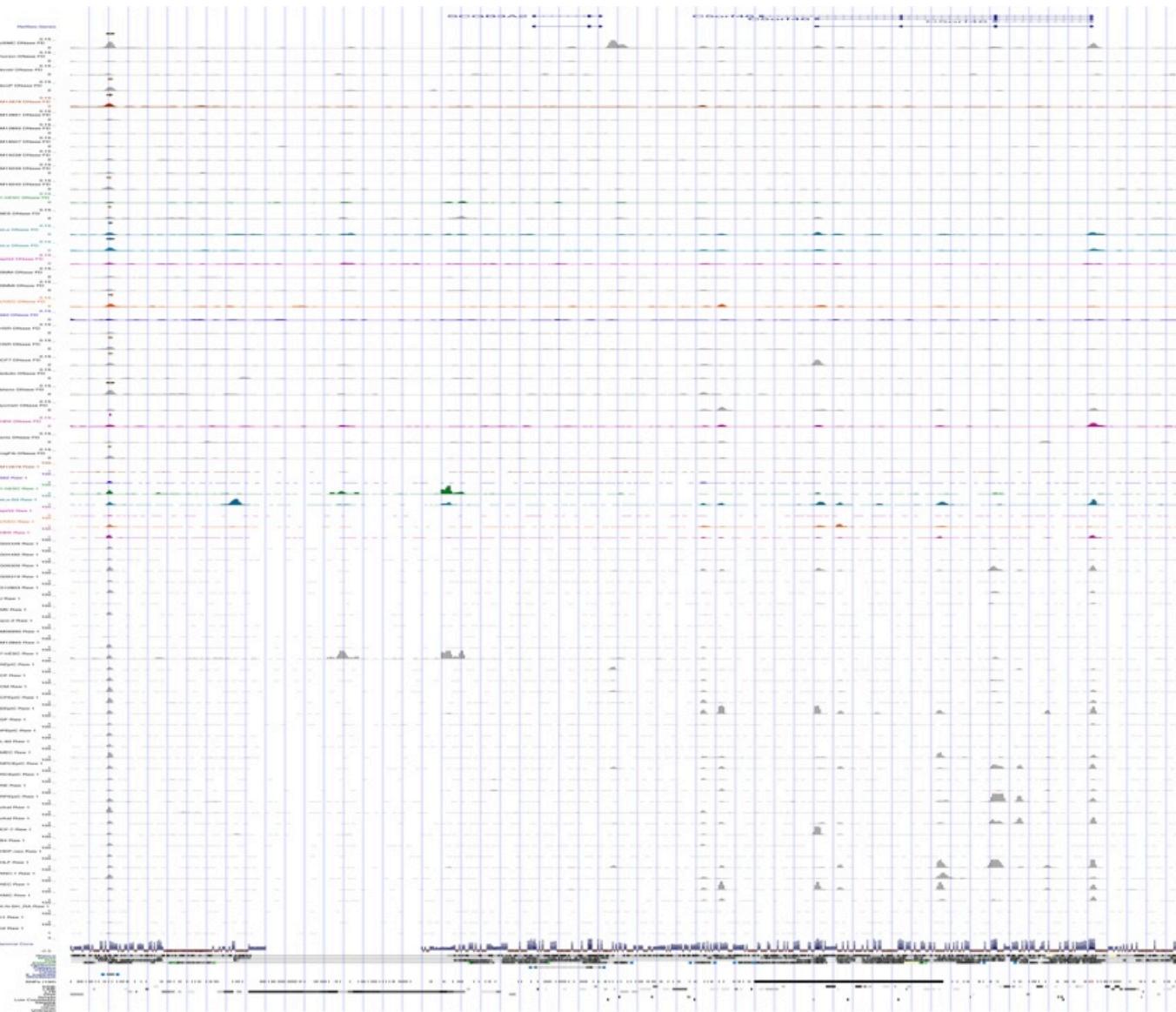


Kunarso et al., *Nat. Genet.*, 2010

One genome... Many epigenomes



Roadmap Epigenomics Consortium et al. *Nature* **518**, 317-330 (2015) doi:10.1038/nature14248



Sequencing human genomes

2001

The
Human
Genome

~ 3 Billion \$

2011

1000
Genomes
Project

~ 10 000 \$

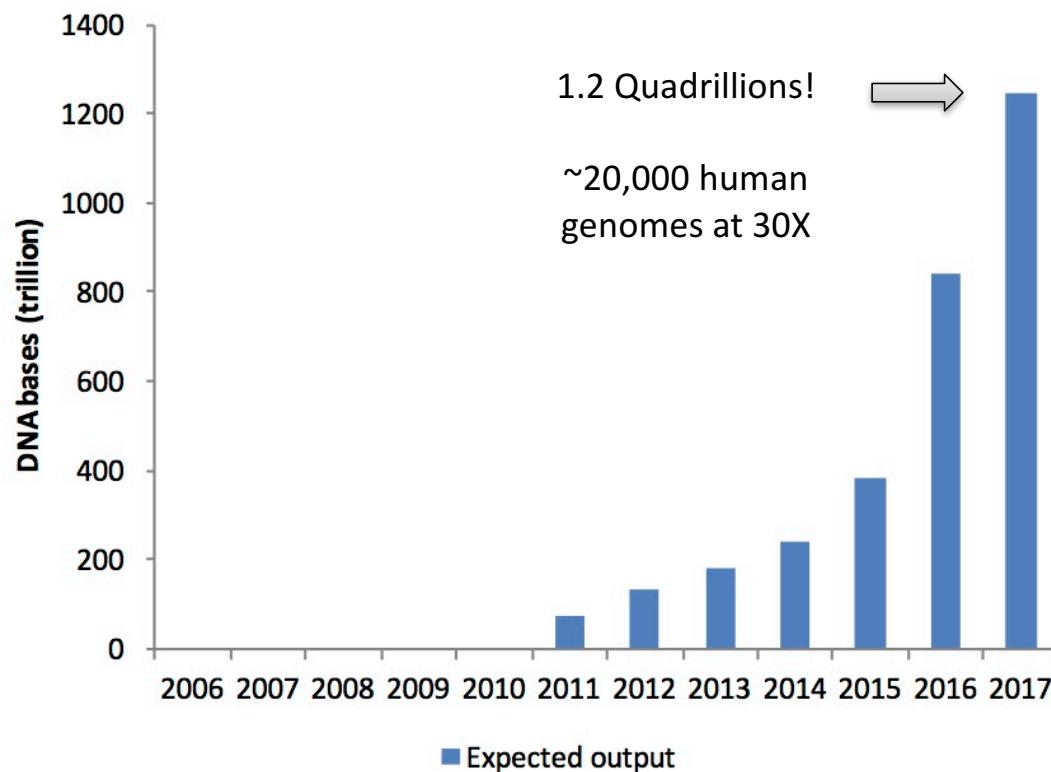
2017

Your
Genome

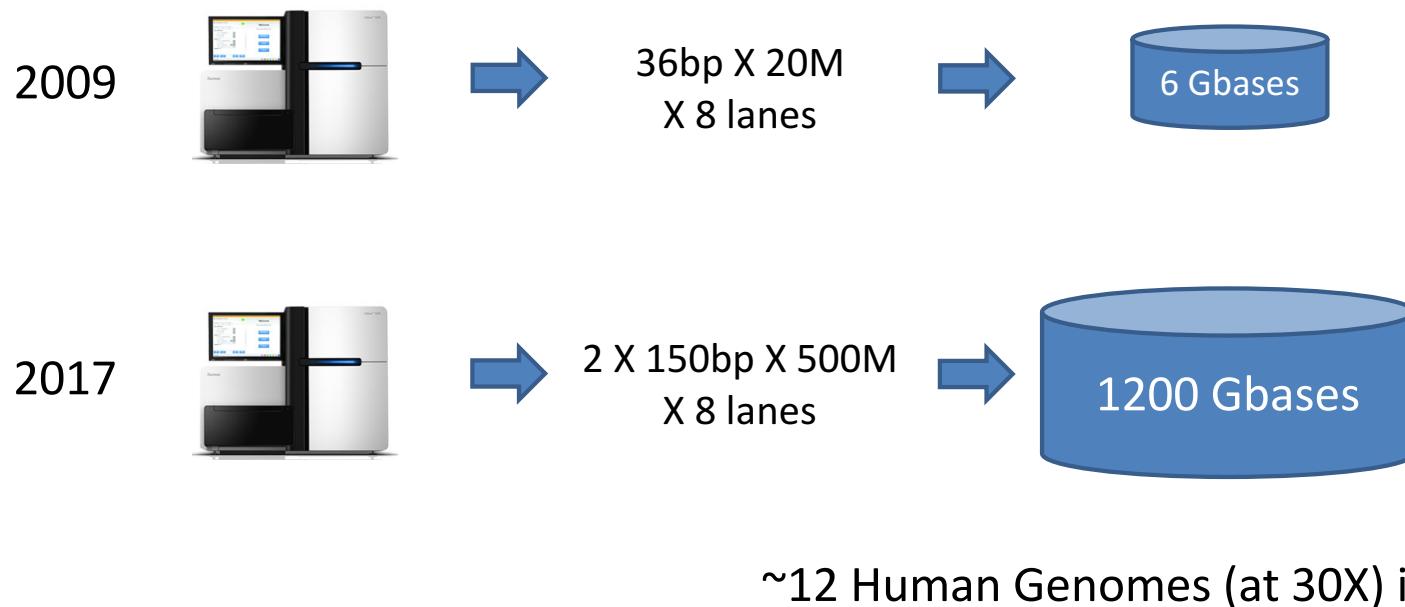
~ 1000 \$

Data Challenges

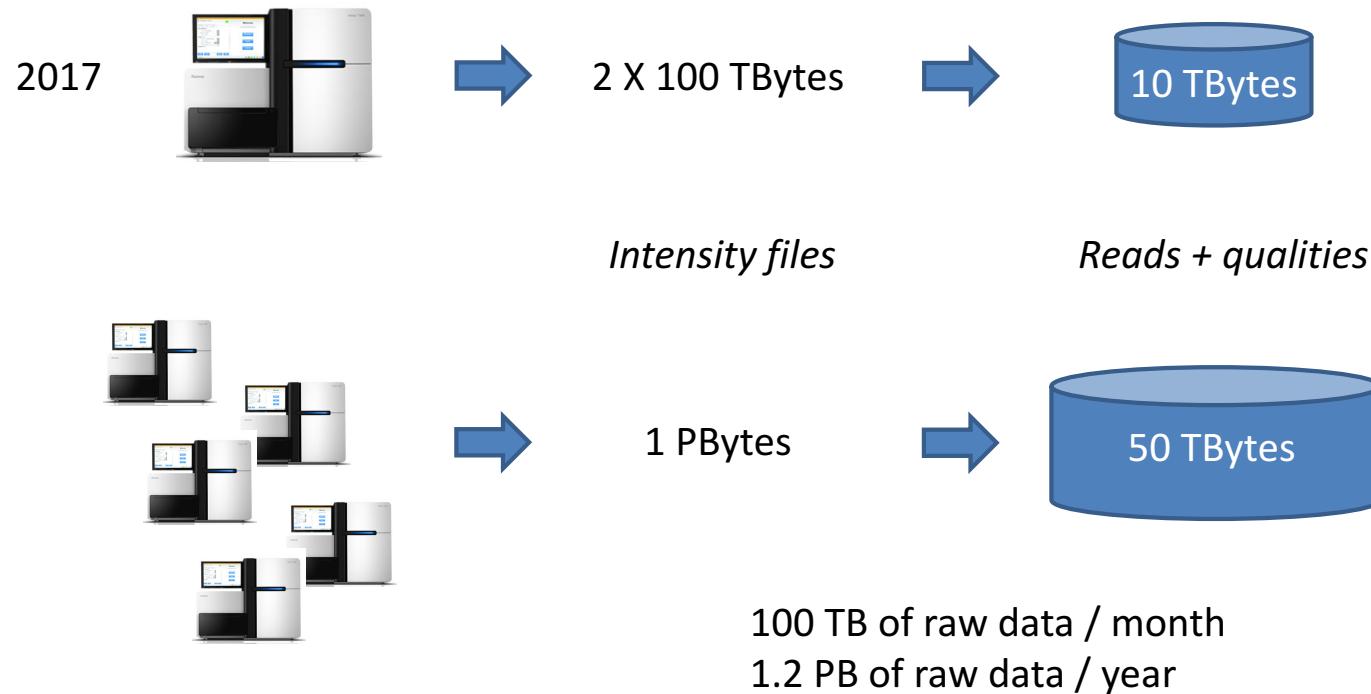
Sequencing capacity at the McGill Genome Center



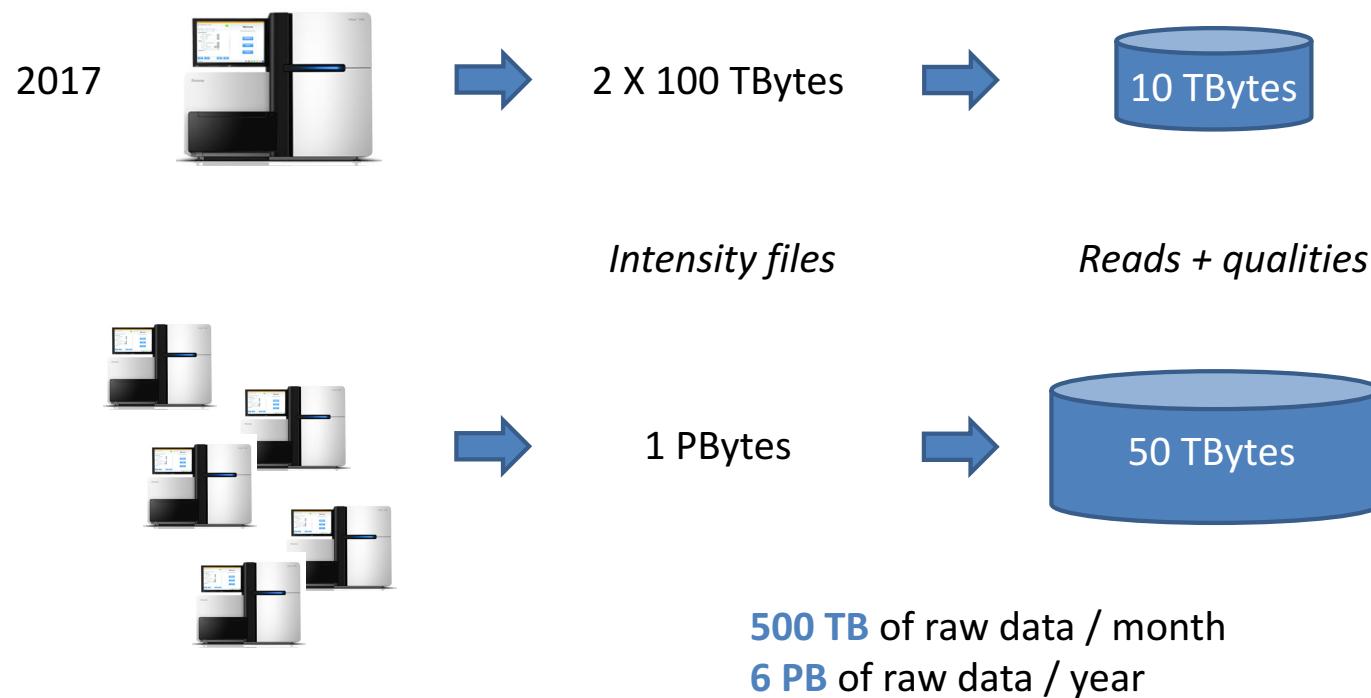
High-throughput sequencing



Big Data



Big Data



Large NGS project

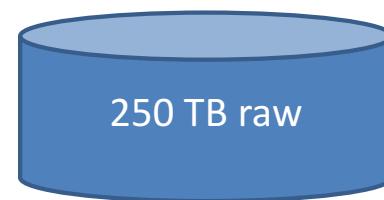
Cancer project with whole genome data:

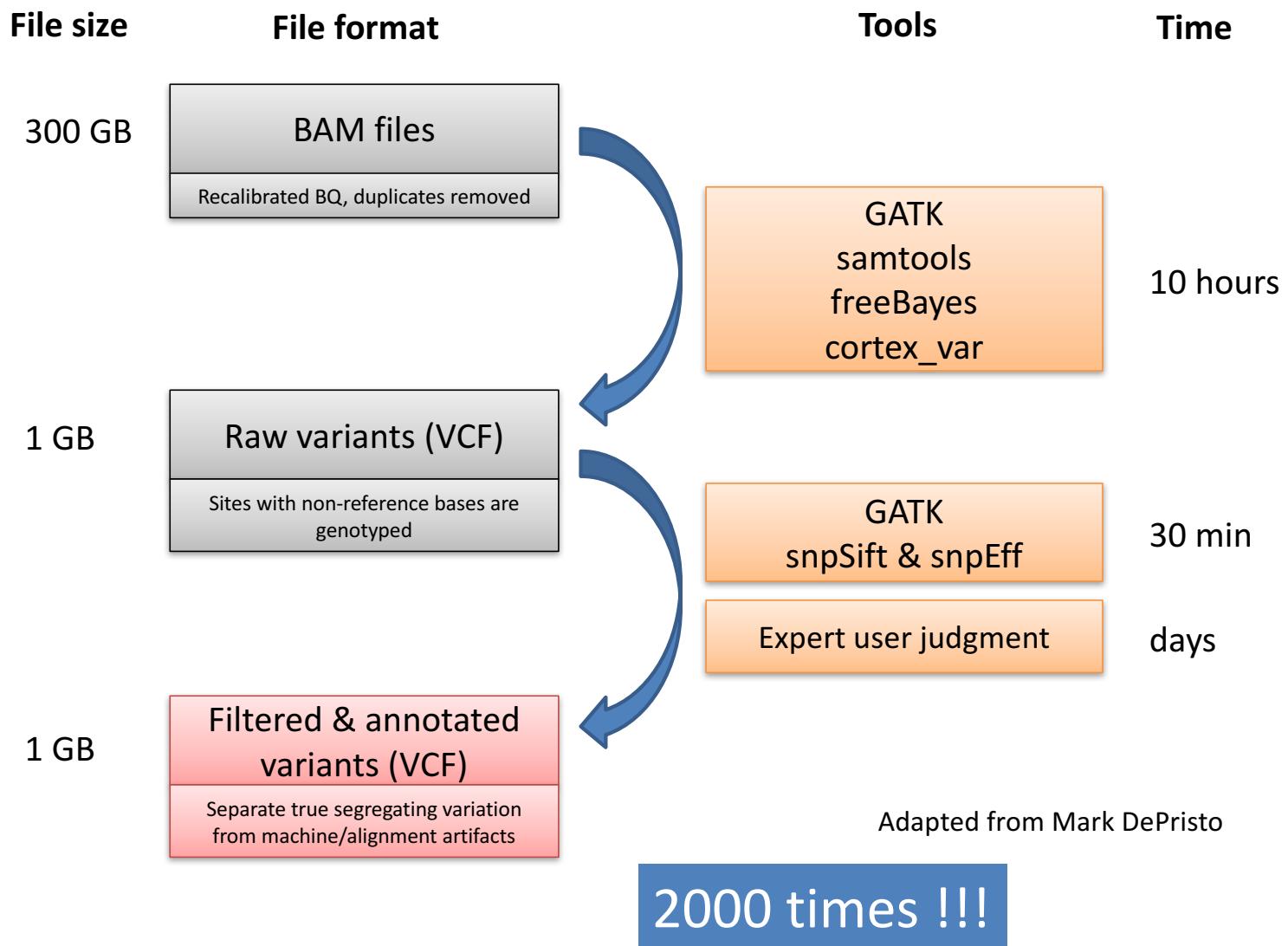
1000 tumors



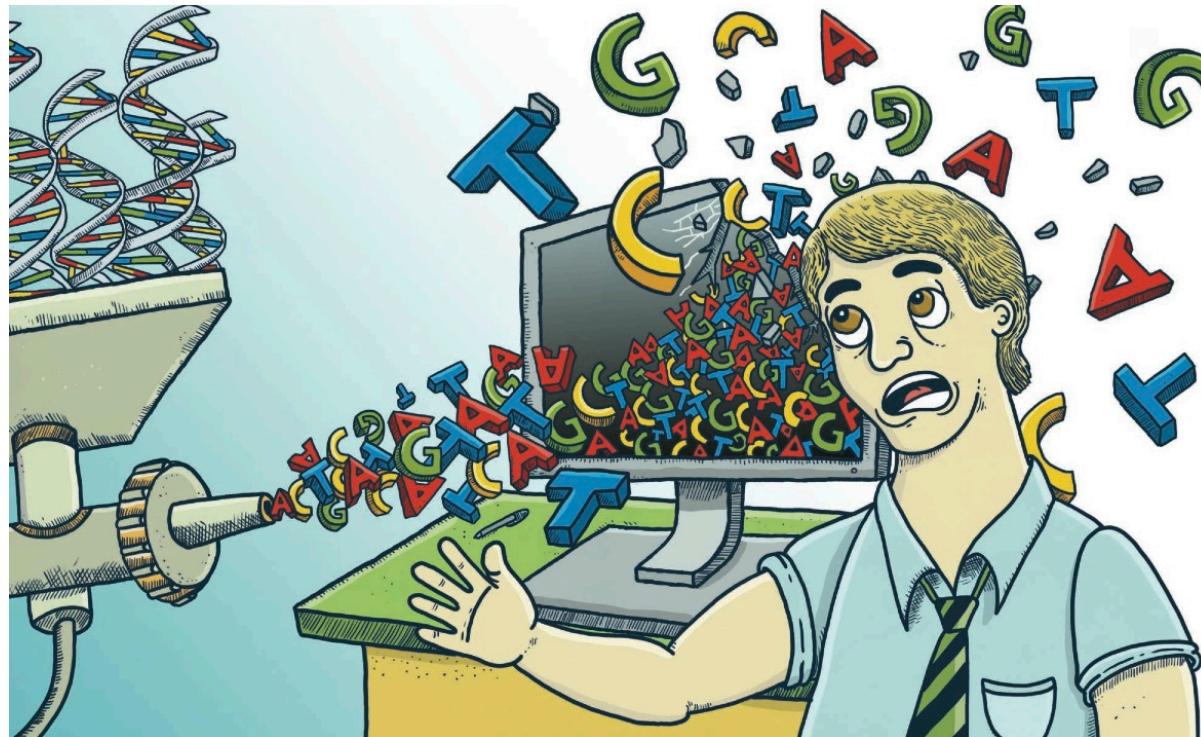
1000 matched-normal

vs





Will computers crash genomics?



Pennisi, Science, 2011

Solutions

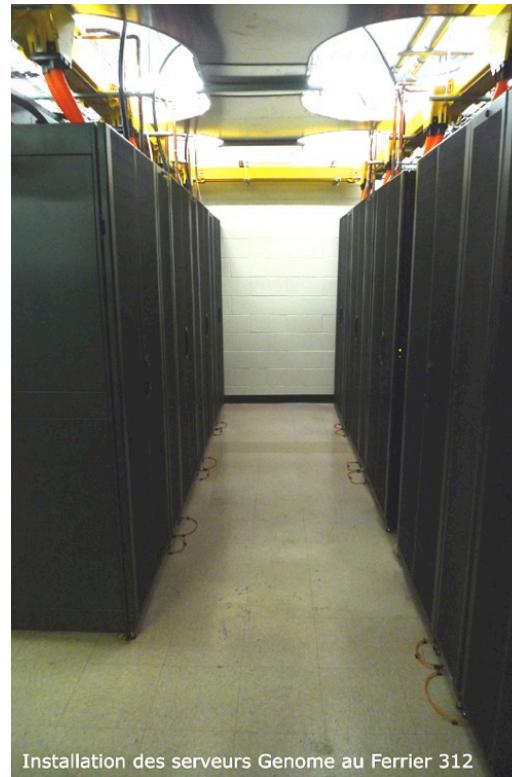
Hardware...

Data center at the Innovation Center

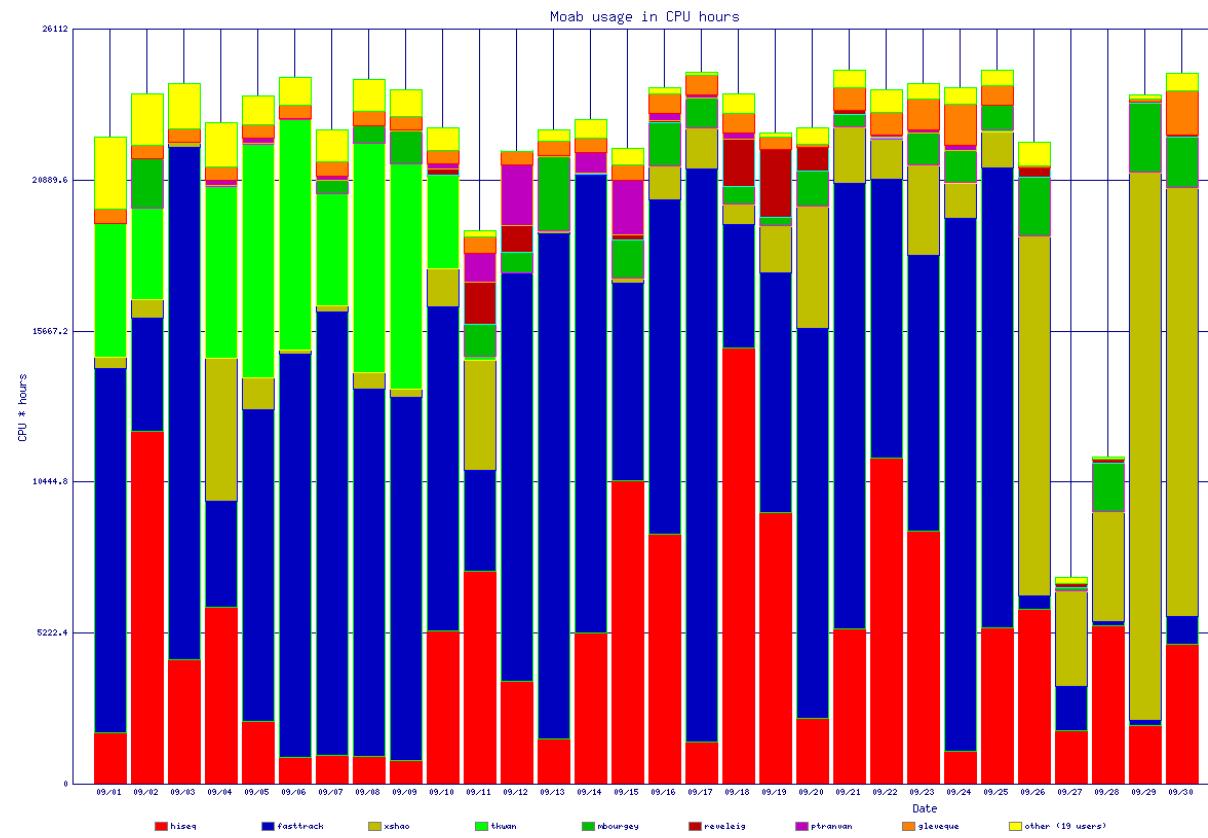


1500 cores
3 PB disk
5 PB tape

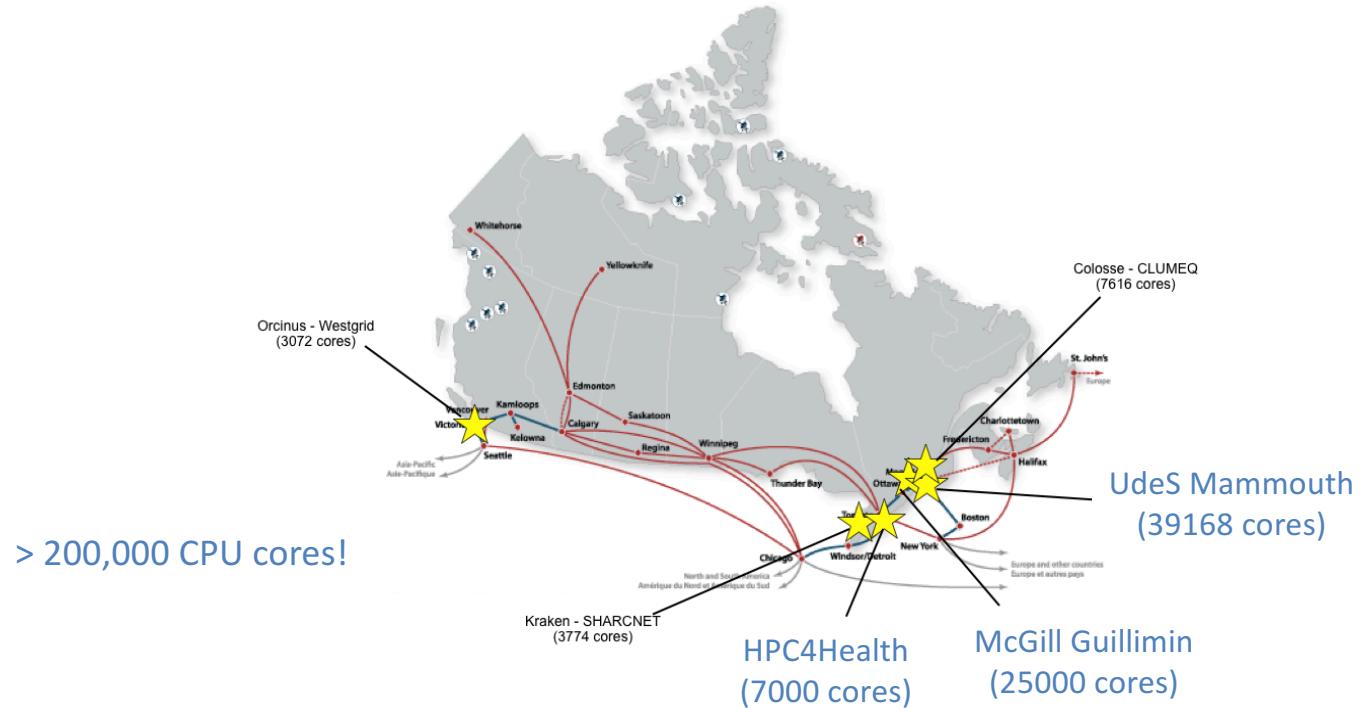
Installed July 2012



Cluster usage



Partnership with Compute Canada



Software...

FastQC

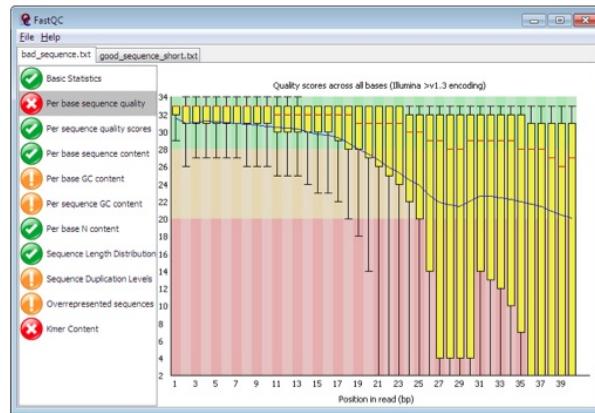
Babraham Bioinformatics

About | People | Services | Projects | Training | Publications

FastQC

Function	A quality control tool for high throughput sequence data.
Language	Java
Requirements	A suitable Java Runtime Environment The Picard BAM/SAM Libraries (included in download)
Code Maturity	Stable. Mature code, but feedback is appreciated.
Code Released	Yes, under GPL v3 or later .
Initial Contact	Simon Andrews

[Download Now](#)



The screenshot shows the FastQC software interface. On the left is a sidebar with various quality control metrics: Basic Statistics (green checkmark), Per base sequence quality (red X), Per sequence quality scores (green checkmark), Per base sequence content (green checkmark), Per base GC content (orange warning), Per sequence GC content (orange warning), Per base N content (green checkmark), Sequence Length Distribution (green checkmark), Sequence Duplication Levels (orange warning), Overrepresented sequences (orange warning), and Kmer Content (red X). The main area displays a bar chart titled "Quality scores across all bases (Illumina >v1.3 encoding)". The y-axis represents quality scores from 2 to 34, and the x-axis represents the position in the read from 1 to 39. The bars are yellow, indicating good quality. A red shaded region highlights the first 15 positions, and a green shaded region highlights the last 15 positions. A black line graph overlays the bars, showing a general decline in quality scores towards the end of the sequence.

FastX

- The FASTX-Toolkit is a collection of command line tools for Short-Reads FASTA/FASTQ files preprocessing (also available via Galaxy, see later).
- The main processing of such FASTA/FASTQ files is mapping (aka aligning) the sequences to reference genomes or other databases using specialized programs.
- It is sometimes more productive to preprocess the FASTA/FASTQ files before mapping the sequences to the genome - manipulating the sequences to produce better mapping results.
- The FASTX-Toolkit tools perform some of these preprocessing tasks.

Adapted from http://hannonlab.cshl.edu/fastx_toolkit/index.html

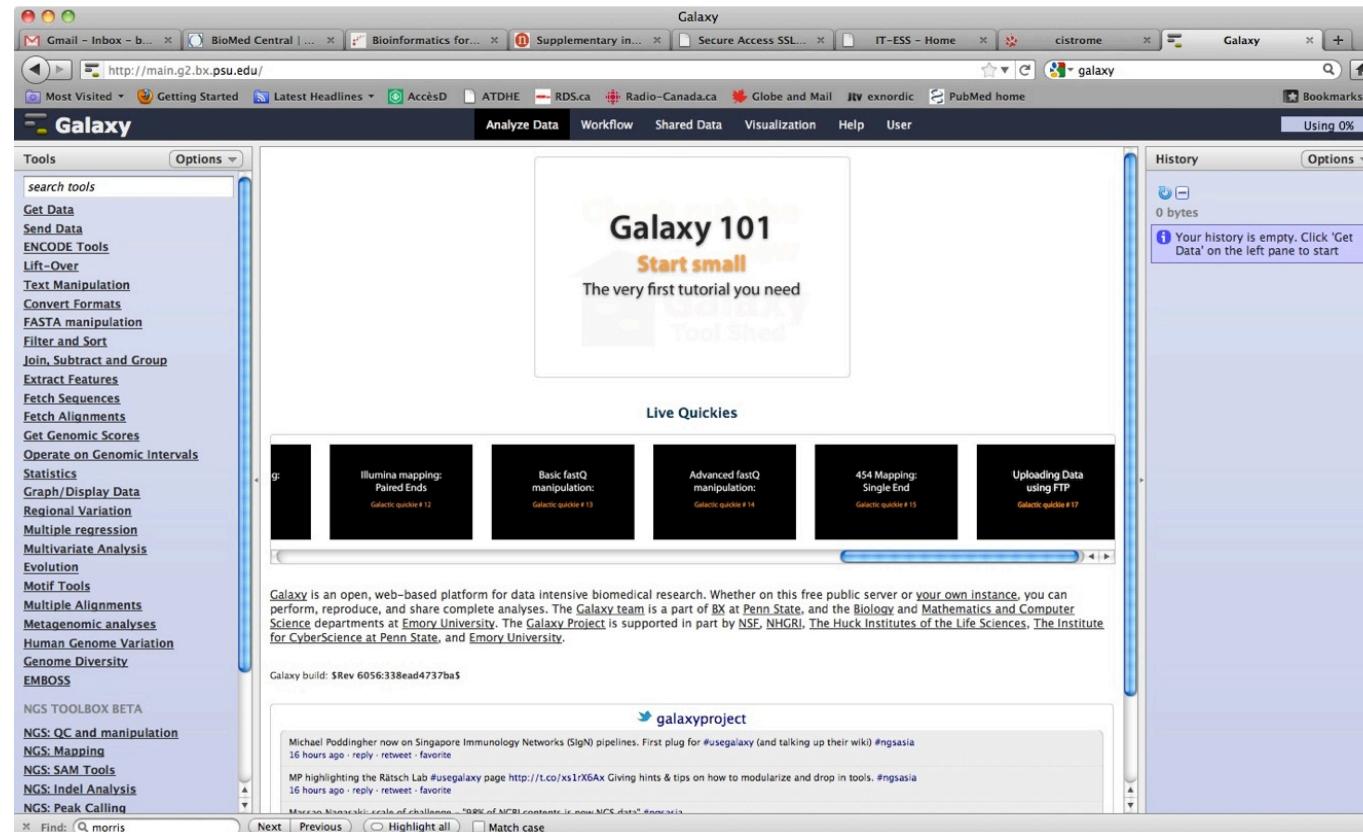
FastX (continued)

Available Tools

- FASTQ-to-FASTA converter : Convert FASTQ files to FASTA files.
- FASTQ Information : Chart Quality Statistics and Nucleotide Distribution
- FASTQ/A Collapser : Collapsing identical sequences in a FASTQ/A file into a single sequence (while maintaining reads counts)
- FASTQ/A Trimmer : Shortening reads in a FASTQ or FASTQ files (removing barcodes or noise).
- FASTQ/A Renamer : Renames the sequence identifiers in FASTQ/A file.
- FASTQ/A Clipper : Removing sequencing adapters / linkers
- FASTQ/A Reverse-Complement : Producing the Reverse-complement of each sequence in a FASTQ/FASTA file.
- FASTQ/A Barcode splitter : Splitting a FASTQ/FASTA files containing multiple samples
- FASTA Formatter : changes the width of sequences line in a FASTA file
- FASTA Nucleotide Changer : Converts FASTA sequences from/to RNA/DNA
- FASTQ Quality Filter : Filters sequences based on quality
- FASTQ Quality Trimmer : Trims (cuts) sequences based on quality
- FASTQ Masker : Masks nucleotides with 'N' (or other character) based on quality

Adapted from http://hannonlab.cshl.edu/fastx_toolkit/index.html

Galaxy



Galaxy NGS

NGS QC and manipulation

- Illumina QC and manipulation
 - FASTQ Groomer convert between various FASTQ quality formats
 - FASTQ splitter on joined paired end reads
 - FASTQ joiner on paired end reads
 - FASTQ Summary Statistics by column
- FASTX-Toolkit for FASTQ data
 - Quality format converter (ASCII-Numeric)
 - Compute quality statistics
 - Draw quality score boxplot
 - Draw nucleotides distribution chart
 - FASTQ to FASTA converter
 - Filter by quality
 - Remove sequencing artifacts
 - Barcode Splitter
 - Clip adapter sequences
 - Collapse sequences
 - Rename sequences
 - Reverse-Complement
 - Trim sequences
- FASTQ QC
 - Fastqc: Fastqc QC using FastQC from Babraham

BEDTools

The BEDTools utilities allow one to address common genomics tasks such as finding feature overlaps and computing coverage. The utilities are largely based on four widely-used file formats: [BED](#), [GFF/GTF](#), [VCF](#), and [SAM/BAM](#). Using BEDTools, one can develop sophisticated pipelines that answer complicated research questions by "streaming" several BEDTools together.

The following are examples of common questions that one can address with BEDTools.

- Intersecting two BED files in search of overlapping features.
- Culling/refining/computing coverage for BAM alignments based on genome features.
- Merging overlapping features.
- Screening for *paired-end* (PE) overlaps between PE sequences and existing genomic features.
- Calculating the depth and breadth of sequence coverage across defined "windows" in a genome.
- Screening for overlaps between "split" alignments and genomic features.

<http://code.google.com/p/bedtools/>

GenAP

The screenshot shows the homepage of the Genetics and Genomics Analysis Platform (GenAP). The header includes the GenAP logo, the tagline "The Computing Gateway for Life Sciences", and links for About, Contact Us, and Logout. The main menu has options for Home, My Projects, My Applications, Tools, Help, and a user profile for Guillaume Bourque. Below the menu is a grid of icons for various features: My Projects, My Applications, My Settings, My Usage, Pipelines, Genome Browser, Manage Files, Public Data, GenAP Hosts, and Help. To the right, there's a "Latest News" section about the official release, and a "My Messages" section with a welcome message for a new project member.

www.genap.ca

- Maintain >80 genomic tools and >18 datasets
- Advanced bioinformatics pipelines
- Easy way to analyze genomics data (Galaxy)
- Automatic distribution of software code (CVFMS)
- All running on Compute Canada resources



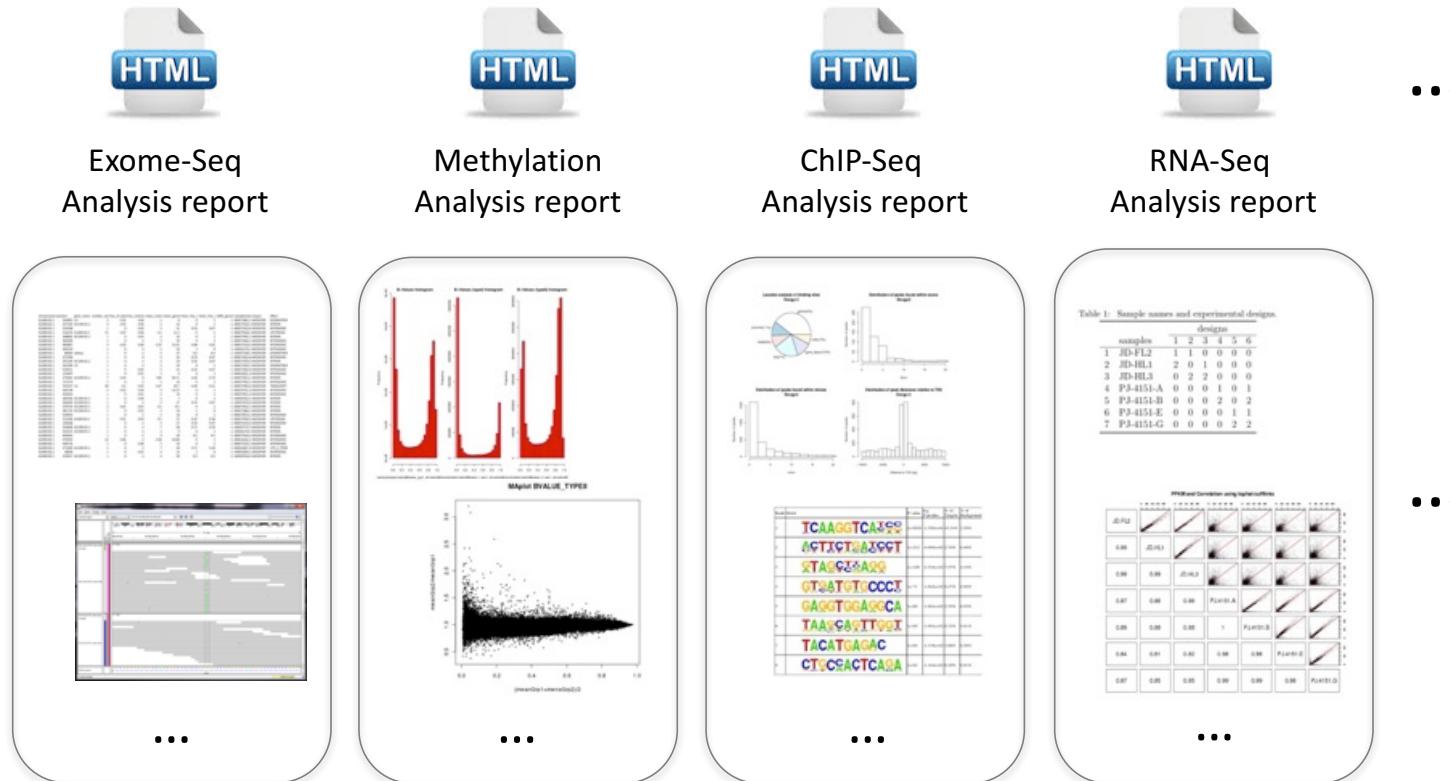
GenAPpipes

- **RNA-Seq** Spliced alignment, QC, differential analysis, isoform analysis, ...
- **ChIP-Seq** Narrow/Wide peaks, Homer, GoSeq, other annotations, ...
- **RNA-Seq Denovo**, differential analysis, QC, transcript annotations, ...
- **DNA-Seq** Alignment, Realignment, MarkDup, Recalibration, SNV, CNV, SV, ...
- **Pacbio Denovo**, bacteria and genomes up to ~50Mb, annotations (in-progress)

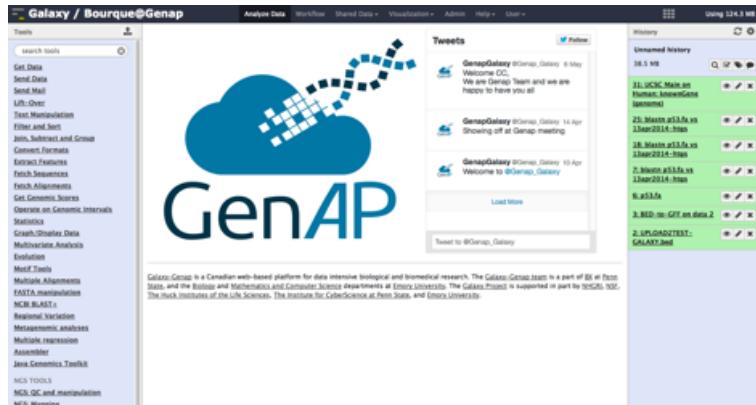
- All pipelines are optimised for the hardware and schedulers at the different cluster sites (the configuration is adjustable)
- All pipelines include an HTML report with the references, explanations and details on the sequencing and analysis



Standardized analysis reports



Private instances of Galaxy on Compute Canada

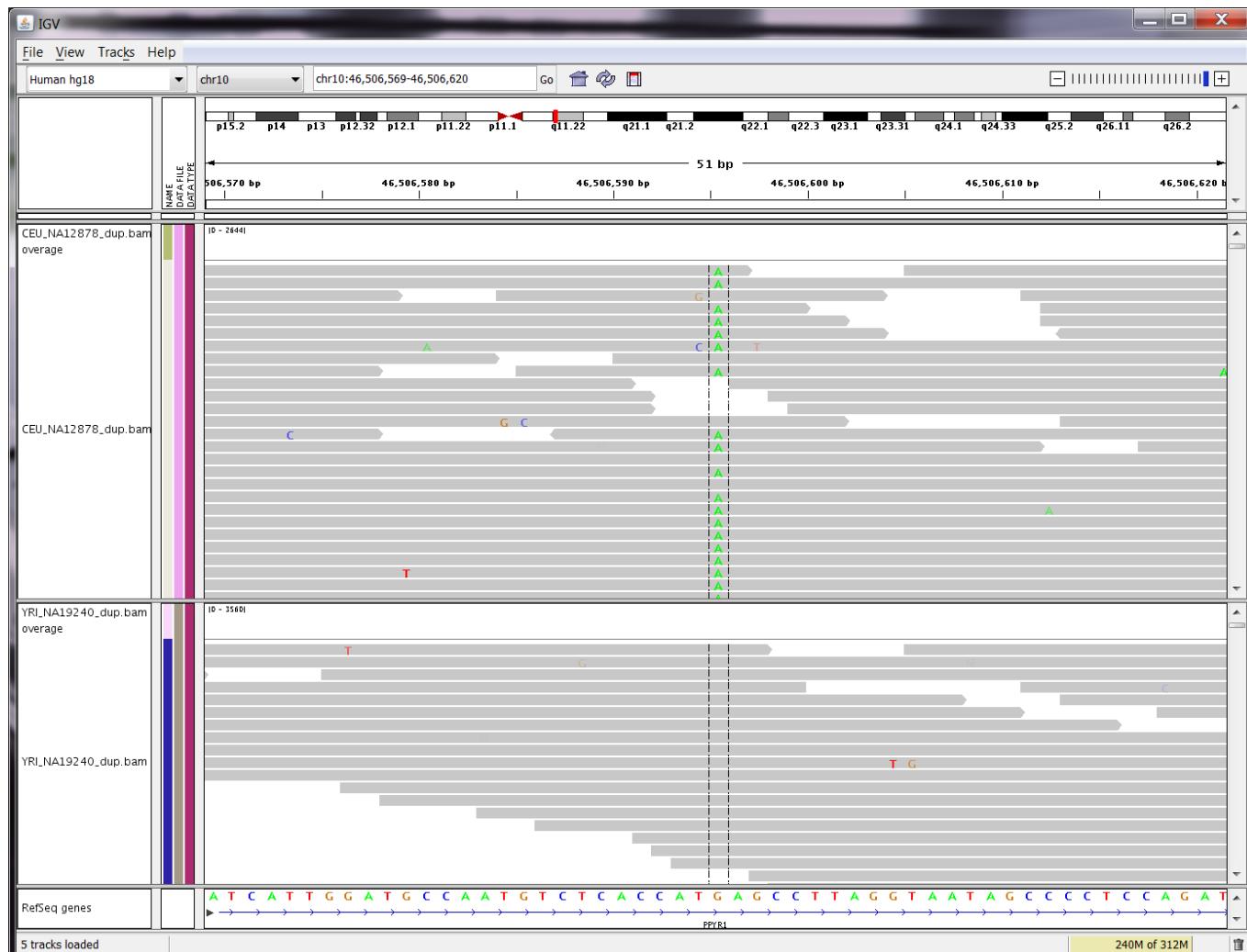


The screenshot shows the Galaxy web interface for the GenAP project. On the left, a sidebar lists various bioinformatics tools and NCS tools. The main area features a large blue cloud icon with the text "GenAP" overlaid. Below the icon, a "Tweets" section displays three recent tweets from the "GenapGalaxy" account. To the right, a "Workflow History" section shows a list of completed workflows, each with a thumbnail, name, and file size (e.g., "21: NCS: Main.maf Human_knowGene.bam.indx", "22: Main.a13fa.vg", etc.).



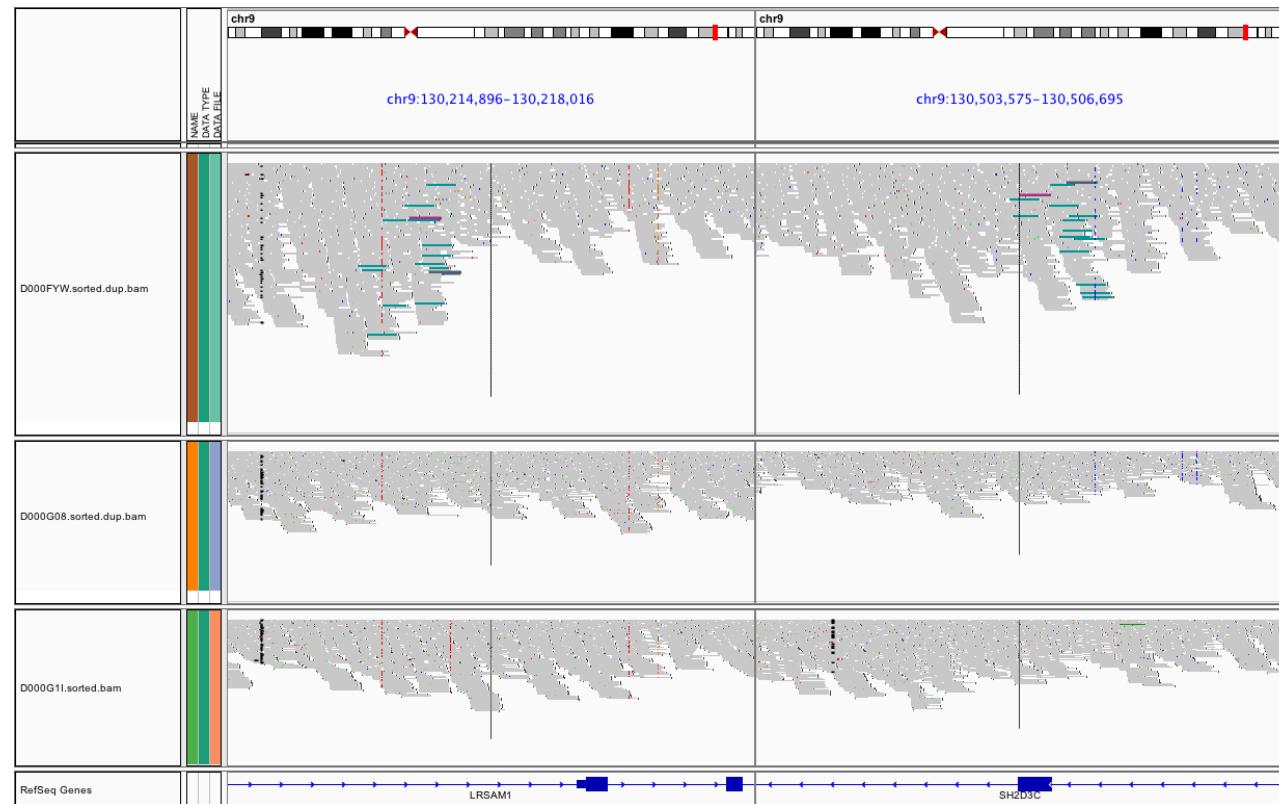
A photograph of a computer lab at the University of Emory. Several people are seated at desks, each with multiple computer monitors. The room has large windows overlooking a cityscape. A presentation is visible on a screen in the background.

- Ability to create/run/share single jobs or complex workflows.
- The user does not need to have programming skills to run jobs.
- Over 1,500 compute nodes and ~40,000 cores available.
- Tools are configured to run in compute nodes with 32,256,512 GB RAM, according to their necessity.



Michael Stromberg, bioinformatics.ca

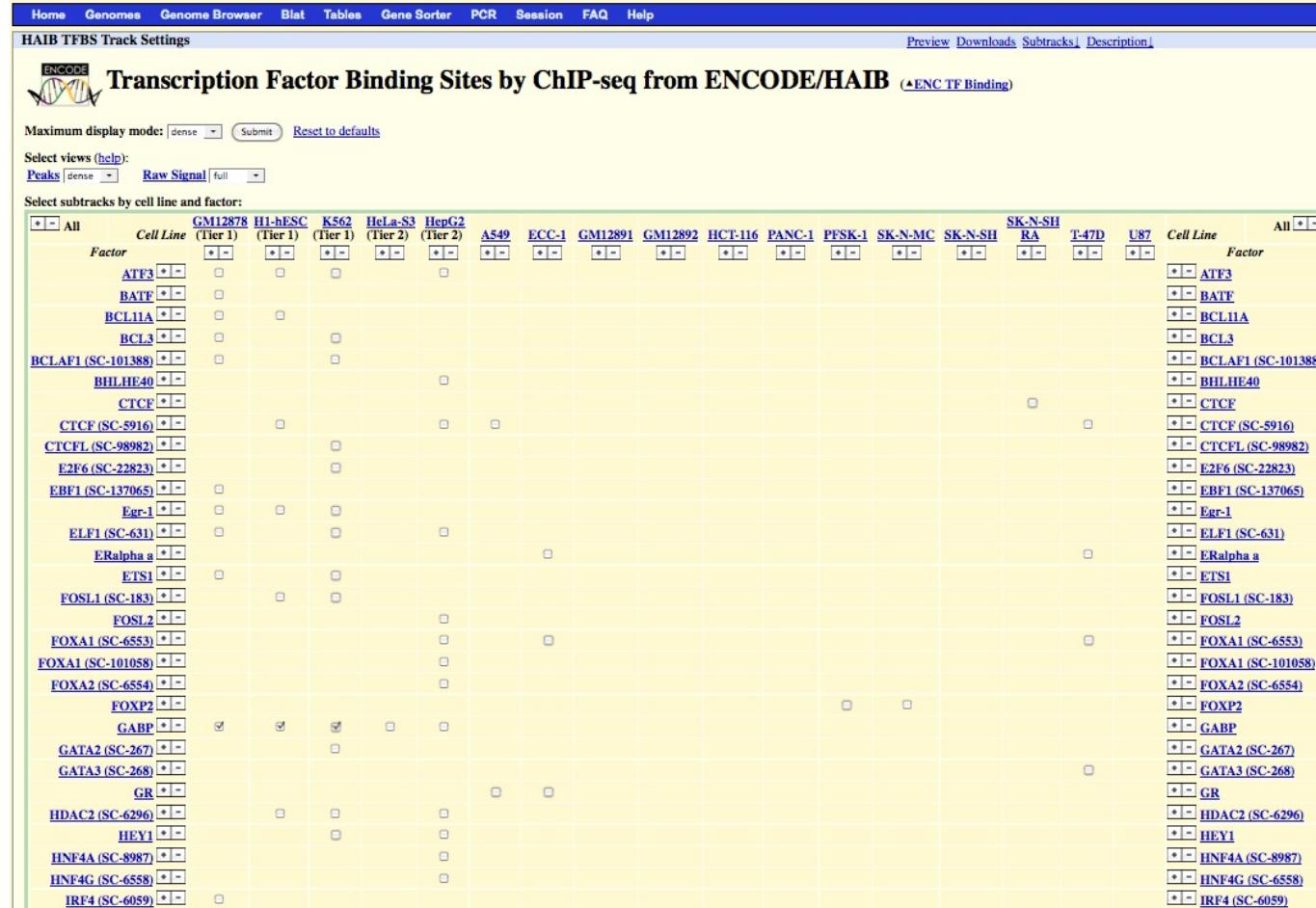
IGV - Fusion gene



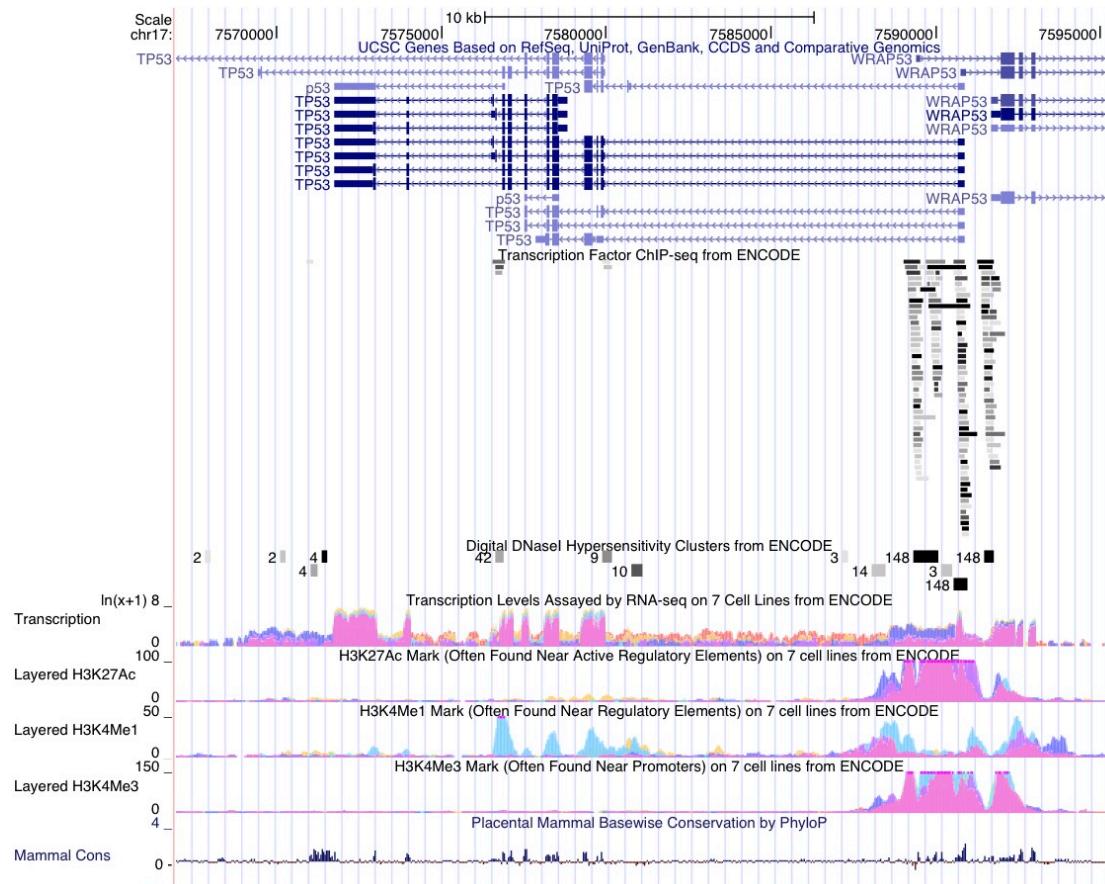
About UCSC Genome Browser

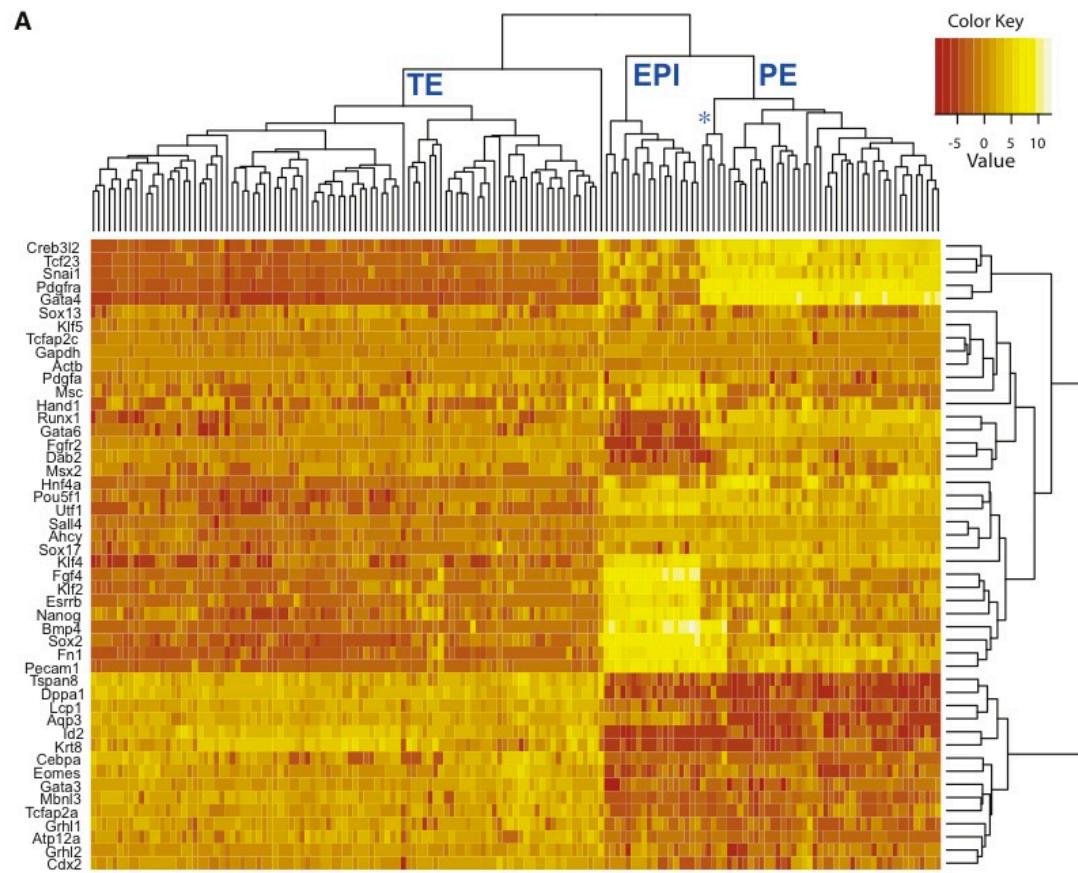
- Browse many Eukaryotic genomes (yeast to human)
- Most annotations are there
- Important evolutionary and variation data representation.
- Very flexible and configurable views
- Graphical and table views (Galaxy also uses this)
- Upload your data into custom tracks and share with colleagues
- Client/server application with it's issues, but a great app!

Some of the ENCODE ChIP-Seq

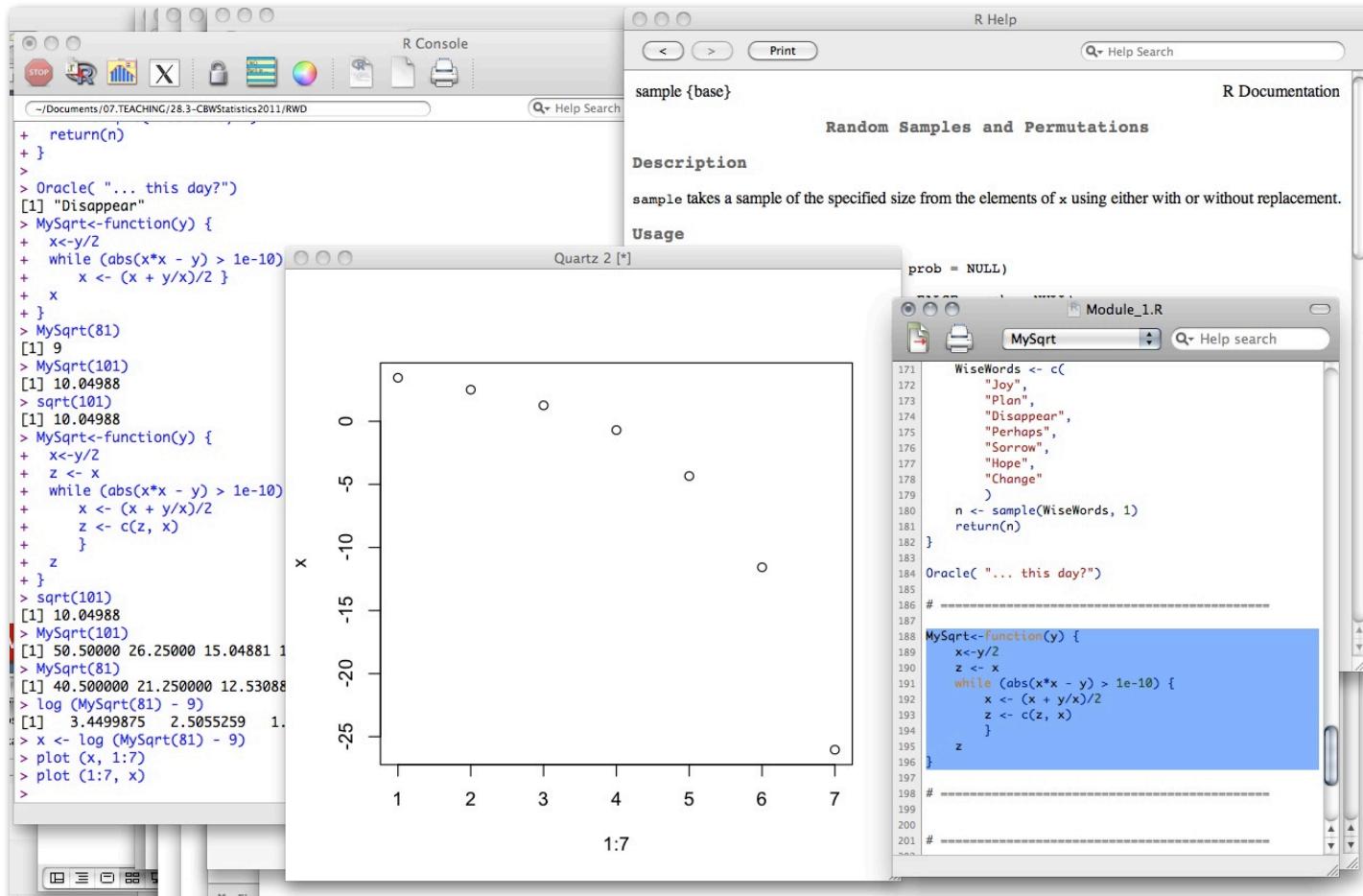


P53 + ENCODE





R



Boris Steipe, bioinformatics.ca



R Graph Gallery



<http://www.r-graph-gallery.com>

WWW

- SEQAnswers <http://seqanswers.com/>
- SEQWiki <http://seqanswers.com/wiki/SEQanswers>
- Mailing lists (e.g. genome@soe.ucsc.edu, Galaxy, etc.)

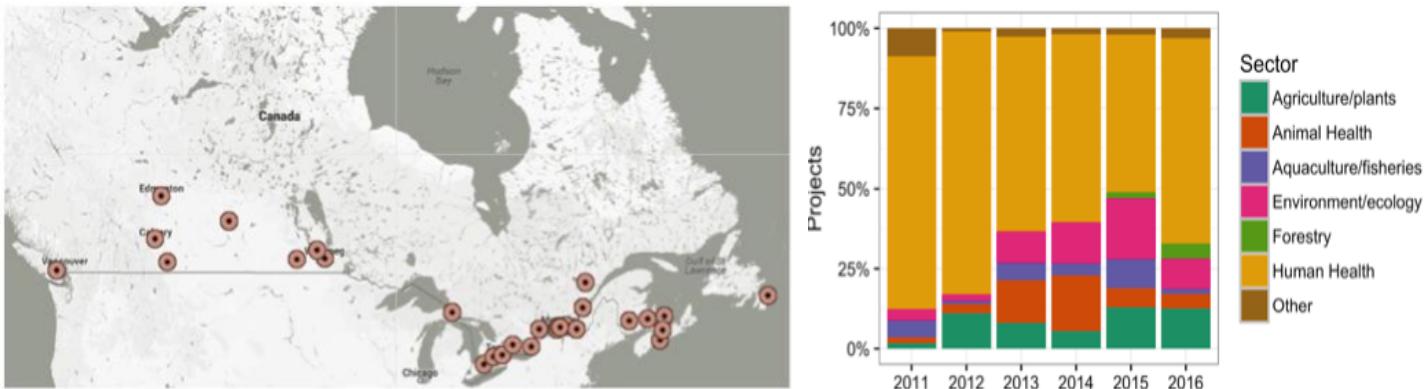
Other...



Canadian Centre for Computational Genomics

- Objectives:
 - Offer state-of-the-art bioinformatics services
 - Install and distribute a bioinformatics software suite on Compute Canada
 - Further develop open-source bioinformatics software solutions
- Lead by:
 - Guillaume Bourque (McGill Univ.)
 - Michael Brudno (Univ. Toronto)
- Launched in 2015 and funded by Genome Canada

Bioinformatics analysis services



- **Fee-for-Service analysis** - 100+ projects per year for 40+ institutions
- **Free consultation** – expert guidance for project design and analysis

Different types of analyses

Services

Whole genome-seq • Exome-seq • ChIP-seq • Marker-based metagenomics

RNA-seq • de novo RNA-seq • Methyl-seq • Shotgun metagenomics

Metatranscriptomics • Cancer Analysis • miRNA-seq • CRISPR • HPC

Large & small genomes assembly and annotation • Expression microarrays

Methylation microarrays • Functional genomics • Pool-seq • GBS • and more

<http://www.computationalgenomics.ca/>

Training activities

- **2017 past workshops**

- CFIA Genomic Virology Workshop, Montreal, Qc
- High-Throughput Biology: From Sequence to Networks, CSHL, NY
- Bioinformatics of Genomic Medicine, Toronto, On
- Informatics on High-Throughput Sequencing Data, Toronto, On
- Epigenomic Data Analysis, Montreal, Qc
- Cancer Genomics workshop, Hinxton, UK
- Kyoto Course on Bioinformatics, Kyoto, JPN

- **Open door session**

- Every Thursday 3-4pm, Genome Center, room 4200
- Meet experts to discuss about your analysis



Canadian Centre for
Computational
Genomics



Acknowledgements

Lab

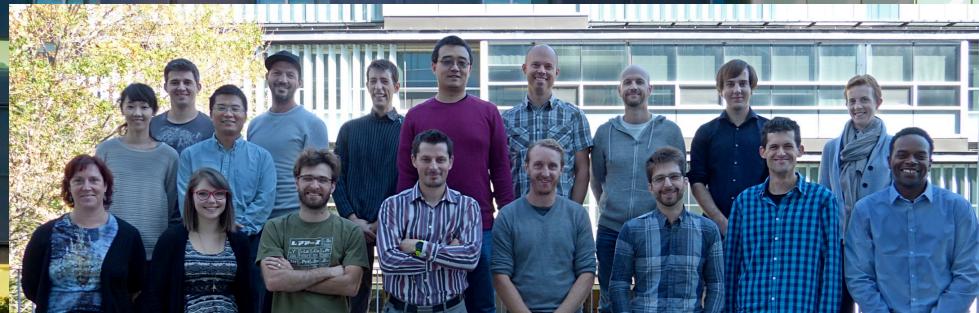
LeeAnn Ramsay
Jean Monlong
Simon Girard
Toby Hocking
Patricia Goerner-Potvin
Joe Sue
David Venuto

Analysis team

Mathieu Bourgey
Gary Lévesque
Robert Eveleigh
Francois Lefebvre
Johanna Sandoval
Pascale Marquis
Edouard Henrion

IHEC Data Portal and GenAP

David Bujold
Catherine Côté
Bryan Caron
Kuang Chung Chen
Simon Nderitu
ME Rousseau
Pierre-Étienne Jacques (UdeS)
David Morais (UdeS)
Carol Gauthier (UdeS)
Alain Veilleux (UdeS)
Maxime Levesque (UdeS)
Michel Barrette (UdeS)



guil.bourque@mcgill.ca

