BIO392
Bioinformatics of Genome Variations

Genome Variation | Function | Data Formats | Resources | Privacy

Michael Baudis **UZH SIB**
Computational Oncogenomics

University of Zurich[UZH]

1992 — Heidelberg

Student of medicine | doctoral thesis in molecular cytogenetics @ DKFZ (Peter Lichter) | resident in clinical hematology/oncology | data, clinical studies & cancer systematics

2001 — Stanford

Post-doc in hemato-pathology (Michael Cleary) | molecular mechanisms of leukemogenesis | transgenic models | expression arrays | systematic cancer genome data collection | *Progenetix* website

2003 — Gainesville

Assistant professor in paediatric haematology | molecular mechanisms of leukemogenesis | focus on bioinformatics for cancer genome data analysis

2006 — Aachen

Research group leader in genetics | genomic array analysis for germline alterations | descriptive analysis of copy number aberration patterns in cancer entities

2007 — Zürich

Professor of bioinformatics @ IMLS (2015) | systematic assembly of oncogenomic data | databases and software tools | patterns in cancer genomes | *arraymap* online resource | GA4GH | SPHN

BIO392: Course Schedule

- Introduction, File Formats & Genome Browsers (Michael Baudis)
  - 2019-09-17 (Tue), 13-17
  - 2019-09-18 (Wed), 09-17
  - 2019-09-19 (Thu), 09-17
  - 2019-09-20 (Fri), 09-17
- Tools & Programmatic Solutions (Izaskun Mallona)
  - 2019-09-24 (Tue), 13-17
  - 2019-09-25 (Wed), 09-17
  - 2019-09-26 (Thu), 09-17
  - 2019-09-27 (Fri), 09-17
- Genome Variants to Modified Proteins (Elif Özkirimli Olmez)
  - 2019-10-01 (Tue), 13-17
  - 2019-10-02 (Wed), 09-17
  - 2019-10-03 (Thu), 09-17
  - 2019-10-04 (Fri), 09-17
- Review & Exam
  - 2019-10-08 (Tue), 13-17
  - 2019-10-09 (Wed), 09-12
    - Written exam
    - Feedback

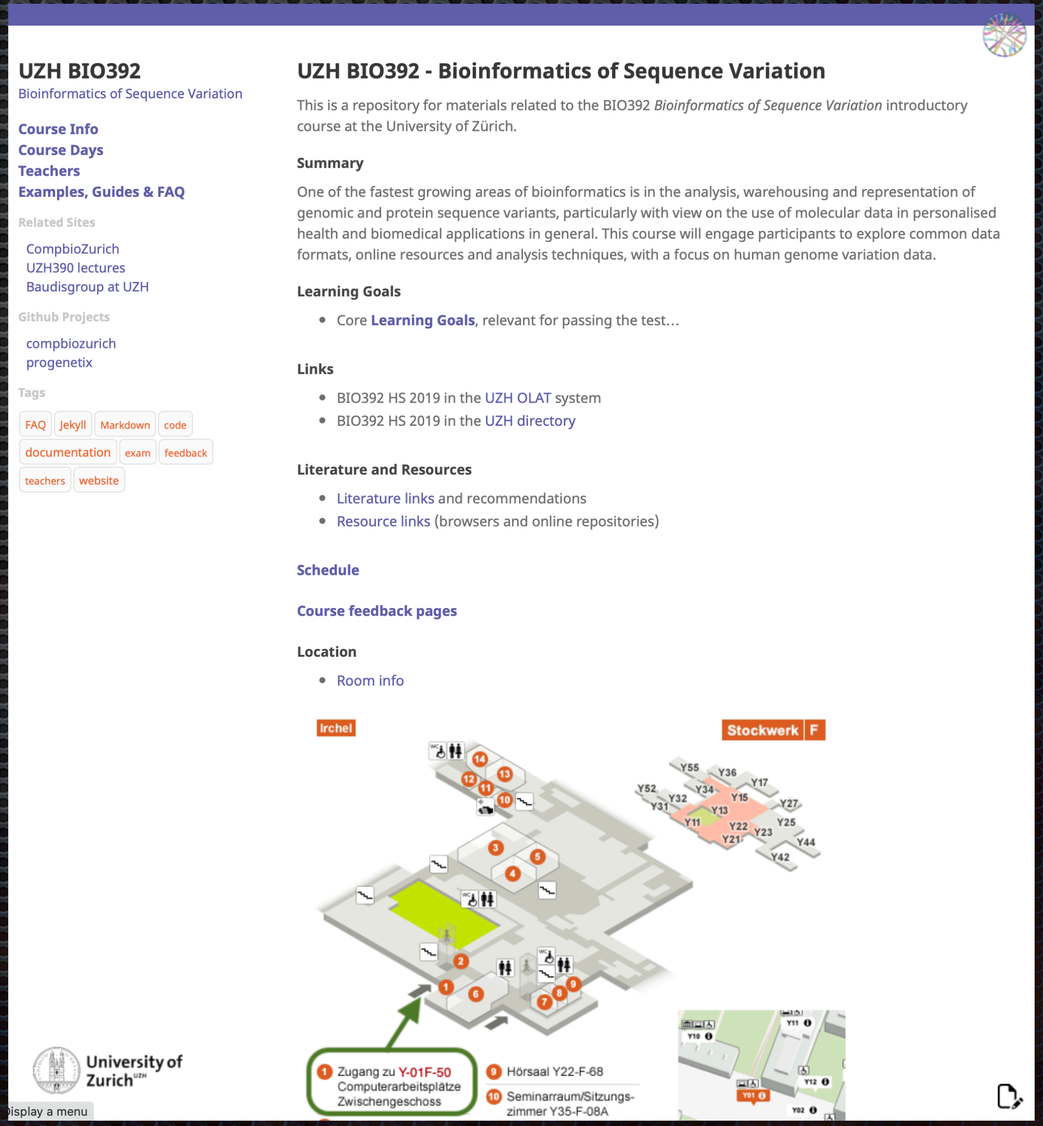
University of Zurich[UZH]

BIO392: Course Resources

University of Zurich[UZH]

- Course repository & website on Github
  - links to articles and information resources
  - downloads

https://compbiozurich.org/UZH-BIO392/
https://github.com/compbiozurich/UZH-BIO392/



Terminal?
UNiX?
Github?
R?
Python?
Perl?
YAML?

Skill assessment

BIO392
Bioinformatics of Genome Variations

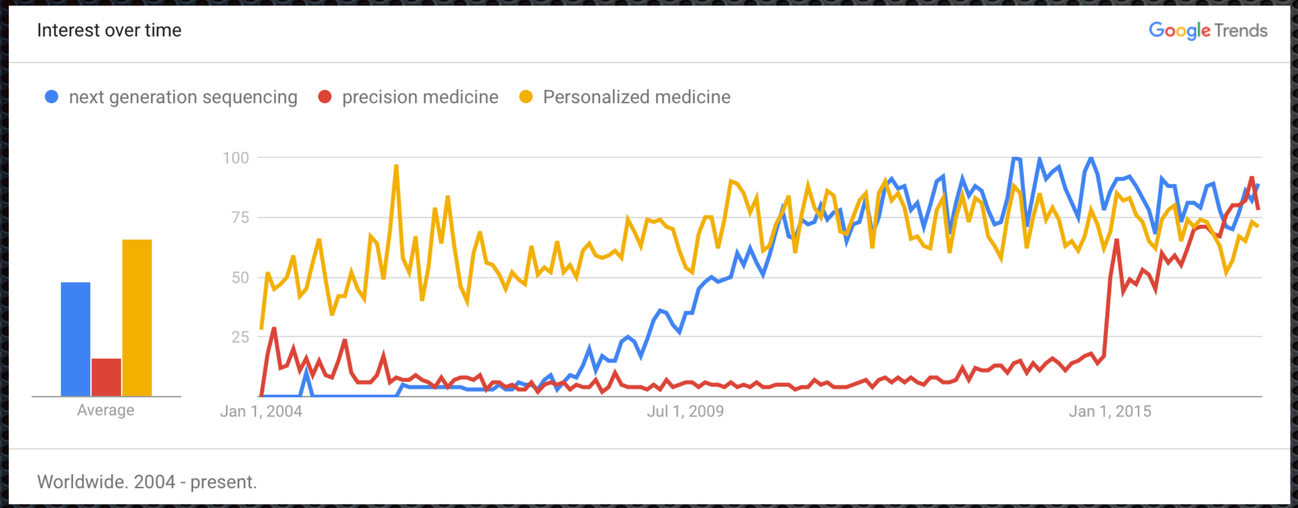
Genomes: Core of "Personalized Health" & "Precision Medicine"

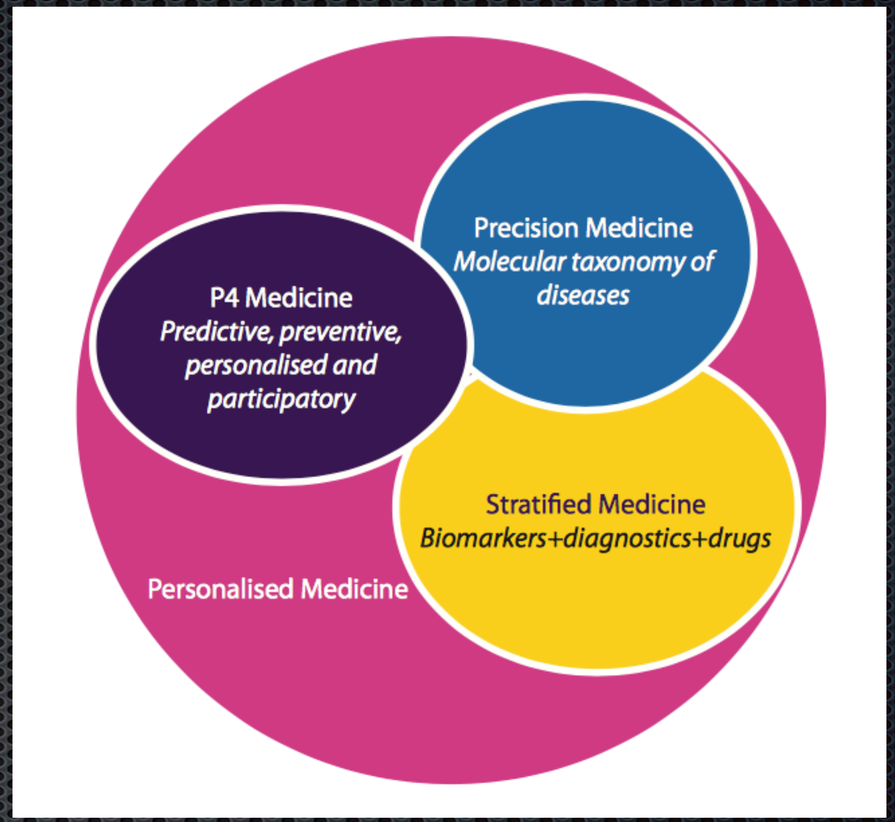
Michael Baudis **UZH SIB**
Computational Oncogenomics

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?



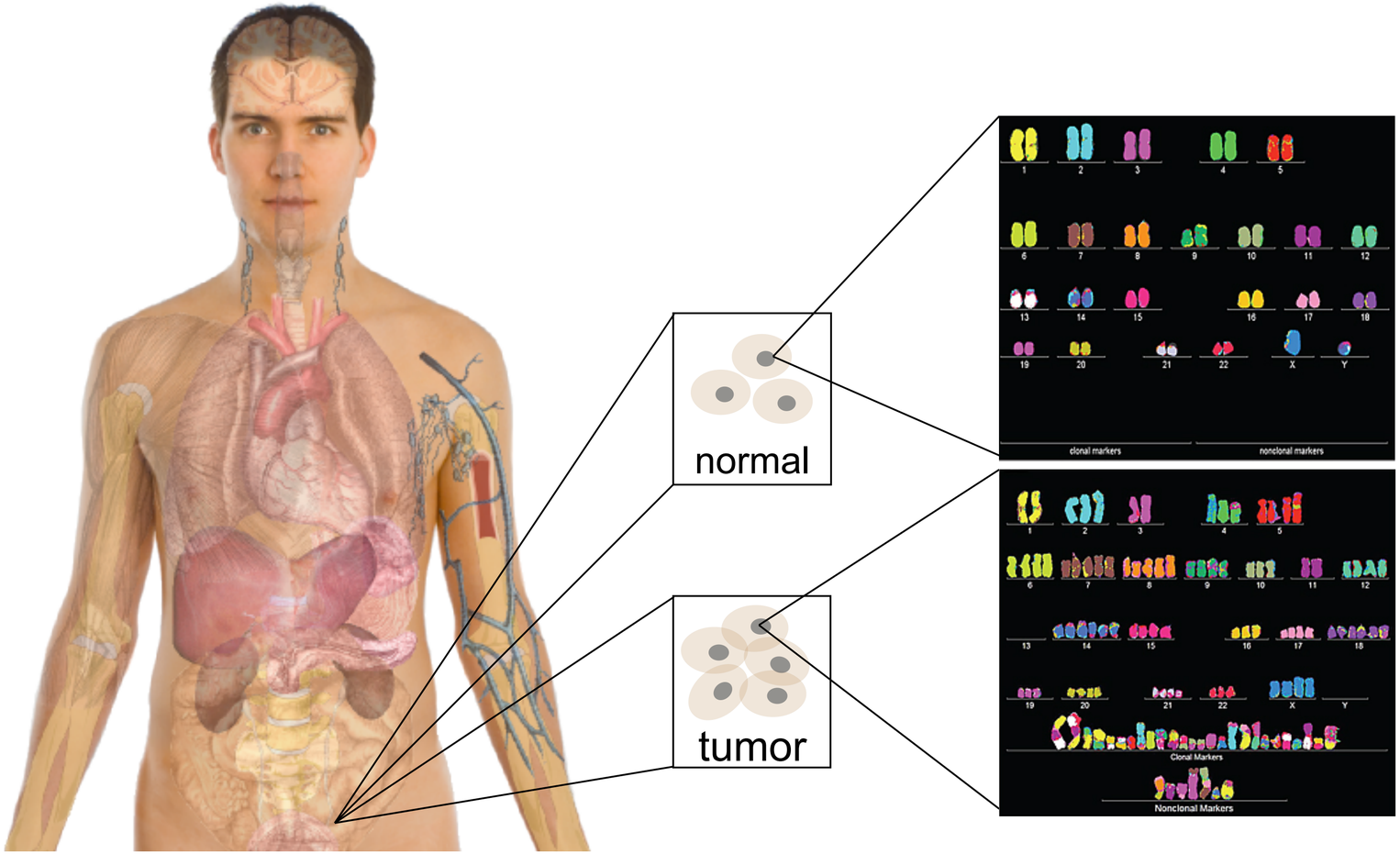


Source: PHG Foundation

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.

Personal Genomics as a Gateway into Biology

Personal genomes soon will become a commonplace part of medical research & eventually treatment (esp. for cancer). They will provide a primary connection for biological science to the general public.



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 resource(s)** & expertise
- participation in reciprocal, international **data sharing** and **biocuration** efforts

Genome analyses at the core of Personalized Health™

Susceptibility, Pharmacogenomics, Classification, Infectious Diseases, Outcome Prediction, Lifestyle ...

doi:10.1038/nature19057

Analysis of protein-coding genetic variation in 60,706 humans

Monkol Lek[1,2,3,4], Konrad J. Karczewski[1,2]*, Eric V. Minikel[1,2,5]*, Kaitlin E. Samocha[1,2,5,6]*, Eric Banks[2], Timothy Fennell[2], Anne H. O'Donnell-Luria[1,2,7], James S. Ware[2,8,9,10,11], Andrew J. Hill[1,2,12], Beryl B. Cummings[1,2,5], Taru Tukiainen[1,2], Daniel P. Birnbaum[2], Jack A. Kosmicki[1,2,6,13], Laramie E. Duncan[1,2,6], Karol Estrada[1,2], Fengmei Zhao[1,2], James Zou[2],

Barkur S. Shastry

SNP alleles in human disease and evolution

DISEASE MECHANISMS

Mechanisms underlying structural variant formation in genomic disorders

Claudia M. B. Carvalho[1,2] and James R. Lupski[1,3,4,5]

Abstract | With the recent burst of technological developments in genomics, and the clinical implementation of genome-wide assays, our understanding of the molecular basis of genomic disorders, specifically the contribution of structural variation to disease burden, is evolving

Rapid whole genome sequencing and precision neonatology

Joshua E. Petrikin, MD[a,], Laurel K. Willig, MD, FAAP[b], Laurie D. Smith, MD, PhD[c], and Stephen F. Kingsmore, MB, BAO, ChB, Dsc, FRCPath[d,e]*

insight progress

Cancer genetics

Bruce A. J. Ponder

Genomic Classification of Cutaneous Melanoma

The Cancer Genome Atlas Network[1,*,*,*]

[1]Cancer Genome Atlas Program Office, National Cancer Institute at NIH, 31 Center Drive, Bldg. 31, Suite 3A20, Bethesda, MD 20892, USA

*Correspondence: irwatson@mdanderson.org (I.R.W.), jgershen@mdanderson.org (J.E.G.), lchin@mdanderson.org (L.C.)

http://dx.doi.org/10.1016/j.cell.2015.05.044

Consequences of genomic diversity in *Mycobacterium tuberculosis*

Mireia Coscolla[a,b], Sebastien Gagneux[a,b,*]

[a] *Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002 Basel, Switzerland*

[b] *University of Basel, Petersplatz 1, Basel 4003, Switzerland*

doi:10.1111/pcn.12128

Psychiatry and Clinical Neurosciences

PCN Frontier Review

Copy-number variation in the pathogenesis of autism spectrum disorder

Emiko Shishido, PhD[1,2,3], Branko Aleksic, MD, PhD[3] and Norio Ozaki, MD, PhD[3]*

the Promotion of Science, Japan

Common gene variants, mortality and extreme longevity in humans

B.T. Heijmans[a,b], R.G.J. Westendorp[b], P.E. Slagboom[a,*]

RESEARCH ARTICLE Open Access

Integrative genome-wide expression profiling identifies three distinct molecular subgroups of renal cell carcinoma with different patient outcome

...nfred Beleut[1,5*], Philip Zimmermann[2], Michael Baudis[3], Nicole Bruni[4], Peter Bühlmann[4], Oliver Laule[2], ...-Duc Luu[1], Wilhelm Gruissem[2], Peter Schraml[1*] and Holger Moch[1]

NEURODEVELOPMENT

Genes, circuits, and precision therapies for autism and related neurodevelopmental disorders

Mustafa Sahin* and Mriganka Sur*

The landscape of somatic copy-number alteration across human cancers

Rameen Beroukhim[1,3,4,5]*, Craig H. Mermel[1,3]*, Dale Porter[8], Guo Wei[1], Soumya Raychaudhuri[1,4], Jerry Donovan[8], Jordi Barretina[1,3], Jesse S. Boehm[1], Jennifer Dobson[1,3], Mitsuyoshi Urashima[9], Kevin T. Mc Henry[8], Reid M. Pinchback[1], Azra H. Ligon[4], Yoon-Jae Cho[6], Leila Haery[1,3], Heidi Greulich[1,3,4,5], Michael Reich[1], Wendy Winckler[1], Michael S. Lawrence[1], Barbara A. Weir[1,3], Kumiko E. Tanaka[1,3], Derek Y. Chiang[1,3,13], Adam J. Bass[1,3,4], Alice Loo[8], Carter Hoffman[1,3], John Prensner[1,3], Ted Liefeld[1], Qing Gao[1], Derek Yecies[3], Sabina Signoretti[3,4], Elizabeth Maher[10], Frederic J. Kaye[11], Hidefumi Sasaki[12], Joel E. Tepper[13], Jonathan A. Fletcher[4], Josep Tabernero[14], José Baselga[14], Ming-Sound Tsao[15], Francesca Demichelis[16], Mark A. Rubin[16], Pasi A. Janne[3,4], Mark J. Daly[1,17], Carmelo Nucera[7], Ross L. Levine[18], Benjamin L. Ebert[1,4,5], Stacey Gabriel[1], Anil K. Rustgi[19], Cristina R. Antonescu[18], Marc Ladanyi[18], Anthony Letai[3], Levi A. Garraway[1,3], Massimo Loda[3,4], David G. Beer[20], Lawrence D. True[21], Aikou Okamoto[22], Scott L. Pomeroy[6], Samuel Singer[18], Todd R. Golub[1,3,23], Eric S. Lander[1,2,5], Gad Getz[1], William R. Sellers[8] & Matthew Meyerson[1,3,5]

RESEARCH ARTICLE Open Access

Chromothripsis-like patterns are recurring but heterogeneously distributed features in a survey of 22,347 cancer genome screens

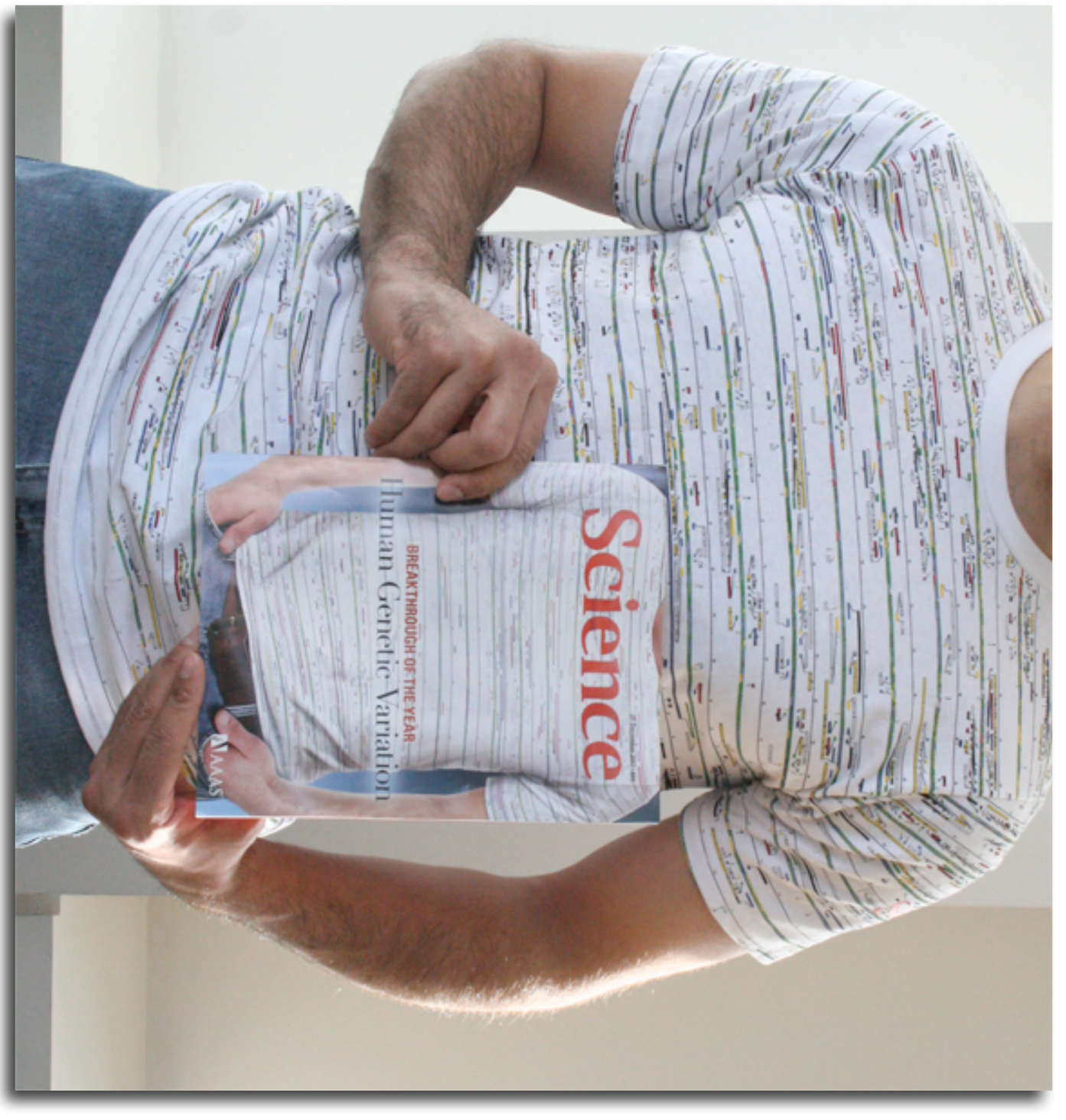
Haoyang Cai[1,2], Nitin Kumar[1,2], Homayoun C Bagheri[3], Christian von Mering[1,2], Mark D Robinson[1,2]* and Michael Baudis[1,2]*

Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib

Thomas J. Lynch, M.D., Daphne W. Bell, Ph.D., Raffaella Sordella, Ph.D., Sarada Gurubhagavatula, M.D., Ross A. Okimoto, B.S., Brian W. Brannigan, B.A., Patricia L. Harris, M.S., Sara M. Haserlat, B.A., Jeffrey G. Supko, Ph.D., Frank G. Haluska, M.D., Ph.D., David N. Louis, M.D., David C. Christiani, M.D., Jeff Settleman, Ph.D., and Daniel A. Haber, M.D., Ph.D.

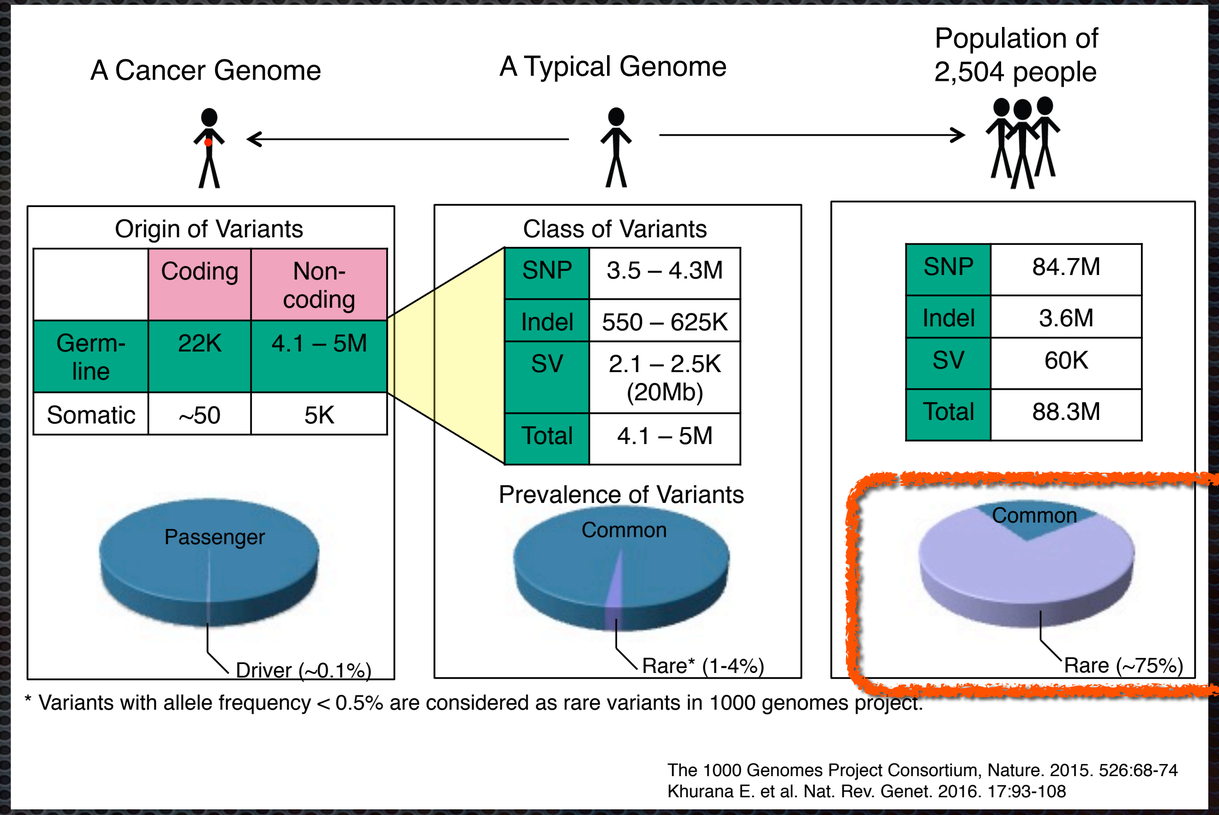
N Engl J Med 2004; 350:2129-2139 | May 20, 2004 | DOI: 10.1056/NEJMoa040938

The trouble with human genome variation



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

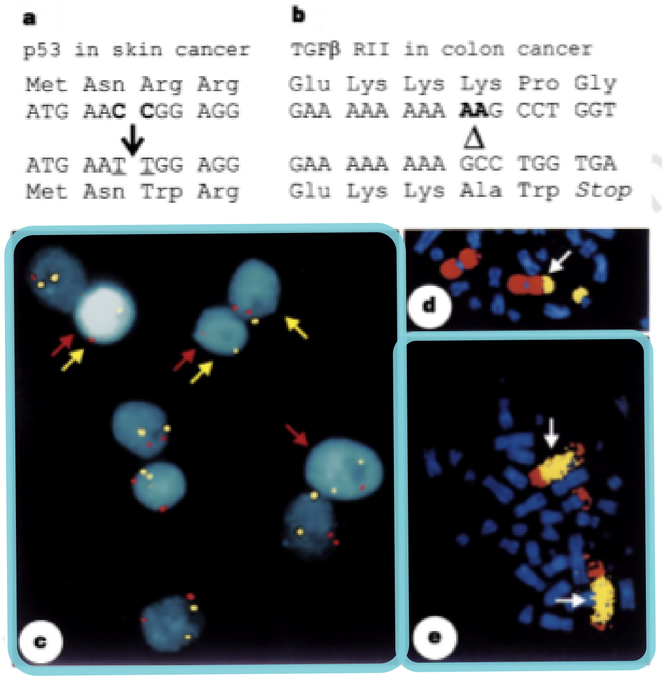
- a typical human genome (~3 billion base pairs) has ~5 million variants
- most of them are "**rare**"; i.e. can only be identified as recurring when sequencing thousands of people
- cancer cells accumulate additional variants, only **few** of which ("**drivers**") are relevant for the disease



The 1000 Genomes Project Consortium, Nature. 2015. 526:68-74
Khurana E. et al. Nat. Rev. Genet. 2016. 17:93-108

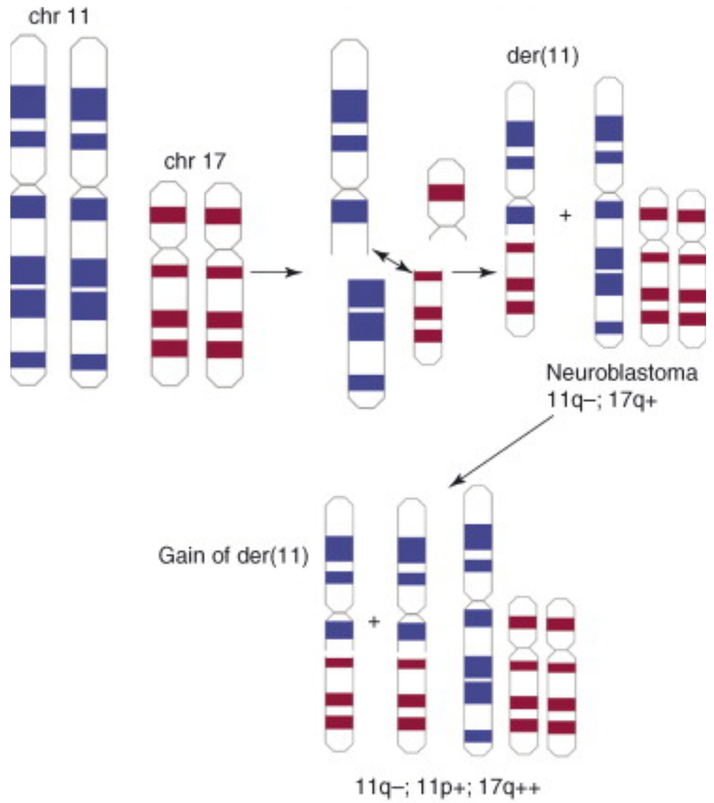
Graphic adapted from Mark Gerstein (GersteinLab.org; @markgerstein)

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)

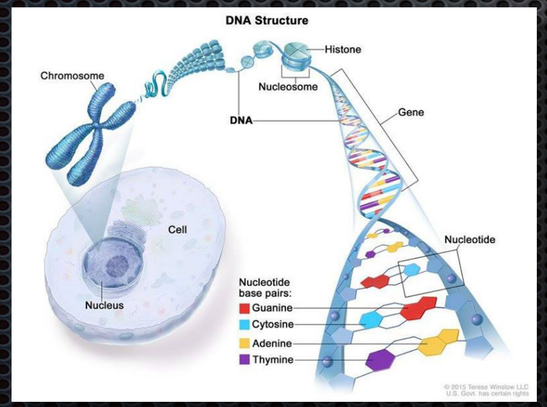


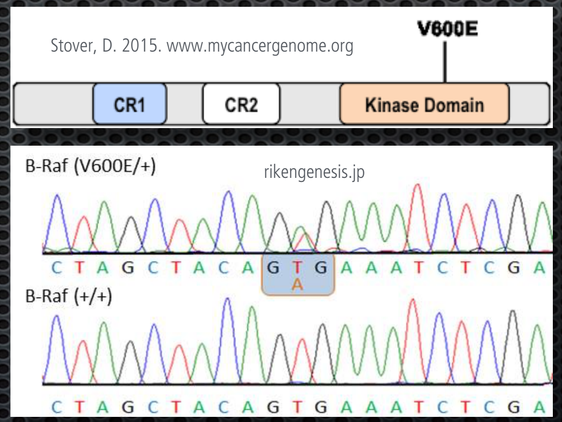
Generation of copy number imbalances in cancer through imbalanced cytogenetic rearrangements - partial deletion of 11q, gain of 11pterq21 and 2 addl. copies of 17q

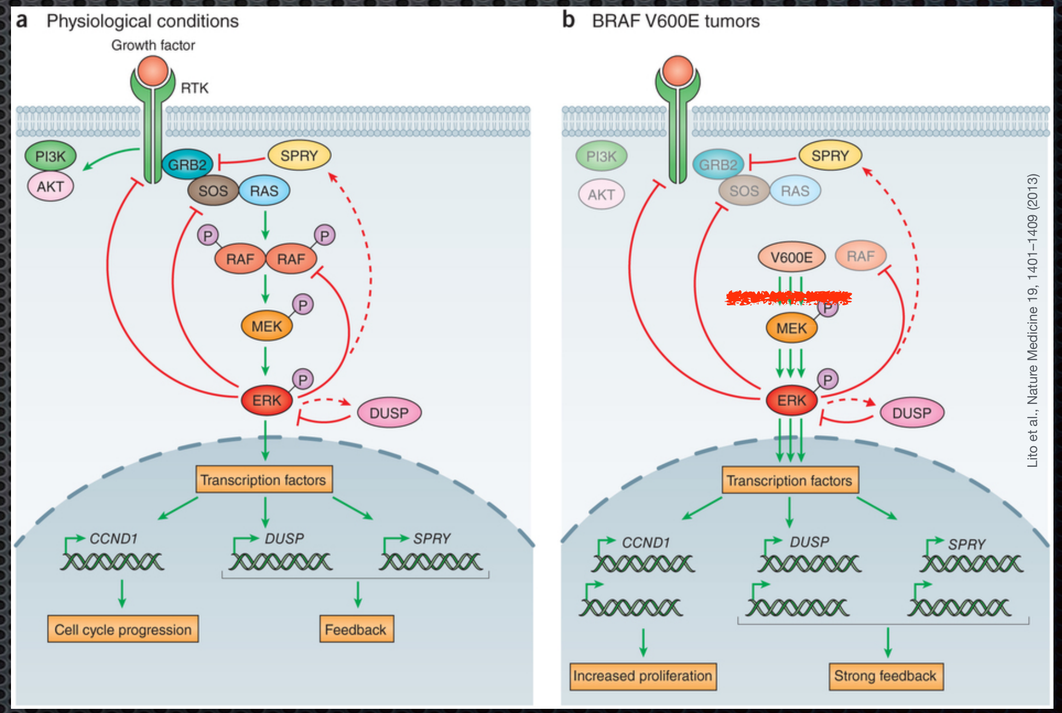
RL Stallings: Are chromosomal imbalances important in cancer? Volume 23, Issue 6, p278–283, 2007

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

- a single nucleotide exchange **T**hymidine > **A**denine 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***)

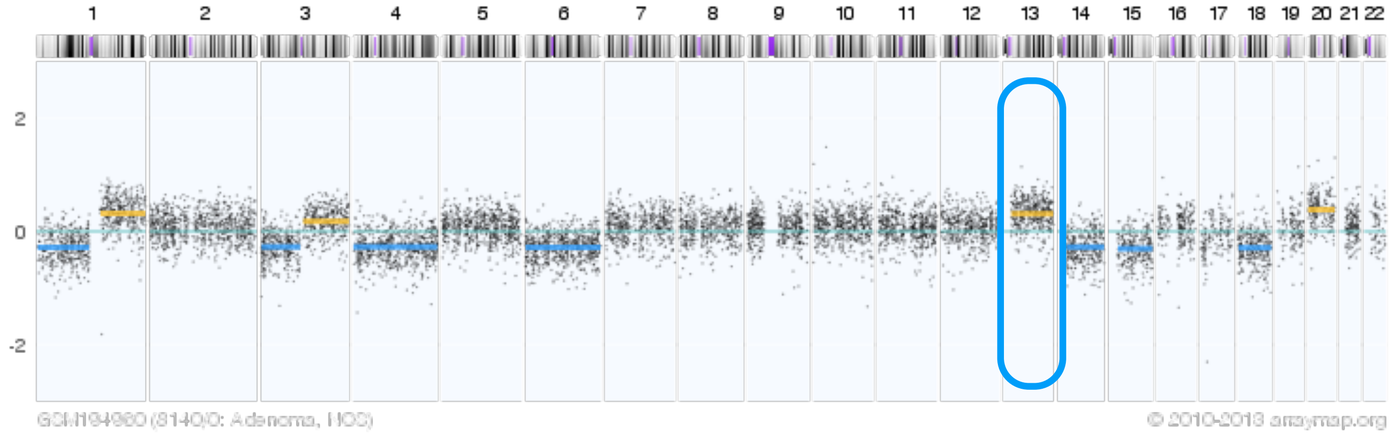




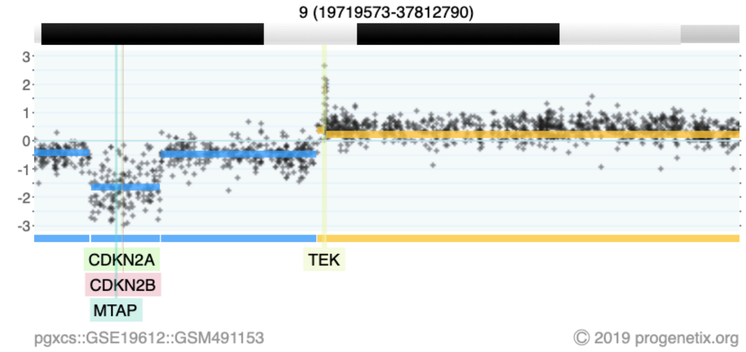


The BRAF V600E mutation leads to continuos phosphorylation of MEK, without the need for receptor based activation of the upstream pathway and loss of inhibitory feedback control.

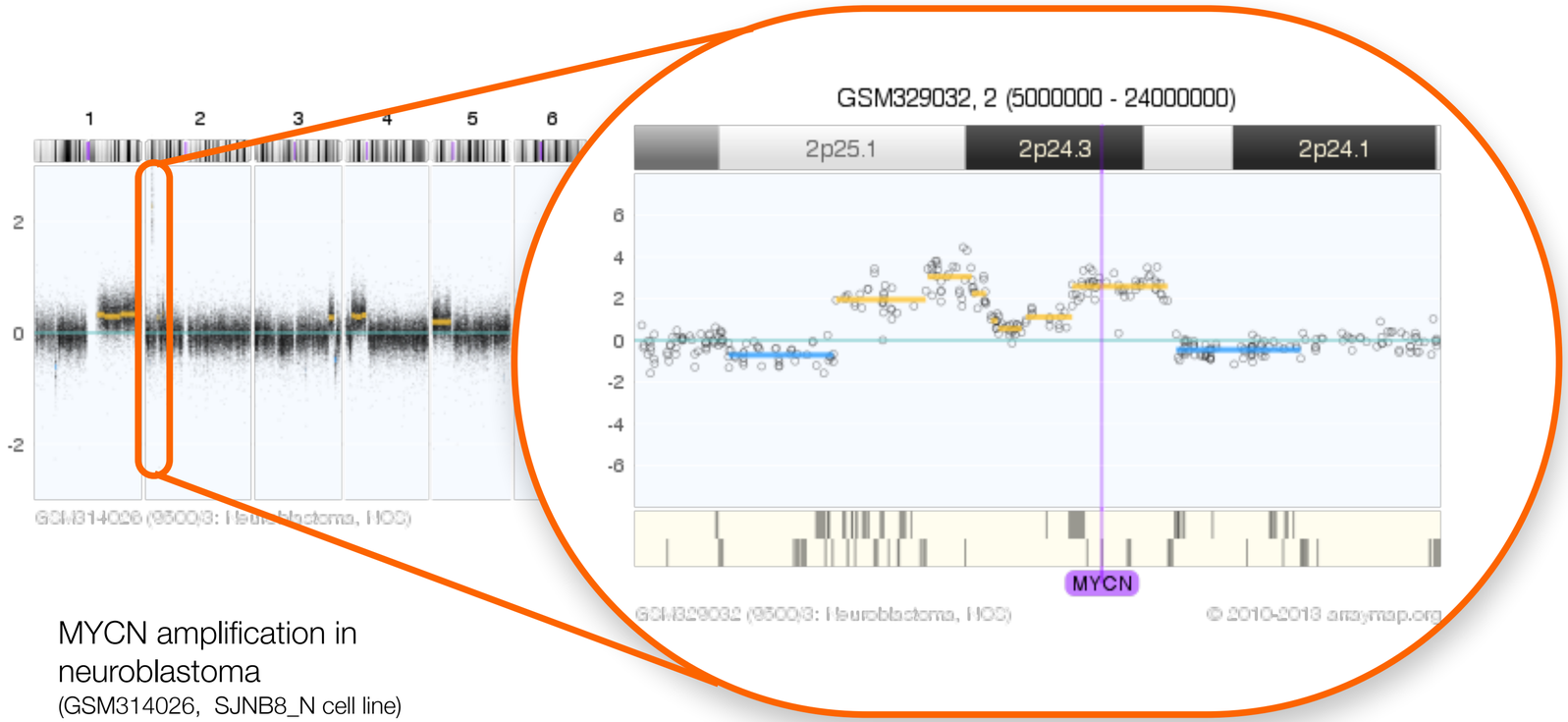
Somatic Copy Number Variations



Gain of chromosome arm 13q in colorectal carcinoma



2-event, homozygous deletion in a Glioblastoma

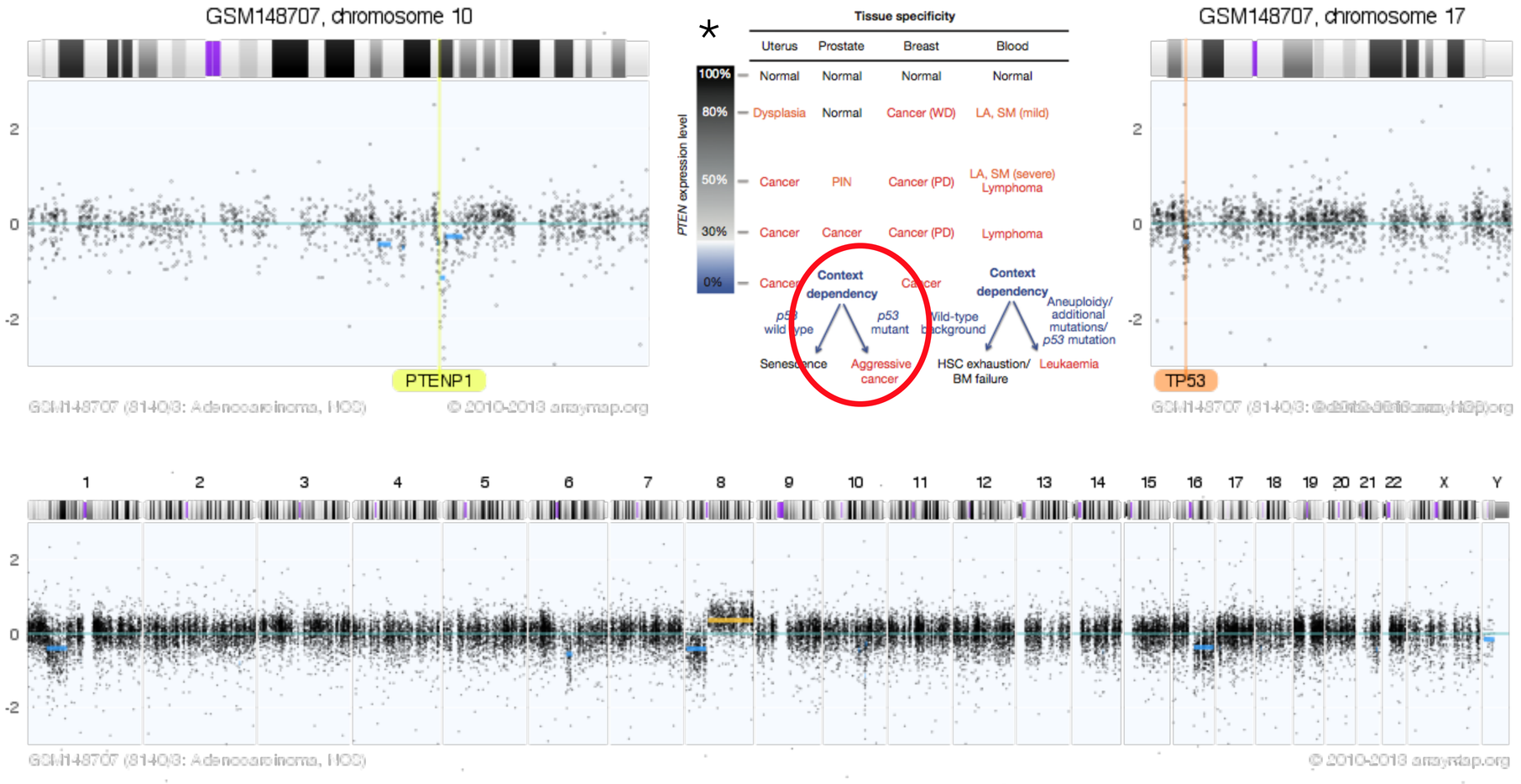


MYCN amplification in neuroblastoma
(GSM314026, SJNB8_N cell line)

low level/**high level** copy number alterations (CNAs)

arrayMap

Gene dosage phenomena beyond simple on/off effects

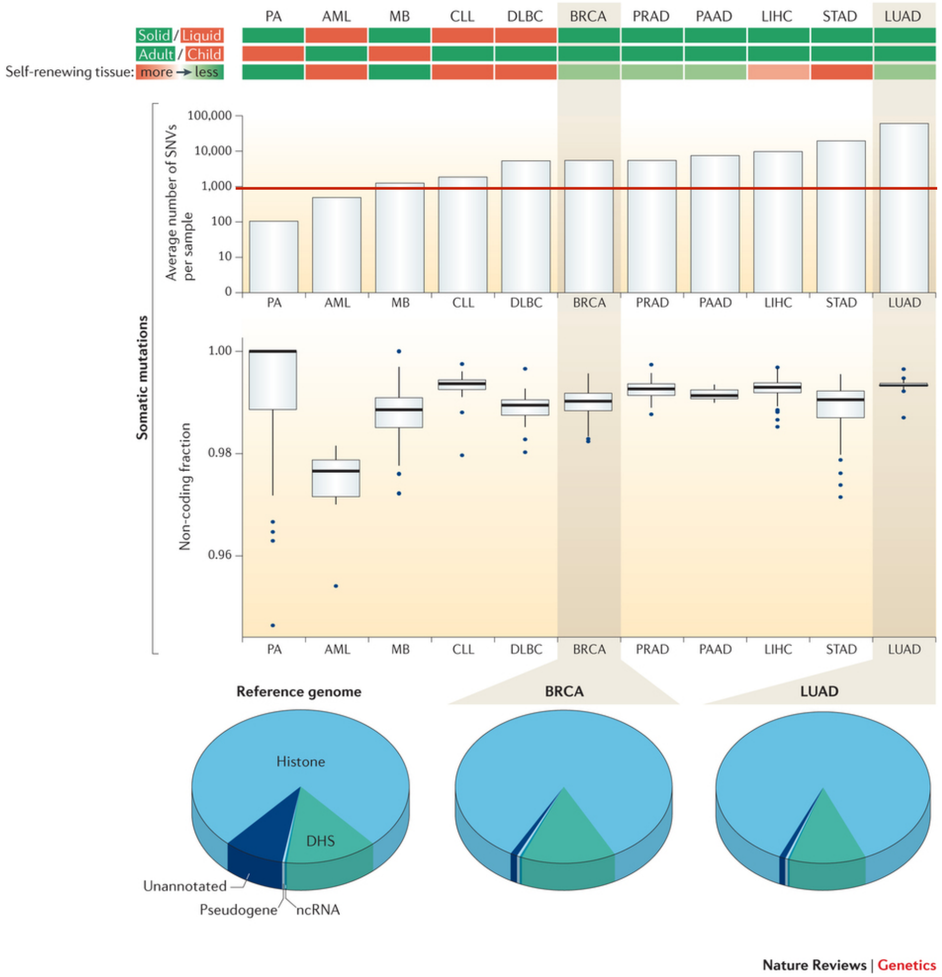


Combined heterozygous deletions involving *PTEN* and TP53 loci in a case of prostate adenocarcinoma
(GSM148707, PMID 17875689, Lapointe *et al*., CancRes 2007)

* A. H. Berger, A. G. Knudson, and P. P. Pandolfi, “A continuum model for tumour suppression,” *Nature*, vol. 476, no. 7359, pp. 163–169, Aug. 2011.

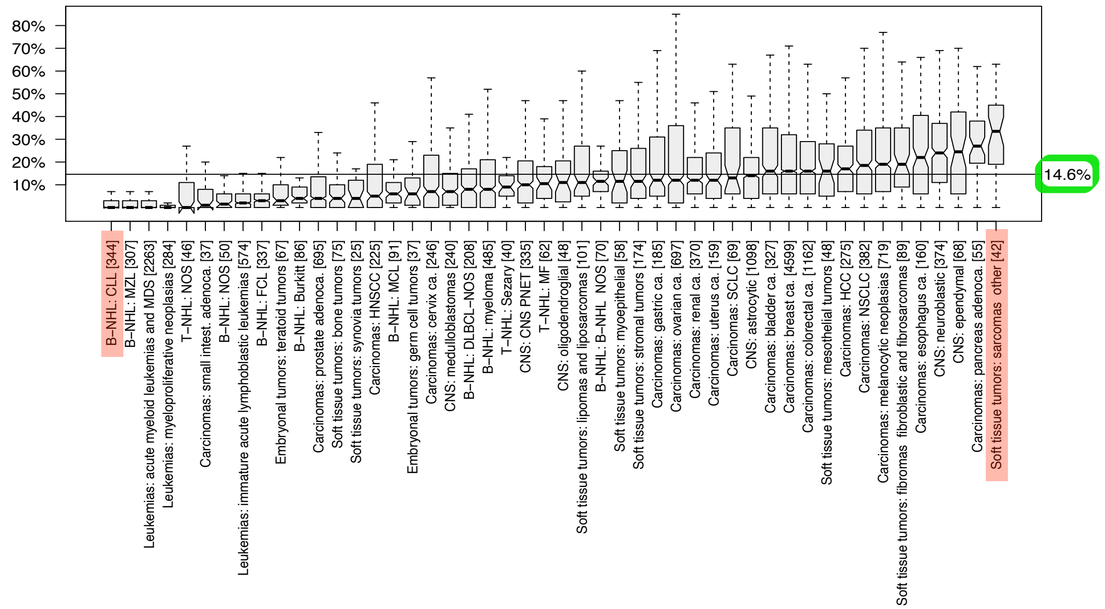
arrayMap

Quantifying Somatic Mutations In Cancer



CANCERS SHOW THOUSANDS OF SINGLE NUCLEOTIDE VARIANTS PER SAMPLE, MOSTLY IN NON-CODING REGIONS

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)

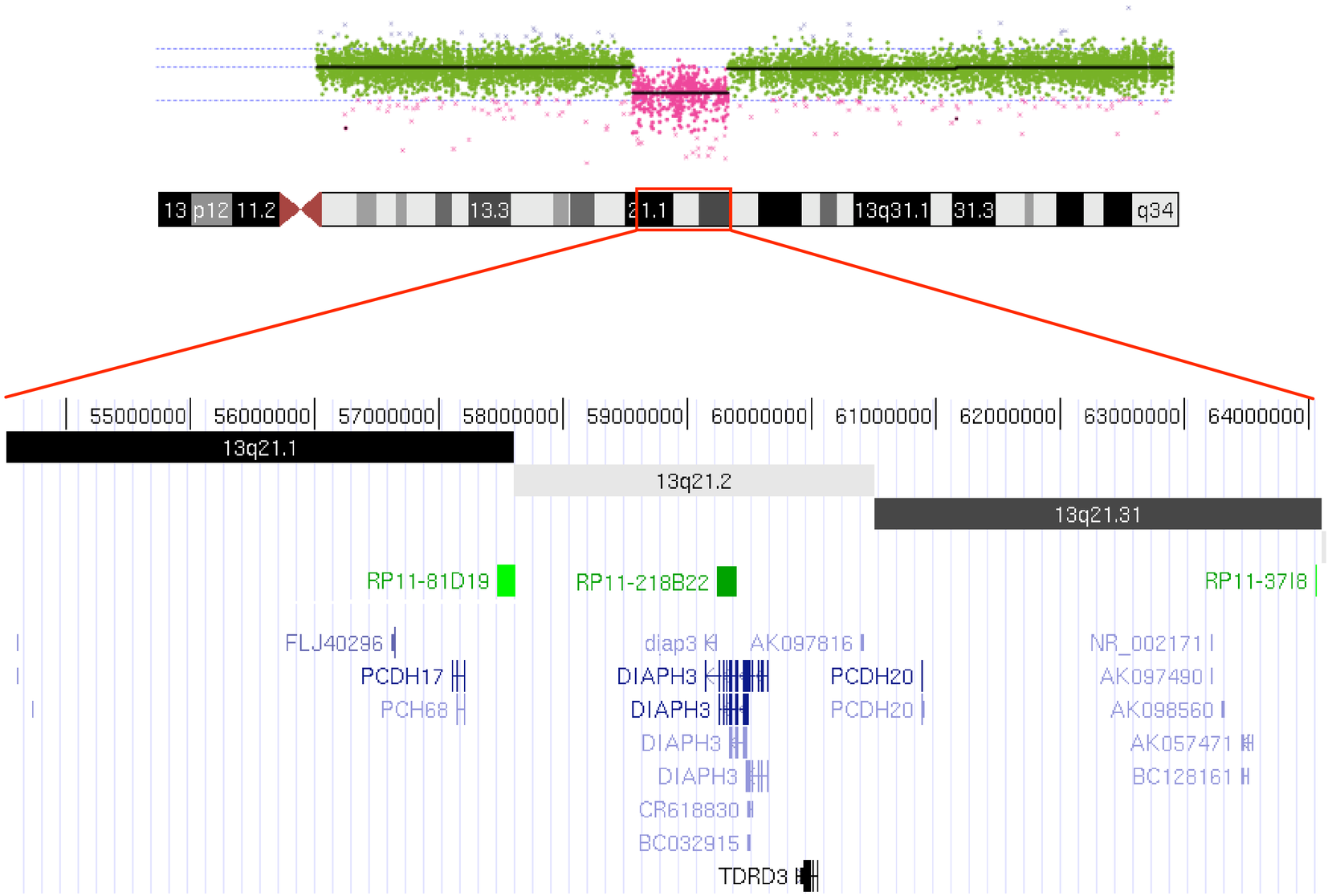


GENOMIC COPY NUMBER IMBALANCES PROVIDE WIDESPREAD SOMATIC VARIANTS IN CANCER

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

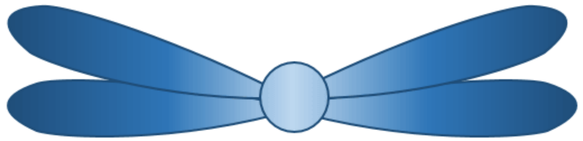
Nobody is perfect (?)

A 10.7 Mb Interstitial Deletion of 13q21 Without Phenotypic Effect Defines a Further Non-Pathogenic Euchromatic Variant
Andreas Roos, Miriam Elbracht, Michael Baudis, Jan Senderek, Nadine Schönherr, Thomas Eggermann, and Herdit M. Schüler
American Journal of Medical Genetics Part A 146A:2417 – 2420 (2008)

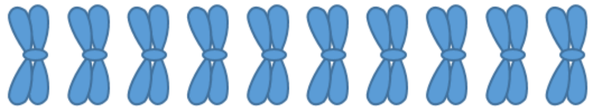


Genome Sequencing

whole genome sequencing (WGS)

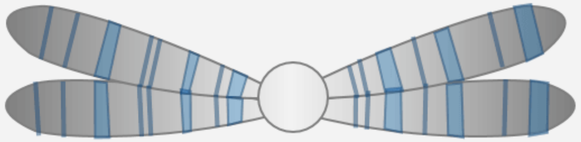


100%



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	~$1,500
time:	12+ years	< 2 days

today:
2017

~$530
~3 days

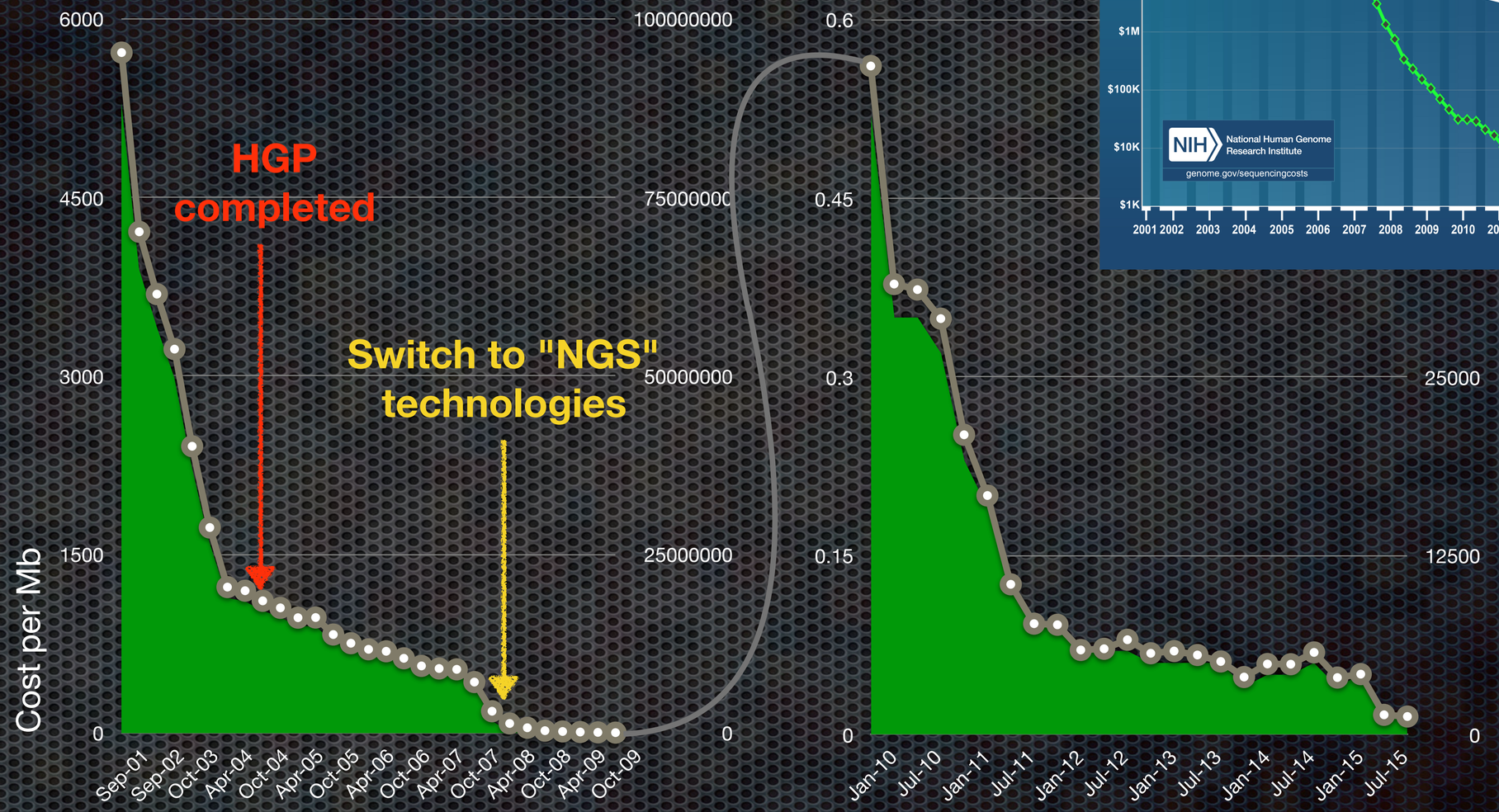
Ryan L. Collins & Brad Wierbowski (2017)

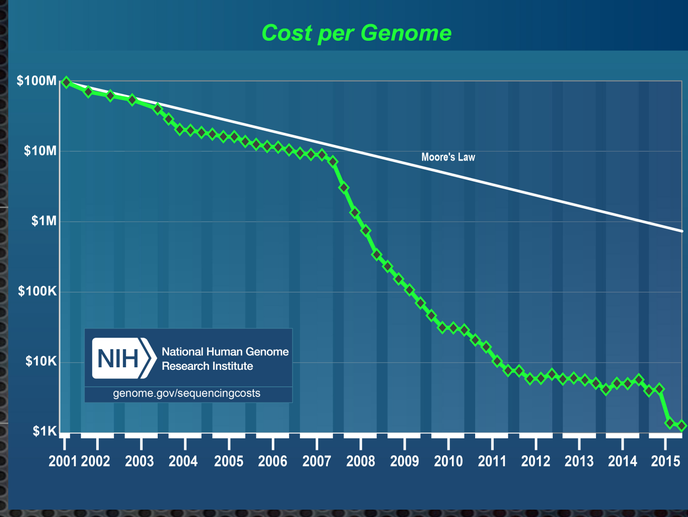
DNA Sequencing Costs: Data

Data from the NHGRI Genome Sequencing Program (GSP)

https://www.genome.gov/sequencingcostsdata/

Date	Cost per Mb	Cost per Genome
Sep-01	JS$ 5'292.39	US$ 95'263'072
Mar-02	JS$ 3'898.64	US$ 70'175'437
Sep-02	JS$ 3'413.80	US$ 61'448'422
Mar-03	JS$ 2'986.20	US$ 53'751'684
Oct-03	JS$ 2'230.98	US$ 40'157'554
Jan-04	JS$ 1'598.91	US$ 28'780'376
Apr-04	JS$ 1'135.70	US$ 20'442'576
Jul-04	JS$ 1'107.46	US$ 19'934'346
Oct-04	JS$ 1'028.85	US$ 18'519'312
Jan-05	US$ 974.16	US$ 17'534'970
Apr-05	US$ 897.76	US$ 16'159'699
Jul-05	US$ 898.90	US$ 16'180'224
Oct-05	US$ 766.73	US$ 13'801'124
Jan-06	US$ 699.20	US$ 12'585'659
Apr-06	US$ 651.81	US$ 11'732'535
Jul-06	US$ 636.41	US$ 11'455'315
Oct-06	US$ 581.92	US$ 10'474'556
Jan-07	US$ 522.71	US$ 9'408'739
Apr-07	US$ 502.61	US$ 9'047'003
Jul-07	US$ 495.96	US$ 8'927'342
Oct-07	US$ 397.09	US$ 7'147'571
Jan-08	US$ 102.13	US$ 3'063'820
Apr-08	US$ 15.03	US$ 1'352'982
Jul-08	US$ 8.36	US$ 752'080
Oct-08	US$ 3.81	US$ 342'502
Jan-09	US$ 2.59	US$ 232'735
Apr-09	US$ 1.72	US$ 154'714
Jul-09	US$ 1.20	US$ 108'065
Oct-09	US$ 0.78	US$ 70'333
Jan-10	US$ 0.52	US$ 46'774
Apr-10	US$ 0.35	US$ 31'512
Jul-10	US$ 0.35	US$ 31'125
Oct-10	US$ 0.32	US$ 29'092
Jan-11	US$ 0.23	US$ 20'963
Apr-11	US$ 0.19	US$ 16'712
Jul-11	US$ 0.12	US$ 10'497
Oct-11	US$ 0.09	US$ 7'743
Jan-12	US$ 0.09	US$ 7'666
Apr-12	US$ 0.07	US$ 5'901
Jul-12	US$ 0.07	US$ 5'985
Oct-12	US$ 0.07	US$ 6'618
Jan-13	US$ 0.06	US$ 5'671
Apr-13	US$ 0.06	US$ 5'826
Jul-13	US$ 0.06	US$ 5'550
Oct-13	US$ 0.06	US$ 5'096
Jan-14	US$ 0.04	US$ 4'008
Apr-14	US$ 0.05	US$ 4'920
Jul-14	US$ 0.05	US$ 4'905
Oct-14	US$ 0.06	US$ 5'731
Jan-15	US$ 0.04	US$ 3'970
Apr-15	US$ 0.05	US$ 4'211
Jul-15	US$ 0.015	US$ 1'363
Oct-15	US$ 0.014	US$ 1'245

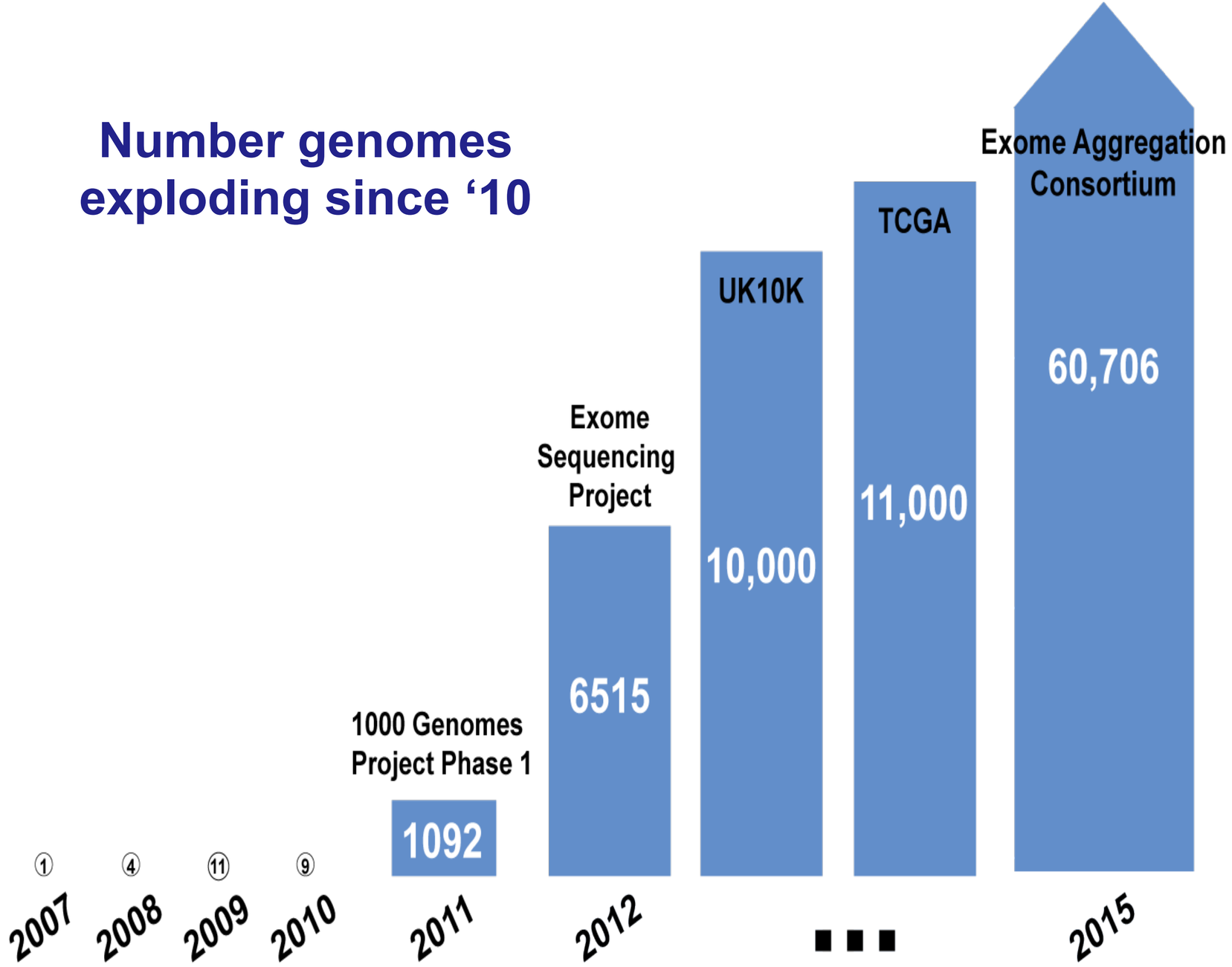




- Labor, administration, management, utilities, reagents, and consumables
- Sequencing instruments and other large equipment (amortized over three years)
- Informatics activities directly related to sequence production (e.g., laboratory information management systems and initial data processing)
- Submission of data to a public database
- Indirect Costs (http://oamp.od.nih.gov/dfas/faq/indirect-costs#difference) as they relate to the above items

Slides freely downloadable from lectures.GersteinLab.org & “tweetable” (via @markgerstein)

Number genomes exploding since ‘10



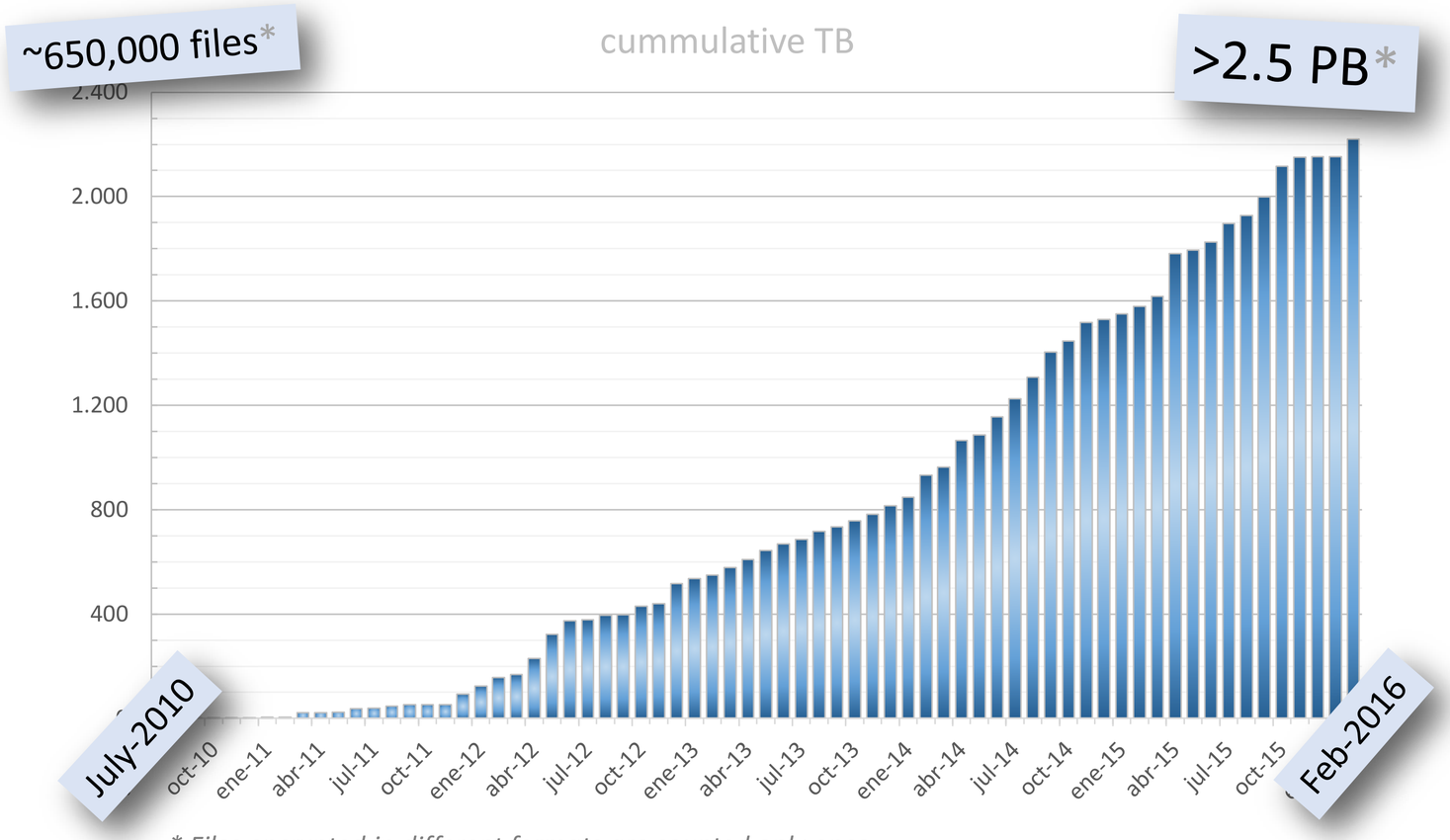
Genomes Everywhere

Large Genome Data Generation, Analysis & Sharing Initiatives

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

Growth of Genome Data Repositories: Example EGA

The EGA contains a growing amount of data



International Cancer Genome Consortium

IRDiRC INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM

UK 10K

wellcome trust sanger institute

** Files encrypted in different formats are counted only once*

Source: Jordi Rambla (CRG Barcelona) 2016

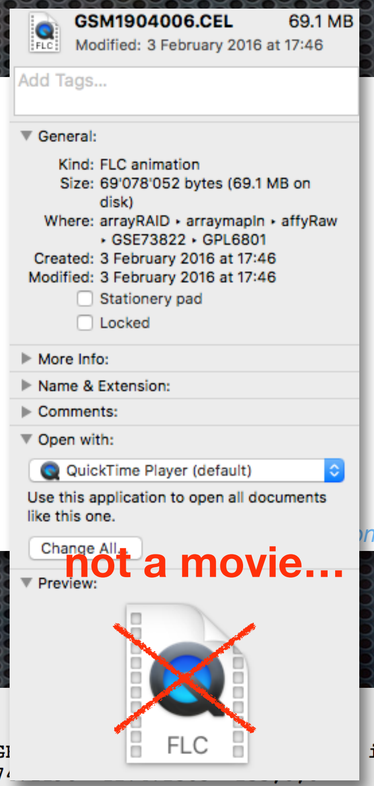
What is a PB, for human genomes? It depends.

- 2 bits per base are sufficient to encode TCGA
  - using 00, 01, 10, 11
  - [TCGA]{3’000’000’000}
  - 2 * 3 * 10^9b = 6, 000, 000, 000 b
  - perfect genome (no overhead): ~715 MB
  - 1PB => ~1’400’000 genomes
- according to Swiss online store (Sep 2019) ~35'000CHF (65x16TB disks)
- this is less than a PhD position per year in Switzerland …
- (real costs are 2x that, + duplication, facilities, service … => ~500’000CHF)
- **However: A single 30x BAM file => 100GB**
- Still: 500’000CHF => 1PB => 10’000 genomes => 50CHF/genome (BAM format)

Bioinformatics: **File Formats**

- **text** or **binary** file formats, optimised for specific types of biological data
- examples from genomics:
  - **BAM** - compressed binary version of Sequence Alignment/ Map (SAM)
  - **BED** (Browser Extensible Data) -flexible way to define the data lines in an genome browser annotation tracks
  - **VCF** (Variant Call Format)

- Axt format
- BAM format
- BED format
- BED detail format
- bedGraph format
- barChart and bigBarChart format
- bigBed format
- bigGenePred table format
- bigPsl table format
- bigMaf table format
- bigChain table format
- bigWig format
- Chain format



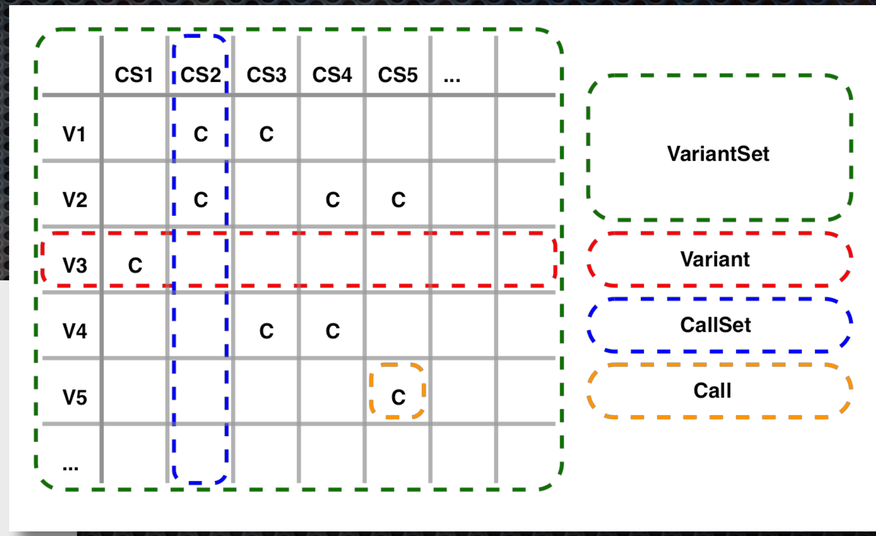
- CRAM format
- GenePred table format
- GFF format
- GTF format
- HAL format
- MAF format
- Microarray format
- Net format
- Personal Genome SNP format
- PSL format
- VCF format
- WIG format

me.ucsc.edu/FAQ/FAQformat.html

```
browser position chr7:127471196-127495720
browser hide all
track name="ItemRGBDemo" description="Item RGB                                    itemRgb="On"
chr7    127471196  127472363  Pos1  0  +  127                                    ,,
chr7    127472363  127473530  Pos2  0  +  127472363  127473530  255,0,0
chr7    127473530  127474697  Pos3  0  +  127473530  127474697  255,0,0
chr7    127474697  127475864  Pos4  0  +  127474697  127475864  255,0,0
chr7    127475864  127477031  Neg1  0  -  127475864  127477031  0,0,255
chr7    127477031  127478198  Neg2  0  -  127477031  127478198  0,0,255
chr7    127478198  127479365  Neg3  0  -  127478198  127479365  0,0,255
chr7    127479365  127480532  Pos5  0  +  127479365  127480532  255,0,0
chr7    127480532  127481699  Neg4  0  -  127480532  127481699  0,0,255
```

BED file example

The VCF file format
Standard for variant representation



Example

```
##fileformat=VCFv4.0
##fileDate=20100707
##source=VCFtools
##reference=NCBI36
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality (phred score)">
##FORMAT=<ID=GL,Number=3,Type=Float,Description="Likelihoods for RR,RA,AA genotypes (R=ref,A=alt)">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##ALT=<ID=DEL,Description="Deletion">
##INFO=<ID=SVTYPE,Number=1,Type=String,Description="Type of structural variant">
##INFO=<ID=END,Number=1,Type=Integer,Description="End position of the variant">
#CHROM POS ID      REF  ALT    QUAL FILTER  INFO                FORMAT     SAMPLE1   SAMPLE2
1       1  .       ACG  A,AT    .    PASS    .                   GT:DP      1/2:13    0/0:29
1       2  rs1     C    T,CT    .    PASS    H2;AA=T             GT:GQ      0|1:100   2/2:70
1       5  .       A    G       .    PASS    .                   GT:GQ      1|0:77    1/1:95
1      100 .       T    <DEL>   .    PASS    SVTYPE=DEL;END=300  GT:GQ:DP   1/1:12:3  0/0:20
```

VCF header

Body

Mandatory header lines

Optional header lines (meta-data about the annotations in the VCF body)

Reference alleles (GT=0)

Alternate alleles (GT>0 is an index to the ALT column)

Phased data (G and C above are on the same chromosome)

Deletion

SNP

Large SV

Insertion

Other event

Variant
Call
Format

- stores the results of a single or multiple interpretations of genome sequencing datasets, in comparison to a reference genome
- standard format for file-based storage of human genome variants

Global Alliance for Genomics & Health

Task: Estimate Storage Requirements for 1000 Genomes

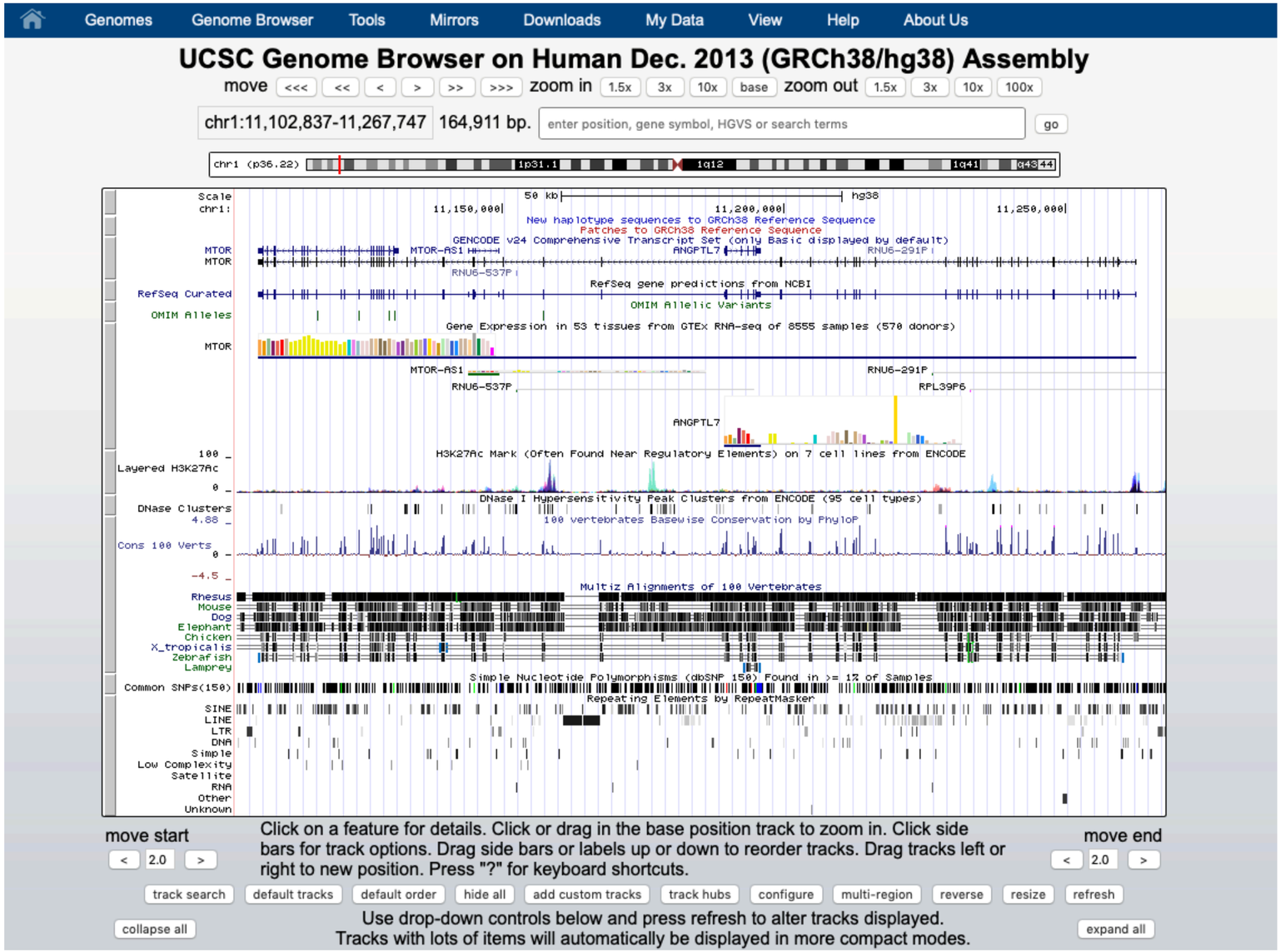
- How much computer storage is required for 1000 Genomes
  - WES & WGS
  - Different file formats
    - SAM
    - BAM
    - VCF
    - FASTA
  - Associated costs
    - Cost afctors
    - Raw Storage costs

Reference Genome Resources...

RESOURCES FOR GENOMICS: UCSC GENOME BROWSER

- Originated from the Human Genome Project
- Most widely used general genome browser
- many default tracks
- many species
- customization with "BED" files

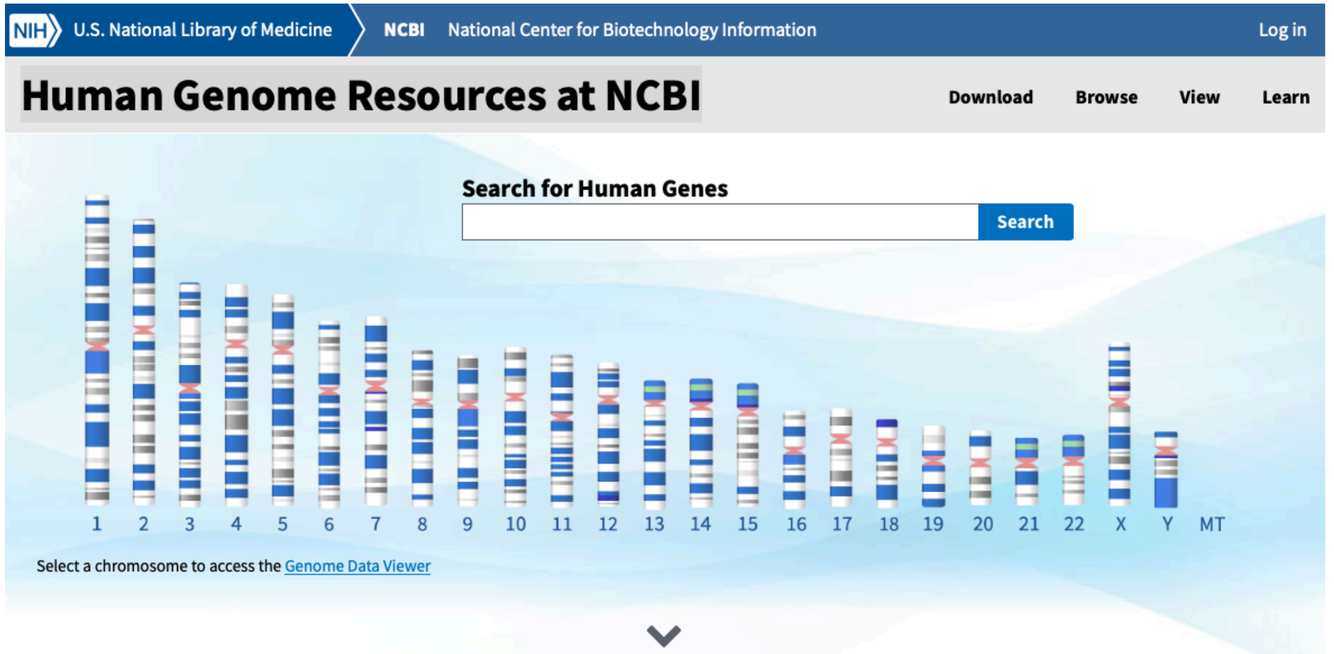
genome.ucsc.edu



RESOURCES FOR GENOMICS: HUMAN GENOME RESOURCES AT NCBI

- Entry point for genome reference data
- Human genome assemblies
- Human variant collections (dbVar, ClinVar, dbSNP) for download

www.ncbi.nlm.nih.gov/projects/genome/guide/human/

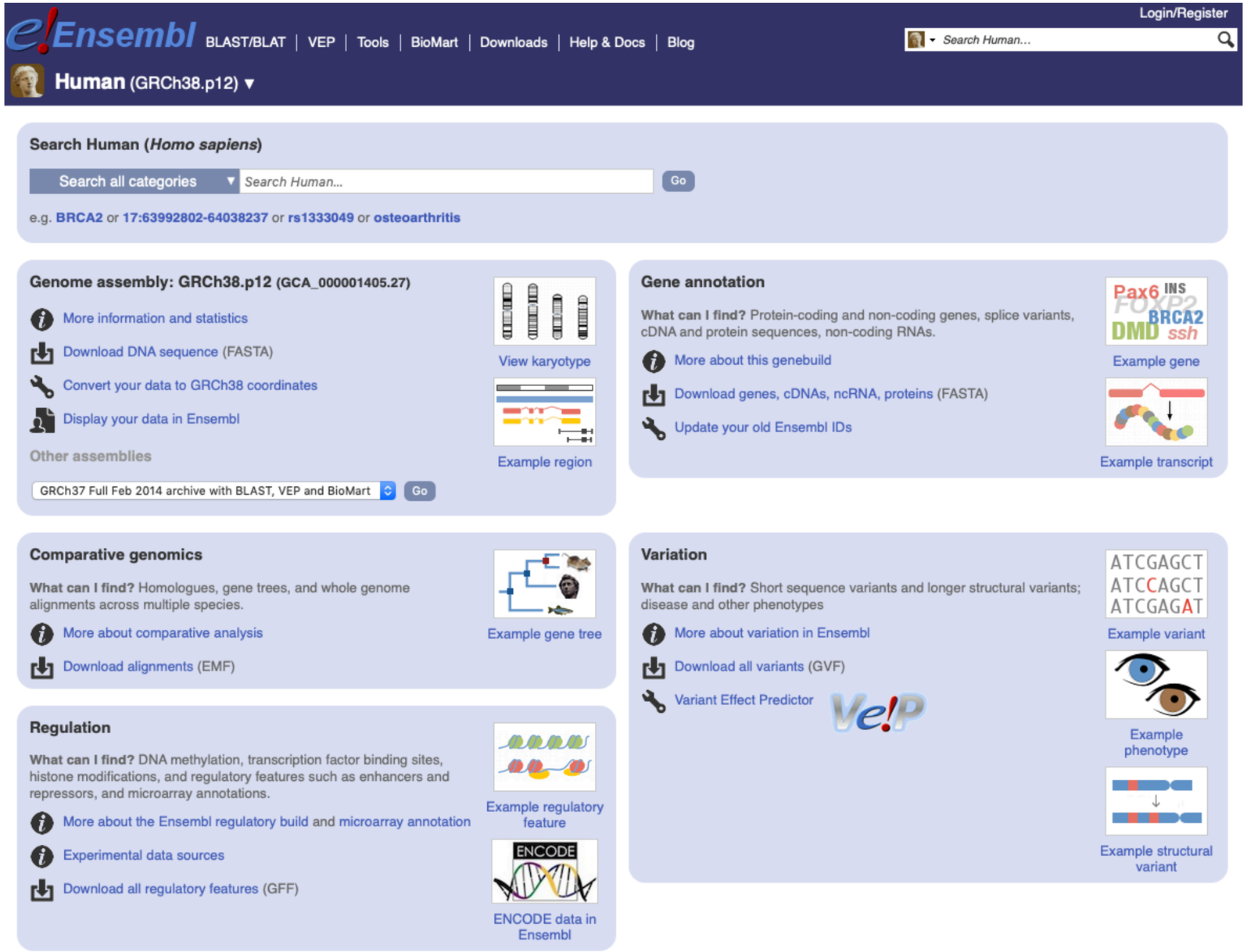


Download	GRCh38	GRCh37
Reference Genome Sequence	Fasta	Fasta
RefSeq Reference Genome Annotation	gff3	gff3
RefSeq Transcripts	Fasta	Fasta
RefSeq Proteins	Fasta	Fasta
ClinVar	vcf	vcf
dbSNP	vcf	vcf
dbVar	vcf	vcf

RESOURCES FOR GENOMICS: ENSEMBL

- Entry point for many genome data services and collections
- Downloads ("BioMart"), REST API

www.ensembl.org/Homo_sapiens/Info/Index



Where to find genome ***variant*** data ...

Reference Resources for Human Genome Variants

- NCBI:dbSNP
  - single nucleotide polymorphisms (SNPs) and multiple small-scale variations
  - including insertions/deletions, microsatellites, non-polymorphic variants
- NCBI:dbVAR
  - genomic structural variation
  - insertions, deletions, duplications, inversions, multinucleotide substitutions, mobile element insertions, translocations, complex chromosomal rearrangements
- NCBI:ClinVar
  - aggregates information about genomic variation and its relationship to human health
- EMBL-EBI:EVA
  - open-access database of all types of genetic variation data from all species
- Ensembl
  - portal for many things genomic...

RESOURCES FOR CANCER GENOMICS

COSMIC
Catalogue of somatic mutations in cancer

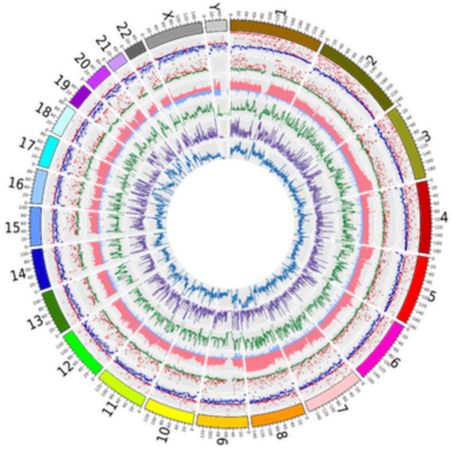
Home ▼ | Resources ▼ | Curation ▼ | Tools ▼ | Data ▼ | News ▼ | Help ▼ | About ▼ | Search COSMIC... | Login ▼

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



Browse the **genomic landscape** of cancer

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 New
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

Cancer Gene Census Update

7 genes have been added to the Cancer Gene Census -

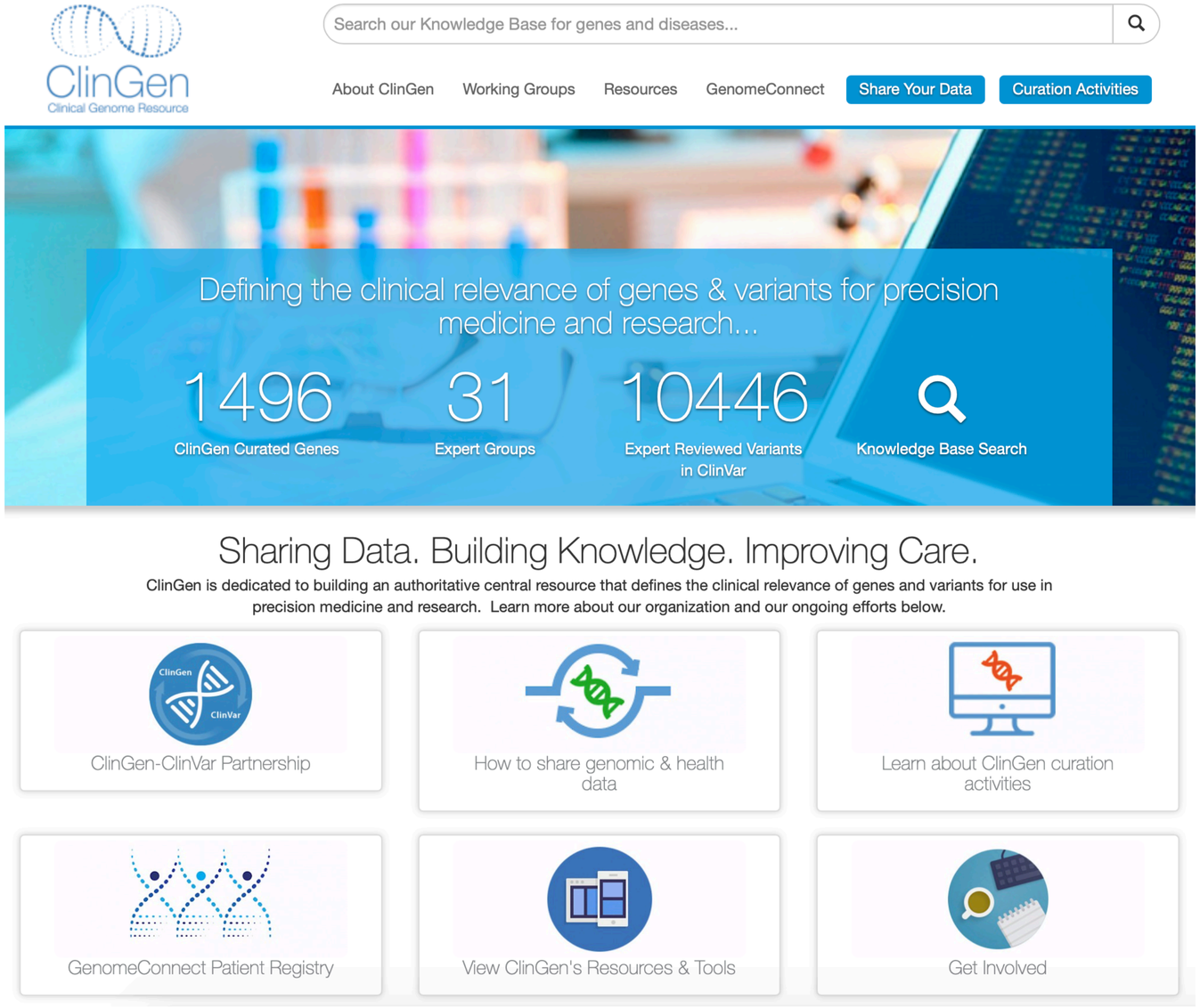
EPAS1 - Endothelial PAS domain protein 1.
PTPRT - Protein tyrosine phosphatase, receptor type T.
PPM1D - Protein phosphatase, Mg2+/Mn2+ dependent 1D.
BTK - Bruton tyrosine kinase.
PREX2 - Phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 2.
TP63 - Tumour protein p63.
QKI - QKI, KH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

For full details, see the Datasheet.

RESOURCES FOR GENOMICS: CLINGEN

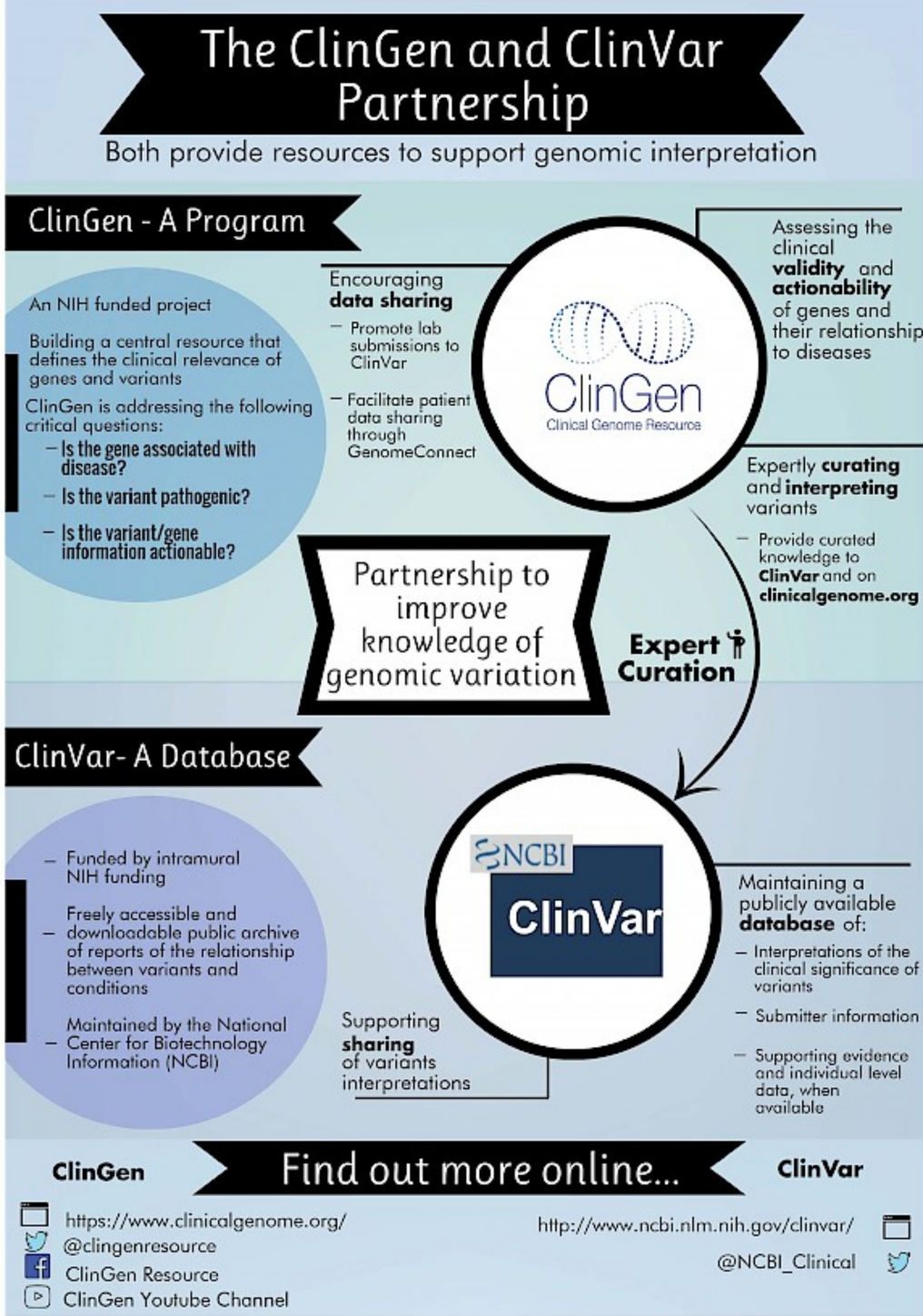
- "The Genomic Variant WG brings together representatives from the Sequence and Structural Variant communities for focused discussions on resolving discrepancies in variant interpretation and creating consistent curation guidelines."
- Interpreted genome variants with disease association

clinicalgenome.org



- ClinVar (an NCBI database/ resource) is used as basis for curated variant <-> disease associations in ClinGen
- ClinGen - a funded project (application/funding limited)
- ClinVar - an internal NIH resource (dependent on political "goodwill")

clinicalgenome.org



RESOURCES FOR CANCER GENOMICS

COSMIC
Catalogue of somatic mutations in cancer

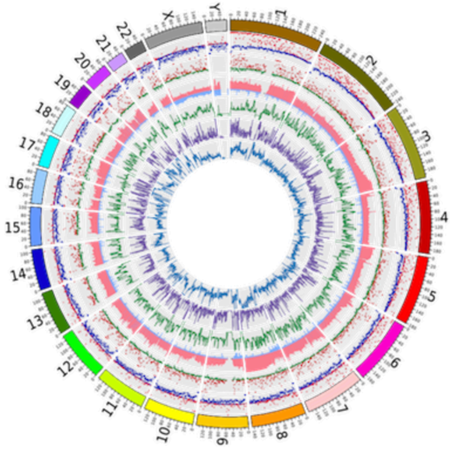
Home ▼ | Resources ▼ | Curation ▼ | Tools ▼ | Data ▼ | News ▼ | Help ▼ | About ▼ | Search COSMIC... | Login ▼

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



Browse the **genomic landscape** of cancer

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 New
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

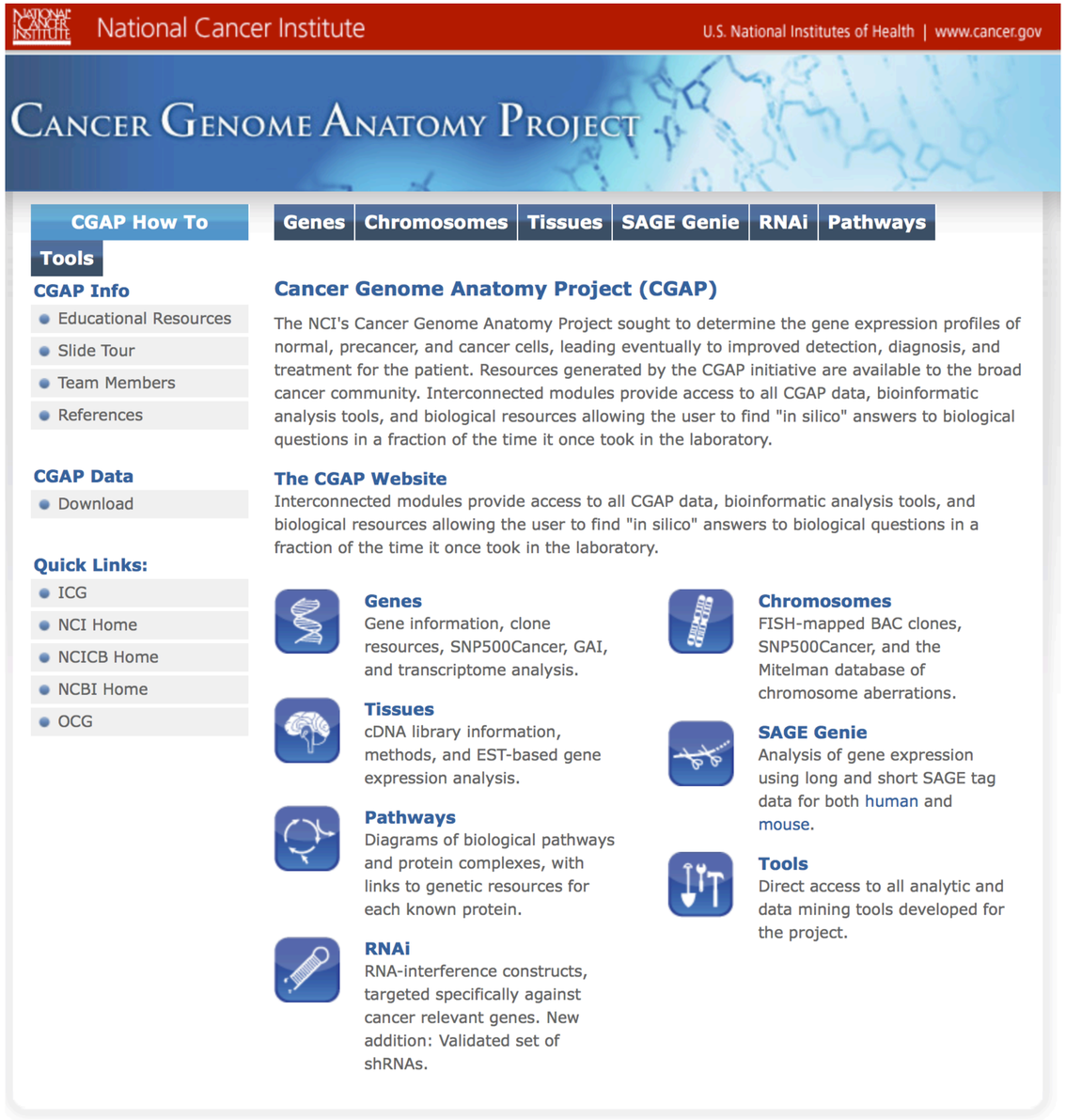
Cancer Gene Census Update

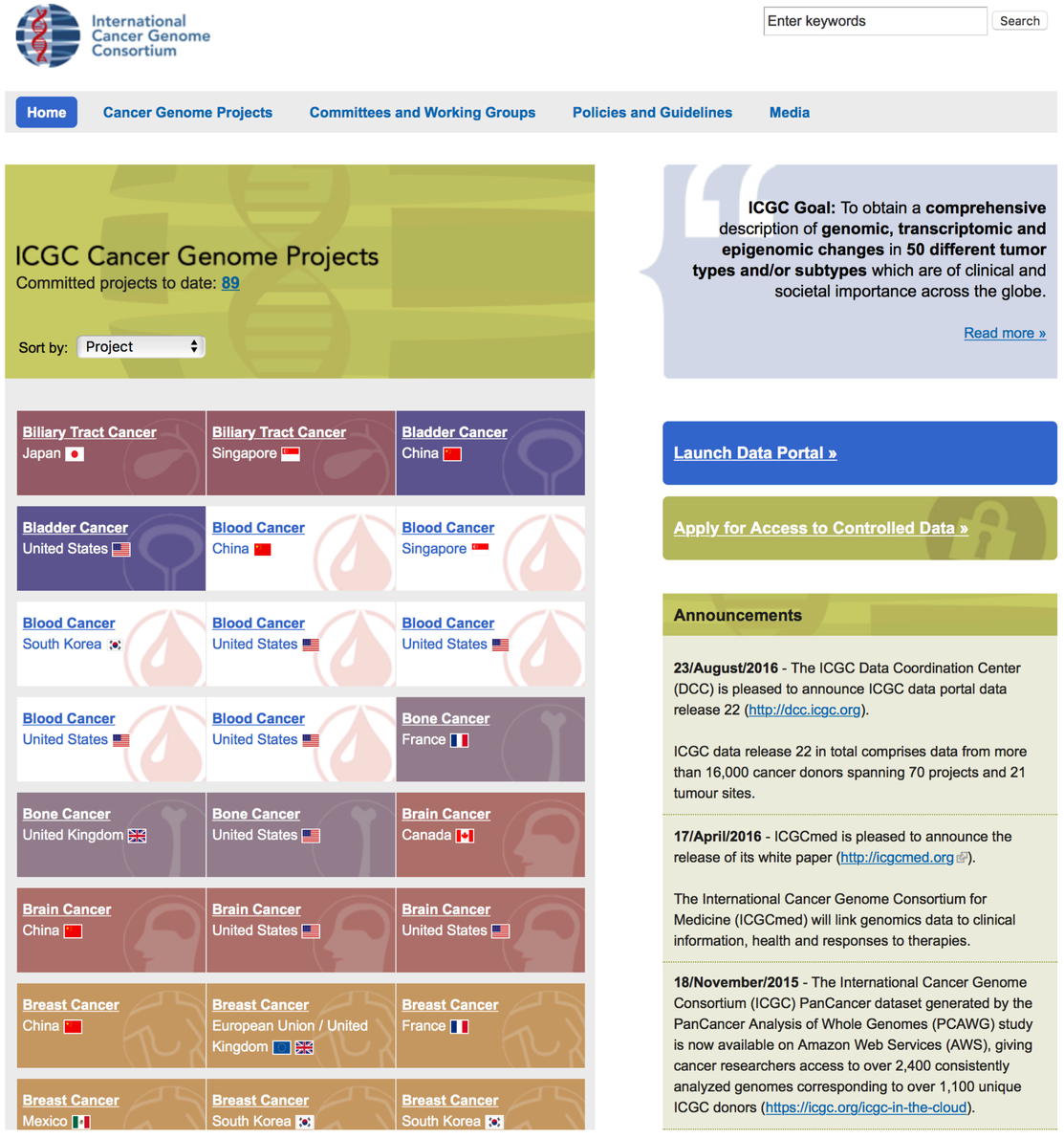
7 genes have been added to the Cancer Gene Census -

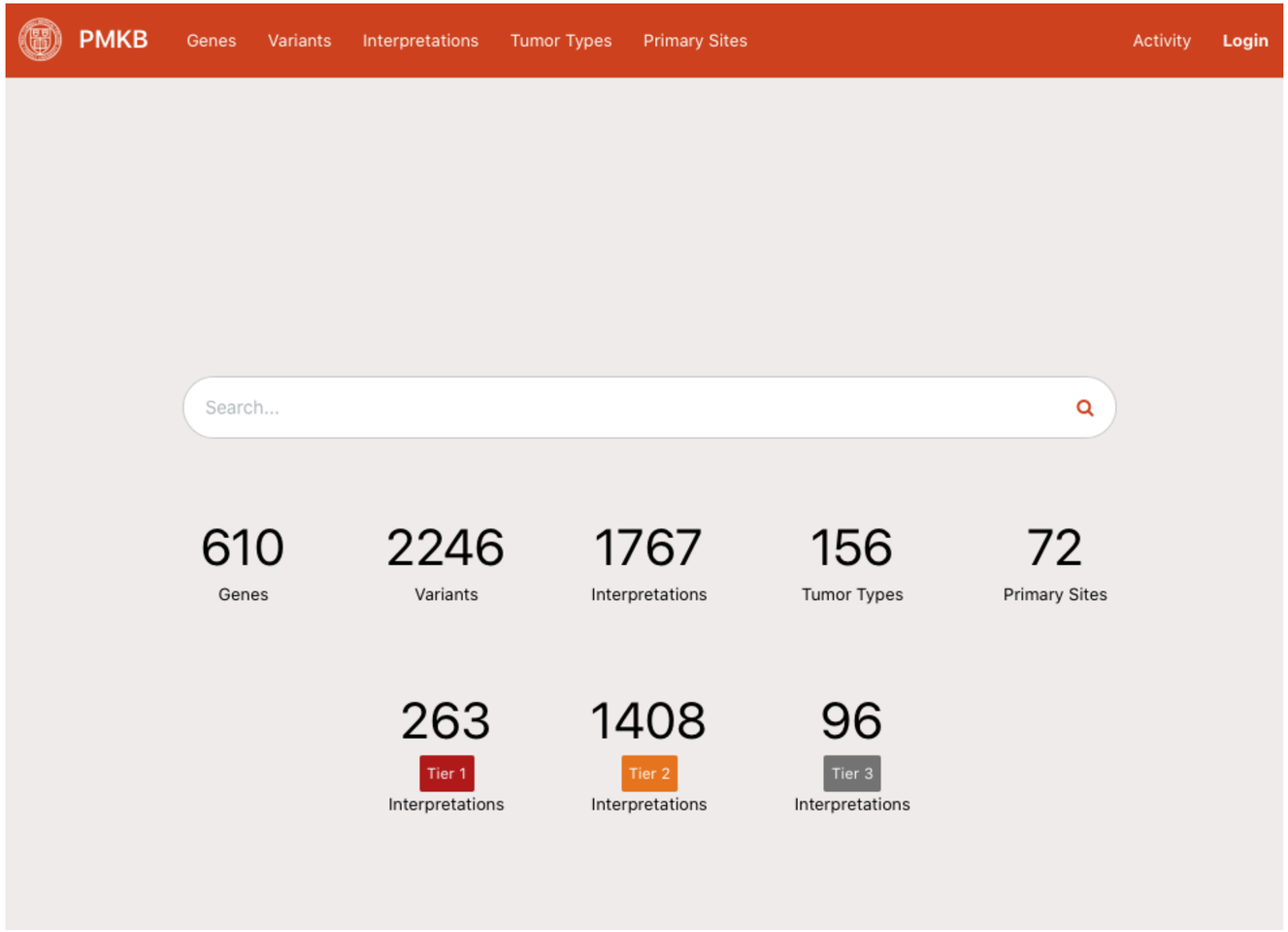
EPAS1 - Endothelial PAS domain protein 1.
PTPRT - Protein tyrosine phosphatase, receptor type T.
PPM1D - Protein phosphatase, Mg2+/Mn2+ dependent 1D.
BTK - Bruton tyrosine kinase.
PREX2 - Phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 2.
TP63 - Tumour protein p63.
QKI - QKI, KH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

For full details, see the Datasheet.

RESOURCES FOR CANCER GENOMICS







About

The Precision Medicine Knowledgebase (PMKB) is a project of the Englander Institute for Precision Medicine (EIPM) at Weill Cornell Medicine.

Weill Cornell Medicine NewYork-Presbyterian

PMKB provides information about clinical interpretations of cancer variants in a structured way. It allows users to browse, submit and edit existing entries for continued growth of the knowledgebase. All changes are reviewed by molecular pathologists and oncologists.

Part of the Variant Interpretation for Cancer Consortium (VICC), a Driver Project of the Global Alliance for Genomic Health (GA4GH)

Global Alliance for Genomics & Health

Academic Partners

San Raffaele University

UCSF University of California San Francisco

University of Bern

Non-Academic Partners

Microsoft

VarSome

Download Information

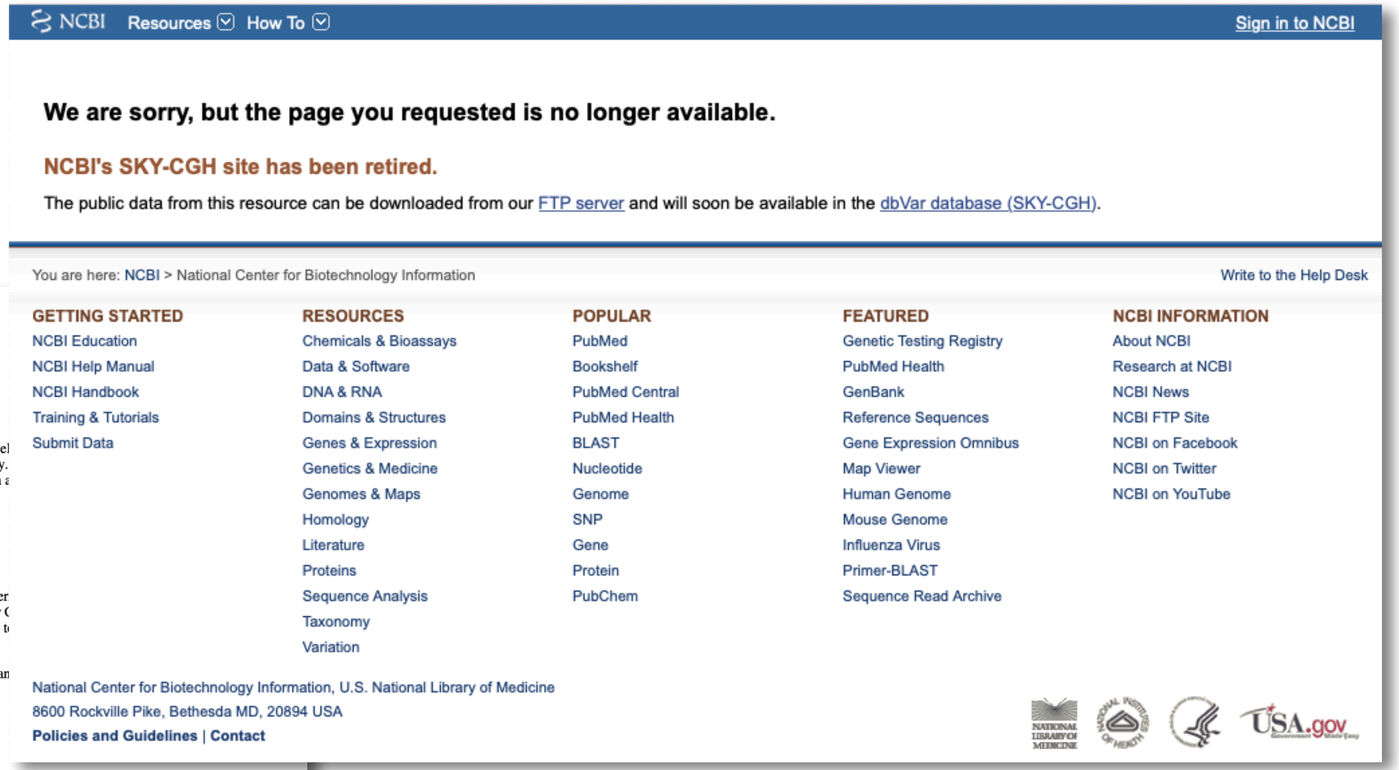
Download All Interpretations (Excel)

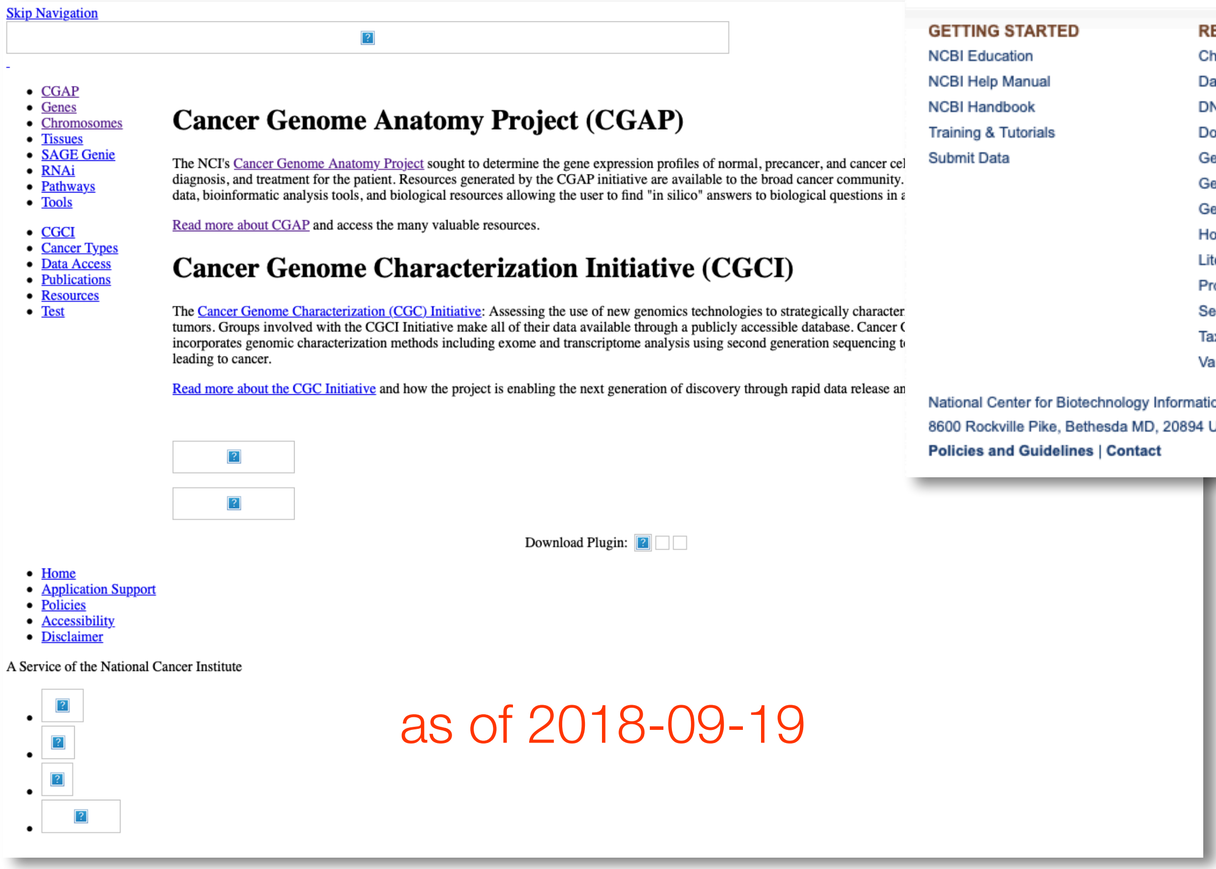
Download All Interpretations (CSV)

https://pmkb.weill.cornell.edu

Weill Cornell Medicine
Englander Institute
for Precision Medicine

RESOURCES FOR GENOMICS - THEY MAY BREAK SOMETIMES ...





as of 2018-09-19

VARIANT RESOURCES FOR CANCER GENOMICS

Resource name	Primary institute	Constituent Knowledge base	Cancer focused	Therapeutic evidence	Predisp. evidence	Diagnostic evidence	Prognostic evidence	Variant emphasis	URL
Cancer Genome Interpreter (CGI)	Institute for Research in Biomedicine, Barcelona, Spain	x	x	x				Somatic	https://www.cancergenomeinterpreter.org/home
Clinical Interpretation of Variants in Cancer (CIViC)	Washington University School of Medicine (WashU)	x	x	x	x	x	x	All variants	www.civicdb.org
JAX Clinical Knowledgebase (CKB)	The Jackson Laboratory	x	x	x	x	x	x	Somatic	https://ckb.jax.org/
Molecular Match	Molecular Match	x	x	x			x	Somatic	https://app.molecularmatch.com/
OncoKB	Memorial Sloan Kettering Cancer Center	x	x	x				Somatic	http://oncokb.org/#/
Precision Medicine Knowledgebase (PMKB)	Weill Cornell Medical College	x	x	x	x	x	x	Somatic	https://pmkb.weill.cornell.edu/
BRCA exchange	GA4GH	x	x		x			Germline	http://brcaexchange.org/
Cancer Driver Log (CanDL)	Ohio State University (OSU) / James Cancer Hospital		x	x				Somatic	https://candl.osu.edu/
Gene Drug Knowledge Database	Synapse		x	x		x	x	Somatic	https://www.synapse.org/#!Synapse:syn2370773/wiki/62707
MatchMiner	Dana-Farber Cancer Institute		x					Somatic	http://bcb.dfci.harvard.edu/knowledge-systems/
COSMIC Drug Resistance Curation	Wellcome Trust Sanger Institute		x	x				Somatic	http://cancer.sanger.ac.uk/cosmic/drug_resistance
My Cancer Genome	Vanderbilt University		x	x		x	x	Somatic	https://www.mycancergenome.org/
NCI Clinical Trials	National Cancer Institute of the National Institutes of Health		x					Somatic	www.cancer.gov/about-cancer/treatment/clinical-trials
Personalized Cancer Therapy Database	MD Anderson Cancer Center		x	x	x	x	x	Somatic	https://pct.mdanderson.org/#/home
ClinGen Knowledge Base	ClinGen				x			Germline	https://www.clinicalgenome.org/resources-tools/
ClinVar	National Center for Biotechnology Information			x	x			All variants	http://www.ncbi.nlm.nih.gov/clinvar/
Pharmacogenomics Knowledgebase (PharmGKB)	Stanford University			x				Germline	https://www.pharmgkb.org/
The Human Gene Mutation Database (HGMD)	Institute of Medical Genetics in Cardiff				x			Germline	http://www.hgmd.cf.ac.uk

Wagner et al. A harmonized meta-knowledgebase of clinical interpretations of cancer genomic variants. Nature Genetics (under review 2018)

arrayMap

Reference resource for copy number variation data in cancer

arrayMap

Search Samples
Search Publications
Progenetix

University of Zurich[UZH]

Citation & Licensing
User Guide
People
Beacon+

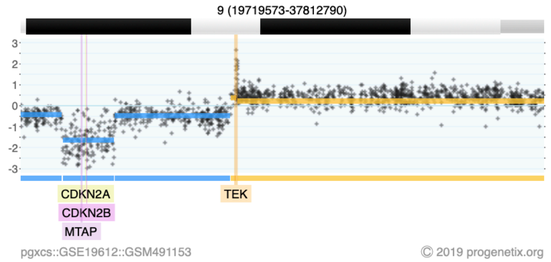
162.158.150.56

visualizing cancer genome array data @ arraymap.org

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:

- 72724 genomic array profiles
- 898 experimental series
- 257 array platforms
- ICD-O 341 ICD-O cancer entities
- 795 publications (Pubmed entries)



Genomic copy number imbalances on chromosome 9 in a case of Glioblastoma (**GSM491153**), indcating, 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 we provide through our Progenetix project.

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

RELATED PUBLICATIONS

Cai H, Gupta S, Rath P, Ai N, Baudis M. arrayMap 2014: an updated cancer genome resource. Nucleic Acids Res. 2015 Jan;43(Database issue). Epub 2014 Nov 26.

Cai, H., Kumar, N., & Baudis, M. 2012. arrayMap: A Reference Resource for Genomic Copy Number Imbalances in Human Malignancies. *PLoS One* 7(5), e36944.

Baudis, M. 2007. Genomic imbalances in 5918 malignant epithelial tumors: An explorative meta-analysis of chromosomal CGH data. *BMC Cancer* 7:226.

Baudis, M. 2006. Online database and bioinformatics toolbox to support data mining in cancer cytogenetics. *Biotechniques* 40, no. 3: 296-272.

Baudis, M, and ML Cleary. 2001. Progenetix.net: an online repository for molecular cytogenetic aberration data. *Bioinformatics* 12, no. 17: 1228-1229.

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.

arrayMap

Progenetix - Cancer CNV Information Resource

- launched online in 2001 as *progenetix.net*
- **curation** of published CNV profiling data
  - originally cCGH and CNV extraction from Mitelman database
  - + aCGH, WES, WGS; - karyotype data
- increasingly focused on representing the "publication landscape" of cancer genome screening - What? Where?
- Genomes:
  - 93640 CNV profiles (cCGH, aCGH, WES, WGS) from 469 cancer types (NCIt & ICD-O mapping)
  - 6'817'645 "CNVs" (i.e. called segments)
- Articles:
  - 3229 registered articles
    - geographic mapping
    - "cancer type" labelling
  - represent 174'530 reported samples

progenetix

Progenetix :: Info

Structural Cancer Genomics Resource
Documentation and Example Pages

News
About...
Documentation
Publications
Data Pages

Related Sites

arrayMap
Baudisgroup @ UZH
Beacon+
SchemaBlocks {S}[B]
ELIXIR Beacon
Baudisgroup Internal

Github Projects

baudisgroup
progenetix
ELIXIR Beacon

Tags

API article code documentation licensing maps statistics tools

Progenetix Publication Collection

The curent page lists publications of whole genome screening experiments in cancer, registered in the Progenetix publication collection.
This page is a *beta* version, intended to replace the original **publications** page.

Show 50 entries

Search:

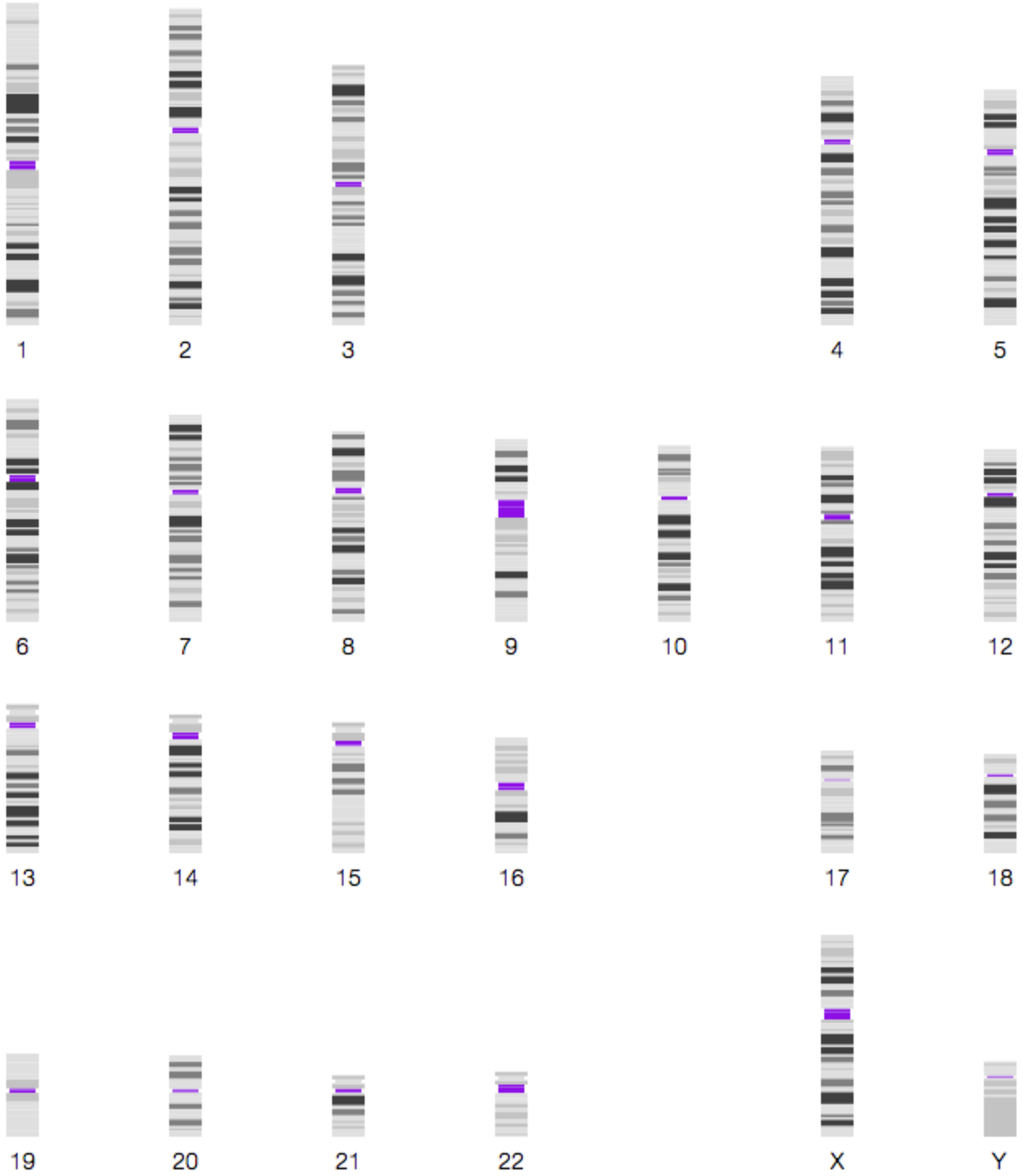
Publication	Samples			
	cCGH	aCGH	WES	WGS
Harada K, Okamoto W, Mimaki S, Kawamoto Y, Bando et al. (2019): Comparative sequence analysis of patient-matched primary colorectal cancer, metastatic, and recurrent metastatic tumors ... BMC Cancer 19(1), 2019 (30898102)	0	0	4	0
Lavrov AV, Chelysheva EY, Adilgereeva EP, Shukhov et al. (2019): Exome, transcriptome and miRNA analysis don't reveal any molecular markers of TKI efficacy in primary CML ... BMC Med Genomics 12(Suppl 2), 2019 (30871622)	0	0	62	0
Zandberg DP, Tallon LJ, Nagaraj S, Sadzewicz LK, Zhang et al. (2019): Intratumor genetic heterogeneity in squamous cell carcinoma of the oral cavity. Head Neck, 2019 (30869813)	0	0	5	0
Heinrich MC, Patterson J, Beadling C, Wang Y, Debiec-Rychter et al. (2019): Genomic aberrations in cell cycle genes predict progression of KIT-mutant gastrointestinal stromal tumors ... Clin Sarcoma Res 9, 2019 (30867899)	0	0	29	0
Jiao J, Sagnelli M, Shi B, Fang Y, Shen Z, Tang T, Dong et al. (2019): Genetic and epigenetic characteristics in ovarian tissues from polycystic ovary syndrome patients with irregular ... BMC Endocr Disord 19(1), 2019 (30866919)	0	0	20	0
Mueller S, Jain P, Liang WS, Kilburn L, Kline C, Gupta et al. (2019): A pilot precision medicine trial for children with diffuse intrinsic pontine glioma - PNOC003: a report from the Pacific ... Int. J. Cancer, 2019 (30861105)	0	0	14	14
Xie SN, Cai YJ, Ma B, Xu Y, Qian P, Zhou JD, Zhao et al. (2019): The genomic mutation spectrums of breast fibroadenomas in Chinese population by whole exome sequencing ... Cancer Med, 2019 (30851086)	0	0	12	0

Showing 1 to 50 of 3,232 entries

Previous 1 2 3 4 5 ... 65 Next

Genome Editions

Sizes | positions | mappings



Chromosome	Basepair length (GRCh38)
1	248'956'422
2	242'193'529
3	198'295'559
4	190'214'555
5	181'538'259
6	170'805'979
7	159'345'973
8	145'138'636
9	138'394'717
10	133'797'422
11	135'086'622
12	133'275'309
13	114'364'328
14	107'043'718
15	101'991'189
16	90'338'345
17	83'257'441
18	80'373'285
19	59'128'983
20	64'444'167
21	46'709'983
22	50'818'468
X	156'040'895
Y	57'227'415
	3'080'419'480

genome.ucsc.edu

cytoBand_UCSC_hg18.txt

chr1	0	2300000	p36.33	gneg
chr1	2300000	5300000	p36.32	gpos25
chr1	5300000	7100000	p36.31	gneg
chr1	7100000	9200000	p36.23	gpos25
chr1	9200000	12600000	p36.22	gneg
chr1	12600000	16100000	p36.21	gpos50
chr1	16100000	20300000	p36.13	gneg
chr1	20300000	23800000	p36.12	gpos25
chr1	23800000	27800000	p36.11	gneg
chr1	27800000	30000000	p35.3	gpos25
chr1	30000000	32200000	p35.2	gneg
chr1	32200000	34400000	p35.1	gpos25
chr1	34400000	39600000	p34.3	gneg
chr1	39600000	43900000	p34.2	gpos25
chr1	43900000	46500000	p34.1	gneg
chr1	46500000	51300000	p33	gpos75
chr1	51300000	56200000	p32.3	gneg
chr1	56200000	58700000	p32.2	gpos50
chr1	58700000	60900000	p32.1	gneg
...	...	...	...	...
chrX	130300000	133500000	q26.2	gpos25
chrX	133500000	137800000	q26.3	gneg
chrX	137800000	140100000	q27.1	gpos75
chrX	140100000	141900000	q27.2	gneg
chrX	141900000	146900000	q27.3	gpos100
chrX	146900000	154913754	q28	gneg
chrY	0	1700000	p11.32	gneg
chrY	1700000	3300000	p11.31	gpos50
chrY	3300000	11200000	p11.2	gneg
chrY	11200000	11300000	p11.1	acen
chrY	11300000	12500000	q11.1	acen
chrY	12500000	14300000	q11.21	gneg
chrY	14300000	19000000	q11.221	gpos50
chrY	19000000	21300000	q11.222	gneg
chrY	21300000	25400000	q11.223	gpos50
chrY	25400000	27200000	q11.23	gneg
chrY	27200000	57772954	q12	gvar

Cytogenetic band sizes

chromosome	band start position	band stop position	cytogenetic band	staining intensity	band size
chr6	63400000	63500000	q11.2	gneg	100000
chr15	64900000	65000000	q22.32	gpos25	100000
chr17	22100000	22200000	p11.1	acen	100000
chrX	65000000	65100000	q11.2	gneg	100000
chrY	11200000	11300000	p11.1	acen	100000
chr17	35400000	35600000	q21.1	gneg	200000
chr3	44400000	44700000	p21.32	gpos50	300000
chr3	51400000	51700000	p21.2	gpos25	300000
chr9	132500000	132800000	q34.12	gpos25	300000
chr13	45900000	46200000	q14.13	gneg	300000
chr15	65000000	65300000	q22.33	gneg	300000
chr1	120700000	121100000	p11.2	gneg	400000
chr8	39500000	39900000	p11.22	gpos25	400000
chr9	72700000	73100000	q21.12	gneg	400000
chr16	69400000	69800000	q22.2	gpos50	400000
chr19	43000000	43400000	q13.13	gneg	400000
chr9	70000000	70500000	q13	gneg	500000
chr20	41100000	41600000	q13.11	gneg	500000
...	...	...	...	...	...
chr9	51800000	60300000	q11	acen	8500000
chrX	76000000	84500000	q21.1	gpos100	8500000
chr11	76700000	85300000	q14.1	gpos100	8600000
chr13	77800000	86500000	q31.1	gpos100	8700000
chr7	77400000	86200000	q21.11	gpos100	8800000
chr8	29700000	38500000	p12	gpos75	8800000
chr3	14700000	23800000	p24.3	gpos100	9100000
chr5	82800000	91900000	q14.3	gpos100	9100000
chr6	104800000	113900000	q21	gneg	9100000
chrX	120700000	129800000	q25	gpos100	9100000
chr9	60300000	70000000	q12	gvar	9700000
chr1	212100000	222100000	q41	gpos100	10000000
chr1	128000000	142400000	q12	gvar	14400000
chr1	69500000	84700000	p31.1	gpos100	15200000
chrY	27200000	57772954	q12	gvar	30572954

Positional genomic data has to be evaluated in the context of the correct edition

Chromosome	Basepairs 2003 (HG16)	Basepairs 2006 (HG18)	Basepairs 2009 (HG19)	Basepairs 2013 (GRCh38)	HG16 => HG19
1	246'127'941	247'249'719	249'250'621	248'956'422	2'828'481
2	243'615'958	242'951'149	243'199'373	242'193'529	-1'422'429
3	199'344'050	199'501'827	198'022'430	198'295'559	-1'048'491
4	191'731'959	191'273'063	191'154'276	190'214'555	-1'517'404
5	181'034'922	180'857'866	180'915'260	181'538'259	503'337
6	170'914'576	170'899'992	171'115'067	170'805'979	-108'597
7	158'545'518	158'821'424	159'138'663	159'345'973	800'455
8	146'308'819	146'274'826	146'364'022	145'138'636	-1'170'183
9	136'372'045	140'273'252	141'213'431	138'394'717	2'022'672
10	135'037'215	135'374'737	135'534'747	133'797'422	-1'239'793
11	134'482'954	134'452'384	135'006'516	135'086'622	603'668
12	132'078'379	132'349'534	133'851'895	133'275'309	1'196'930
13	113'042'980	114'142'980	115'169'878	114'364'328	1'321'348
14	105'311'216	106'368'585	107'349'540	107'043'718	1'732'502
15	100'256'656	100'338'915	102'531'392	101'991'189	1'734'533
16	90'041'932	88'827'254	90'354'753	90'338'345	296'413
17	81'860'266	78'774'742	81'195'210	83'257'441	1'397'175
18	76'115'139	76'117'153	78'077'248	80'373'285	4'258'146
19	63'811'651	63'811'651	59'128'983	59'128'983	-4'682'668
20	63'741'868	62'435'964	63'025'520	64'444'167	702'299
21	46'976'097	46'944'323	48'129'895	46'709'983	-266'114
22	49'396'972	49'691'432	51'304'566	50'818'468	1'421'496
X	153'692'391	154'913'754	155'270'560	156'040'895	2'348'504
Y	50'286'555	57'772'954	59'373'566	57'227'415	6'940'860
	3'070'128'059	**3'080'419'480**	**3'095'677'412**	**3'088'781'199**	**18'653'140**

University of Zurich[UZH]

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

arraymap.org

progenetix.org

info.baudisgroup.org

sib.swiss/baudis-michael

imls.uzh.ch/en/research/baudis

Global Alliance
for Genomics & Health