

# BIO392

# Bioinformatics of Genome Variations

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

Michael Baudis **UZH SIB**  
Computational Oncogenomics



University of  
Zurich<sup>UZH</sup>

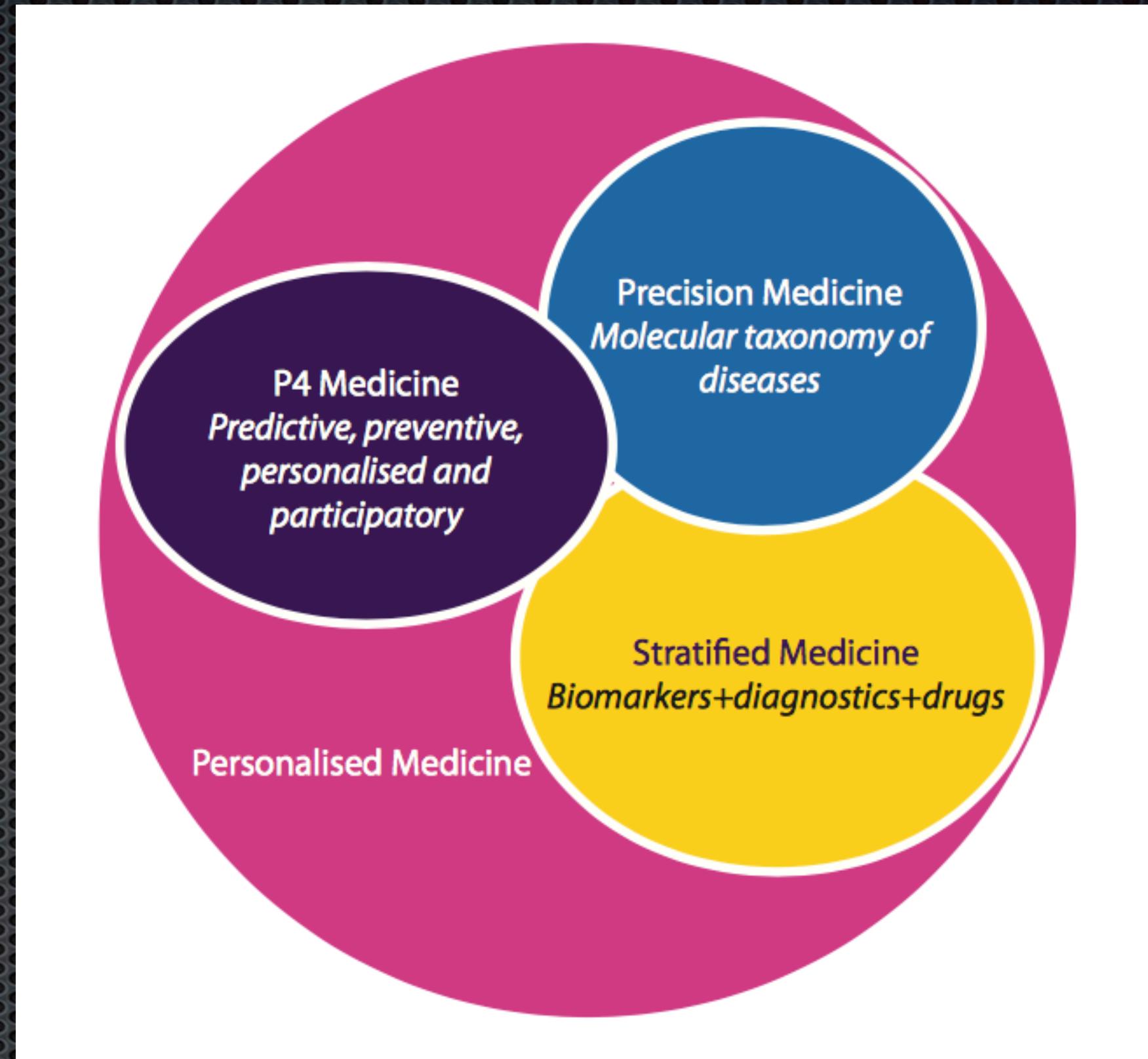
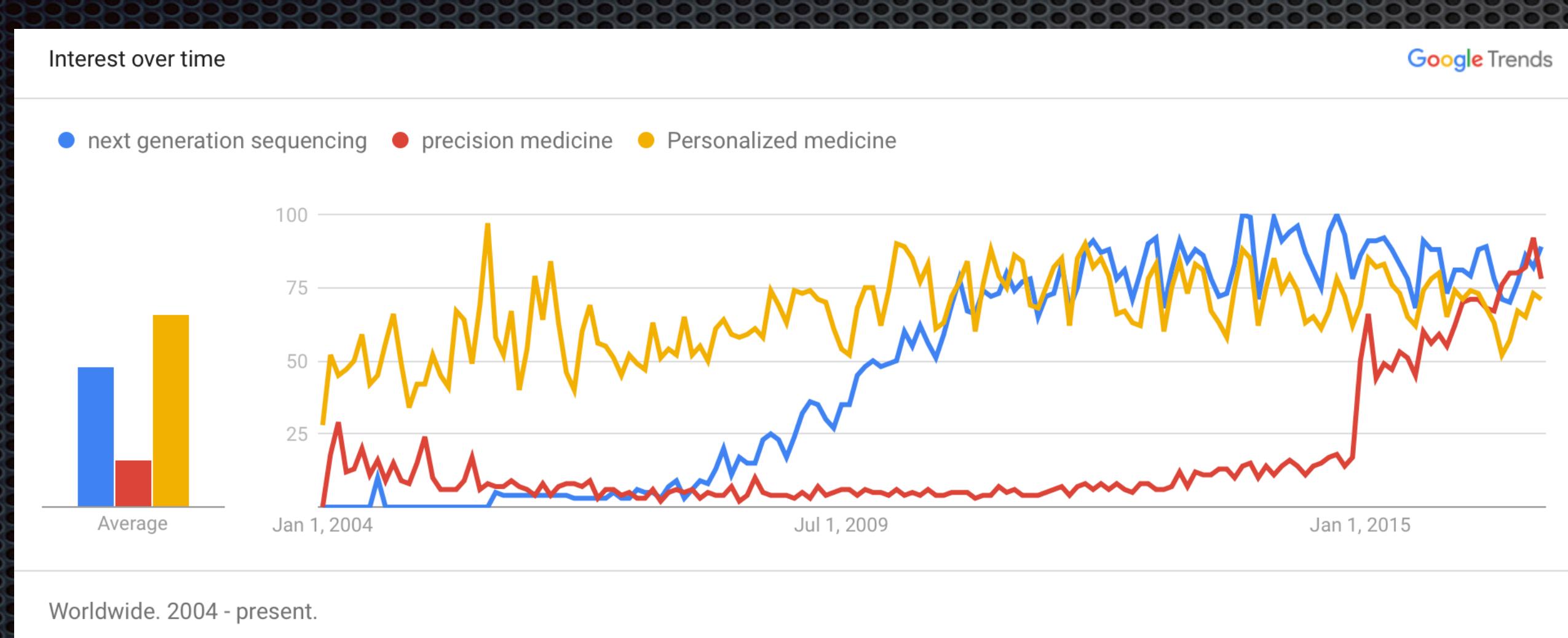
# BIO392: Course Schedule

	Tue Sep 15	Wed Sep 16	Thu Sep 17	Fri Sep 18	Tue Sep 22	Wed Sep 23	Thu Sep 24	Fri Sep 25	Tue Sep 29	Wed Sep 30	Thu Oct 1	Fri Oct 2	Tue Oct 6	Wed Oct 7
<b>09:00 - 10:00</b>		Michael: short introductory lecture about genome variation	Izaskun: Terminal / Unix / Files	Michael: genomic file formats		Izaskun: File typers / manipulations	Bo: Introduction to UCSC genome browser	Genome versions & liftover (continued)		Intro online platform Rahel?	Intro to R ...	Genome variants & populations		Exam
<b>10:00 - 11:00</b>		Github exercise: create user specific directories & upload/edit test files using Markdown	Izaskun: Terminal / Unix / Files	ToDo: File formats, calculations		Izaskun: File typers / manipulations	Bo: Introduction to UCSC genome browser	A taste of BLAST and scripting		GATK online course	DNAcopy ... exercise	plink		Exam
<b>11:00 - 12:00</b>			Izaskun: SIB online introduction to Unix			Izaskun: File typers / manipulations	Bo: Introduction to UCSC genome browser	A taste of BLAST and scripting		GATK online course	DNAcopy ... exercise			Exam
<b>13:00 - 14:00</b>	* Room information * Administrative - discuss times/days - exam	Literature (genome variant review papers ...)	Izaskun: SIB online introduction to Unix	ToDo: File formats, calculations	Recap W1; Q&A	Izaskun: file typers/manipulations reading/literature	genome browser related exercises	Python, R ...	Michael: genomic variants w/ respect to schedule etc.	GATK online course	Progenetix as tool for CNV frequencies etc.		Recap; Q&A; genomic privacy (?)	
<b>14:00 - 15:00</b>	Tina Siegenthaler: technical introduction (room,					Literature (genome analysis techniques ...)		read up/look up other genome resources, writing about use cases	Q&A, (Exercises???)			CNV literature etc.	Populations / Ethnicities /	
<b>15:00 - 16:30</b>	* explore course site * create Github accounts and forward to bio392@compbiozurich.org	Exercise: Q&A using Github				Genome technologies - brief notes about usage scenarios, pro & con								



# 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?

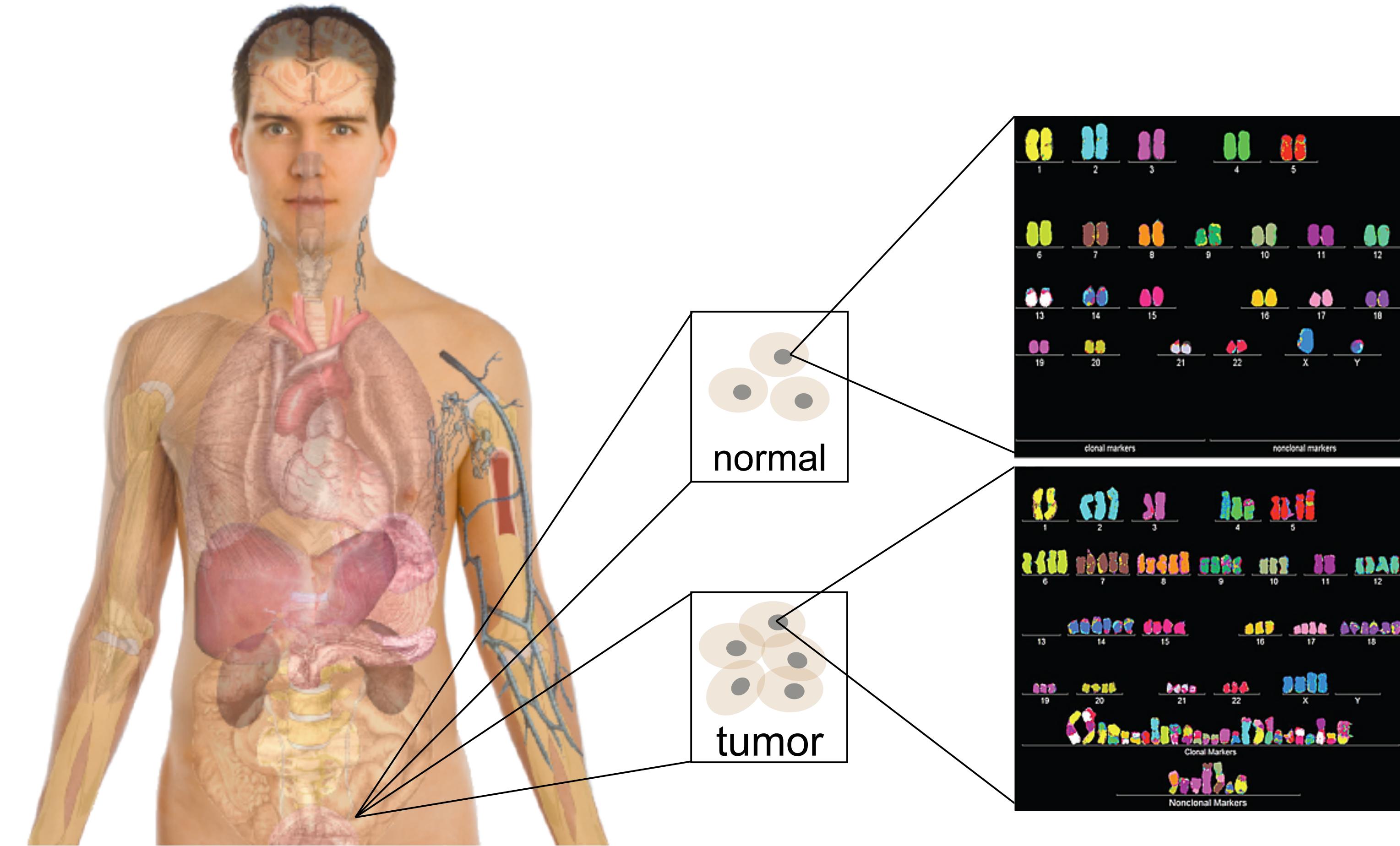


Source: PHG Foundation

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

Rapid whole genome sequencing and precision neonatology

Joshua E. Petrikis, MD<sup>a,\*</sup>, Laurel K. Willig, MD, FAAP<sup>b</sup>, Laurie D. Smith, MD, PhD<sup>c</sup>, and Stephen F. Kingsmore, MB, BAO, ChB, Dsc, FRCPath<sup>d,e</sup>

Barkur S. Shastry

SNP alleles in human disease and evolution



insight progress

Cancer genetics

Bruce A. J. Ponder

DISEASE MECHANISMS

## Mechanisms underlying structural variant formation in genomic disorders

Claudia M. B. Carvalho<sup>1,2</sup> and James R. Lupski<sup>1,3,4,5</sup>

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

## Genomic Classification of Cutaneous Melanoma

The Cancer Genome Atlas Network<sup>1,\*,\*</sup>

<sup>1</sup>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>

## PCN Frontier Review

doi:10.1111/pcn.12128

Copy-number variation in the pathogenesis of autism spectrum disorder

Emiko Shishido, PhD<sup>1,2,3</sup>, Branko Aleksić, MD, PhD<sup>3</sup> and Norio Ozaki, MD, PhD<sup>3,\*</sup>

RESEARCH ARTICLE

Open Access

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

Haoyang Cai<sup>1,2</sup>, Nitin Kumar<sup>1,2</sup>, Homayoun C Bagheri<sup>3</sup>, Christian von Mering<sup>1,2</sup>, Mark D Robinson<sup>1,2\*</sup>

Common gene variants, mortality and extreme longevity in humans

B.T. Heijmans<sup>a,b</sup>, R.G.J. Westendorp<sup>b</sup>, P.E. Slagboom<sup>a,\*</sup>

RESEARCH ARTICLE

Open Access

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

Alfred Beletz<sup>1,5\*</sup>, Philip Zimmermann<sup>2</sup>, Michael Baudis<sup>3</sup>, Nicole Brun<sup>4</sup>, Peter Bühlmann<sup>4</sup>, Oliver Laule<sup>2</sup>, Duc Luu<sup>1</sup>, Wilhelm Gruissem<sup>2</sup>, Peter Schraml<sup>1,\*</sup> and Holger Moch<sup>1</sup>

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

NEURODEVELOPMENT

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

Mustafa Sahin\* and Mridanka Sur\*

## The landscape of somatic copy-number alteration across human cancers

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

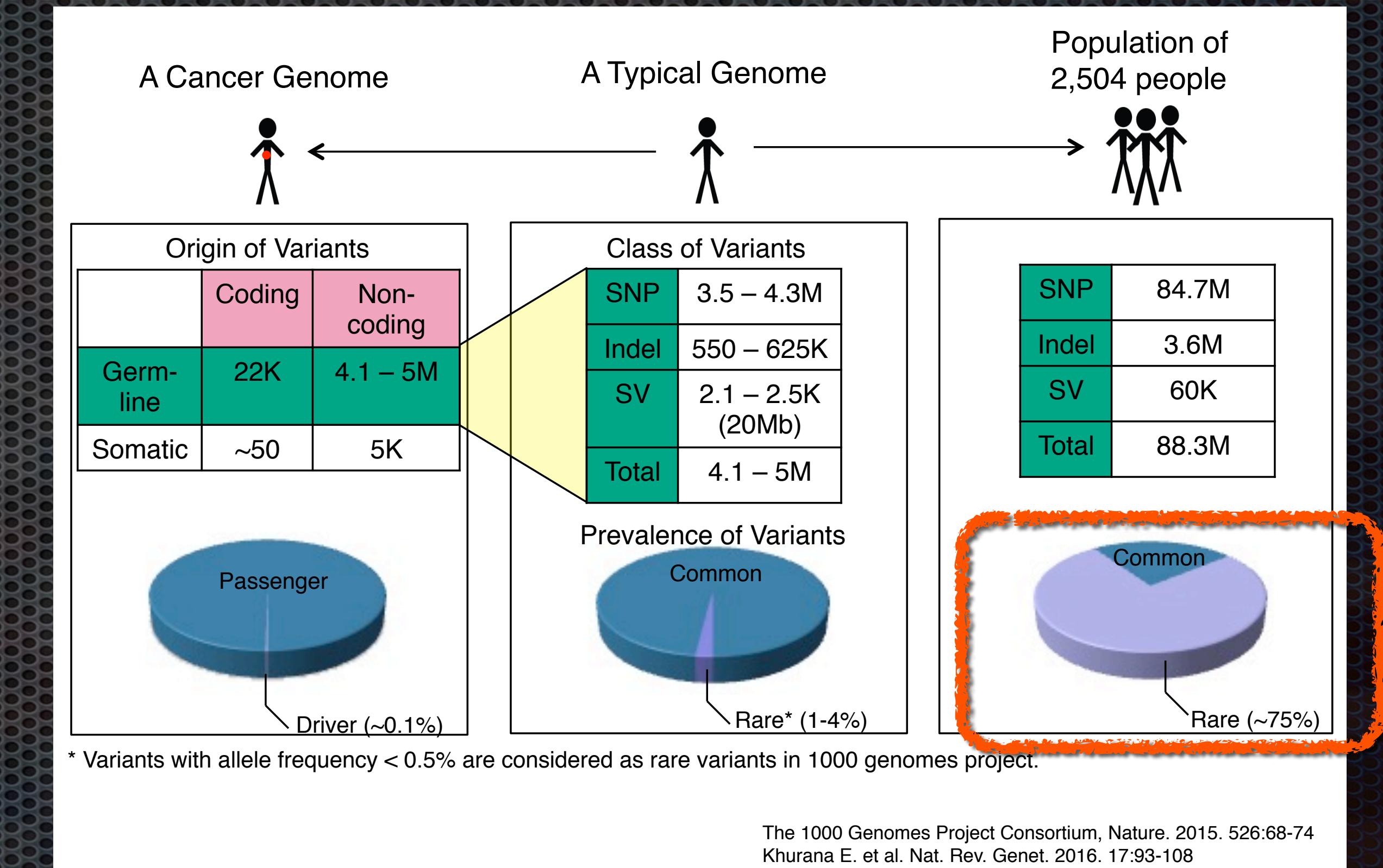
The trouble with human genome variation

---



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

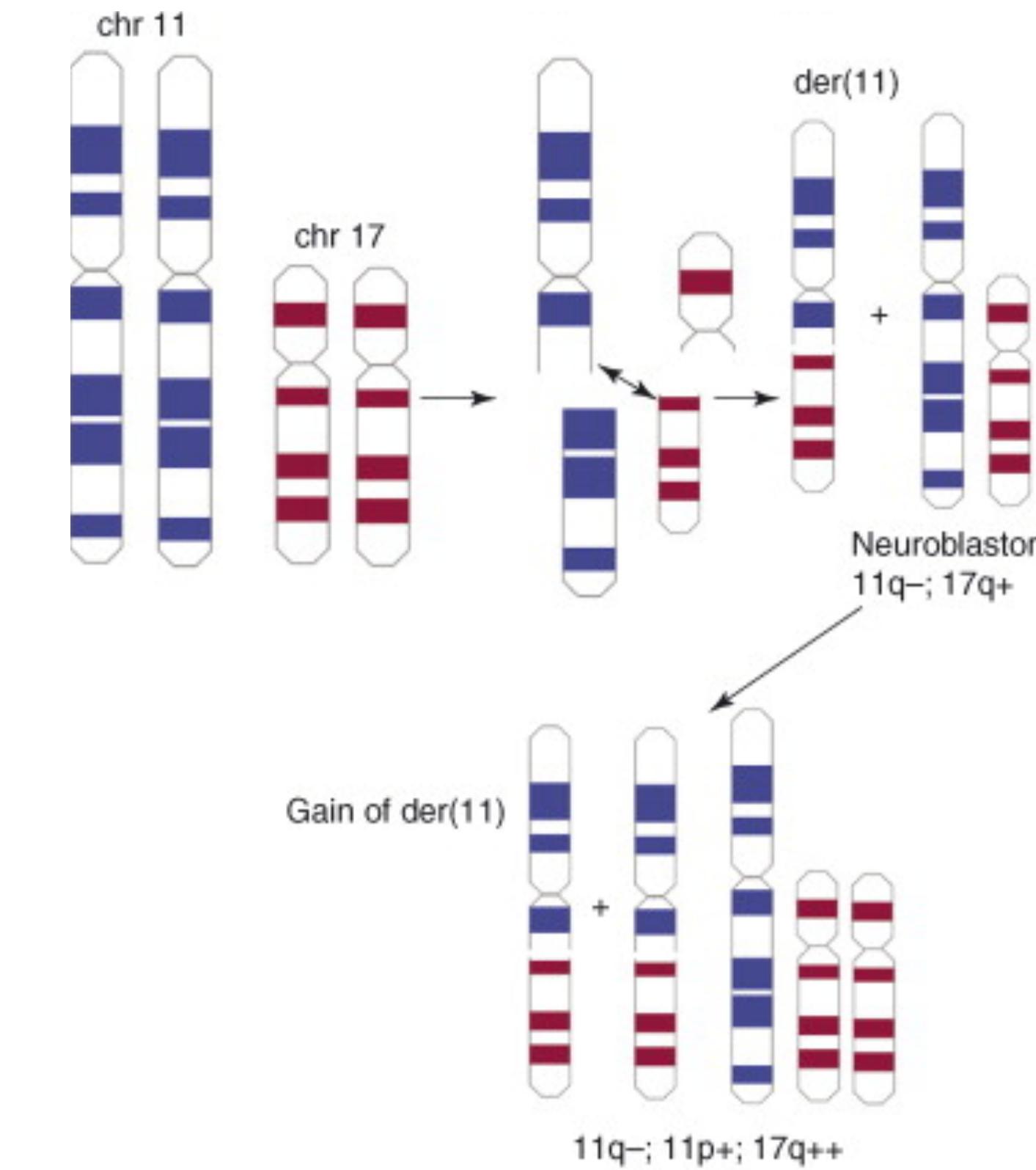
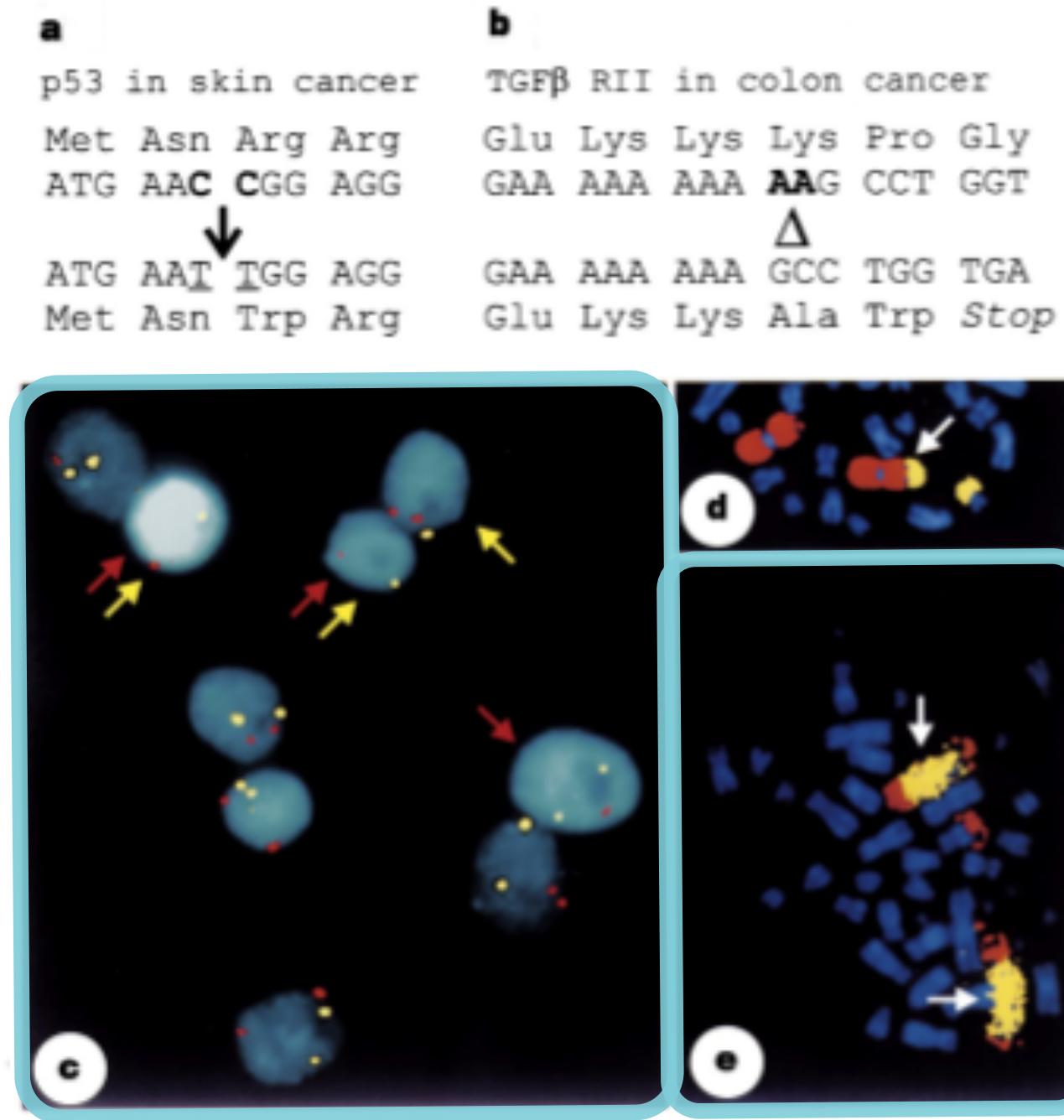
- a typical human genome (~3 billion base pairs) has ~5 million variants
- most of them are "**rare**"; i.e. can only be identified as recurring when sequencing thousands of people
- cancer cells accumulate additional variants, only **few** of which ("**drivers**") are relevant for the disease



Graphic adapted from Mark Gerstein ([GersteinLab.org](http://GersteinLab.org); @markgerstein)

# Mutations & genomic rearrangements in cancer

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



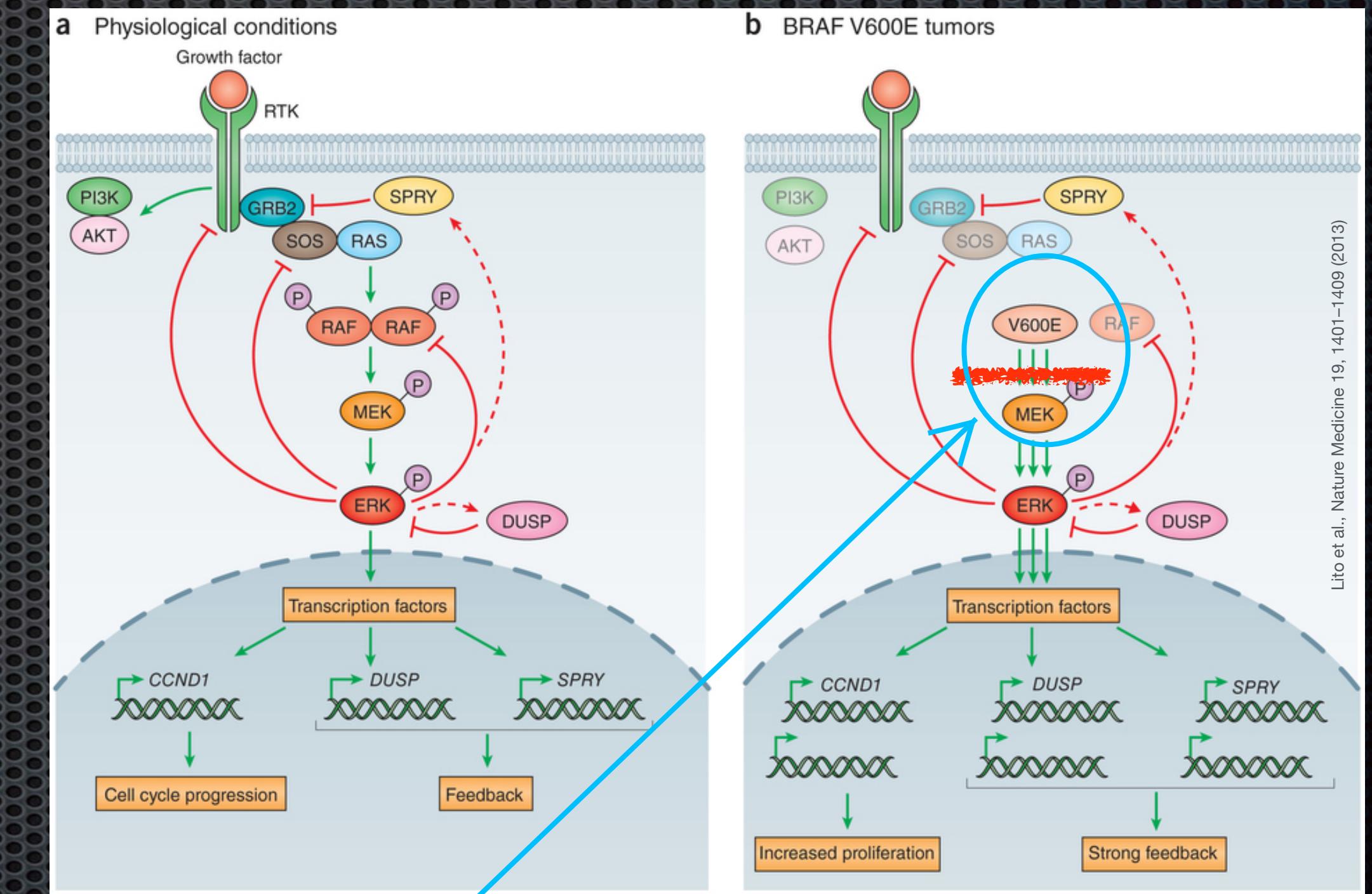
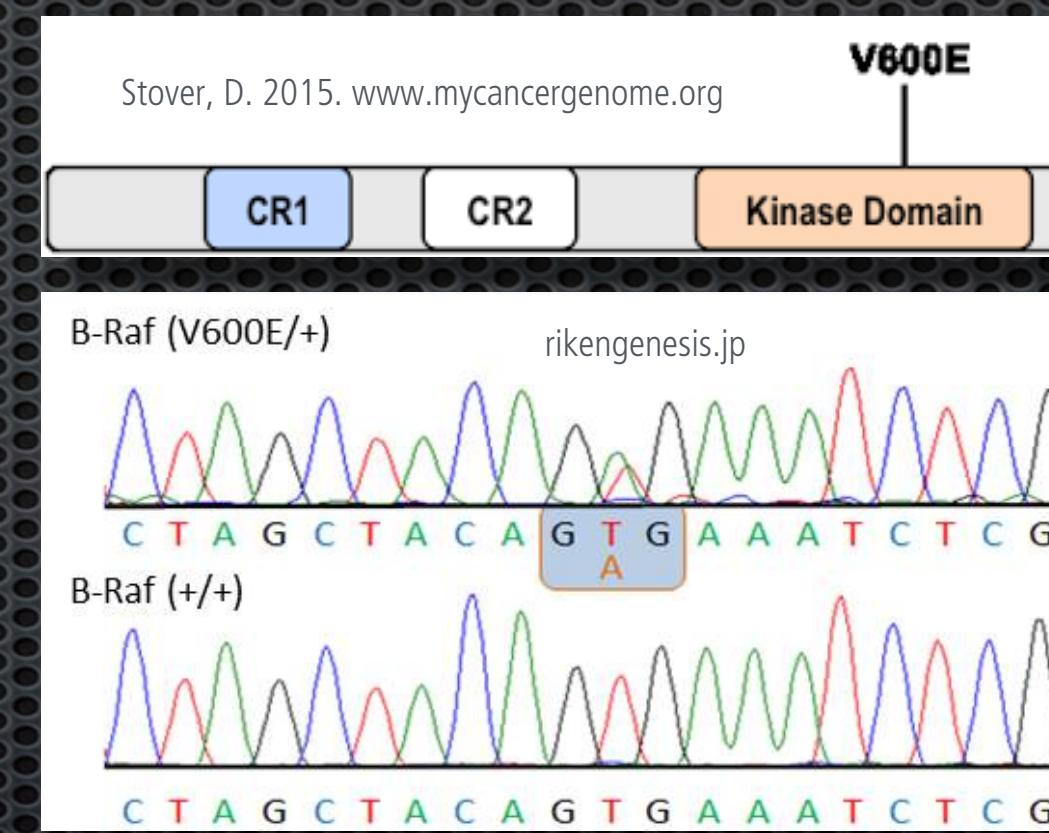
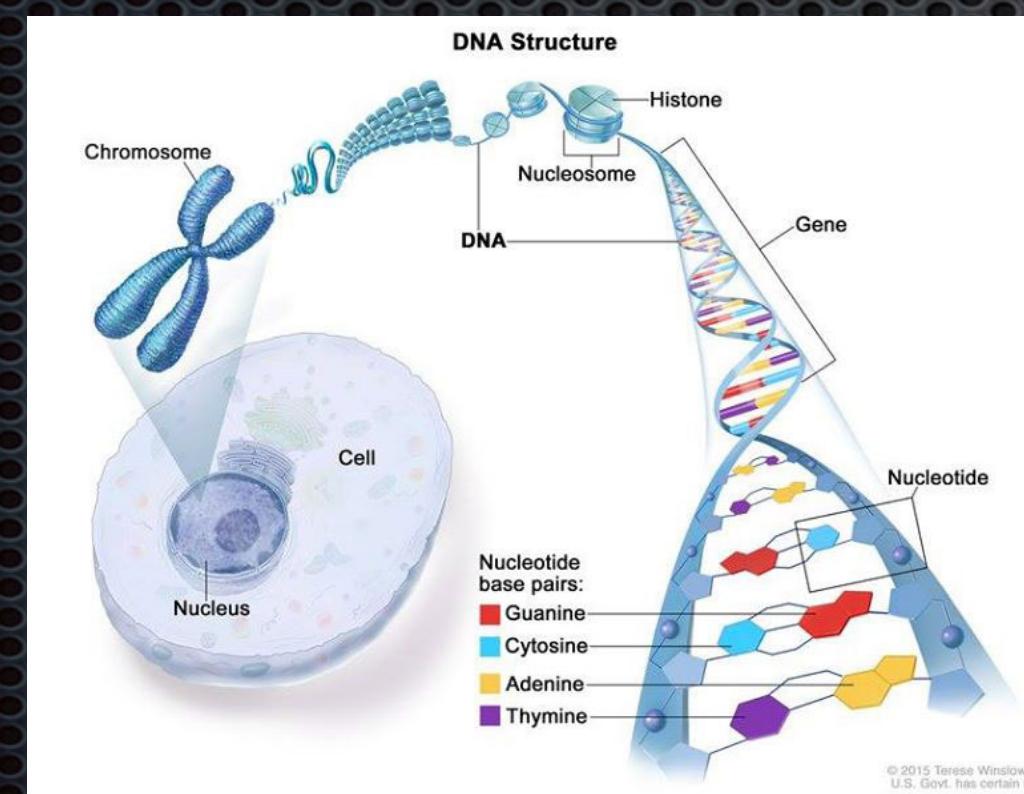
- a. small mutation (di-pyrimidine exchange) at p53 in Xeroderma pigmentosum patient)
- b. two-base deletion in *TGFB* in a colorectal cancer patient with mismatch repair deficiency
- c. chromosomal losses (FISH; red=3, yellow=12) in CRC
- d. t(1;17) in neuroblastoma, whole-chromosomal painting
- e.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

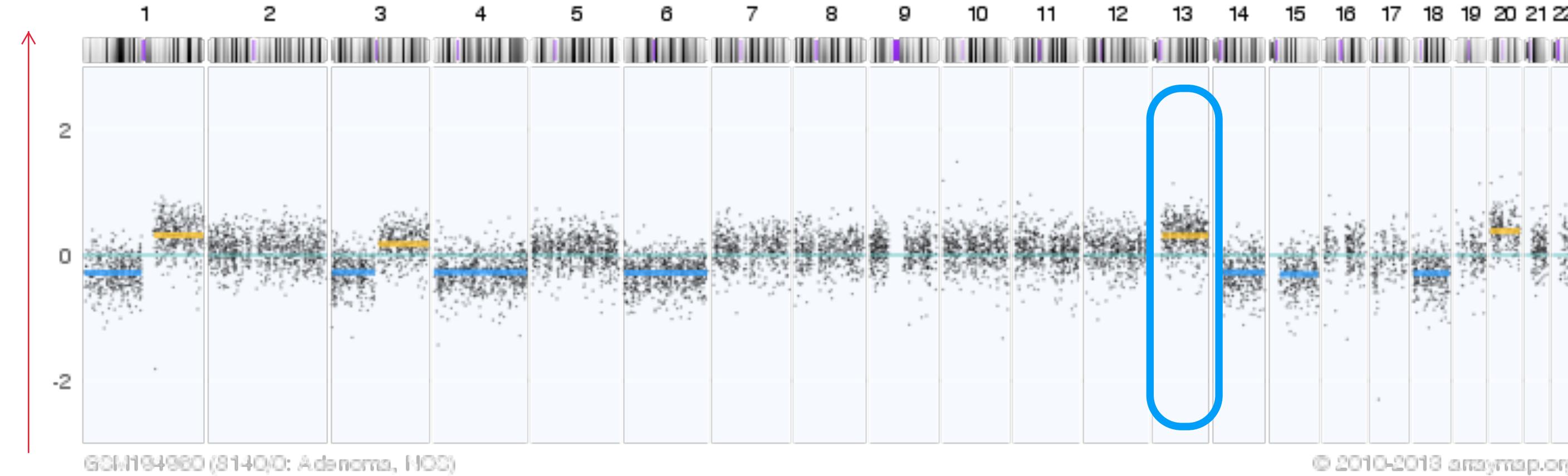
# BRAF V600E (c.1799T>A) Mutation Oncogene Activation by Single Nucleotide Alteration

- a single nucleotide exchange 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**)

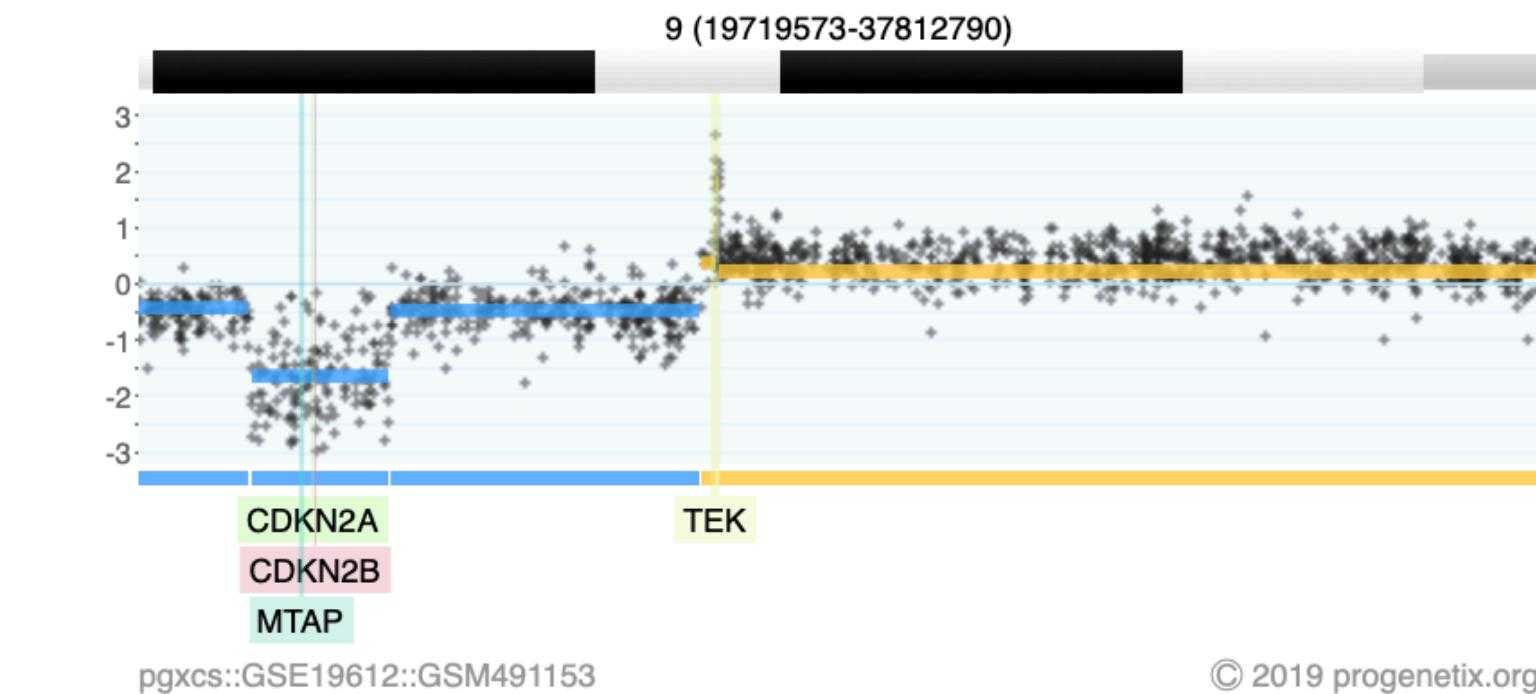


The BRAF V600E mutation leads to continuous phosphorylation of MEK, without the need for receptor based activation of the upstream pathway and loss of inhibitory feedback control.

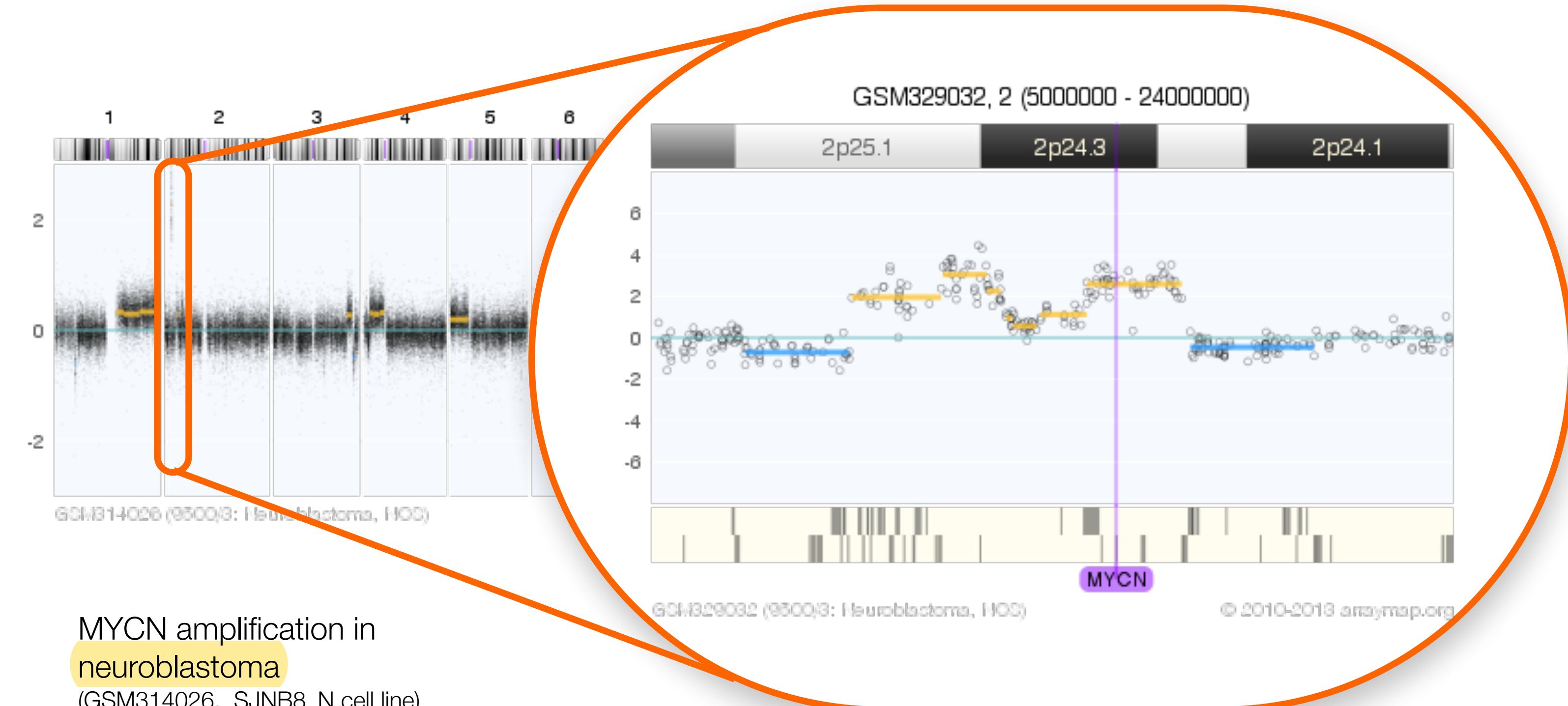
# Somatic Copy Number Variations



Gain of chromosome arm 13q in colorectal carcinoma



2-event, homozygous deletion in a Glioblastoma



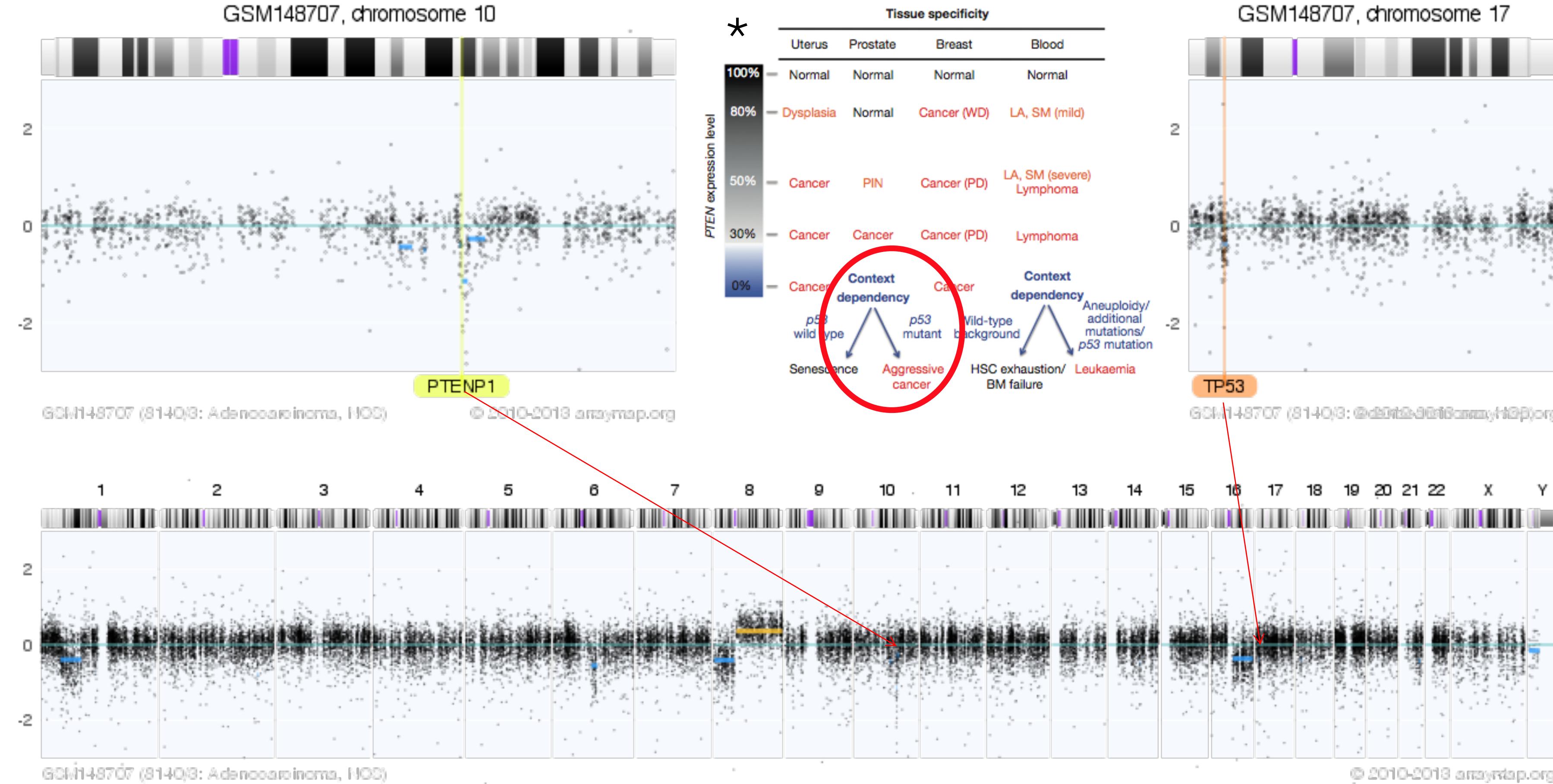
MYCN amplification in neuroblastoma  
(GSM314026, SJNB8\_N cell line)

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

arrayMap



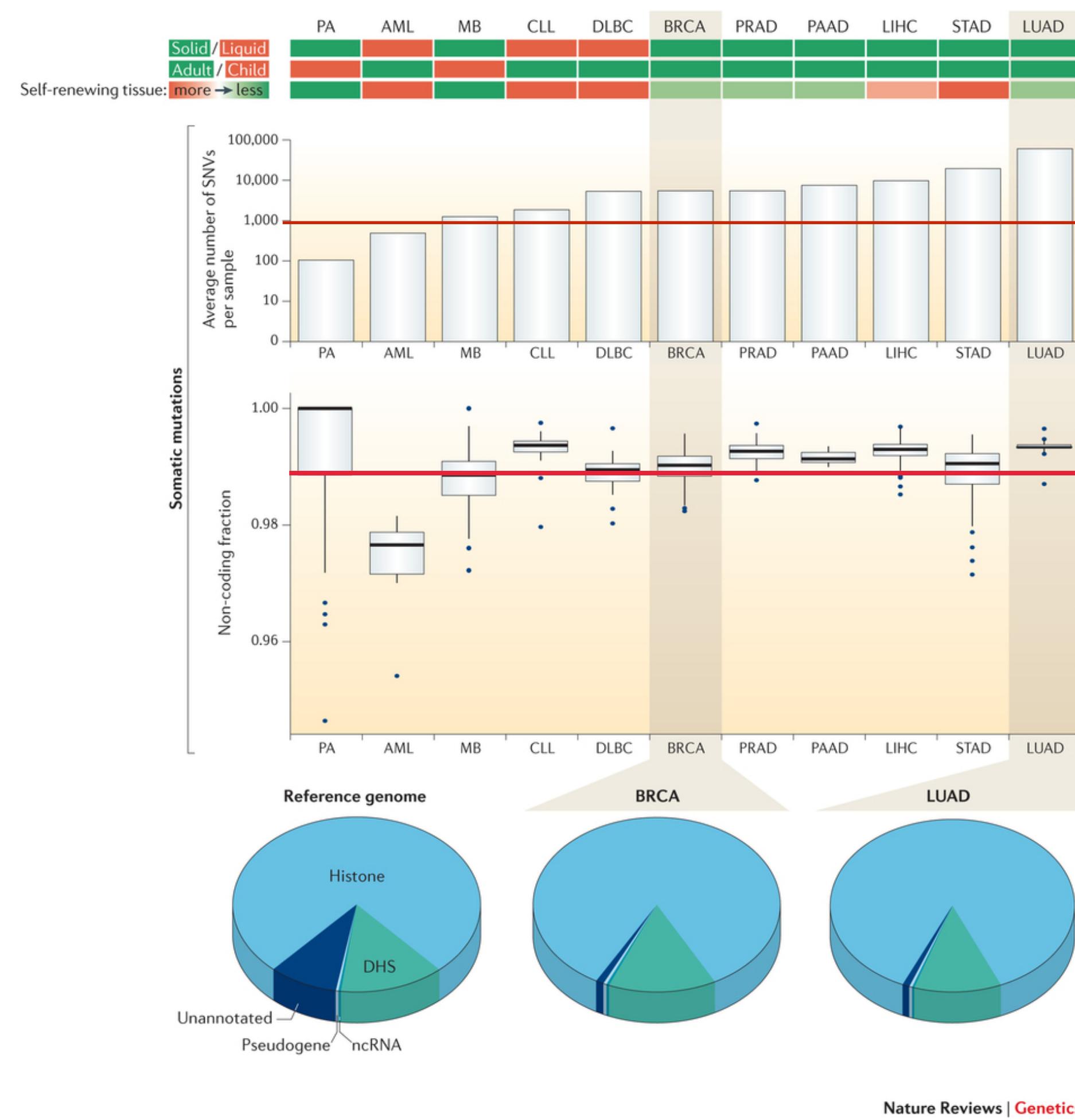
# Gene dosage phenomena beyond simple on/off effects



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.

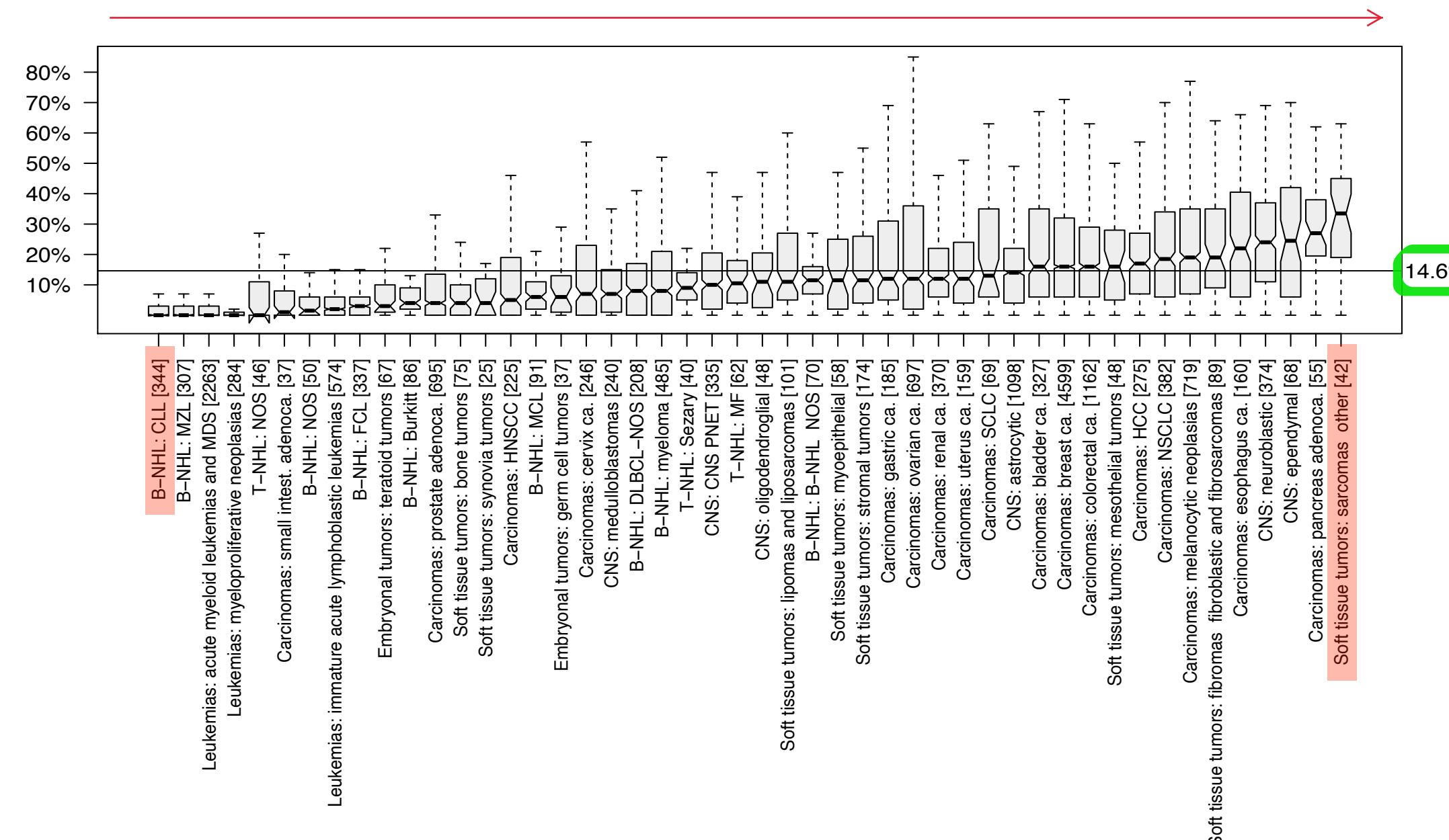




CANCERS SHOW THOUSANDS OF SINGLE NUCLEOTIDE VARIANTS PER SAMPLE, MOSTLY IN NON-CODING REGIONS

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))

# Quantifying Somatic Mutations In Cancer

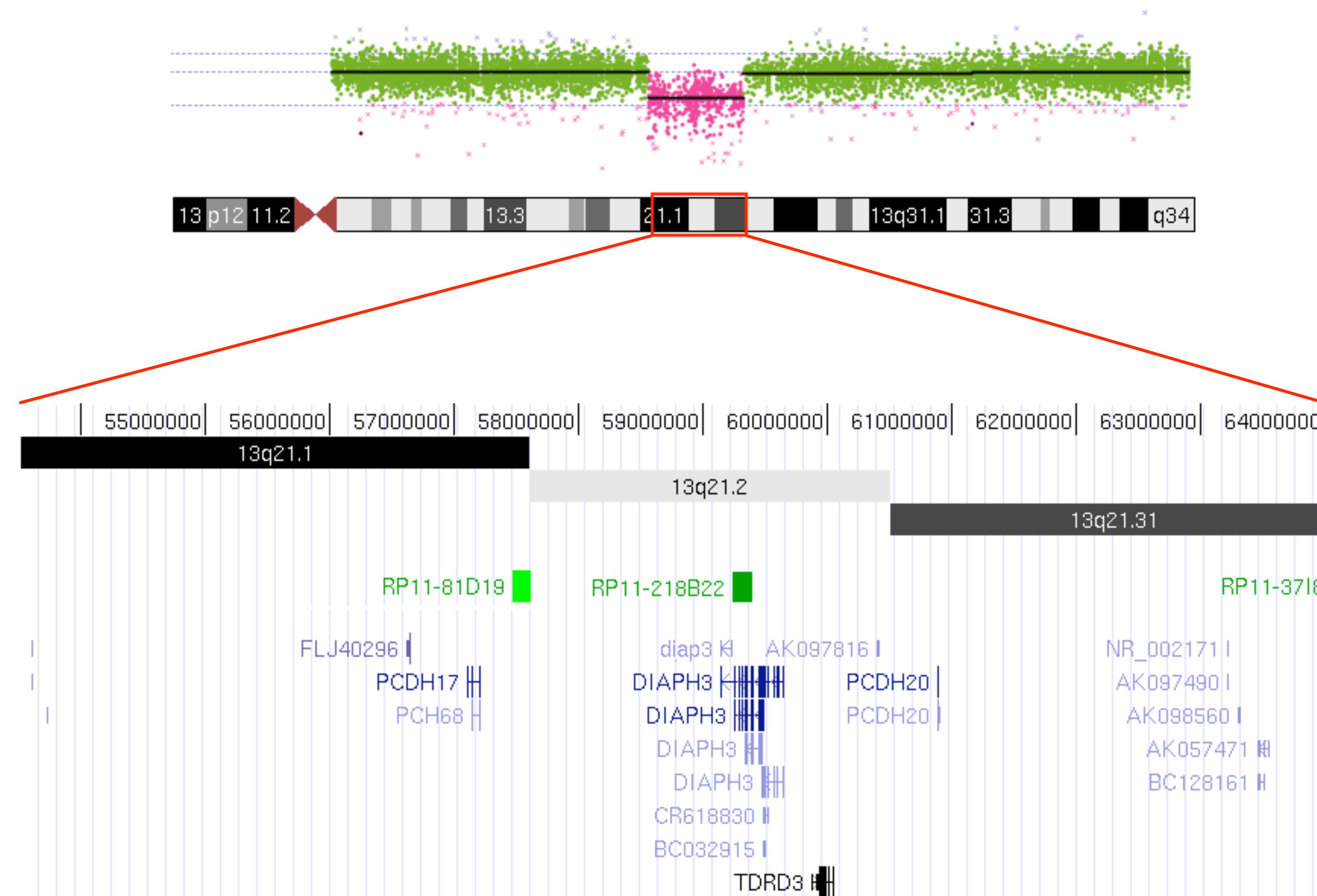


GENOMIC COPY NUMBER IMBALANCES PROVIDE WIDESPREAD SOMATIC VARIANTS IN CANCER

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

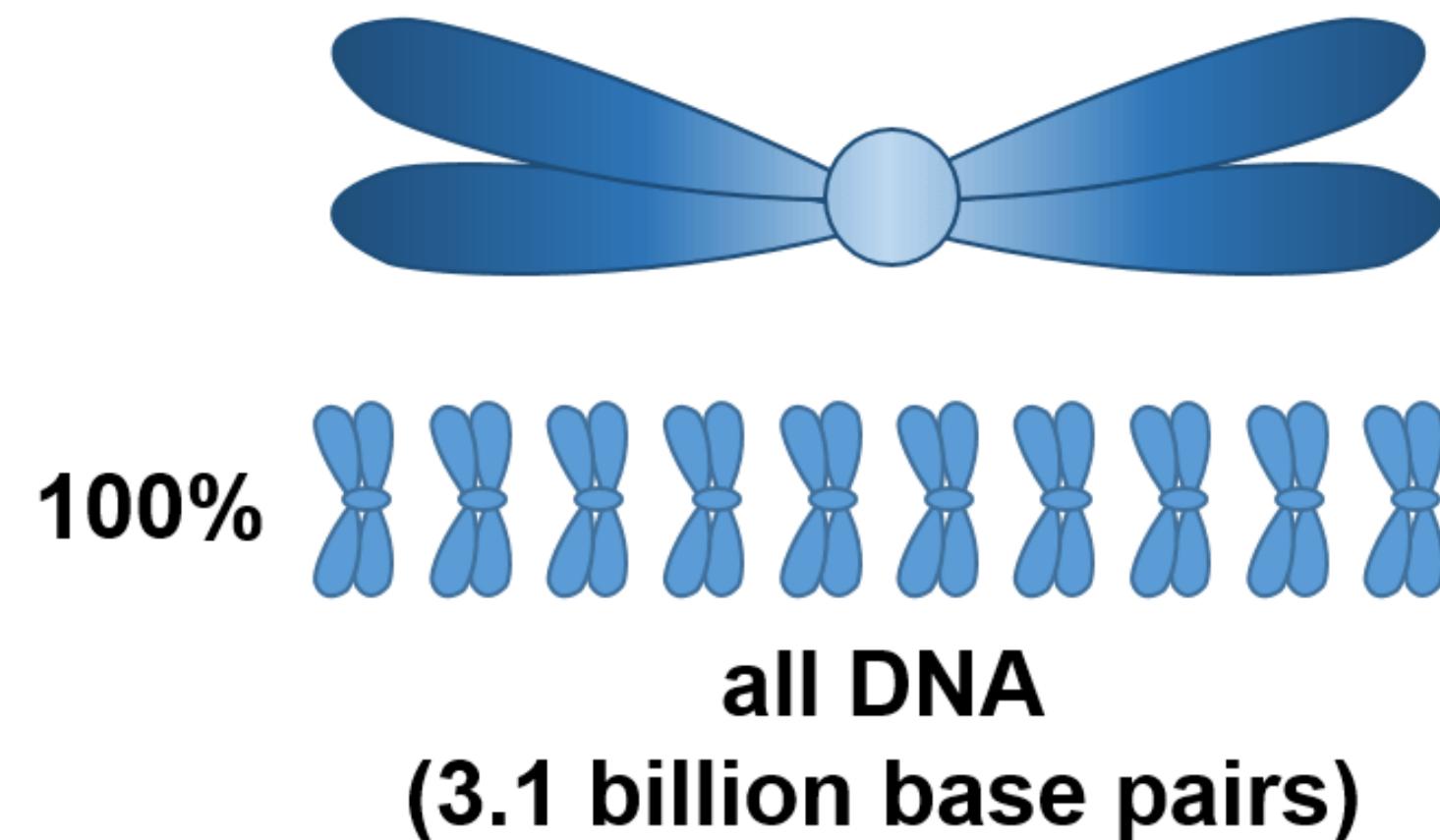
# Nobody is perfect (?)

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 Eggemann, and Herdit M. Schüler  
*American Journal of Medical Genetics Part A* 146A:2417 – 2420 (2008)

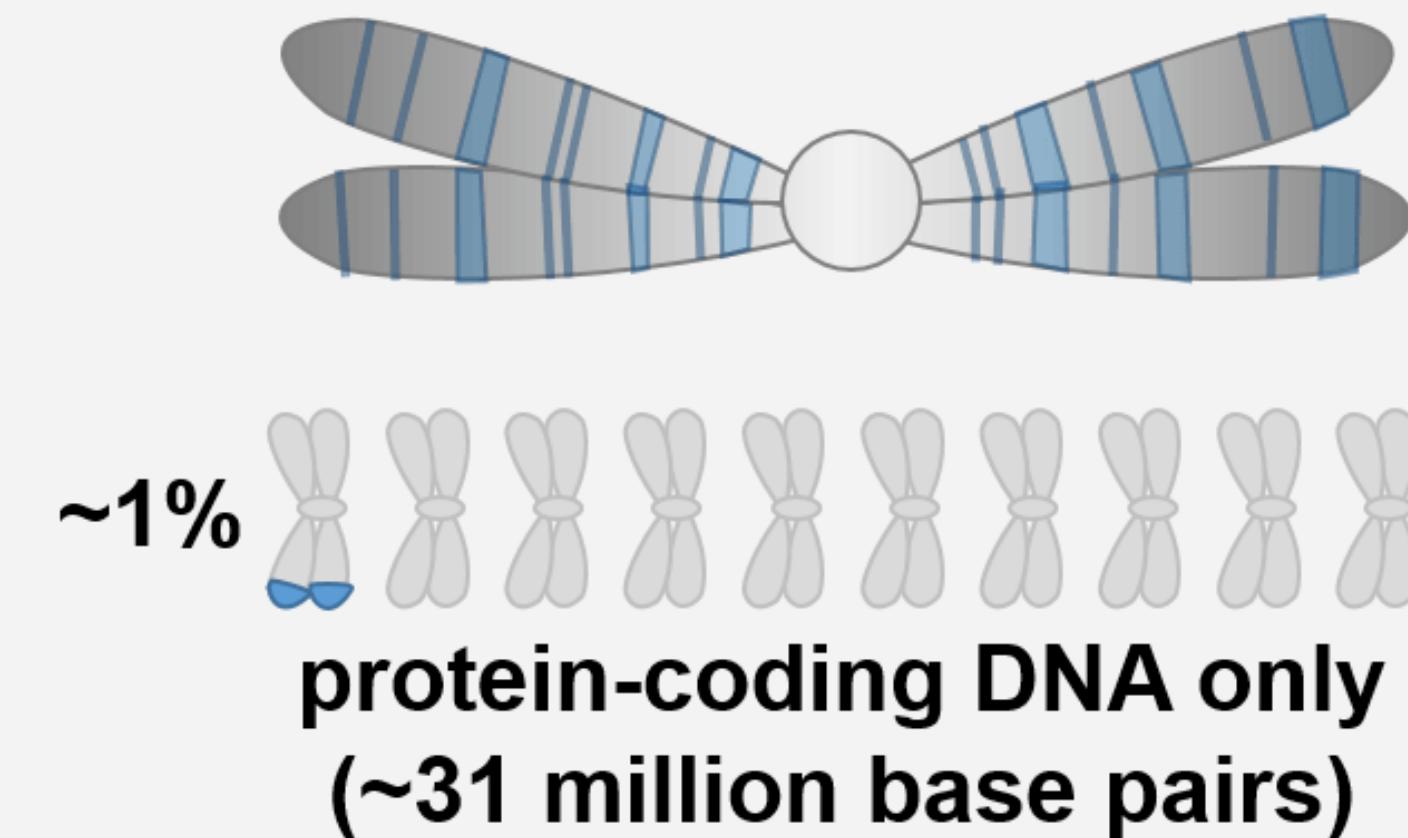


# Genome Sequencing

## whole genome sequencing (WGS)



## exome sequencing



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

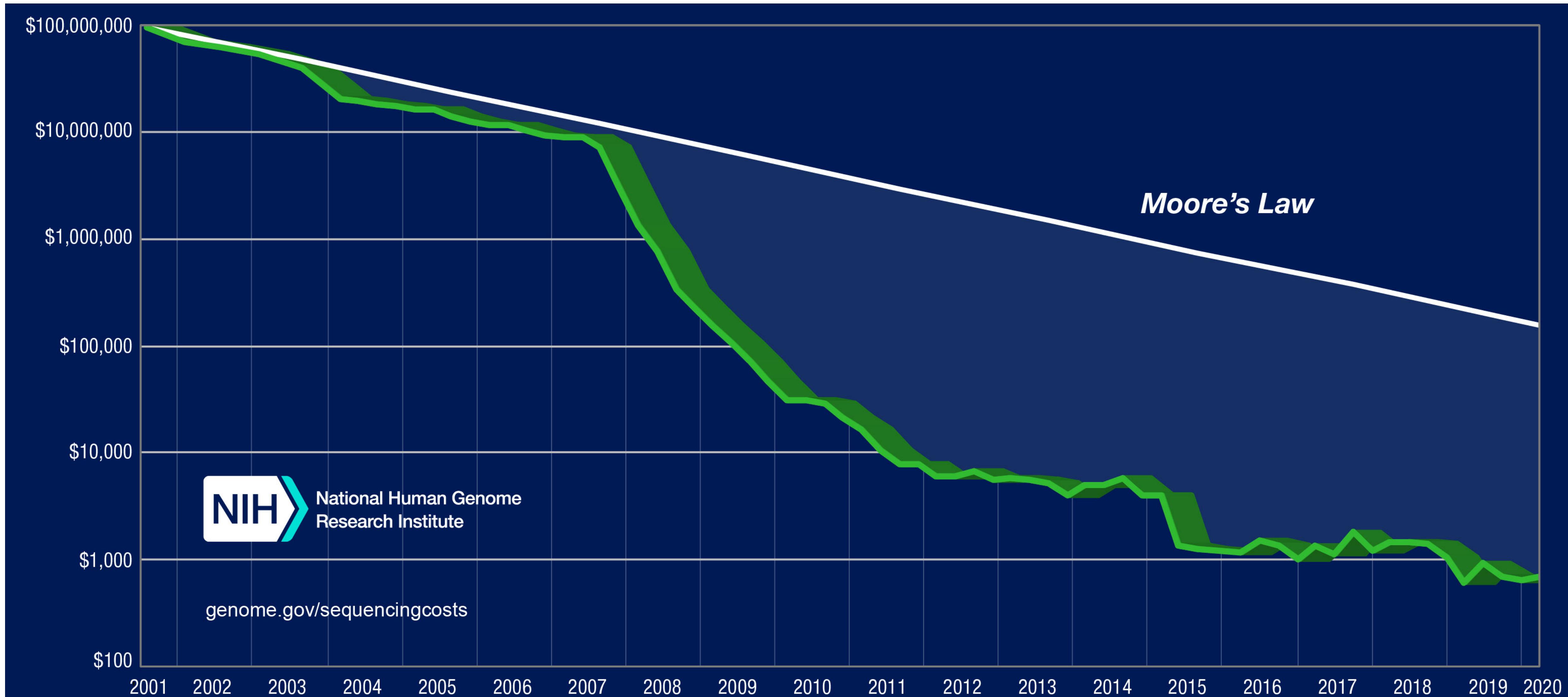
today:

2017

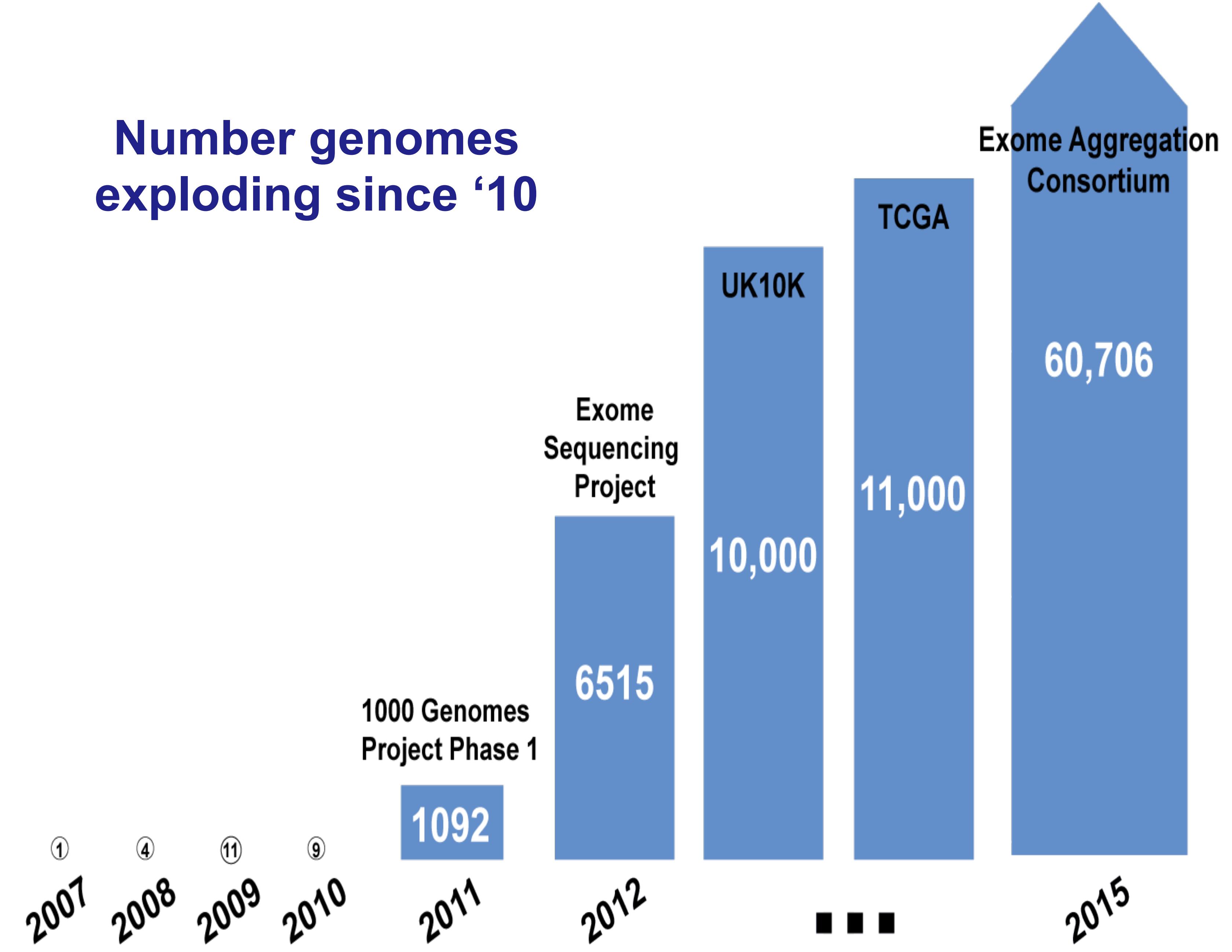
~\$530

~3 days

# The Cost of Sequencing a Human Genome



## Number genomes exploding since ‘10



# Genomes Everywhere

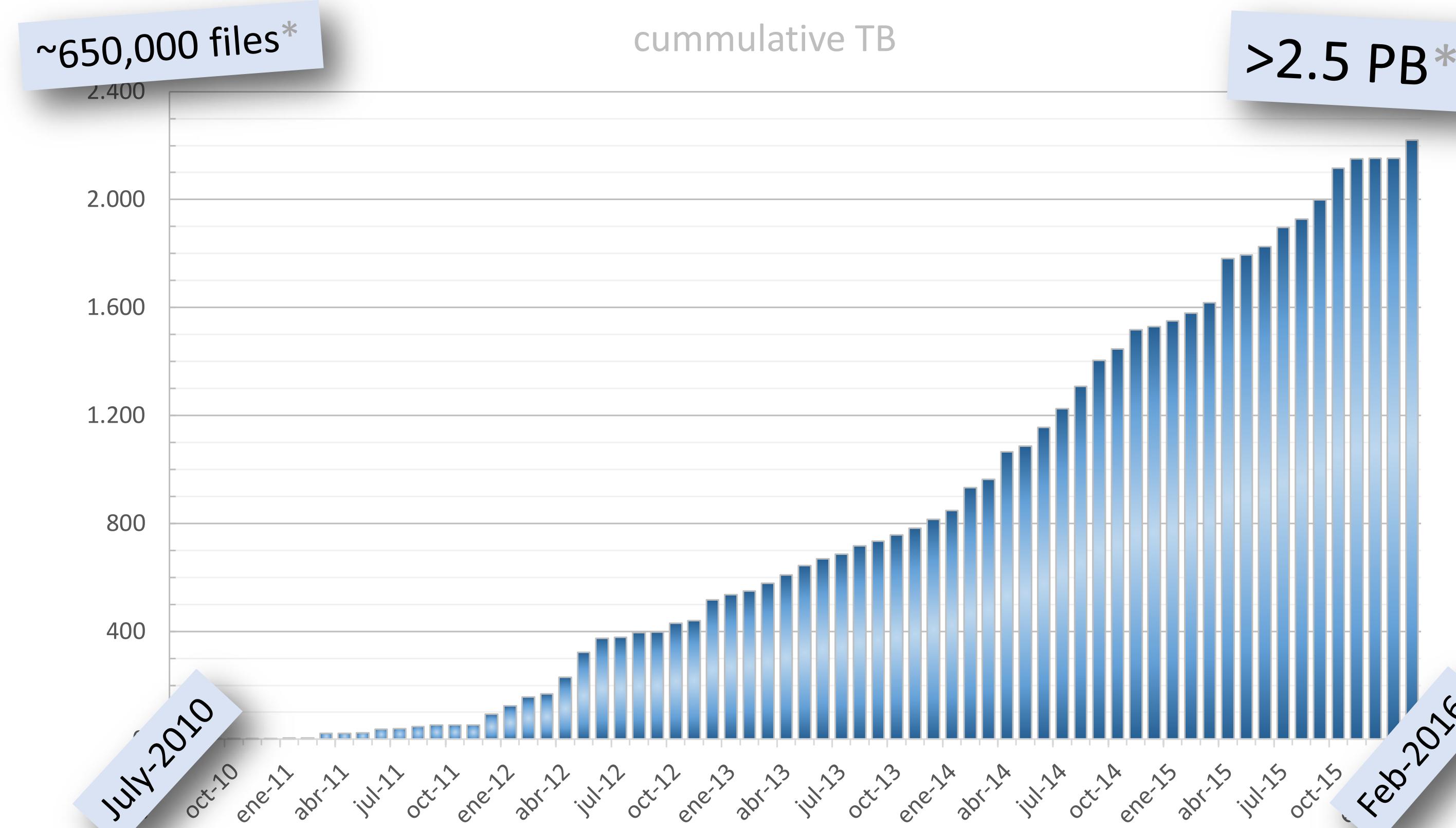
Organization / Initiative: Name	Organization / Initiative: Category	Cohort
100K Wellness Project	Research Project	107 unaffected individuals (scaling up to 100,000)
23andMe	Organization	>1 million customers (>80% consented to research)
Actionable Cancer Genome Initiative (ACGI)	Data-Sharing Project	Goal: 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

# National Medical Genome Projects and Cohorts (2018)



# Growth of Genome Data Repositories: Example EGA

The EGA contains a growing amount of data

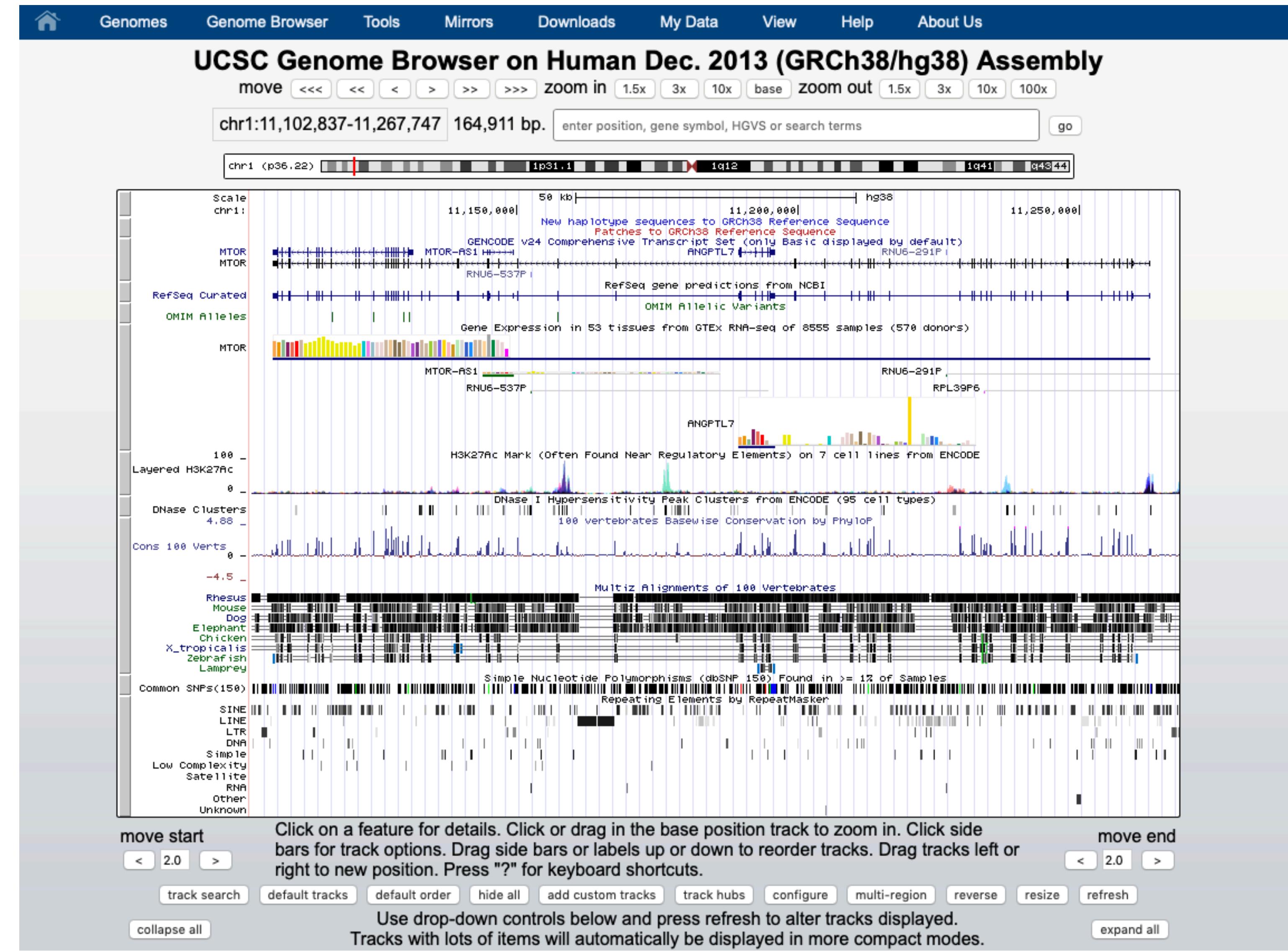


Reference Genome Resources... |

## 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](http://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/](http://www.ncbi.nlm.nih.gov/projects/genome/guide/human/)

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Log in

### Human Genome Resources at NCBI

Download Browse View Learn

Search for Human Genes  Search

Select a chromosome to access the [Genome Data Viewer](#)

▼

### Download

	GRCh38	GRCh37
Reference Genome Sequence	<a href="#">Fasta</a>	<a href="#">Fasta</a>
RefSeq Reference Genome Annotation	<a href="#">gff3</a>	<a href="#">gff3</a>
RefSeq Transcripts	<a href="#">Fasta</a>	<a href="#">Fasta</a>
RefSeq Proteins	<a href="#">Fasta</a>	<a href="#">Fasta</a>
ClinVar	<a href="#">vcf</a>	<a href="#">vcf</a>
dbSNP	<a href="#">vcf</a>	<a href="#">vcf</a>
dbVar	<a href="#">vcf</a>	<a href="#">vcf</a>

## RESOURCES FOR GENOMICS: ENSEMBL

- ▶ Entry point for many genome data services and collections

- ▶ Downloads ("BioMart"), REST API

[www.ensembl.org/  
Homo sapiens/Info/Index](http://www.ensembl.org/Homo_sapiens/Info/Index)

The screenshot shows the Ensembl Human GRCh38.p12 genome browser. At the top, there's a navigation bar with links to BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and a Blog. A search bar is also present. Below the navigation, it says "Human (GRCh38.p12) ▾". The main content area has several sections:

- Search Human (*Homo sapiens*)**: Includes a search bar for "Search all categories" and "Search Human...", a "Go" button, and a note about search terms like BRCA2 or rs1333049.
- Genome assembly: GRCh38.p12 (GCA\_000001405.27)**: Includes links for "More information and statistics", "Download DNA sequence (FASTA)", "Convert your data to GRCh38 coordinates", and "Display your data in Ensembl". It also shows a "View karyotype" icon and an "Example region" icon.
- Gene annotation**: Includes links for "More about this genebuild", "Download genes, cDNAs, ncRNA, proteins (FASTA)", and "Update your old Ensembl IDs". It also shows an "Example gene" icon (Pax6 INS FOXP2 BRCA2 DMD ssh).
- Comparative genomics**: Includes links for "More about comparative analysis" and "Download alignments (EMF)". It shows an "Example gene tree" icon.
- Variation**: Includes links for "More about variation in Ensembl", "Download all variants (GVF)", and "Variant Effect Predictor". It shows an "Example variant" icon (ATCGAGCT ATCCAGCT ATCGAGAT).
- Regulation**: Includes links for "More about the Ensembl regulatory build and microarray annotation", "Experimental data sources", and "Download all regulatory features (GFF)". It shows an "Example regulatory feature" icon (ENCODE data in Ensembl) and an "Example structural variant" icon (Ve!P).



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**

**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**

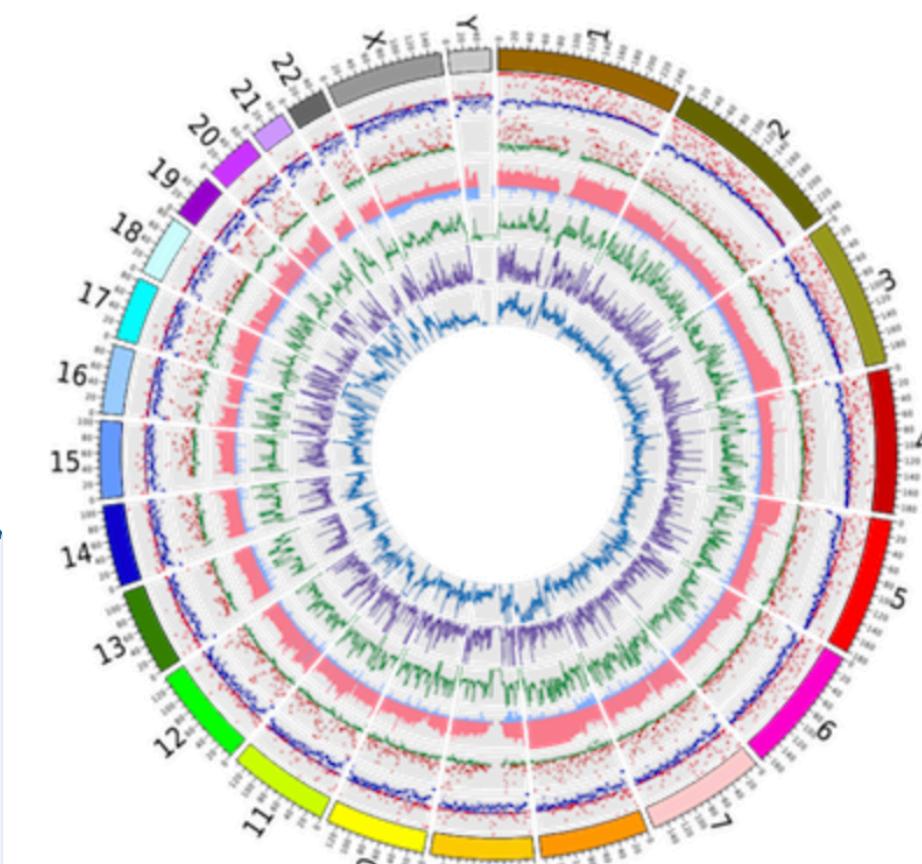
*High quality curation by expert postdoctoral scientists*

- 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, Mg<sup>2+</sup>/Mn<sup>2+</sup> 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

The screenshot shows the ClinGen Clinical Genome Resource website. At the top right is a search bar with the placeholder "Search our Knowledge Base for genes and diseases..." and a magnifying glass icon. Below the search bar are navigation links: About ClinGen, Working Groups, Resources, GenomeConnect, Share Your Data (highlighted in blue), and Curation Activities. The main banner features a blue background with a blurred image of laboratory glassware and a computer screen displaying genetic data. The text "Defining the clinical relevance of genes & variants for precision medicine and research..." is centered above three large numbers: 1496 (ClinGen Curated Genes), 31 (Expert Groups), and 10446 (Expert Reviewed Variants in ClinVar). To the right of these numbers is a magnifying glass icon labeled "Knowledge Base Search". Below the banner, the tagline "Sharing Data. Building Knowledge. Improving Care." is displayed, followed by a description of ClinGen's mission. Six call-to-action boxes are arranged in a grid at the bottom:

- ClinGen-ClinVar Partnership (Icon: DNA helix inside a circle)
- How to share genomic & health data (Icon: DNA helix inside a circular arrow)
- Learn about ClinGen curation activities (Icon: Computer monitor with DNA helix)
- GenomeConnect Patient Registry (Icon: Three DNA helices)
- View ClinGen's Resources & Tools (Icon: Computer monitor with multiple windows)
- Get Involved (Icon: Computer keyboard, mouse, and notepad)

[clinicalgenome.org](http://clinicalgenome.org)

# The ClinGen and ClinVar Partnership

Both provide resources to support genomic interpretation

- ▶ 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")

### ClinGen - A Program

An NIH funded project

Building a central resource that defines the clinical relevance of genes and variants

ClinGen is addressing the following critical questions:

- Is the gene associated with disease?
- Is the variant pathogenic?
- Is the variant/gene information actionable?

Encouraging data sharing

- Promote lab submissions to ClinVar
- Facilitate patient data sharing through GenomeConnect



Assessing the clinical **validity** and **actionability** of genes and their relationship to diseases

### ClinVar- A Database

Funded by intramural NIH funding

Freely accessible and downloadable public archive of reports of the relationship between variants and conditions

Maintained by the National Center for Biotechnology Information (NCBI)



Maintaining a publicly available **database** of:

- Interpretations of the clinical significance of variants
- Submitter information
- Supporting evidence and individual level data, when available

[clinicalgenome.org](http://clinicalgenome.org)

ClinGen

Find out more online...

ClinVar

# 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**

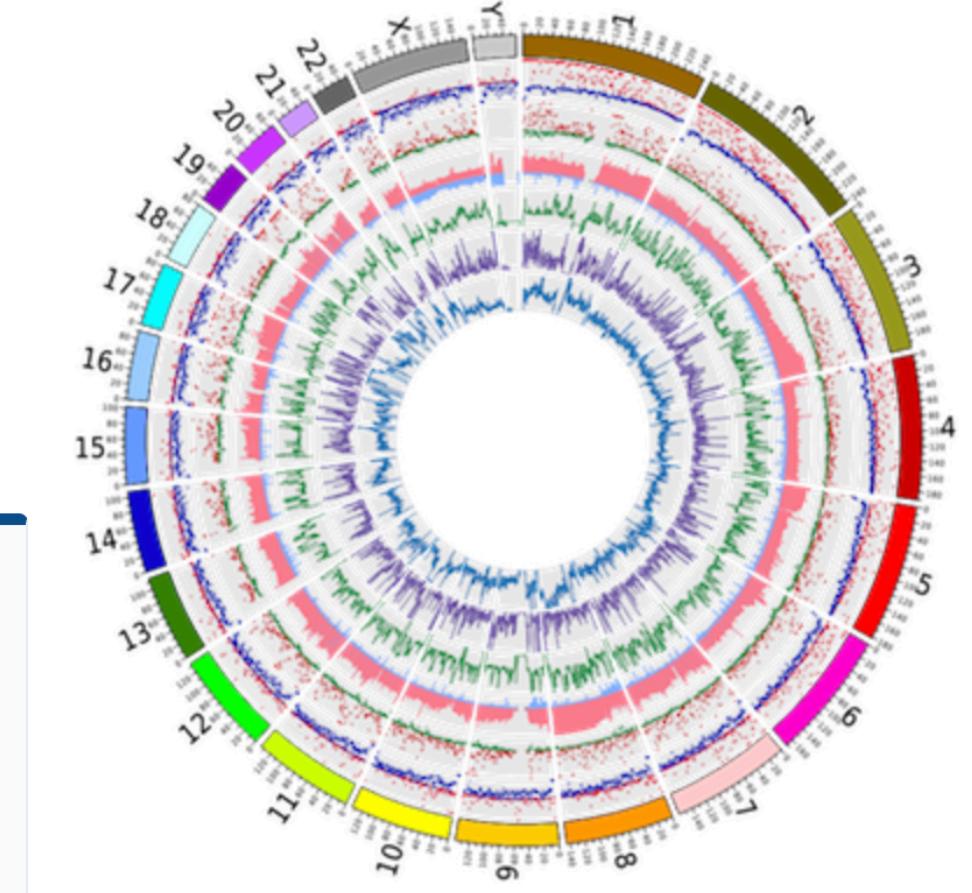
*High quality curation by expert postdoctoral scientists*

- Drug Resistance 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



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, Mg<sup>2+</sup>/Mn<sup>2+</sup> 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

National Cancer Institute U.S. National Institutes of Health | www.cancer.gov

## CANCER GENOME ANATOMY PROJECT

**CGAP How To**

**Tools**

**CGAP Info**

- Educational Resources
- Slide Tour
- Team Members
- References

**CGAP Data**

**Quick Links:**

- ICG
- NCI Home
- NCICB Home
- NCBI Home
- OCG

**Genes** **Chromosomes** **Tissues** **SAGE Genie** **RNAi** **Pathways**

### Cancer Genome Anatomy Project (CGAP)

The NCI's Cancer Genome Anatomy Project sought to determine the gene expression profiles of normal, precancer, and cancer cells, leading eventually to improved detection, diagnosis, and treatment for the patient. Resources generated by the CGAP initiative are available to the broad cancer community. Interconnected modules provide access to all CGAP data, bioinformatic analysis tools, and biological resources allowing the user to find "in silico" answers to biological questions in a fraction of the time it once took in the laboratory.

### The CGAP Website

Interconnected modules provide access to all CGAP data, bioinformatic analysis tools, and biological resources allowing the user to find "in silico" answers to biological questions in a fraction of the time it once took in the laboratory.

**Genes** Gene information, clone resources, SNP500Cancer, GAI, and transcriptome analysis.

**Chromosomes** FISH-mapped BAC clones, SNP500Cancer, and the Mitelman database of chromosome aberrations.

**Tissues** cDNA library information, methods, and EST-based gene expression analysis.

**Pathways** Diagrams of biological pathways and protein complexes, with links to genetic resources for each known protein.

**RNAi** RNA-interference constructs, targeted specifically against cancer relevant genes. New addition: Validated set of shRNAs.

International Cancer Genome Consortium

Home Cancer Genome Projects Committees and Working Groups Policies and Guidelines Media

### ICGC Cancer Genome Projects

Committed projects to date: 89

Sort by: Project

Biliary Tract Cancer Japan	Biliary Tract Cancer Singapore	Bladder Cancer China
Bladder Cancer United States	Blood Cancer China	Blood Cancer Singapore
Blood Cancer South Korea	Blood Cancer United States	Blood Cancer United States
Blood Cancer United States	Blood Cancer United States	Bone Cancer France
Bone Cancer United Kingdom	Bone Cancer United States	Brain Cancer Canada
Brain Cancer China	Brain Cancer United States	Brain Cancer United States
Breast Cancer China	Breast Cancer European Union / United Kingdom	Breast Cancer France
Breast Cancer Mexico	Breast Cancer South Korea	Breast Cancer South Korea

**ICGC Goal:** To obtain a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types and/or subtypes which are of clinical and societal importance across the globe.

[Read more »](#)

**Launch Data Portal »**

**Apply for Access to Controlled Data »**

**Announcements**

**23/August/2016** - The ICGC Data Coordination Center (DCC) is pleased to announce ICGC data portal data release 22 (<http://dcc.icgc.org>).

ICGC data release 22 in total comprises data from more than 16,000 cancer donors spanning 70 projects and 21 tumour sites.

**17/April/2016** - ICGCmed is pleased to announce the release of its white paper (<http://icgcmed.org>).

The International Cancer Genome Consortium for Medicine (ICGCmed) will link genomics data to clinical information, health and responses to therapies.

**18/November/2015** - The International Cancer Genome Consortium (ICGC) PanCancer dataset generated by the PanCancer Analysis of Whole Genomes (PCAWG) study is now available on Amazon Web Services (AWS), giving cancer researchers access to over 2,400 consistently analyzed genomes corresponding to over 1,100 unique ICGC donors (<https://icgc.org/icgc-in-the-cloud>).



PMKB

Genes Variants

Interpretations Tumor Types

Primary Sites

Activity

Login

## About

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



Weill Cornell Medicine



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

Search...



610

Genes

2246

Variants

1767

Interpretations

156

Tumor Types

72

Primary Sites

263

Tier 1

Interpretations

1408

Tier 2

Interpretations

96

Tier 3

Interpretations

## Academic Partners



San Raffaele University



University of California San Francisco



University of Bern

## Non-Academic Partners



Microsoft



## 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 ...

NCBI Resources How To Sign in to NCBI

We are sorry, but the page you requested is no longer available.

NCBI's SKY-CGH site has been retired.

The public data from this resource can be downloaded from our [FTP server](#) and will soon be available in the [dbVar database \(SKY-CGH\)](#).

You are here: NCBI > National Center for Biotechnology Information Write to the Help Desk

Skip Navigation

GETTING STARTED RESOURCES POPULAR FEATURED NCBI INFORMATION

NCBI Education	Chemicals & Bioassays	PubMed	Genetic Testing Registry	About NCBI
NCBI Help Manual	Data & Software	Bookshelf	PubMed Health	Research at NCBI
NCBI Handbook	DNA & RNA	PubMed Central	GenBank	NCBI News
Training & Tutorials	Domains & Structures	PubMed Health	Reference Sequences	NCBI FTP Site
Submit Data	Genes & Expression	BLAST	Gene Expression Omnibus	NCBI on Facebook
	Genetics & Medicine	Nucleotide	Map Viewer	NCBI on Twitter
	Genomes & Maps	Genome	Human Genome	NCBI on YouTube
	Homology	SNP	Mouse Genome	
	Literature	Gene	Influenza Virus	
	Proteins	Protein	Primer-BLAST	
	Sequence Analysis	PubChem	Sequence Read Archive	
	Taxonomy			
	Variation			

Cancer Genome Anatomy Project (CGAP)

The NCI's [Cancer Genome Anatomy Project](#) sought to determine the gene expression profiles of normal, precancer, and cancer cells for diagnosis, and treatment for the patient. Resources generated by the CGAP initiative are available to the broad cancer community. Data, bioinformatic analysis tools, and biological resources allowing the user to find "in silico" answers to biological questions in a timely manner.

[Read more about CGAP](#) and access the many valuable resources.

Cancer Genome Characterization Initiative (CGCI)

The [Cancer Genome Characterization \(CGC\) Initiative](#): Assessing the use of new genomics technologies to strategically characterize tumors. Groups involved with the CGCI Initiative make all of their data available through a publicly accessible database. Cancer CGCI incorporates genomic characterization methods including exome and transcriptome analysis using second generation sequencing to identify genetic changes leading to cancer.

[Read more about the CGC Initiative](#) and how the project is enabling the next generation of discovery through rapid data release and analysis.

Download Plugin: [Windows](#) [Mac OS X](#) [Linux](#)

National Center for Biotechnology Information, U.S. National Library of Medicine  
8600 Rockville Pike, Bethesda MD, 20894 USA  
[Policies and Guidelines](#) | [Contact](#)

NATIONAL LIBRARY OF MEDICINE NATIONAL INSTITUTES OF HEALTH USA.gov

A Service of the National Cancer Institute

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

# arrayMap

## Reference resource for copy number variation data in cancer



Search Samples

Search Publications

Progenetix



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:

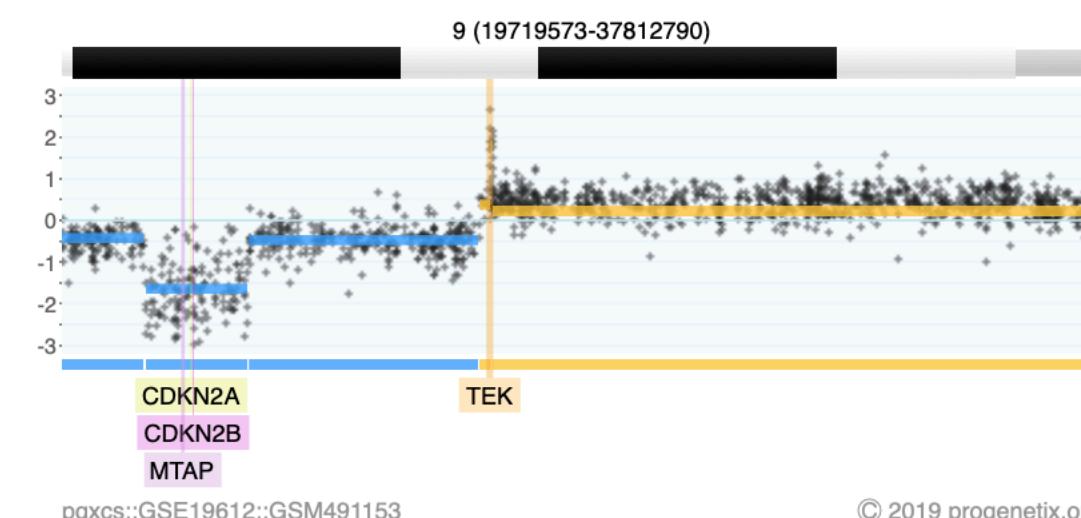
72724 genomic array profiles

898 experimental series

257 array platforms

341 ICD-O cancer entities

795 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 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. [\[PubMed\]](#)

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

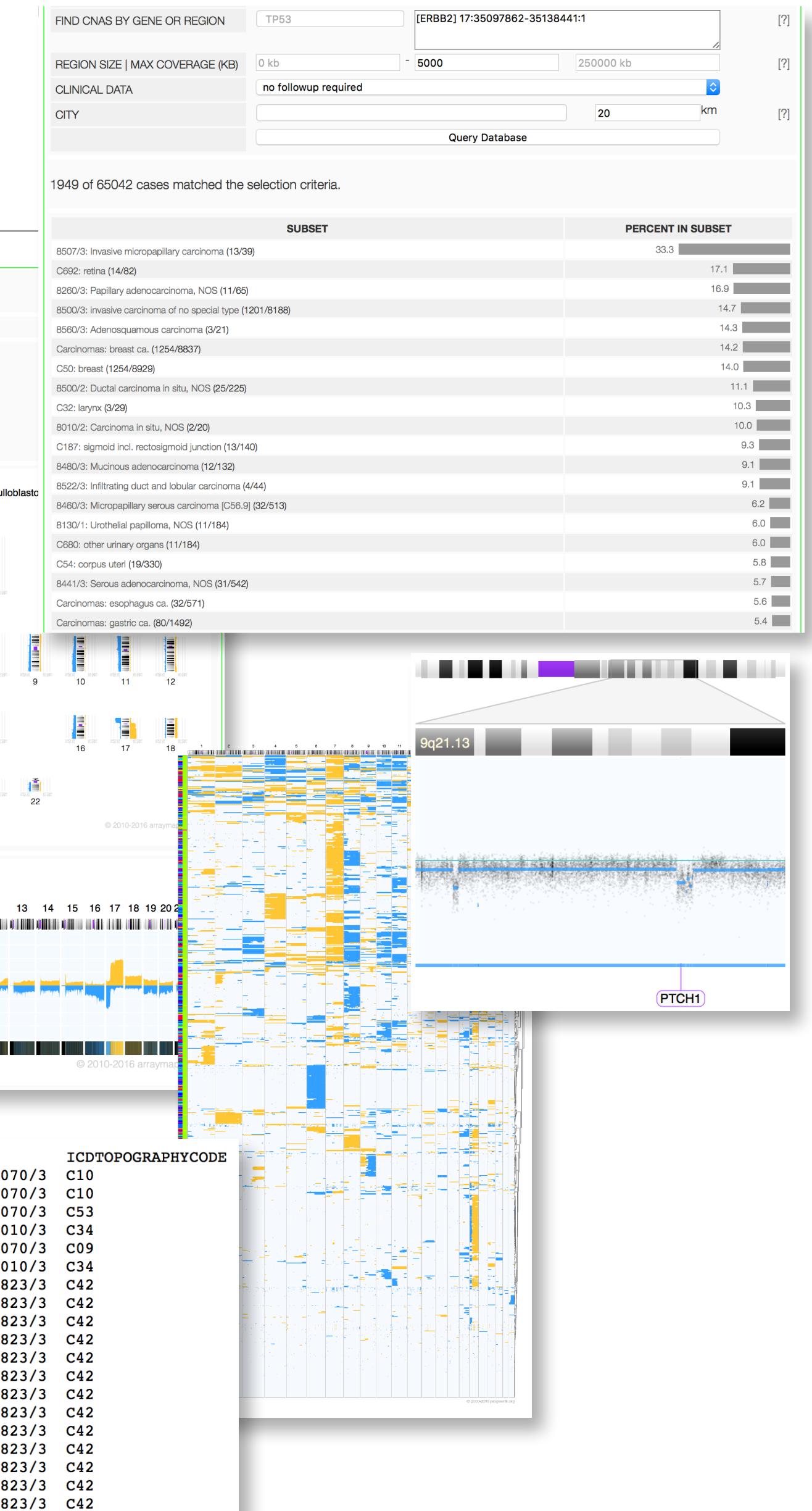
Baudis, M. 2007. Genomic imbalances in 5918 malignant epithelial tumors: An explorative meta-analysis of chromosomal CGH data. *BMC Cancer* 7:226. [\[PubMed\]](#)

Baudis, M. 2006. Online database and bioinformatics toolbox to support data mining in cancer cytogenetics. *Biotechniques* 40, no. 3: 296-272. [\[PubMed\]](#)

Baudis, M, and ML Cleary. 2001. Progenetix.net: an online repository for molecular cytogenetic aberration data. *Bioinformatics* 12, no. 17: 1228-1229. [\[PubMed\]](#)

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.

© 2000 - 2019 Michael Baudis, refreshed 2019-06-12T21:00:19Z in 6.00s 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.



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 :: 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 current 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

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 spectra 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

