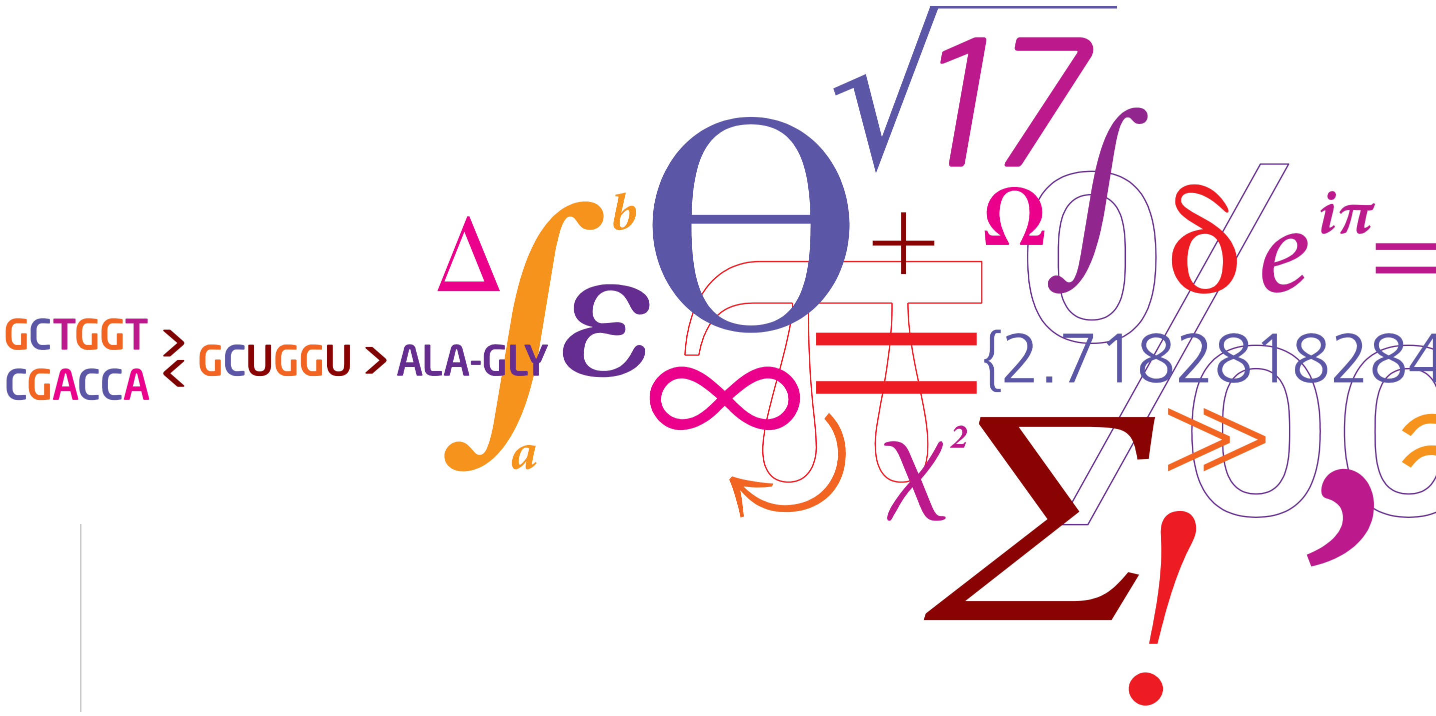
Technical University
of Denmark

DTU

Understanding cancer genomics

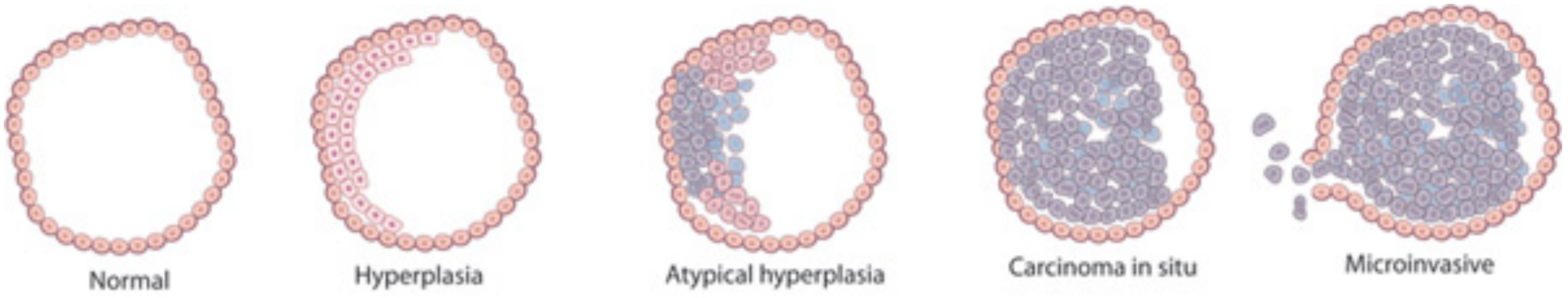
Adrian Otamendi Laspiur, Research Assistant (iCOPE)
Original slides: Jose MG Izarzugaza



DTU Bioinformatics
Department of Bio and Health Informatics

What is cancer?

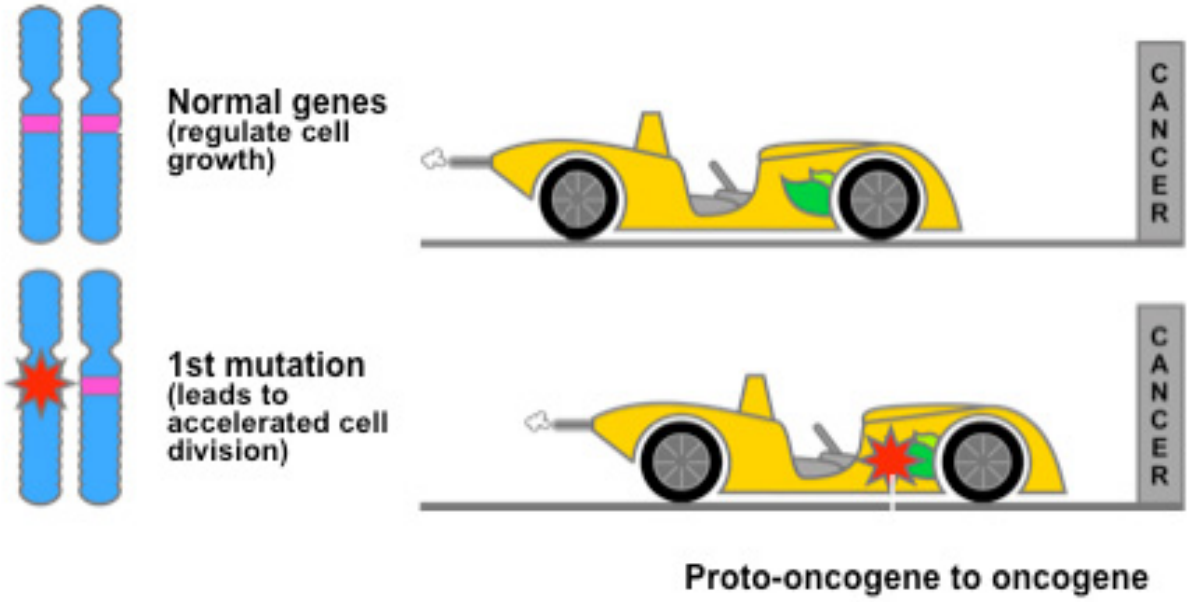
The disease caused by an uncontrolled division of *abnormal* cells in a part of the body



http://www.nature.com/scitable/topicpage/cell-division-and-cancer-14046590

Need for speed: Oncogenes vs Tumour suppressors

Oncogenes:
- Mutated proto-oncogenes
- Turn abnormal cell growth on
- 70 protooncogenes
- gain of function genes
- primarily somatic activated
- [throttle pedal in a car]



"Oncogenes are mutated genes whose PRESENCE can stimulate the development of cancer"

Examples: HER-2/neu. RAS, MYC, SRC, hTERT

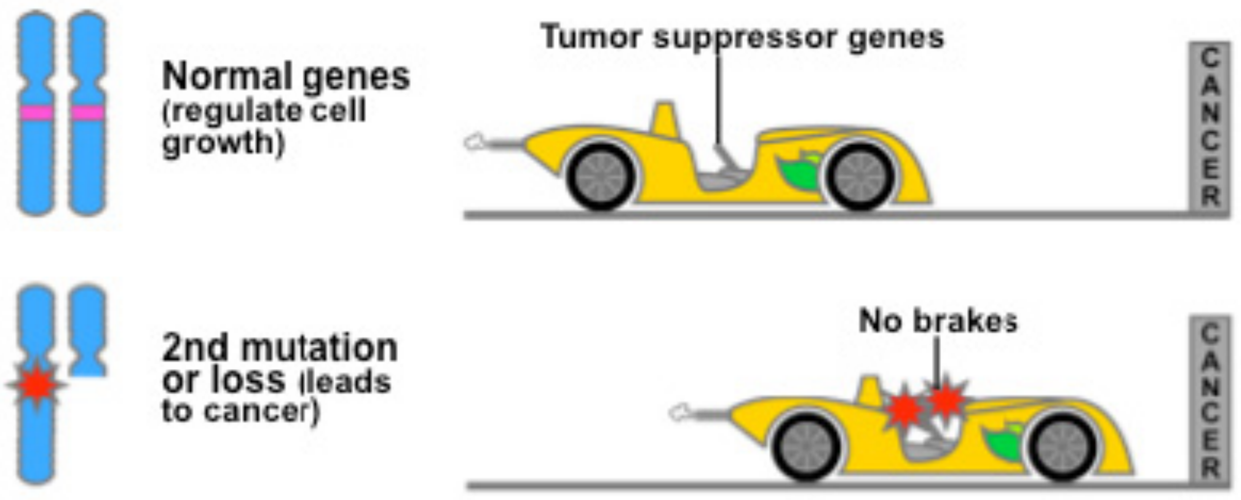
RAS, MYC, SRC are protein kinases ➔ Cell cycle regulation

(Adapted from Nat. Cancer Inst.)

Need for speed: Oncogenes vs Tumour suppressors

Tumour suppressor genes:
- Stop the cell cycle, G1 phase
- Slow the cell cycle before S phase
- Can induce apoptosis

- primarily somatic de-activated
- loss-of-function mutations
- [brake pedal in a car]



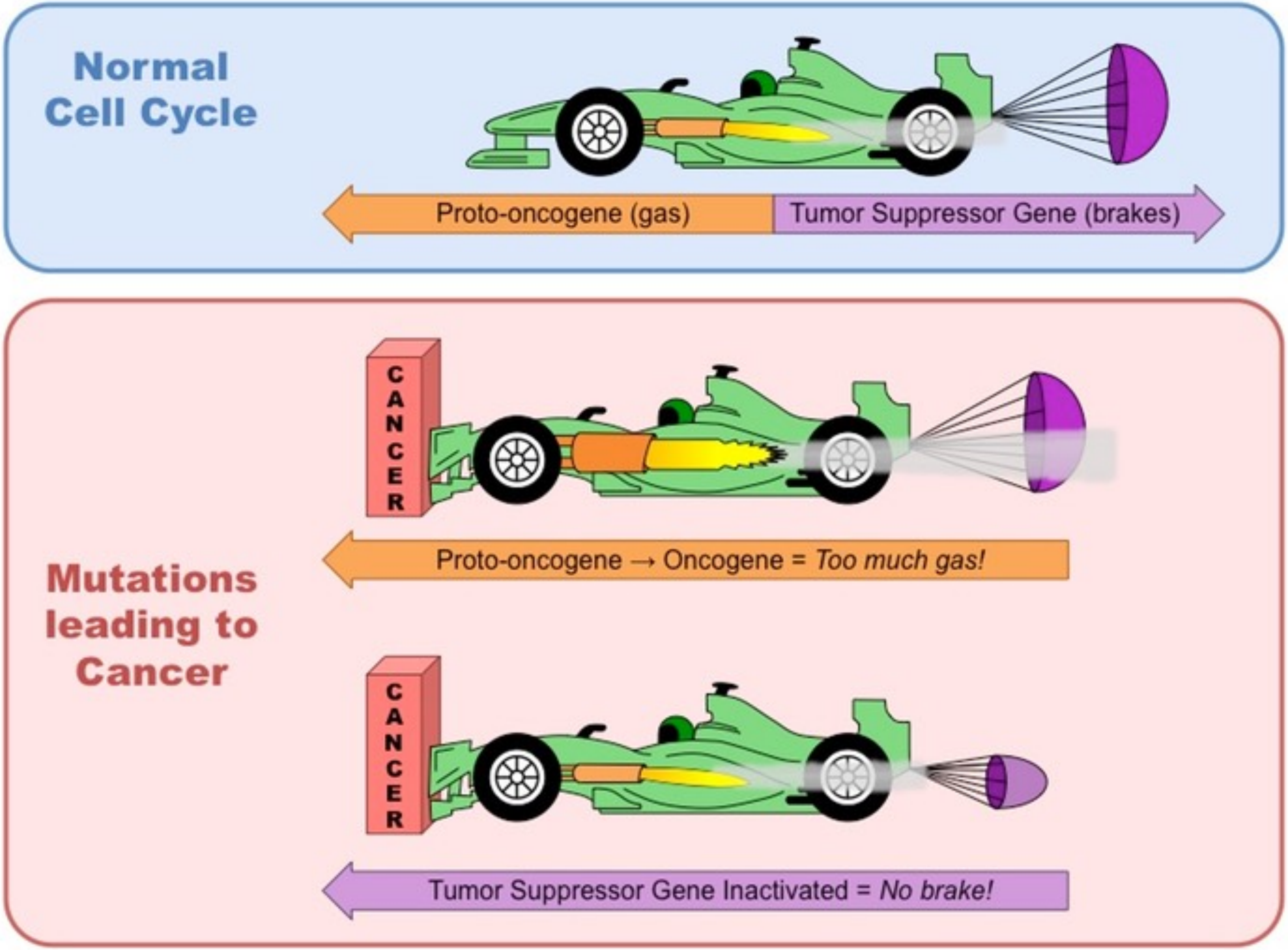
"Tumour suppressors are normal genes whose ABSENCE can stimulate the development of cancer"

Examples: p53, Rb, APC

Sometimes, a single functional copy (heterozygous) is enough to prevent cancer

(Adapted from Nat. Cancer Inst.)

Need for speed: Oncogenes vs Tumour suppressors


Normal Cell Cycle
Proto-oncogene (gas)
Tumor Suppressor Gene (brakes)
Mutations leading to Cancer
CANCER
Proto-oncogene → Oncogene = *Too much gas!*
CANCER
Tumor Suppressor Gene Inactivated = *No brake!*

(ib.bioninja.com.au)

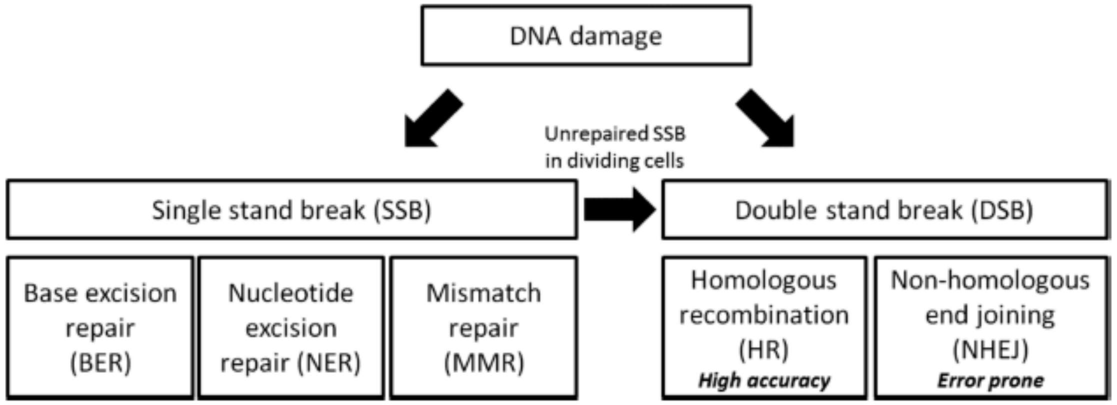
Need for speed: DNA repair genes

DNA damage repair genes
- Correct damage during DNA duplication
- Active in cell cycle, primarily G2
- After DNA replication, before Chr divides

- loss-of-function mutations ➔ increased mutation burden

Examples: BRCA1 and BRCA2 in breast cancer

Also, mDDRG in hereditary colon cancer



DNA damage

Unrepaired SSB in dividing cells

Single stand break (SSB)

Double stand break (DSB)

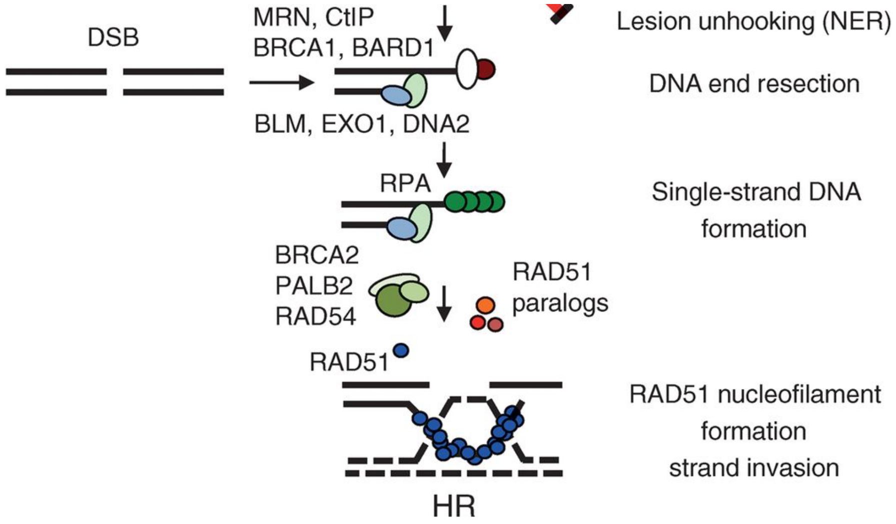
Base excision repair (BER)

Nucleotide excision repair (NER)

Mismatch repair (MMR)

Homologous recombination (HR) *High accuracy*

Non-homologous end joining (NHEJ) *Error prone*

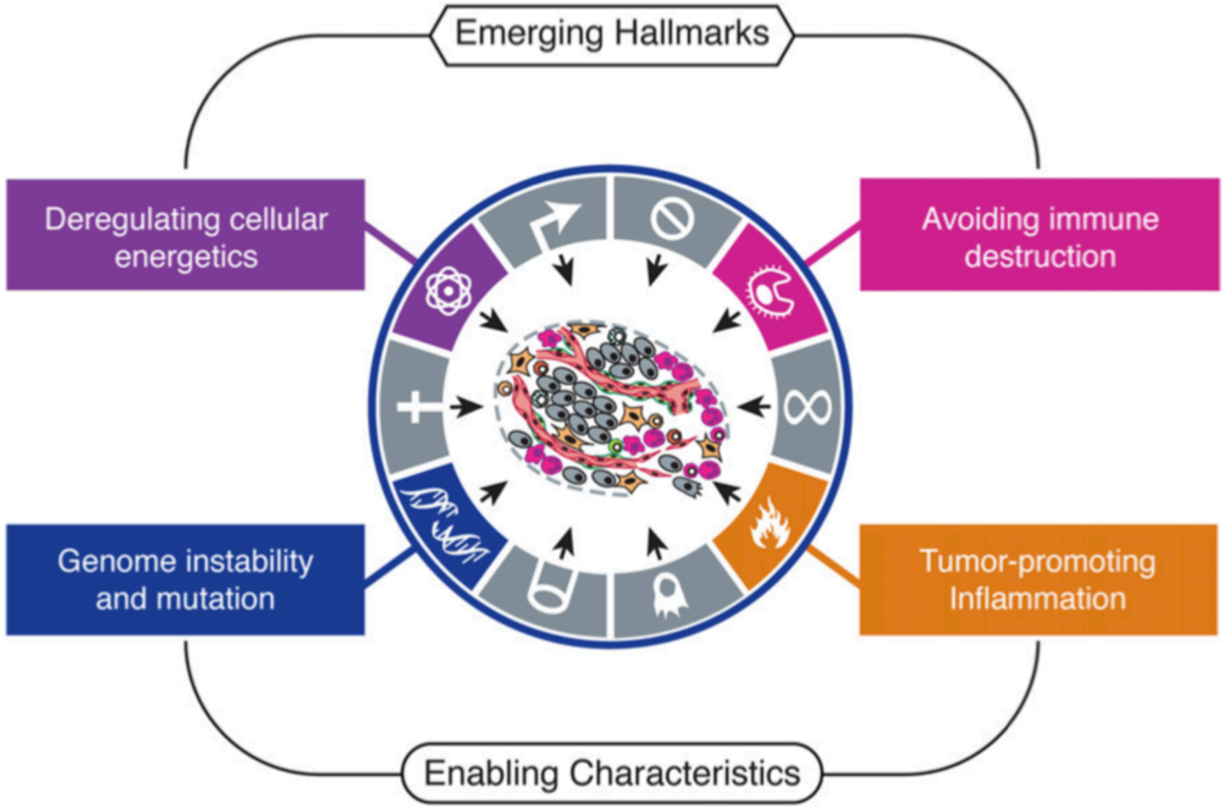


(Adapted from Nat. Cancer Inst.)

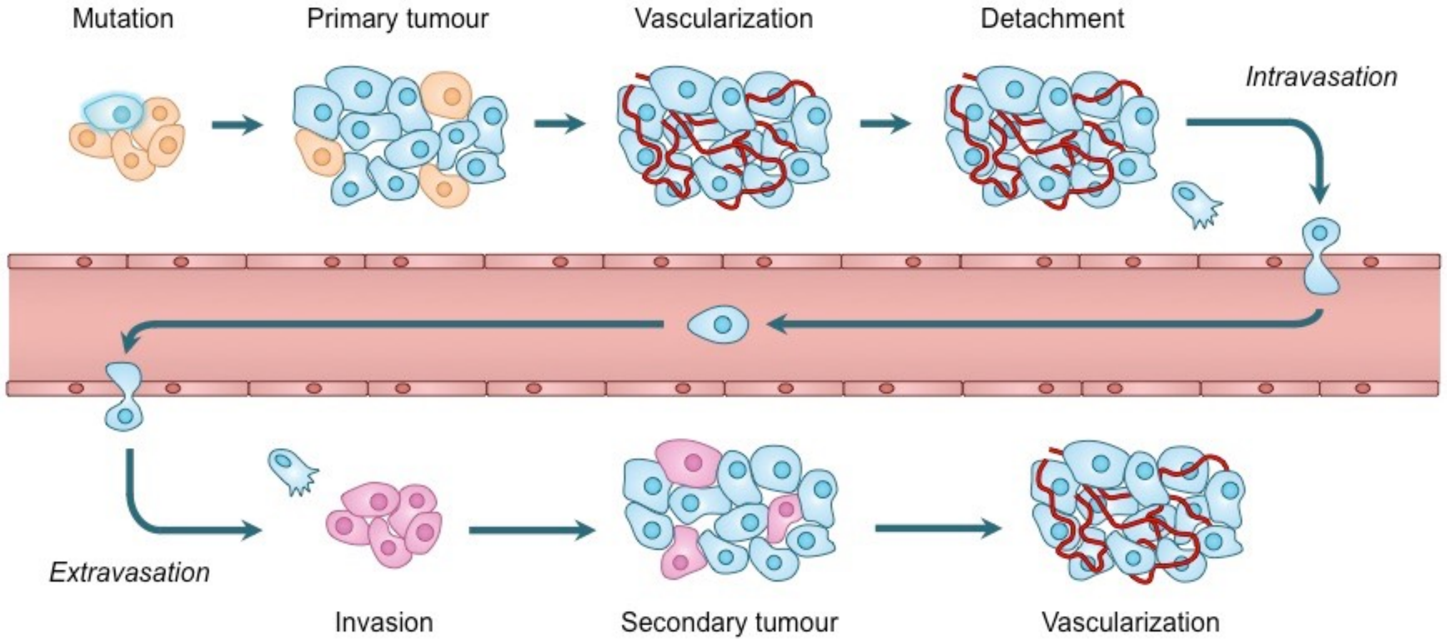
The hallmarks of cancer

Acquire functional capabilities

- Sustaining proliferative signaling
- Evading growth suppressors
- Resisting cell death
- Enabling replicative immortality
- Inducing angiogenesis
- Activating invasion and metastasis
- Emerging Hallmarks
- Enabling characteristics


Emerging Hallmarks
Deregulating cellular energetics
Avoiding immune destruction
Genome instability and mutation
Tumor-promoting Inflammation
Enabling Characteristics

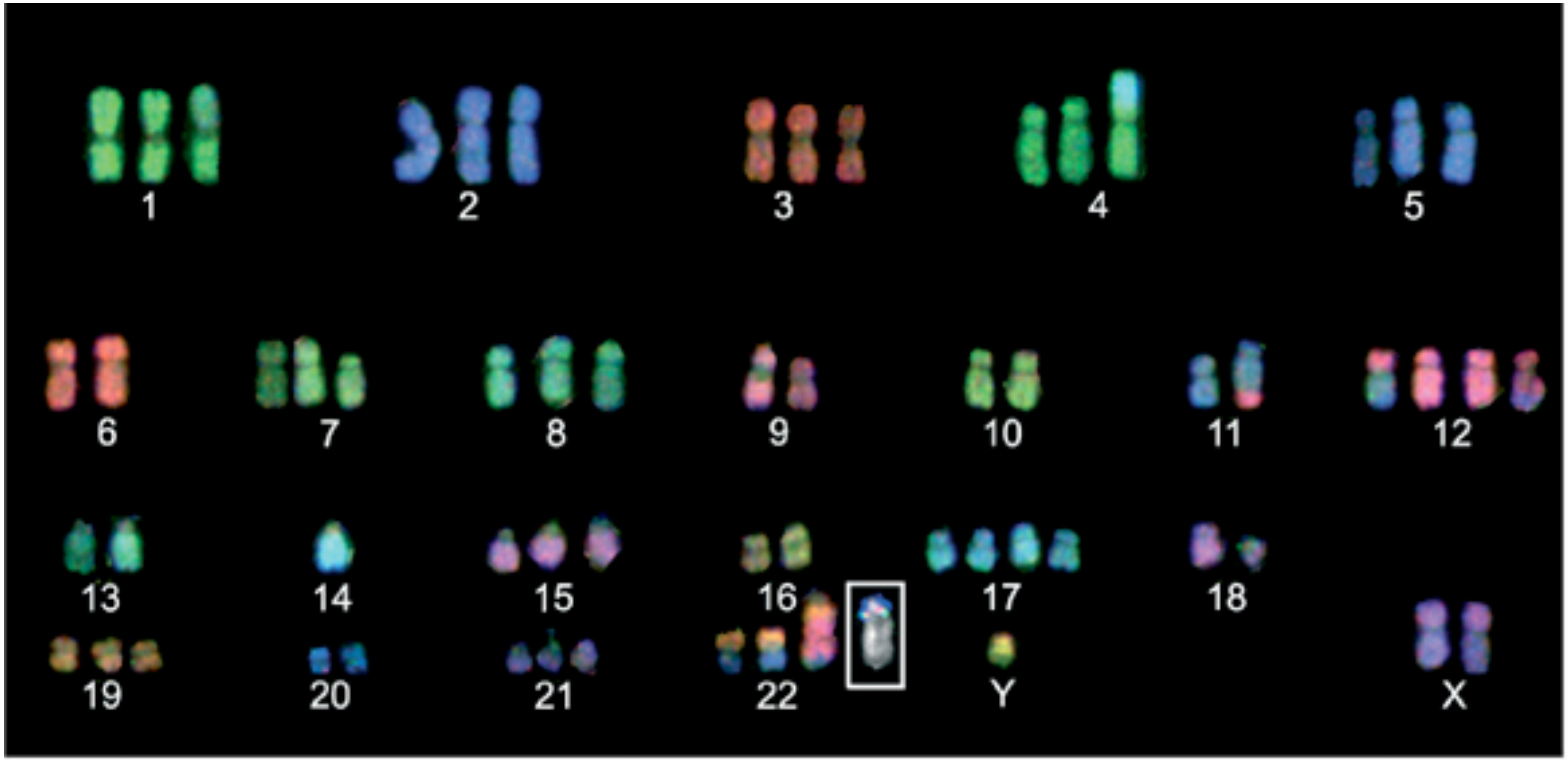
The hallmarks o cancer
Hanahan and Weinberg, Cell 2011



(ib.bioninja.com.au)

What is cancer?

Cancer is a genetic disease: **chromosomal aberrations**



Spectral karyotyping

Chromosomal gain, loss
Translocation, inversion
Focal amplification

What is cancer?

Cancer is a genetic disease: **point mutations**

Substitution
Insertion
Deletion

KRAS-wt

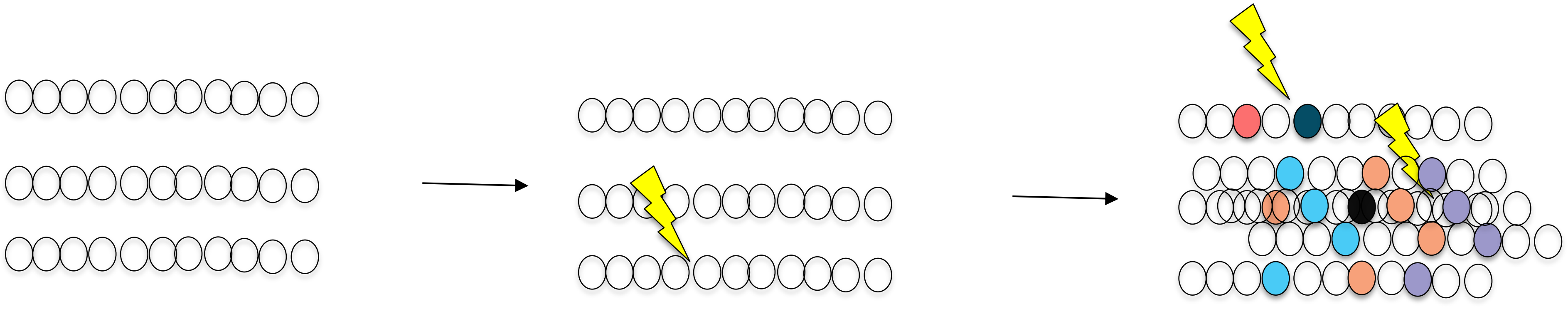
ATGACTGAATATAAACTTGTGGTAGTTGGAGCT**GGT**GGCGTAGGCAAG...
-M--T--E--Y--K--L--V--V--V--G--A--**G**--G--V--G--K-...

KRAS-G12D

ATGACTGAATATAAACTTGTGGTAGTTGGAGCT**GAT**GGCGTAGGCAAG...
-M--T--E--Y--K--L--V--V--V--G--A--**D**--G--V--G--K-...

- Frequent d**river mutation** for tumors of the lung, colon, etc.
- **Predicts lack of benefit** from EGFR inhibitors

The drivers and passengers



Normal tissue

Mutation

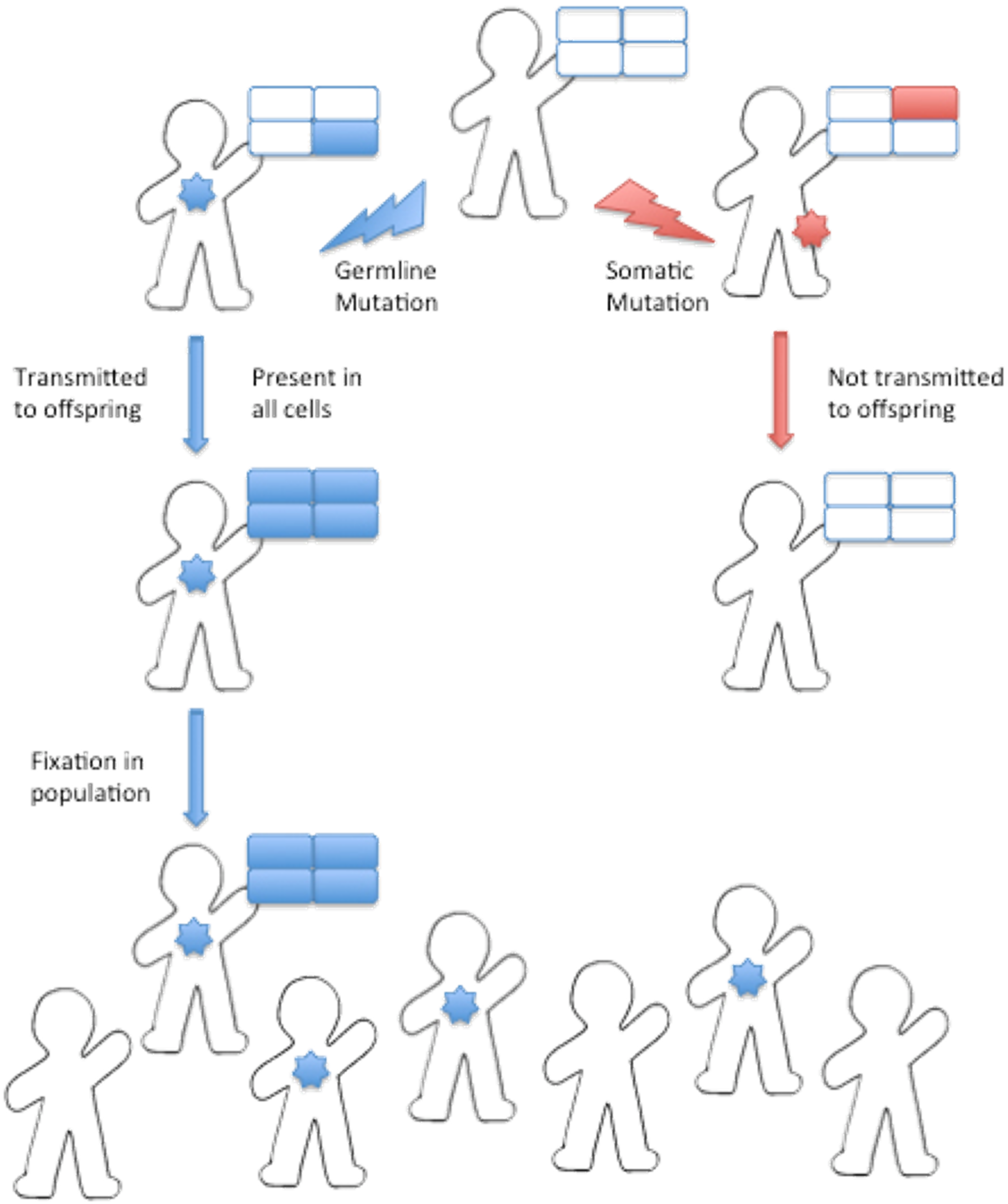
Cancer

⚡ Driver mutation

● ● Passenger mutations

Driver → Confers selective advantage
Disease associated, pathogenic
Passenger → Present in the clonal progenitor

Germline vs Somatic mutations



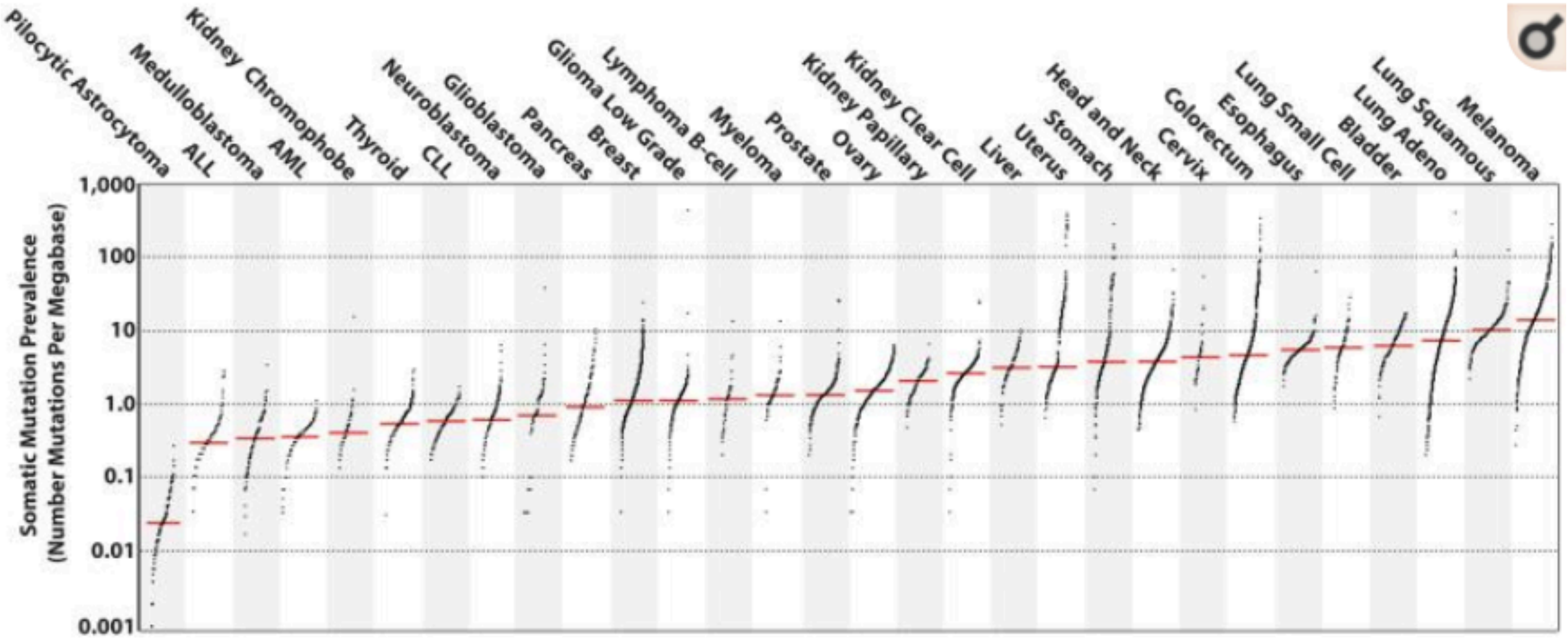
Germline Mutations
Present in **all** cells
Transmitted to offspring
Fixate in population (SNP)

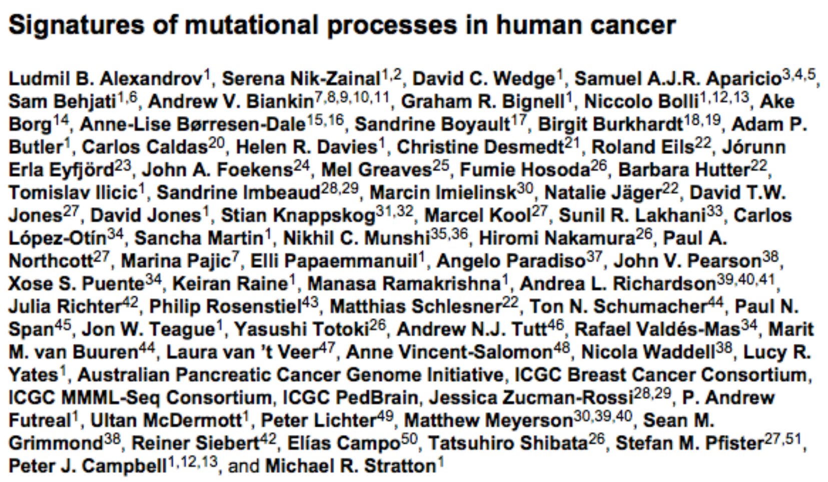
Somatic Mutations
-Present only in **some** cells
-**Not transmitted** to offspring
-Do **not fixate** in population

“Cancer is not a single disease, but rather 150+ different diseases.”

Prof. Dr. Mariano Barbacid, former director of the Spanish National Cancer Research Center and discoverer of the first oncogene, RAS.

Number of somatic mutations in different cancer types



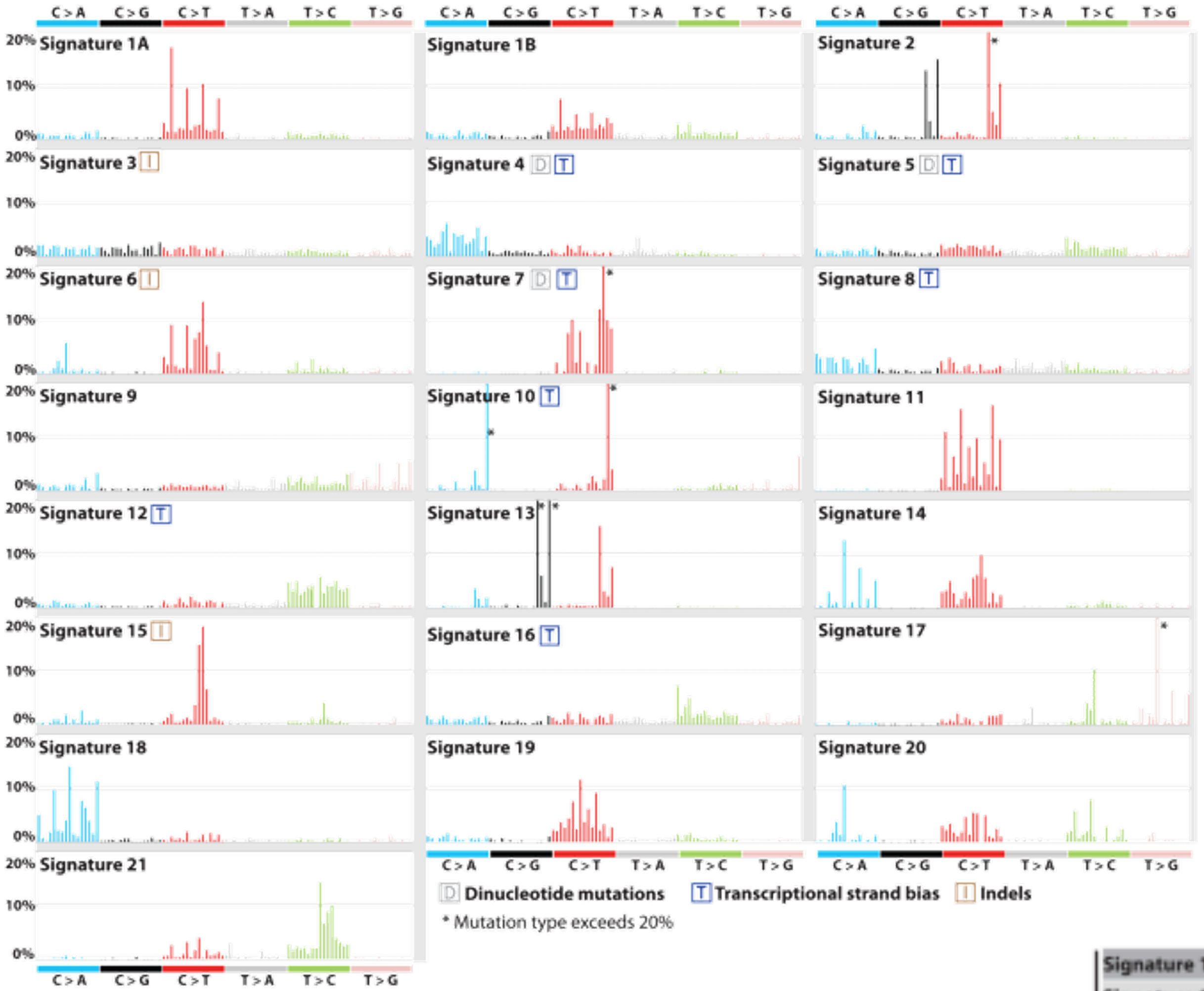


Signatures of mutational processes in human cancer

Ludmil B. Alexandrov[1], Serena Nik-Zainal[1,2], David C. Wedge[1], Samuel A.J.R. Aparicio[3,4,5], Sam Behjati[1,6], Andrew V. Biankin[7,8,9,10,11], Graham R. Bignell[1], Niccolo Bolli[1,12,13], Ake Borg[14], Anne-Lise Børresen-Dale[15,16], Sandrine Boyault[17], Birgit Burkhardt[18,19], Adam P. Butler[1], Carlos Caldas[20], Helen R. Davies[1], Christine Desmedt[21], Roland Eils[22], Jórunn Erla Eyfjörd[23], John A. Foekens[24], Mel Greaves[25], Fumie Hosoda[26], Barbara Hutter[22], Tomislav Ilicic[1], Sandrine Imbeaud[28,29], Marcin Imielinsk[30], Natalie Jäger[22], David T.W. Jones[27], David Jones[1], Stian Knappskog[31,32], Marcel Kool[27], Sunil R. Lakhani[33], Carlos López-Otín[34], Sancha Martin[1], Nikhil C. Munshi[35,36], Hiromi Nakamura[26], Paul A. Northcott[27], Marina Pajic[7], Elli Papaemmanuil[1], Angelo Paradiso[37], John V. Pearson[38], Xose S. Puente[34], Keiran Raine[1], Manasa Ramakrishna[1], Andrea L. Richardson[39,40,41], Julia Richter[42], Philip Rosenstiel[43], Matthias Schlesner[22], Ton N. Schumacher[44], Paul N. Span[45], Jon W. Teague[1], Yasushi Totoki[26], Andrew N.J. Tutt[46], Rafael Valdés-Mas[34], Marit M. van Buuren[44], Laura van 't Veer[47], Anne Vincent-Salomon[48], Nicola Waddell[38], Lucy R. Yates[1], Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain, Jessica Zucman-Rossi[28,29], P. Andrew Futreal[1], Ultan McDermott[1], Peter Lichter[49], Matthew Meyerson[30,39,40], Sean M. Grimmond[38], Reiner Siebert[42], Elías Campo[50], Tatsuhiro Shibata[26], Stefan M. Pfister[27,51], Peter J. Campbell[1,12,13], and Michael R. Stratton[1]

Nature et al., 2013

Mutation signatures in different cancer types



E.g.: Signature 4

- Smoking induced mutations

- Lung cancers

Validated Mutational Signatures

		ALL	AML	Bladder	Breast	Cervix	CLL	Colorectum	Esophagus	Glioblastoma	Glioma Low Grade	Head and Neck	Kidney Chromophobe	Kidney Clear Cell	Kidney Papillary	Liver	Lung Adeno	Lung Small Cell	Lung Squamous	Lymphoma B-cell	Medulloblastoma	Melanoma	Myeloma	Neuroblastoma	Ovary	Pancreas	Pilocytic Astrocytoma	Prostate	Stomach	Thyroid	Uterus
		2	2	4	5	2	3	3	4	2	4	4	2	2	2	6	4	2	3	5	3	3	2	2	2	4	2	2	6	2	6
Signature 1A	7		🟩							🟩	🟩		🟩								🟩							🟩	🟩		
Signature 1B	19	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩			🟩		🟩		🟩	🟩			🟩		🟩		🟩	🟩	🟩	🟩				🟩
Signature 2	16	🟩		🟩	🟩	🟩	🟩		🟩			🟩			🟩		🟩		🟩	🟩			🟩			🟩			🟩	🟩	🟩
Signature 3	3				🟩																				🟩	🟩					
Signature 4	5											🟩				🟩	🟩	🟩	🟩												
Signature 5	9			🟩							🟩				🟩		🟩		🟩	🟩	🟩		🟩							🟩	
Signature 6	9							🟩	🟩		🟩		🟩	🟩		🟩										🟩		🟩			🟩
Signature 7	2											🟩										🟩									
Signature 8	2				🟩																🟩										
Signature 9	2						🟩													🟩											
Signature 10	2							🟩																							🟩
Signature 11	2									🟩												🟩									
Signature 12	2															🟩															🟩
Signature 13	2			🟩	🟩																										
Signature 14	2										🟩																				🟩
Signature 15	2																	🟩											🟩		
Signature 16	1															🟩															
Signature 17	4								🟩							🟩				🟩									🟩		
Signature 18	1																							🟩							
Signature 19	1																										🟩				
Signature 20	1																												🟩		
Signature 21	1																												🟩		
Other					🟩			🟩		🟩			🟩	🟩		🟩	🟩		🟩			🟩						🟩	🟩		

🟩 Mutational signature present 🟦 Total validated mutational signatures in a cancer type 🟥 Total cancer types in which a signature is op

Alexandrov, Nature et al., 2013

Why study cancer genomes?

For the researcher:

- Identify recurrent mutations that represent druggable targets
- Identify specific mutations or patterns that predict benefit from specific drugs
- Study the evolutionary process -- mutation, selection

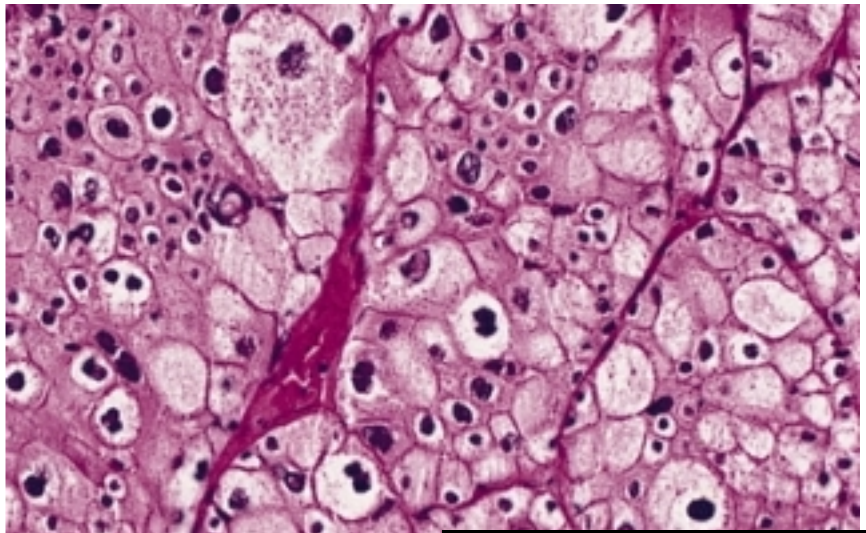
For the cancer patient:

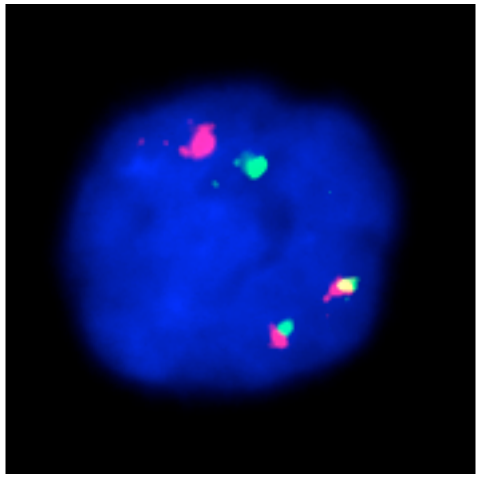
Identify “actionable” mutations - inform treatment decisions

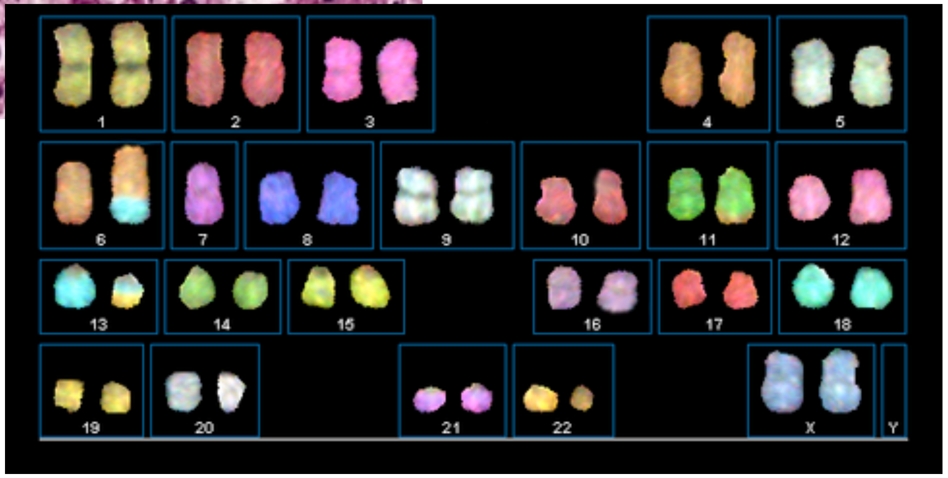
Aid in diagnosis

Characterising a tumour specimen

Measured in individual cells:
- Cellular/tissue morphology
- Protein expression
- Gene copy number (FISH)
- Karyotype







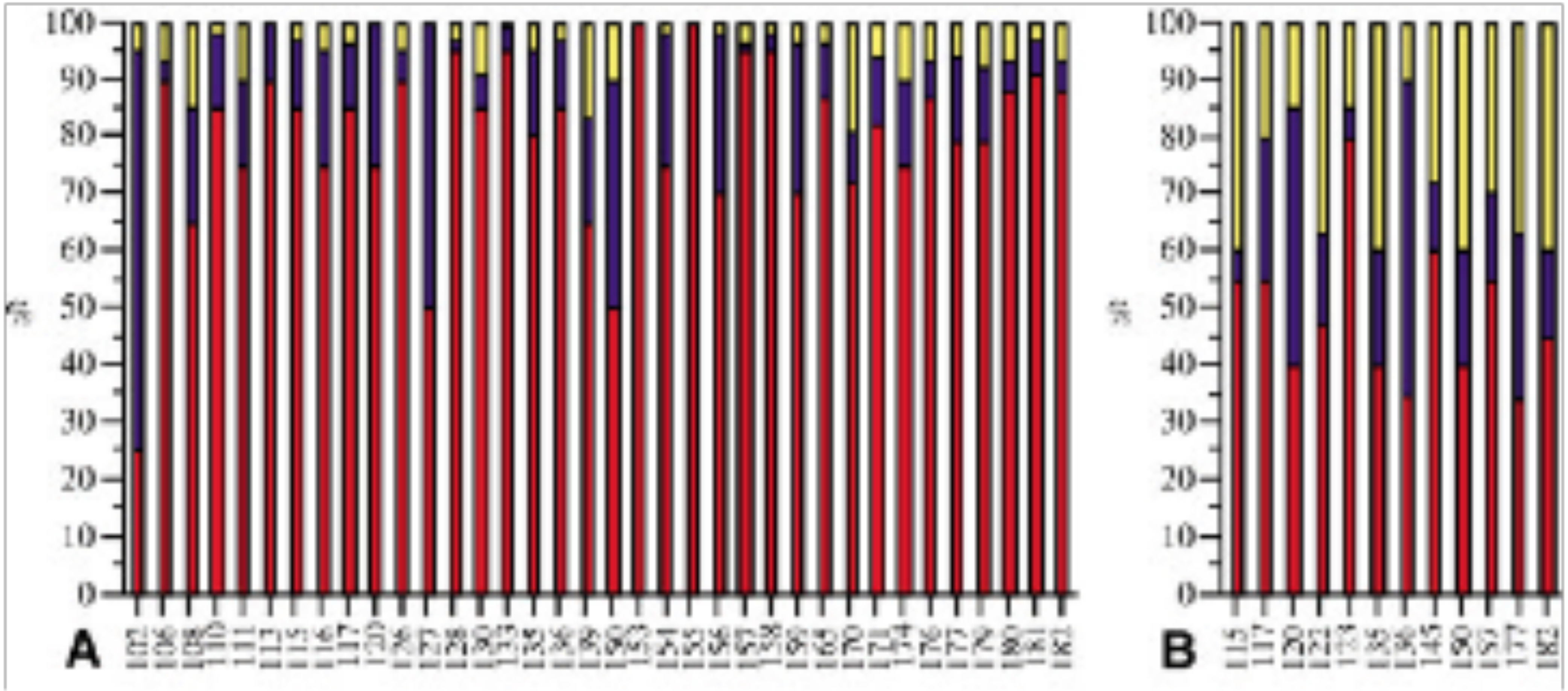
Measured in bulk tissue (usually):
- Gene expression (from microarray)
- Copy number profile (from SNP array)
- WES/WGS/RNA-seq

Bulk tissue includes non-tumor cells!

```
@HS21_6684:1:1306:6031:9563#14
CTTCCGATCTGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
+HS21_6684:1:1306:6031:9563#14
9((A=CA;2FDEEE>E=IIIIIIGGGIIHEF?CEHFDIGIIIGGGEGGHHHHIHFBBBGHEIHHFDHDDFDD?@@
```

Caution 1: Tumour specimens are not just tumour cells

Histological characterization of a set of tumor specimens



Other stromal*

Lymphocytes

Tumor

Fine-needle aspiration (FNA)

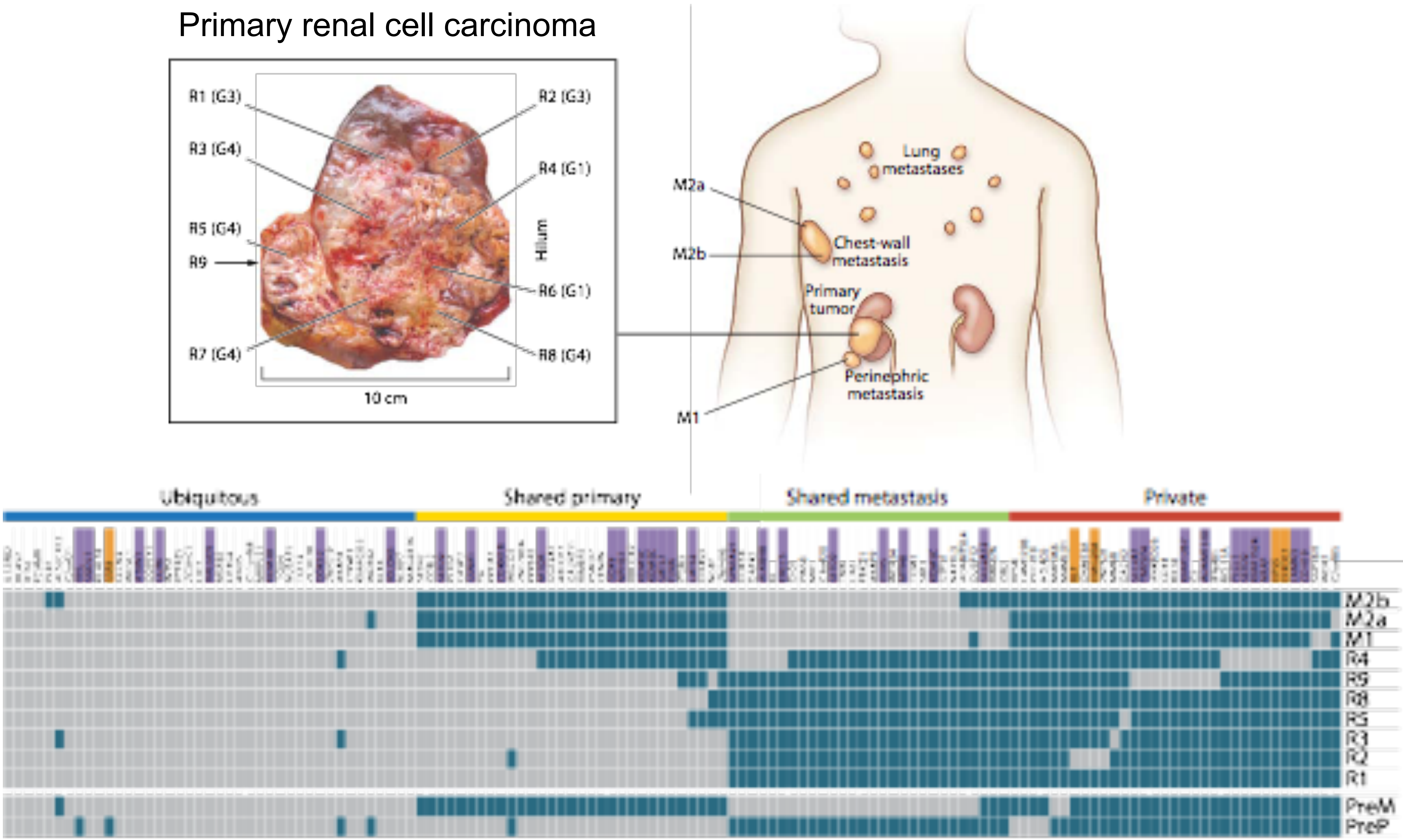
Core needle biopsy

* Other stromal = fibroblasts, endothelials, histocytes, adipocytes

W. F. Symmans *et al*., *Cancer* 97:2960 (2003)

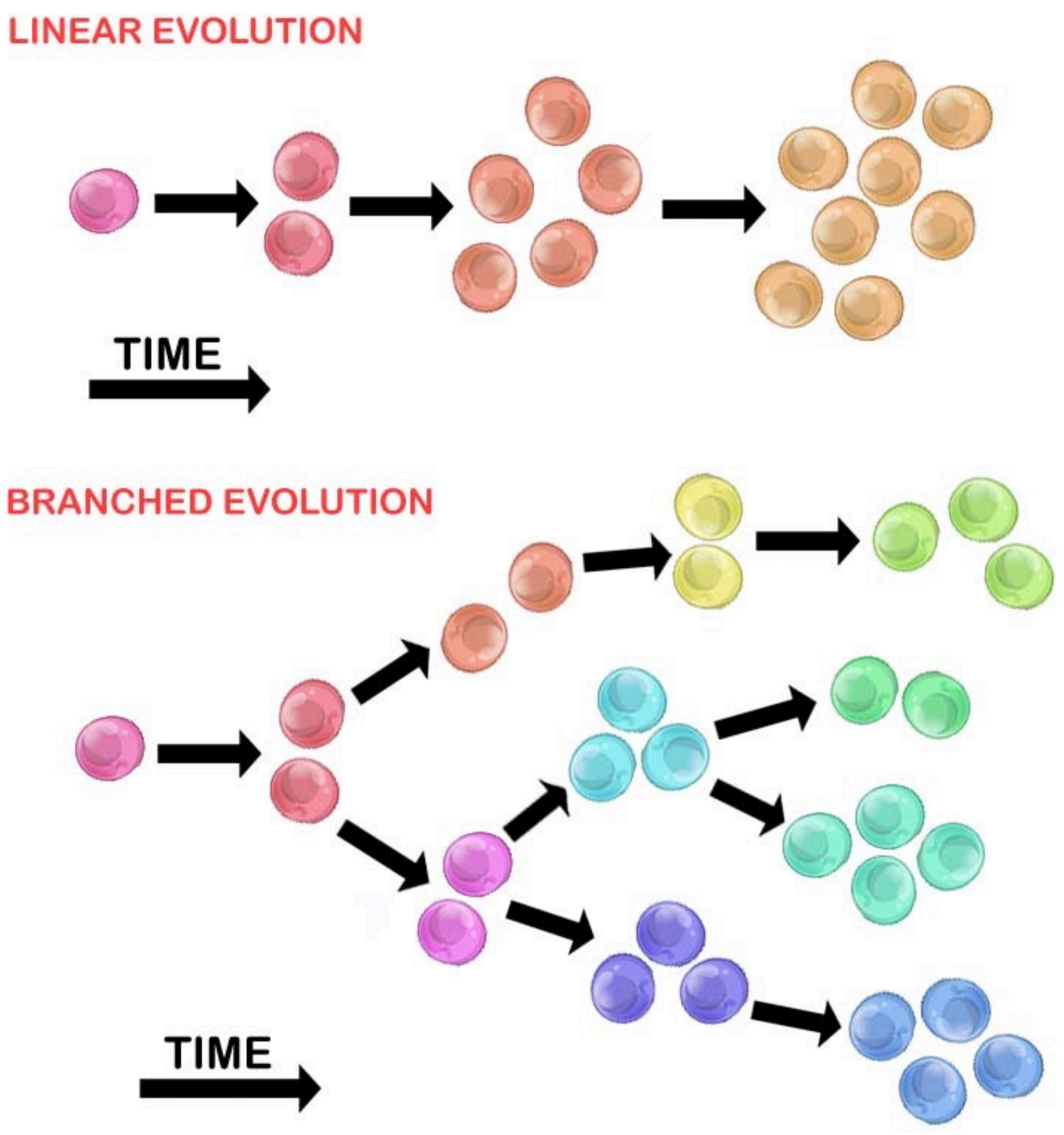
Caution 2: Tumour heterogeneity

Primary renal cell carcinoma



Gerlinger et al. (2012) *NEJM*

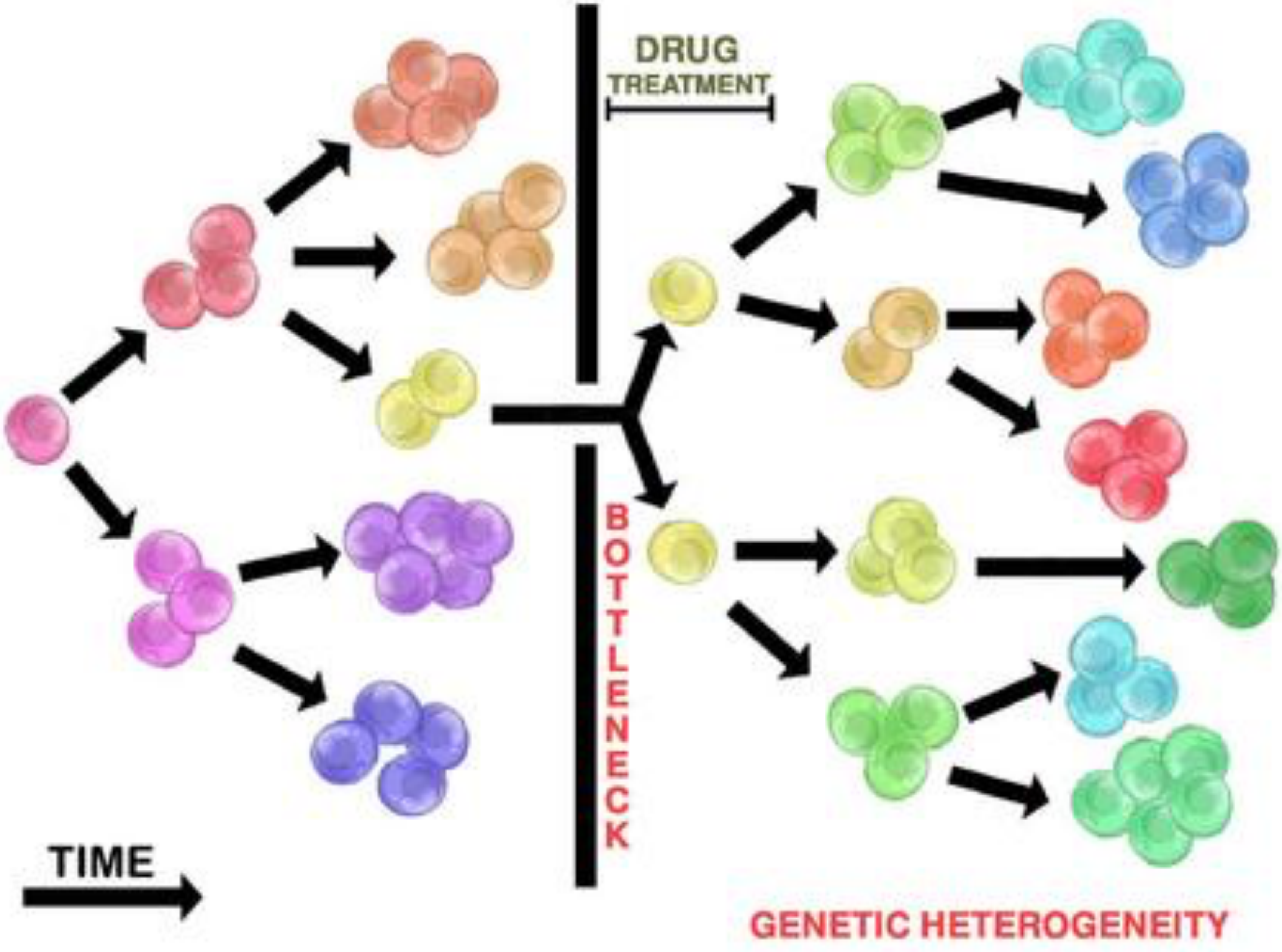
Caution 2: Tumour heterogeneity



- Tumour heterogeneity ➔ differential subclonal evolution

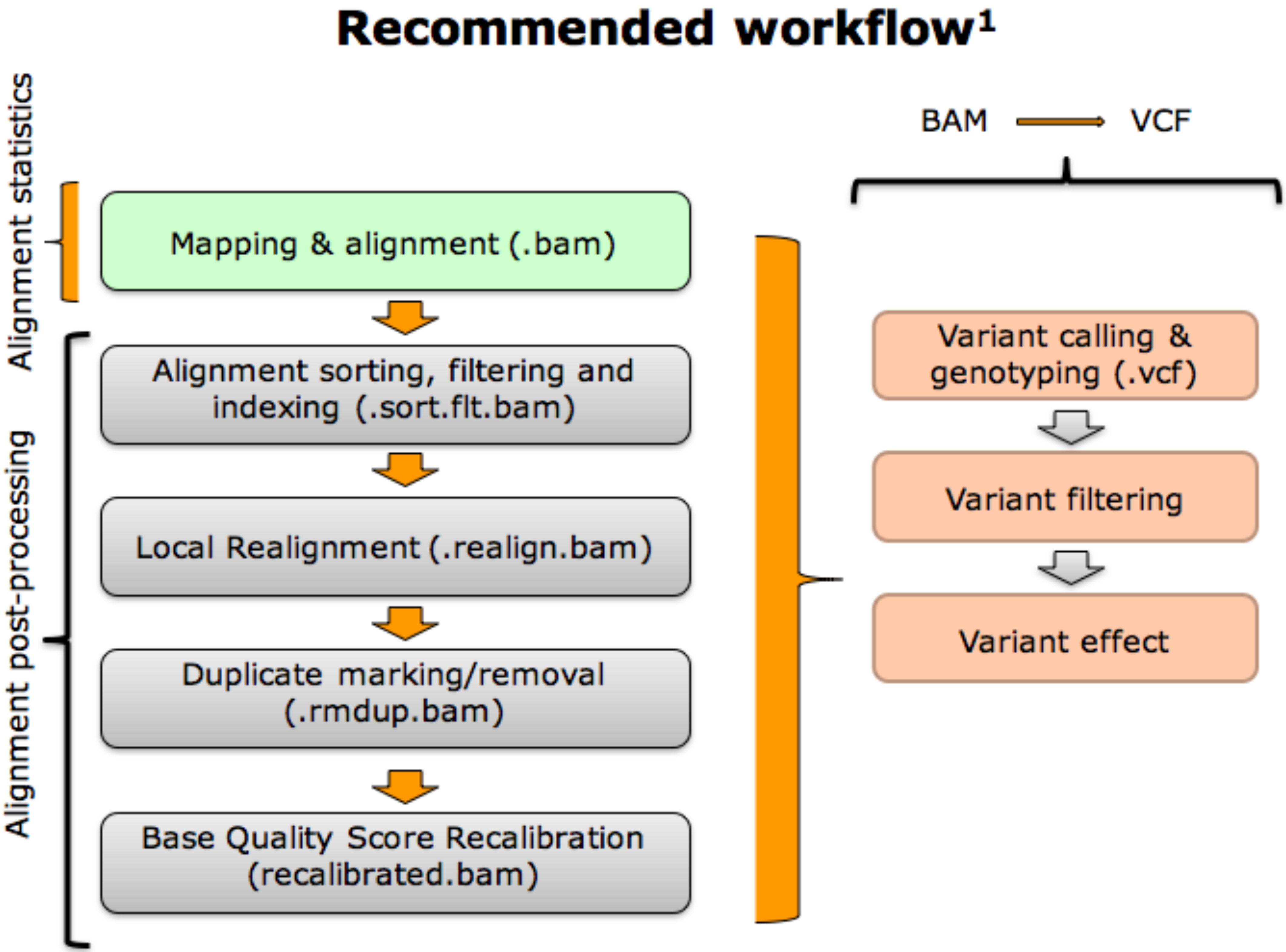
- Clones accumulate different mutations as they diverge.

Treatment can re-shape tumour heterogeneity



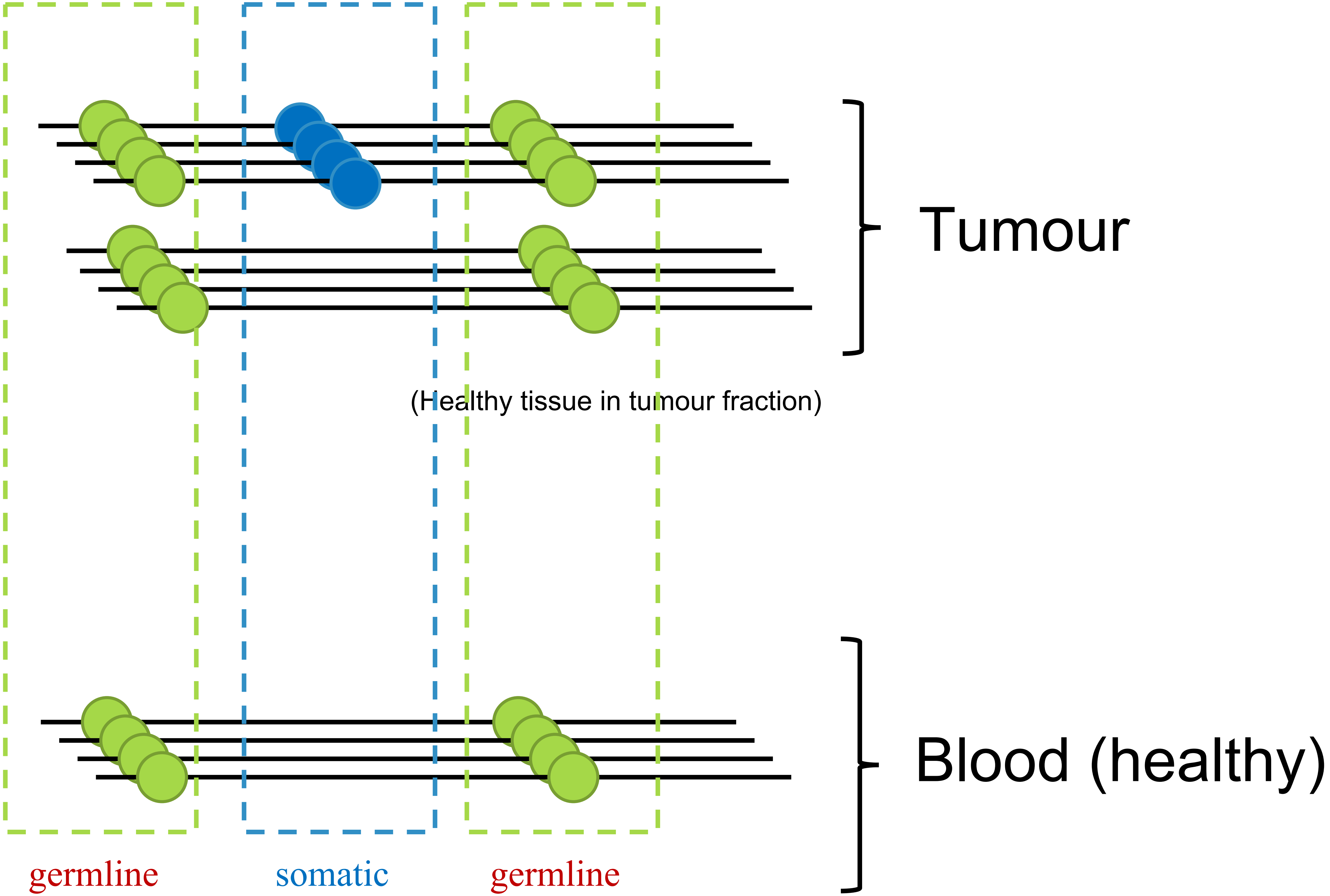
How to identify somatic mutations in a tumor

Variant calling pipeline



Recommended workflow[1]
Alignment statistics
Alignment post-processing
Mapping & alignment (.bam)
Alignment sorting, filtering and indexing (.sort.flt.bam)
Local Realignment (.realign.bam)
Duplicate marking/removal (.rmdup.bam)
Base Quality Score Recalibration (recalibrated.bam)
BAM → VCF
Variant calling & genotyping (.vcf)
Variant filtering
Variant effect

Matched samples for variant calling



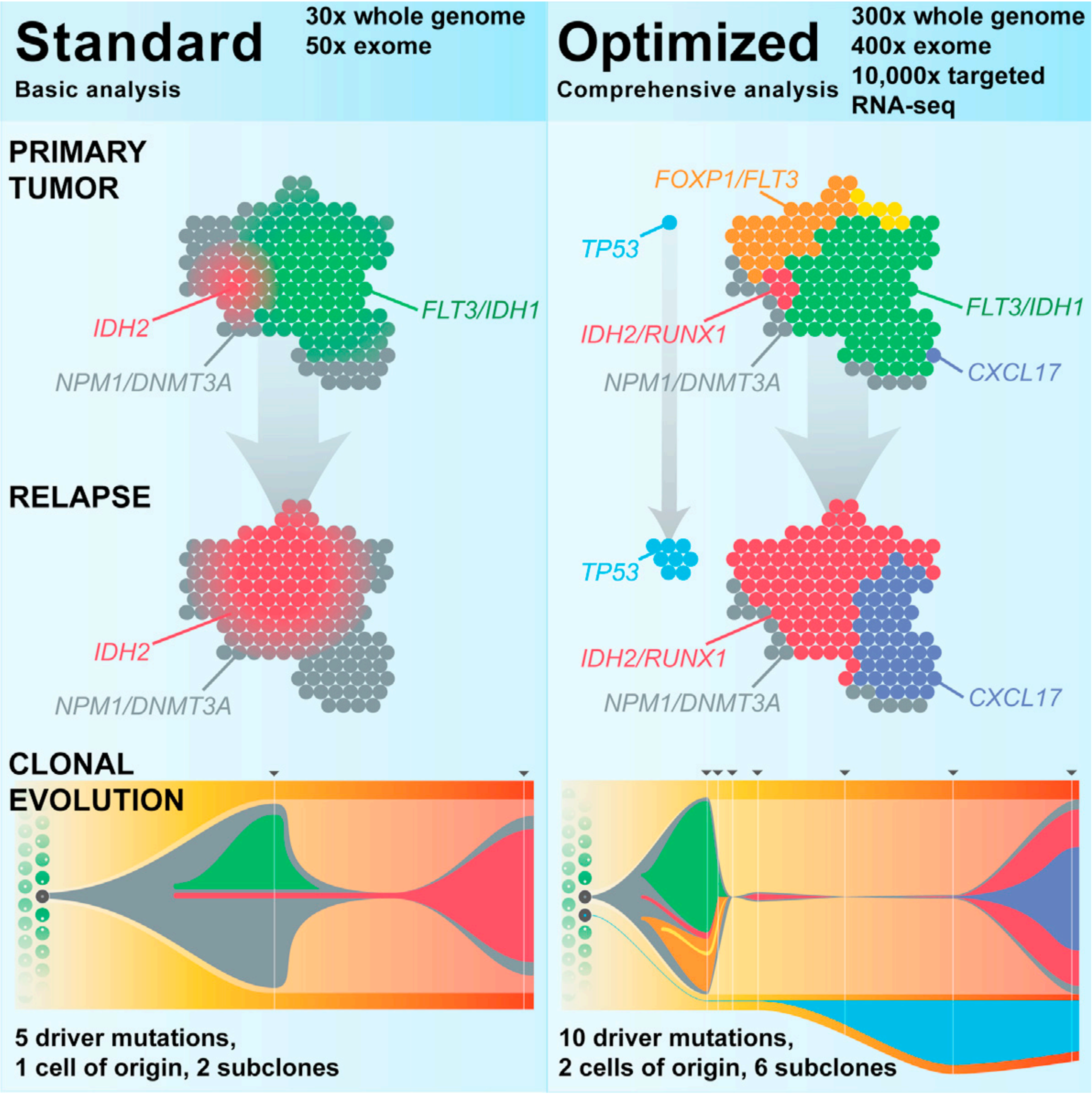
Somatic mutation calling vs.“regular” variant calling

1. We are interested in somatic mutations: differences between the **tumor genome and the normal genome** (NOT the reference genome).

2. The tumor data represents a **mixture** of reads from tumor cells and from normal cells, so we need **deeper sequencing** and **more sensitive analysis** to detect variants.

3. Tumors are often heterogeneous, and relevant mutations may be present at low allelic frequency. So we need **even deeper** sequencing.

Also: we are often interested in copy number changes, translocations, and clonal architecture



M. Griffith *et al.*, *Cell Systems* (2015)

Cancer gene panel amplicon sequencing

TruSeq Amplicon - Cancer Panel Gene List

ABL1	EGFR	GNAS	MLH1	RET
AKT1	ERBB2	HNF1A	MPL	SMAD4
ALK	ERBB4	HRAS	NOTCH1	SMARCB1
APC	FBXW7	IDH1	NPM1	SMO
ATM	FGFR1	JAK2	NRAS	SRC
BRAF	FGFR2	JAK3	PDGFRA	STK11
CDH1	FGFR3	KDR	PIK3CA	TP53
CDKN2A	FLT3	KIT	PTEN	VHL
CSF1R	GNA11	KRAS	PTPN11	
CTNNB1	GNAQ	MET	RB1	

Understanding variation in –omics times

Traditionally

1 Mutation
=
1 Disease

Phenotype
Function
Mechanism

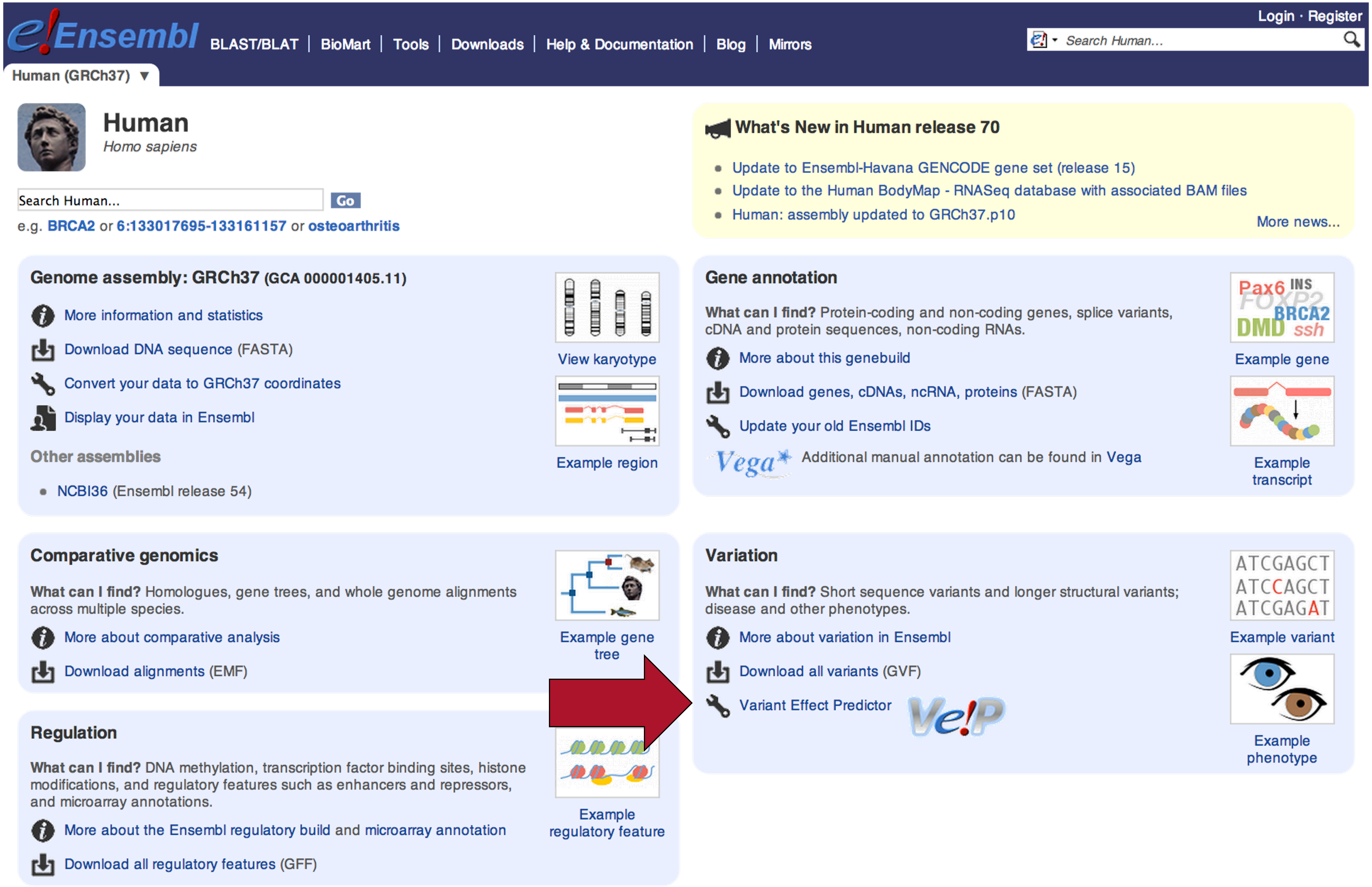
Lots of hard work

Now (High Throughput Sequencing, NGS)

X Mutations
In
Y Patients
And
Z Conditions

Prediction of
Pathogenicity /
Unfeasible
Prioritization

http://www.ensembl.org/Homo_sapiens/Info/Index



Human

Homo sapiens

Search Human… Go

e.g. **BRCA2** or **6:133017695-133161157** or **osteoarthritis**

What's New in Human release 70

- Update to Ensembl-Havana GENCODE gene set (release 15)
- Update to the Human BodyMap - RNASeq database with associated BAM files
- Human: assembly updated to GRCh37.p10

More news...

Genome assembly: GRCh37 (GCA 000001405.11)

More information and statistics

Download DNA sequence (FASTA)

Convert your data to GRCh37 coordinates

Display your data in Ensembl

Other assemblies

- NCBI36 (Ensembl release 54)

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

More about this genebuild

Download genes, cDNAs, ncRNA, proteins (FASTA)

Update your old Ensembl IDs

Additional manual annotation can be found in Vega

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

More about comparative analysis

Download alignments (EMF)

Variation

What can I find? Short sequence variants and longer structural variants; disease and other phenotypes.

More about variation in Ensembl

Download all variants (GVF)

Variant Effect Predictor

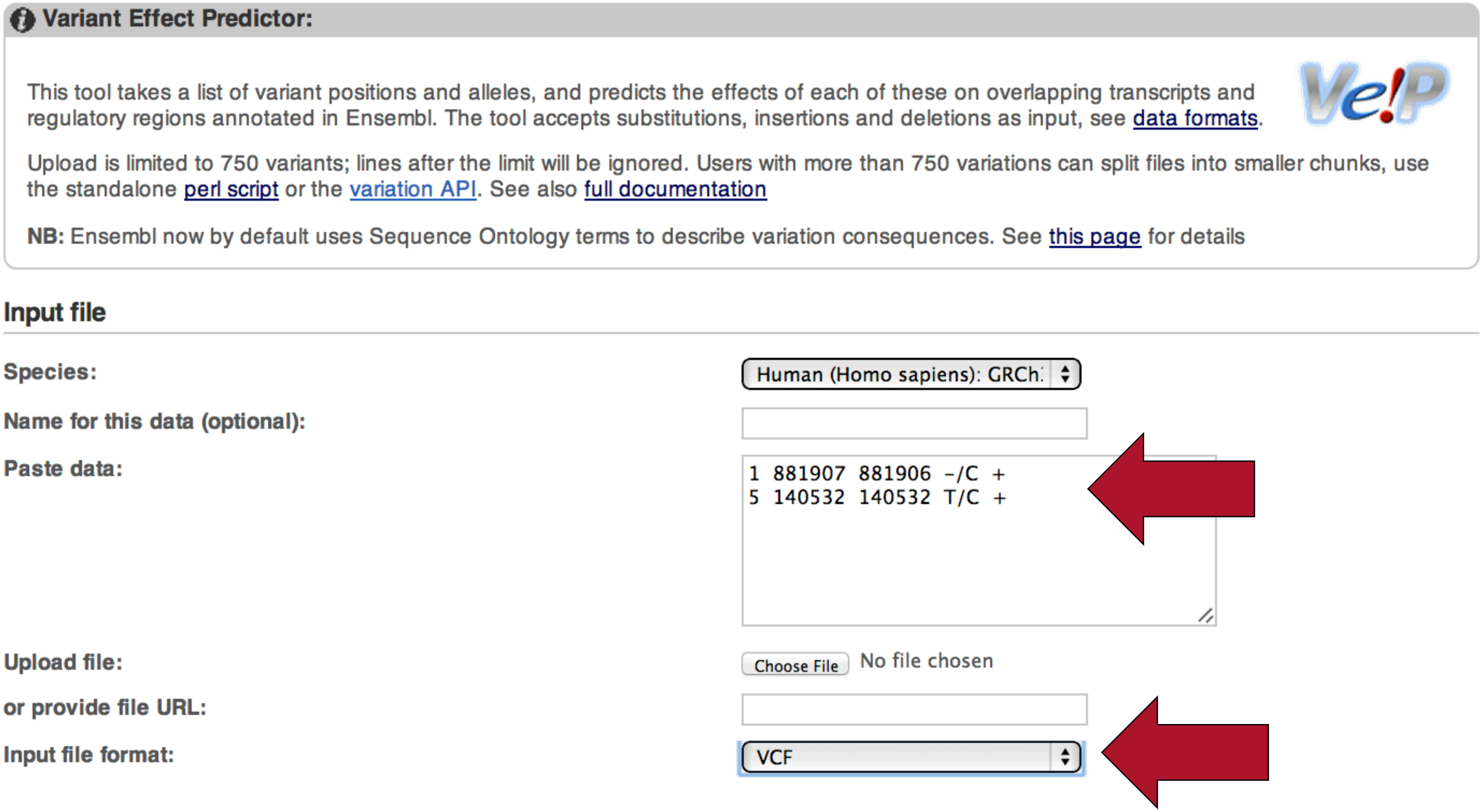
Regulation

What can I find? DNA methylation, transcription factor binding sites, histone modifications, and regulatory features such as enhancers and repressors, and microarray annotations.

More about the Ensembl regulatory build and microarray annotation

Download all regulatory features (GFF)

Ensembl Variant Effect Predictor (I)



Variant Effect Predictor:

This tool takes a list of variant positions and alleles, and predicts the effects of each of these on overlapping transcripts and regulatory regions annotated in Ensembl. The tool accepts substitutions, insertions and deletions as input, see data formats.

Upload is limited to 750 variants; lines after the limit will be ignored. Users with more than 750 variations can split files into smaller chunks, use the standalone perl script or the variation API. See also full documentation

NB: Ensembl now by default uses Sequence Ontology terms to describe variation consequences. See this page for details

Input file

Species: Human (Homo sapiens): GRCh:

Name for this data (optional):

Paste data:

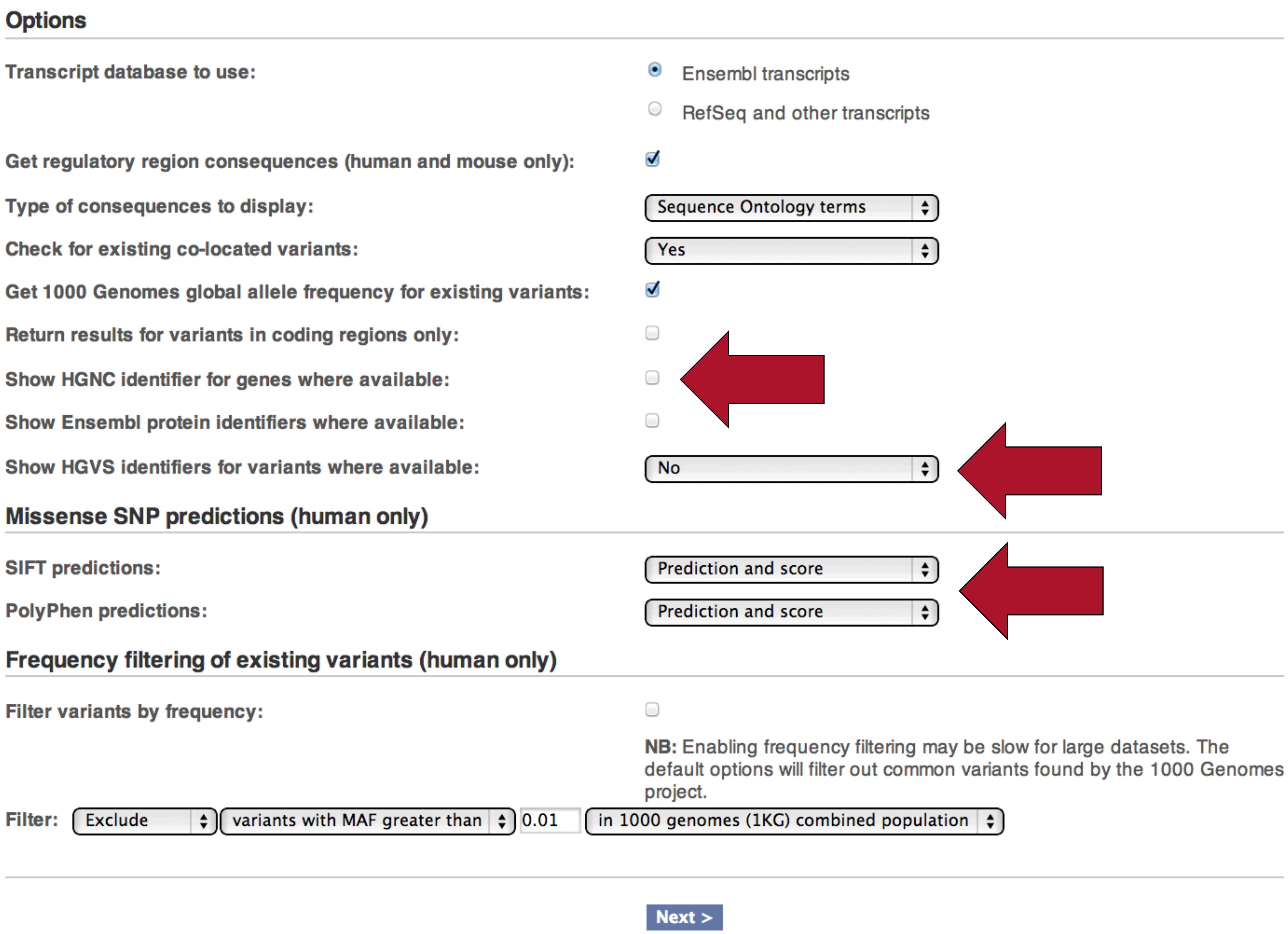
```
1 881907 881906 -/C +
5 140532 140532 T/C +
```

Upload file: Choose File No file chosen

or provide file URL:

Input file format: VCF

Ensembl Variant Effect Predictor (II)



Options

Transcript database to use:
● Ensembl transcripts
○ RefSeq and other transcripts

Get regulatory region consequences (human and mouse only): ☑

Type of consequences to display: Sequence Ontology terms

Check for existing co-located variants: Yes

Get 1000 Genomes global allele frequency for existing variants: ☑

Return results for variants in coding regions only: □

Show HGNC identifier for genes where available: □

Show Ensembl protein identifiers where available: □

Show HGVS identifiers for variants where available: No

Missense SNP predictions (human only)

SIFT predictions: Prediction and score

PolyPhen predictions: Prediction and score

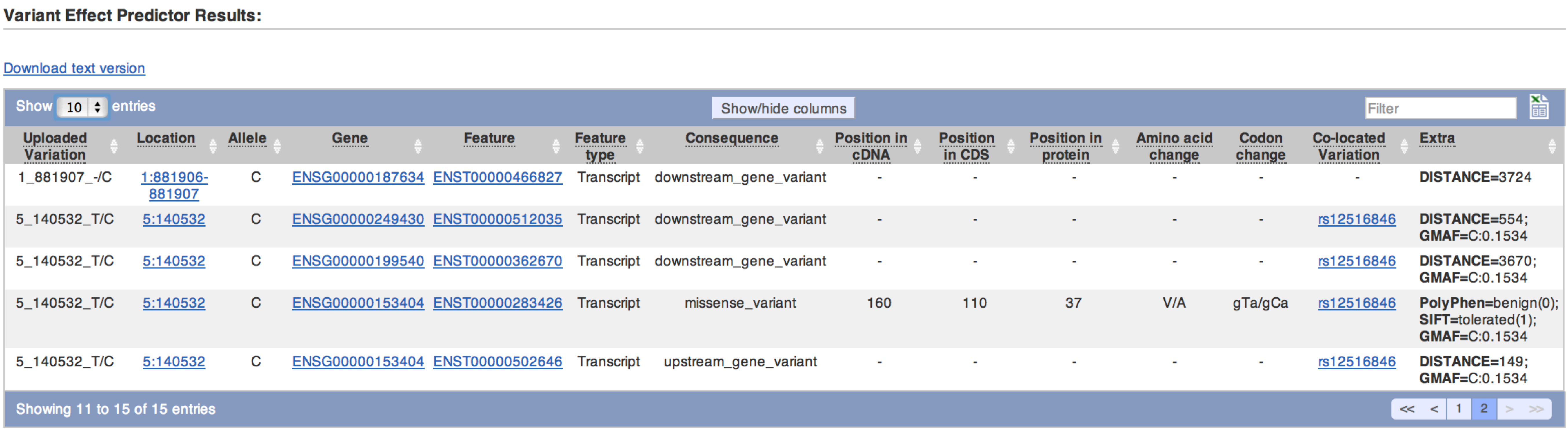
Frequency filtering of existing variants (human only)

Filter variants by frequency: □

NB: Enabling frequency filtering may be slow for large datasets. The default options will filter out common variants found by the 1000 Genomes project.

Filter: Exclude | variants with MAF greater than | 0.01 | in 1000 genomes (1KG) combined population

Ensembl Variant Effect Predictor (Results)

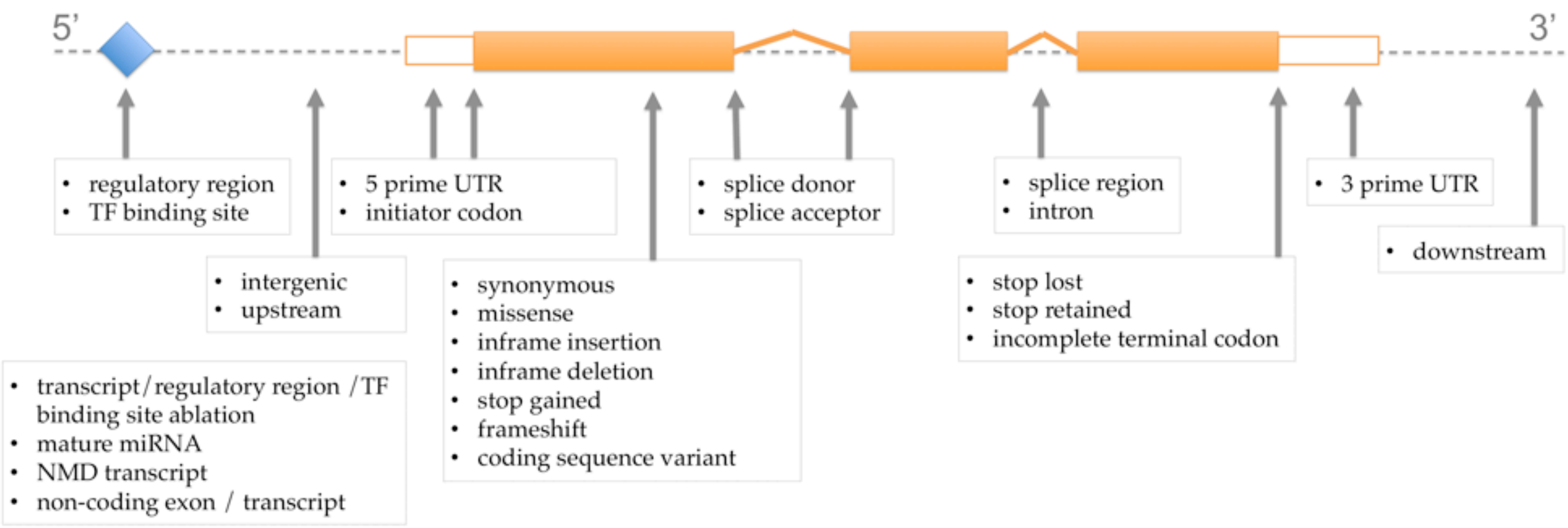


Variant Effect Predictor Results:

Download text version

Uploaded Variation	Location	Allele	Gene	Feature	Feature type	Consequence	Position in cDNA	Position in CDS	Position in protein	Amino acid change	Codon change	Co-located Variation	Extra
1_881907_-/C	1:881906-881907	C	ENSG00000187634	ENST00000466827	Transcript	downstream_gene_variant	-	-	-	-	-	-	DISTANCE=3724
5_140532_T/C	5:140532	C	ENSG00000249430	ENST00000512035	Transcript	downstream_gene_variant	-	-	-	-	-	rs12516846	DISTANCE=554; GMAF=C:0.1534
5_140532_T/C	5:140532	C	ENSG00000199540	ENST00000362670	Transcript	downstream_gene_variant	-	-	-	-	-	rs12516846	DISTANCE=3670; GMAF=C:0.1534
5_140532_T/C	5:140532	C	ENSG00000153404	ENST00000283426	Transcript	missense_variant	160	110	37	V/A	gTa/gCa	rs12516846	PolyPhen=benign(0); SIFT=tolerated(1); GMAF=C:0.1534
5_140532_T/C	5:140532	C	ENSG00000153404	ENST00000502646	Transcript	upstream_gene_variant	-	-	-	-	-	rs12516846	DISTANCE=149; GMAF=C:0.1534

Showing 11 to 15 of 15 entries



- regulatory region
- TF binding site

- intergenic
- upstream

- 5 prime UTR
- initiator codon

- synonymous
- missense
- inframe insertion
- inframe deletion
- stop gained
- frameshift
- coding sequence variant

- splice donor
- splice acceptor

- splice region
- intron

- stop lost
- stop retained
- incomplete terminal codon

- 3 prime UTR

- downstream

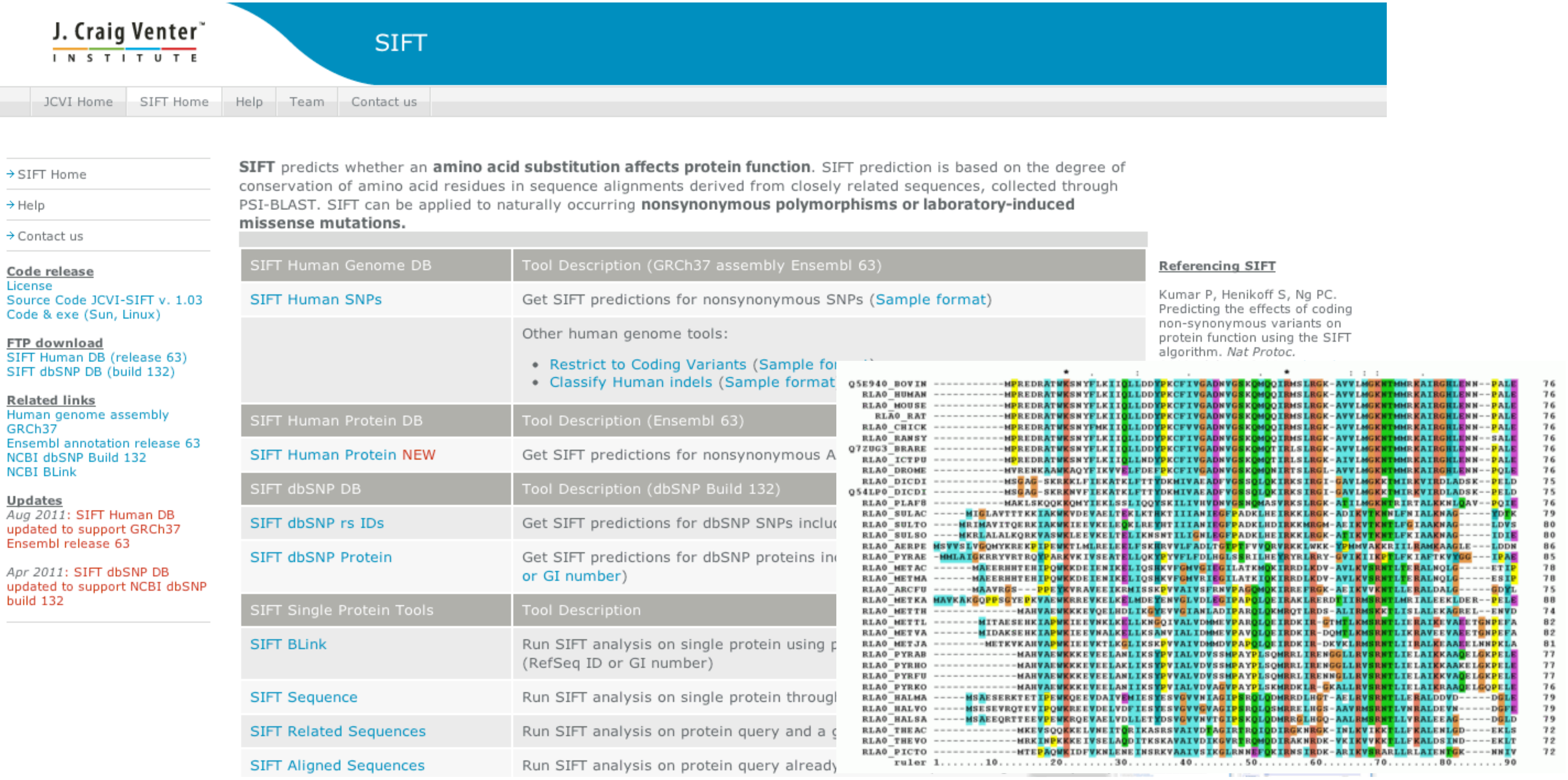
- transcript/regulatory region /TF binding site ablation
- mature miRNA
- NMD transcript
- non-coding exon / transcript

Predictors: SIFT

Predicting Deleterious Amino Acid Substitutions

Pauline C. Ng and Steven Henikoff

http://sift.jcvi.org/



J. Craig Venter INSTITUTE

SIFT

JCVI Home | SIFT Home | Help | Team | Contact us

→SIFT Home
→Help
→Contact us

Code release
License
Source Code JCVI-SIFT v. 1.03
Code & exe (Sun, Linux)

FTP download
SIFT Human DB (release 63)
SIFT dbSNP DB (build 132)

Related links
Human genome assembly GRCh37
Ensembl annotation release 63
NCBI dbSNP Build 132
NCBI BLink

Updates
Aug 2011: SIFT Human DB updated to support GRCh37 Ensembl release 63

Apr 2011: SIFT dbSNP DB updated to support NCBI dbSNP build 132

SIFT predicts whether an **amino acid substitution affects protein function**. SIFT prediction is based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences, collected through PSI-BLAST. SIFT can be applied to naturally occurring **nonsynonymous polymorphisms or laboratory-induced missense mutations.**

SIFT Human Genome DB	Tool Description (GRCh37 assembly Ensembl 63)
SIFT Human SNPs	Get SIFT predictions for nonsynonymous SNPs (Sample format)
	Other human genome tools: • Restrict to Coding Variants (Sample fo • Classify Human indels (Sample format
SIFT Human Protein DB	Tool Description (Ensembl 63)
SIFT Human Protein NEW	Get SIFT predictions for nonsynonymous A
SIFT dbSNP DB	Tool Description (dbSNP Build 132)
SIFT dbSNP rs IDs	Get SIFT predictions for dbSNP SNPs inclu
SIFT dbSNP Protein	Get SIFT predictions for dbSNP proteins in or GI number)
SIFT Single Protein Tools	Tool Description
SIFT BLink	Run SIFT analysis on single protein using (RefSeq ID or GI number)
SIFT Sequence	Run SIFT analysis on single protein throug
SIFT Related Sequences	Run SIFT analysis on protein query and a
SIFT Aligned Sequences	Run SIFT analysis on protein query alread

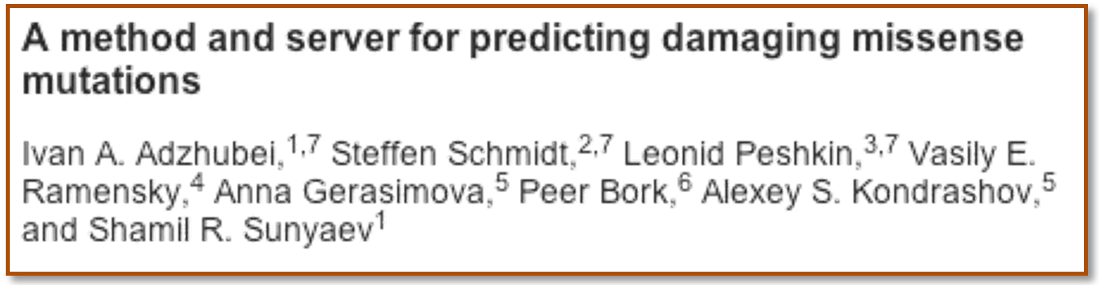
Referencing SIFT

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.*

- Based on the degree of conservation in a multiple sequence alignment (MSA)
  - MSA generated from PSI-BLAST results (closely related sequences)
    - Deleterious if SIFT ≤ 0.05

Predictors: Polyphen-2

http://genetics.bwh.harvard.edu/pph2/



A method and server for predicting damaging missense mutations

Ivan A. Adzhubei,[1,7] Steffen Schmidt,[2,7] Leonid Peshkin,[3,7] Vasily E. Ramensky,[4] Anna Gerasimova,[5] Peer Bork,[6] Alexey S. Kondrashov,[5] and Shamil R. Sunyaev[1]

MACHINE LEARNING

- Naïve Bayes Classifier

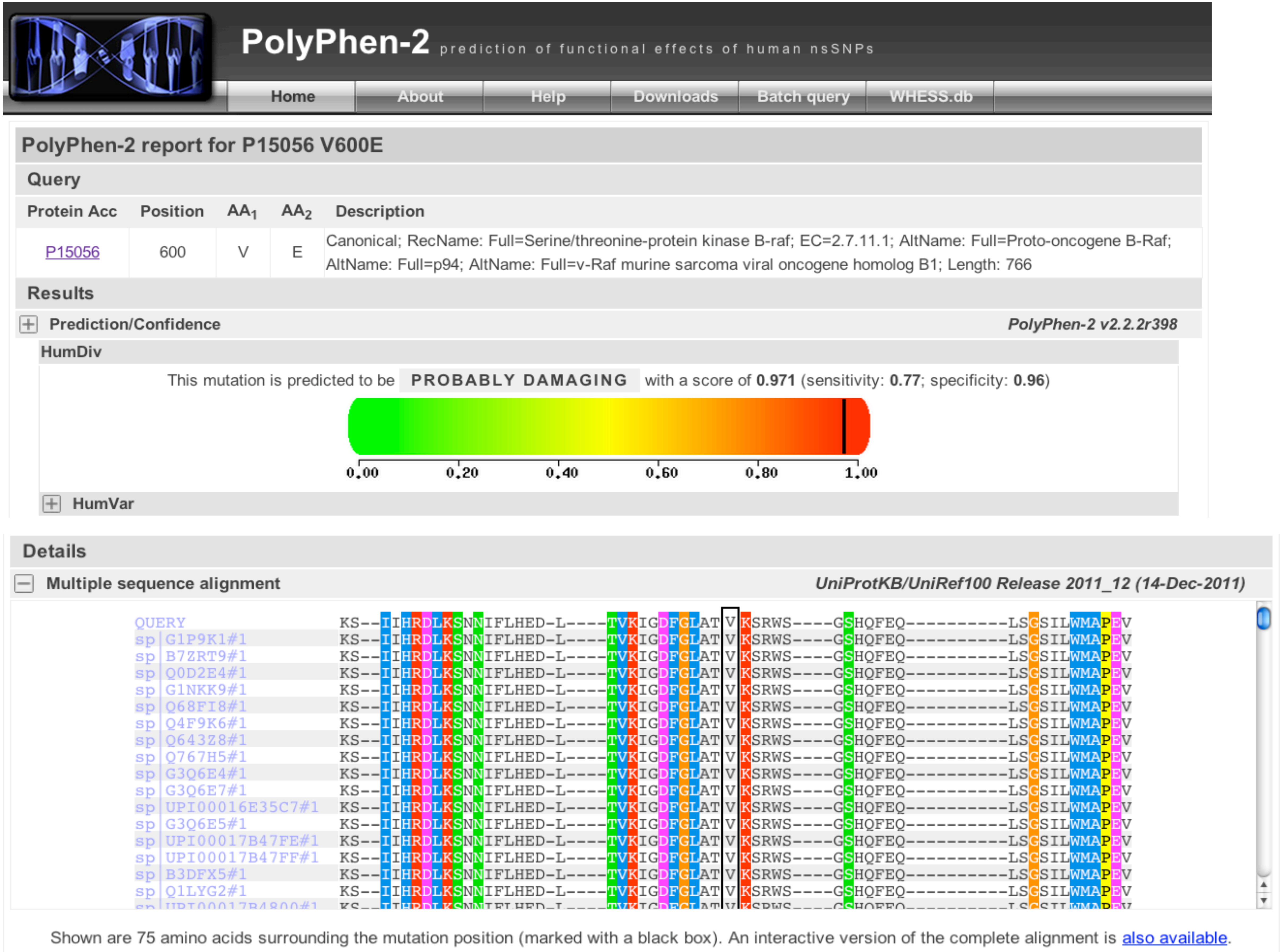
SEQUENCE BASED FEATURES

- Importance of site: DISULFID, CROSSLNK, BINDING, ACT_SITE, LIPID, METAL, SITE, MOD_RES, CARBOHYD, NON_STD…
- Importance of region: TRANSMEM, INTRAMEM, COMPBIAS, REPEAT, COILED, SIGNAL, PROPEP…
- PSIC conservation score

STRUCTURE BASED FEATURES

- Likeness to destroy hydrophobic core, electrostatic interactions, interactions with ligands, or other important features of proteins

Predictors: Polyphen-2



PolyPhen-2 report for P15056 V600E

Query

Protein Acc	Position	AA_1	AA_2	Description
P15056	600	V	E	Canonical; RecName: Full=Serine/threonine-protein kinase B-raf; EC=2.7.11.1; AltName: Full=Proto-oncogene B-Raf; AltName: Full=p94; AltName: Full=v-Raf murine sarcoma viral oncogene homolog B1; Length: 766

Results

⊞ Prediction/Confidence *PolyPhen-2 v2.2.2r398*

HumDiv

This mutation is predicted to be **PROBABLY DAMAGING** with a score of **0.971** (sensitivity: **0.77**; specificity: **0.96**)

⊞ HumVar

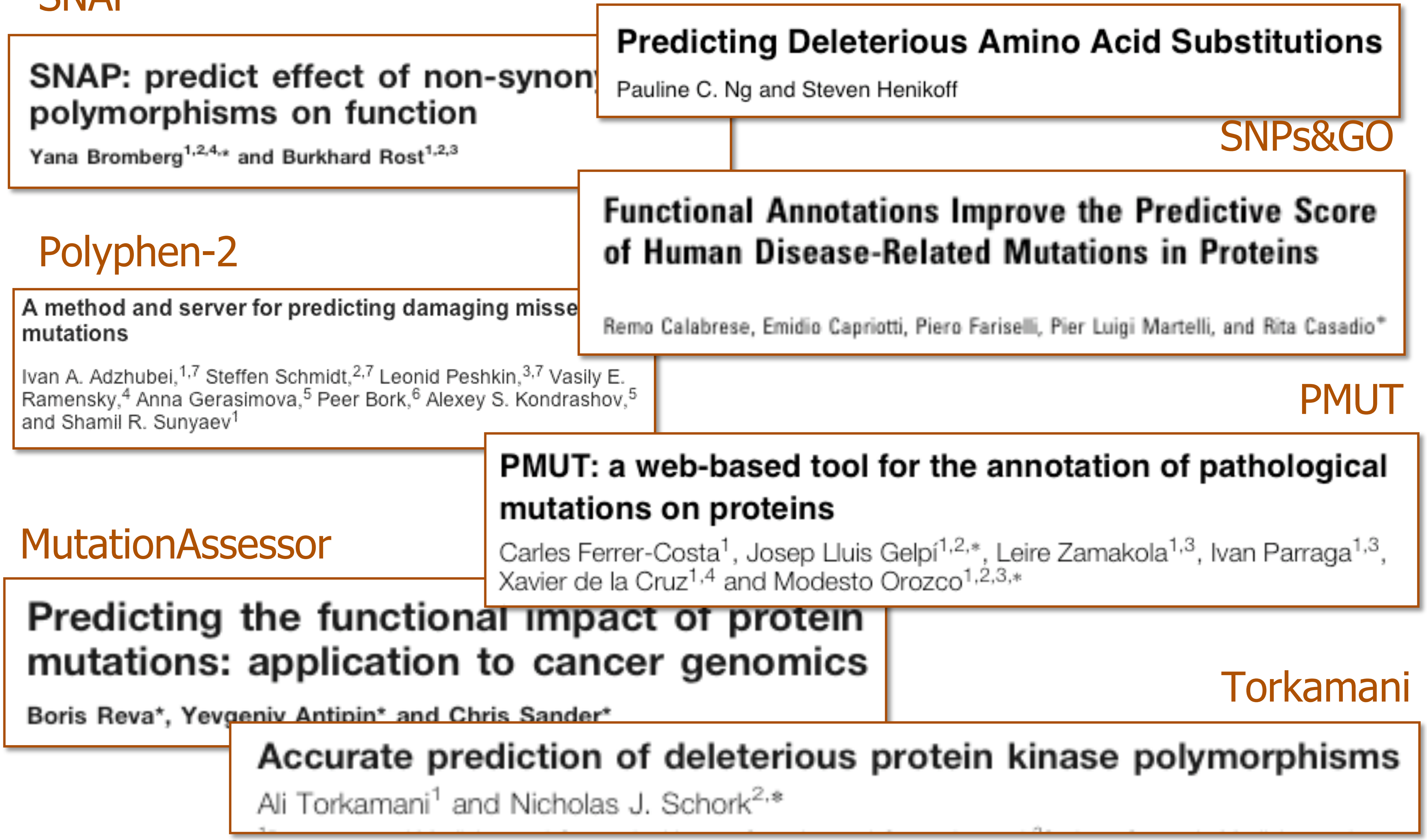
Details

⊟ Multiple sequence alignment *UniProtKB/UniRef100 Release 2011_12 (14-Dec-2011)*

```
QUERY              KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|G1P9K1#1        KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|B7ZRT9#1        KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|Q0D2E4#1        KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|G1NKK9#1        KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|Q68FI8#1        KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|Q4F9K6#1        KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|Q643Z8#1        KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|Q767H5#1        KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|G3Q6E4#1        KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|G3Q6E7#1        KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|UPI00016E35C7#1 KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|G3Q6E5#1        KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|UPI00017B47FE#1 KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|UPI00017B47FF#1 KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|B3DFX5#1        KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|Q1LYG2#1        KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
sp|UPI00017B4800#1 KS--IIHRDLKSNNIFLHED-L----TVKIGDFGLAT V KSRWS----GSHQFEQ----------LSGSILWMAPEV
```

Shown are 75 amino acids surrounding the mutation position (marked with a black box). An interactive version of the complete alignment is also available.

Automatic methods to predict the pathogenicity of mutations



SNAP

SNAP: predict effect of non-synon polymorphisms on function

Yana Bromberg[1,2,4,*] and Burkhard Rost[1,2,3]

SIFT

Predicting Deleterious Amino Acid Substitutions

Pauline C. Ng and Steven Henikoff

SNPs&GO

Functional Annotations Improve the Predictive Score of Human Disease-Related Mutations in Proteins

Remo Calabrese, Emidio Capriotti, Piero Fariselli, Pier Luigi Martelli, and Rita Casadio*

Polyphen-2

A method and server for predicting damaging misse mutations

Ivan A. Adzhubei,[1,7] Steffen Schmidt,[2,7] Leonid Peshkin,[3,7] Vasily E. Ramensky,[4] Anna Gerasimova,[5] Peer Bork,[6] Alexey S. Kondrashov,[5] and Shamil R. Sunyaev[1]

PMUT

PMUT: a web-based tool for the annotation of pathological mutations on proteins

Carles Ferrer-Costa[1], Josep Lluis Gelpí[1,2,*], Leire Zamakola[1,3], Ivan Parraga[1,3], Xavier de la Cruz[1,4] and Modesto Orozco[1,2,3,*]

MutationAssessor

Predicting the functional impact of protein mutations: application to cancer genomics

Boris Reva*, Yevgeniy Antipin* and Chris Sander*

Torkamani

Accurate prediction of deleterious protein kinase polymorphisms

Ali Torkamani[1] and Nicholas J. Schork[2,*]

Some of the (many) methods implemented during the last decade