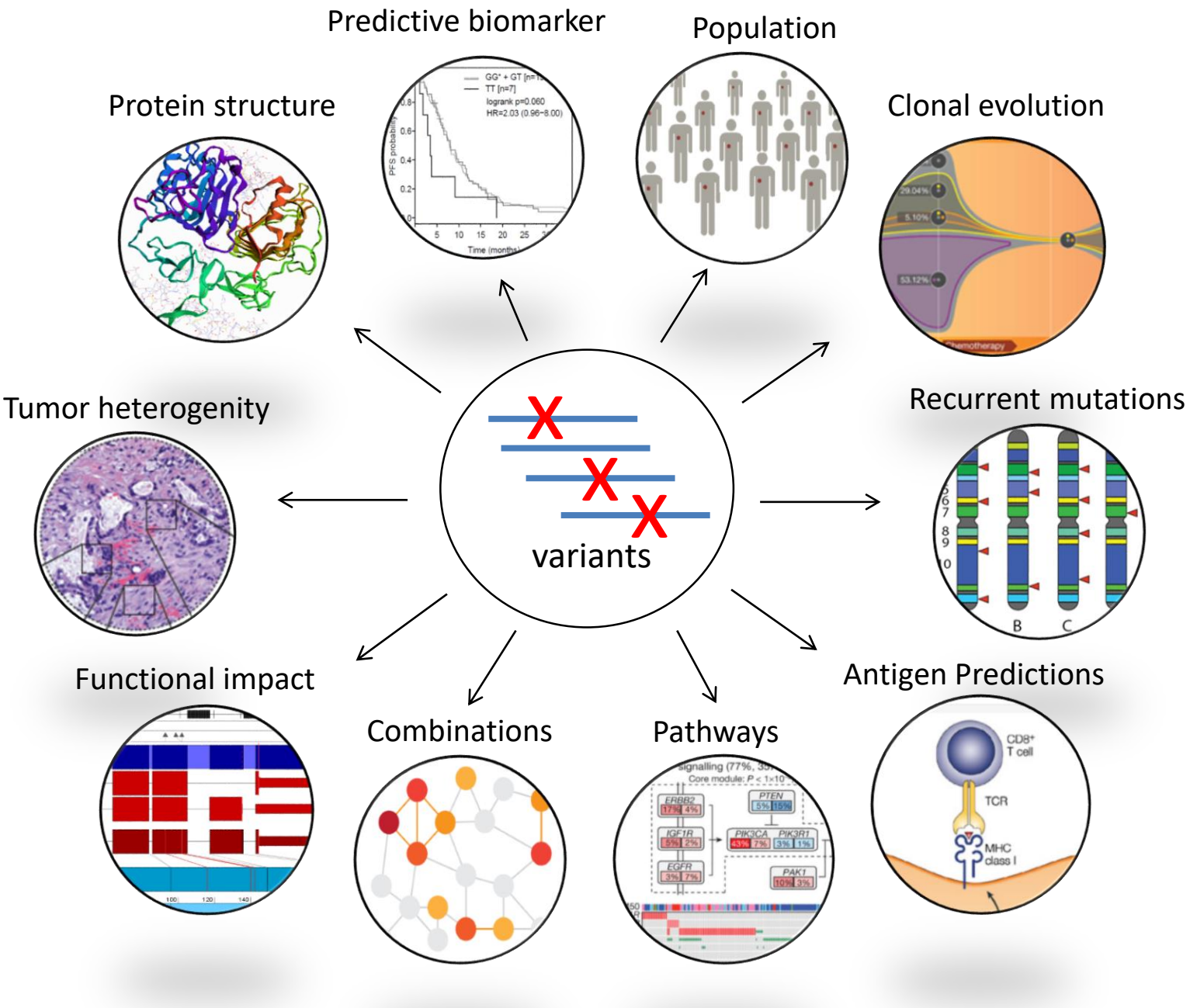
WM8 Bioinformatics

Analyses and interpretation of DNA variants

Hubert Hackl

Analyses and interpretation of DNA variants



Definitions

Mutation: any changes (physical events) made in the sequence of DNA also called variants

Point mutation: any changes made to a single nucleotide in the DNA sequence also called single nucleotide variant (SNV)

Single-nucleotide polymorphism (SNP): mutations or variants in context of a population (previously an arbitrary cutoff of >1% was used)

Indel: short (only a few bases) insertion or deletion in the DNA sequence

Structural DNA variation: long insertions deletion translocations of the DNA (genome, chromosomes)

Fusion gene: a hybrid gene formed from two previously separate genes as a result of translocation in the cancer genome (e.g. BCR-ABL in leukemia).

Definitions

Non-synonymous mutation: amino acid sequence of protein is altered

Synonymous mutation: amino acid sequence of protein is not altered

Functional variant: affects the molecular function of a protein (as a gain, loss or switch of function).

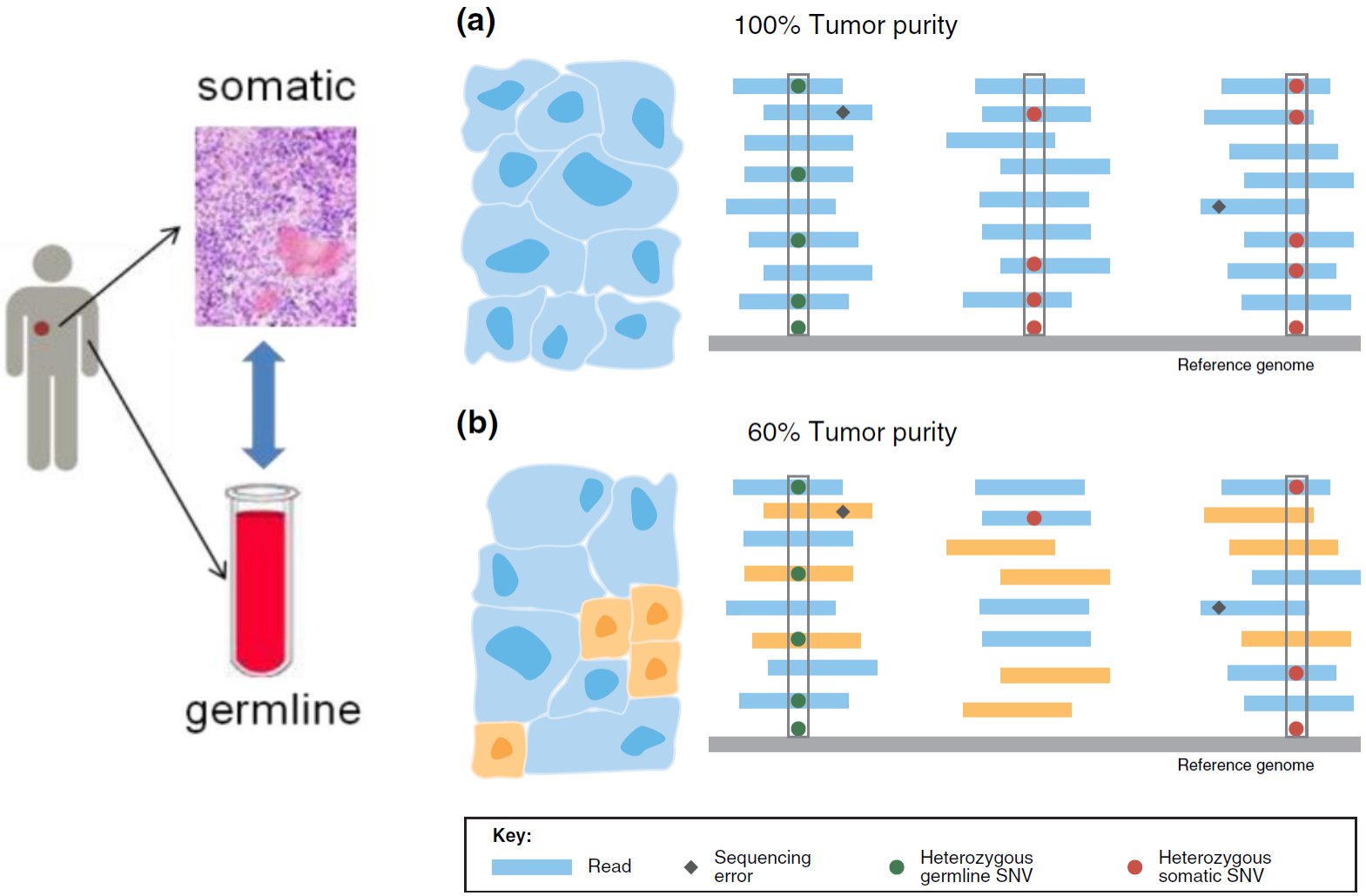
Nonfunctional variant: not appreciably affect the molecular function of a protein.

Driver variant: selective advantage to a particular tumor cell, i.e. often mutations in oncogenes or tumor suppressor genes

Passenger variant: no selective advantage

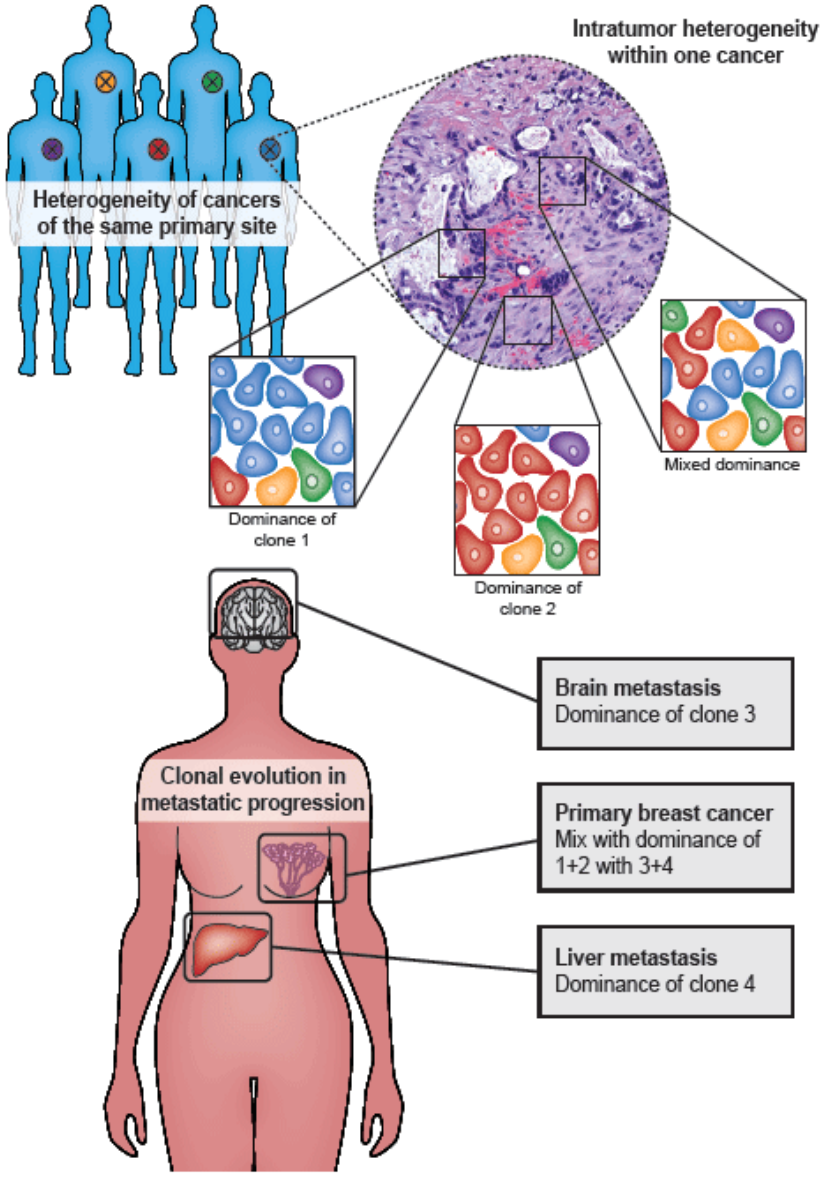
*A substitution might dramatically affect the function of a protein **without** providing any selective advantage to the tumor (it is a functional passenger variant).*

Somatic mutation detection in tumor samples

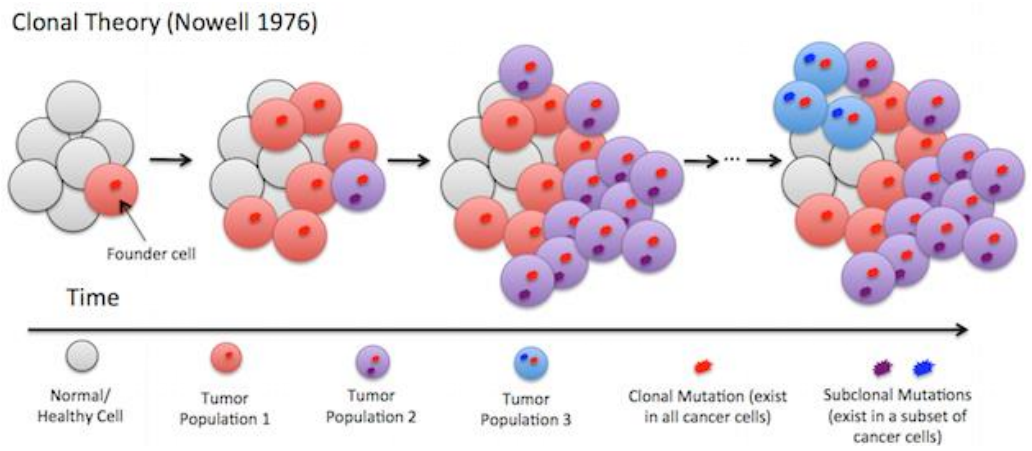


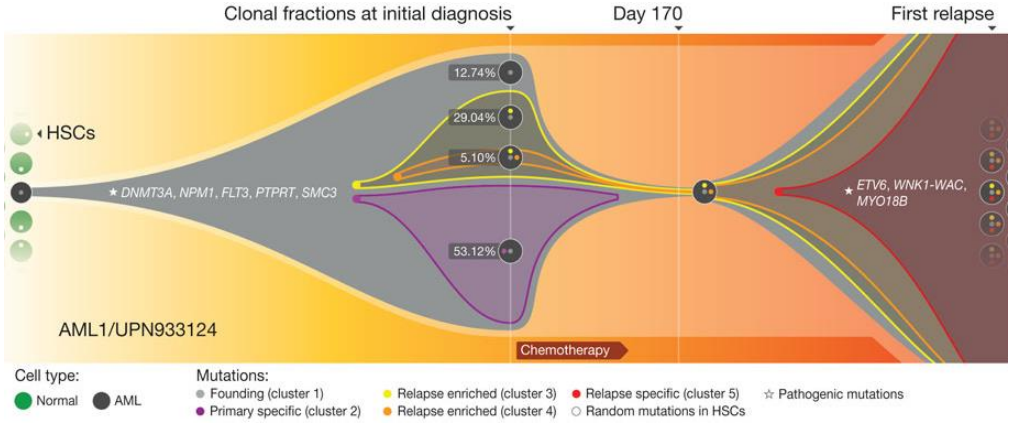
Raphael et al. Genome Med. 2014

Intratumor heterogenity and clonal evolution



Intratumor heterogenity and clonal evolution





Number of chromosomes in a cell
Diploid: proportion of cancer cells have the same number of chromosomes as normal, healthy cells (2 sets 23 each)
Aneuploid: too many or too few chromosomes, cancer may be more aggressive

Tumor purity: percentage of all cancer cells in tumor sample

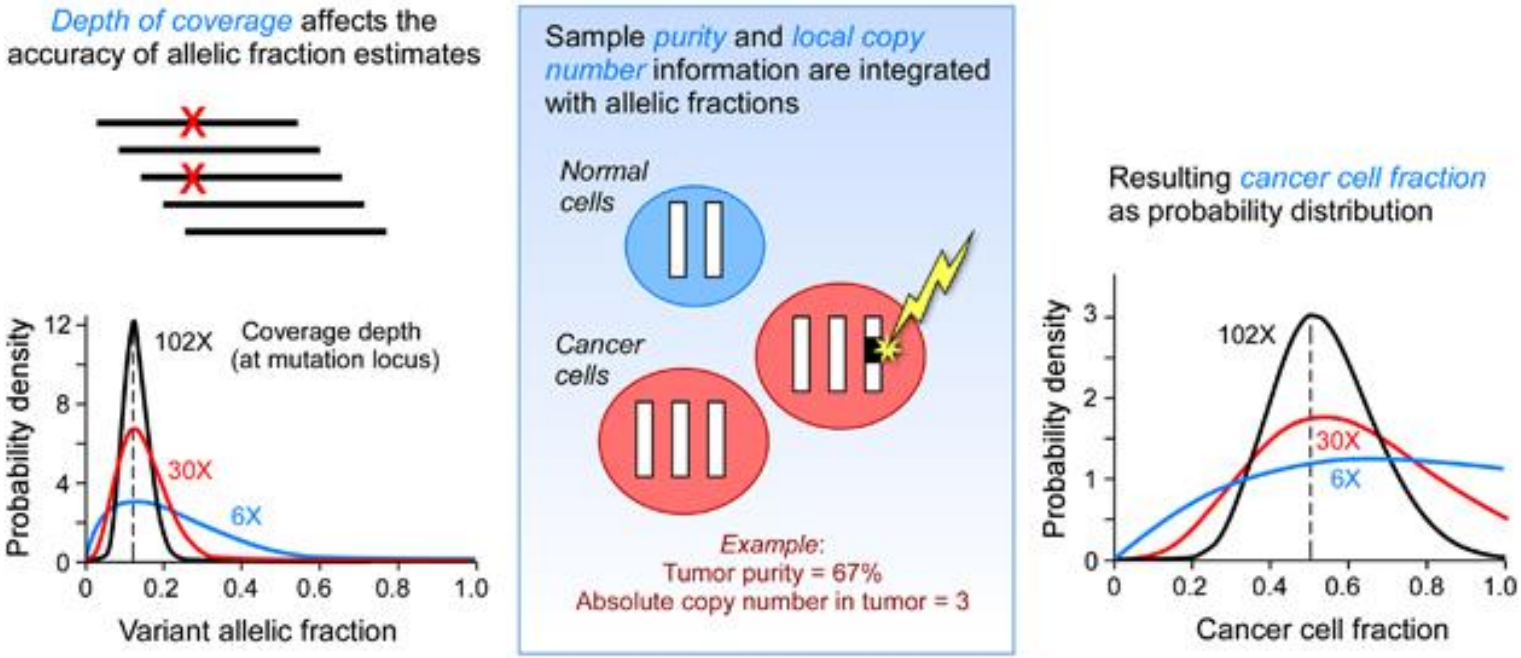
Variant allele fraction: Fraction of alleles (DNA molecules) from a locus that carry a mutation (also the expected fraction of supporting reads)

Variant allele frequency: Frequency of variant allele in population

Clonal: all cells in a tumor sample have a specific mutation
Subclonal: only a subgroup of all cells in a tumor sample have a specific mutation

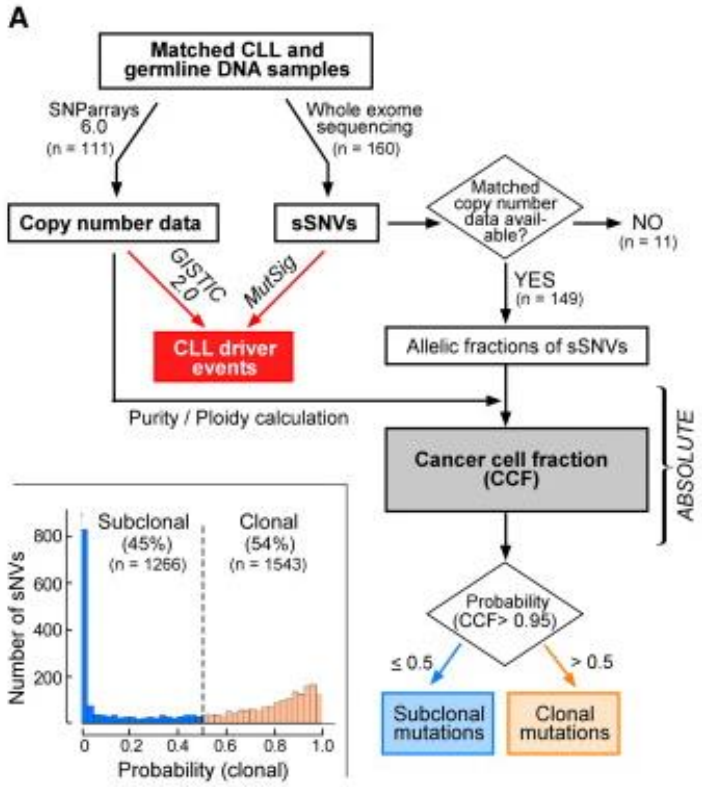
Cancer cell fraction: fraction of cancer cells with mutation

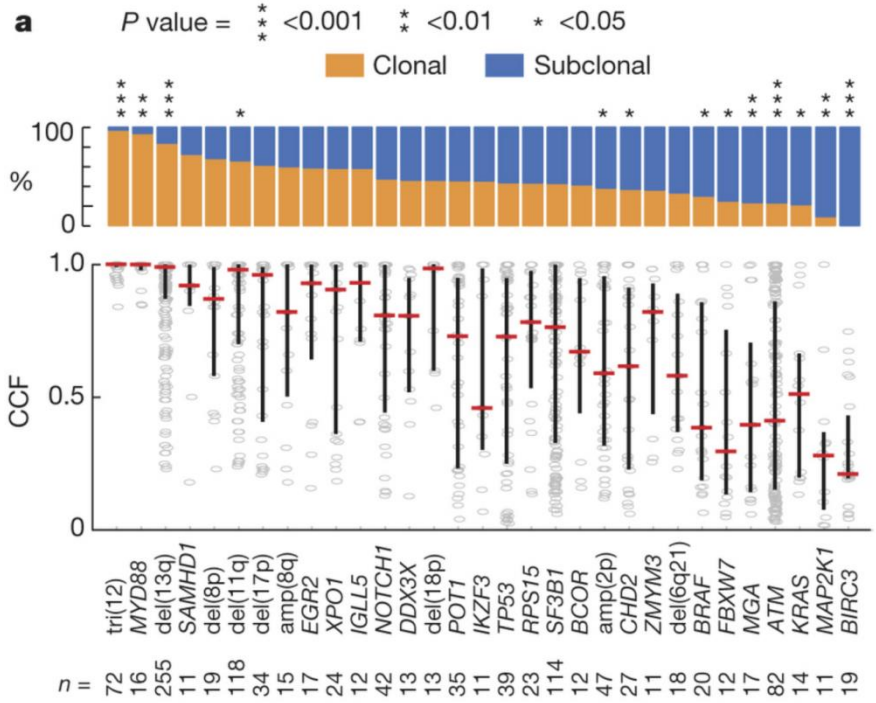
Depth of coverage



Landau et al , Leukemia, 2013

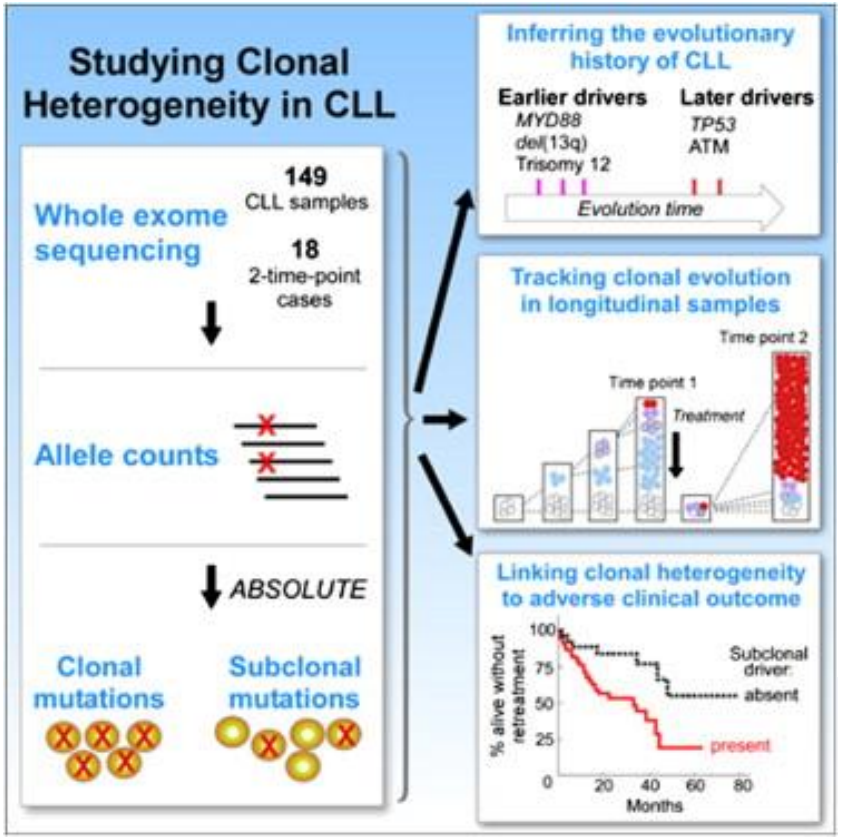
Clonal-subclonal variants; Cancer cell fraction (CCF)

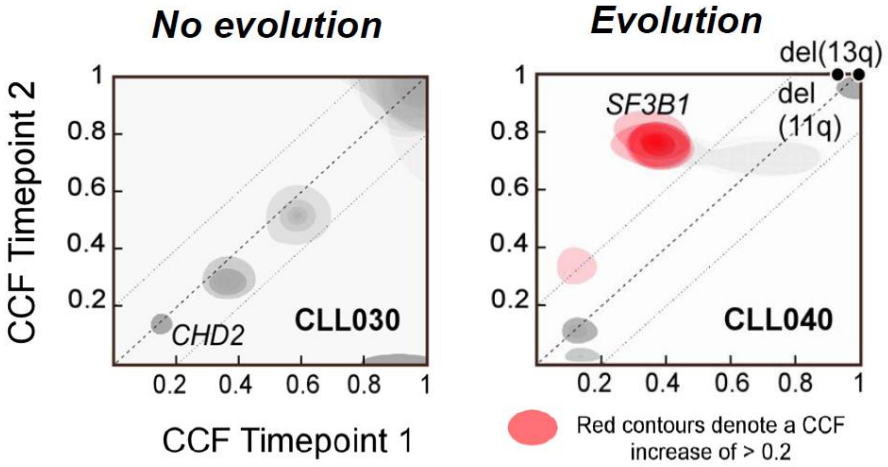




Landau et al , Nature, 2015

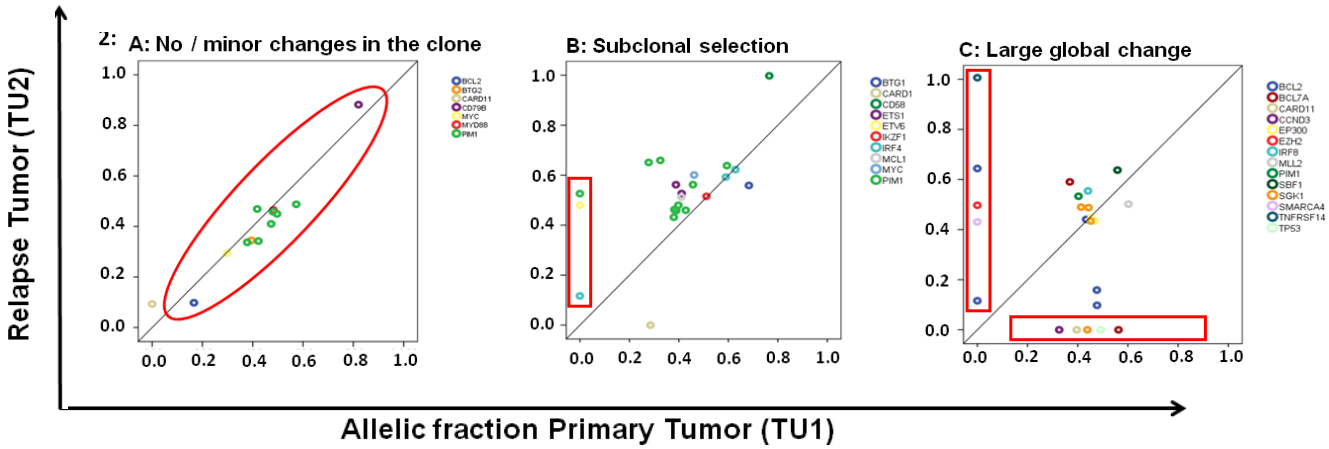
Patterns of clonal evolution

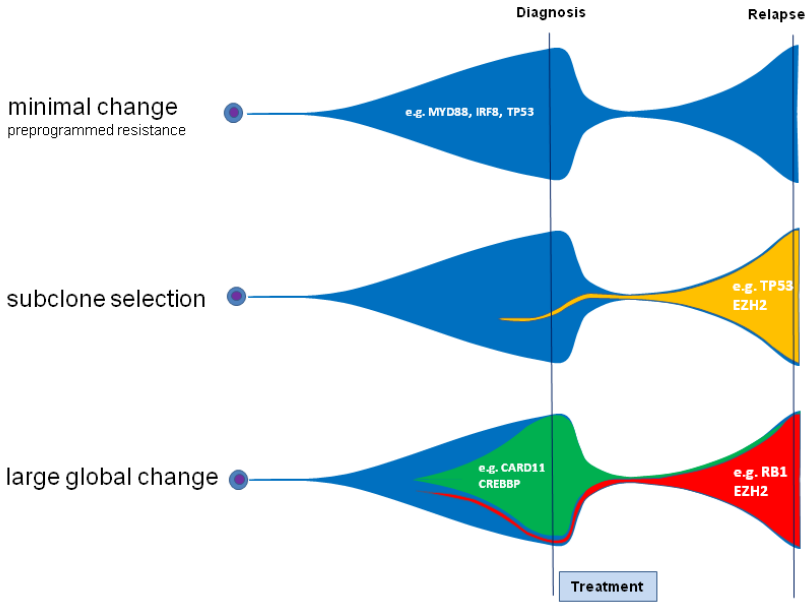




Landau et al , Nature, 2015

Example: Targeted exome sequencing in DBLCL





Melchardt et al, Oncotarget, 2016 (in press)

Databases for variants/mutations/SNPs

dbSNP
The Human Genome Mutation Database (HGMD)
OMIM
SNPedia
1000 Genomes
HapMap
GnomAD

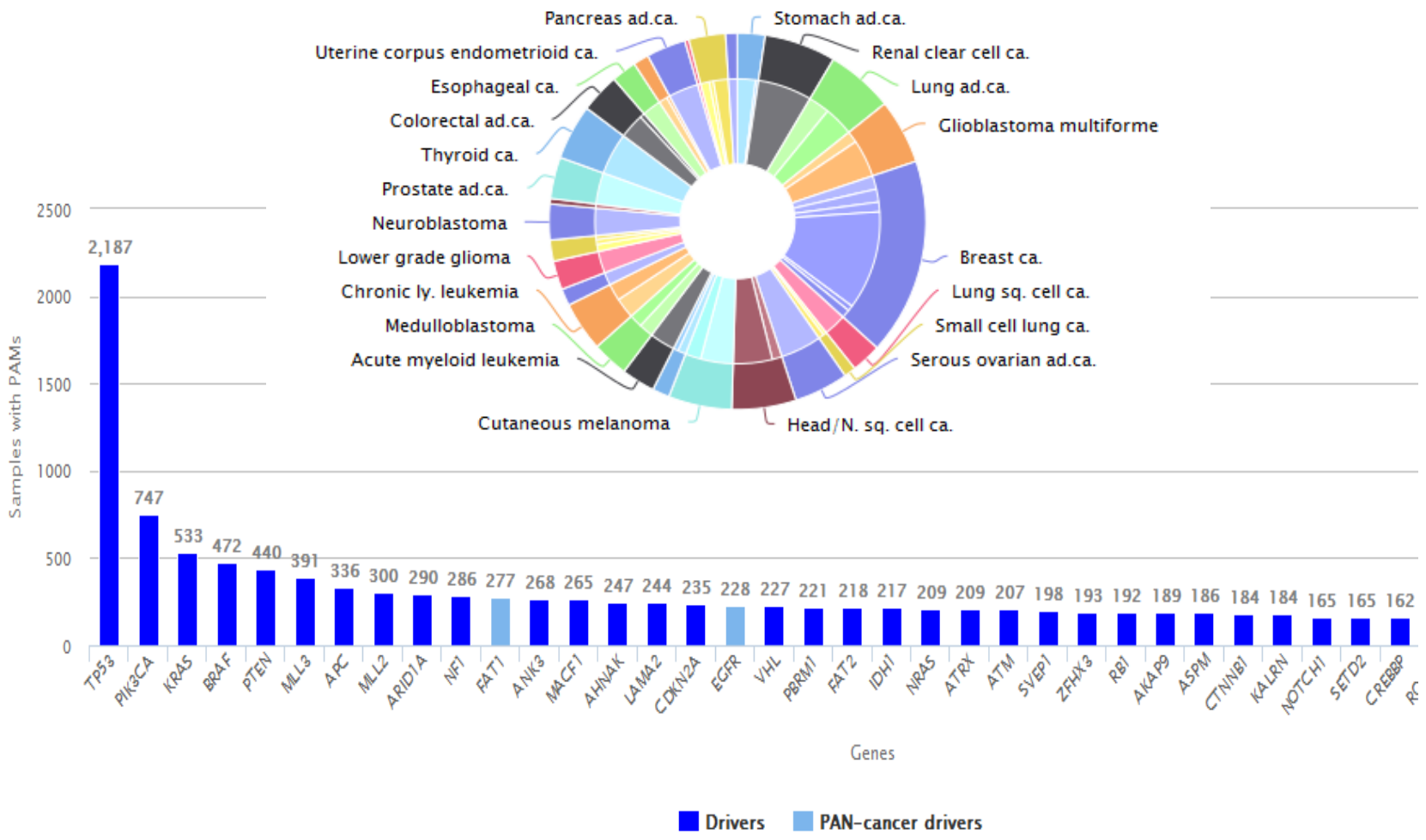
Specific for cancer:

Catalogue of Somatic Mutations in Cancer (COSMIC)
IntOGen
cBioPortal

Clinical Interpretation of Variants in Cancer

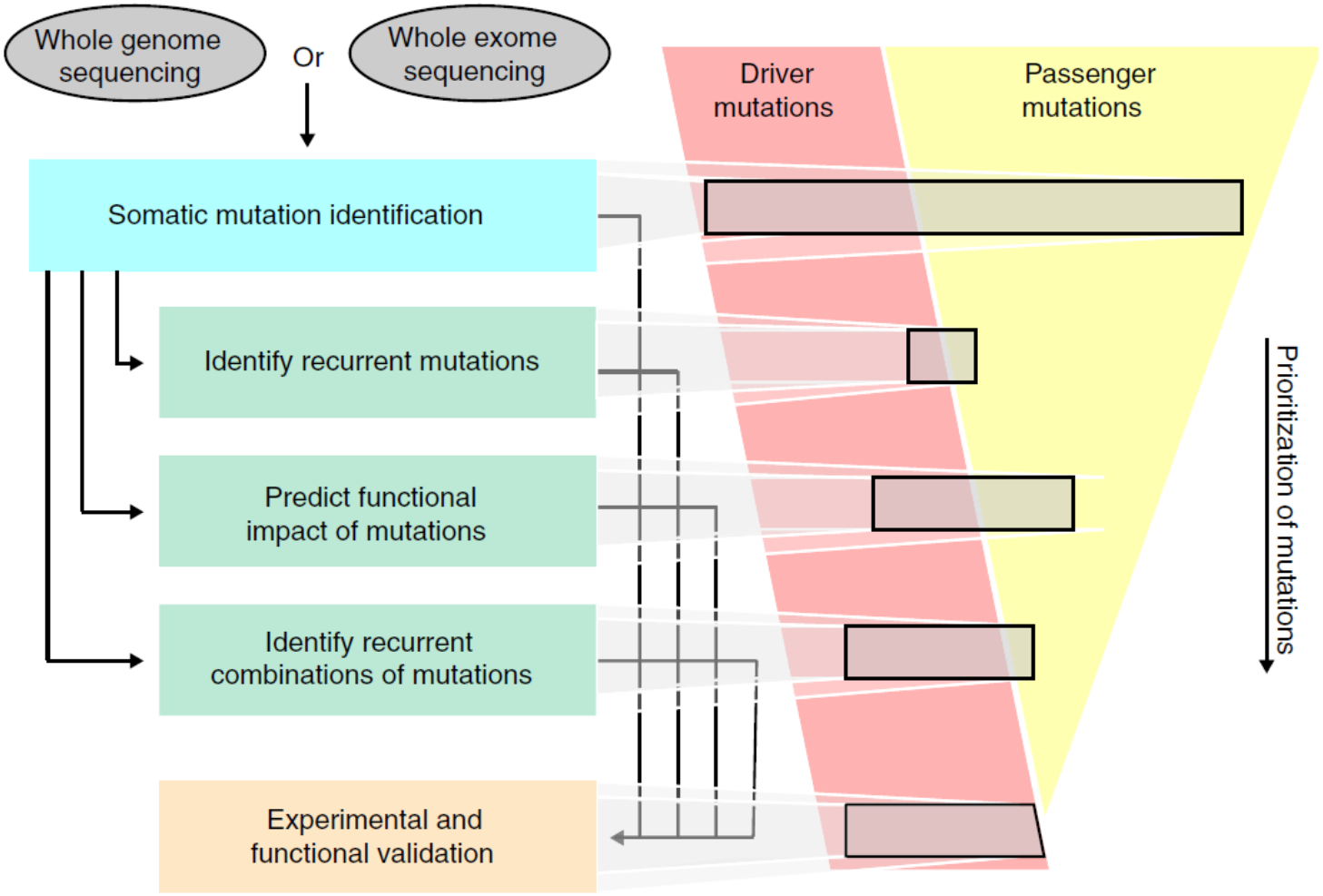
BaseSpace Knowledge Network
Cancer Driver Log (CanDL)
Cancer Genome Interpreter
Clinical Interpretation of Variants in Cancer (CIViC)
ClinVar
COSMIC Drug Resistance Curation
Database of Curated Mutations (DoCM)
Database of evidence for precision oncology (DEPO)
Gene Drug Knowledge Database
Genomics of Drug Sensitivity (GDSC)
HemOnc
JAX Clinical Knowledgebase
My Cancer Genome
OncoKB
Personalized Cancer Therapy (PCT)
Pharmacogenomics Knowledgebase (PharmGKB)
Precision Medicine Knowledgebase (PMKB)
Therapeutic target database
Variant Interpretation for Cancer Consortium (VICC)

Long tail distribution of cancer genes



http://www.intogen.org

Identifying driver mutations



Raphael et al. Genome Med. 2014

Map mutations to functional genomic features and annotation

Oncotator (Funcotator)
Ensembl Variant Effect Predictor (VEP)
ANNOVAR
snpEff

GPViz (visualization)

Oncotator

What is Oncotator?

Oncotator is a web application for annotating human genomic point mutations and indels with data relevant to cancer researchers. Annotations are aggregated from the following resources:

Genomic Annotations

- Gene, transcript, and functional consequence annotations using GENCODE for hg19.
- Reference sequence around a variant.
- GC content around a variant.
- Human DNA Repair Gene annotations from Wood et al.

Protein Annotations

- Site-specific protein annotations from UniProt.
- Functional impact predictions from dbNSFP.

Cancer Variant Annotations

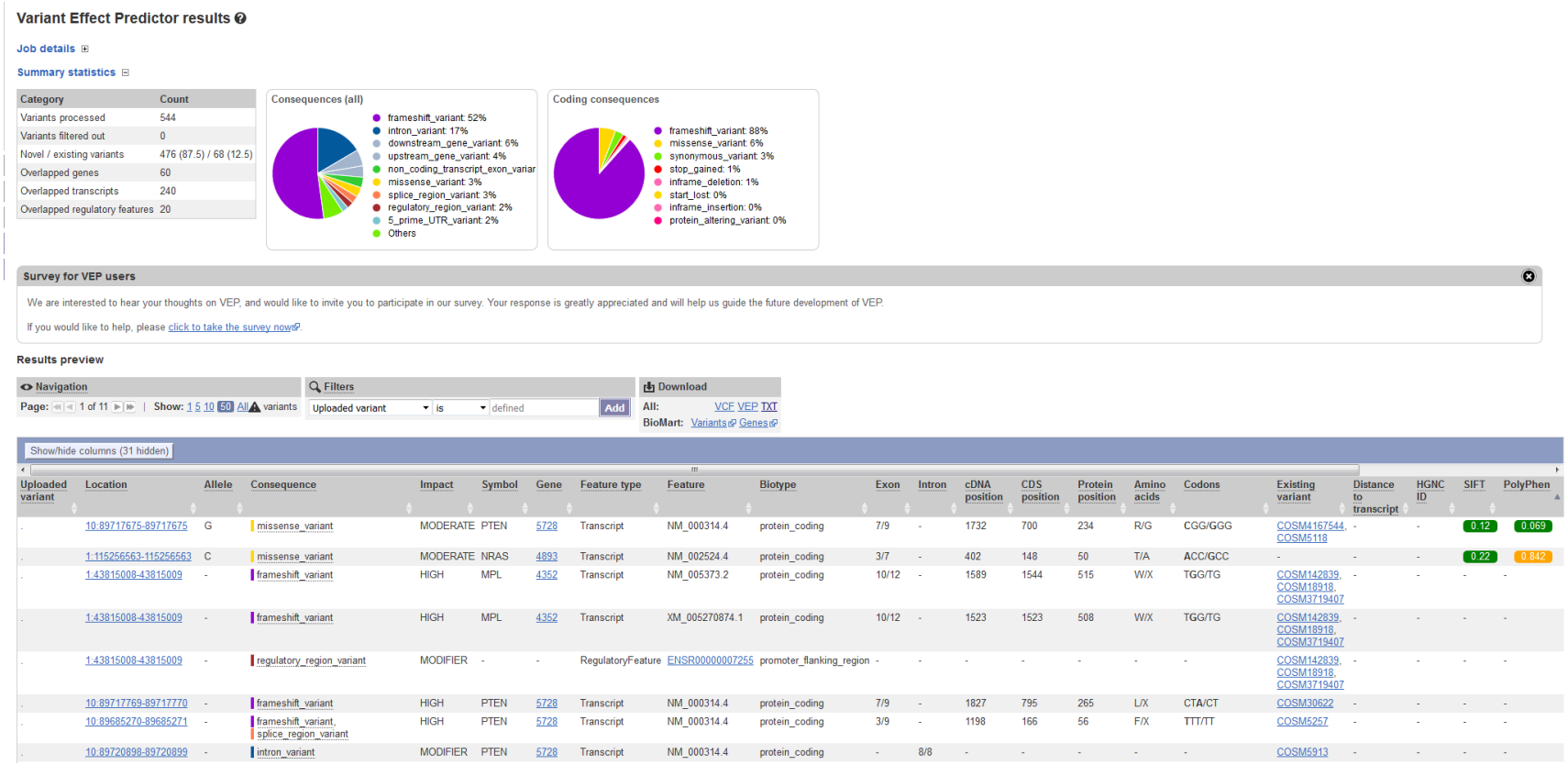
- Observed cancer mutation frequency annotations from COSMIC.
- Cancer gene and mutation annotations from the Cancer GenCensus.
- Overlapping mutations from the Cancer Cell Line Encyclopedia.
- Cancer gene annotations from the Familial Cancer Database.
- Cancer variant annotations from ClinVar.

Non-Cancer Variant Annotations

- Common SNP annotations from dbSNP.
- Variant annotations from 1000 Genomes.
- Variant annotations from NHLBI GO Exome Sequencing Project (ESP).

http://www.broadinstitute.org/oncotator/

Ensembl Variant Effect Predictor (VEP)



Variant Effect Predictor results

Job details

Summary statistics

Category	Count
Variants processed	544
Variants filtered out	0
Novel / existing variants	476 (87.5) / 68 (12.5)
Overlapped genes	60
Overlapped transcripts	240
Overlapped regulatory features	20

Consequences (all)

- frameshift_variant: 52%
- intron_variant: 17%
- downstream_gene_variant: 6%
- upstream_gene_variant: 4%
- non_coding_transcript_exon_variant
- missense_variant: 3%
- splice_region_variant: 3%
- regulatory_region_variant: 2%
- 5_prime_UTR_variant: 2%
- Others

Coding consequences

- frameshift_variant: 88%
- missense_variant: 6%
- synonymous_variant: 3%
- stop_gained: 1%
- inframe_deletion: 1%
- start_lost: 0%
- inframe_insertion: 0%
- protein_altering_variant: 0%

Survey for VEP users

We are interested to hear your thoughts on VEP, and would like to invite you to participate in our survey. Your response is greatly appreciated and will help us guide the future development of VEP.

If you would like to help, please click to take the survey now.

Results preview

Navigation — Page: 1 of 11 | Show: 1 5 10 50 All variants

Filters — Uploaded variant is defined Add

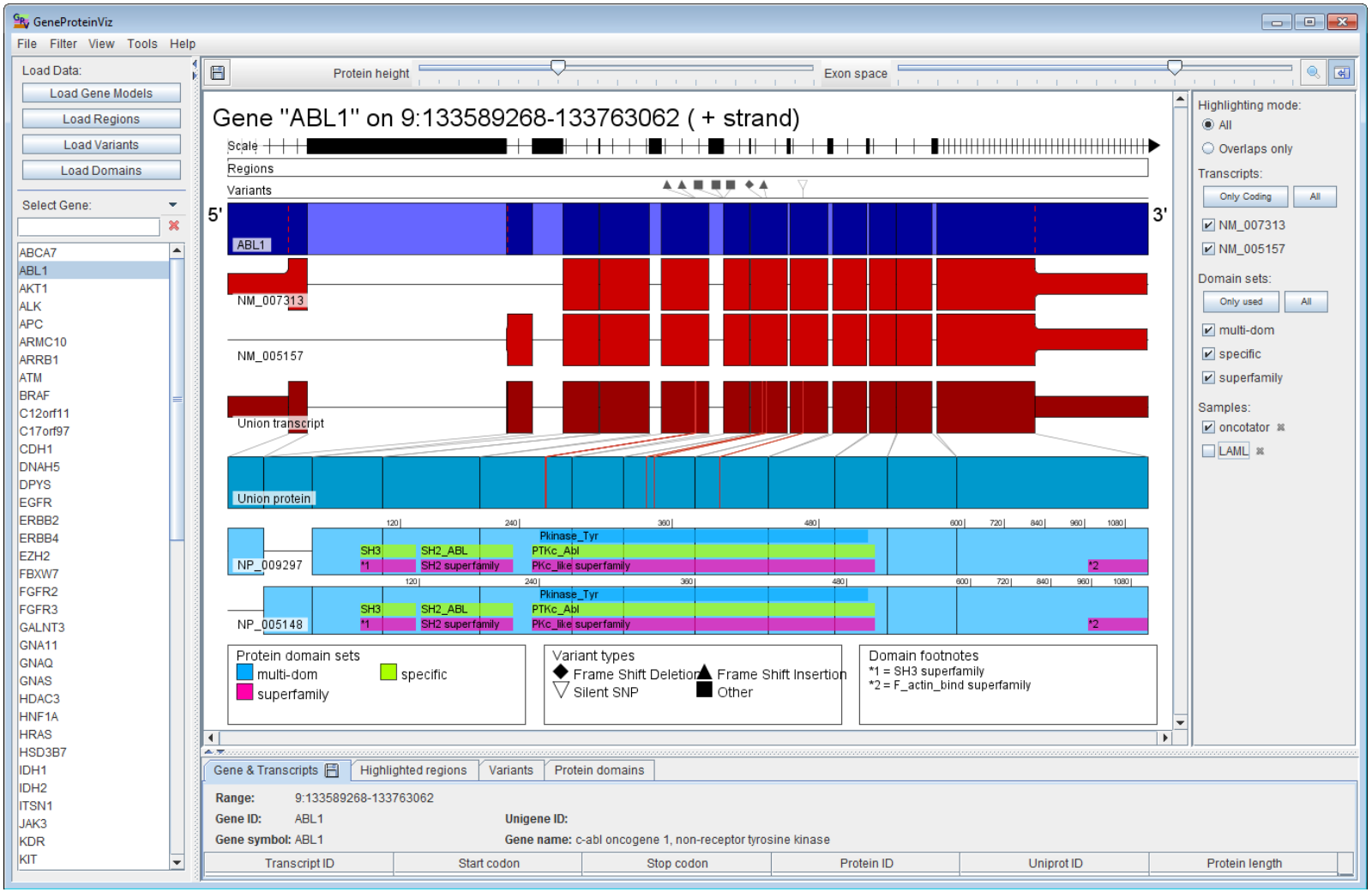
Download — All: VCF VEP TXT; BioMart: Variants Genes

Show/hide columns (31 hidden)

Uploaded variant	Location	Allele	Consequence	Impact	Symbol	Gene	Feature type	Feature	Biotype	Exon	Intron	cDNA position	CDS position	Protein position	Amino acids	Codons	Existing variant	Distance to transcript	HGNC ID	SIFT	PolyPhen
.	10:89717675-89717675	G	missense_variant	MODERATE	PTEN	5728	Transcript	NM_000314.4	protein_coding	7/9	-	1732	700	234	R/G	CGG/GGG	COSM4167544, COSM5118	-	-	0.12	0.069
.	1:115256563-115256563	C	missense_variant	MODERATE	NRAS	4893	Transcript	NM_002524.4	protein_coding	3/7	-	402	148	50	T/A	ACC/GCC	-	-	-	0.22	0.842
.	1:43815008-43815009	-	frameshift_variant	HIGH	MPL	4352	Transcript	NM_005373.2	protein_coding	10/12	-	1589	1544	515	W/X	TGG/TG	COSM142839, COSM18918, COSM3719407	-	-	-	-
.	1:43815008-43815009	-	frameshift_variant	HIGH	MPL	4352	Transcript	XM_005270874.1	protein_coding	10/12	-	1523	1523	508	W/X	TGG/TG	COSM142839, COSM18918, COSM3719407	-	-	-	-
.	1:43815008-43815009	-	regulatory_region_variant	MODIFIER	-	-	RegulatoryFeature	ENSR00000007255	promoter_flanking_region	-	-	-	-	-	-	-	COSM142839, COSM18918, COSM3719407	-	-	-	-
.	10:89717769-89717770	-	frameshift_variant	HIGH	PTEN	5728	Transcript	NM_000314.4	protein_coding	7/9	-	1827	795	265	L/X	CTA/CT	COSM30622	-	-	-	-
.	10:89685270-89685271	-	frameshift_variant, splice_region_variant	HIGH	PTEN	5728	Transcript	NM_000314.4	protein_coding	3/9	-	1198	166	56	F/X	TTT/TT	COSM5257	-	-	-	-
.	10:89720898-89720899	-	intron_variant	MODIFIER	PTEN	5728	Transcript	NM_000314.4	protein_coding	-	8/8	-	-	-	-	-	COSM5913	-	-	-	-

http://grch37.ensembl.org/Homo_sapiens/Tools/VEP

GPViz



http://icbi.at/software/gpviz/gpviz.shtml

Snajder et al, Bioinformatics 2013

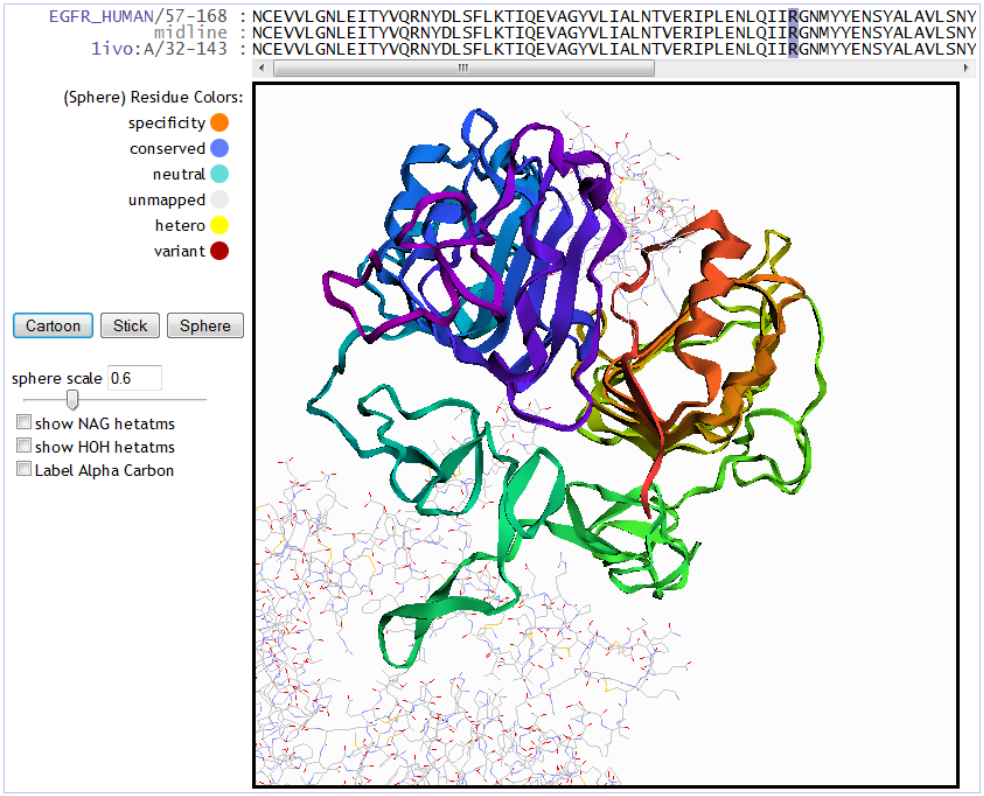
Effect on the function of a protein or regulatory element (functional impact)

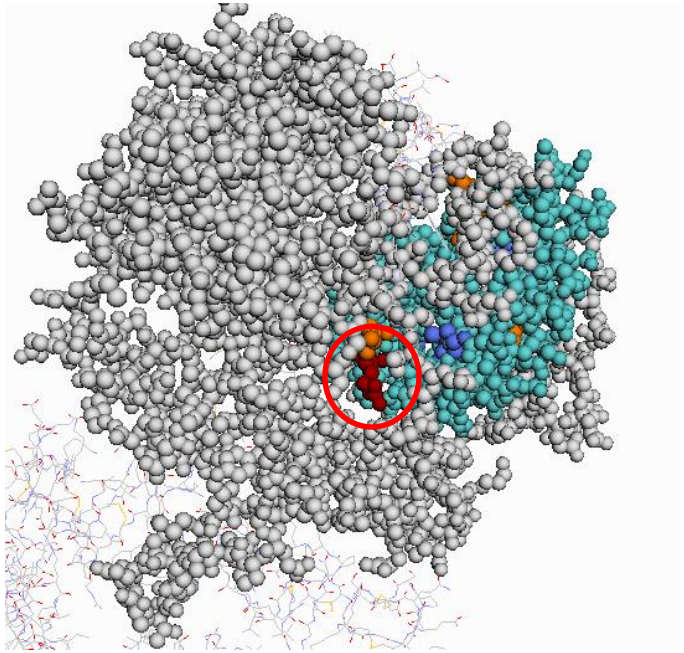
SIFT
PolyPhen2
CADD
CHASM/VEST
TransFIC
PredictSNP
MutationAssessor

MutationAssessor

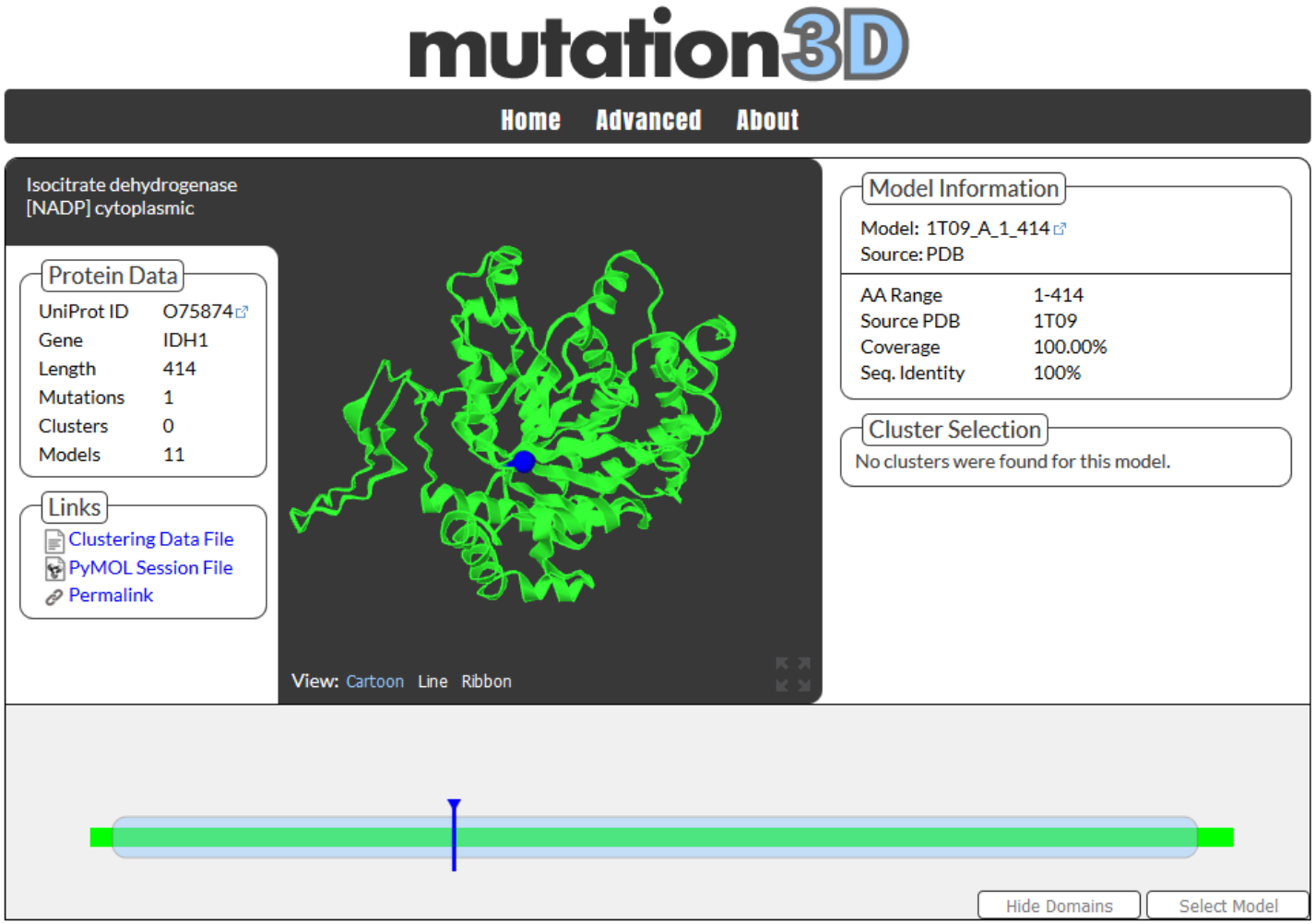
Chr Pos RefAll AltAll

Mutation	AA variant	Gene	MSA	PDB	Func. Impact	FI score	Uniprot	Refseq	MSA height
2,212288939,C,G	G936A	ERBB4	msa	pdb	medium	2.4	ERBB4_HUMAN	XP_005246432	1684

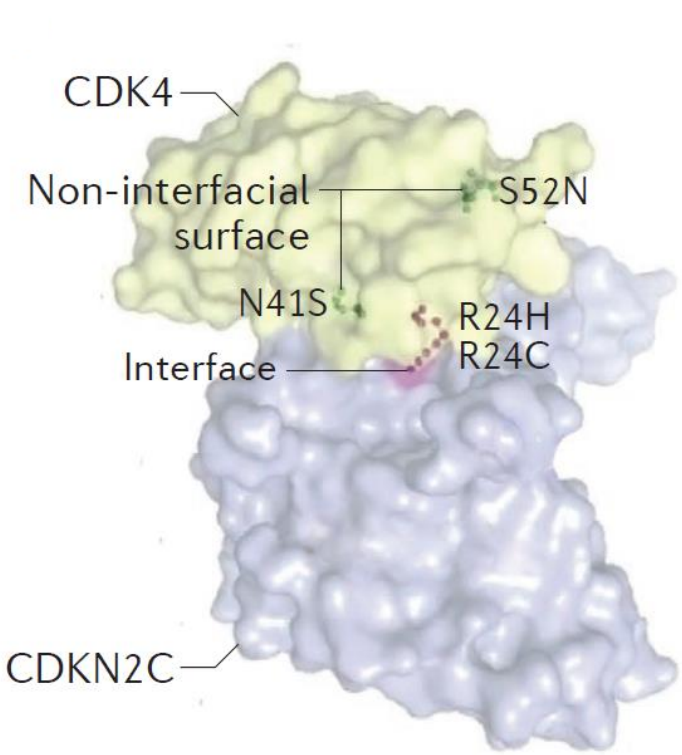


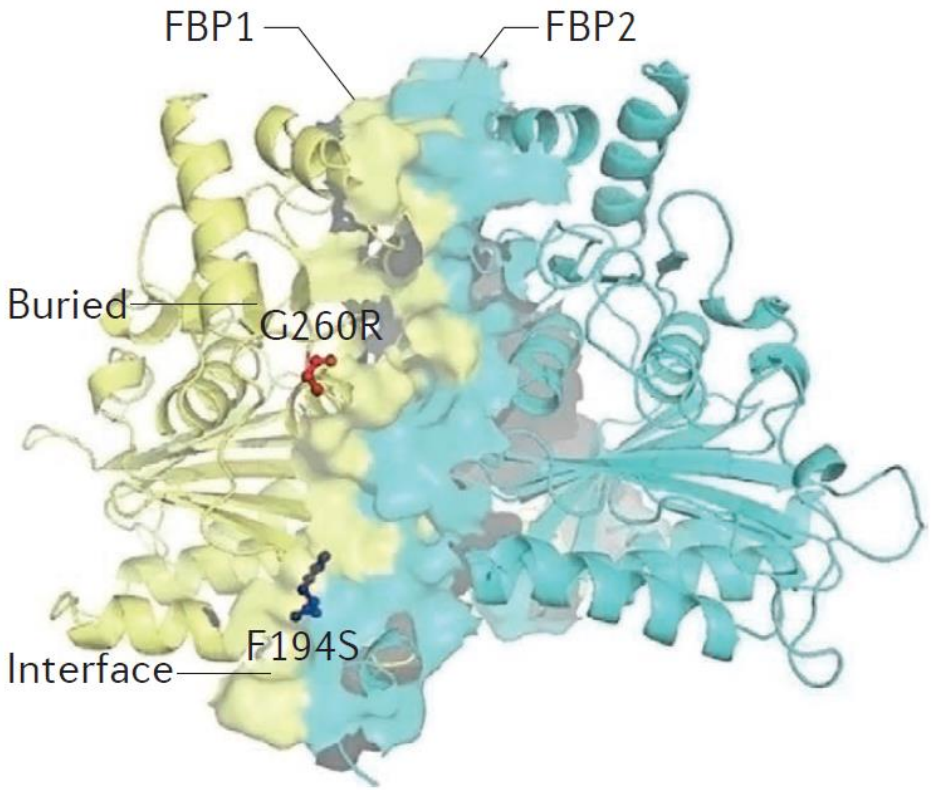


Mutation3D


mutation3D
Home
Advanced
About
Isocitrate dehydrogenase [NADP] cytoplasmic
Protein Data
UniProt ID O75874
Gene IDH1
Length 414
Mutations 1
Clusters 0
Models 11
Links
Clustering Data File
PyMOL Session File
Permalink
View: Cartoon Line Ribbon
Model Information
Model: 1T09_A_1_414
Source: PDB
AA Range 1-414
Source PDB 1T09
Coverage 100.00%
Seq. Identity 100%
Cluster Selection
No clusters were found for this model.
Hide Domains
Select Model

Effects on molecular interactions (edgetic mutations)



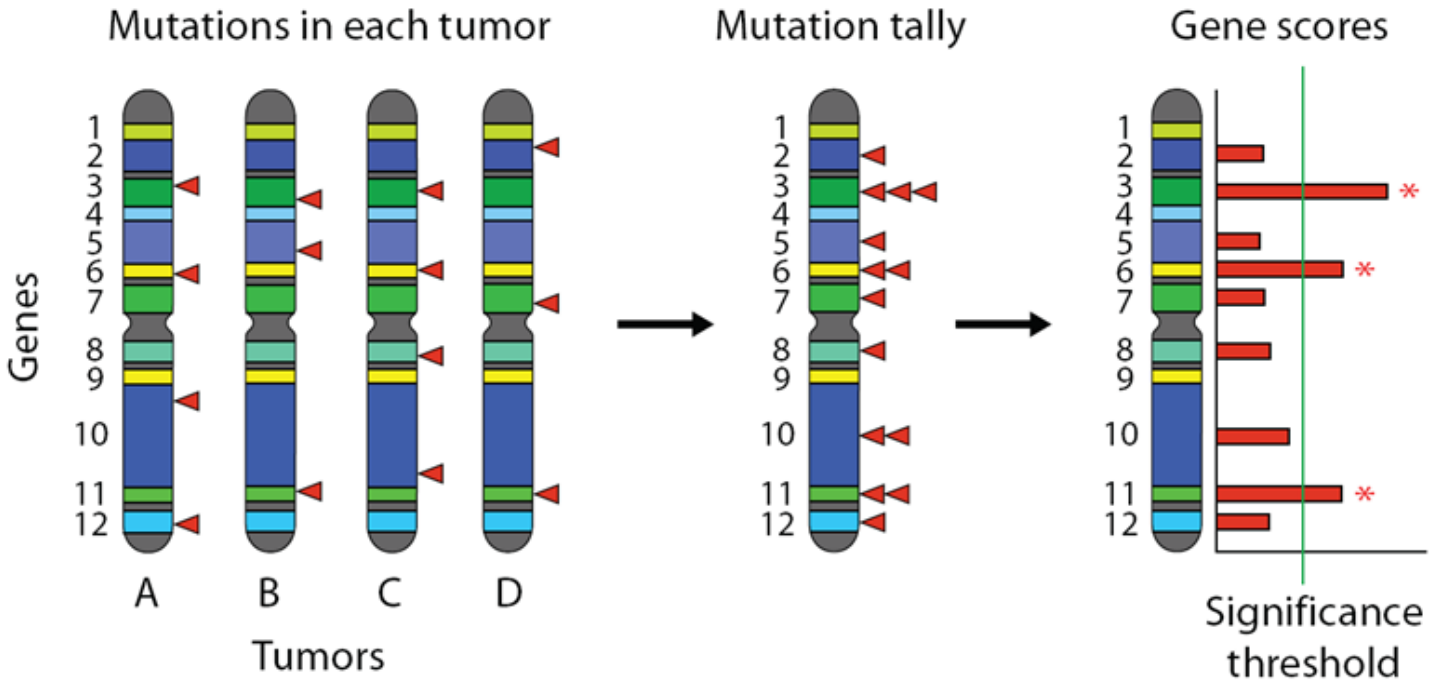


Yi et al. Nature Rev Genet. 2017

Detect mutations, genes and pathways recurrently altered or with overrepresentation of functional mutations

MutSig, MutSigCV
OncodriveFM
ActiveDriver
MuSIC
Mutually Exclusive Modules in cancer (MEMo)
HotNet

MutSig (MutSigCV)

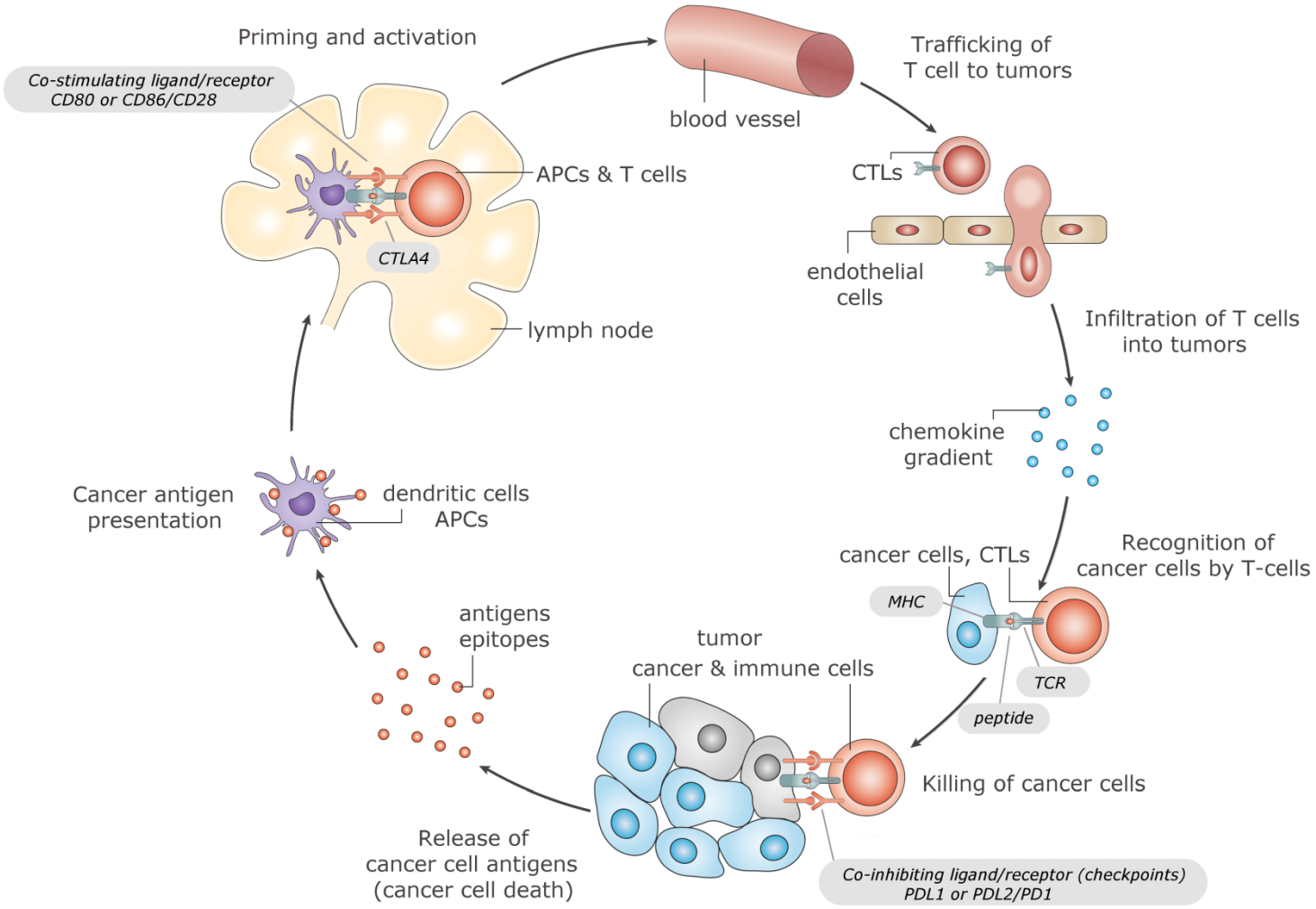


MutSig builds a model of the background mutation processes (BMR) that were at work during formation of the tumors, and it analyzes the mutations of each gene to identify genes that were mutated more often than expected by chance, given the background model.

MutSigCV (CV for 'covariate') improves the BMR estimation by pooling data from 'neighbor' genes with similar genomic properties such as DNA replication time, chromatin state (open/closed), and general level of transcription activity.

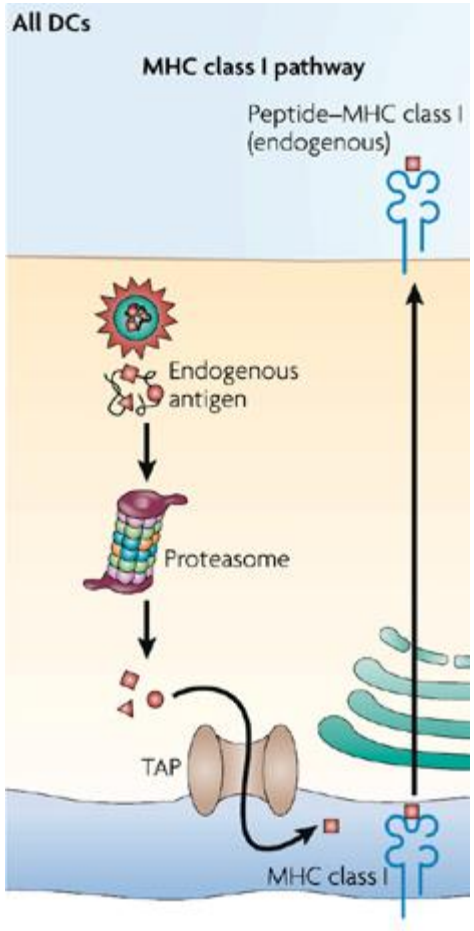
Lawrence, M. et al. Nature 2013

Antigen prediction (cancer immunity cycle)



Hackl H et al, Nature Reviews Genetics, 2016

Antigen prediction



In human MHC is called HLA

Somatic mutation in tumor ⟶ Peptide (length 8-11)

HLA-type ↘

Peptide ↗

NetMHCpan ⟶ epitope?

http://www.cbs.dtu.dk/services/NetMHCpan/

Predictive marker

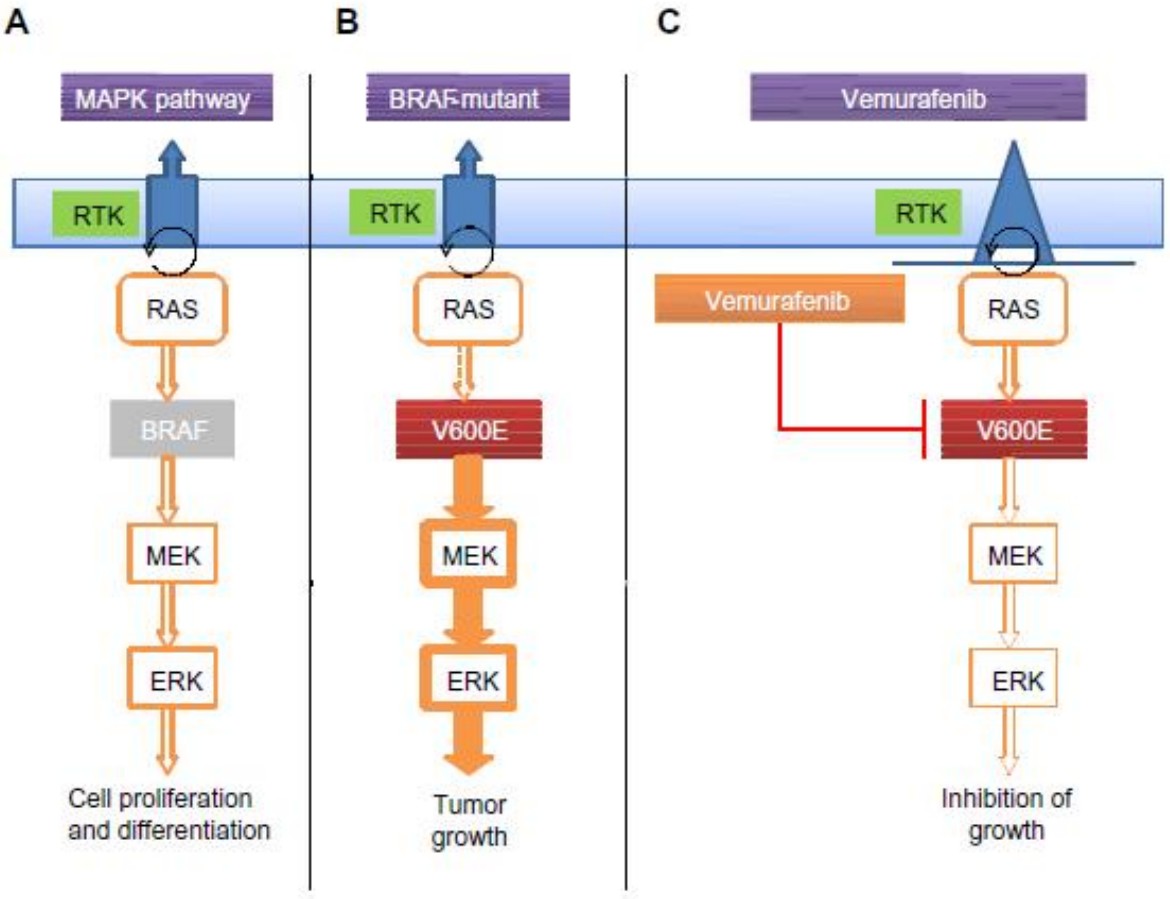
A **predictive marker** is a particular protein or gene that indicates sensitivity or resistance to a specific therapy.

The use of predictive markers is becoming increasingly relevant in cancer therapy as it allows for better identification of patients who will respond positively to the therapy.

Expression of estrogen and progesterone receptors can determine the benefits of hormone therapy, whilst the benefit of treating breast cancer patients with herceptin (Trastuzumab) is determined by the expression of HER2.

Predictive marker

Variants (mutations) might also used as predictive marker e.g. BRAF V600E mutations in melanoma:



Example: EDN1-variant as predictive marker for therapy with Bevacizumab in metastatic breast cancer

Bevacizumab is a recombinant humanized monoclonal antibody that blocks angiogenesis by inhibiting vascular endothelial growth factor A (VEGF-A). VEGF-A is a chemical signal that stimulates angiogenesis in a variety of diseases, especially in cancer.

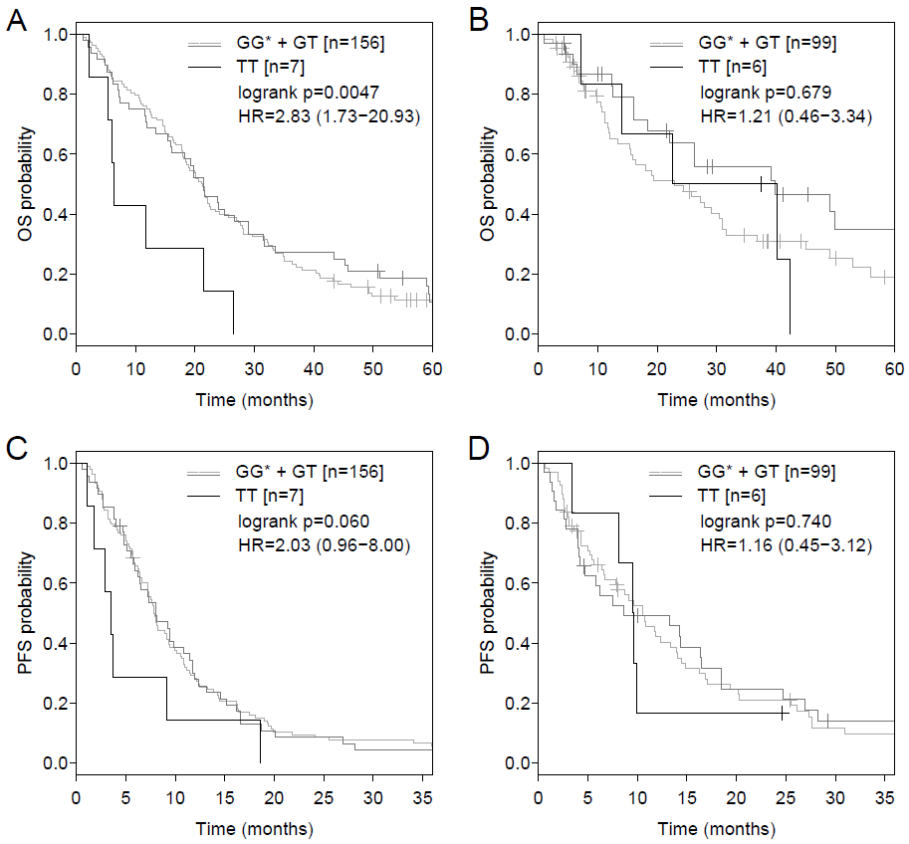
Genotype of 10 SNPs associated with VEGF and hypertension tested

Genotype					Bevacizumab cohort			
SNP	GENE	wt (+/+)	+/-	-/-	wt (+/+)	+/-	-/-	P*
rs13207351	VEGF-152 G/A	GG	GA	AA	92	63	8	0.62
rs2010963	VEGF-634 G/C	GG	GC	CC	13	85	65	0.055
rs3025039	VEGF-936 C/T	CC	CT	TT	113	41	7	0.30
rs1570360	VEGF-1154 G/A	GG	GA	AA	31	36	54	4.8×10^{-5}
rs833061	VEGF-1498 C/T	CC	CT	TT	35	89	39	0.29
rs699947	VEGF-2578 C/A	AA	CA	CC	33	89	41	0.28
rs9582036	VEGFR-1 C/A	CC	CA	AA	81	80	2	5.7×10^{-4}
rs13333226	UMOD A/G	GG	AG	AA	109	49	5	0.94
rs3754777	STK39 A/G	GG	AG	AA	121	37	5	0.45
rs5370	EDN1 G/T	GG	GT	TT	108	48	7	0.73
					Control cohort			
rs-5370	EDN1 G/T	GG	GT	TT	67	32	6	0.56

* Deviation from Hardy-Weinberg equilibrium (HWE)

Gampenrieder SP et al., Pharmacogenomics J, 2016

Example: EDN1-variant as predictive marker



This is a germline mutation, so can tested in blood!

Gampenrieder SP et al., Pharmacogenomics J, 2016