WM8- Bioinformatics Exercises

Analyses and interpretation of DNA variants

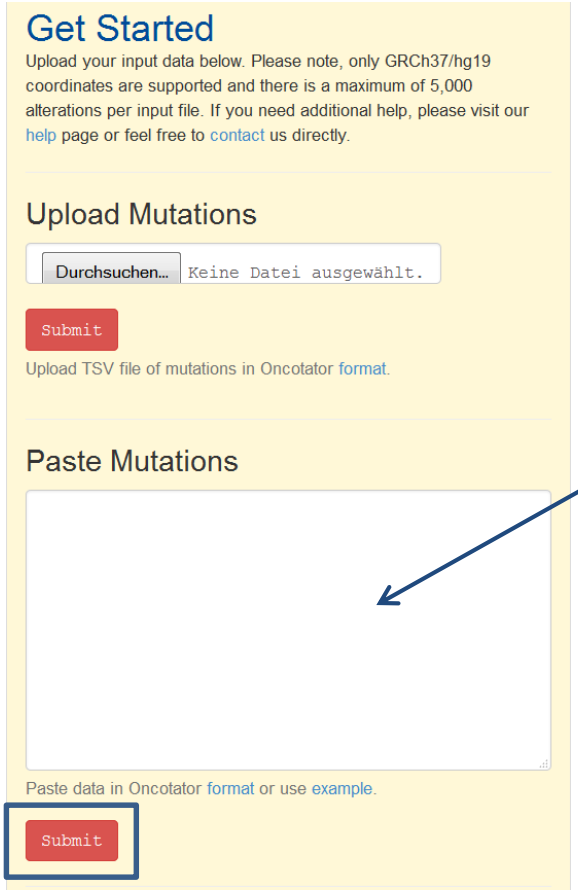
Hubert Hackl

Exercises

1. VEP, Oncotator (Funcotator)
2. MutSigCV
3. GPViz
4. IntoGen, cBioPortal, FannsDB (Condel, Cancer Genome Interpreter)
5. MutationAssessor, Mutation3D
6. Predict neoantigens (NetMHCpan 4.1)
7. Get correct qvalues for MutSigCV results of DLBCL patients, as for targeted sequencing (get_sig.R)
8. Get variant allele frequencies and compare with mapped reads in the IGV (get_VAF.R)
9. Compare MAF (VCF) files for variants on the same position (same type of variant, and same alternative allele) (compare_MAF.R)
10. Statistics and plot diagrams using R (maftools, MAF.R,stats.R)

Oncotator

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

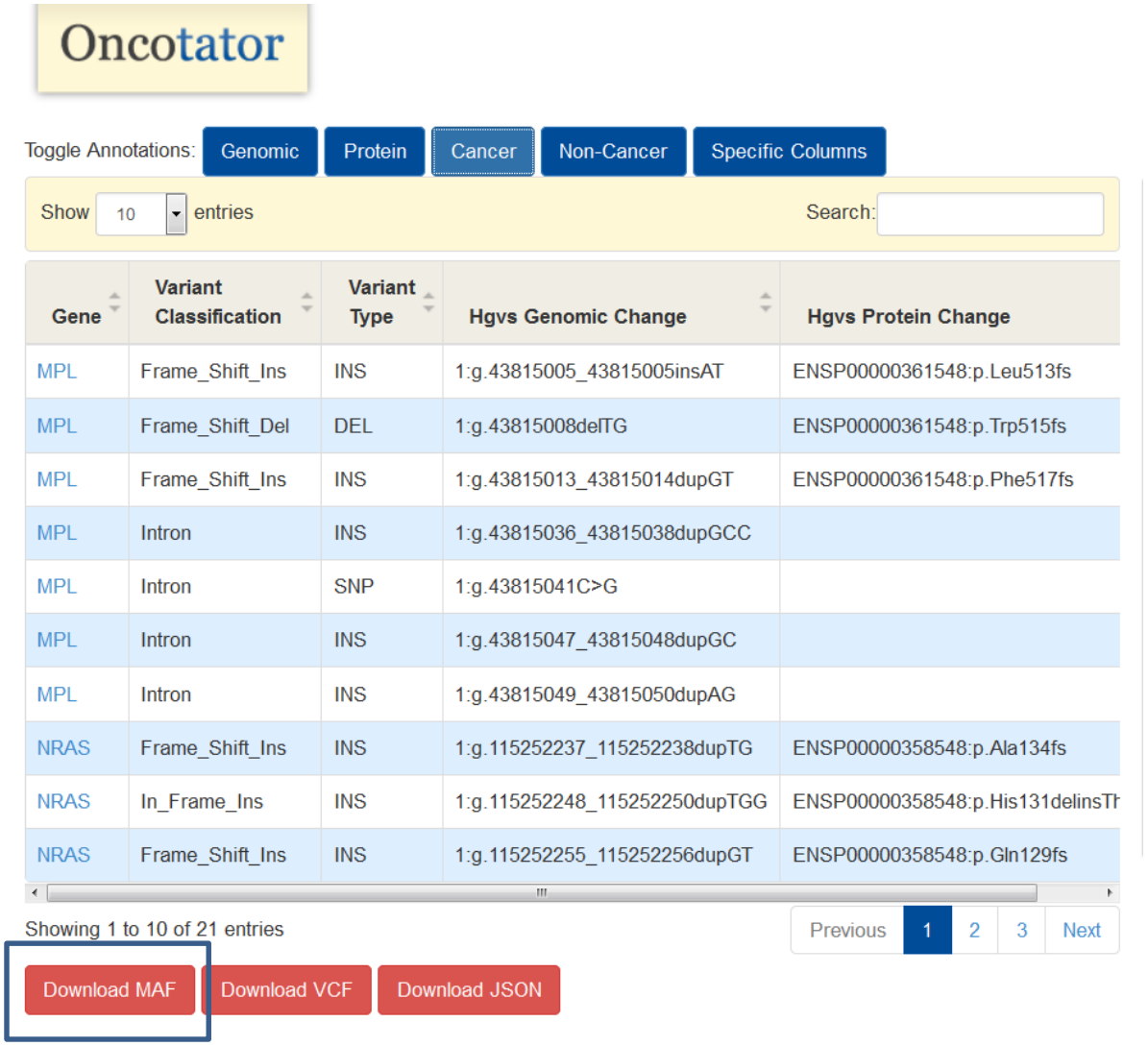


jurkat_mutations.vcf

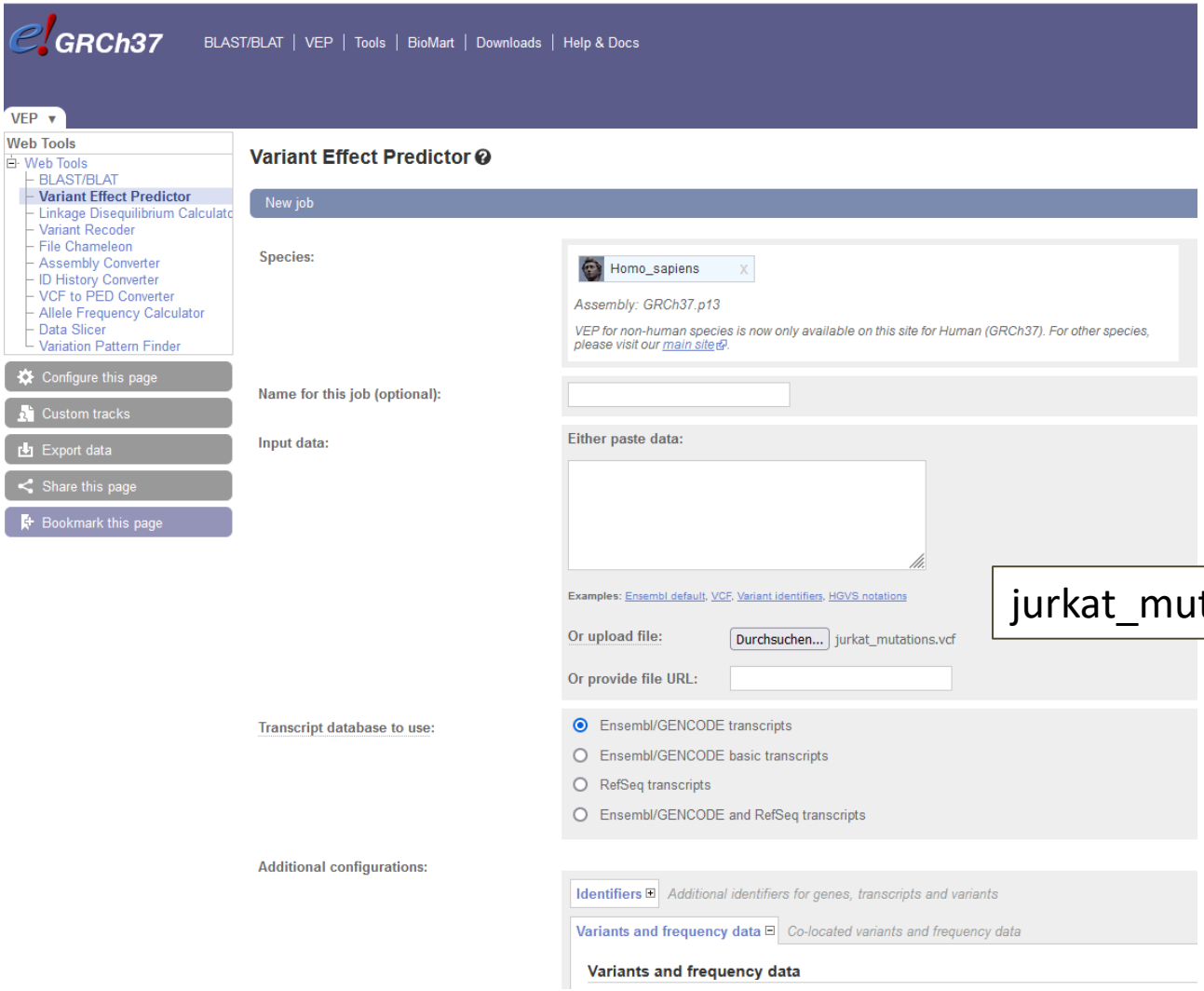
chr1	43815005	43815005	A	AT
chr1	43815008	43815008	TG	T
chr1	43815013	43815013	G	GT
chr1	43815036	43815036	G	GCC
chr1	43815041	43815041	C	G
chr1	43815047	43815047	G	GC
chr1	43815049	43815049	A	AG
chr1	115252237	115252237	T	TG

Now succeeded by Funcotator (GATK)

Oncotator

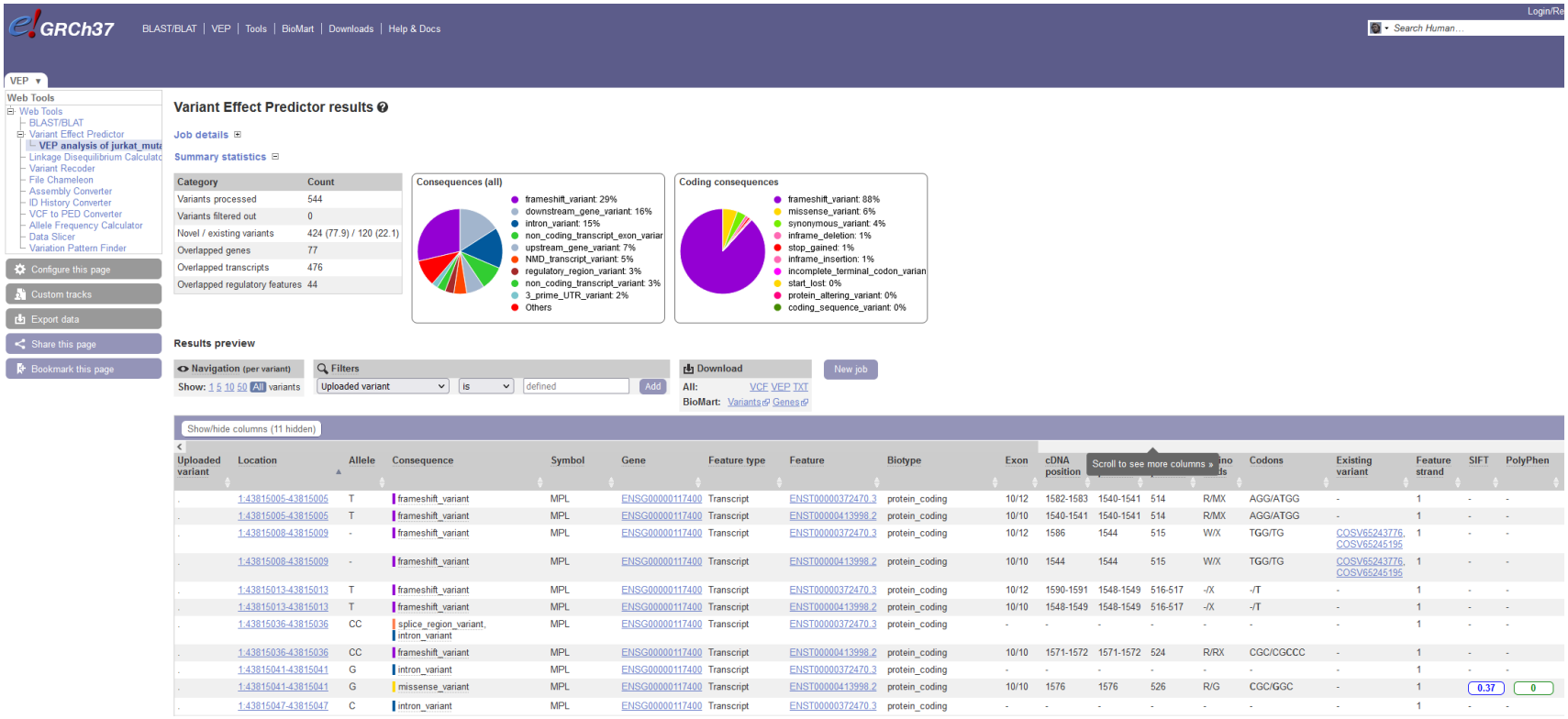

Oncotator
Toggle Annotations:
Genomic
Protein
Cancer
Non-Cancer
Specific Columns
Show 10 entries
Search:
Gene
Variant Classification
Variant Type
Hgvs Genomic Change
Hgvs Protein Change
MPL Frame_Shift_Ins INS 1:g.43815005_43815005insAT ENSP00000361548:p.Leu513fs
MPL Frame_Shift_Del DEL 1:g.43815008delTG ENSP00000361548:p.Trp515fs
MPL Frame_Shift_Ins INS 1:g.43815013_43815014dupGT ENSP00000361548:p.Phe517fs
MPL Intron INS 1:g.43815036_43815038dupGCC
MPL Intron SNP 1:g.43815041C>G
MPL Intron INS 1:g.43815047_43815048dupGC
MPL Intron INS 1:g.43815049_43815050dupAG
NRAS Frame_Shift_Ins INS 1:g.115252237_115252238dupTG ENSP00000358548:p.Ala134fs
NRAS In_Frame_Ins INS 1:g.115252248_115252250dupTGG ENSP00000358548:p.His131delinsTh
NRAS Frame_Shift_Ins INS 1:g.115252255_115252256dupGT ENSP00000358548:p.Gln129fs
Showing 1 to 10 of 21 entries
Previous
1
2
3
Next
Download MAF
Download VCF
Download JSON

Variant effect predictor (VEP)



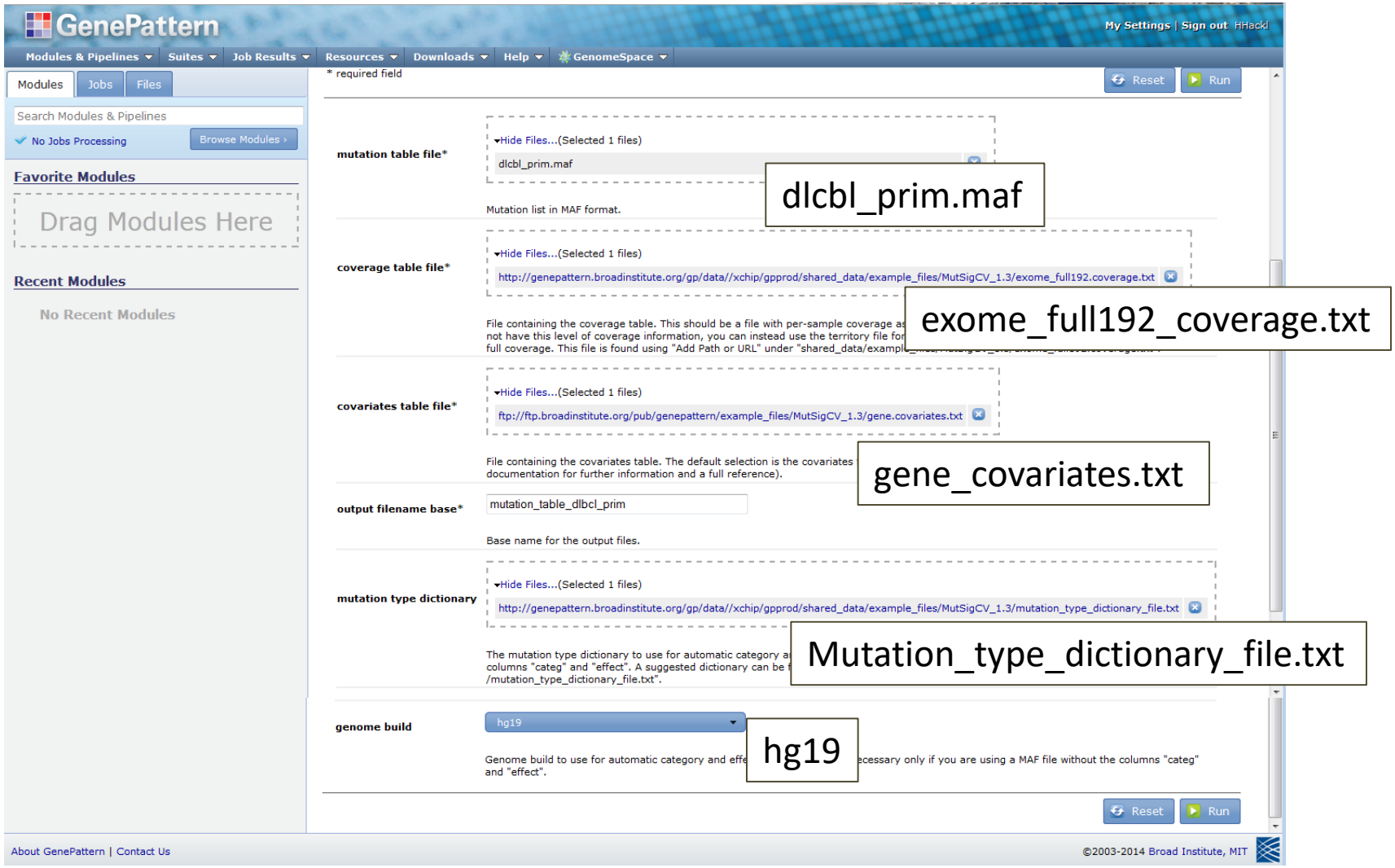
jurkat_mutations.vcf

Variant effect predictor (VEP)



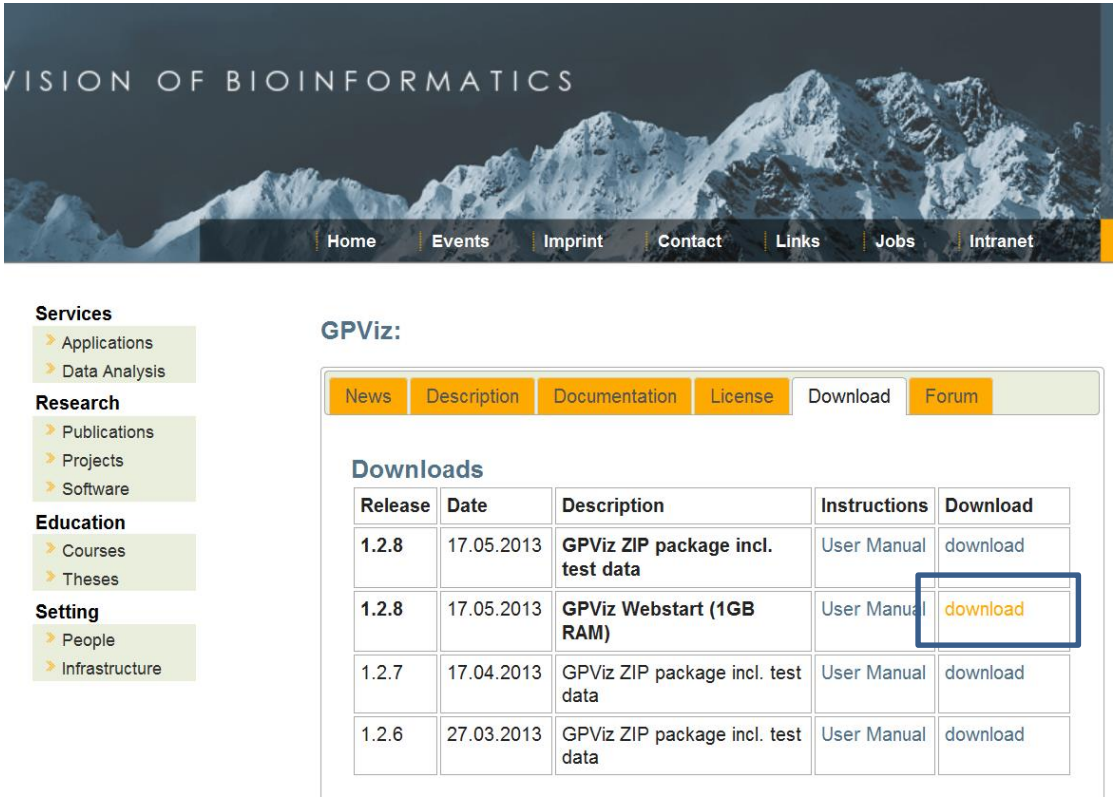
MutSigCV

https://www.broadinstitute.org/cancer/cga/mutsig_run (see key.txt)

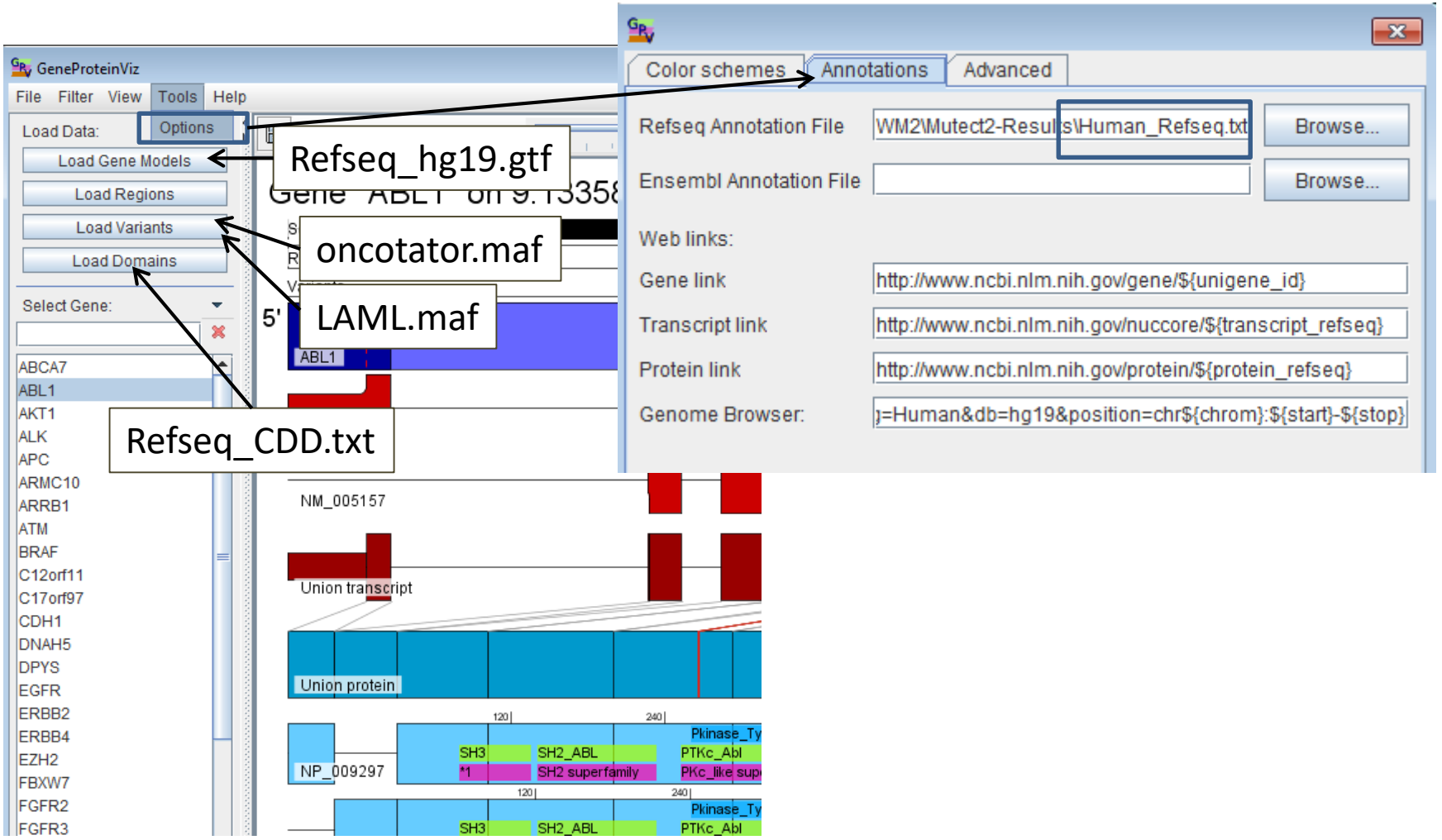


GPViz

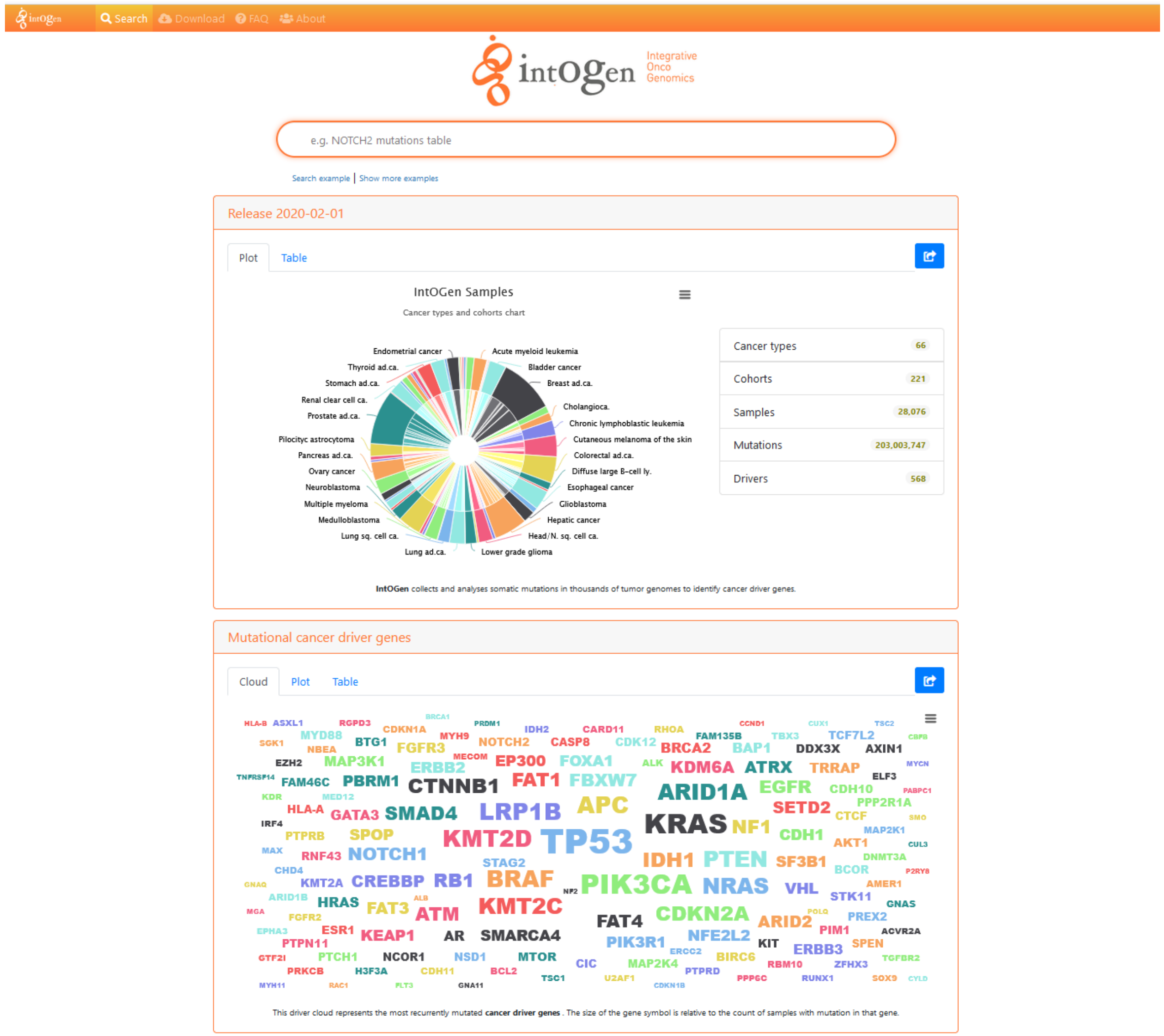
http://icbi.at/software/gpviz/gpviz.shtml



Do not update JAVA!


GeneProteinViz
File Filter View Tools Help
Load Data:
Options
Load Gene Models
Load Regions
Load Variants
Load Domains
Select Gene:
ABCA7
ABL1
AKT1
ALK
APC
ARMC10
ARRB1
ATM
BRAF
C12orf11
C17orf97
CDH1
DNAH5
DPYS
EGFR
ERBB2
ERBB4
EZH2
FBXW7
FGFR2
FGFR3
Refseq_hg19.gtf
oncotator.maf
LAML.maf
Refseq_CDD.txt
Color schemes
Annotations
Advanced
Refseq Annotation File
WM2\Mutect2-Results\Human_Refseq.txt
Browse...
Ensembl Annotation File
Browse...
Web links:
Gene link
http://www.ncbi.nlm.nih.gov/gene/${unigene_id}
Transcript link
http://www.ncbi.nlm.nih.gov/nuccore/${transcript_refseq}
Protein link
http://www.ncbi.nlm.nih.gov/protein/${protein_refseq}
Genome Browser:
}=Human&db=hg19&position=chr${chrom}:${start}-${stop}
5'
ABL1
NM_005157
Union transcript
Union protein
120
240
Pkinase_Ty
SH3
SH2_ABL
PTKc_Abl
NP_009297
*1
SH2 superfamily
PKc_like sup

IntOGen



intOGen
Search
Download
FAQ
About
intOGen Integrative Onco Genomics
e.g. NOTCH2 mutations table
Search example | Show more examples
Release 2020-02-01
Plot
Table
IntOGen Samples
Cancer types and cohorts chart
Endometrial cancer
Acute myeloid leukemia
Thyroid ad.ca.
Bladder cancer
Stomach ad.ca.
Breast ad.ca.
Renal clear cell ca.
Cholangioca.
Prostate ad.ca.
Chronic lymphoblastic leukemia
Pilocityc astrocytoma
Cutaneous melanoma of the skin
Pancreas ad.ca.
Colorectal ad.ca.
Ovary cancer
Diffuse large B-cell ly.
Neuroblastoma
Esophageal cancer
Multiple myeloma
Glioblastoma
Medulloblastoma
Hepatic cancer
Lung sq. cell ca.
Head/N. sq. cell ca.
Lung ad.ca.
Lower grade glioma
Cancer types 66
Cohorts 221
Samples 28,076
Mutations 203,003,747
Drivers 568
IntOGen collects and analyses somatic mutations in thousands of tumor genomes to identify cancer driver genes.
Mutational cancer driver genes
Cloud
Plot
Table
TP53
PIK3CA
KRAS
KMT2D
LRP1B
APC
ARID1A
CTNNB1
SMAD4
NOTCH1
BRAF
NRAS
KMT2C
PTEN
IDH1
CDKN2A
FAT1
FBXW7
EGFR
ATRX
KDM6A
SETD2
CREBBP
RB1
ATM
KEAP1
SMARCA4
FAT4
ARID2
PIK3R1
NFE2L2
HRAS
FAT3
SPOP
GATA3
PBRM1
EP300
FOXA1
ERBB2
MAP3K1
FGFR3
NOTCH2
CASP8
BRCA2
BAP1
DDX3X
AXIN1
This driver cloud represents the most recurrently mutated cancer driver genes . The size of the gene symbol is relative to the count of samples with mutation in that gene.

FannsDB

FannsDB is a database for Functional ANnotations for Non Synonymous SNVs which contains precalculated scores for several predictors.

Documentation

Download

Condel

Condel is a method to assess the outcome of non-synonymous SNVs using a CONsensus DELeteriousness score that combines various tools (MutationAssessor, FATHMM). This is the second version of Condel which includes an update of the combined tools and a new web interface. If you use Condel, please cite us:

Improving the Assessment of the Outcome of Nonsynonymous SNVs with a Consensus Deleteriousness Score, Condel (2011) Abel González-Pérez and Nuria López-Bigas, American Journal of Human Genetics 10.1016/j.ajhg.2011.03.004 Download PDF

The last update of CONDEL was released on 2014. No new updates are planned for the future.

▶ Query

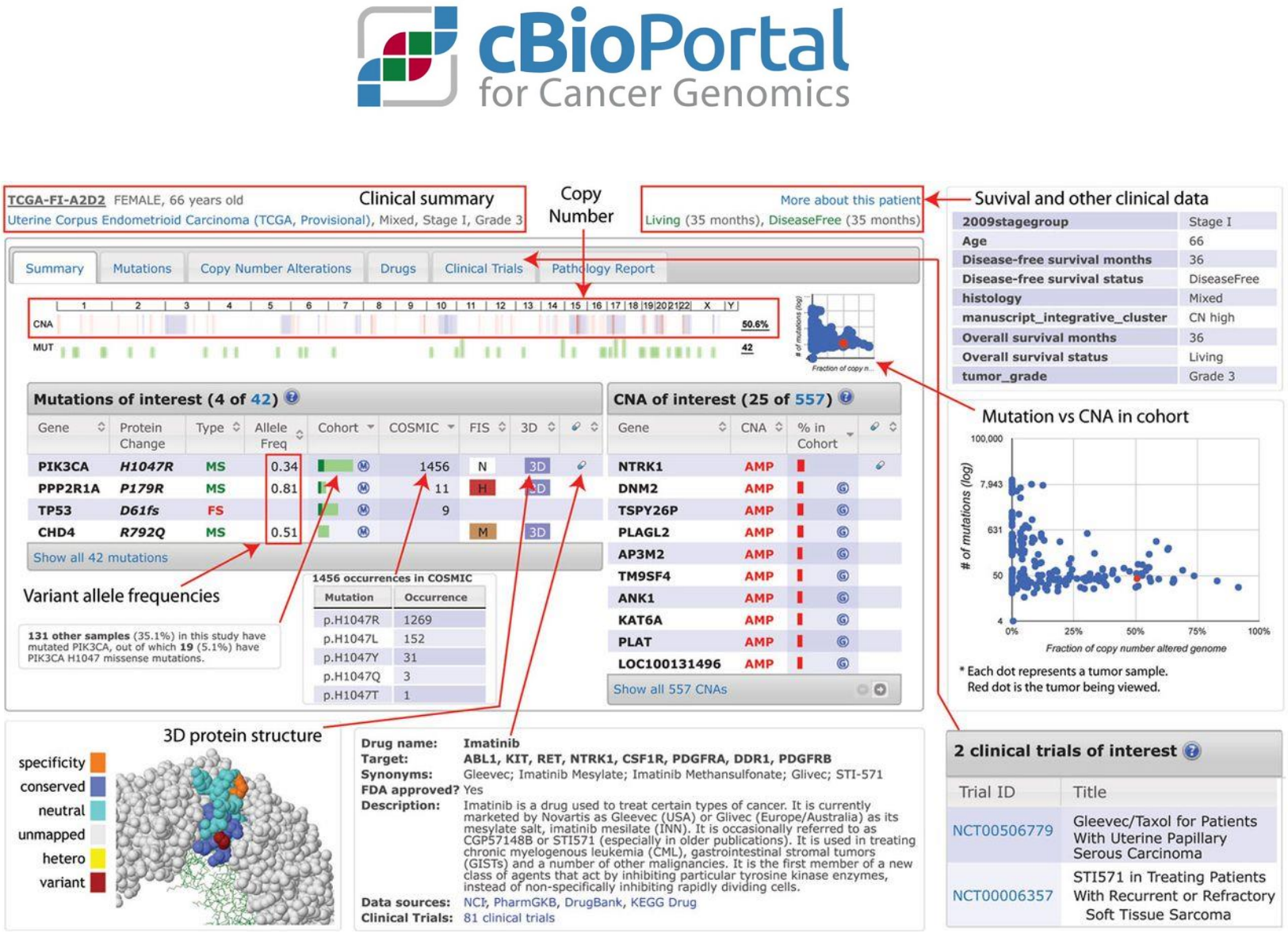
TransFIC

TransFIC (TRANSformed Functional Impact for Cancer) is a method to transform Functional Impact scores taking into account the differences in basal tolerance to germline SNVs of genes that belong to different functional classes. This transformation allows to use the scores provided by well-known tools (e.g. SIFT, Polyphen2, MutationAssessor) to rank the functional impact of cancer somatic mutations. Mutations with greater TransFIC are more likely to be cancer drivers. If you use TransFIC, please cite us:

Improving the prediction of the functional impact of cancer mutations by baseline tolerance transformation (2012) Gonzalez-Perez A, Deu-Pons J and Lopez-Bigas N. Genome Medicine. 4:89 doi:10.1186/gm390s Read

We have stopped supporting TRANSFIC. To evaluate the oncogenic potential of cancer mutations, please use the Cancer Genome Interpreter.

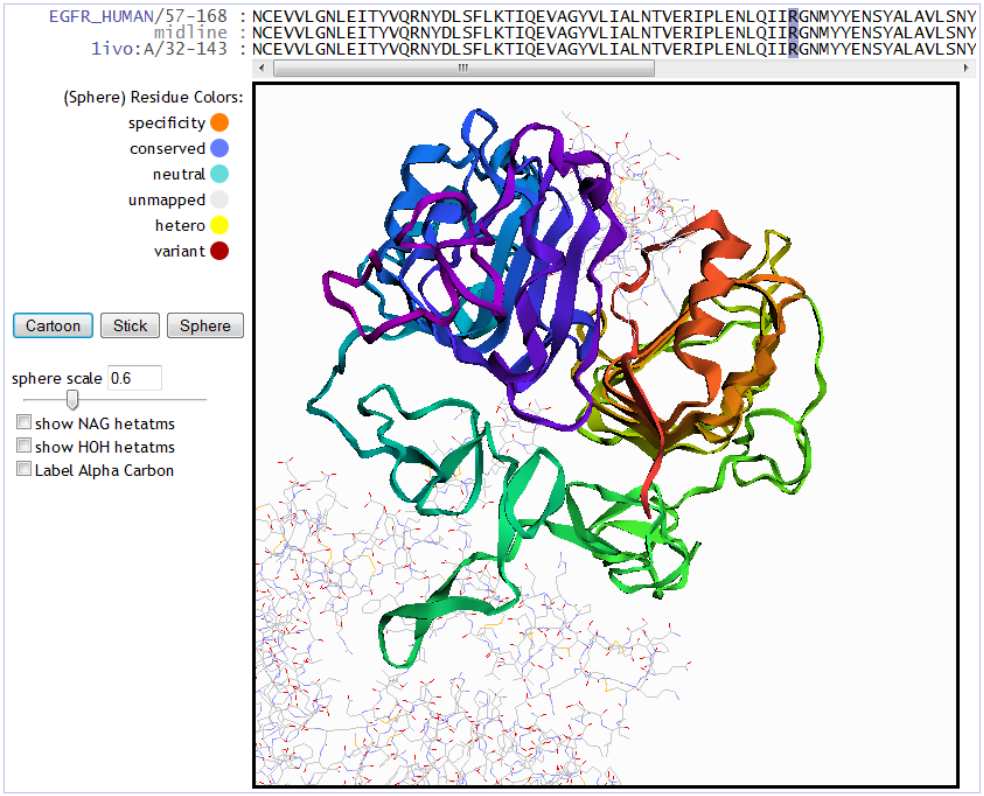
▶ Query

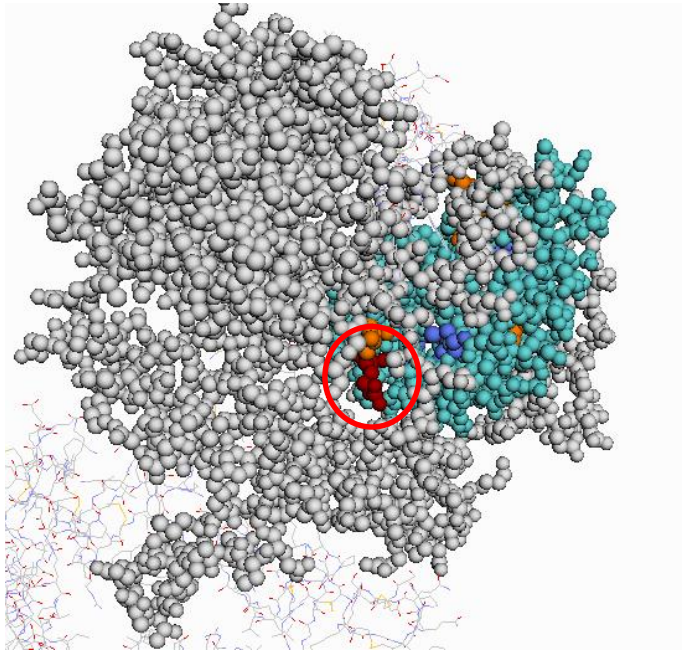

cBioPortal
for Cancer Genomics
TCGA-FI-A2D2 FEMALE, 66 years old
Uterine Corpus Endometrioid Carcinoma (TCGA, Provisional), Mixed, Stage I, Grade 3
Clinical summary
Copy Number
More about this patient
Living (35 months), DiseaseFree (35 months)
Suvival and other clinical data
2009stagegroup Stage I
Age 66
Disease-free survival months 36
Disease-free survival status DiseaseFree
histology Mixed
manuscript_integrative_cluster CN high
Overall survival months 36
Overall survival status Living
tumor_grade Grade 3
Summary
Mutations
Copy Number Alterations
Drugs
Clinical Trials
Pathology Report
CNA
MUT
50.6%
42
Mutations of interest (4 of 42)
Gene
Protein Change
Type
Allele Freq
Cohort
COSMIC
FIS
3D
PIK3CA H1047R MS 0.34 1456 N 3D
PPP2R1A P179R MS 0.81 11 H 3D
TP53 D61fs FS 9
CHD4 R792Q MS 0.51 M 3D
Show all 42 mutations
Variant allele frequencies
131 other samples (35.1%) in this study have mutated PIK3CA, out of which 19 (5.1%) have PIK3CA H1047 missense mutations.
1456 occurrences in COSMIC
Mutation
Occurrence
p.H1047R 1269
p.H1047L 152
p.H1047Y 31
p.H1047Q 3
p.H1047T 1
CNA of interest (25 of 557)
Gene
CNA
% in Cohort
NTRK1 AMP
DNM2 AMP
TSPY26P AMP
PLAGL2 AMP
AP3M2 AMP
TM9SF4 AMP
ANK1 AMP
KAT6A AMP
PLAT AMP
LOC100131496 AMP
Show all 557 CNAs
Mutation vs CNA in cohort
100,000
7,943
631
50
4
of mutations (log)
0%
25%
50%
75%
100%
Fraction of copy number altered genome
* Each dot represents a tumor sample.
Red dot is the tumor being viewed.
3D protein structure
specificity
conserved
neutral
unmapped
hetero
variant
Drug name: Imatinib
Target: ABL1, KIT, RET, NTRK1, CSF1R, PDGFRA, DDR1, PDGFRB
Synonyms: Gleevec; Imatinib Mesylate; Imatinib Methansulfonate; Glivec; STI-571
FDA approved? Yes
Description: Imatinib is a drug used to treat certain types of cancer. It is currently marketed by Novartis as Gleevec (USA) or Glivec (Europe/Australia) as its mesylate salt, imatinib mesilate (INN). It is occasionally referred to as CGP57148B or STI571 (especially in older publications). It is used in treating chronic myelogenous leukemia (CML), gastrointestinal stromal tumors (GISTs) and a number of other malignancies. It is the first member of a new class of agents that act by inhibiting particular tyrosine kinase enzymes, instead of non-specifically inhibiting rapidly dividing cells.
Data sources: NCI, PharmGKB, DrugBank, KEGG Drug
Clinical Trials: 81 clinical trials
2 clinical trials of interest
Trial ID
Title
NCT00506779 Gleevec/Taxol for Patients With Uterine Papillary Serous Carcinoma
NCT00006357 STI571 in Treating Patients With Recurrent or Refractory Soft Tissue Sarcoma

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

Home Advanced About

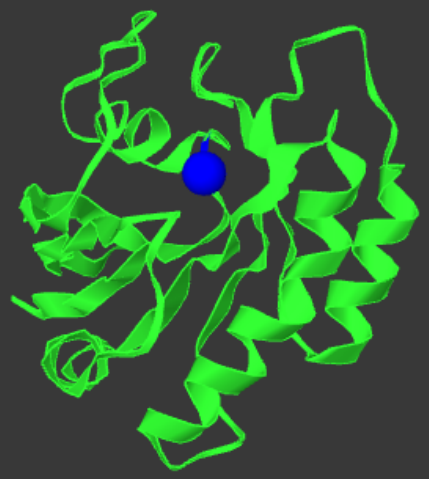
GTPase KRas

Protein Data

UniProt ID	P01116
Gene	KRAS
Length	189
Mutations	1
Clusters	0
Models	24

Links

Clustering Data File
PyMOL Session File
Permalink



View: Cartoon Line Ribbon

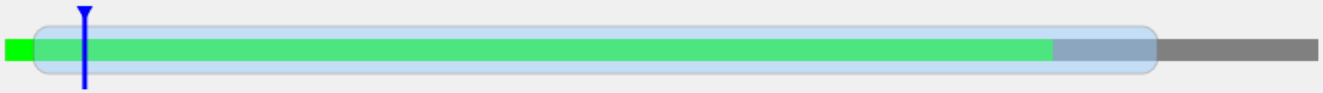
Model Information

Model: 3GFT_A_1_150
Source: PDB

AA Range	1-150
Source PDB	3GFT
Coverage	79.37%
Seq. Identity	100%

Cluster Selection

No clusters were found for this model.



Hide Domains Select Model

Cornell University
Weill Institute for Cell and Molecular Biology

Yu Group
Computational Suite

Visit the Yu Lab website for contact information and for other tools from the Yu Group Computational Suite

NetMHCpan 4.1

Submission | Instructions | Output format | Motif viewer | Abstract | Evaluation data sets | Version history

INPUT TYPE: Fasta

***Paste** a single sequence or several sequences in FASTA format into the field below:*

ASTPGHTIIYEAVCLHNDRTTIP

*... or **upload** a file in FASTA format directly from your local disk:*

Durchsuchen... Keine Datei ausgewählt.

*... or **load** some sample data:*

Load Data

PEPTIDE LENGTH:

You may select multiple lengths

8mer peptides
9mer peptides
10mer peptides
11mer peptides

SELECT SPECIES/LOCI:

HLA supertype representative

***Select** Allele(s) (max 20 per submission)*

HLA-A*01:01 (A1)
HLA-A*02:01 (A2)
HLA-A*03:01 (A3)
HLA-A*24:02 (A24)
HLA-A*26:01 (A26)

Some questions

1. How many somatic variants have you found
2. Are those point mutations, indels or other type
3. Are those synonymous, non-synonymous, or none coding
4. Changes in protein sequence? (p.V412A)
5. How many in intron, 3'UTR, cds,5'UTR
6. Type of variant (missense mut, frame shift del, frame shift ins, etc)
7. How many variants
8. Where in the 3D structure of the protein are the variants
9. Are these known variants (dbSNP, COSMIC)
10. Is there a functional impact (SIFT, PolyPhen2)
11. Are variants cancer drivers
12. Which genes/exons are studied (info about the panel, and why this was chosen)
13. Variants per megabase of studied DNA sequence

Question

14. Which transcripts and protein isoforms are covered by the variants (NM_,NP_,ENST, ENSP)
15. Which chromsome and position of the variant
16. Reference allele, alternative allele
17. Are there common variants to previously sequenced Melanoma cell lines
18. Are some of the genes found significantly and recurrently mutated in a cohort of 28 diffuse large B cell lymphoma patients by targeted exome sequencing
19. Are there somatic variants in the Melanoma TCGA cohort patients at the same positions. What are the allele frequency in this cohort?
20. Are some variants from the same cancer signalling pathway