

GeneGrid: finding disease-causing variants in NGS data

BTEP Exome-Seq Workshop
Justin Lack, CCBR
Feb 22, 2017

Slides provided by Christian Zinser, Genomatix GmbH

Basis

Genomic variants like SNPs or InDels are of major interest to biologists and clinicians

Identifying causal variants is crucial for the diagnostics of rare and common diseases

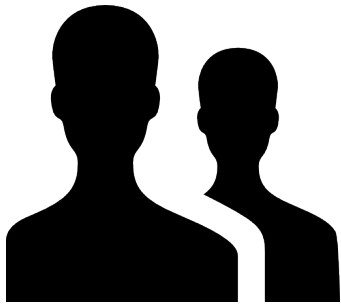
With NGS technology it is possible to detect millions of variants within an individual genome

Which are the relevant ones?



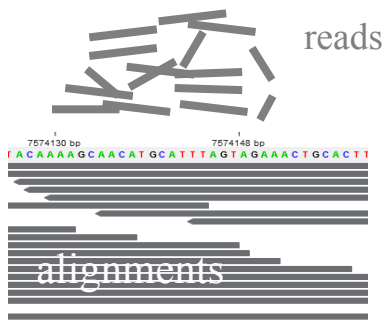
General workflow

Patient sampling



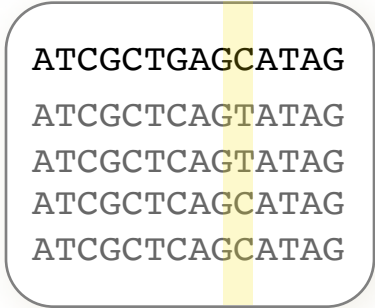
Sequencing library

Sequencing and alignment



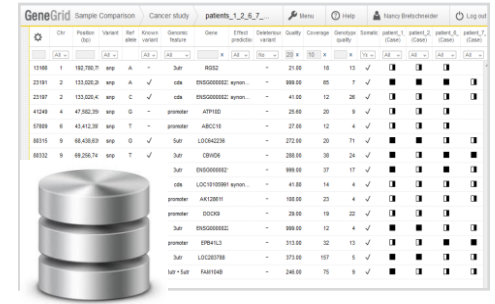
BAM file with read positions

Variant calling



VCF file with variant information

Variant analysis in GeneGrid



Chr	Position (bp)	Variant Ref alt	Struc variant	Gene	Struc product	Quality	Coverage	Genotype quality	Gene	strand_1 (Case)	strand_2 (Case)	strand_1 (Control)	strand_2 (Control)
13	182,782,7	wsp	A -	Sub	RSD2	-	21/65	18	13	✓			
21	191,252,3	wsp	A ✓	exon	ENSG00000218812	synon	-	98/95	85	7	✓	■	■
21	193,253,4	wsp	C ✓	exon	ENSG00000218812	synon	-	41/65	12	28	✓	■	■
4	41,582,26	wsp	G -	promoter	ATP12B	-	25/65	25	9	✓	■	■	■
8	43,412,38	wsp	T -	promoter	ABCC19	-	27/65	12	4	✓	■	■	■
8	48,438,83	wsp	G ✓	Sub	LOC1042328	-	272/50	20	71	✓	■	■	■
8	48,438,83	wsp	G ✓	Sub	CRH3G	-	288/50	20	24	✓	■	■	■
8	48,256,74	wsp	T ✓	Sub	CRH3G	-	98/95	37	17	✓	■	■	■
8	48,256,74	wsp	T ✓	Sub	ENSG00000218812	-	98/95	37	17	✓	■	■	■
8	48,256,74	wsp	T ✓	exon	LOC101058911	synon	-	41/65	14	4	✓	■	■
1	108,128,11	wsp	A	promoter	AA128B11	-	185/65	23	4	✓	■	■	■
1	108,128,11	wsp	A	promoter	ENSG00000218812	-	28/65	19	22	✓	■	■	■
1	108,128,11	wsp	A	Sub	ENSG00000218812	-	98/95	12	4	✓	■	■	■
1	108,128,11	wsp	A	promoter	EPHA4L3	-	213/65	52	15	✓	■	■	■
1	108,128,11	wsp	A	Sub	LOC101058911	-	375/65	167	9	✓	■	■	■
1	108,128,11	wsp	A	Sub	ENSG00000218812	-	248/65	75	9	✓	■	■	■

Short list of potentially interesting variants

Finding the needle in the haystack

Variant caller

VCF file
SNPs and small InDels

up to millions of variants

select non-synonymous variants

keep only rare variants

e.g. 1000 Genomes Project

GeneGrid

filter variants on inheritance patterns or case/control

e.g. autosomal recessive

select genes based on patient's phenotype

e.g. COSMIC, LitInspector disease

rank

e.g. Blosum62, SIFT, PhyloP

10s of causal variants

Selected GeneGrid filter fields

Field	Description
Known gene	Variant is in an annotated gene
Genotype quality	Variant confidence score; higher is better; max 999
gAF	Global minor allele frequency from 1000 Genomes Project; also separate for African, American, Asian, European; <0.01: rare
espMAF	Minor allele frequency from Exome Sequencing Project; also separate for African American, European American; <0.01: rare
exacAF	Alternative allele frequency from the Exome Aggregation Consortium (ExAC) project; <0.01: rare
BLOSUM	AA substitution score; <0: rare; >0: common
SIFT	Sorting Intolerant From Tolerant: predicts effect of AA substitution on protein function; <0.05: affects function
SIFT pred	Yes: damaging prediction based on SIFT score
PhyloP	DNA conservation score; >0: conserved
GERP	Genomic Evolutionary Rate Profiling; RS Score; >0: fewer substitutions than expected -> evolutionary constraint

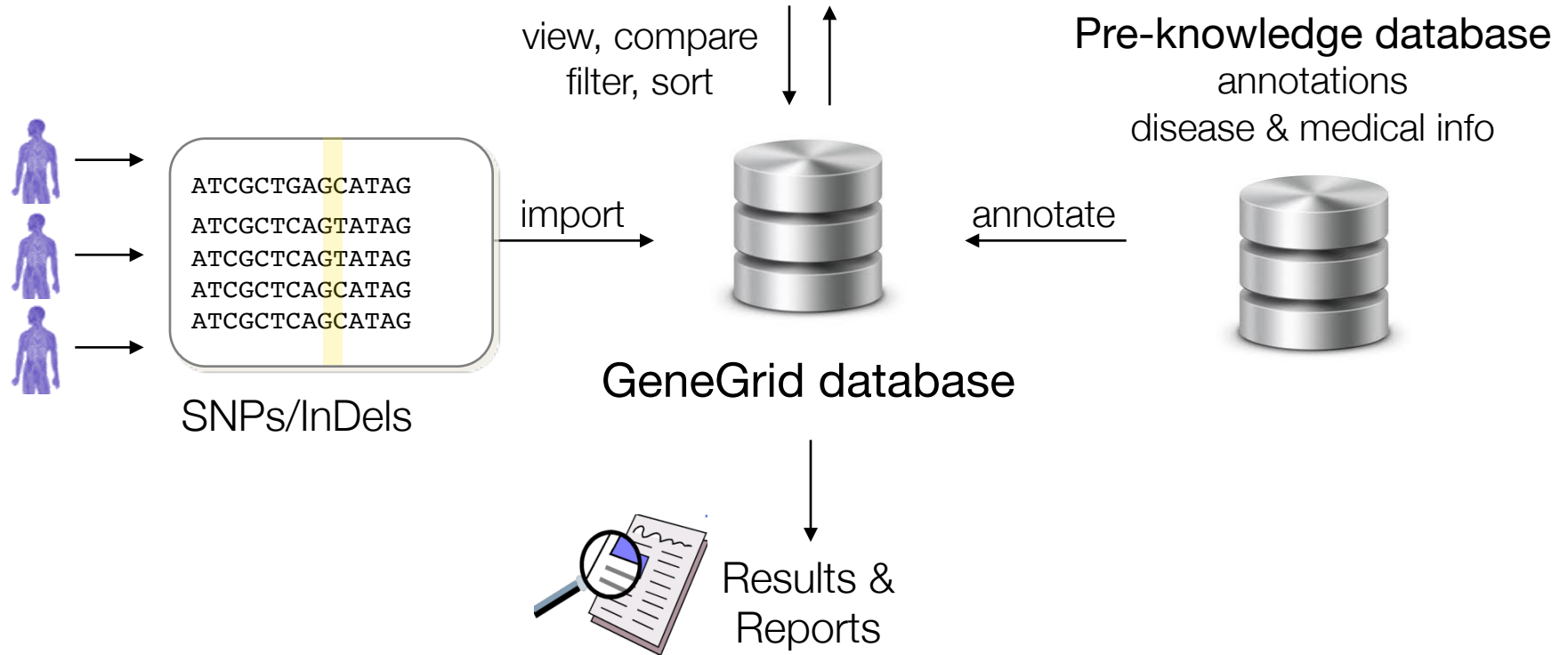
Selected GeneGrid filter fields

Field	Description
SiPhy Pi, Omega	Overlap with evolutionary constrained elements (two parameter sets)
Literature diseases	Associated diseases based on gene co-citations
Literature tissues	Associated tissues based on gene co-citations
Somatic mutation tissues	Associated tissues; source: COSMIC
Clinical diseases	Associated clinical diseases; source: ClinVar and OMIM
Diagnostic tests	Number of diagnostic tests available for gene
Diagnostic diseases	Associated disease terms based on diagnostic gene tests (gene level)
Clinical significance	Summary of clinical significance based on ClinVar database (position level)
GO processes	Associated GO terms for the domain of cellular components
GO functions	Associated GO terms for the domain of molecular functions
GO components	Associated GO terms for the domain of biological processes

The Genomatix GeneGrid technology

Chr	Position (bp)	Variant	Ref allele	Known variant	Genomic feature	Gene	Known gene	Effect prediction	Deleterious variant	Quality	Coverage	pAF	Compound heterozygosity	Offspring inheritance	II-1 (Case)	II-4 (Control)	II-5 (Control)
1	752.721	snp	A	✓	promoter-interg.	FAM87B	✓	-	✓	204.00	1	0.68	-	het	■	■	■
2	752.894	snp	T	✓	cds	FAM87B	✓	missense	✓	205.00	1	0.74	-	het	■	■	■
3	809.876	snp	A	✓	cds	FAM41C	✓	missense	✓	287.00	8	0.14	-	het	■	■	■
4	847.250	snp	G	✓	intron (cs)	LOC284800	✓	-	✓	76.00	1	1.00	-	het	■	■	■
5	848.738	snp	C	✓	cds	LOC284800	✓	missense	✓	15.66	4	0.16	-	denovo	■	■	■
6	849.998	snp	A	✓	3ut-intron (cs)	LOC284800	✓	-	✓	73.00	1	0.81	-	het	■	■	■
7	850.780	snp	C	✓	intergenic	-	-	-	✓	40.13	2	0.58	-	recal	■	■	■
8	852.875	snp	C	✓	3ut-intron (cs)	LOC109130417	✓	splice-site	✓	72.00	1	0.51	-	het	■	■	■
9	852.954	snp	T	✓	3ut-intron (cs)	LOC109130417	✓	-	✓	79.00	1	0.75	-	het	■	■	■
10	853.954	snp	C	✓	3ut-intron (cs)	LOC109130417	✓	-	✓	46.13	2	0.51	-	recal	■	■	■
11	856.319	snp	G	✓	intron (cs)	ENSG00000268170	✓	-	✓	111.00	1	0.96	-	het	■	■	■
12	856.319	snp	G	✓	intron (cs)	SAMD11	✓	-	✓	111.00	1	0.96	-	het	■	■	■
13	856.511	ins	C...	-	promoter	ENSG00000268170	-	-	✓	96.00	1	-	-	het	■	■	■
14	856.511	ins	C...	-	intron (cs)	SAMD11	✓	-	✓	96.00	1	-	-	het	■	■	■
15	871.334	snp	G	✓	intron (cs)	SAMD11	✓	-	✓	292.00	8	0.51	-	het	■	■	■
16	873.314	snp	G	✓	3ut+cds	SAMD11	✓	synonymous	✓	49.50	4	0.05	-	ref	■	■	■
17	879.676	snp	G	✓	3ut	NOC2L	✓	-	✓	315.00	14	0.90	-	het	■	■	■
18	879.676	snp	G	✓	3ut	SAMD11	✓	-	✓	315.00	14	0.90	-	het	■	■	■
19	879.687	snp	T	✓	3ut	NOC2L	✓	-	✓	309.00	13	0.93	-	het	■	■	■
20	879.687	snp	T	✓	3ut	SAMD11	✓	-	✓	309.00	13	0.93	-	het	■	■	■

available online



Data sources

Internal sources

Variant annotation	Genomatix
Genome annotation	EIDorado
Text mining (PubMed)	LitInspector
Combined thesaurus (MeSH, NCI, UMLS)	Genomatix Thesaurus
Pathways and networks	GePS

Data sources

External sources

Alleles and allele frequencies

dbSNP
 1000 Genomes Project
 ESP6500
 ExAC

Protein effect predictions

BLOSUM
 SIFT

Evolutionary conservation

PhyloP
 GERP
 29 Mammals Project

Diagnostic annotation

GTR

Phenotype annotation

ClinVar
 COSMIC
 OMIM

Gene ontology

GO

Sequence variant description

HGVS

Regulatory regions

Ensembl Regulatory Build

GeneGrid example 1

Trio analysis

Leber congenital amaurosis

Inherited eye disease

Onset at birth or in early childhood

Blindness or impaired vision with loss of central vision

Genetically heterogeneous

e.g. LCA1 caused by homozygous mutation in GUCY2D

LCA2: RPE65

LCA3: SPATA7

LCA4: AIPL1

etc. (18 types known)

Leber congenital amaurosis

Published in final edited form as:

Nat Genet. 2012 September ; 44(9): 1040–1045. doi:10.1038/ng.2361.

***NMNAT1* mutations cause Leber congenital amaurosis**

Marni J Falk^{1,2,22}, **Qi Zhang**^{3,4,22}, **Eiko Nakamaru-Ogiso**⁵, **Chitra Kannabiran**⁶, **Zoe Fonseca-Kelly**^{3,4}, **Christina Chakarova**⁷, **Isabelle Audo**^{8,9,10,11}, **Donna S Mackay**⁷, **Christina Zeitz**^{8,9,10}, **Arundhati Dev Borman**^{7,12}, **Magdalena Staniszewska**^{3,4}, **Rachna Shukla**⁶, **Lakshmi Palavalli**⁶, **Saddek Mohand-Said**^{8,9,10,11}, **Naushin H Waseem**⁷, **Subhadra Jalali**^{6,13}, **Juan C Perin**¹⁴, **Emily Place**^{1,3,4}, **Julian Ostrovsky**¹, **Rui Xiao**¹⁵, **Shomi S Bhattacharya**^{7,16}, **Mark Consugar**^{3,4}, **Andrew R Webster**^{7,12}, **José-Alain Sahel**^{8,9,10,11,17,18}, **Anthony T Moore**^{7,12,19}, **Eliot L Berson**⁴, **Qin Liu**^{3,4}, **Xiaowu Gai**^{20,21,23}, and **Eric A. Pierce**^{3,4,23}

¹Division of Human Genetics, Department of Pediatrics, The Children's Hospital of Philadelphia and University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA

²Division of Child Development and Metabolic Disease, Department of Pediatrics, The Children's Hospital of Philadelphia and University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA

³Ocular Genomics Institute, Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, Massachusetts, USA

⁴Berman-Gund Laboratory for the Study of Retinal Degenerations, Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, Massachusetts, USA

⁵Department of Biochemistry and Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA

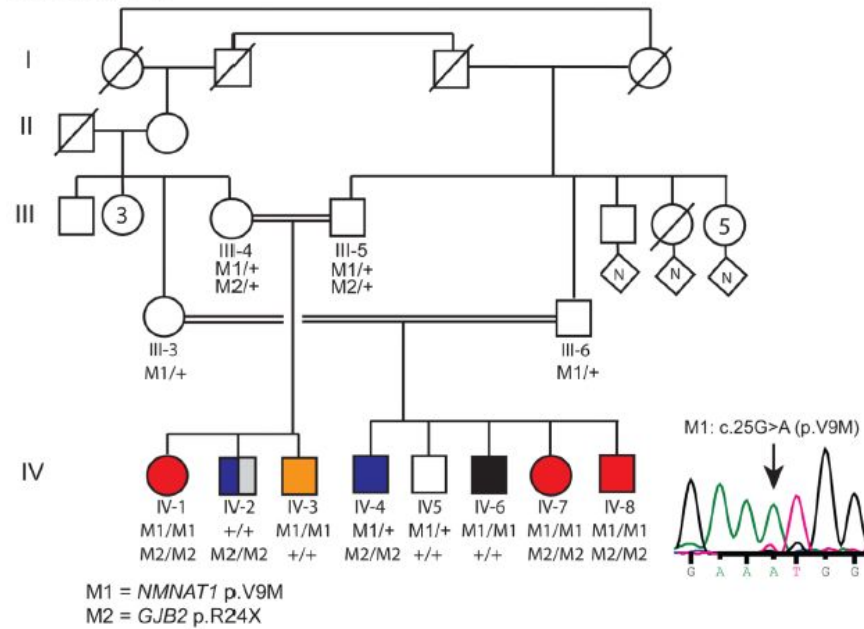
⁶Kallam Anji Reddy Molecular Genetics Laboratory, LV Prasad Eye Institute (LVPEI), Kallam Anji Reddy Campus, LV Prasad Marg, Hyderabad, India

⁷Institute of Ophthalmology, University College of London, London, UK

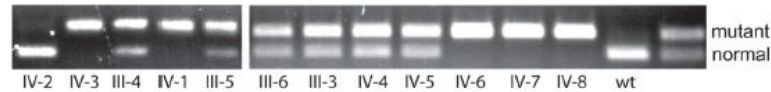
⁸Institut National de la Santé et de la Recherche Médicale, U968, Paris, France

Consanguineous family

a Family 047



b Genotyping 047



Mucopolipidosis type III
 Deafness
 LCA, global developmental delay, autism



LCA, congenital deafness, global developmental delay, autism

Step 1: load and annotate VCF files

Getting started

The Genomatix GeneGrid technology enables you to quickly reduce millions of small variants to the few or even the single relevant one(s). All known & novel SNPs in your results can be annotated using our extensive annotation. You can filter the list for those variants of interest to you, perform trio analyses, compare case and control sets (using multiple samples) or identify somatic SNPs within minutes. [Read more](#)

Variant Annotation



Load your VCF files with samples into GeneGrid to be automatically annotated.



Sample Comparison



Find the relevant small variants and identify disease-causing mutations by comparing samples.

Genome Browser



Browse the human genome in context of your variants of interest and explore publicly available data.

Pathway System (GePS)



Browse, search and load canonical pathways and visualize affected genes on pathway level.

Step 1: load and annotate VCF files

GeneGrid Variant Annotation
Samples

Filter samples

Import samples & annotate variants

Select your input file and import samples and automatically annotate the variants.

Note: The required input file format is the **VCF format**. The genomic positions of the variants must match the human NCBI build 37 / hg19. [Read more](#)

Step 1: Define pre-filter settings for import:

Exome filter

Minimum coverage:

Hint: Pre-filters are optional and can be used to reduce the number of variants that will be imported. [Read more](#)

Step 2: Select the variant file from your computer:

Keine Datei ausgewählt

Name	Date modified	Type
patient_7.exome.dup.mq20.filtered.min.vcf.gz	05.09.2014 09:01	WinRAR-Archiv
trio.exome.mq20.min.vcf.gz	05.09.2014 08:53	WinRAR-Archiv

f.gz

Variant Annotation (1 running)

- The VCF file **trio.exome.mq20.min.vcf.gz** of size 3.6 MB was uploaded *today*. The analysis has been running for *5 minutes*. It can take up to several hours depending on the size of the input data. Upon completion of the analysis you will receive a notification by email.

Associate alignment files

Step 2: sample comparison

Getting started

The Genomatix GeneGrid technology enables you to quickly reduce millions of small variants to the few or even the single relevant one(s). All known & novel SNPs in your results can be annotated using our extensive annotation. You can filter the list for those variants of interest to you, perform trio analyses, compare case and control sets (using multiple samples) or identify somatic SNPs within minutes. [Read more](#)

Variant Annotation



Load your VCF files with samples into GeneGrid to be automatically annotated.

Sample Comparison



Find the relevant small variants and identify disease-causing mutations by comparing samples.



Genome Browser



Browse the human genome in context of your variants of interest and explore publicly available data.

Pathway System (GePS)



Browse, search and load canonical pathways and visualize affected genes on pathway level.

Step 2: sample comparison

Sample ID	Input file	Sample	Number of non-ref variants	Class	Activated	Associated alignments	Exome filter	Minimum coverage
<input type="text"/> x	<input type="text"/> x	<input type="text"/> x	<input type="text"/> x	All	All	All	All	<input type="text"/> x
Open	LCA047_Trio_Demo.vcf	III-5	76,762	medium	✓	✓	-	1
1277	LCA047_Trio_Demo.vcf	III-4			✓	✓	-	1
1276	LCA047_Trio_Demo.vcf	IV-1			✓	✓	-	1

Compare samples

Step 1: Select the type of comparison study:

Step 2: Assign the samples to the groups:

Offspring (1 assigned)

	Sample	#	ID
1	IV-1	77,084	1276

Parents (2 assigned)

	Sample	#	ID
1	III-4	41,217	1277
2	III-5	76,762	1278

Study name:

Step 1: Select the type of comparison study:

Step 2: Assign the samples to the groups:

Hint: Just drag an activated sample from the table on the right side and drop it in one of the two groups below. [Read more](#)

Offspring (0 assigned, requires 1 more)

	Sample	#	ID

Parents (0 assigned, requires 2 more)

	Sample	#	ID

Study name:

Step 3: filtering

Filter variants

Add column:

Known gene:

Deleterious variant:

gAF:

IV-1 (Case):

III-4 (Control):

III-5 (Control):

gAF:

IV-1 (Case):

III-4 (Control):

III-5 (Control):

Filter variants

Add column:

Known gene:

Deleterious variant:

gAF:

IV-1 (Case):

III-4 (Control):

III-5 (Control):

Genomic feature	Gene	Known gene	Effect prediction	Deleterious variant	Quality	Coverage	gAF	Compound heterozygosity	Offspring inheritance	IV-1 (Case)	III-4 (Control)
All		Yes	All	Yes			0.01	All	All	Homozygous	Heterozygous
3utr • intron • cds	NADK	✓	insertion	✓	304.00	2	-	-	hom	■	□
intron (cs) • cds	TAS1R1	✓	missense	✓	511.00	4	0.01	-	hom	■	□
cds	NMNAT1	✓	missense	✓	728.00	33	-	-	hom	■	□
3utr • intron (cs) • ...	TPO	✓	missense	✓	608.00	10	-	-	hom	■	□
cds	OR5K4	✓	nonsense	✓	203.00	11	-	-	hom	■	□
cds	OR5K3	✓	frameshift	✓	200.00	13	-	-	hom	■	□
cds • promoter	ZNF518B	✓	missense	✓	621.00	14	-	-	hom	■	□
cds	CC2D2A	✓	missense	✓	586.00	7	-	-	hom	■	□
cds	SHISA3	✓	missense	✓	593.00	9	-	-	hom	■	□
intron (cs) • dono...	MCC	✓	splice-site	✓	356.00	5	-	-	hom	■	□
cds	MRAP2	✓	missense	✓	572.00	13	< 0.01	-	hom	■	□
5utr • exon (no or...	KIAA1009	✓	deletion	✓	728.00	16	-	-	hom	■	□
intron (cs) • intro...	GPR126	✓	nonsense	✓	571.00	4	-	-	hom	■	□
intron (cs)	GUSB	✓	splice-site	✓	553.00	3	-	-	hom	■	□
intron • cds	ATP6V0A4	✓	missense	✓	434.00	3	< 0.01	-	hom	■	□
intron (cs) • cds	KLRG2	✓	missense	✓	402.00	2	-	-	hom	■	□
cds	RP1L1	✓	missense	✓	664.00	12	-	-	hom	■	□
intron (cs)	ADHFE1	✓	splice-site	✓	434.00	3	-	-	hom	■	□
5utr • cds	TERF1	✓	missense	✓	497.00	8	-	-	hom	■	□
intron (cs)	BAG1	✓	splice-site	✓	505.00	9	-	-	hom	■	□
5utr • intron • cds	FAM205A	✓	missense	✓	421.00	4	-	-	hom	■	□

Viewing: 1 to 44 | Filtered: 44 | Total variants: 113,697

Filter history and template filters

Speed up the analysis and facilitate sample comparisons

Filter history

Template filter

1 Store the currently active filter settings as a *reusable template* to quickly filter variants in other result sets.

2 seconds ago

Deleterious variant • K diseases • gAF • IV-1 (Control)

4 minutes ago

Deleterious variant • K (Case) • III-4 (Control)

49 minutes ago

Unfiltered

Template title:

Save
Manage all

Available templates

19 CancerSomatic

Genotype quality • Deleterious variant • gAF • Quality • III-4 (Control) • Somatic

28 MeinFilter44

Deleterious variant • gAF • III-4 (Control) • III-5 (Control) • Offspring inheritance

21 CancerSomaticBlymph

Coverage • Genotype quality • Deleterious variant • Literature tissues • gAF • Quality • Somatic

27 LCATrio_IntellDisab

Deleterious variant • gAF • OMIM diseases • III-4 (Control) • III-5 (Control) • Offspring inheritance

Combine filters that are routinely used, e.g.

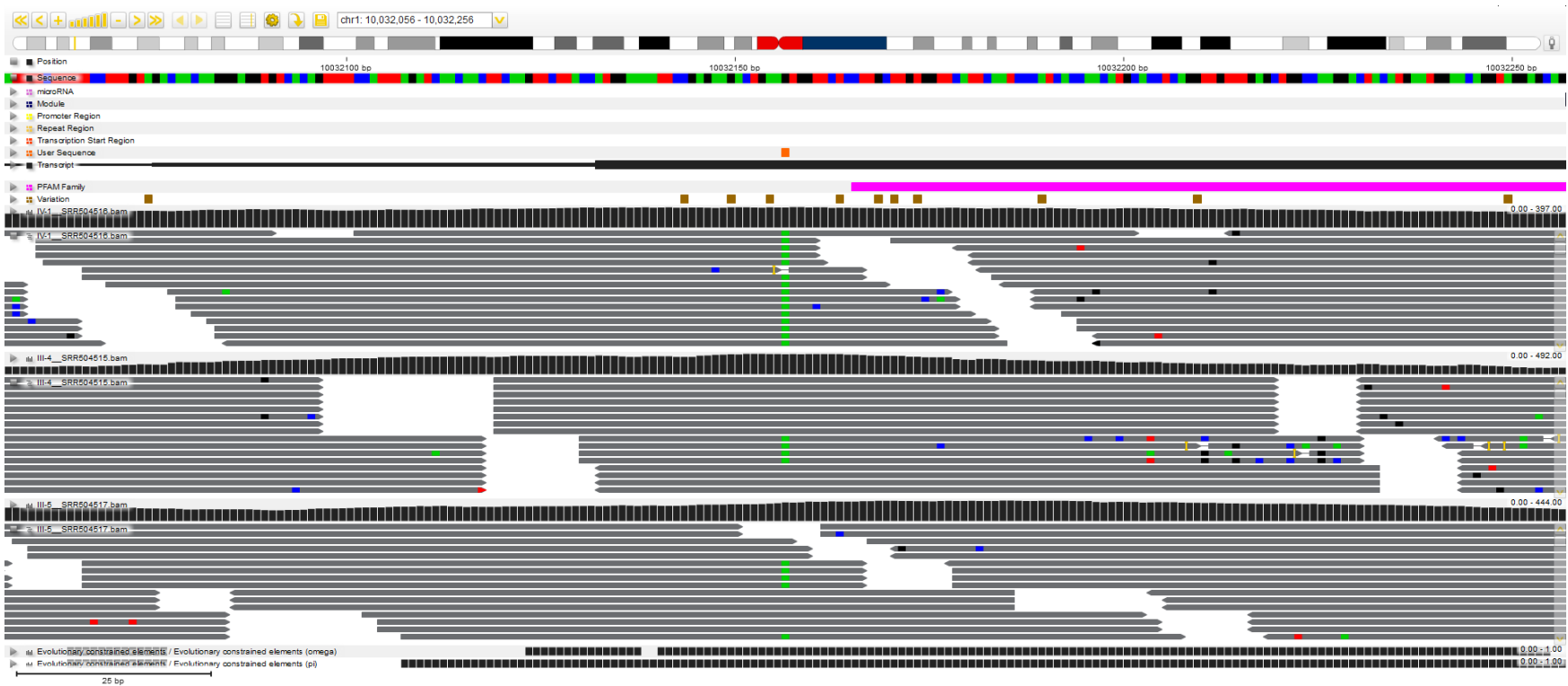
- Minimum read coverage
- Genotype quality
- Allele frequency (gAF)
- Phenotype annotation
- Inheritance patterns
- Case / control genotypes



and save as template for use in other comparisons

Vizualisation in genome browser

Settings	Chr	Position (bp)	Variant	Ref allele	Known variant	Genomic feature	Gene	Known gene	Effect prediction	Deleterious variant	Quality	Coverage	gAF	Literature diseases		
	All		All		All	All		x	All	All		x	x	0.01	x	Leber Congenital Amaurosis [Leber Congenital Amaurosis, 335
Browse	1	10,032,156	snp	G	-	cds	NMNAT1	✓	missense	✓	728.00	33		Frontotemporal Dementia • Leber Congenital Amaurosis • Anet		



Additional annotation / filter columns

Variant description

 Alt2 allele

Evolutionary conservation

 PhyloP
 GERP

Clinical and diagnostic annotation

 Somatic mutation tissues

Filter variants

Add column:

Known gene = Yes x

Deleterious variant = Yes x

gAF ≤ 0.01 x

Literature diseases ~ Leber Congenital Amaurosis [Leber C] x

IV-1 (Case) = Homozygous x

III-4 (Control) = Heterozygous x

III-5 (Control) = Heterozygous x

Search

Filter history

Template filter

Pathway analysis

Report generator

Export variants

Ref allele	Known variant	Genomic feature	Gene	Known gene	Effect prediction	Deleterious variant	Quality	Coverage	gAF	Literature diseases	Compound heterozygosity	Offspring inheritance	IV-1 (Case)
G	-	cds	NMNAT1	✓	missense	✓	728.00	33	0.01	Leber Congenital Amaurosis [Leber Congenital Amaurosis, 339527]	x	All	Homozyg

Sample details (3)			Transcript effects (7)		dbSNP	ClinVar (1)	COSMIC	Literature diseases (26)				ClinVar diseases (2)		OMIM diseases (1)	Gene details		Literature tissues (22)		
Sample	Variant	Chr	Position (bp)	Variant	Zygosity	Ref	Alt	Alt2 allele	Genotype	Gene	Known gene	Effect prediction	Deleterious	Quality	Coverage	Ref coverage	Alt coverage	Genotype quality	
IV-1	691	1	10,032,156	snp	hom	G	A		■	NMNAT1	✓	missense	✓	728.00	115	1	114	127	
III-4	272	1	10,032,156	snp	het	G	A		■	NMNAT1	✓	missense	✓	728.00	33	14	19	127	
III-5	693	1	10,032,156	snp	het	G	A		■	NMNAT1	✓	missense	✓	728.00	98	59	39	127	

© 2014 [Privacy](#) [Terms](#) [Contact](#)

Viewing: 1 to 1 | Filtered: 1 | Total variants: 113,697

SIFT
 SIFT pred

Somatic mutation frequency

Annotation detail views and links

Sample details (3)
Transcript effects (7)
dbSNP
ClinVar (1)
COSMIC
Literature diseases (26)
ClinVar diseases (2)
OMIM diseases (1)
Gene details
Literature tissues (22)

Sample	Variant	Chr	Position (bp)	Variant	Zygoty	Ref	Alt	Alt2 allele	Genotype	Gene	Known gene	Effect
IV-1												
III-4												
III-5												

NCBI Resources How To
Sign in to NCBI

ClinVar

ClinVar

Search ClinVar for gene symbols, HGVS expressions, conditions, and more

Advanced

Home
About
Data use and maintenance
Using the website
How to submit
Statistics
FTP site

Show additional filters
Display Settings: Tabular, 20 per page, Sorted by Default order
Send to:

Gene

Select ...

Clinical significance

Pathogenic (20)

Review status

Multiple submitters (1)

Single submitter (19)

Method type

Literature only (9)

Clinical testing (11)

Molecular consequence

Missense (8)

Variation type

Copy gain (2)

Copy loss (9)

Deletion (9)

Insertion (2)

Single nucleotide (9)

Results: 1 to 20 of 21

	Gene(s)	Condition(s)	Frequency	Clinical significance (Last reviewed)	Review status	Chr	Location (GRCh38)
1.	NM_022787.3(NMNAT1):c.695G>A (p.Arg232Lys)	Malignant melanoma		not provided	not classified by submitter	1	9982556
2.	GRCh38/hg38 1p36.22-36.21(chr1:9406722-12852772)x1	See cases		Pathogenic/Likely pathogenic (Aug 12, 2011)	classified by single submitter	1	9406722 - 12852772
3.	GRCh38/hg38 1p36.23-36.13(chr1:9034671-16441465)x1	See cases		Pathogenic/Likely pathogenic (Aug 12, 2011)	classified by single submitter	1	9034671 - 16441465
4.	GRCh38/hg38 1p36.23-36.22(chr1:7165036-13111056)x1	See cases		Pathogenic/Likely pathogenic (Aug 12, 2011)	classified by single submitter	1	7165036 - 13111056
5.	GRCh38/hg38 1p36.31-36.22(chr1:5682528-10863843)x1	See cases		Pathogenic/Likely pathogenic (Aug 12, 2011)	classified by single submitter	1	5682528 - 10863843
6.	GRCh38/hg38 1p36.32-36.21(chr1:4898439-13111056)x1	See cases		Pathogenic/Likely pathogenic (Aug 12, 2011)	classified by single submitter	1	4898439 - 13111056

Sample details

ID	Label
2	Ma

[View](#)
[Clear all](#)
[Show additional filters](#)

Report generator

Report generator

Generate a report for the currently filtered variants (must *not exceed* 10 variants) appearing in the right table.

Depending on the number of variants please allow a few minutes for report generation.

Report title:

NMNAT1

Generate

Reset

NMNAT1

October 27, 2014

DNA variants

Summary

This report consists of one variant:

1. NMNAT1 | SNV | c.25G>A / p.(Val9Met) | pathogenic

p.(Val9Met) in NMNAT1

Variant description

The indicated single nucleotide variant (SNV) is located on chromosome 1 at position 10,032,156 bp. It overlaps the coding sequence of at least one transcript of gene *NMNAT1*. The reference allele for this variant is G, whereas the alternative allele is A.

Variant quality (smallest value across all samples in comparison)

- Minimal depth of coverage (COV): 33 reads
- Minimal quality value for the assertion of the alternative allele (QUAL): 728.00
- Minimal conditional genotype quality for this site being a variant (GQ): 127

Case distribution (1/1)

- Homozygous variant:
 1. IV-1 (COV: 115 reads, QUAL: 728.00, GQ: 127)

Control distribution (2/2)

- Heterozygous variant:
 1. III-4 (COV: 33 reads, QUAL: 728.00, GQ: 127)
 2. III-5 (COV: 98 reads, QUAL: 728.00, GQ: 127)

Predicted molecular effects on protein

This variant is predicted to be a missense mutation which alters the protein's amino acid from valine (Val) to methionine (Met). The prediction for p.(Val9Met) is based on 7 annotated transcripts for that gene locus. The BLOSUM62 substitution matrix reports a score of 1 for this alteration.

Known variant

The variant was not previously reported in dbSNP.

Clinical significance

The mutation g.10032156G>A was clinically associated with Leber congenital amaurosis 9 and classified as pathogenic (ID RCV000030771).

Allele frequencies in populations

There was no alternative allele frequency listed for this variant in the 1000 genomes data set. There was no minor allele frequency listed for this variant in the NHLBI GO Exome Sequencing Project (ESP6500) data set.

Computational protein effect prediction

The amino acid substitution is predicted to be damaging (SIFT score: 0.01). The sequence alteration is evaluated to be disease-causing (MutationTaster score: 0.94). The prediction that this mutation is in highly conserved amino acids within protein-coding sequences is neutral (LRT score: < 0.01). The accuracy of these tools is unknown.

Evolutionary conservation

Variant overlaps with evolutionary constrained element (detected using SiPhy- ω and SiPhy- π statistics). The conservation across 28 species is described with PhyloP (score: 0.98). GERP identifies constrained elements in multiple alignments by quantifying substitution deficits (score: 2.99).

NMNAT1

General information

Data source: EIDorado


Gene description

nicotinamide nucleotide adenylyltransferase 1 (ID 64802)
NMNAT1 is also known as *LCA9*, *NMNAT*, *NMNAT1*, *PNAT1* and located on locus 1p36.22.

Brief summary

This gene encodes an enzyme which catalyzes a key step in the biosynthesis of the coenzyme NAD. The encoded protein is one of several nicotinamide nucleotide adenylyltransferases. Studies in *Drosophila* and mammalian neurons have shown the encoded protein can confer protection to damaged neurons. This protection requires enzymatic activity which increases NAD levels and activates a nuclear deacetylase which is the protective molecule. Pseudogenes of this gene are located on chromosomes 1, 3, 4, 14 and 15. [provided by RefSeq, Dec 2011]

Export of annotated variants

 Export variants 

Export the currently filtered list of variants appearing in the right table.

File format:

TSV

VCF

##Number: Consecutively numbered variants
 ##Contig: Contig numbers in karyotypic order, Genomatix
 ##Position: Position on the contig, Customer
 ##Type: Type of variation, Customer
 ##Reference: Sequence for reference, Customer
 ##KnownVariant: Known variant, dbSNP, Build 138
 ##Function: Genomic feature, Eldorado, May 2013
 ##Gene: Preferred symbol for gene, Eldorado, May 2013
 ##Category: Predicted effects, Genomatix
 ##IsDeleterious: Deleterious variant effect prediction, Genomatix
 ##Quality: Quality value for the assertion of the alternative allele (minimum), Customer
 ##Coverage: Coverage (minimum), Customer
 ##MafGenomes: Alternative allele frequency from the whole 1000 genomes project, 1000 Genomes Project, Apr 2012
 ##LiteratureDiseases: Associated disease terms based on literature mining by Genomatix (gene level), LitInspector, Sep 2014
 ##ClinVarDiseases: Associated clinical disease terms based on ClinVar database (gene level), ClinVar, Oct 2014
 ##OmicDiseases: Associated disease terms based on OMIM database (gene level), OMIM, Oct 2014
 ##SampleCase: IV-1 (Case)
 ##SampleControl: III-4 (Control)
 ##SampleControl: III-5 (Control)

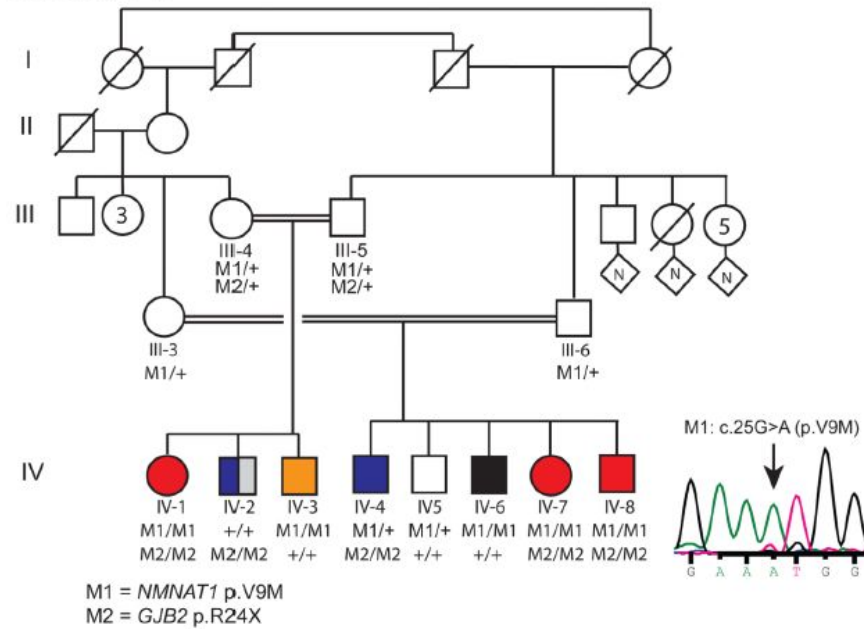
Number	Contig	ChrL	Contig	Cont	Position	Type	Reference	KnownVariat	Function	Gene_GeneL	Gene_GeneS	Category	IsDeleteriou	Quality	Coverage	MafGenome	LiteratureDiseases	ClinVarDiseases	OmicDiseases	SampleCase	SampleCont	SampleControl_2
290	1	1	1684347	insertion	CCCTCCTCCT	0	3utr,intron,c	65220	NADK	insertion	1	304	2	0	Tuberculosis:Dysplasti	Lung cancer			hom	het	het	
607	1	1	6637056	snp	G	1	intron_cs,cd	80835	TAS1R1	missense	1	511	4	0.0093	Ageusia:Herpes Zoster	Malignant melanoma			hom	het	het	
746	1	1	10032156	snp	G	0	cds	64802	NMNAT1	missense	1	728	33	0	Frontotemporal Deme	Leber congenital am	Leber congenital am	hom	het	het	het	
11097	2	2	1440078	snp	C	1	3utr,intron_c	7173	TPO	missense	1	608	10	0	Multiple System Atrop	Deficiency of iodide	Thyroid dyshormono	hom	het	het	het	
21791	3	3	98073591	deletion	TAAAAAAA	0	cds	403278	OR5K4	nonsense	1	203	11	0		Malignant melanoma			hom	het	het	
21793	3	3	98110406	insertion	GAAAAAAA	1	cds	403277	OR5K3	frameshift	1	200	13	0					hom	het	het	
25841	4	4	10446604	snp	T	0	cds,promote	85460	ZNF518B	missense	1	621	14	0		Malignant melanoma			hom	het	het	
25872	4	4	15542617	snp	C	1	cds	57545	CC2D2A	missense	1	586	7	0	Meckel syndrome type	Meckel syndrome, ty	COACH syndrome:Mi	hom	het	het	het	
26408	4	4	42403128	snp	C	1	cds	152573	SHISA3	missense	1	593	9	0	Neoplasms:Cell Transf	Malignant melanoma			hom	het	het	
31745	5	5	112676428	snp	C	1	intron_cs,cs	4163	MCC	splice-site	1	356	5	0	Colorectal carcinoma:N	Carcinoma of colon:L	Colorectal cancer, so	hom	het	het	het	
39823	6	6	84798956	snp	G	1	cds	112609	MRAP2	missense	1	572	13	0.0005	Obesity:Adrenocortica	Malignant melanom	Obesity, susceptibili	hom	het	het	het	
39832	6	6	84896313	deletion	TTTCTT	0	5utr,exon,cd	22832	KIAA1009	deletion	1	728	16	0	Communicable Diseases				hom	het	het	
40963	6	6	142738314	deletion	CTCTTTTC	0	intron_cs,int	57211	GPR126	nonsense	1	571	4	0	Peripheral neuropathy	Lung cancer			hom	het	het	
44143	7	7	65444359	insertion	TGAGAG	0	intron_cs	2990	GUSB	splice-site	1	553	3	0	Mucopolysaccharidosis	Mucopolysaccharido	Mucopolysaccharido	hom	het	het	het	
46432	7	7	138424359	snp	T	1	intron,cds	50617	ATP6V0A4	missense	1	434	3	0.0009	Renal Tubular Acidosis	Renal tubular acidosi	Renal tubular acidosi	hom	het	het	het	
46467	7	7	139164446	snp	G	1	intron_cs,cd	346689	KLRG2	missense	1	402	2	0	Prostate Cancer				hom	het	het	
47886	8	8	10467589	snp	T	1	cds	94137	RP1L1	missense	1	664	12	0	Occult macular dystrop	Malignant melanom	Occult macular dystri	hom	het	het	het	
49527	8	8	67369047	snp	T	1	intron_cs	137872	ADHF1E1	splice-site	1	434	3	0	D-2-hydroxyglutaric aci	Malignant melanoma			hom	het	het	
49733	8	8	73937076	snp	T	0	5utr,cds	7013	TERF1	missense	1	497	8	0	Low Grade Prostatic	Intraepithelial Neoplasia	Acute leukemias:A	hom	het	het	het	
52162	9	9	33255933	snp	A	0	intron_cs	573	BAG1	splice-site	1	505	9	0	Breast cancer:Invasive	Breast Carcinoma:Periampullary	Adenocar	hom	het	het	het	

GeneGrid example 2

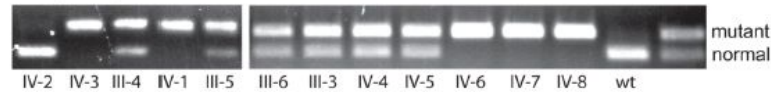
Familial autism analysis

Consanguineous family

a Family 047



b Genotyping 047



Mucopolipidosis type III
 Deafness
 LCA, global developmental delay, autism



LCA, congenital deafness, global developmental delay, autism

Step 2: sample comparison

GeneGrid Variant Annotation
Samples

Filter samples

Import samples & annotate variants

Compare samples

Step 1: Select the type of comparison study:

Trio Cancer **Other**

Step 2: Assign the samples to the groups:

Hint: Just drag an activated sample from the table on the right side and drop it in one of the two groups below. [Read more](#)

Case (0 affected, requires 1 more)

	Name	#	ID

Control (0 not affected)

	Name	#	ID

Study name:

Submit

ID	Input file	Name	Number of non-ref variants	Class	
3072	15KSa_N_ACAGTG_L003...	15KSa_N_ACAGTG_L003_1	907	small	
3071	15KSa_N_ACAGTG_L003...	15KSa_N_ACAGTG_L003	7,893	small	
2614	quartet.variants.annotated.vcf	ISDBM322018_mother	165,359	medium	
2613	quartet.variants.annotated.vcf	ISDBM322017_daughter	142,106	medium	
2612	quartet.variants.annotated.vcf	ISDBM322016_father	162,143	medium	
2611	quartet.variants.annotated.vcf	ISDBM322015_son	170,032	medium	
1861	patient_2.vcf	patient2_tumor	325,272	medium	
1860	patient_2.vcf	patient2_normal	214,581	medium	
1859	patient_6.vcf	patient6_tumor	204,988	medium	
1858	patient_6.vcf	patient6_normal	221,519	medium	
1857	patient_14.vcf	patient14_tumor	203,282	medium	
1856	patient_14.vcf	patient14_normal	195,854	medium	
1855	patient_7.vcf	patient7_tumor	171,276	medium	
1854	patient_7.vcf	patient7_normal	167,692	medium	
1603	LCA047_All_mincov10_sa...	IV-3	Case	65,911	medium
1602	LCA047_All_mincov10_sa...	III-5	Control	60,716	medium
1601	LCA047_All_mincov10_sa...	IV-1	Case	58,079	medium
1600	LCA047_All_mincov10_sa...	III-4	Control	76,209	medium
1599	LCA047_All_mincov10_sa...	IV-2	Control	55,412	medium

Step 2: sample comparison

	Chr	Position (bp)	Variant	Ref allele	Alt allele	Known variant	Genomic feature	Gene
<i>Variant description</i>			<i>Evolutionary conservation</i>			<i>Gene ontology</i>		
<input type="checkbox"/>			<input type="checkbox"/>				<input type="checkbox"/>	
<i>Feature annotation</i>			<i>Regulatory annotation</i>			<i>Comparison summary</i>		
<input checked="" type="checkbox"/>			<input type="checkbox"/>				<input checked="" type="checkbox"/>	
<i>Predicted molecular effects on protein</i>			<i>Experimental evidence</i>					
<input type="checkbox"/>			<input type="checkbox"/>				<input type="checkbox"/>	
<i>Variant quality</i>								
<input type="checkbox"/>			<input type="checkbox"/>					

IV-2 (Control)

≠ Homozygous x

III-4 (Control)

≠ Homozygous x

III-5 (Control)

≠ Homozygous x

Search

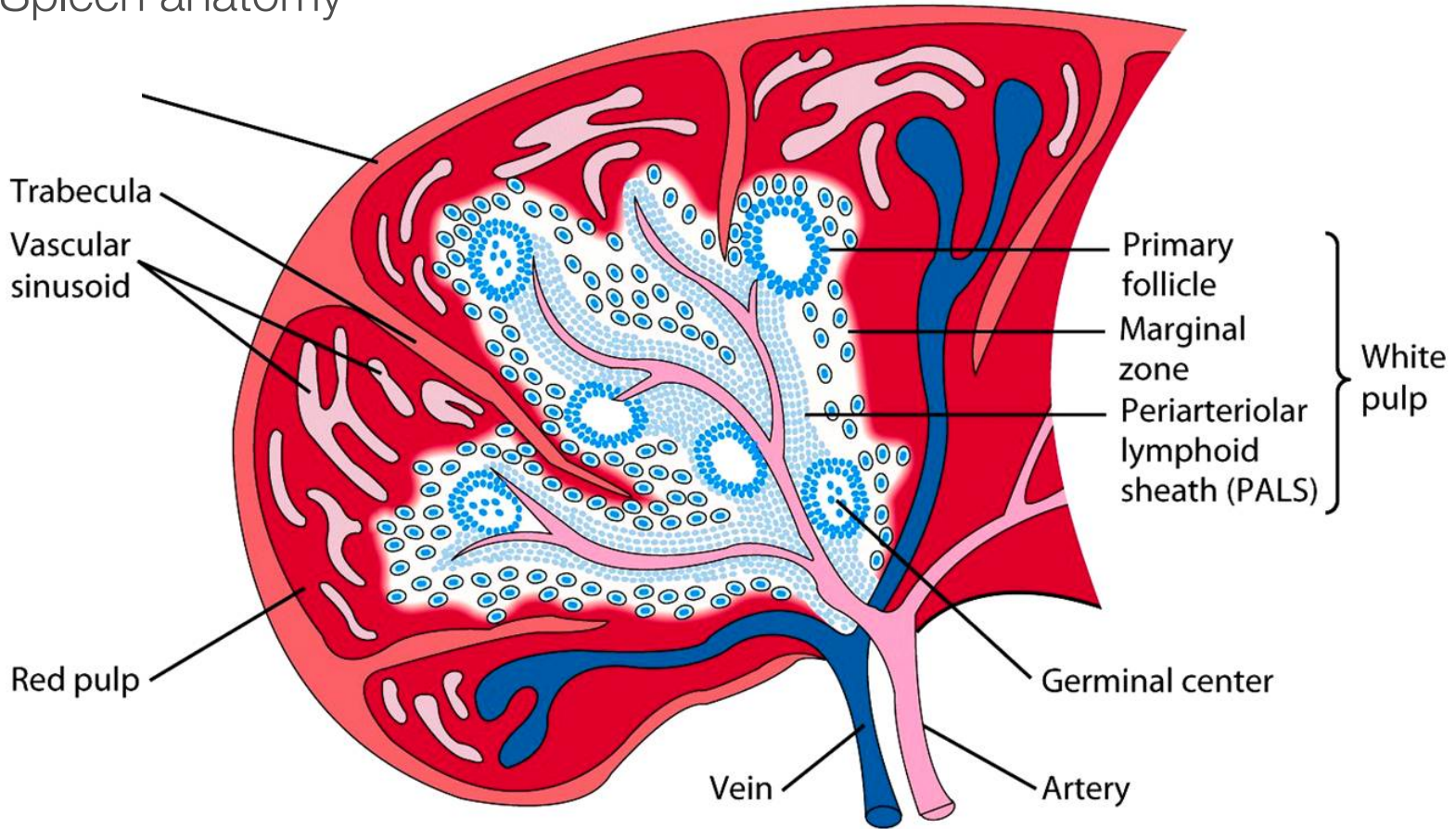
Reset

GeneGrid example 3

Cancer analysis

Splenic marginal zone lymphoma (SMZL)

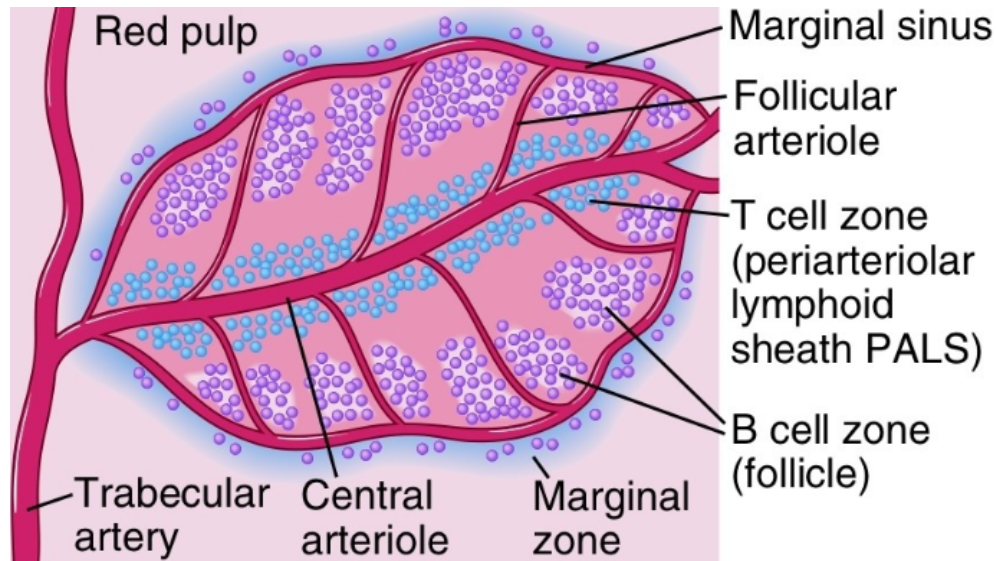
Spleen anatomy



Splenic marginal zone lymphoma (SMZL)

B cell lymphoma consisting of small lymphocytes and larger blasts

- Starts in white pulp of the spleen
- Invades mantle zone of splenic follicles
- Erodes marginal zone
- Ultimately invades red pulp of the spleen



Frequent clonal rearrangements of immunoglobulin genes

Splenic marginal zone lymphoma (SMZL)



Leukemia (2014) **28**, 1334–1340

© 2014 Macmillan Publishers Limited All rights reserved 0887-6924/14

www.nature.com/leu

ORIGINAL ARTICLE

Whole-exome sequencing in splenic marginal zone lymphoma reveals mutations in genes involved in marginal zone differentiation

N Martínez¹, C Almaraz¹, JP Vaqué¹, I Varela², S Derdak³, S Beltran³, M Mollejo⁴, Y Campos-Martin⁴, L Agueda³, A Rinaldi⁵, I Kwee^{5,6,7}, M Gut³, J Blanc³, D Oscier⁸, JC Strefford⁹, J Martinez-Lopez¹⁰, A Salar¹¹, F Sole¹², JL Rodriguez-Peralto¹³, C Diez-Tascón¹⁴, JF García¹⁵, M Fraga¹⁶, E Sebastián¹⁷, J Alvés¹⁸, J Menárguez¹⁹, J González-Carrero²⁰, LF Casado⁴, M Bayes³, F Bertoni^{5,21}, I Gut³ and MA Piris^{1,22}

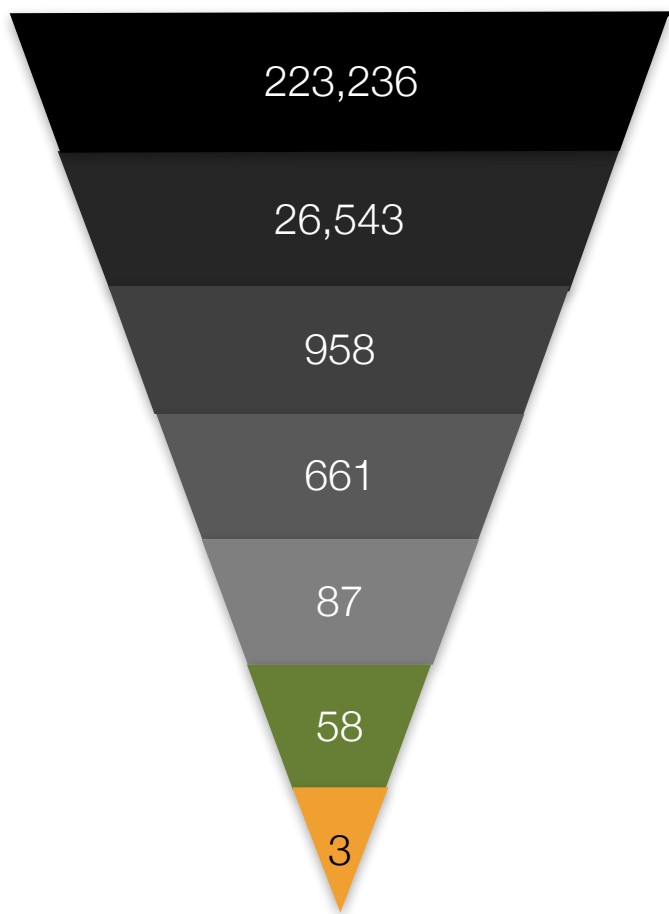
Splenic marginal zone lymphoma (SMZL) is a B-cell neoplasm whose molecular pathogenesis remains fundamentally unexplained, requiring more precise diagnostic markers. Previous molecular studies have revealed 7q loss and mutations of nuclear factor κ B (NF- κ B), B-cell receptor (BCR) and Notch signalling genes. We performed whole-exome sequencing in a series of SMZL cases. Results confirmed that SMZL is an entity distinct from other low-grade B-cell lymphomas, and identified mutations in multiple genes involved in marginal zone development, and others involved in NF- κ B, BCR, chromatin remodelling and the cytoskeleton.

Leukemia (2014) **28**, 1334–1340; doi:10.1038/leu.2013.365

Keywords: whole-exome sequencing; splenic; lymphoma; marginal zone

Analysis example: SMZL in patient 7

Filter summary



total number

somatic variants

select deleterious variants

select rare variants (gAF < 1%)

quality and genotype quality > 30

only high confidence predicted protein effects

select genes associated with
"Splenic Marginal Zone B-Cell Lymphoma" in the
literature

Filtering result

GeneGrid Sample Comparison Pair study patient7 Menu Help Christian Zinser Log out

Filter variants

Add column:

Deleterious variant

= Yes x

Low confidence

= No x

Quality

≥ 30 x

Genotype quality

≥ 30 x

gAF

≤ 0.01 x

Literature diseases

~ Splenic Marginal Zone B-Cell Lymphoc x

Somatic

= Yes x

Filter history

Template filter

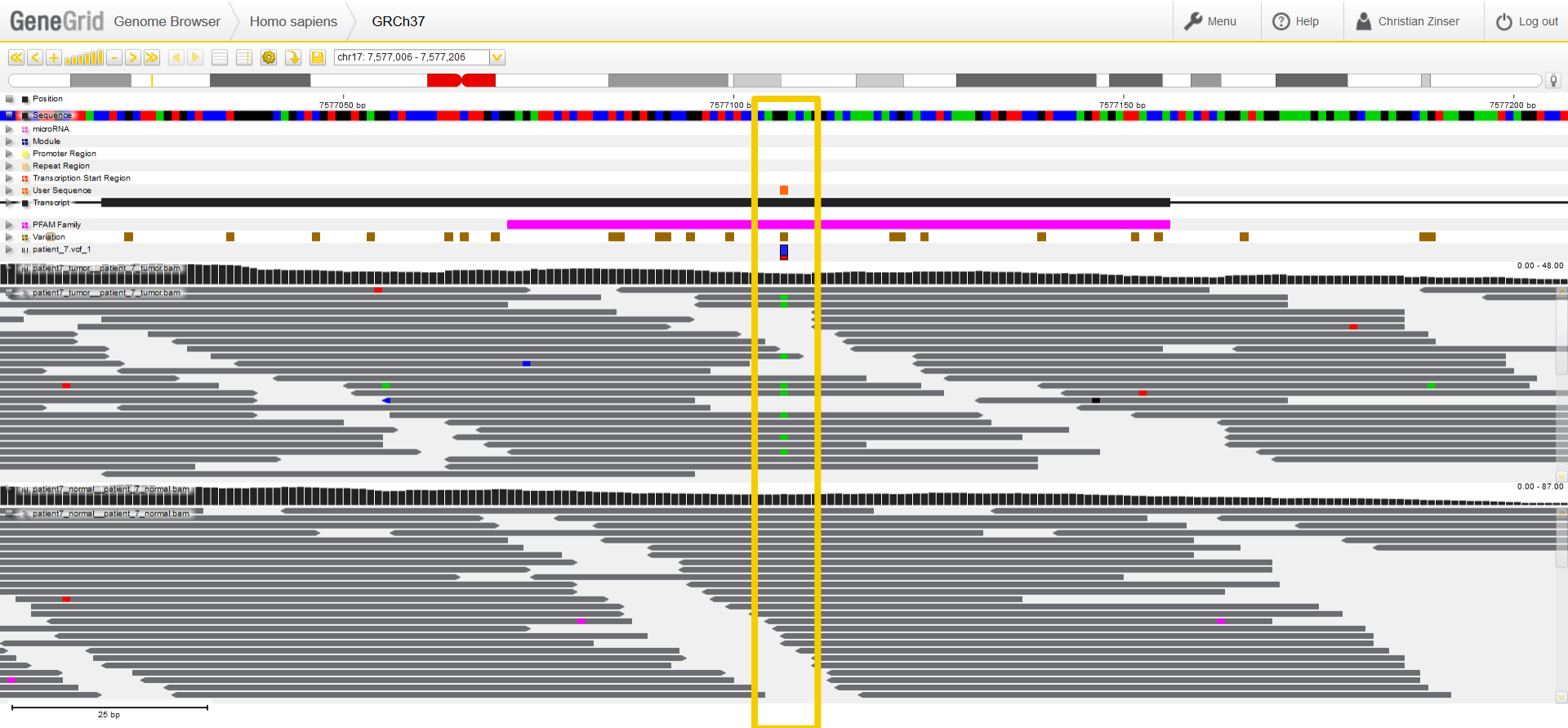
Pathway analysis

Report generator

Export variants

Variant	Ref allele	Alt allele	Known variant	Genomic feature	Gene	Known gene	Effect prediction	Deleterious variant	Low confidence	Consensus variation	Quality	Coverage	Genotype quality	gAF	exacAF	Regulatory evidences	Literature diseases	Diff. pairwise somatic	Somatic	patient7_tumor (Case)	patient7_normal (Control)	
40234	snv	C	G	-	cds • ex...	MYD88	✓	missense	✓	Ser219Cys	222.00	59	99	0.01	x	All	Splenic Marginal Zone B-Cell Lymphoc	x	1	✓	■	□
179694	snv	G	A	-	cds • ex...	TP53	✓	missense	✓	Pro278Ser	152.00	25	99				Li-Fraumeni Synd...	x	1	✓	■	□
179697	del	TTCT	TT	-	cds • ex...	TP53	✓	frameshift	✓	Arg209LysfsTer6	64.50	17	99				Li-Fraumeni Synd...	x	1	✓	■	□

Visualization



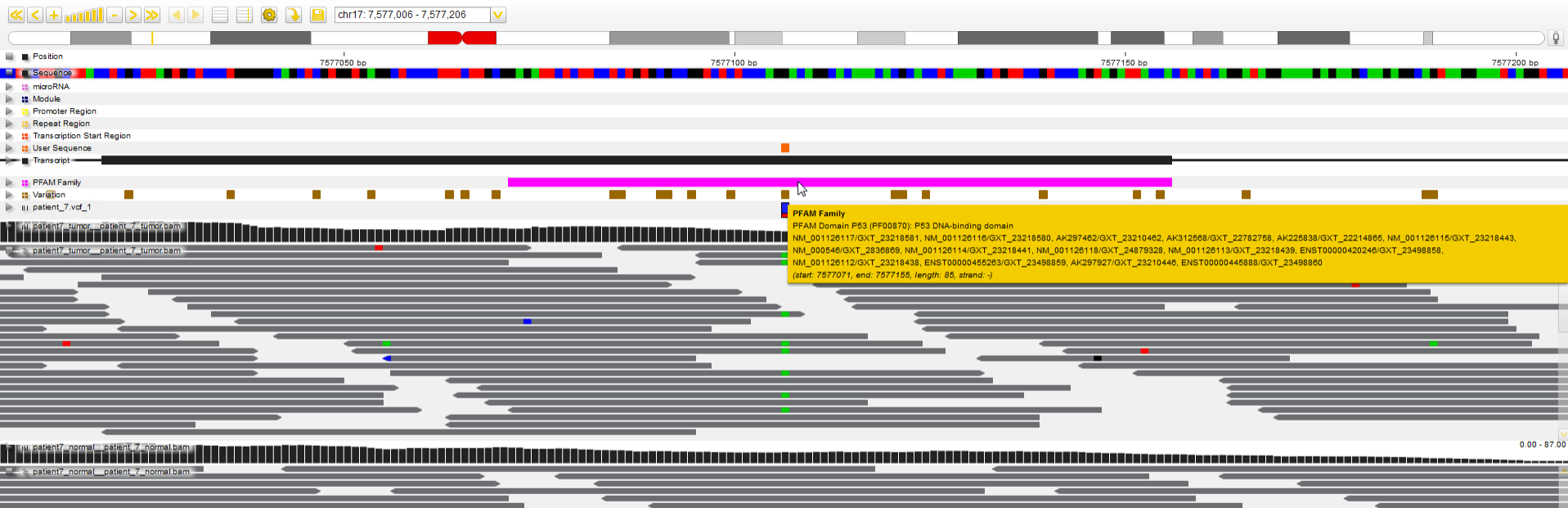
Read coverage at SNP position in TP53 for tumor and control

Visualization



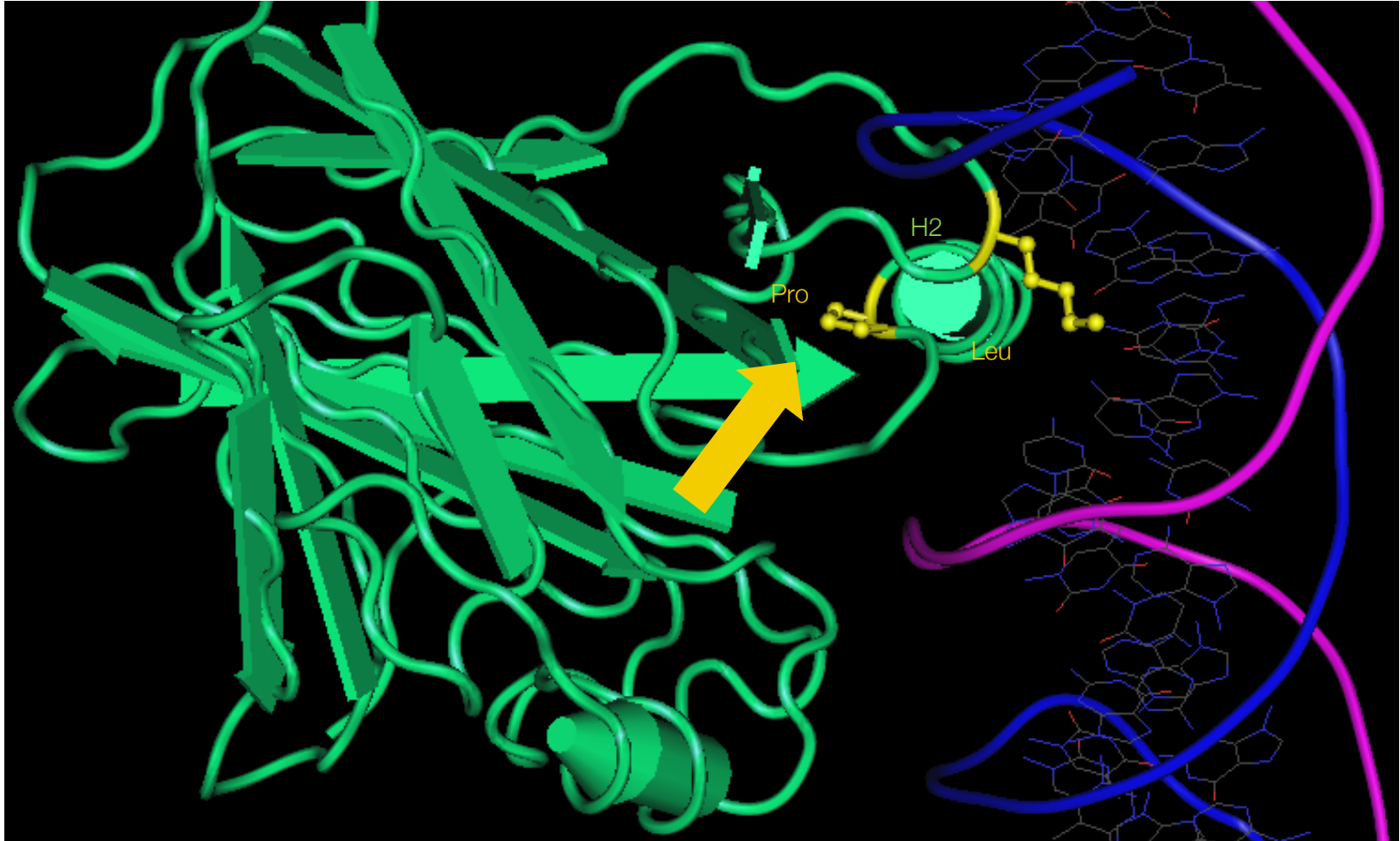
The SNP is located in the DNA-binding domain

Visualization



Sample details (2)			Transcript effects (22)				dbSNP	ClinVar	Somatic mutations (19)		COSMIC (12)		Literature diseases (100+)			Clinical diseases (39)		Gene details		Citations (94)	
Ref allele	Alt allele	Ref protein sequence	Alt protein sequence	Gene	Transcript accession	Transcript version	Strand	Type	Genomic feature	Effect prediction	Deleterious	Low confidence	Variation coding	Variation protein	Coding position (bp)	Relative coding position	BLOSUM	SIFT score			
G	A	P	S	TP53	ENST00000455263	2	-	protein coding	cds • exon	missense	✓	-	832C>T	Pro278Ser	832	0.80	-1	0.02			
G	A	P	S	TP53	ENST00000509690	1	-	protein coding	cds • exon	missense	✓	✓	436C>T	Pro146Ser	436	0.73	-1	0.03			
G	A	P	S	TP53	NM_001126112	2	-	protein coding	cds • exon	missense	✓	-	832C>T	Pro278Ser	832	0.70	-1	0.03			
G	A	P	S	TP53	NM_001276761	1	-	protein coding	cds • exon	missense	✓	-	715C>T	Pro239Ser	715	0.67	-1	0.02			
G	A	P	S	TP53	NM_000546	5	-	protein coding	cds • exon	missense	✓	-	832C>T	Pro278Ser	832	0.70	-1	0.03			
G	A	P	S	TP53	NM_001276760	1	-	protein coding	cds • exon	missense	✓	-	715C>T	Pro239Ser	715	0.67	-1	0.02			

P53 core domain in complex with DNA



Source: NCBI structure, MMDB ID: 106061

- ▶ the identified mutation could affect DNA binding or DNA affinity