



GeneGrid: finding disease-causing variants in NGS data

Susan M. Dombrowski, PhD
26 Oct 2018

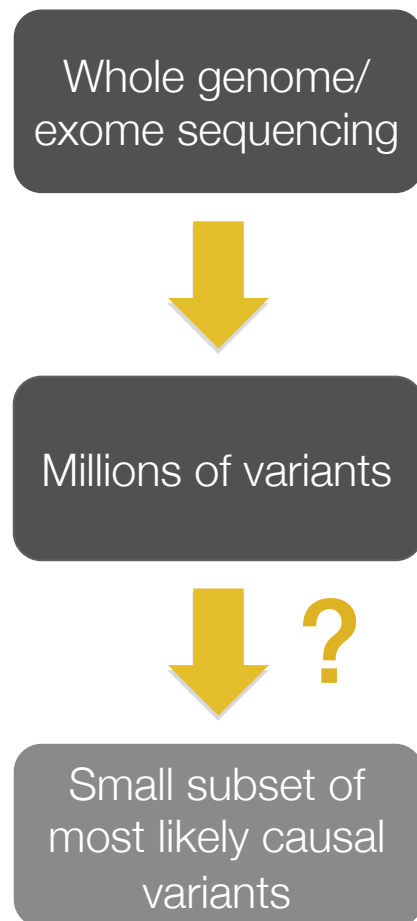
Basis

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

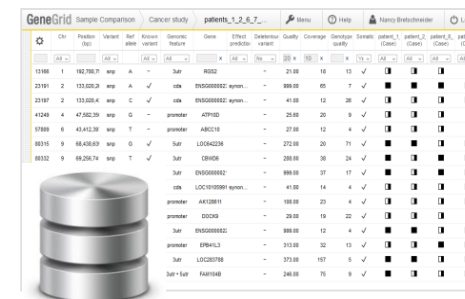
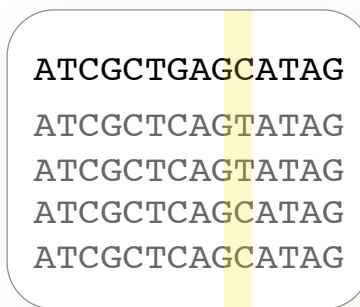
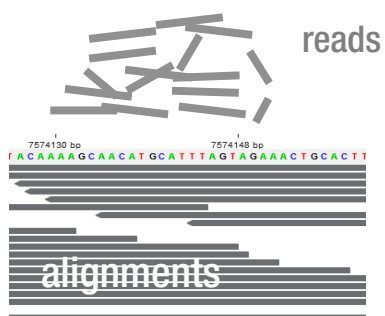
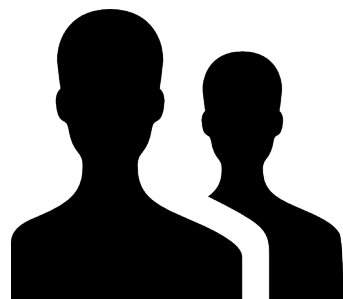
Identifying causal variants is crucial for the diagnostics of rare and common diseases

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

Which are the relevant ones?



General workflow



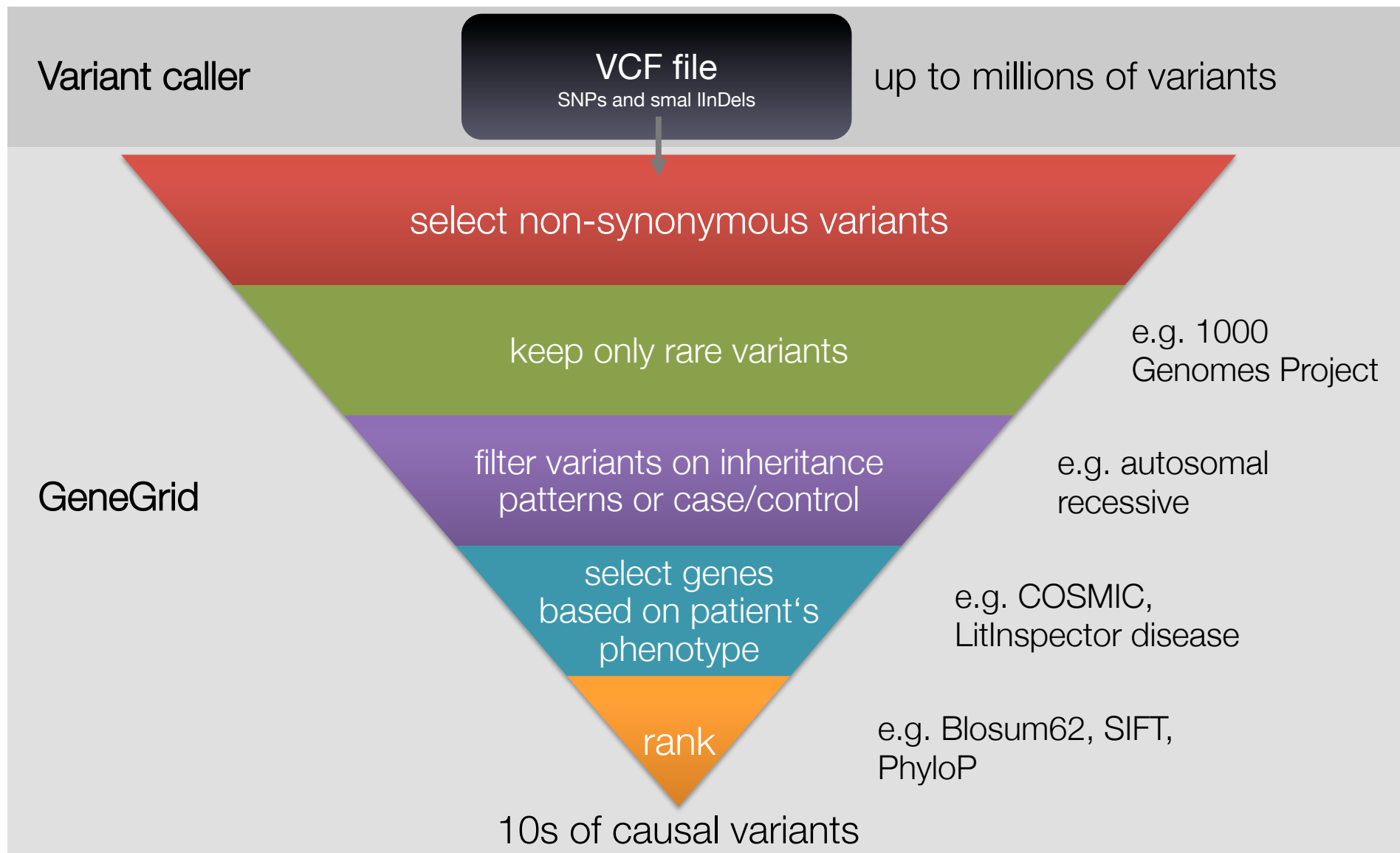
Sequencing library

BAM file with read positions

VCF file with variant information

Short list of potentially interesting variants

Finding the needle in the haystack



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

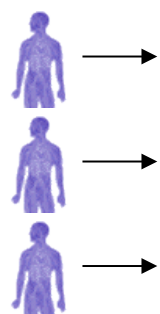
Selected GeneGrid filter fields

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

available online

Chr	Position (bp)	Variant	Ref allele	Known variant	Genomic feature	Gene	Known gene	Effect prediction	Deleterious variant	Quality	Coverage	gAF	Compound heterozygous	Offspring inheritance	9-1 (Case)	9-4 (Control)	9-5 (Control)
1	752,721	snp	A	✓	promoter + interg...	FAM87B	✓	-	-	294.00	1	0.68	-	-	■	■	■
2	752,854	snp	T	✓	-	FAM87B	✓	missense	✓	205.00	1	0.74	-	-	■	■	■
3	899,876	snp	A	✓	cds	FAM41C	✓	missense	✓	287.00	8	0.14	-	-	■	■	■
4	847,250	snp	G	✓	intron (cs)	LOC284600	✓	-	-	76.00	1	1.00	-	-	■	■	■
5	848,738	snp	C	✓	cds	LOC284600	✓	missense	✓	15.65	4	0.15	-	-	■	■	■
6	849,998	snp	A	✓	3'ut + intron (cs)	LOC284600	✓	-	-	73.00	1	0.81	-	-	■	■	■
7	850,760	snp	C	✓	intergenic	-	-	-	-	40.13	2	0.58	-	-	■	■	■
8	852,875	snp	C	✓	3'ut + intron (cs)	LOC100130417	✓	splice site	✓	72.00	1	0.51	-	-	■	■	■
9	852,954	snp	T	✓	3'ut + intron (cs)	LOC100130417	✓	-	-	79.00	1	0.75	-	-	■	■	■
10	853,954	snp	C	✓	3'ut + intron (cs)	LOC100130417	✓	-	-	45.13	2	0.51	-	-	■	■	■
11	856,319	snp	G	✓	intron (cs)	ENSG00000269171	✓	-	-	111.00	1	0.96	-	-	■	■	■
12	856,319	snp	G	✓	intron (cs)	SAMD11	✓	-	-	111.00	1	0.96	-	-	■	■	■
13	856,511	ins	C...	-	promoter	ENSG00000269171	-	-	-	95.00	1	-	-	-	■	■	■
14	856,511	ins	C...	-	intron (cs)	SAMD11	✓	-	-	95.00	1	-	-	-	■	■	■
15	871,334	snp	G	✓	intron (cs)	SAMD11	✓	-	-	282.00	8	0.51	-	-	■	■	■
16	876,314	snp	G	✓	3'ut + cds	SAMD11	✓	synonymous	-	49.50	4	0.05	-	-	■	■	■
17	876,676	snp	G	✓	3'ut	NOCDL	✓	-	-	315.00	14	0.90	-	-	■	■	■
18	876,676	snp	G	✓	3'ut	SAMD11	✓	-	-	315.00	14	0.90	-	-	■	■	■
19	876,687	snp	T	✓	3'ut	NOCDL	✓	-	-	309.00	13	0.93	-	-	■	■	■
20	876,687	snp	T	✓	3'ut	SAMD11	✓	-	-	309.00	13	0.93	-	-	■	■	■



ATCGCTGAGCATAG
 ATCGCTCAGTATAG
 ATCGCTCAGTATAG
 ATCGCTCAGCATAG
 ATCGCTCAGCATAG

SNPs/InDels

view, compare
filter, sort



GeneGrid database

Pre-knowledge database
 annotations
 disease & medical info



import

annotate



Results & Reports

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 (LCA)

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

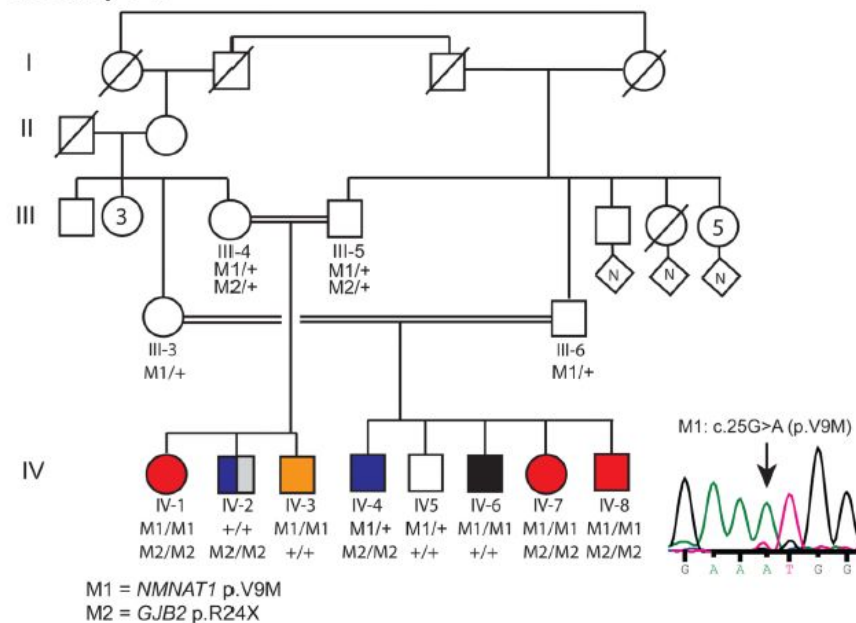
⁶Kallam Anji Reddy Molecular Genetics Laboratory, LV Prasad Eye Institute (LVPEI), Kallam Anji Reddy Campus, LV Prasad Marg, Hyderabad, India

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

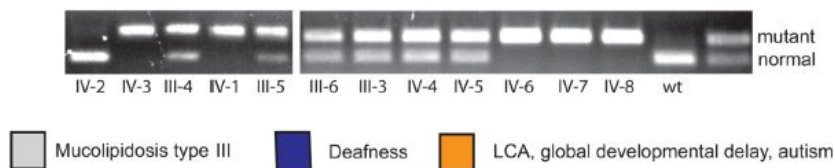
⁸Institut National de la Santé et de la Recherche Médicale, U968, Paris, France

Consanguineous family

a Family 047



b Genotyping 047



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

Filter samples
GeneGrid Variant Upload LCA047_Trio_Demo.38(1).vcf.gz

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 match the human genome build GRCh37/hg19 or GRCh38/hg38. [Read more](#)

Step 1: Select import workspace:

MyProject (private) ▼

Step 2: 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 3: Select the variant file from your computer:

Browse...

 No file selected.

Submit
Reset

Compare samples

Associate alignment files

Variant Upload Job

Your analysis has been submitted successfully. The input file `LCA047_Trio_Demo.38(1).vcf.gz` was uploaded to GeneGrid. The analysis can take up to several hours depending on the size of the input data. The output will become available in the [Result Management](#).

Thank you for using GeneGrid.

32%

Analysis is running...

Refresh

Analysis progress

2018-10-19 15:34:03	Processing variants on chromosome 5
2018-10-19 15:34:01	Processing variants on chromosome 4
2018-10-19 15:33:58	Processing variants on chromosome 3
2018-10-19 15:33:54	Processing variants on chromosome 2
2018-10-19 15:33:49	Processing variants on chromosome 1
2018-10-19 15:33:49	Annotating variants
2018-10-19 15:33:48	Compiling general statistics
2018-10-19 15:33:48	Detected genome build GRCh38
2018-10-19 15:33:48	Matched 25 contigs with variants
2018-10-19 15:33:47	Validating variants with genome reference sequence
2018-10-19 15:33:46	Creating index for VCF file
2018-10-19 15:33:44	Sorting VCF file
2018-10-19 15:33:43	Detected input file of size 3 MB
2018-10-19 15:33:43	Preparing input file

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:

Trio Cancer Other

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 3: filtering



GeneGrid example: trio

*Apply filtering criteria to search for low frequency (rare) variants that are deleterious and found in known (annotated) genes.

*This leaves the top 6% variants in the list.

The screenshot displays the GeneGrid interface. On the left is a 'Filter variants' panel with the following settings:

- Known gene: Yes
- Deleterious variant: Yes
- gAF: ≤ 0.01
- IV-1 (Case): Homozygous
- III-4 (Control): Heterozygous
- III-5 (Control): Heterozygous

Buttons for 'Search' and 'Reset' are at the bottom of the filter panel. The main table on the right shows a list of variants with columns: Gene symbol, Known gene, Effect prediction, Deleterious variant, Consensus variation, Quality, Coverage, gAF, exacAF, Regulatory evidences, Compound heterozygosity, Offspring inheritance, and genotypes for IV-1_1 (Case), III-5_2 (Control), and III-4_1 (Control). The status bar at the bottom right indicates 'Viewing: 1 to 45', 'Filtered: 45', and 'Total variants: 121,936'.

Filter history and template filters

Speed up the analysis and facilitate sample comparisons

Filter variants

Filter history

Today		
3 seconds ago	121,936	100.00%
<i>Unfiltered</i>		
31 seconds ago	106	0.09%
exacAF • Deleterious variant • Known gene • IV-1_1 (Case) • III-5_2 (Control)		
7 minutes ago	45	0.04%
exacAF • Deleterious variant • Known gene • IV-1_1 (Case) • III-5_2 (Control) • III-4_1 (Control)		
7 minutes ago	655	0.54%
exacAF • Deleterious variant • Known gene • IV-1_1 (Case)		
7 minutes ago	3,361	2.76%
exacAF • Deleterious variant • Known gene		
7 minutes ago	18,455	15.13%
Deleterious variant • Known gene		
7 minutes ago	113,625	93.18%
Known gene		

Template filter

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

Template title:

Save Manage all

Available templates

135 TrioAdvancedFilter

Clinical significance • Diff. between groups • Offspring inheritance

filters that are routinely used, e.g.

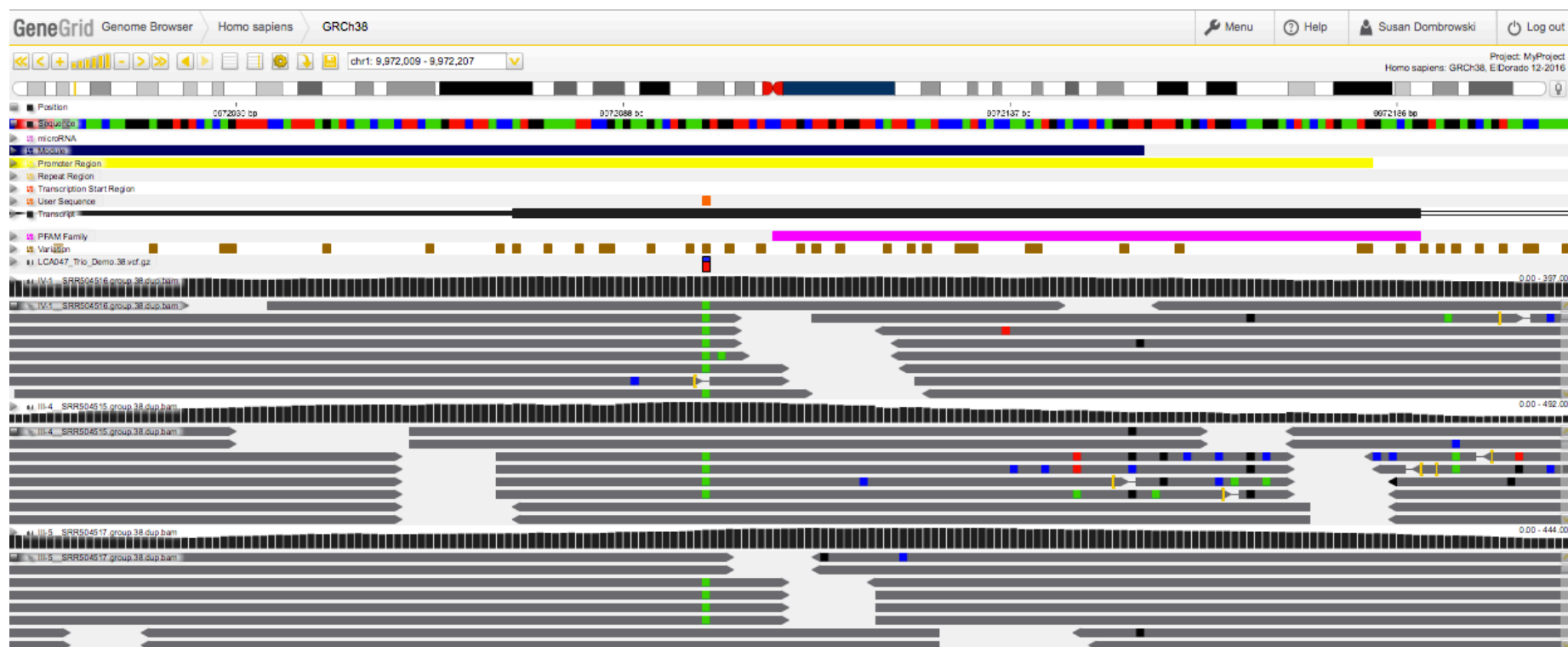
- low read coverage
- low mapping quality
- low sequencing frequency (gAF)
- missing gene annotations
- Mendelian inheritance patterns
- missing control genotypes



as template for use in other comparisons

Visualization in the Genomatix Genome Browser

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	Yes	x	x	0.01 x	Leber Congenital Amaurosis [Leber Congenital Amaurosis, 33]
1	10,032,156	snp	G	-	cds	NMNAT1	✓	missense	✓	728.00	33		Frontotemporal Dementia • Leber Congenital Amaurosis • Anet



Additional annotation / filter columns: Link-outs to transcript annotation, dnSNP, ClinVar, COSMIC, Biomedical literature and Clinical information.

Variant	Ref allele	Alt allele	Known variant	Genomic feature	Gene symbol	Known gene	Effect prediction	Deleterious variant	Consensus variation	Quality	Coverage	gAF	exacAF	Regulatory evidences	Compound heterozygo	Offspring inheritance	IV-1 (Case)	III-4 (Control)	III-5 (Control)		
771	snv	G	A	✓	ods • exon	NMNAT1	✓	missense	✓	Val9Met	728.00	35					-	hom	■	■	■

Sample	Variant	Chr	Position (bp)	Variant	Zygoty	Ref	Alt	Alt2 allele	Genotype	Gene symbol	Known gene	Effect prediction	Deleterious	Quality	Coverage	Ref coverag	Alt coverag	Ref fraction	Alt fraction	gAF
III-4	396	1	9,972,098	snv	het	G	A		■	NMNAT1	✓	missense	✓	728.00	35	16	19	0.46	0.54	
III-5	694	1	9,972,098	snv	het	G	A		■	NMNAT1	✓	missense	✓	728.00	98	59	39	0.60	0.40	
IV-1	709	1	9,972,098	snv	hom	G	A		■	NMNAT1	✓	missense	✓	728.00	116	1	115	0.01	0.99	

Diff. between groups
 Diff. in case group

Annotation detail views and links: ClinVar

ClinVar

Search

[Create alert](#)
[Advanced](#)

Home | About ▾ | Access ▾ | Help ▾ | Submit ▾ | Statistics ▾ | FTP ▾

Gene Tabular ▾ 100 per page ▾ Sort by Location ▾

Customize this list...

Clinical significance

- Conflicting interpretations (0)
- Benign (0)
- Likely benign (2)
- Uncertain significance (0)
- Likely pathogenic (1)
- Pathogenic (12)
- Risk factor (0)

Review status

- Practice guideline (0)
- Expert panel (0)
- Multiple submitters (1)
- Single submitter (4)
- At least one star (5)
- Conflicting interpretations (0)

Allele origin

- Germline (14)
- De novo (0)
- Somatic (0)

Method type

- Research (1)
- Literature only (9)
- Clinical testing (6)


Search results


Items: 14

	Variation <i>Location</i>	Gene(s)	Condition(s)	Clinical significance <small>(Last reviewed)</small>	Review status
1.	<input type="checkbox"/> NMNAT1, TRP169TER	NMNAT1	Leber congenital amaurosis 9	Pathogenic <small>(Sep 1, 2012)</small>	no assertion criteria provided
2.	<input type="checkbox"/> NM_022787.3(NMNAT1):c.25G>A (p.Val9Met) <small>GRCh37: Chr1:10032156 GRCh38: Chr1:9972098</small>	NMNAT1	Leber congenital amaurosis 9	Pathogenic <small>(Sep 1, 2012)</small>	no assertion criteria provided
3.	<input type="checkbox"/> NM_022787.3(NMNAT1):c.115+3A>G <small>GRCh37: Chr1:10032249 GRCh38: Chr1:9972191</small>	NMNAT1	Leber congenital amaurosis 9	Likely benign <small>(Jun 23, 2017)</small>	criteria provided, single submitter
4.	<input type="checkbox"/> NM_022787.3(NMNAT1):c.451G>T (p.Val151Phe) <small>GRCh37: Chr1:10042370 GRCh38: Chr1:9982312</small>	NMNAT1	Leber congenital amaurosis 9	Pathogenic <small>(Sep 1, 2012)</small>	no assertion criteria provided
5.	<input type="checkbox"/> NM_022787.3(NMNAT1):c.457C>G (p.Leu153Val)	NMNAT1	Leber congenital amaurosis 9	Pathogenic <small>(Sep 1, 2012)</small>	no assertion criteria provided

Report generator

GeneGrid Variant Report NMNAT1

 Print report

 Download report ▼

The report is available for download in the PDF format.

Note: The result should be communicated by a human geneticist or by a genetic counselor.

Additionally, the PDF report can be sent as attachment directly to your mail address.

NMNAT1

October 19, 2018

DNA variants

Summary

This report consists of 4 variants (ordered by genomic position):

1. *NMNAT1* | snv | rs387907294 | 25G>A / Val9Met | pathogenic
2. *ACTN3* | snv | rs1815739 | 1858C>T / Arg620Ter | 1729C>T / Arg577Ter | benign / pathogenic /
3. *GJB2* | snv | rs104894396 | 71G>A / Trp24Ter | pathogenic
4. *ARSA* | snv | rs743616 | 1178C>G / Thr393Ser | 920C>G / Thr307Ser | 46C>G / Thr16Ser | pathogenic / benign

Val9Met in *NMNAT1*

Variant description

The indicated snv is located on chromosome 1 at position 9,972,098 bp. It overlaps the coding sequence of at least one transcript of gene *NMNAT1*. It overlaps a non-coding transcript (without an open reading frame or incomplete annotation) for 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): 35 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: 116 reads, QUAL: 728.00, GQ: 127)

Control distribution (2/2)

- Heterozygous variant:
 1. III-4 (COV: 35 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 Val9Met is based on 9 annotated transcripts for that gene locus. The BLOSUM62 substitution matrix reports a score of 1 for this alteration.

Known variant

The variant is reported in dbSNP (ID rs387907294).

Clinical significance

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
##OimDiseases: Associated disease terms based on OMIM database (gene level), OMIM, Oct 2014
##SampleCase: IV-1 (Case)
##SampleControl: III-4 (Control)
##SampleControl: III-5 (Control)
```

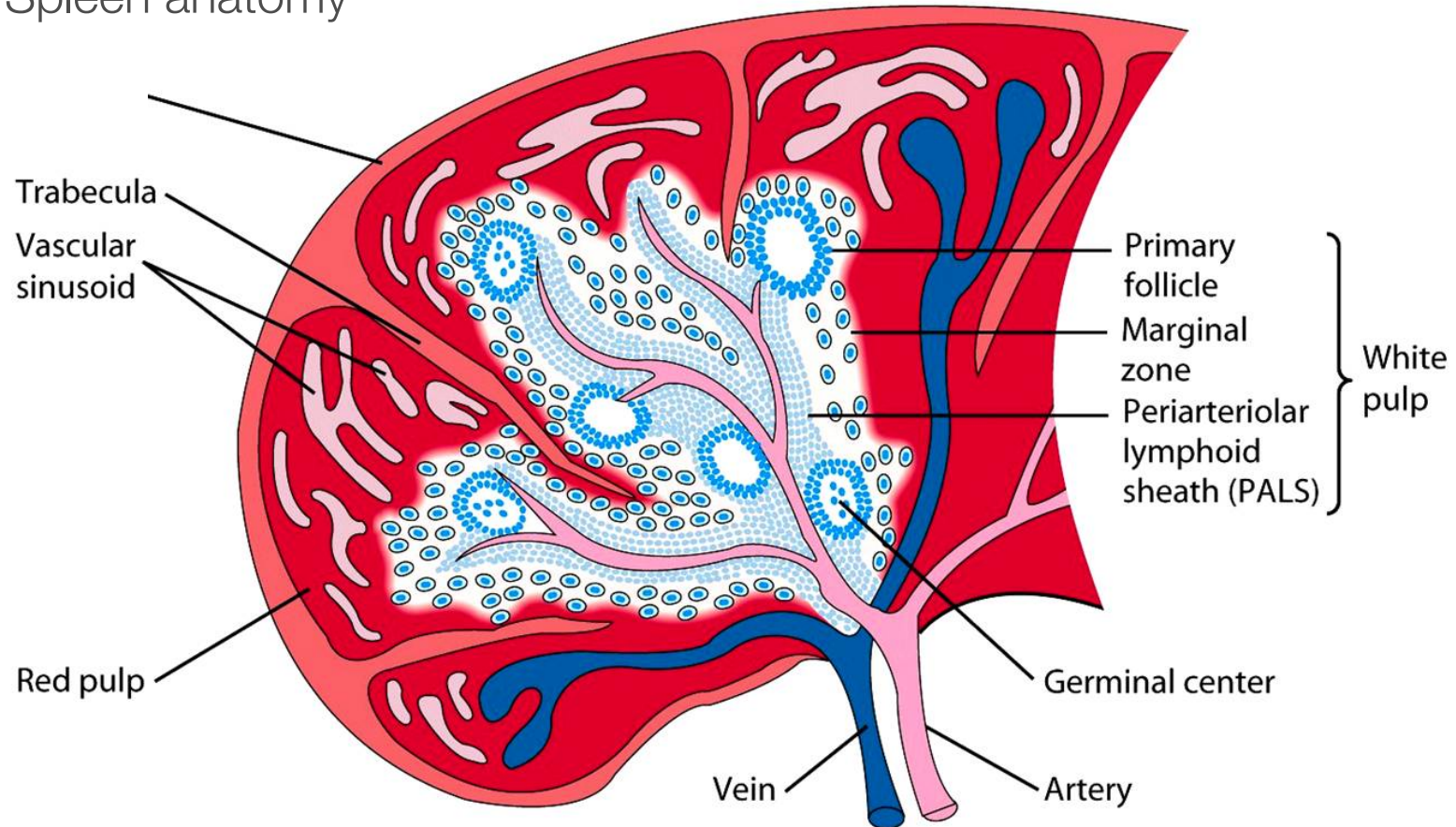
Number	Contig	Chr	Contig_Cont	Position	Type	Reference	KnownVariat	Function	Gene_GeneI	Gene_GeneS	Category	IsDeleteriou	Quality	Coverage	MafGenome	LiteratureDiseases	ClinVarDiseases	OimDiseases	SampleCase	SampleConti	SampleControl_2
290	1	1	1	1684347	insertion	CCCTCCTCCT	0	3utr,intron,c	65220	NADK	insertion	1	304	2	0	Tuberculosis:Dysplastii	Lung cancer		hom	het	het
607	1	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	1	10032156	snp	G	0	cds	64802	NMNAT1	missense	1	728	33	0	Frontotemporal Deme	Leber congenital am	Leber congenital am	hom	het	het
11097	2	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
21791	3	3	3	98073591	deletion	TAAAAAAA	0	cds	403278	OR5K4	nonsense	1	203	11	0		Malignant melanoma		hom	het	het
21793	3	3	3	98110406	insertion	GAAAAAAA	1	cds	403277	OR5K3	frameshift	1	200	13	0				hom	het	het
25841	4	4	4	10446604	snp	T	0	cds,promote	85460	ZNF518B	missense	1	621	14	0		Malignant melanoma		hom	het	het
25872	4	4	4	15542617	snp	C	1	cds	57545	CC2D2A	missense	1	586	7	0	Meckel syndrome type	Meckel syndrome, ty	COACH syndrome:M	hom	het	het
26408	4	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	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
39823	6	6	6	84798956	snp	G	1	cds	112609	MRAP2	missense	1	572	13	0.0005	Obesity:Adrenocortica	Malignant melanom	Obesity, susceptibili	hom	het	het
39832	6	6	6	84896313	deletion	TTTCTT	0	5utr,exon,cd	22832	KIAA1009	deletion	1	728	16	0	Communicable Diseases			hom	het	het
40963	6	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	7	65444359	insertion	TGAGAG	0	intron_cs	2990	GUSB	splice-site	1	553	3	0	Mucopolysaccharidosis:	Mucopolysaccharido	Mucopolysaccharido	hom	het	het
46432	7	7	7	138424359	snp	T	1	intron_cds	50617	ATPGV0A4	missense	1	434	3	0.0009	Renal Tubular Acidosis	Renal tubular acidosi	Renal tubular acidosi	hom	het	het
46467	7	7	7	139164446	snp	G	1	intron_cs,cd	346689	KLRG2	missense	1	402	2	0	Prostate Cancer			hom	het	het
47886	8	8	8	10467589	snp	T	1	cds	94137	RP11	missense	1	664	12	0	Occult macular dystrop	Malignant melanom	Occult macular dystri	hom	het	het
49527	8	8	8	67369047	snp	T	1	intron_cs	137872	ADHFE1	splice-site	1	434	3	0	D-2-hydroxyglutaric aci	Malignant melanoma		hom	het	het
49733	8	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
52162	9	9	9	33255933	snp	A	0	intron_cs	573	BAG1	splice-site	1	505	9	0	Breast cancer:Invasive	Breast Carcinoma:Peri	ampullary Adenocar	hom	het	het

GeneGrid example 2

Cancer analysis

Splenic marginal zone lymphoma (SMZL)

Spleen anatomy

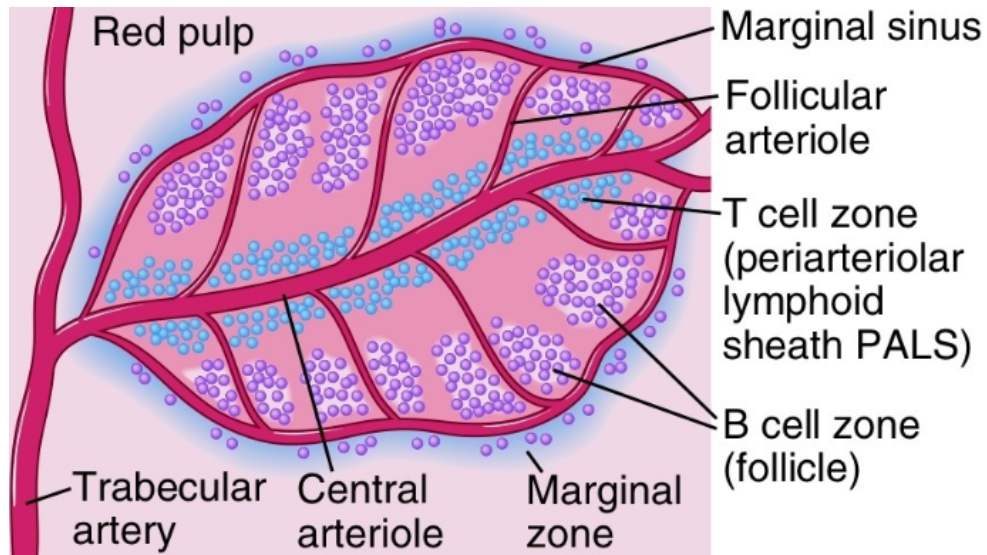


Source: Northern Arizona University <http://www2.nau.edu/~fpm/immunology/spleen1.jpg>

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 Alvéz¹⁸, J Menárguez¹⁹, J González-Carrero²⁰, LF Casado⁴, M Bayes³, F Bertoni^{5,21}, I Gut³ and MA Piris^{1,22}

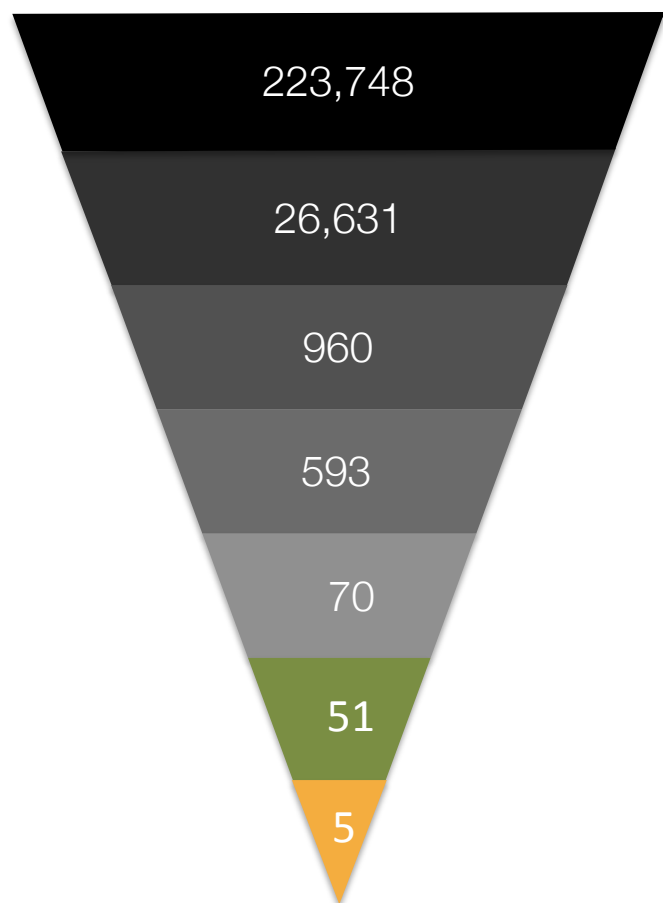
Splenic marginal zone lymphoma (SMZL) is a B-cell neoplasm whose molecular pathogenesis remains fundamentally unexplained, requiring more precise diagnostic markers. Previous molecular studies have revealed 7q loss and mutations of nuclear factor κ B (NF- κ B), B-cell receptor (BCR) and Notch signalling 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 (exacAF < 1%)

quality and genotype quality > 30

only high-confidence predicted protein effects (SIFT < .05)

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

Filtering result

GeneGrid Sample Comparison Pair study patient7_tumor_vs_normal

Filter variants ▼

Add column:

Deleterious variant

– Yes x

Quality

≥ 30 x

Genotype quality

≥ 30 x

exacAF

≤ .01 x

SIFT

≤ .05 x

Literature diseases

~ B Cell Lymphoma [B-cell lymphom] x

Somatic

– Yes x

The currently active filter returned 5 out of 223,748 total rows.

Variant	Ref allele	Alt allele	Known variant	Genomic feature	Gene symbol	Known gene	Effect prediction	Deleterious variant	Consensus variation	Quality	Coverage	Genotype quality	gAF	exacAF	SIFT
40261	snv	C	G	-	cds • exon • i...	MYD88	✓	missense	✓	Ser219Cys	222.00	59	99		0.01
161947	snv	A	G	✓	cds • exon • ...	IGHV1-69	✓	missense	✓	Val10Ala	222.00	98	99		0.01
180094	snv	G	A	-	cds • exon • i...	TP53	✓	missense	✓	Pro278Ser	152.00	25	99		0.02
180097	del	TTCT	TT	-	cds • exon • i...	TP53	✓	frameshift	✓	Arg209Lysfs...	64.50	17	99		
197660	del	TGC	T	-	cds • 5utr • e...	LYL1	✓	frameshift	✓	Ser68LysfsT...	72.50	8	38		

Visualization in the Genomatix Genome Browser



Read coverage at SNP position chr17: 7,577,106 (GRCh37)
in TP53 for tumor and control

Visualization



The SNP is located in the P53 DNA-binding domain

Visualization in the Genomatix Genome Browser

GeneGrid Genome Browser | Homo sapiens | GRCh37 | chr17: 7,577,006 - 7,577,206

Menu | Help | Christian Zinser | Log out

Position: 7577050 bp, 7577100 bp, 7577150 bp, 7577200 bp

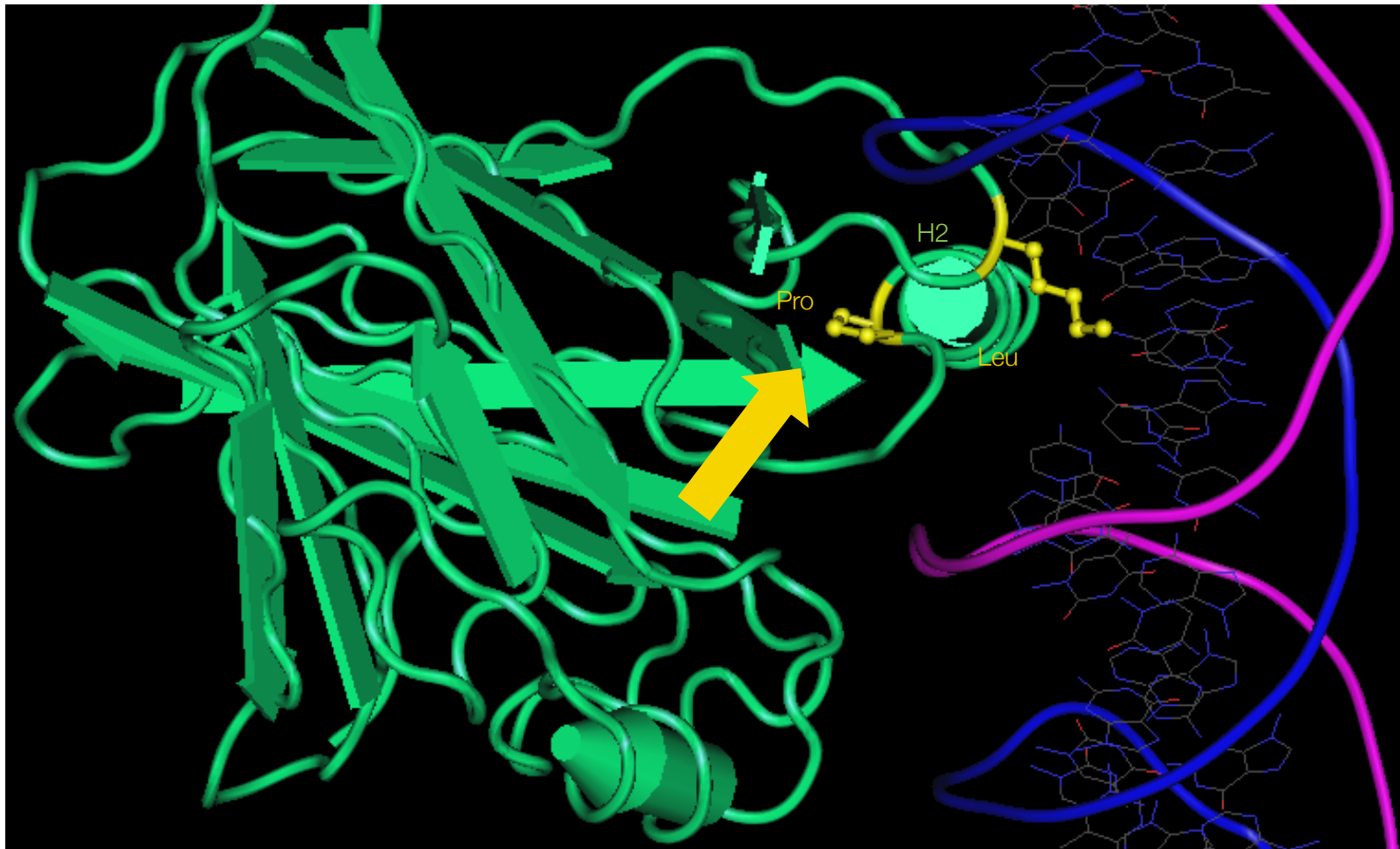
Sequence | microRNA | Module | Promoter Region | Repeat Region | Transcription Start Region | User Sequence | Transcript

PFAM Family | Variation | patient7_tumor_patient7_tumorbam

PFAM Family
 PFAM Domain P53 (PF00870): P53 DNA-binding domain
 NM_001126117/GXT_23218581, NM_001126116/GXT_23218580, AK297462/GXT_23210462, AK312568/GXT_22782758, AK225838/GXT_22214865, NM_001126115/GXT_23218443, NM_000549/GXT_2896889, NM_001126114/GXT_23218441, NM_001126118/GXT_24879328, NM_001126113/GXT_23218439, ENST00000420246/GXT_23498858, NM_001126112/GXT_23218435, ENST00000455263/GXT_23498859, AK297927/GXT_23210446, ENST00000445888/GXT_23498860
 (start: 7577071, end: 7577155, length: 85, strand: -)

Sample details (2)		Transcript effects (66)				dbSNP (1)	ClinVar	Somatic mutations (20)	COSMIC (12)	Literature diseases (100+)	Clinical diseases (44)		Gene details	Citations (100)	Literature tissues (100+)	Canonical		
Ref allele	Alt allele	Ref protein sequence	Alt protein sequence	Gene symbol	Transcript accession	Transcript version	Strand	Type	Genomic feature	Effect prediction	Deleterious	Low confidence	Variation coding	Variation protein	Coding position	Relative coding	BLOSUM	SIFT score
G	A	P	S	TP53	NM_001126112	2	-	protein cod	cds + exon	missense	✓	-	832C>T	Pro278Ser	832	0.70	-1	0.03
G	A	P	S	TP53	NM_000546	5	-	protein cod	cds + exon	missense	✓	-	832C>T	Pro278Ser	832	0.70	-1	0.03
G	A	P	S	TP53	NM_001126113	2	-	protein cod	cds + exon	missense	✓	-	832C>T	Pro278Ser	832	0.80	-1	0.02
G	A	P	S	TP53	NM_001126114	2	-	protein cod	cds + exon	missense	✓	-	832C>T	Pro278Ser	832	0.81	-1	0.02
G	A	P	S	TP53	ENST00000269:	4	-	protein cod	cds + exon	missense	✓	-	832C>T	Pro278Ser	832	0.70	-1	0.03
G	A	P	S	TP53	ENST00000359:	4	-	protein cod	cds + exon	missense	✓	-	832C>T	Pro278Ser	832	0.81	-1	0.02
G	A	P	S	TP53	ENST00000420:	2	-	protein cod	cds + exon	missense	✓	-	832C>T	Pro278Ser	832	0.81	-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