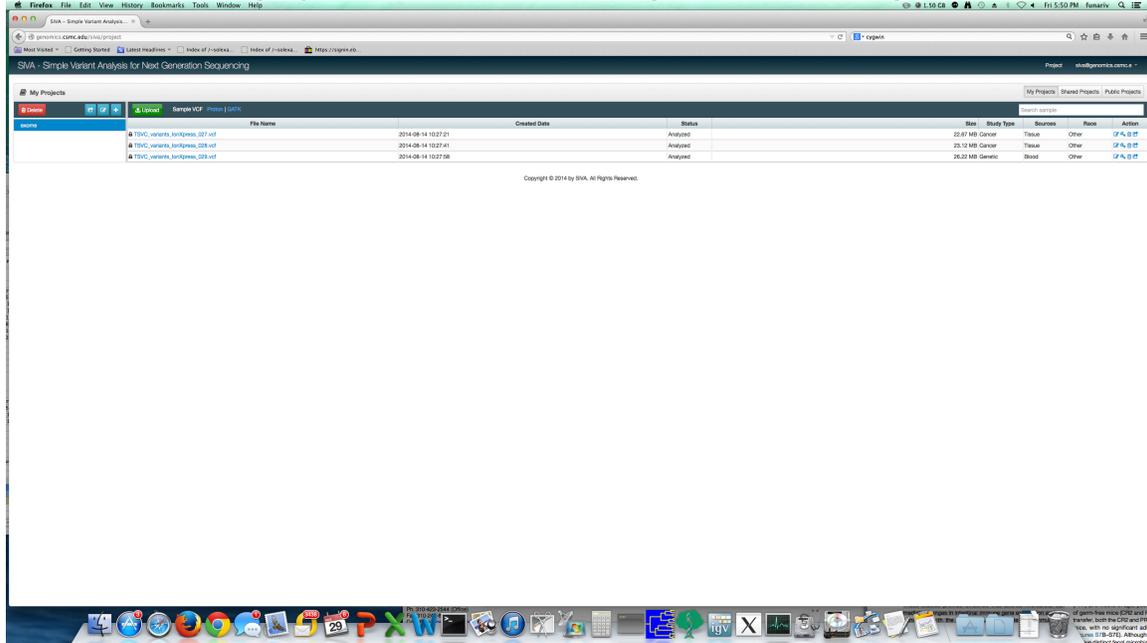


Below is an example of how one might use SiVA to interrogate exome or targeted gene panels. This is meant as an illustrated guide not a detailed manual on exome analysis, approaches or strategies. This guide assumes a basic level of genetic and genomic knowledge.

Log on to SiVA with a popular web browser (e.g.firefox) at genomics.csmc.edu/siva/
 After successful login you should find something like the following screenshot:



Upload and edit sample details:

To upload a sample, click the green **“Upload”** button to upload a FASTA file or .vcf file (make sure to look at the Sample VCF for an acceptable .vcf file format).



After you have uploaded your file. You will see (from left to right)

- 1) **“My projects”** directory (all projects registered to the user name),
- 2) **“File names”** (Files in the selected project directory),
- 3) **“Created date”** (Date file was uploaded
- 4), **“Status”** (if file upload has completed will read “Analyzed”);
- 5) **“Size”** (sizes of files uploaded),
- 6) **“Study Type”** (user determined at upload either Cancer or Genetic analysis”)
- 7) **“Sources”** (user determined at upload either Tissue or Blood)
- (8) **“Race”** (user determined at upload)

9) **“Action”** (user selectable options: **“Edit”**, **“Make public”**, **“Delete”**, **“Share”** (with another user))

Select **“Edit”** to edit file name details.

name	File Name	Created Date	Status	Size	Study Type	Sources	Race	Action
	T5VC_variants_hc10988_021.vcf	2014-08-14 10:27:21	Analyzed	22.07 MB	Cancer	Tissue	Other	[Edit] [Share] [Delete]
	T5VC_variants_hc10988_020.vcf	2014-08-14 10:27:41	Analyzed	22.13 MB	Cancer	Tissue	Other	[Edit] [Share] [Delete]
	T5VC_variants_hc10988_020.vcf	2014-08-14 10:27:58	Analyzed	26.22 MB	Genetic	Blood	Other	[Edit] [Share] [Delete]

Select **“Make public”** to make file publically viewable

name	File Name	Created Date	Status	Size	Study Type	Sources	Race	Action
	T5VC_variants_hc10988_021.vcf	2014-08-14 10:27:21	Analyzed	22.07 MB	Cancer	Tissue	Other	[Edit] [Share] [Delete] [Make public]
	T5VC_variants_hc10988_020.vcf	2014-08-14 10:27:41	Analyzed	22.13 MB	Cancer	Tissue	Other	[Edit] [Share] [Delete]
	T5VC_variants_hc10988_020.vcf	2014-08-14 10:27:58	Analyzed	26.22 MB	Genetic	Blood	Other	[Edit] [Share] [Delete]

Select **“Delete”** to delete file

name	File Name	Created Date	Status	Size	Study Type	Sources	Race	Action
	T5VC_variants_hc10988_021.vcf	2014-08-14 10:27:21	Analyzed	22.07 MB	Cancer	Tissue	Other	[Edit] [Share] [Delete]
	T5VC_variants_hc10988_020.vcf	2014-08-14 10:27:41	Analyzed	22.13 MB	Cancer	Tissue	Other	[Edit] [Share] [Delete]
	T5VC_variants_hc10988_020.vcf	2014-08-14 10:27:58	Analyzed	26.22 MB	Genetic	Blood	Other	[Edit] [Share] [Delete]

Select **“Share”** to share with a colleague or Principle investigator who is already registered. When you share a variant results sample, please enter in the email or registered user, if the user is registered the user name will autopopulate.

name	File Name	Created Date	Status	Size	Study Type	Sources	Race	Action
	T5VC_variants_hc10988_021.vcf	2014-08-14 10:27:21	Analyzed	22.07 MB	Cancer	Tissue	Other	[Edit] [Share] [Delete]
	T5VC_variants_hc10988_020.vcf	2014-08-14 10:27:41	Analyzed	22.13 MB	Cancer	Tissue	Other	[Edit] [Share] [Delete]
	T5VC_variants_hc10988_020.vcf	2014-08-14 10:27:58	Analyzed	26.22 MB	Genetic	Blood	Other	[Edit] [Share] [Delete]

Search Variants

Click on one sample name to view results and to search variants.



After clicking on one sample you can view in graphical detail a summary of variants



You can mouse over each section to get more information of the variants.



You can search for a Comment in the **Search** window above the Distribution of variants pie graph this is helpful if you are a principle investigator reviewing comments from a group member maybe who reviewed the variants previously.



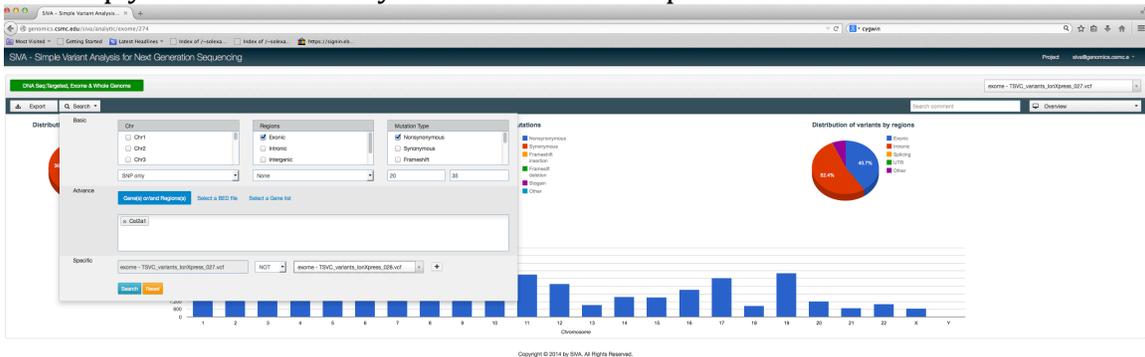
Once you upload the .VCF file and its annotated you can click **Export** on the far left (next to the **Search** key). Or you can search it and export only the searched values.



You can then search or filter all the variants in the sample (e.g. 50K) to something that is more representative of the candidate mutation list (e.g. 1-50 SNPs).

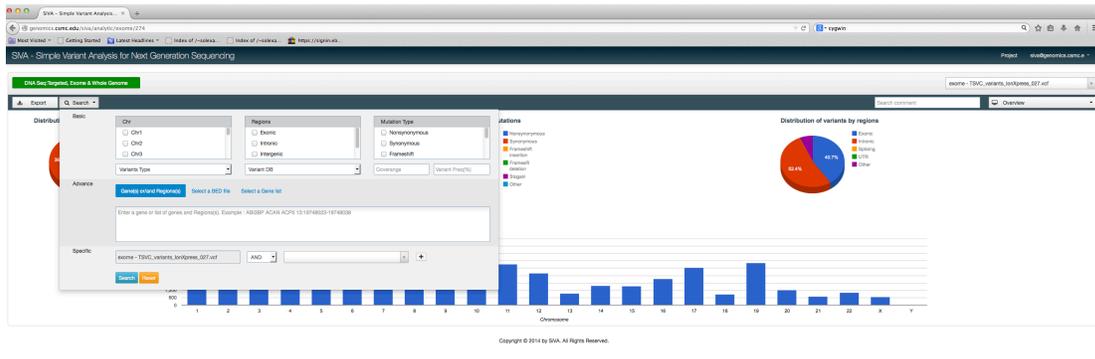


After you click search, a Search tab opens (see below) that includes many fields that will help you narrow down your variants to true positives or candidate mutations.



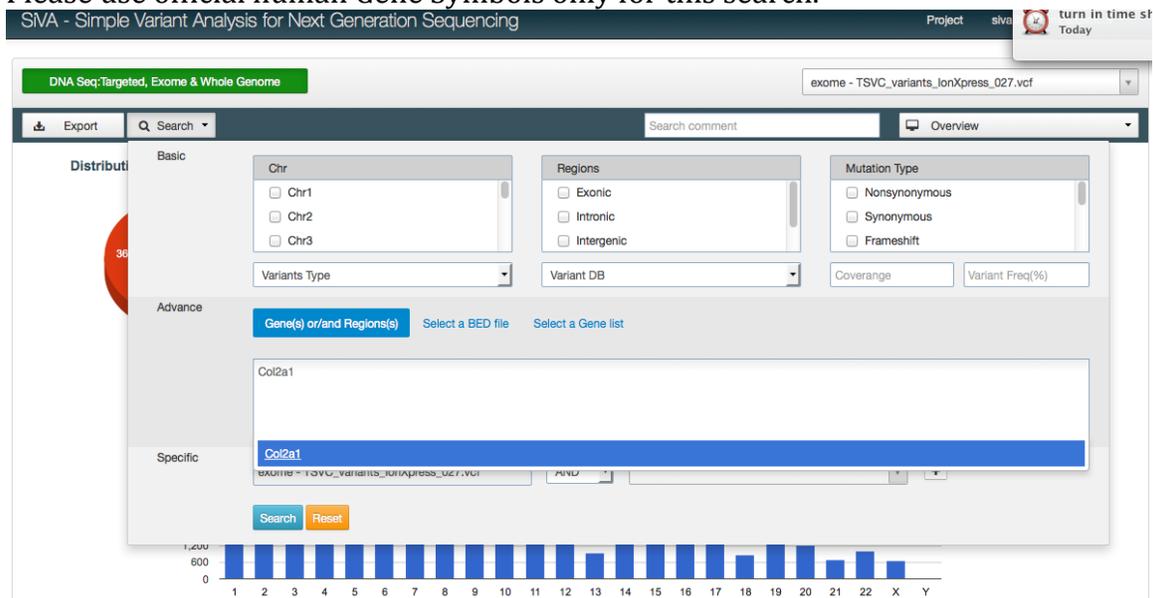
Basic Search

In the search Table, select or enter reasonable filters that will result in a query that selects for only candidate mutations. For Example, in the case below we are searching a sample for a mutation that is inherited and not found in a sibling. Specifically, in the search tab we have selected to narrow our search to Exons (by checking “Exons” in “Regions” box), only Non-Synonymous variants (by checking “non-synonymous” in “Mutation type” box). Then narrowed the candidates by only looking at rare SNPs, by pulling down the “variants type” and “variant db” so that we selected “SNPs” (not indels) and “none” (for not in dbSNP). Then we enriched or selected for the heterozygous variants that represented at least 35 percent of the reads with at least a total coverage of 20.



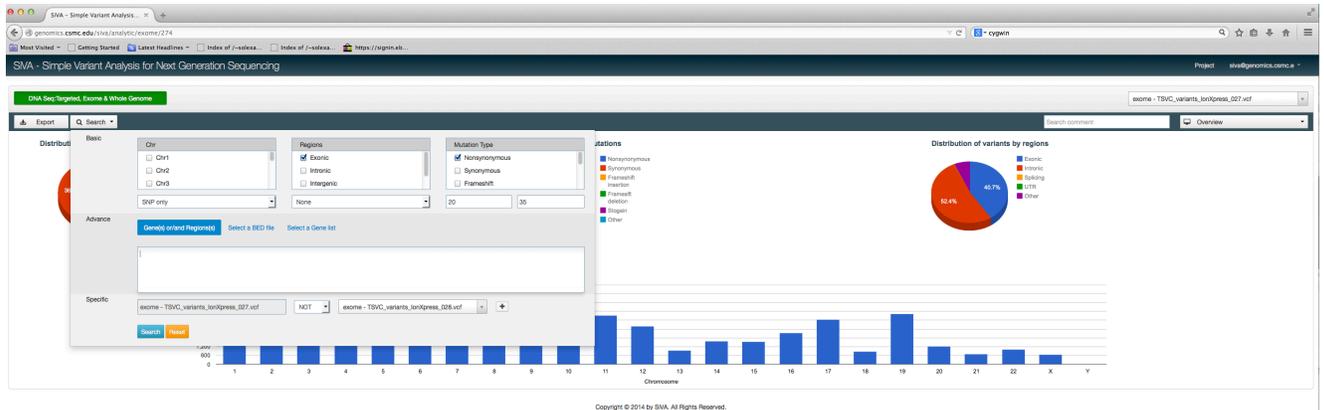
Advanced SNP filtering/Analysis

Under the “Basic” SNP Filtering section is the “Advanced” SNP Search Window. In this window you can construct a list of genes that are routinely searched and upload the list to your account for future searches. This is helpful for routine searches. If you have 1 or 2 genes you just want to check then you can type the names of the genes in the window (As seen below in the screenshot where “Col2a1 was entered). Please use official human Gene Symbols only for this search.



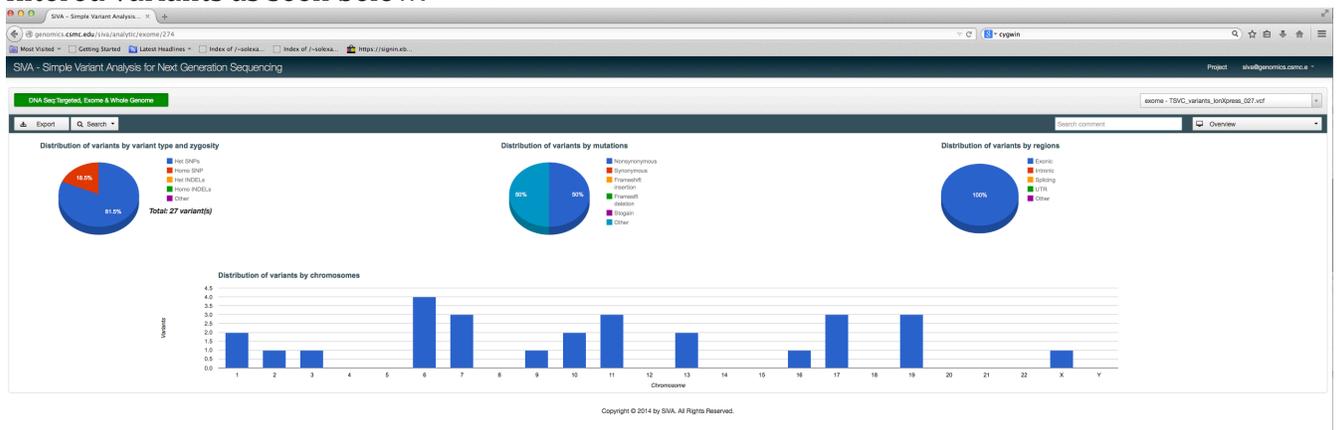
Variants “Specific” to one sample

Under the Advanced SNP filtering there is a “**Specific**” filtering window to select for Variants that are uniquely found in one sample and/or not another sample. The Boolean query can be added sequentially with as many samples that are present in the users account. In this case, we selected not found in another sample to find a mutation that was unique and rare and not found in another sequencing sample. This can be very useful for separating out rare mutations from batch sequencing errors that may be in unrelated samples, as much as it is used to identify “**denovo**” mutations found in an unaffected parent for example.



Dynamic Graphical user interfaces rapidly shows Summary of Filtering

After filtering the graphical interface is updated with the information from the filtered variants as seen below.



Variant Annotation Details to further identify true candidate mutations.
 After viewing the graphic summary, one can research the variant annotation in more detail by selecting the tab on the far right labeled **“Overview”**. When the tab is pulled down there reveals multiple annotation tabs.

A. Variant Detail tab.

If you select Variant Detail you will see the following information

Gene	Chr	Start	Ref	Obs	Region	Mutation	Coverage	VQ(%)	Zygosity	altBP	VD
CPDYL1	chr6	20017712	A	G	exonic	non synonymous SNV	32	37	Het		2
PVRG	chr7	8981781	G	G	exonic	non synonymous SNV	79	91	Het		2
TNFR22B	chr19	1010629	G	A	exonic	non synonymous SNV	36	97	Homo		2
MFRP	chr11	11813386	A	G	exonic	non synonymous SNV	20	35	Het		1
ENAH1	chr3	5202082	T	G	exonic	non synonymous SNV	22	40	Het		1
SNARE2	chr17	7807382	C	G	exonic	non synonymous SNV	38	64	Het		2
NFK2	chr7	8624170	A	G	exonic	non synonymous SNV	32	41	Het		1
SLC25A4B	chr11	6014482	C	G	exonic	non synonymous SNV	37	52	Het		2
TMEM178	chr2	9102391	T	A	exonic	non synonymous SNV	34	62	Het		2
DNCE	chr17	7129845	T	G	exonic	non synonymous SNV	61	100	Homo		2
ADCY1	chr7	4570503	G	C	exonic	non synonymous SNV	38	58	Het		2
ARIS	chr11	3693960	T	C	exonic	non synonymous SNV	70	50	Het		1
CHTFP	chr1	15301768	A	G	exonic	non synonymous SNV	27	100	Homo		2
ZNFVND15	chr17	4648890	G	C	exonic	non synonymous SNV	61	42	Het		2
KCNKG	chr6	3918824	T	G	exonic	non synonymous SNV	67	100	Homo		2

B. The second tab is “Annotation and Databases” and provides rich annotation of the variant frequency and AA changes.

Gene	Chr	Start	Ref	Obs	Trn. ID	CDS mut	AA mut	ESP 6500	100g	Coverage	BedPip
CPDYL1	chr6	20017712	A	G	NA_000070	LAA171G	p.L61A				
PVRG	chr7	8981781	G	C	NA_000070	C520C	p.S66A			365	0.96
TNFR22B	chr19	1010629	G	A	NA_00103328	C1158T	p.A529V				
MFRP	chr11	11813386	A	G	NA_00104453	G1120T	p.C429F			258	
ENAH1	chr3	5202082	T	C	NA_011326	C766G	p.T516L			271	
SNARE2	chr17	7807382	C	G	NA_00144888	C586G	p.P627R			329	
NFK2	chr7	8624170	A	G	NA_002023	C487G	p.T135A			650	
SLC25A4B	chr11	6014482	C	G	NA_00207049	C200C	p.C100R			397	
TMEM178	chr2	9102391	T	A	NA_00122566	C431T	p.E106V				
DNAJ2	chr17	7129845	T	G	NA_004423	C418G	p.G564R			476	
ADCY1	chr7	4570503	G	C	NA_001118	G2008G	p.C1068S				
ARIS	chr11	3693960	T	C	NA_000017	L484G	p.K289R				
CHTFP	chr1	15301768	A	G	NA_00105612	C453G	p.M185V			545	
ZNFVND15	chr17	4648890	G	C	NA_00126608	G4120T	p.S413R			827	
KCNKG	chr6	3918824	T	G	NA_000140	G4134G	p.S46W				

C. The Next tab Functional Prediction includes the results from several prediction algorithms on the effect of the variant on the mutation.

Gene	Chr	Start	Ref	Obs	LRS_SPT	LRS_PolyPhen2	LRS_LTR	Multiplier	LRS_PhyP	GERP++
CPDYL1	chr6	20017712	A	G						
PVRG	chr7	8981781	G	C	Damaging	Probably damaging	NA	N	C	2.22
TNFR22B	chr19	1010629	G	A	Tolerated	Probably damaging	Unknown	N	C	1.89
MFRP	chr11	11813386	A	G	Damaging	Probably damaging	Neutral	NA	C	3.69
ENAH1	chr3	5202082	T	C	Damaging	Probably damaging	Detrimental	N	C	4.27
SNARE2	chr17	7807382	C	G	Damaging	Probably damaging	Neutral	Disease_causin	C	2.88
NFK2	chr7	8624170	A	G	Tolerated	Berign	Neutral	Disease_causin	N	2.84
SLC25A4B	chr11	6014482	C	G	Tolerated	Berign	Neutral	N	N	2.06
TMEM178	chr2	9102391	T	A						
DNAJ2	chr17	7129845	T	G	Tolerated	NA	Detrimental	Disease_causin	N	-4.11
ADCY1	chr7	4570503	G	C	NA	NA	Neutral	N	C	3.49
ARIS	chr11	3693960	T	C	Tolerated	Berign	Detrimental	N	N	-0.76
CHTFP	chr1	15301768	A	G	Tolerated	Berign	Detrimental	N	C	4.42
ZNFVND15	chr17	4648890	G	C	Tolerated	Probably damaging	Detrimental	N	C	5.47
KCNKG	chr6	3918824	T	G	Tolerated	Berign	Neutral	N	N	-0.38

Drilling down into Gene Biology and Gene annotation details._____

After viewing the variants list, one can right click on the hypertext gene symbol to learn more information about the gene available in different databases like OMIM, ClinVar, etc. As seen below in the pop up menu in the screen shot.

Gene	Chr	Start	Ref	Obs	LIB_SFT	LIB_PolyPhen	LIB_LTR	McRater	LIB_PhyP	GERP++
SRSF11	chr6	5817712	A	G	Tolerated	Probably damaging	NA	N	C	2.22
OMIM - Online Mendelian Inheritance in Man		60817821	G	C	Damaging	Probably damaging	NA	N	C	2.22
ClinVar		1101295	G	A	Tolerated	Probably damaging	Unknown	N	C	1.86
HGNC - HGNC Gene Nomenclature Committee		118126860	A	G	Damaging	Probably damaging	Neutral	NA	C	3.95
HGDP - Human Gene and Protein Database		58155482	T	C	Damaging	Probably damaging	Deleterious	N	C	4.27
dbSNP - dbSNP Cancer Genomics Portal (Integrated with TCGA)		79577522	C	G	Damaging	Probably damaging	Neutral	Disease causing	C	2.93
SLC5A45	chr11	8581719	A	G	Tolerated	Benign	Neutral	N	N	2.94
SLC5A45	chr11	8514482	C	G	Tolerated	Benign	Neutral	N	N	2.95
TM7SF188	chr2	9787381	T	A	Tolerated	NA	Deleterious	Disease causing	N	-4.11
HAL2	chr17	7129640	T	G	Tolerated	NA	Neutral	N	C	3.49
ADCY1	chr7	4573353	G	C	NA	NA	Neutral	N	C	3.49
ARIS	chr11	368990	T	C	Tolerated	Benign	Neutral	N	N	-0.76
CHOP	chr1	15381768	A	G	Tolerated	Benign	Deleterious	N	C	4.43
ZNF525	chr17	466950	G	C	Tolerated	Probably damaging	Deleterious	N	C	5.47
KCNK9	chr6	3918824	T	G	Tolerated	Benign	Neutral	N	N	-0.38

Finally, one can drill down into more variant detail by looking at the position of the variant in respect to other genomic data in UCSC genome browser. By left clicking on the hypertext position of the variant (as seen below in the screen shot) one can open a new window that opens at the variant position in the human genome in the UCSC genome browser.

Gene	Chr	Start	Ref	Obs	LIB_SFT	LIB_PolyPhen	LIB_LTR	McRater	LIB_PhyP	GERP++
SRSF11	chr6	5817712	A	G	Tolerated	Probably damaging	NA	N	C	2.22
JNYB	chr7	89119	G	C	Damaging	Probably damaging	NA	N	C	1.93
TRAF3IP3	chr19	31208	A	Tolerated	Probably damaging	Unknown	N	C	1.93	
MFRP	chr11	118123	G	C	Damaging	Probably damaging	Deleterious	NA	C	3.85
STAB1	chr3	59102	G	C	Damaging	Probably damaging	Deleterious	N	C	4.27
SH3BP2	chr17	79175	G	C	Damaging	Probably damaging	Neutral	Disease causing	C	2.93
MFSD2	chr7	85147	G	Tolerated	Benign	Neutral	Disease causing	N	N	2.94
SLC5A45	chr11	85146	G	Tolerated	Benign	Neutral	N	N	2.95	
TM7SF188	chr2	9787381	T	A	Tolerated	NA	Deleterious	Disease causing	N	-4.11
HAL2	chr17	7129640	T	G	Tolerated	NA	Neutral	N	C	3.49
ADCY1	chr7	4573353	G	C	NA	NA	Neutral	N	C	3.49
ARIS	chr11	368990	T	C	Tolerated	Benign	Neutral	N	N	-0.76
CHOP	chr1	15381768	A	G	Tolerated	Benign	Deleterious	N	C	4.43
ZNF525	chr17	466950	G	C	Tolerated	Probably damaging	Deleterious	N	C	5.47
KCNK9	chr6	3918824	T	G	Tolerated	Benign	Neutral	N	N	-0.38

This will open a new window allowing the user to overlay many other types of genomic information, including snp variants, expression, conservation, position in the gene etc. to help the user understand the true effect of the mutation/variant in context to the surrounding genome and other genomic information (See screen shot below).

