1390 Shorebird Way
Mountain View, CA 94043
www.23andme.com

Exome Results & Raw Data Summary

Generated on: June 20, 2012

Congratulations! Your exome has been sequenced and your data is ready for you to download. We have also included this overview of your data to get you started on your exome exploration. Here are a few important points about your exome data:

- Two types of files are available for download: 1) the aligned sequencing reads in BAM format, 2) a file containing variant calls (VCF file).
- The raw data VCF file is a preliminary draft of your exome. Our ability to call variants, especially indels, is greatly improved with each additional exome added to our database. Moreover we will build upon this protocol to include additional steps such as custom treatment of the sex chromosomes. To this end we will update your VCF file at the end of the pilot. We will contact you when this data is available.

Your exome at a glance:

The Exome Service is a pilot project, and this report contains preliminary data only. 23andMe does not represent that all of this information is accurate. **In this report we have used 1000 Genome Project data to report frequencies of variants to determine how common or rare a particular variant is.** We have also only provided information about a subset of the many gene-disrupting variants present in the human genome, in a chosen set of genes. Sequencing was performed such that the total number of bases read was at least 80X the size of the exome. As described in the Exome Terms of Use, 23andMe will not be providing the reports and explanations that 23andMe typically provides to customers with respect to their genotyping results for this data. 23andMe Services are for research, informational, and educational use only. We do not provide medical advice. Please keep in mind that genetic information you share with others could be used against your interests.

Your exome in numbers

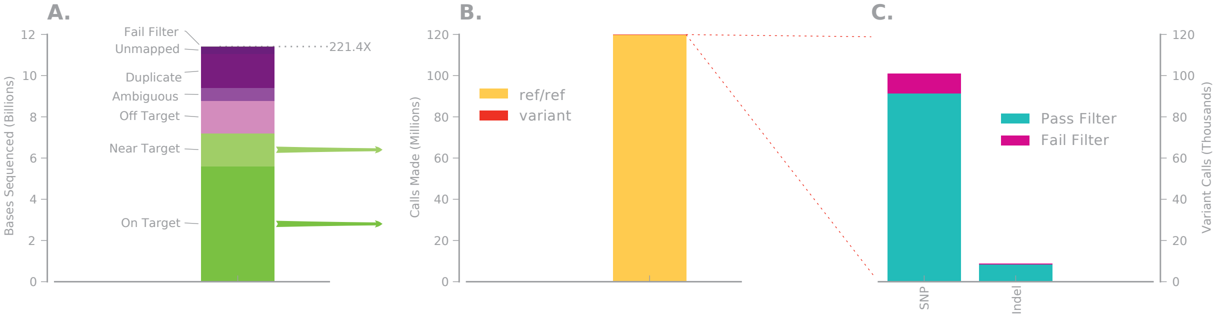


Figure 1: Getting from raw reads to called variants. A) The number of bases obtained by sequencing your exome. The top line indicates total coverage. B) Total number of called bases in your exome. The vast majority are the same as the reference genome. C) An expansion of the small sliver of variants depicted in B. These are the variants present in your VCF file.

Welcome to your exome. Your exome is the 50 million DNA bases of your genome containing the information necessary to encode all your proteins. Your exome data consists of two parts, the raw data (both aligned and unaligned Illumina reads, fig1A) and a draft of the variants present in your exome (fig1C). While this draft is provisional and we will be improving upon it, we wanted to allow you to dig in to your exome as soon as possible so you can tell us what you think is important and should be included.

To create the first draft of your exome we implemented the Broad Institute's "Best Practice" protocol for exome sequencing analysis. You can read a detailed description of it here (for brief summary see Appendix).

Characterizing your variants

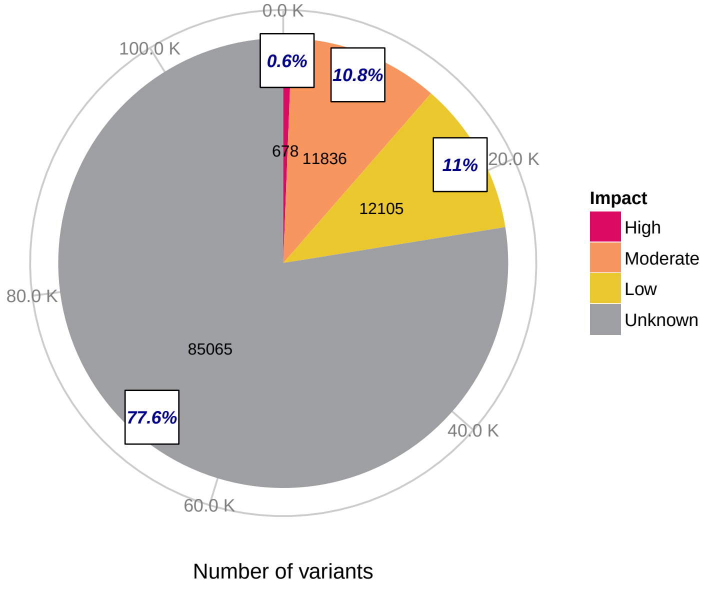


Figure 2: Predicting impact of variants on gene function. An overview of your variants and their predicted impact on gene function.

The variants in your VCF file are the positions in your genome that differ from the reference genome. Most of these variants are likely to be functionally neutral and unlikely to cause any severe disorders. Pinpointing genuine disease mutations is still challenging and we used a number of software tools to identify those that may be functionally important. We estimated the impact a variant has on gene function based on the severity of its effect on the gene product:

High impact:

Frame shift Insertion or deletion of bases, not multiple of 3.

Splice site Variant at the 'splicing site' may disrupt the consensus splicing site sequence.

Stop gain Premature termination of peptides, which would disable protein function.

Start loss Loss of the start codon.

Stop loss Loss of the stop codon.

Moderate impact:

Nonsynonymous substitution Non-conservative change altering an amino acid in a protein.

Codon insertion or deletion Insertion or deletion of bases, multiple of 3.

Low impact:

Synonymous substitution Variant that does not alter the amino acid sequence due to codon degeneracy.

Start gain Variant resulting in the gain of a start codon.

Synonymous stop Variant changing one stop codon into another.

Unknown impact: Variants unlikely to affect gene products.

How rare are your variants?

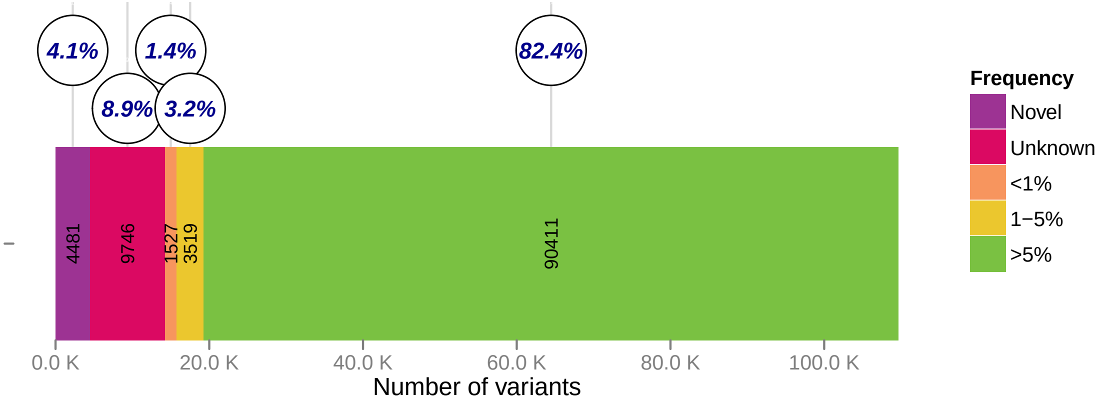


Figure 3: Variant frequencies. The allele frequencies of the variants in your exome. Unknown: allele is present in a public database but no frequency data was available.

One of the advantages of exome sequencing is that we can detect sequence variants that are unique to you! By comparing your variants to all those that have been discovered so far, we can divide your variants into the following categories:

- **novel** variant hasn't been observed in current public sequence databases
- **unknown** variant has been observed in public databases but allelic frequency has not been calculated and therefore is not available
- **rare** variant with allelic frequency <1%
- **somewhat rare** variant with frequency 1-5%
- **common** frequency of the variant is greater than 5%

One of the most comprehensive human variation public datasets is maintained by the 1000 Genomes Project. We use 1000 Genomes Project data (project release: 08-26-2011) to report frequencies of alleles found in your exome, including reporting if it is absent from the public database (*i.e.* a novel variant).

Filtering your variants

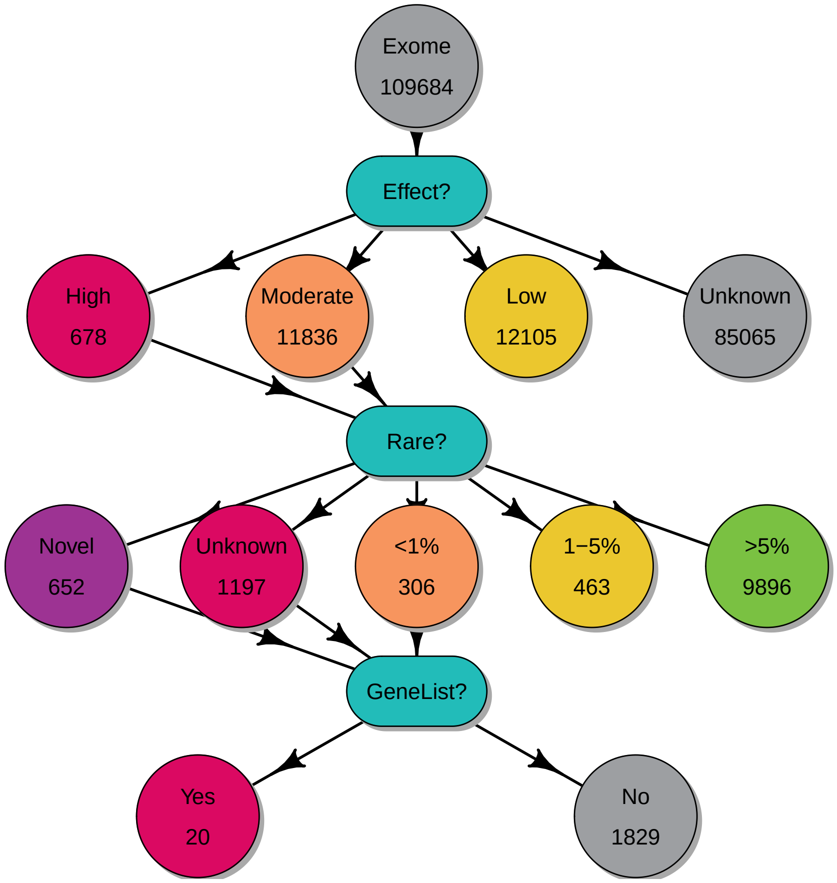


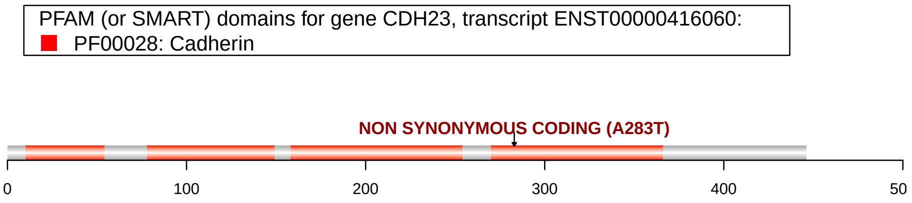
Figure 4: Variant filtering decision tree. A graphical representation of the filtering process that was used to generate your short list of variants of interest.

Most sequence variants in your exome are likely to be neutral and do not cause any severe disorders. A filtering process is often undertaken to prioritize variants discovered through sequencing. To identify potentially interesting and relevant variants with potential functional effects (contributing to disease and other phenotypes of interest) we used three consecutive filters, depicted in the figure above: (1) effect of the variant on the gene product; (2) allele frequency of the variant; (3) location of the variant in one of 592 genes involved in Mendelian disorders (at this point we also exclude indels and variants on the sex chromosomes).

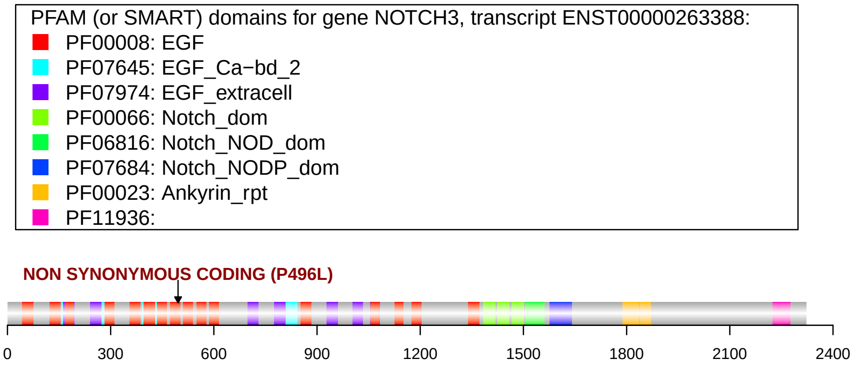
We hope you find this initial list of variants interesting and that it will help you in your journey through your exome. This short list of variants only scratches the surface of what your genome contains and is just the beginning of where your data can take you. Have fun!

List of selected variants

Variant 1:	**Gene:** CDH23 **Your genotype:** G/A	**Location:** chr10:73377112
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00420	**dbSNP:** rs143282422
Quality:	**Genotype quality:** 99	**Coverage depth:** 222
Details:	**Gene description:** cadherin-related 23 **Transcript:** ENST00000416060 **EntrezId:** 64072 **UniProt:** Q9H251	 **AA change:** A283T **EnsemblId:** ENSG00000107736 **OMIM:** 605516



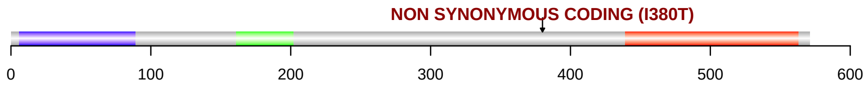
Variant 2:	**Gene:** NOTCH3 **Your genotype:** A/A	**Location:** chr19:15299051
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00610	**dbSNP:** rs11670799
Quality:	**Genotype quality:** 36.11	**Coverage depth:** 16
Details:	**Gene description:** notch 3 **Transcript:** ENST00000263388 **EntrezId:** 4854 **UniProt:** Q9UM47	 **AA change:** P496L **EnsemblId:** ENSG00000074181 **OMIM:** 600276



Variant 3:	**Gene:** MEFV **Your genotype:** A/G **Location:** chr16:3293880	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00730	**dbSNP:** rs11466045
Quality:	**Genotype quality:** 99	**Coverage depth:** 51
Details:	**Gene description:** Mediterranean fever	
	Transcript: ENST00000536379	**AA change:** I380T
	EntrezId: 4210	**EnsemblId:** ENSG00000103313
	UniProt: O15553	**OMIM:** 608107

PFAM (or SMART) domains for gene MEFV, transcript ENST00000536379:

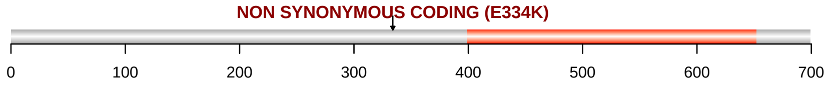
- 🟦 PF02758: Pyrin
- 🟩 PF00643: Znf_B−box
- 🟥 PF00622: SPRY_rcpt



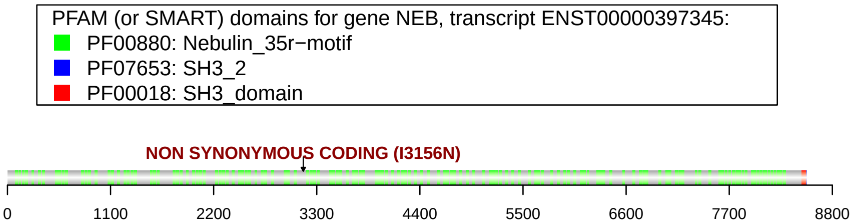
Variant 4:	**Gene:** GLE1 **Your genotype:** G/A **Location:** chr9:131287573	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00460	**dbSNP:** rs138310419
Quality:	**Genotype quality:** 99	**Coverage depth:** 59
Details:	**Gene description:** GLE1 RNA export mediator homolog (yeast)	
	Transcript: ENST00000309971	**AA change:** E334K
	EntrezId: 2733	**EnsemblId:** ENSG00000119392
	UniProt: Q53GS7	**OMIM:** 603371

PFAM (or SMART) domains for gene GLE1, transcript ENST00000309971:

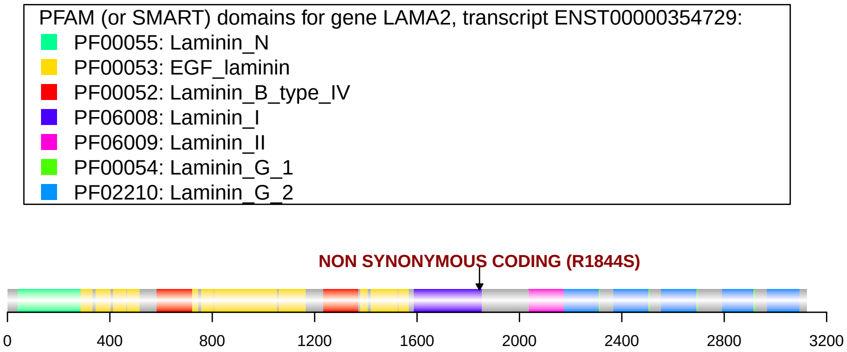
- 🟥 PF07817: GLE1



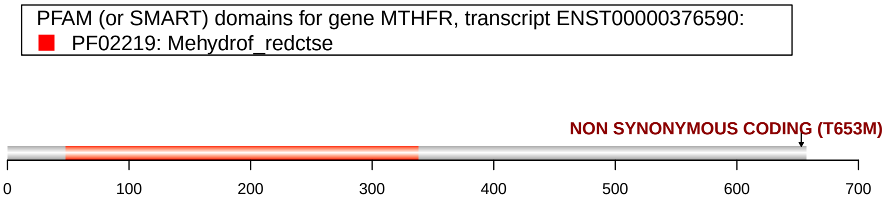
Variant 5:	**Gene:** NEB **Your genotype:** A/T **Location:** chr2:152487808	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00530	**dbSNP:** rs145770770
Quality:	**Genotype quality:** 99	**Coverage depth:** 227
Details:	**Gene description:** nebulin **Transcript:** ENST00000397345 **EntrezId:** 4703 **UniProt:** P20929	 **AA change:** I3156N **EnsemblId:** ENSG00000183091 **OMIM:** 161650



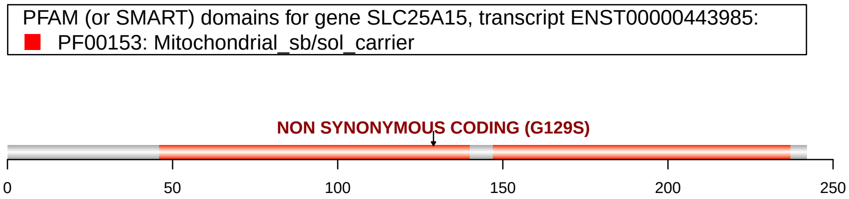
Variant 6:	**Gene:** LAMA2 **Your genotype:** C/A **Location:** chr6:129722453	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00640	**dbSNP:** rs56173620
Quality:	**Genotype quality:** 99	**Coverage depth:** 100
Details:	**Gene description:** laminin, alpha 2 **Transcript:** ENST00000354729 **EntrezId:** 3908 **UniProt:** P24043	 **AA change:** R1844S **EnsemblId:** ENSG00000196569 **OMIM:** 156225



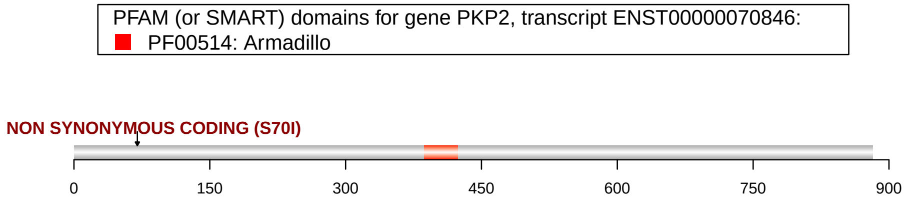
Variant 7:	**Gene:** MTHFR **Your genotype:** G/A	**Location:** chr1:11850750
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00730	**dbSNP:** rs35737219
Quality:	**Genotype quality:** 99	**Coverage depth:** 127
Details:	**Gene description:** methylenetetrahydrofolate reductase (NAD(P)H) **Transcript:** ENST00000376590 **EntrezId:** 4524 **UniProt:** P42898	 **AA change:** T653M **EnsemblId:** ENSG00000177000 **OMIM:** 607093


PFAM (or SMART) domains for gene MTHFR, transcript ENST00000376590:
🟥 PF02219: Mehydrof_redctse

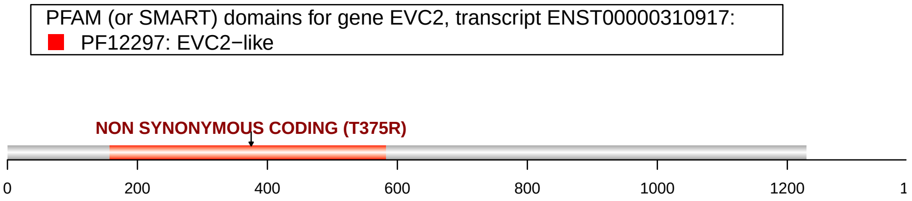
Variant 8:	**Gene:** SLC25A15 **Your genotype:** G/A	**Location:** chr13:41381542
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 5e-04	**dbSNP:** rs151239794
Quality:	**Genotype quality:** 99	**Coverage depth:** 174
Details:	**Gene description:** solute carrier family 25 (mitochondrial carrier; ornithine transporter) member 15 **Transcript:** ENST00000443985 **EntrezId:** 10166 **UniProt:** Q9Y619	 **AA change:** G129S **EnsemblId:** ENSG00000102743 **OMIM:** 603861


PFAM (or SMART) domains for gene SLC25A15, transcript ENST00000443985:
🟥 PF00153: Mitochondrial_sb/sol_carrier

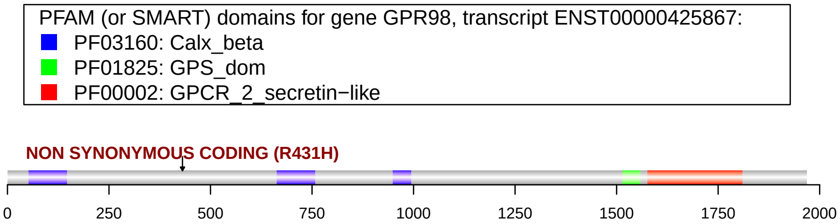
Variant 9:	**Gene:** PKP2 **Your genotype:** C/A **Location:** chr12:33049457	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00970	**dbSNP:** rs75909145
Quality:	**Genotype quality:** 99	**Coverage depth:** 52
Details:	**Gene description:** plakophilin 2 **Transcript:** ENST00000070846 **EntrezId:** 5318 **UniProt:** Q99959	 **AA change:** S70I **EnsemblId:** ENSG00000057294 **OMIM:** 602861


PFAM (or SMART) domains for gene PKP2, transcript ENST00000070846:
🟥 PF00514: Armadillo
NON SYNONYMOUS CODING (S70I)
0
150
300
450
600
750
900

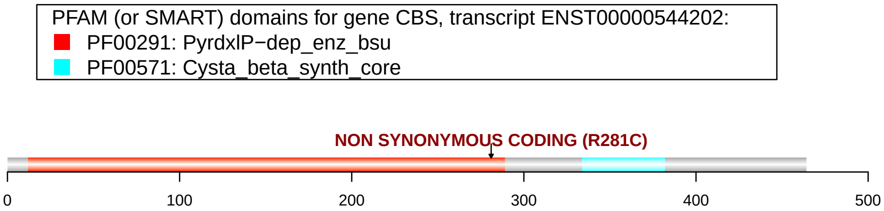
Variant 10:	**Gene:** EVC2 **Your genotype:** G/C **Location:** chr4:5642347	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00320	**dbSNP:** rs141287105
Quality:	**Genotype quality:** 99	**Coverage depth:** 250
Details:	**Gene description:** Ellis van Creveld syndrome 2 **Transcript:** ENST00000310917 **EntrezId:** 132884 **UniProt:** Q86UK5	 **AA change:** T375R **EnsemblId:** ENSG00000173040 **OMIM:** 607261


PFAM (or SMART) domains for gene EVC2, transcript ENST00000310917:
🟥 PF12297: EVC2−like
NON SYNONYMOUS CODING (T375R)
0
200
400
600
800
1000
1200
1

Variant 11:	**Gene:** GPR98 **Your genotype:** G/A **Location:** chr5:90086955	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00280	**dbSNP:** rs41304892
Quality:	**Genotype quality:** 99	**Coverage depth:** 196
Details:	**Gene description:** G protein-coupled receptor 98	
	Transcript: ENST00000425867	**AA change:** R431H
	EntrezId: 84059	**EnsemblId:** ENSG00000164199
	UniProt: Q8WXG9	**OMIM:** 602851

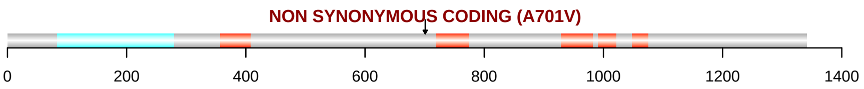


Variant 12:	**Gene:** CBS **Your genotype:** G/A **Location:** chr21:44480591	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00100	**dbSNP:** rs117687681
Quality:	**Genotype quality:** 99	**Coverage depth:** 211
Details:	**Gene description:** cystathionine-beta-synthase	
	Transcript: ENST00000544202	**AA change:** R281C
	EntrezId: 875	**EnsemblId:** ENSG00000160200
	UniProt: P35520	**OMIM:** 613381



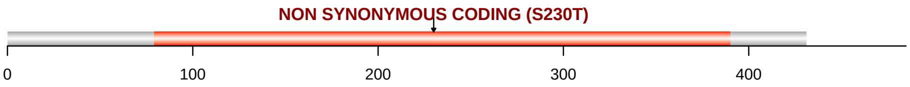
Variant 13:	**Gene:** ADAMTS13 **Your genotype:** C/T	**Location:** chr9:136307825
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00960	**dbSNP:** rs41314453
Quality:	**Genotype quality:** 99	**Coverage depth:** 121
Details:	**Gene description:** ADAM metallopeptidase with thrombospondin type 1 motif, 13	
	Transcript: ENST00000356589	**AA change:** A701V
	EntrezId: 11093	**EnsemblId:** ENSG00000160323
	UniProt: Q76LX8	**OMIM:** 604134

PFAM (or SMART) domains for gene ADAMTS13, transcript ENST00000356589:
🟦 PF01421: Peptidase_M12B
🟥 PF00090: Thrombospondin_1_rpt



Variant 14:	**Gene:** KRT18 **Your genotype:** G/C	**Location:** chr12:53345296
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.0028;0.0028	**dbSNP:** rs58472472, rs140469050
Quality:	**Genotype quality:** 99	**Coverage depth:** 131
Details:	**Gene description:** keratin 18	
	Transcript: ENST00000388835	**AA change:** S230T
	EntrezId: 3875	**EnsemblId:** ENSG00000111057
	UniProt: P05783	**OMIM:** 148070

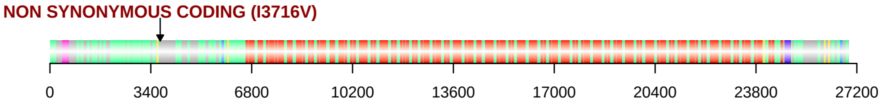
PFAM (or SMART) domains for gene KRT18, transcript ENST00000388835:
🟥 PF00038: FALSE



Variant 15:	**Gene:** TTN **Your genotype:** T/C **Location:** chr2:179605725	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00930	**dbSNP:** rs34070843
Quality:	**Genotype quality:** 99	**Coverage depth:** 145
Details:	**Gene description:** titin	
	Transcript: ENST00000356127	**AA change:** I3716V
	EntrezId: 7273	**EnsemblId:** ENSG00000155657
	UniProt: Q8WZ42	**OMIM:** 188840

PFAM (or SMART) domains for gene TTN, transcript ENST00000356127:

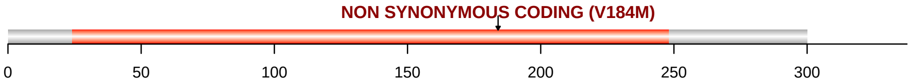
- 🟩 PF07679: Ig_I−set
- 🟪 PF09042: Titin_Z
- 🟨 PF00047: Immunoglobulin
- 🟦 PF07686: Ig_V−set
- 🟥 PF00041: FN_III
- 🟩 PF00069: Se/Thr_kinase−like_dom
- 🟪 PF07714: Ser−Thr/Tyr_kinase



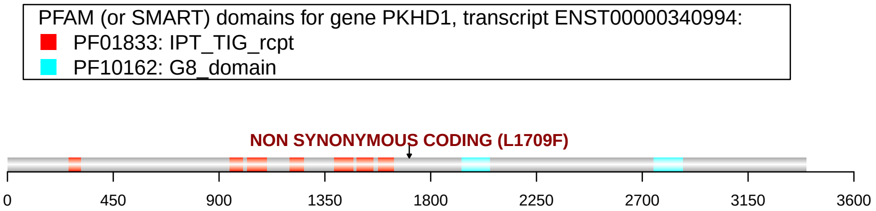
Variant 16:	**Gene:** SGSH **Your genotype:** C/T **Location:** chr17:78184601	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00780	**dbSNP:** rs62620232
Quality:	**Genotype quality:** 99	**Coverage depth:** 204
Details:	**Gene description:** N-sulfoglucosamine sulfohydrolase	
	Transcript: ENST00000534910	**AA change:** V184M
	EntrezId: 6448	**EnsemblId:** ENSG00000181523
	UniProt: P51688	**OMIM:** 605270

PFAM (or SMART) domains for gene SGSH, transcript ENST00000534910:

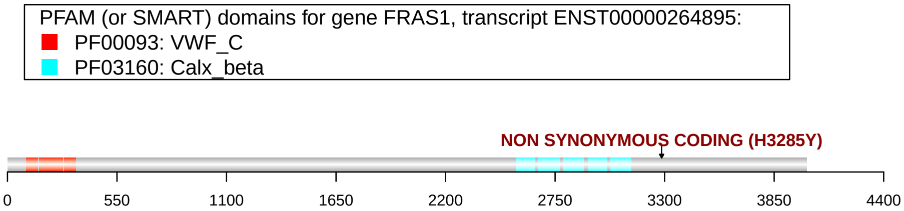
- 🟥 PF00884: Sulfatase



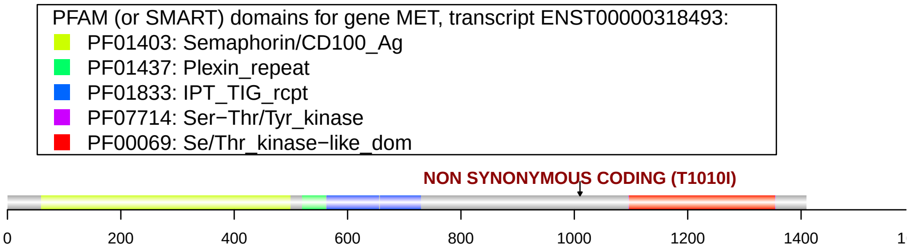
Variant 17:	**Gene:** PKHD1 **Your genotype:** G/A	**Location:** chr6:51889483
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00140	**dbSNP:** rs45517932
Quality:	**Genotype quality:** 99	**Coverage depth:** 205
Details:	**Gene description:** polycystic kidney and hepatic disease 1 (autosomal recessive)	
	Transcript: ENST00000340994	**AA change:** L1709F
	EntrezId: 5314	**EnsemblId:** ENSG00000170927
	UniProt: P08F94	**OMIM:** 606702


PFAM (or SMART) domains for gene PKHD1, transcript ENST00000340994:
🟥 PF01833: IPT_TIG_rcpt
🟦 PF10162: G8_domain
NON SYNONYMOUS CODING (L1709F)
0
450
900
1350
1800
2250
2700
3150
3600

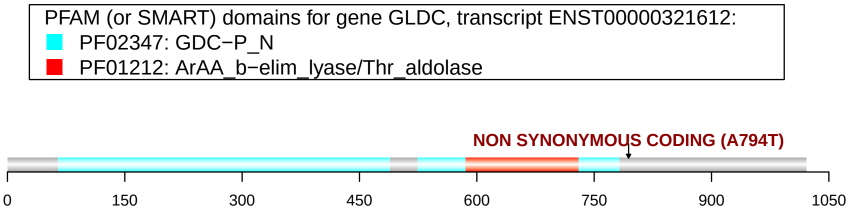
Variant 18:	**Gene:** FRAS1 **Your genotype:** C/T	**Location:** chr4:79432500
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00140	**dbSNP:** NA
Quality:	**Genotype quality:** 99	**Coverage depth:** 242
Details:	**Gene description:** Fraser syndrome 1	
	Transcript: ENST00000264895	**AA change:** H3285Y
	EntrezId: 80144	**EnsemblId:** ENSG00000138759
	UniProt: Q86XX4	**OMIM:** 607830


PFAM (or SMART) domains for gene FRAS1, transcript ENST00000264895:
🟥 PF00093: VWF_C
🟦 PF03160: Calx_beta
NON SYNONYMOUS CODING (H3285Y)
0
550
1100
1650
2200
2750
3300
3850
4400

Variant 19:	**Gene:** MET **Your genotype:** C/T **Location:** chr7:116411990	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00550	**dbSNP:** rs56391007
Quality:	**Genotype quality:** 99	**Coverage depth:** 36
Details:	**Gene description:** met proto-oncogene (hepatocyte growth factor receptor)	
	Transcript: ENST00000318493	**AA change:** T1010I
	EntrezId: 4233	**EnsemblId:** ENSG00000105976
	UniProt: P08581	**OMIM:** 164860


PFAM (or SMART) domains for gene MET, transcript ENST00000318493:
🟨 PF01403: Semaphorin/CD100_Ag
🟩 PF01437: Plexin_repeat
🟦 PF01833: IPT_TIG_rcpt
🟪 PF07714: Ser−Thr/Tyr_kinase
🟥 PF00069: Se/Thr_kinase−like_dom
NON SYNONYMOUS CODING (T1010I)
0
200
400
600
800
1000
1200
1400
1

Variant 20:	**Gene:** GLDC **Your genotype:** C/T **Location:** chr9:6553445	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00380	**dbSNP:** rs141933811
Quality:	**Genotype quality:** 99	**Coverage depth:** 70
Details:	**Gene description:** glycine dehydrogenase (decarboxylating)	
	Transcript: ENST00000321612	**AA change:** A794T
	EntrezId: 2731	**EnsemblId:** ENSG00000178445
	UniProt: P23378	**OMIM:** 238300


PFAM (or SMART) domains for gene GLDC, transcript ENST00000321612:
🟦 PF02347: GDC−P_N
🟥 PF01212: ArAA_b−elim_lyase/Thr_aldolase
NON SYNONYMOUS CODING (A794T)
0
150
300
450
600
750
900
1050

Appendix

To create the first draft of your exome we implemented the Broad Institute's "Best Practice" protocol for exome sequencing analysis. You can read a detailed description of it here, however a brief summary of it follows:

1. We took your raw reads and aligned them against the reference genome (these are the alignments available in the BAM file of the encrypted download).
2. We used these alignments to identify probable contamination (unaligned reads) and artifacts of sample preparation (PCR duplicates) which are then removed from subsequent steps.
3. From this point on we focus on the reads that align either to one of the exons or within the regions 250 bases up and downstream of it.
4. To improve the quality of the alignments we carry out a more accurate alignment of the reads that overlap known indels or are likely to contain indels themselves.
5. We also recalibrate the base quality scores of the reads to bring them in line with the empirically-determined values.
6. Using these realigned+recalibrated reads we generate allele calls at every position with enough high-quality data and filter out those that are homozygous for the allele present in the reference genome (the vast majority of these are at such a high frequency in the population they're unlikely to be interesting). The remaining SNP and indel calls (variants) are the ones available in the VCF file that you downloaded.
7. As yet no sequencing technology is 100% accurate and the highly duplicated nature of the human genome makes variant calling a challenging task. Consequently, a small proportion of the variant calls in your VCF are likely to be incorrect. To reduce this proportion we applied the filters recommended by the Broad Institute to remove technical artifacts. Variants that pass all filters are marked in your VCF file with a PASS. As the exome pilot progresses and we gather more data we will be able to use more advanced techniques identify potential errors and improve the quality of your exome.