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

Exome Results & Raw Data Summary

Generated on: 4/26/2012

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

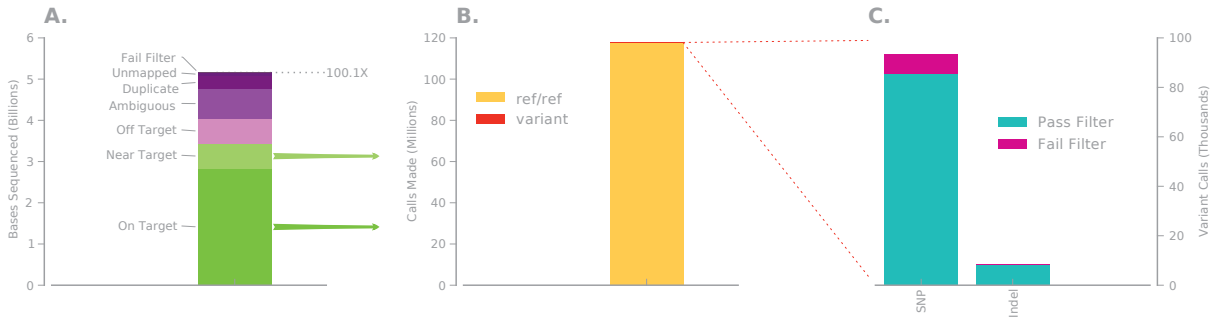


Figure 1: Getting from raw reads to called variants. A) The number of bases obtained by sequencing your exome. The top line indicates total coverage. B) Total number of called bases in your exome. The vast majority are the same as the reference genome. C) An expansion of the small sliver of variants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

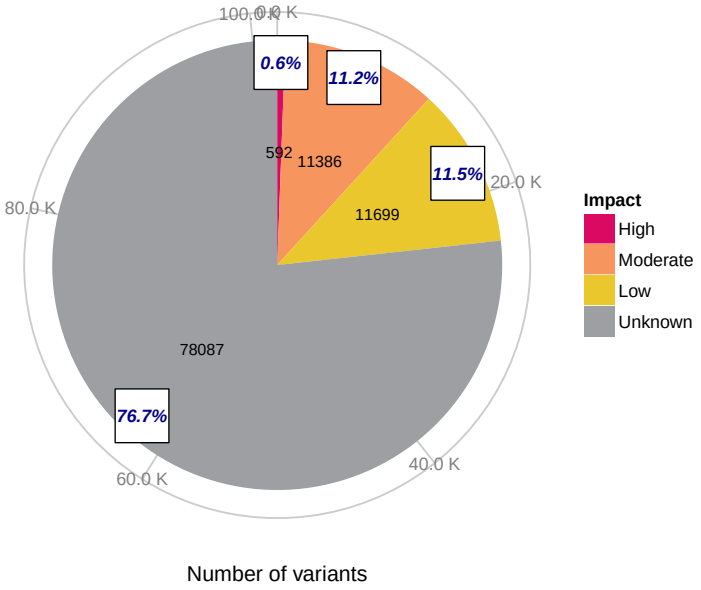


Figure 2: Predicting impact of variants on gene function. An overview of your variants and their predicted impact on gene function.

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?

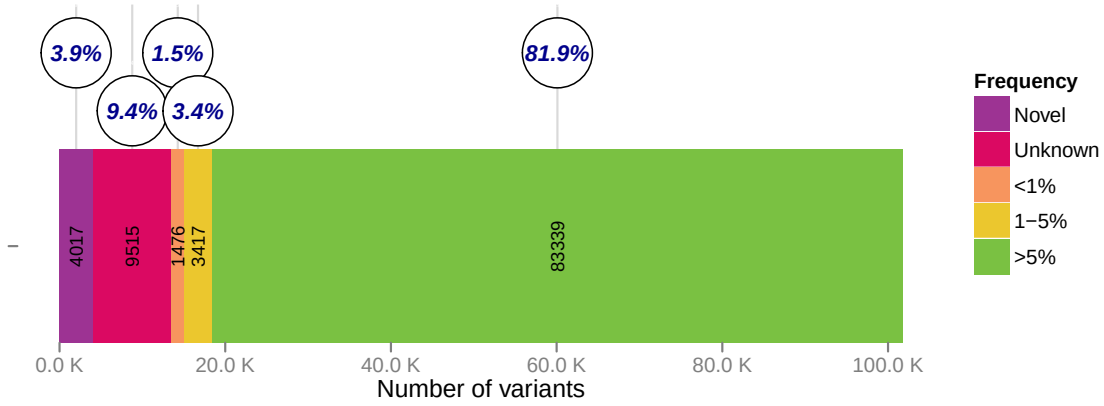


Figure 3: Variant frequencies. The allele frequencies of the variants in your exome. Unknown: allele is present in a public database but no frequency data was available.

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

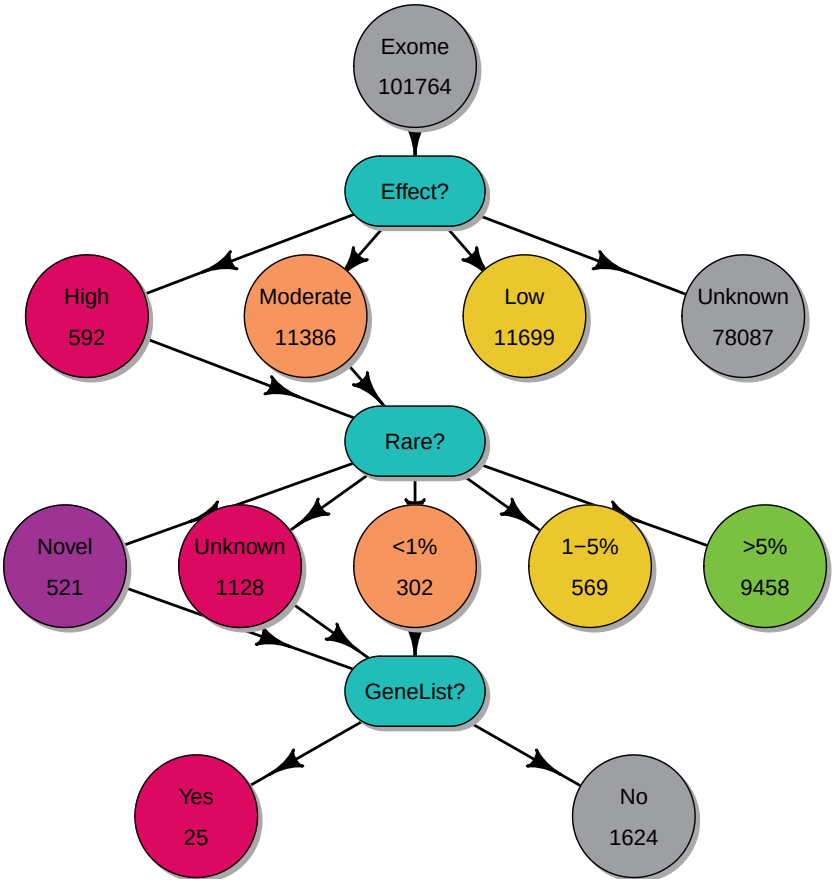


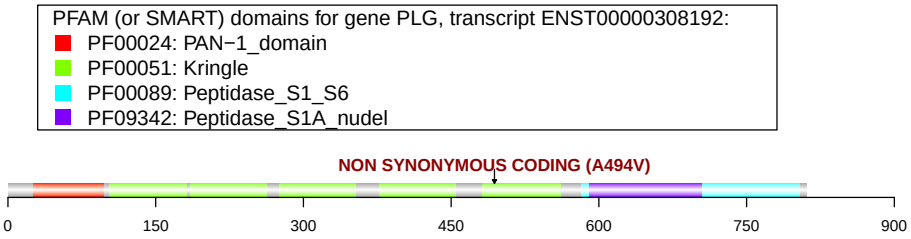
Figure 4: Variant filtering decision tree. A graphical representation of the filtering process that was used to generate your short list of variants of interest.

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).

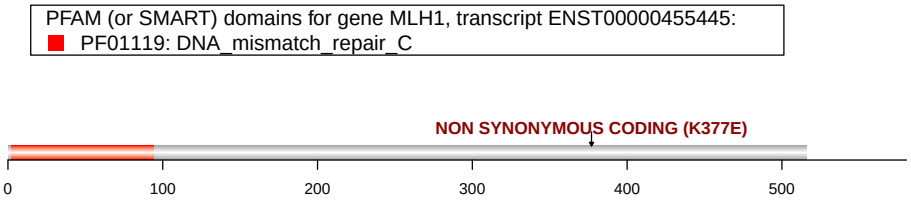
We hope you find this initial list of variants interesting and that it will help you in your journey through your exome. This short list of variants only scratches the surface of what your genome contains and is just the beginning of where your data can take you. Have fun!

List of selected variants

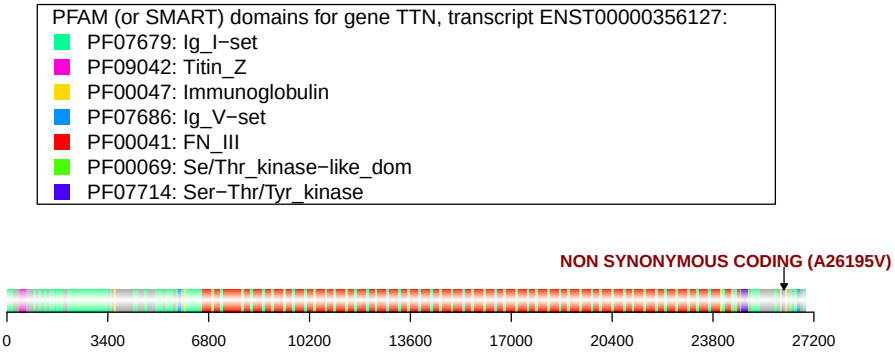
Variant 1:	**Gene:** PLG **Your genotype:** C/T **Location:** chr6:161152819	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00960	**dbSNP:** rs4252128
Quality:	**Genotype quality:** 99	**Coverage depth:** 77
Details:	**Gene description:** plasminogen	
	Transcript: ENST00000308192	**AA change:** A494V
	EntrezId: 5340	**EnsemblId:** ENSG00000122194
	UniProt: P00747	**OMIM:** 173350



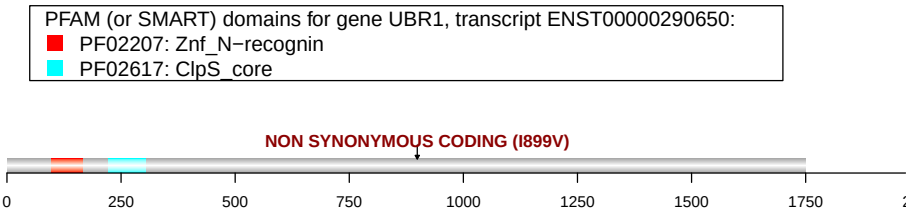
Variant 2:	**Gene:** MLH1 **Your genotype:** A/G **Location:** chr3:37089130	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00410	**dbSNP:** rs35001569
Quality:	**Genotype quality:** 99	**Coverage depth:** 96
Details:	**Gene description:** mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)	
	Transcript: ENST00000455445	**AA change:** K377E
	EntrezId: 4292	**EnsemblId:** ENSG00000076242
	UniProt: P40692	**OMIM:** 120436



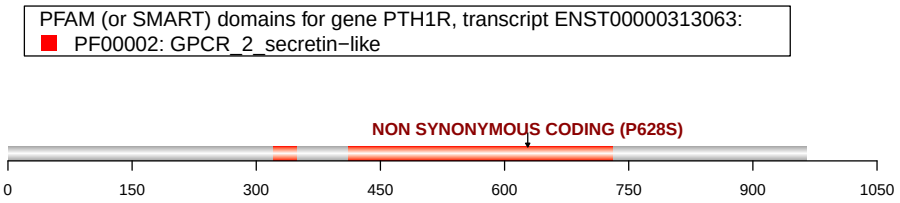
Variant 3:	**Gene:** TTN **Your genotype:** G/A **Location:** chr2:179395554	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00530	**dbSNP:** rs66961115
Quality:	**Genotype quality:** 99	**Coverage depth:** 105
Details:	**Gene description:** titin	
	Transcript: ENST00000356127	**AA change:** A26195V
	EntrezId: 7273	**EnsemblId:** ENSG00000155657
	UniProt: Q8WZ42	**OMIM:** 188840



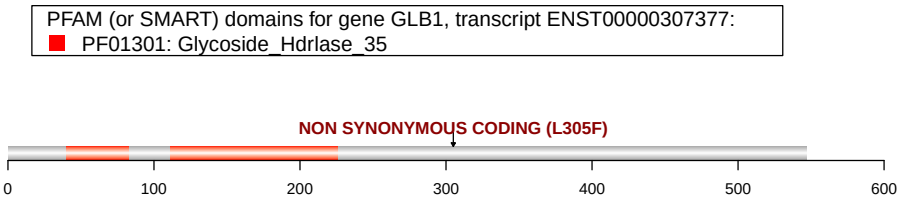
Variant 4:	**Gene:** UBR1 **Your genotype:** T/C **Location:** chr15:43317071	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00960	**dbSNP:** rs35069201
Quality:	**Genotype quality:** 99	**Coverage depth:** 93
Details:	**Gene description:** ubiquitin protein ligase E3 component n-recognin 1	
	Transcript: ENST00000290650	**AA change:** I899V
	EntrezId: 197131	**EnsemblId:** ENSG00000159459
	UniProt: Q8IWV7	**OMIM:** 605981



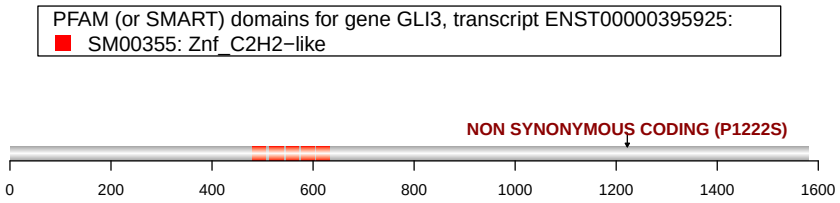
Variant 5:	**Gene:** PTH1R **Your genotype:** C/T **Location:** chr3:46943229	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00190	**dbSNP:** rs144293126
Quality:	**Genotype quality:** 99	**Coverage depth:** 40
Details:	**Gene description:** parathyroid hormone 1 receptor	
	Transcript: ENST00000313063	**AA change:** P628S
	EntrezId: 5745	**EnsemblId:** ENSG00000160801
	UniProt: Q03431	**OMIM:** 168468



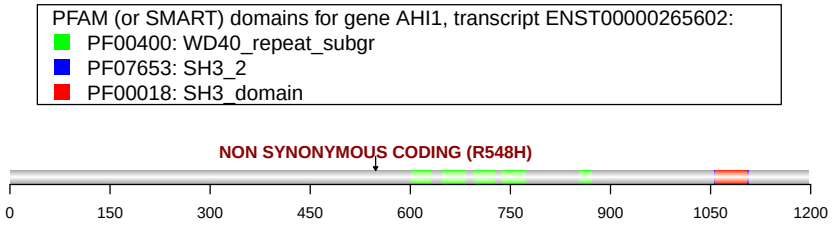
Variant 6:	**Gene:** GLB1 **Your genotype:** G/A **Location:** chr3:33059981	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00600	**dbSNP:** rs34421970
Quality:	**Genotype quality:** 99	**Coverage depth:** 49
Details:	**Gene description:** galactosidase, beta 1	
	Transcript: ENST00000307377	**AA change:** L305F
	EntrezId: 2720	**EnsemblId:** ENSG00000170266
	UniProt: P16278	**OMIM:** 611458



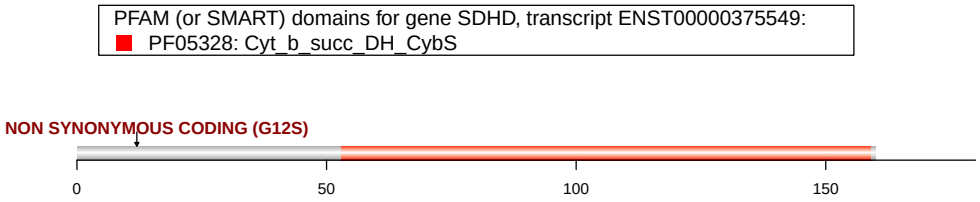
Variant 7:	**Gene:** GLI3 **Your genotype:** G/A **Location:** chr7:42005007	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00520	**dbSNP:** rs118149040
Quality:	**Genotype quality:** 99	**Coverage depth:** 40
Details:	**Gene description:** GLI family zinc finger 3	
	Transcript: ENST00000395925	**AA change:** P1222S
	EntrezId: 2737	**EnsemblId:** ENSG00000106571
	UniProt: P10071	**OMIM:** 165240



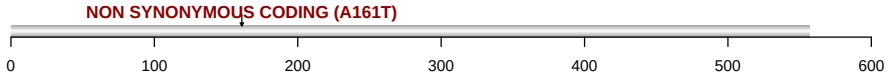
Variant 8:	**Gene:** AHI1 **Your genotype:** C/T **Location:** chr6:135768282	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00930	**dbSNP:** rs35433555
Quality:	**Genotype quality:** 99	**Coverage depth:** 111
Details:	**Gene description:** Abelson helper integration site 1	
	Transcript: ENST00000265602	**AA change:** R548H
	EntrezId: 54806	**EnsemblId:** ENSG00000135541
	UniProt: Q8N157	**OMIM:** 608894



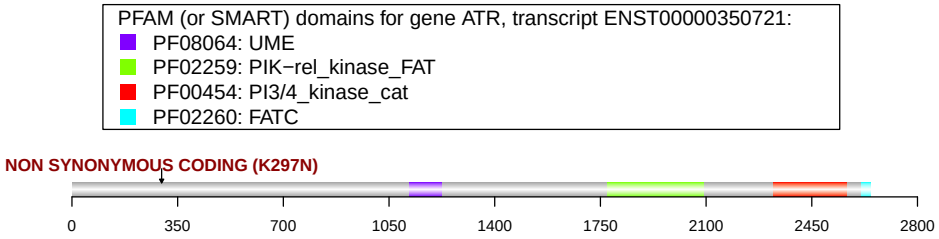
Variant 9:	**Gene:** SDHD **Your genotype:** G/A **Location:** chr11:111957665	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00640	**dbSNP:** rs34677591
Quality:	**Genotype quality:** 92.65	**Coverage depth:** 16
Details:	**Gene description:** succinate dehydrogenase complex, subunit D, integral membrane protein	
	Transcript: ENST00000375549	**AA change:** G12S
	EntrezId: 6392	**EnsemblId:** ENSG00000204370
	UniProt: O14521	**OMIM:** 602690



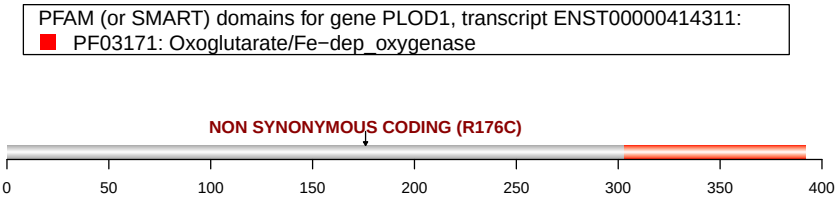
Variant 10:	**Gene:** SEPN1 **Your genotype:** G/A **Location:** chr1:26135116	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00800	**dbSNP:** rs115852080
Quality:	**Genotype quality:** 99	**Coverage depth:** 73
Details:	**Gene description:** selenoprotein N, 1	
	Transcript: ENST00000354177	**AA change:** A161T
	EntrezId: 57190	**EnsemblId:** ENSG00000162430
	UniProt: Q9NZV5	**OMIM:** 606210



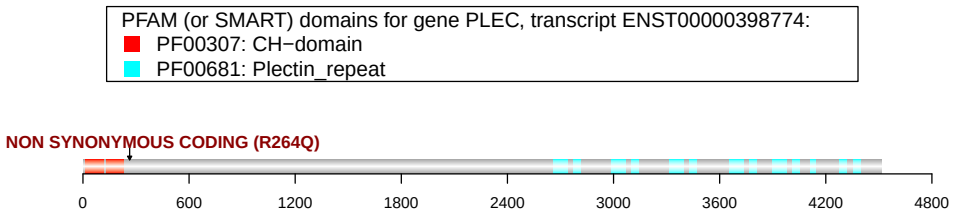
Variant 11:	**Gene:** ATR **Your genotype:** C/G **Location:** chr3:142281353	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00550	**dbSNP:** rs2229033
Quality:	**Genotype quality:** 99	**Coverage depth:** 69
Details:	**Gene description:** ataxia telangiectasia and Rad3 related	
	Transcript: ENST00000350721	**AA change:** K297N
	EntrezId: 545	**EnsemblId:** ENSG00000175054
	UniProt: Q13535	**OMIM:** 601215



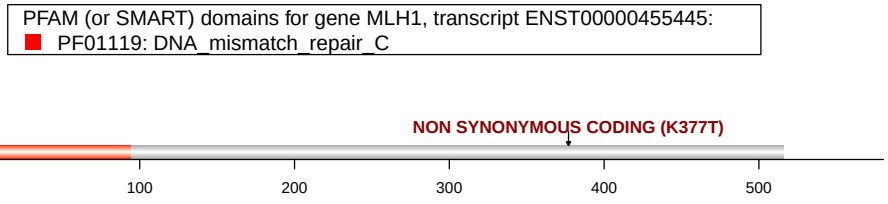
Variant 12:	**Gene:** PLOD1 **Your genotype:** C/T **Location:** chr1:12025600	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00280	**dbSNP:** rs138490756
Quality:	**Genotype quality:** 99	**Coverage depth:** 59
Details:	**Gene description:** procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1	
	Transcript: ENST00000414311	**AA change:** R176C
	EntrezId: 5351	**EnsemblId:** ENSG00000083444
	UniProt: Q02809	**OMIM:** 153454



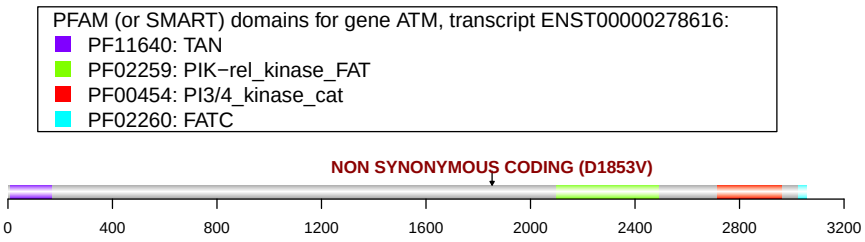
Variant 13:	Gene: PLEC Your genotype: C/T Location: chr8:145009036	
Effect:	Impact: NON SYNONYMOUS CODING	Type: MODERATE
Frequency:	1KGenomes: 0.00610	dbSNP: rs138924815
Quality:	Genotype quality: 99	Coverage depth: 18
Details:	Gene description: plectin Transcript: ENST00000398774 EntrezId: 5339 UniProt: Q15149	 AA change: R264Q EnsemblId: ENSG00000178209 OMIM: 601282



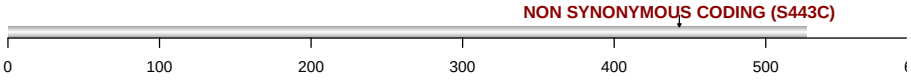
Variant 14:	Gene: MLH1 Your genotype: A/C Location: chr3:37089131	
Effect:	Impact: NON SYNONYMOUS CODING	Type: MODERATE
Frequency:	1KGenomes: 0.00410	dbSNP: rs63750449
Quality:	Genotype quality: 99	Coverage depth: 97
Details:	Gene description: mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)	
	Transcript: ENST00000455445 EntrezId: 4292 UniProt: P40692	AA change: K377T EnsemblId: ENSG00000076242 OMIM: 120436



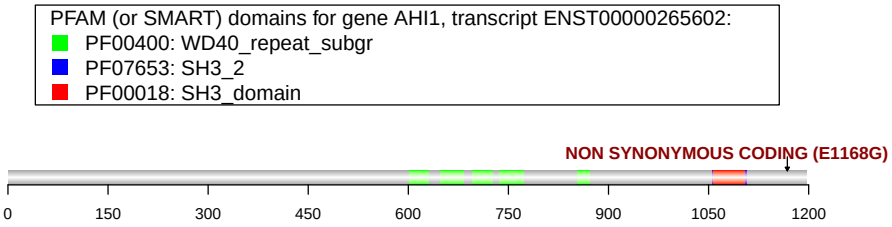
Variant 15:	**Gene:** ATM **Your genotype:** A/T **Location:** chr11:108175463	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00270	**dbSNP:** rs1801673
Quality:	**Genotype quality:** 99	**Coverage depth:** 42
Details:	**Gene description:** ataxia telangiectasia mutated	
	Transcript: ENST00000278616	**AA change:** D1853V
	EntrezId: 472	**EnsemblId:** ENSG00000149311
	UniProt: Q13315	**OMIM:** 607585



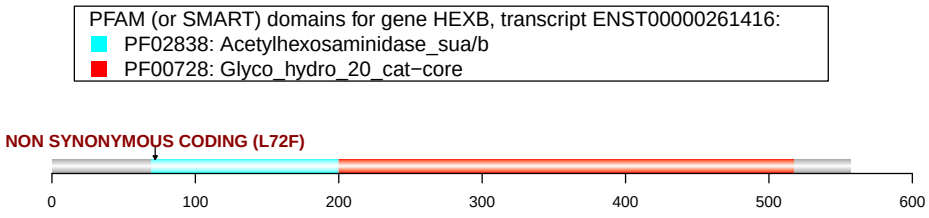
Variant 16:	**Gene:** TSEN54 **Your genotype:** C/G **Location:** chr17:73519758	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00870	**dbSNP:** rs150169668
Quality:	**Genotype quality:** 99	**Coverage depth:** 41
Details:	**Gene description:** tRNA splicing endonuclease 54 homolog (S. cerevisiae)	
	Transcript: ENST00000333213	**AA change:** S443C
	EntrezId: 283989	**EnsemblId:** ENSG00000182173
	UniProt: Q7Z6J9	**OMIM:** 608755



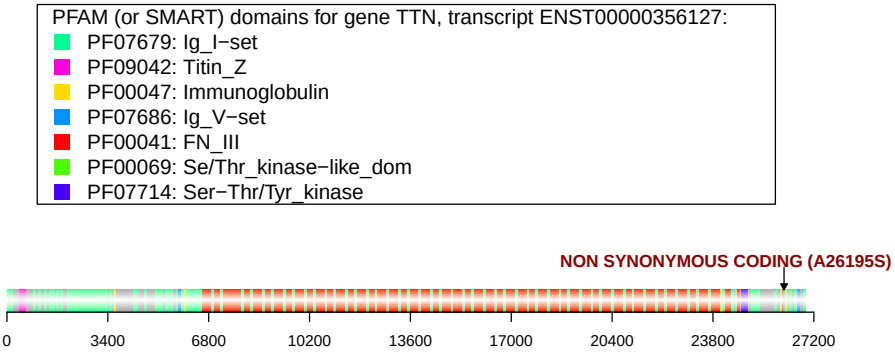
Variant 17:	Gene: AHI1 Your genotype: T/C Location: chr6:135611646	
Effect:	Impact: NON SYNONYMOUS CODING	Type: MODERATE
Frequency:	1KGenomes: 5e-04	dbSNP: NA
Quality:	Genotype quality: 99	Coverage depth: 112
Details:	Gene description: Abelson helper integration site 1	
	Transcript: ENST00000265602	AA change: E1168G
	EntrezId: 54806	EnsemblId: ENSG00000135541
	UniProt: Q8N157	OMIM: 608894



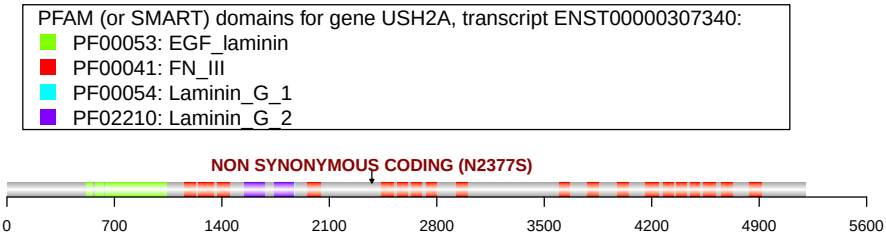
Variant 18:	Gene: HEXB Your genotype: C/T Location: chr5:73981299	
Effect:	Impact: NON SYNONYMOUS CODING	Type: MODERATE
Frequency:	1KGenomes: 0.00580	dbSNP: rs147155126
Quality:	Genotype quality: 99	Coverage depth: 37
Details:	Gene description: hexosaminidase B (beta polypeptide)	
	Transcript: ENST00000261416	AA change: L72F
	EntrezId: 3074	EnsemblId: ENSG00000049860
	UniProt: P07686	OMIM: 606873



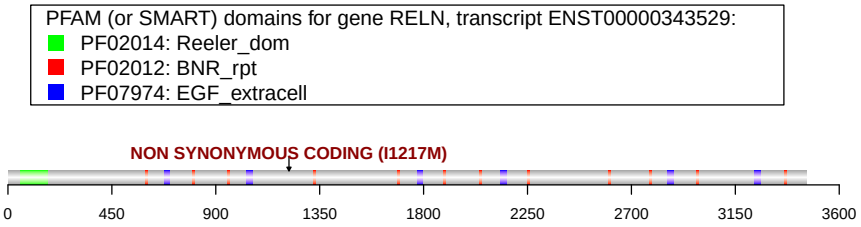
Variant 19:	**Gene:** TTN **Your genotype:** C/A **Location:** chr2:179395555	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00530	**dbSNP:** rs67254537
Quality:	**Genotype quality:** 99	**Coverage depth:** 105
Details:	**Gene description:** titin	
	Transcript: ENST00000356127	**AA change:** A26195S
	EntrezId: 7273	**EnsemblId:** ENSG00000155657
	UniProt: Q8WZ42	**OMIM:** 188840



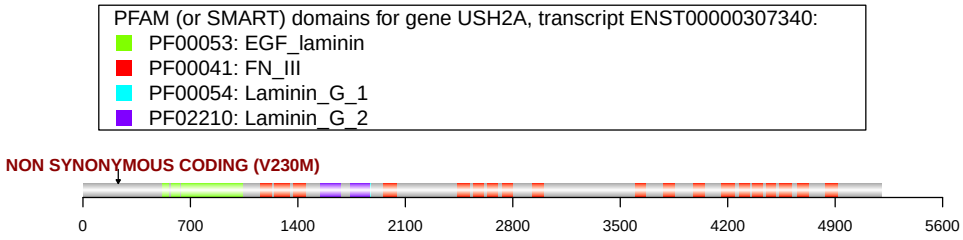
Variant 20:	**Gene:** USH2A **Your genotype:** T/C **Location:** chr1:216108128	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00410	**dbSNP:** rs111033394
Quality:	**Genotype quality:** 99	**Coverage depth:** 45
Details:	**Gene description:** Usher syndrome 2A (autosomal recessive, mild)	
	Transcript: ENST00000307340	**AA change:** N2377S
	EntrezId: 7399	**EnsemblId:** ENSG00000042781
	UniProt: O75445	**OMIM:** 608400



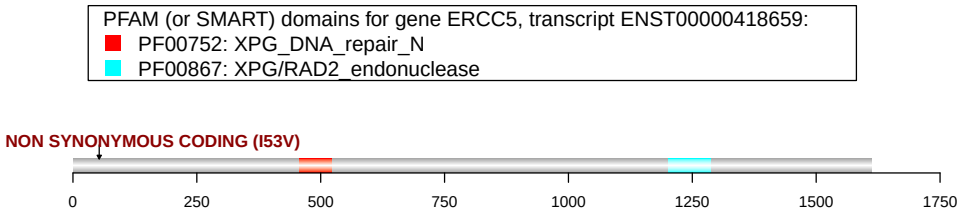
Variant 21:	**Gene:** RELN **Your genotype:** G/C **Location:** chr7:103234828	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00130	**dbSNP:** rs56342240
Quality:	**Genotype quality:** 99	**Coverage depth:** 105
Details:	**Gene description:** reelin	
	Transcript: ENST00000343529	**AA change:** I1217M
	EntrezId: 5649	**EnsemblId:** ENSG00000189056
	UniProt: P78509	**OMIM:** 600514



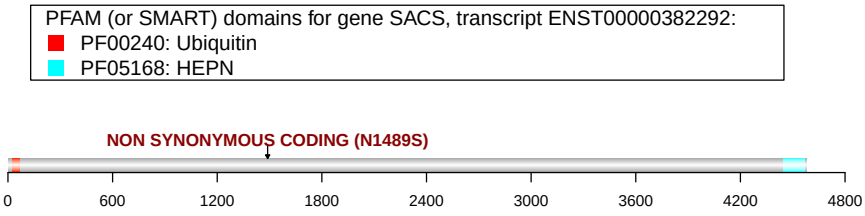
Variant 22:	**Gene:** USH2A **Your genotype:** C/T **Location:** chr1:216538391	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00870	**dbSNP:** rs45500891
Quality:	**Genotype quality:** 99	**Coverage depth:** 31
Details:	**Gene description:** Usher syndrome 2A (autosomal recessive, mild)	
	Transcript: ENST00000307340	**AA change:** V230M
	EntrezId: 7399	**EnsemblId:** ENSG00000042781
	UniProt: O75445	**OMIM:** 608400



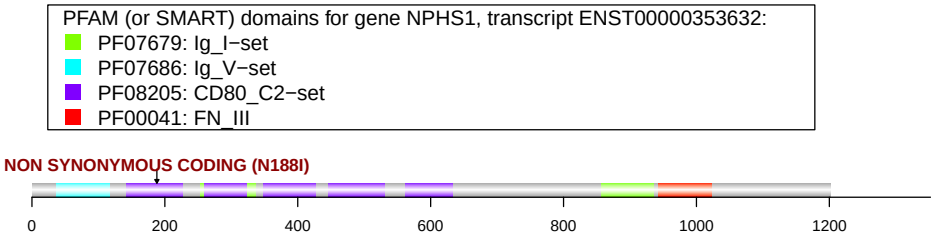
Variant 23:	**Gene:** ERCC5 **Your genotype:** A/G **Location:** chr13:103459861	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00230	**dbSNP:** rs41281670
Quality:	**Genotype quality:** 99	**Coverage depth:** 55
Details:	**Gene description:** excision repair cross-complementing rodent repair deficiency, complementation group 5	
	Transcript: ENST00000418659	**AA change:** I53V
	EntrezId: 2073	**EnsemblId:** ENSG00000134899
	UniProt: P28715	**OMIM:** 133530



Variant 24:	**Gene:** SACS **Your genotype:** T/C **Location:** chr13:23913549	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00460	**dbSNP:** rs147099630
Quality:	**Genotype quality:** 99	**Coverage depth:** 64
Details:	**Gene description:** spastic ataxia of Charlevoix-Saguenay (sacsin)	
	Transcript: ENST00000382292	**AA change:** N1489S
	EntrezId: 26278	**EnsemblId:** ENSG00000151835
	UniProt: Q9NZJ4	**OMIM:** 604490



Variant 25:	**Gene:** NPHS1 **Your genotype:** T/A **Location:** chr19:36341311	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00290	**dbSNP:** rs145125791
Quality:	**Genotype quality:** 99	**Coverage depth:** 40
Details:	**Gene description:** nephrosis 1, congenital, Finnish type (nephrin)	
	Transcript: ENST00000353632	**AA change:** N188I
	EntrezId: 4868	**EnsemblId:** ENSG00000161270
	UniProt: O60500	**OMIM:** 602716



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.