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

Exome Results & Raw Data Summary

Generated on: October 12, 2012

Congratulations on being a part of 23andMe's Exome Pilot! Earlier this year, we provided the raw sequencing data and an initial report of your processed results. Since then, we've been working on improving our analysis and generating a final report to summarize your exome. Here are some important points about your final report:

- Your final data comes in the form of two files: 1) the variant call file (VCF) that contains information about the positions where you differ from the human reference genome (ie. variants), 2) a BED file containing the genomic regions where we could confidently assess your genotype including positions where you match the reference genome. Both of these files are viewable using a text editor.
- The final VCF file provided is improved over the initial one. In this version, we identified variants based on the data of all people in the exome pilot, and updated variant quality estimates based on known variation. This allows us to better identify and filter your variants, please see the appendix for more details.

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

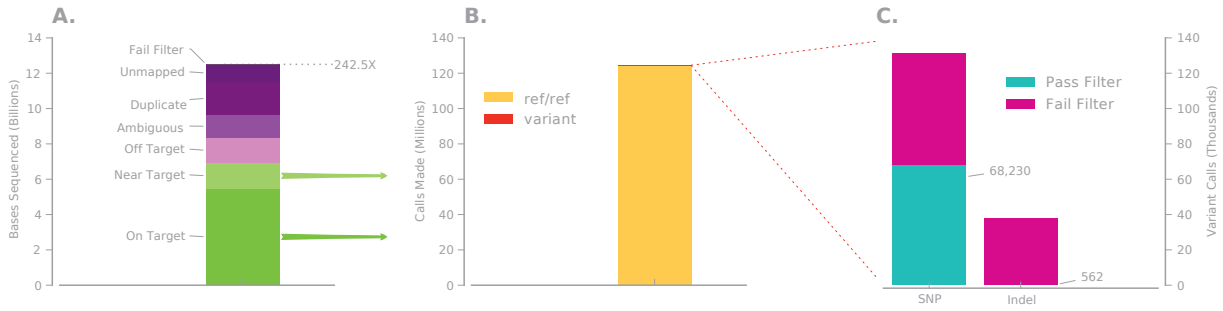


Figure 1: Getting from raw reads to called variants. A) The number of bases obtained by sequencing your exome. The top line indicates total coverage. B) Total number of called bases in your exome. The vast majority are the same as the reference genome. C) An expansion of the small sliver of variants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

Welcome to your exome, the 50 million DNA bases of your genome encoding all your proteins. This data begins as a collection of raw reads which are then aligned against the reference genome (Figure 1A). We analyze the regions where multiple reads overlap to detect where your DNA sequence differs from the reference. In most positions, you will match the reference sequence exactly (Figure 1B), but the small number of variants where you differ are collected into a final VCF file (Figure 1C). The figures in this report are based on the variants that pass all filters.

There are many approaches to this process. We implemented the Broad Institute's "Best Practice" protocol for exome sequence analysis (see Appendix). You can read a detailed description of it here.

Characterizing your variants

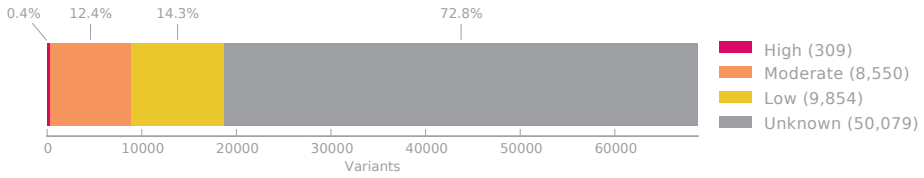


Figure 2: Predicting impact of variants on gene function. An overview of your variants and their predicted impact on gene function.

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, a 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:

All Variant falls either in an intron, UTR, non-coding transcript or up-/downstream of a gene. These variants are less likely to impact the amino-acid sequence of the protein, however may affect other elements of gene expression.

How rare are your variants?

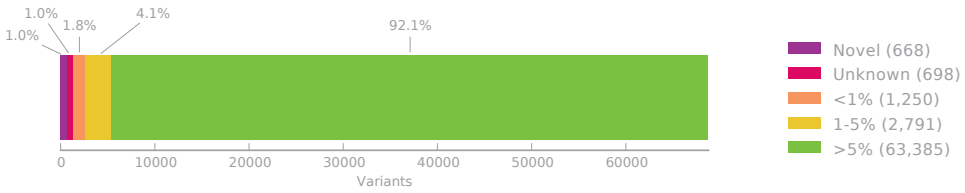


Figure 3: Variant frequencies. The allele frequencies of the variants in your exome. Unknown: allele is present in a public database but no frequency data was available.

One of the advantages of exome sequencing is that we can detect sequence variants that are unique to you! We compared your variants to dbSNP (build 135) and the variants detected by the 1000 Genomes Project (release: 08-26-2011) to divide your variants into the following categories:

- **novel** variant has not been observed in either database
- **unknown** variant has been observed in dbSNP but not the 1000 Genomes dataset and therefore no allele frequency is available
- **rare** variant with an allele frequency $<1\%$
- **somewhat rare** variant with a frequency 1-5%
- **common** frequency of the variant is greater than 5%

Comparing your variants

Now that we have data for everybody in the exome pilot we can see how you compare to the other participants. In the following series of figures we divide your variants into different categories and plot the number of variants in each category as bar chart. We then overlay a Box Plot showing a summary of the equivalent distribution for all exomes in the pilot.

There are many different ways that we could compare the data, here are the ones that we found to be the most informative:

Impact
: Figure 4 breaks down your variants by their predicted impact on gene function.

Effect
: Figure 5 takes your high-impact variants and further classifies them according to their predicted effect on the gene product.

Location
: Figure 6 looks at the location of your variants relative to the coding sequence.

Frequency
: Figure 7 looks at the allele frequencies of your variants.

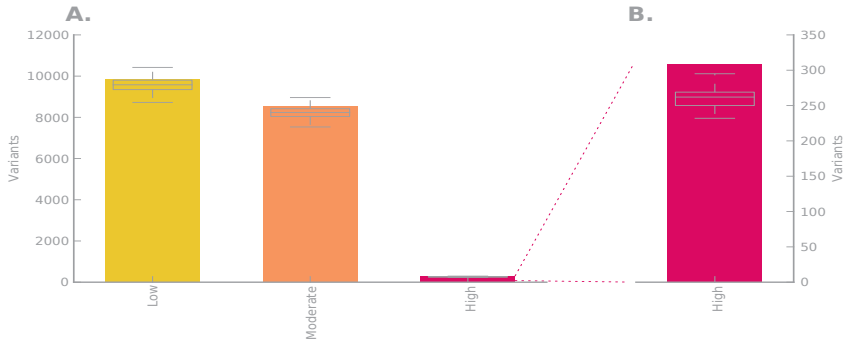


Figure 4: A comparison of the predicted impact of your variants. A) A breakdown of your variants into Low, Medium and High predicted impact (those with Unknown impact not shown). B) Zoom-in of variants predicted to have high impact.

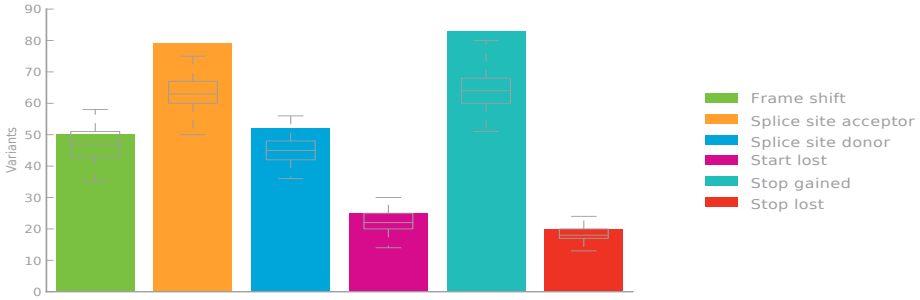


Figure 5: A comparison of the predicted effect of your high-impact variants. Your high-impact variants classifed according to their predicted effect on the gene product.

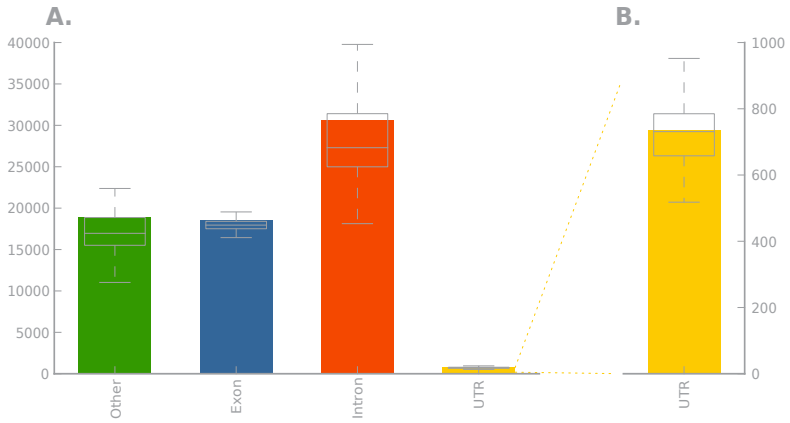


Figure 6: A comparison of the location of your variants relative to the coding sequence. A) Your classified according to whether they overlap the coding portion of a transcript (Exon), the non-coding portion of a transcript (UTR) or an intron. Variants that are either upstream or downstream of a gene or in non-coding transcripts are classified as 'Other'. B) Zoom-in of variants located in the UTRs.

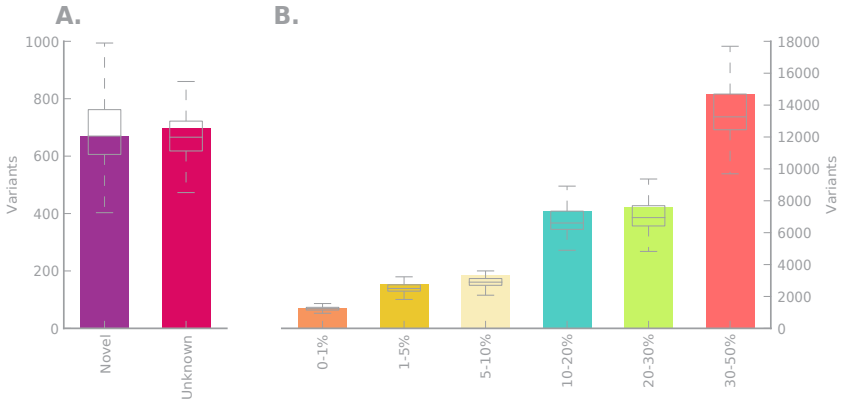


Figure 7: A comparison of the allele frequencies of your variants. A) The number of variants in your exome that are not present in one of the public databases (Novel) and those with no allele frequency in the 1000 Genomes Project (Unknown). B) The remainder of your variants with an allele frequency $< 50\%$ categorized by frequency.

Filtering your variants

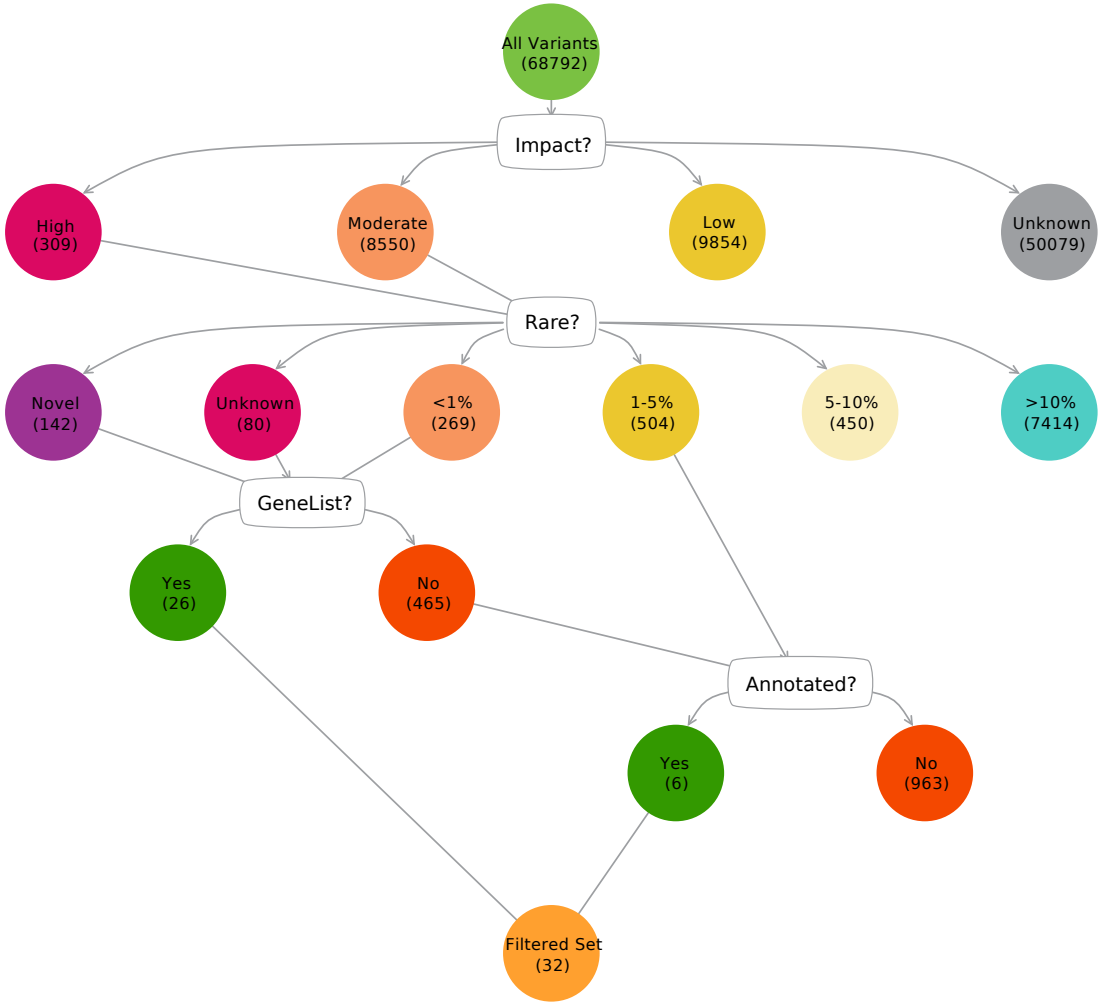


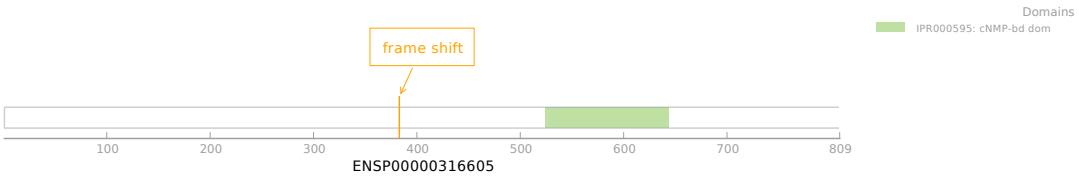
Figure 8: Variant filtering decision tree. A graphical representation of the filtering process that was used to generate your short list of variants of interest.

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 variants with potential functional effects (such as contributing to disease or other phenotypes of interest) we used four consecutive filters, depicted in the figure above: (1) impact 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; (4) annotated in dbSNP as either pathogenic or probably pathogenic.

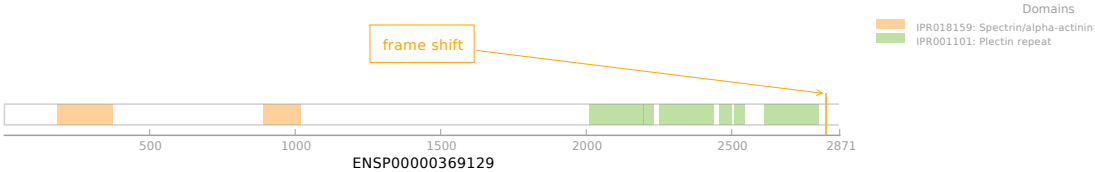
We hope you find this initial list of variants interesting and that it will help you in your journey through your exome. This short list of variants only scratches the surface of what your genome contains and is just the beginning of where your data can take you. Have fun!

List of selected variants

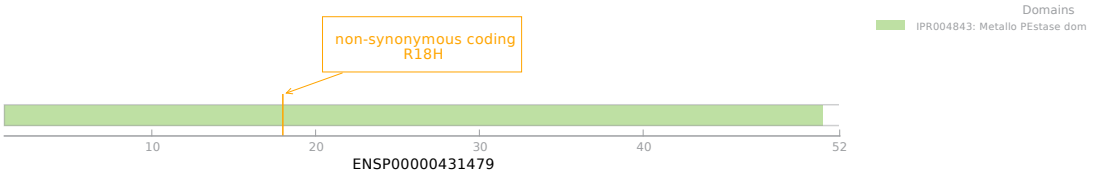
Variant 1:	**Gene:** CNGB3 **Your genotype:** AG/A **Location:** chr8:87656008	
Effect:	FRAME SHIFT	**Type:**HIGH
Frequency:	**1KGenomes:** NA	**dbSNP:** NA
Quality:	**Genotype quality:** 99.00	**Coverage depth:** 121
Details:	**Gene description:** cyclic nucleotide gated channel beta 3	
	Transcript: ENST00000320005	**AA change:** NA
	EntrezId: 54714	**EnsemblId:** ENSG00000170289
	UniProt: Q9NQW8	**OMIM:** 605080



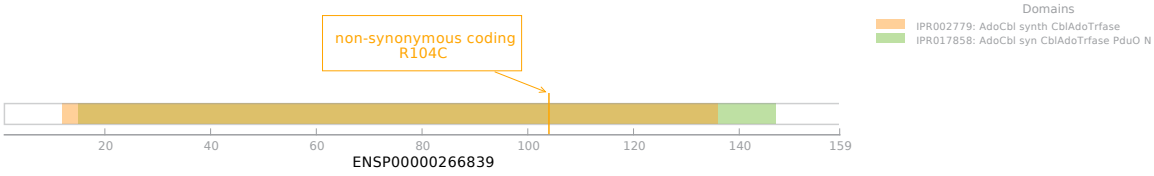
Variant 2:	**Gene:** DSP **Your genotype:** G/GC **Location:** chr6:7585967	
Effect:	FRAME SHIFT	**Type:**HIGH
Frequency:	**1KGenomes:** NA	**dbSNP:** NA
Quality:	**Genotype quality:** 99.00	**Coverage depth:** 229
Details:	**Gene description:** desmoplakin	
	Transcript: ENST00000379802	**AA change:** NA
	EntrezId: 1832	**EnsemblId:** ENSG00000096696
	UniProt: P15924	**OMIM:** 125647



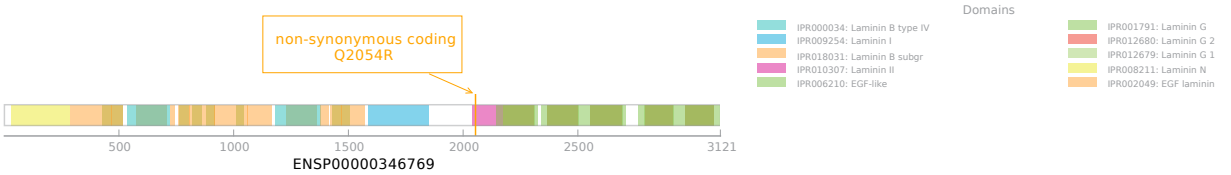
Variant 3:	Gene: SMPD1 Your genotype: G/A Location: chr11:6413167	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0005	dbSNP: rs1803161
Quality:	Genotype quality: 99.00	Coverage depth: 138
Details:	Gene description: sphingomyelin phosphodiesterase 1, acid lysosomal	
	Transcript: ENST00000530395	AA change: R18H
	EntrezId: 6609	EnsemblId: ENSG00000166311
	UniProt: NA	OMIM: 607608



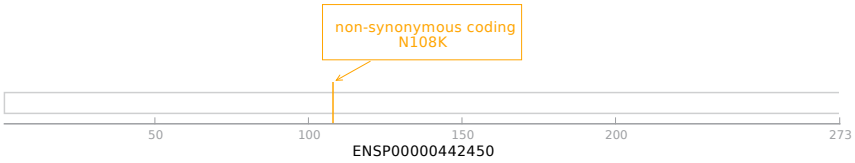
Variant 4:	Gene: MMAB Your genotype: G/A Location: chr12:109998846	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0009	dbSNP: NA
Quality:	Genotype quality: 99.00	Coverage depth: 52
Details:	Gene description: methylmalonic aciduria (cobalamin deficiency) cblB type	
	Transcript: ENST00000266839	AA change: R104C
	EntrezId: 326625	EnsemblId: ENSG00000139428
	UniProt: Q96EY8	OMIM: 607568



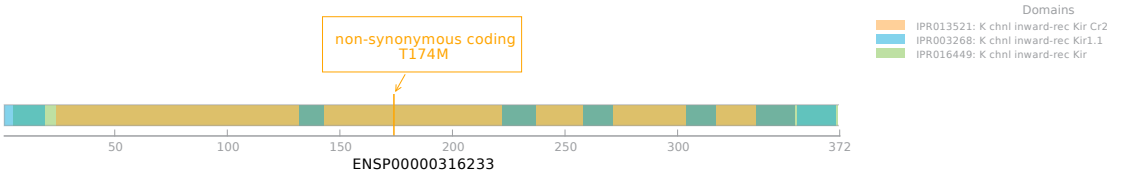
Variant 5:	Gene: LAMA2 Your genotype: A/G Location: chr6:129762036	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0023	dbSNP: rs56035053
Quality:	Genotype quality: 99.00	Coverage depth: 88
Details:	Gene description: laminin, alpha 2 Transcript: ENST00000354729 EntrezId: 3908 UniProt: NA	 AA change: Q2054R EnsemblId: ENSG00000196569 OMIM: 156225



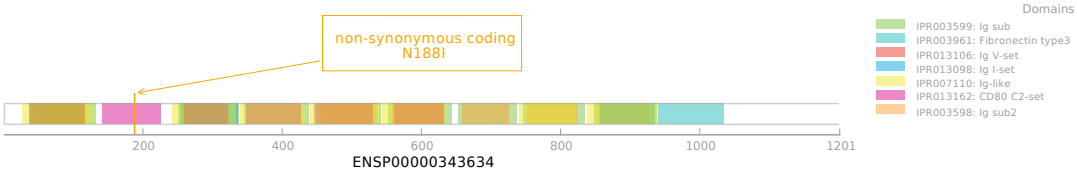
Variant 6:	Gene: CLN5 Your genotype: C/A Location: chr13:77574606	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0027	dbSNP: rs138611001
Quality:	Genotype quality: 99.00	Coverage depth: 63
Details:	Gene description: ceroid-lipofuscinosis, neuronal 5 Transcript: ENST00000535238 EntrezId: 1203 UniProt: NA	 AA change: N108K EnsemblId: ENSG00000102805 OMIM: 608102



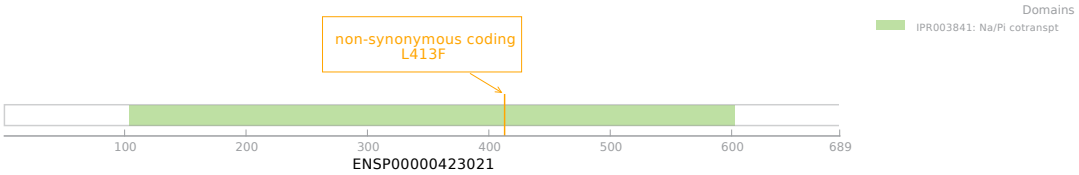
Variant 7:	Gene: KCNJ1 Your genotype: G/A Location: chr11:128709618	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0027	dbSNP: rs117535913
Quality:	Genotype quality: 99.00	Coverage depth: 123
Details:	Gene description: potassium inwardly-rectifying channel, subfamily J, member 1	
	Transcript: ENST00000324036	AA change: T174M
	EntrezId: 3758	EnsemblId: ENSG00000151704
	UniProt: NA	OMIM: 600359



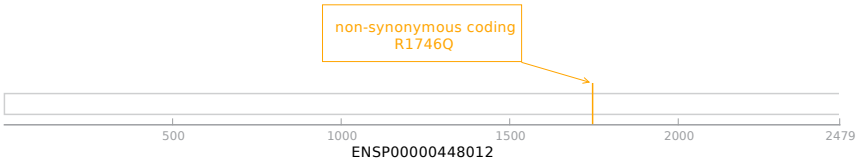
Variant 8:	Gene: NPHS1 Your genotype: T/A Location: chr19:36341311	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0029	dbSNP: rs145125791
Quality:	Genotype quality: 99.00	Coverage depth: 68
Details:	Gene description: nephrosis 1, congenital, Finnish type (nephrin)	
	Transcript: ENST00000353632	AA change: N188I
	EntrezId: 4868	EnsemblId: ENSG00000161270
	UniProt: NA	OMIM: 602716



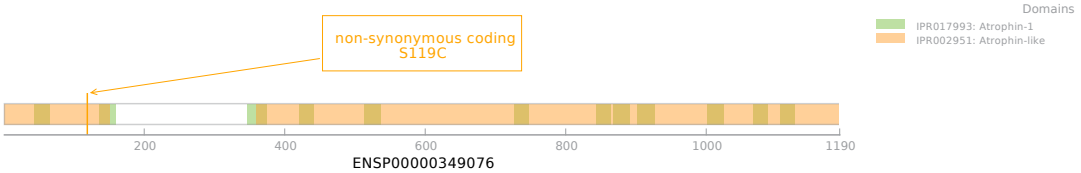
Variant 9:	Gene: SLC34A2 Your genotype: G/C Location: chr4:25675943	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0032	dbSNP: rs76404281
Quality:	Genotype quality: 99.00	Coverage depth: 85
Details:	Gene description: solute carrier family 34 (sodium phosphate), member 2	
	Transcript: ENST00000503434	AA change: L413F
	EntrezId: 10568	EnsemblId: ENSG00000157765
	UniProt: NA	OMIM: 604217



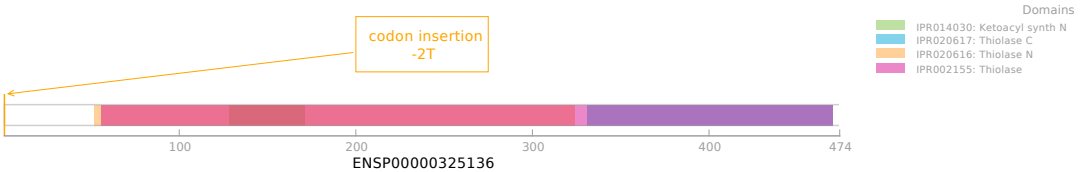
Variant 10:	Gene: CEP290 Your genotype: C/T Location: chr12:88472996	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0033	dbSNP: rs61941020
Quality:	Genotype quality: 99.00	Coverage depth: 126
Details:	Gene description: centrosomal protein 290kDa	
	Transcript: ENST00000552810	AA change: R1746Q
	EntrezId: 80184	EnsemblId: ENSG00000198707
	UniProt: NA	OMIM: 610142



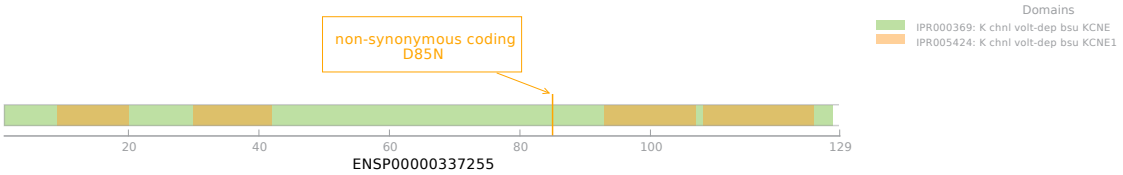
Variant 11:	**Gene:** ATN1 **Your genotype:** A/T **Location:** chr12:7044785	
Effect:	NON-SYNONYMOUS CODING	**Type:**MODERATE
Frequency:	**1KGenomes:** NA	**dbSNP:** NA
Quality:	**Genotype quality:** 99.00	**Coverage depth:** 81
Details:	**Gene description:** atrophin 1 **Transcript:** ENST00000356654 **EntrezId:** 1822 **UniProt:** P54259	 **AA change:** S119C **EnsemblId:** ENSG00000111676 **OMIM:** 607462



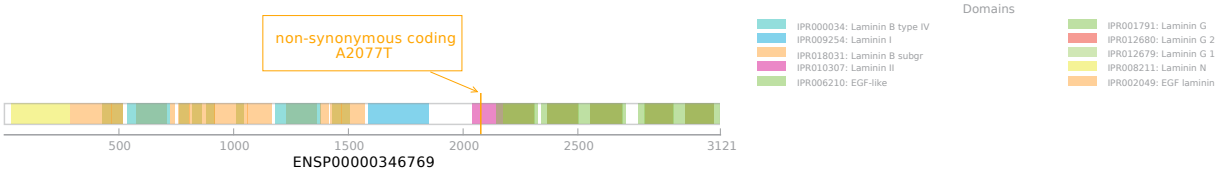
Variant 12:	**Gene:** HADHB **Your genotype:** GACT/GACT **Location:** chr2:26477125	
Effect:	CODON INSERTION	**Type:**MODERATE
Frequency:	**1KGenomes:** NA	**dbSNP:** rs3839049
Quality:	**Genotype quality:** 99.00	**Coverage depth:** 154
Details:	**Gene description:** hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), beta subunit **Transcript:** ENST00000317799 **EntrezId:** 3032 **UniProt:** P55084	 **AA change:** -2T **EnsemblId:** ENSG00000138029 **OMIM:** 143450



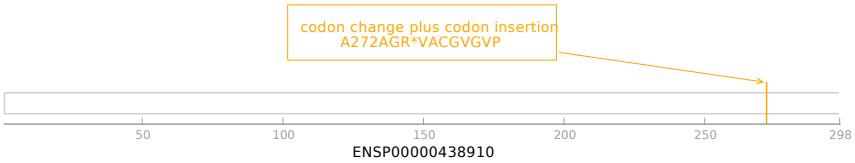
Variant 13:	Gene: KCNE1 Your genotype: C/T Location: chr21:35821680	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0049	dbSNP: rs1805128
Quality:	Genotype quality: 99.00	Coverage depth: 90
Details:	Gene description: potassium voltage-gated channel, Isk-related family, member 1	
	Transcript: ENST00000337385	AA change: D85N
	EntrezId: 3753	EnsemblId: ENSG00000180509
	UniProt: P15382	OMIM: 176261



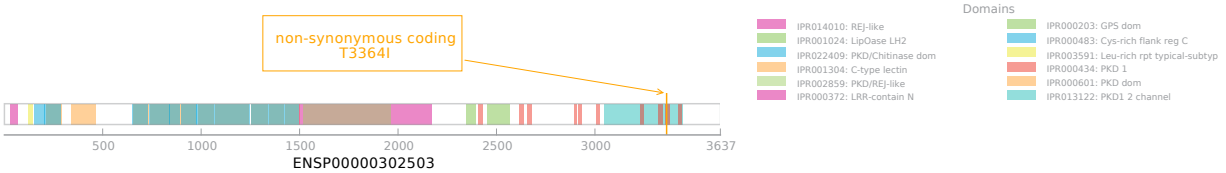
Variant 14:	Gene: LAMA2 Your genotype: G/A Location: chr6:129762104	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: NA	dbSNP: rs142264176
Quality:	Genotype quality: 99.00	Coverage depth: 61
Details:	Gene description: laminin, alpha 2	
	Transcript: ENST00000354729	AA change: A2077T
	EntrezId: 3908	EnsemblId: ENSG00000196569
	UniProt: NA	OMIM: 156225



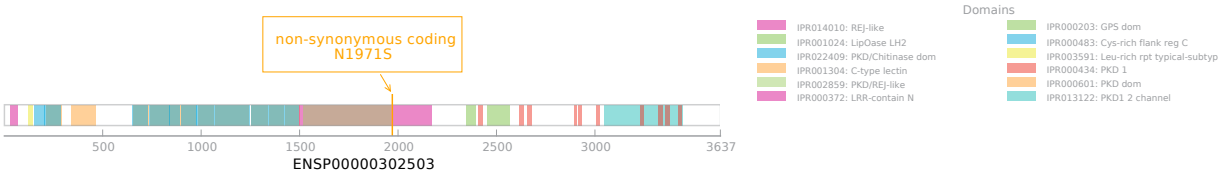
Variant 15:	**Gene:** MLC1 **Your genotype:** A/AGCACCCCCACCCCACAGGCCACTCACCTCCCCG **Location:** chr22:50502469	
Effect:	CODON CHANGE PLUS CODON INSERTION	**Type:**MODERATE
Frequency:	**1KGenomes:** NA	**dbSNP:** NA
Quality:	**Genotype quality:** 89.98	**Coverage depth:** 42
Details:	**Gene description:** megalencephalic leukoencephalopathy with subcortical cysts 1	
	Transcript: ENST00000535444	**AA change:** A272AGR*VACGVGVP
	EntrezId: 23209	**EnsemblId:** ENSG00000100427
	UniProt: NA	**OMIM:** 605908



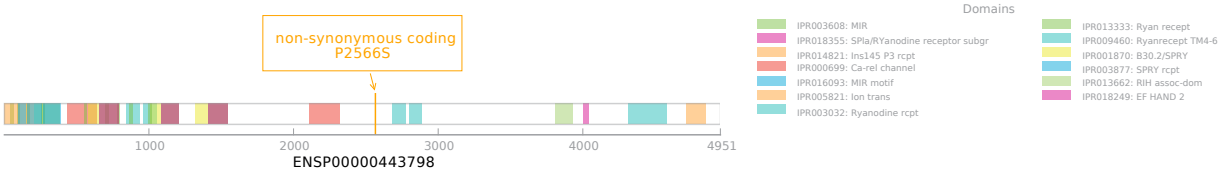
Variant 16:	**Gene:** PKD1 **Your genotype:** G/A **Location:** chr16:2140724	
Effect:	NON-SYNONYMOUS CODING	**Type:**MODERATE
Frequency:	**1KGenomes:** NA	**dbSNP:** NA
Quality:	**Genotype quality:** 99.00	**Coverage depth:** 59
Details:	**Gene description:** polycystic kidney disease 1 (autosomal dominant)	
	Transcript: ENST00000306101	**AA change:** T3364I
	EntrezId: 5310	**EnsemblId:** ENSG00000008710
	UniProt: NA	**OMIM:** 601313



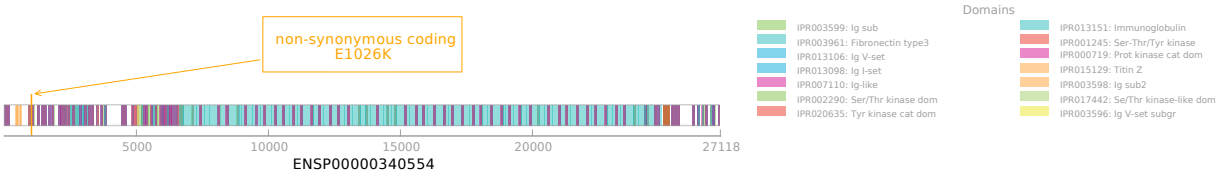
Variant 17:	Gene: PKD1 Your genotype: T/C Location: chr16:2155870	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: NA	dbSNP: NA
Quality:	Genotype quality: 99.00	Coverage depth: 44
Details:	Gene description: polycystic kidney disease 1 (autosomal dominant) Transcript: ENST00000306101 EntrezId: 5310 UniProt: NA	 AA change: N1971S EnsemblId: ENSG00000008710 OMIM: 601313



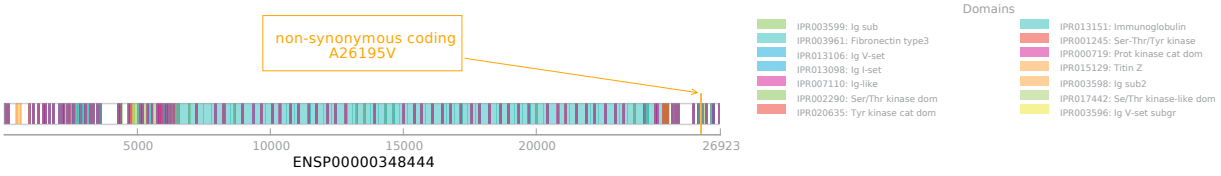
Variant 18:	Gene: RYR2 Your genotype: C/T Location: chr1:237814721	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: NA	dbSNP: NA
Quality:	Genotype quality: 99.00	Coverage depth: 76
Details:	Gene description: ryanodine receptor 2 (cardiac) Transcript: ENST00000542537 EntrezId: 6262 UniProt: NA	 AA change: P2566S EnsemblId: ENSG00000198626 OMIM: 180902



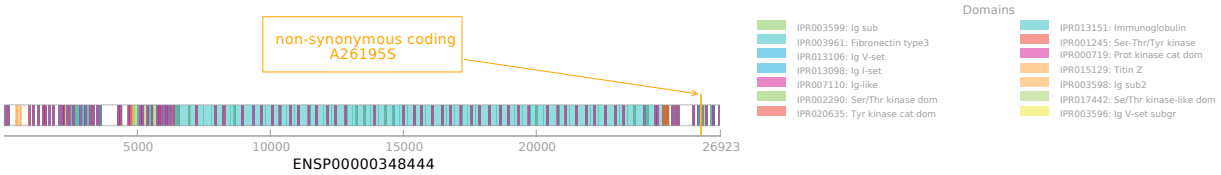
Variant 19:	Gene: TTN Your genotype: C/T Location: chr2:179647105	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: NA	dbSNP: NA
Quality:	Genotype quality: 99.00	Coverage depth: 67
Details:	Gene description: titin Transcript: ENST00000342175 EntrezId: 7273 UniProt: NA	AA change: E1026K EnsemblId: ENSG00000155657 OMIM: 188840



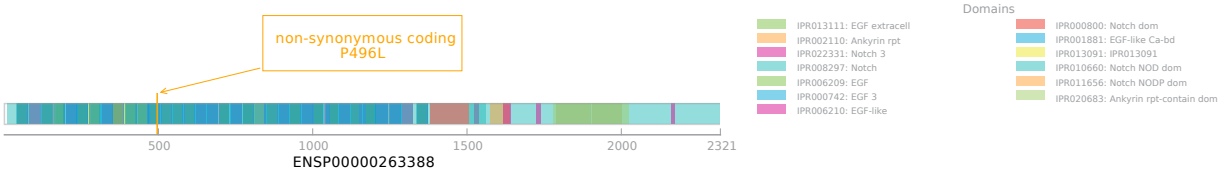
Variant 20:	Gene: TTN Your genotype: G/A Location: chr2:179395554	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0053	dbSNP: rs66961115
Quality:	Genotype quality: 99.00	Coverage depth: 201
Details:	Gene description: titin Transcript: ENST00000356127 EntrezId: 7273 UniProt: NA	AA change: A26195V EnsemblId: ENSG00000155657 OMIM: 188840



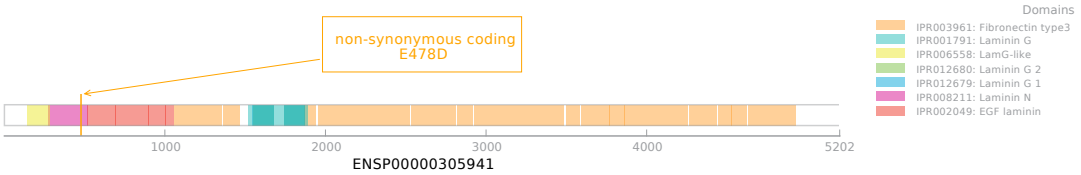
Variant 21:	Gene: TTN Your genotype: C/A Location: chr2:179395555	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0053	dbSNP: rs67254537
Quality:	Genotype quality: 99.00	Coverage depth: 201
Details:	Gene description: titin Transcript: ENST00000356127 EntrezId: 7273 UniProt: NA	AA change: A26195S EnsemblId: ENSG00000155657 OMIM: 188840



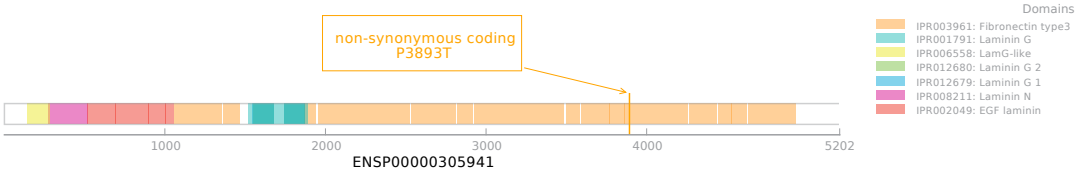
Variant 22:	Gene: NOTCH3 Your genotype: G/A Location: chr19:15299051	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0061	dbSNP: rs11670799
Quality:	Genotype quality: 49.59	Coverage depth: 14
Details:	Gene description: notch 3 Transcript: ENST00000263388 EntrezId: 4854 UniProt: Q9UM47	AA change: P496L EnsemblId: ENSG00000074181 OMIM: 600276



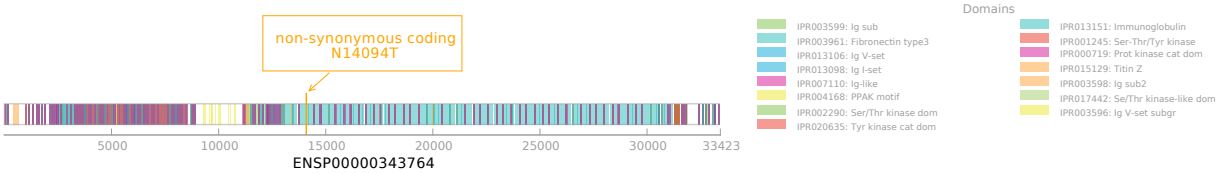
Variant 23:	Gene: USH2A Your genotype: C/G Location: chr1:216496932	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0073	dbSNP: rs35730265
Quality:	Genotype quality: 99.00	Coverage depth: 122
Details:	Gene description: Usher syndrome 2A (autosomal recessive, mild)	
	Transcript: ENST00000307340	AA change: E478D
	EntrezId: 7399	EnsemblId: ENSG00000042781
	UniProt: O75445	OMIM: 608400



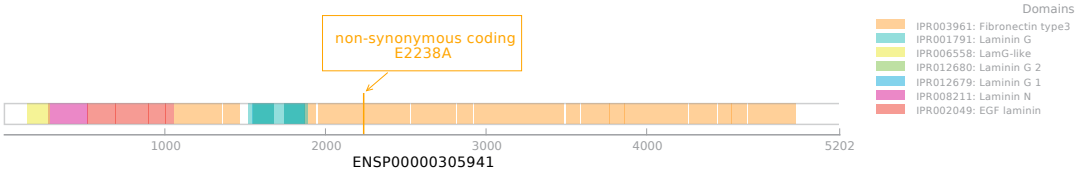
Variant 24:	Gene: USH2A Your genotype: G/T Location: chr1:215914751	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0078	dbSNP: rs41303285
Quality:	Genotype quality: 99.00	Coverage depth: 181
Details:	Gene description: Usher syndrome 2A (autosomal recessive, mild)	
	Transcript: ENST00000307340	AA change: P3893T
	EntrezId: 7399	EnsemblId: ENSG00000042781
	UniProt: O75445	OMIM: 608400



Variant 25:	Gene: TTN Your genotype: T/G Location: chr2:179477267	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0083	dbSNP: rs36043230
Quality:	Genotype quality: 99.00	Coverage depth: 180
Details:	Gene description: titin Transcript: ENST00000342992 EntrezId: 7273 UniProt: NA	AA change: N14094T EnsemblId: ENSG00000155657 OMIM: 188840



Variant 26:	Gene: USH2A Your genotype: T/G Location: chr1:216166454	
Effect:	NON-SYNONYMOUS CODING	Type:MODERATE
Frequency:	1KGenomes: 0.0097	dbSNP: rs41277212
Quality:	Genotype quality: 99.00	Coverage depth: 198
Details:	Gene description: Usher syndrome 2A (autosomal recessive, mild) Transcript: ENST00000307340 EntrezId: 7399 UniProt: O75445	AA change: E2238A EnsemblId: ENSG00000042781 OMIM: 608400



Variant 27:	**Gene:** KRT85 **Your genotype:** C/T **Location:** chr12:52760957	
dbSNP:	**Annotation Present**	
Frequency:	**1KGenomes:** 0.0266	**dbSNP:** rs61630004
Quality:	**Genotype quality:** 99.00	**Coverage depth:** 116
Details:	**Gene description:** keratin 85 **Transcript:** ENST00000257901	
	EntrezId: 3891	**EnsemblId:** ENSG00000135443
	UniProt: P78386	**OMIM:** 602767

Variant 28:	**Gene:** TPMT **Your genotype:** C/T **Location:** chr6:18139228	
dbSNP:	**Annotation Present**	
Frequency:	**1KGenomes:** 0.0174	**dbSNP:** rs1800460
Quality:	**Genotype quality:** 99.00	**Coverage depth:** 146
Details:	**Gene description:** thiopurine S-methyltransferase **Transcript:** ENST00000309983	
	EntrezId: 7172	**EnsemblId:** ENSG00000137364
	UniProt: P51580	**OMIM:** 187680

Variant 29:	**Gene:** TPMT **Your genotype:** T/C **Location:** chr6:18130918	
dbSNP:	**Annotation Present**	
Frequency:	**1KGenomes:** 0.0462	**dbSNP:** rs1142345
Quality:	**Genotype quality:** 99.00	**Coverage depth:** 98
Details:	**Gene description:** thiopurine S-methyltransferase **Transcript:** ENST00000309983	
	EntrezId: 7172	**EnsemblId:** ENSG00000137364
	UniProt: P51580	**OMIM:** 187680

Variant 30:	**Gene:** CRELD1 **Your genotype:** C/T **Location:** chr3:9985136	
dbSNP:	**Annotation Present**	
Frequency:	**1KGenomes:** 0.0005	**dbSNP:** rs28942091
Quality:	**Genotype quality:** 99.00	**Coverage depth:** 169
Details:	**Gene description:** cysteine-rich with EGF-like domains 1 **Transcript:** ENST00000326434	
	EntrezId: 78987	**EnsemblId:** ENSG00000163703
	UniProt: NA	**OMIM:** 607170

Variant 31:	**Gene:** SPTA1 **Your genotype:** G/T **Location:** chr1:158624528	
dbSNP:	**Annotation Present**	
Frequency:	**1KGenomes:** 0.0177	**dbSNP:** rs35948326
Quality:	**Genotype quality:** 99.00	**Coverage depth:** 103
Details:	**Gene description:** spectrin, alpha, erythrocytic 1 (elliptocytosis 2) **Transcript:** ENST00000368147	
	EntrezId: 6708	**EnsemblId:** ENSG00000163554
	UniProt: NA	**OMIM:** 182860

Variant 32:	**Gene:** IL23R **Your genotype:** G/A **Location:** chr1:67705958	
dbSNP:	**Annotation Present**	
Frequency:	**1KGenomes:** 0.0329	**dbSNP:** rs11209026
Quality:	**Genotype quality:** 99.00	**Coverage depth:** 71
Details:	**Gene description:** interleukin 23 receptor **Transcript:** ENST00000395227	
	EntrezId: 149233	**EnsemblId:** ENSG00000162594
	UniProt: NA	**OMIM:** 607562

Appendix

To create the final draft of your exome we added some additional steps from the Broad Institute's "Best Practice" protocol aimed at increasing both the sensitivity and specificity of the variant calls returned to you. In the description that follows, steps 1–5 are unchanged from your first report:

1. We took your raw reads and aligned them against the reference genome (these are the alignments available in the BAM file of the first 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. At this point the protocol begins to differ from that used to generate the first draft of your exome. We now generate allele calls for all exome pilot participants simultaneously. By integrating data from multiple individuals we can more accurately detect i) variants in low coverage regions and ii) signatures of technical artifacts that might lead to incorrect variant calls. In addition we generate a BED file of all confidently called positions in the genome, which can be used in conjunction with the VCF file to determine where you are likely to be homozygous for the allele represented in the reference genome.
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 a technique developed at the Broad Institute known as Variant Quality Score Recalibration (VQSR). On top of this we applied the following cutoffs: i) $GQ \geq 30$, ii) $DP \geq 10$, iii) variant not on one of the sex-chromosomes. Variants that pass all filters are marked in your VCF file with a PASS, those that fail a filter are marked with the filters that they failed.
8. We then use snpEff to predict the functional impact of each variant on each gene that it may affect. Note that due to the existence of alternative transcription start/end points and alternative splicing a variant can have different effects on different products of the same gene. To simplify analysis we used GATK to select the highest-impact effect for each variant (see here for details).
9. We also annotate each variant with its allele frequency in the 1000 Genome's Project if available.