

CLINICAL APPLICATION OF WHOLE EXOME SEQUENCING FOR RARE DISEASES DIAGNOSTICS

TUGCE BOZKURT, M.Sc.

tugce.bozkurt@live.acibadem.edu.tr

OUTLINE

- Rare Diseases
- Whole Exome Sequencing (WES)
 - WES Data Analysis
 - Variant Prioritization Strategies
- A real-life case
- Conclusion

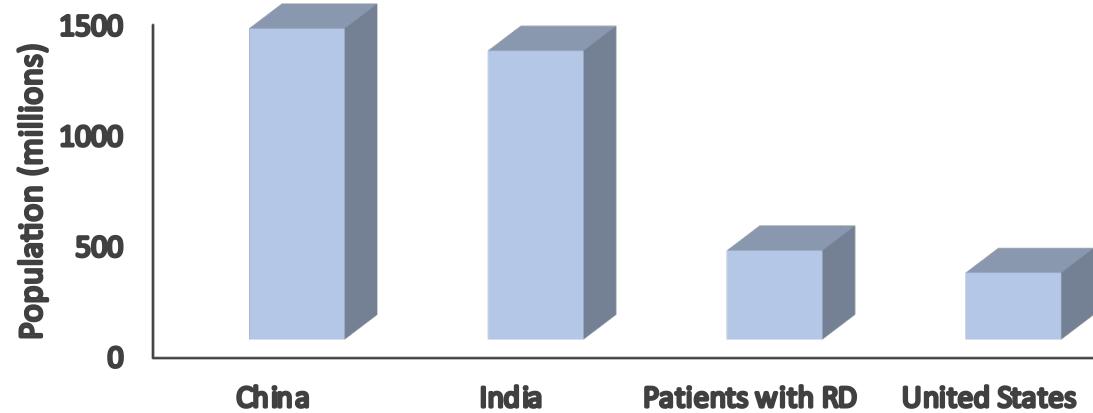
RARE DISEASES

- Rare diseases (RDs) are generally defined as a health condition that affects a small proportion of individuals in a population. The prevalence definition per 100 000;



RARE DISEASES

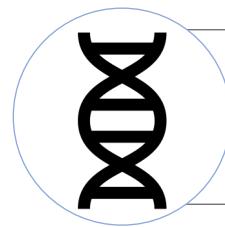
- According to World Health Organisation (WHO) report, it is estimated that nearly 400 million people suffer from RDs on a global scale.



It means that if all the people with rare diseases lived in one country, it would be the 3rd most crowded country in the world

RARE DISEASES

>6,000
distinct
phenotypes



71.9% of them arise from a genetic origin



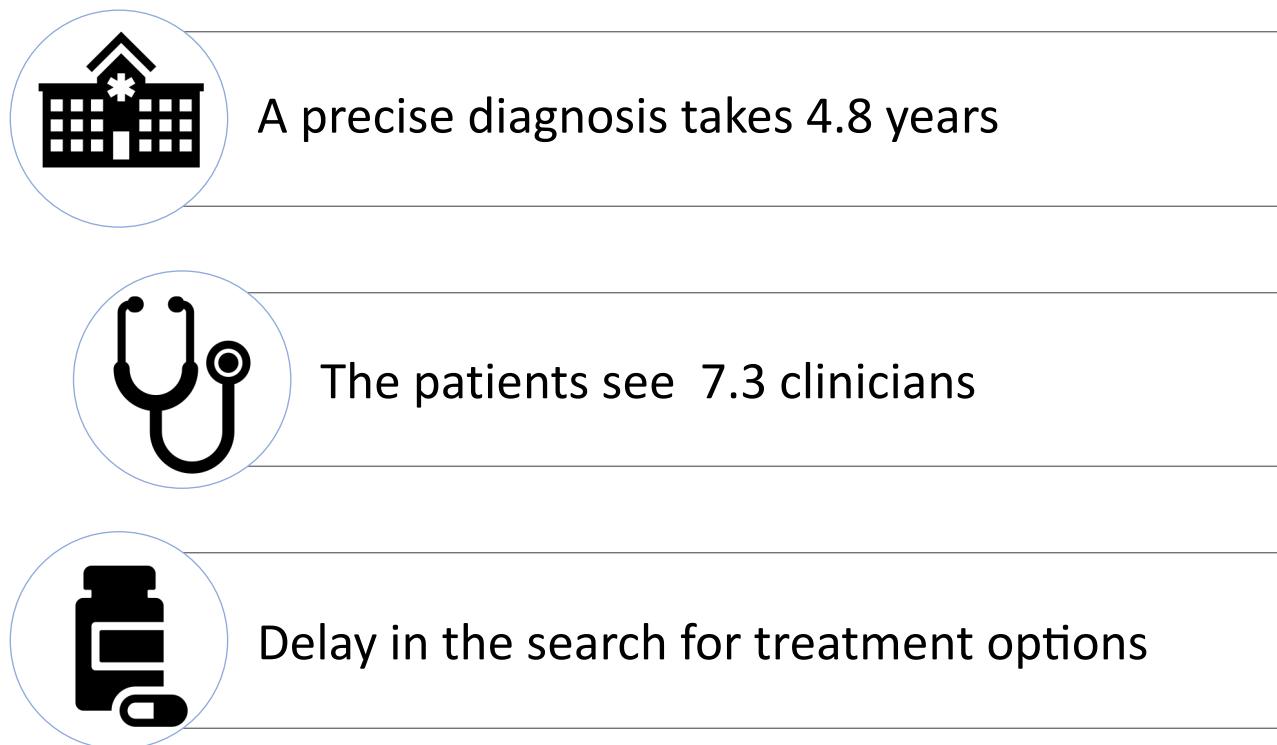
~30% of them still lack diagnostic definitions



69.9% of them are pediatric onset

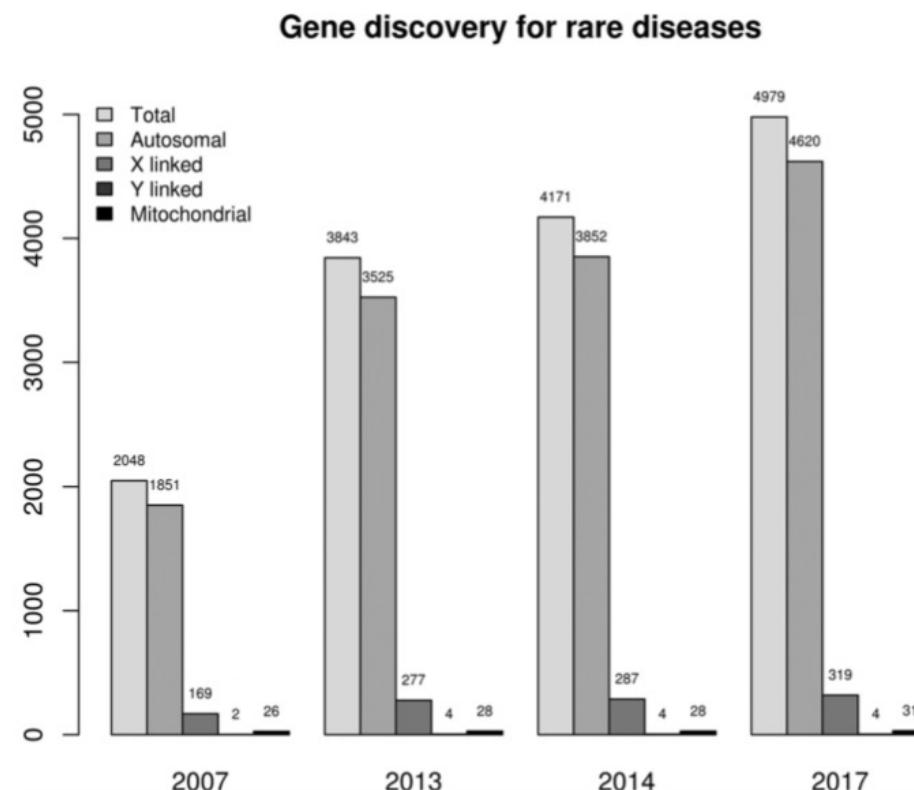
RARE DISEASES

The rarity of conditions
+
Heterogeneity
+
Incomplete knowledge
↓
Challenges in the diagnosis



RARE DISEASES

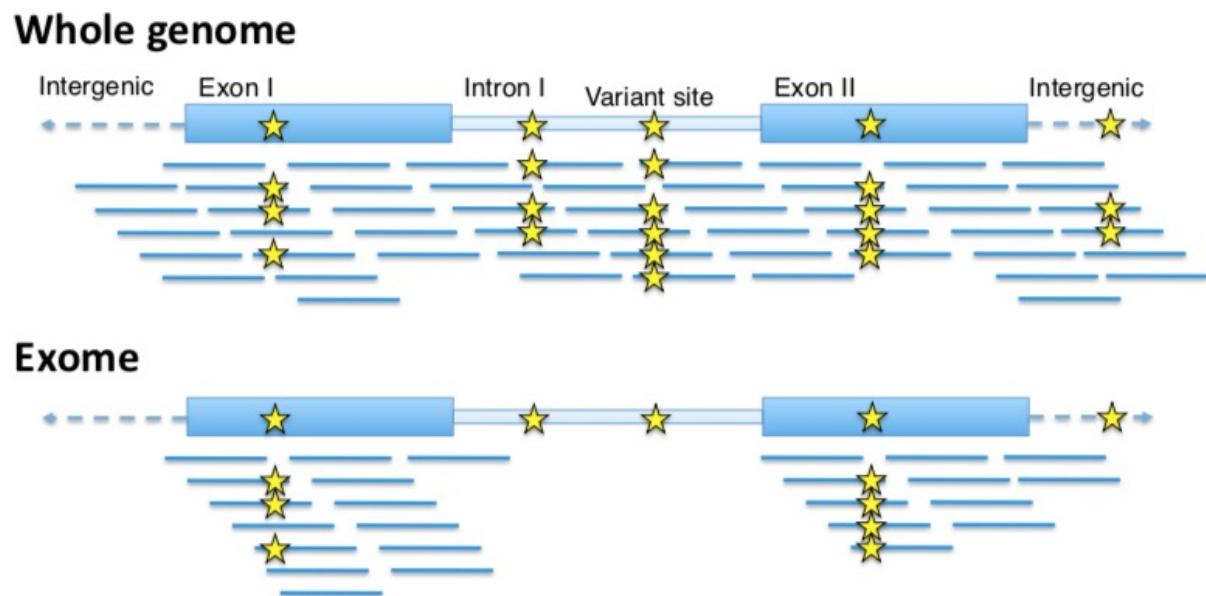
- Next Generation Sequencing-based approaches have become precise and unbiased ways to detect disease-causing variations.



Fernández-Marmiesse A. et.al, *Current Medicinal Chemistry*, 2018

WES is one of the NGS methods that feasible approach for identifying variants underlying rare Mendelian disorders.

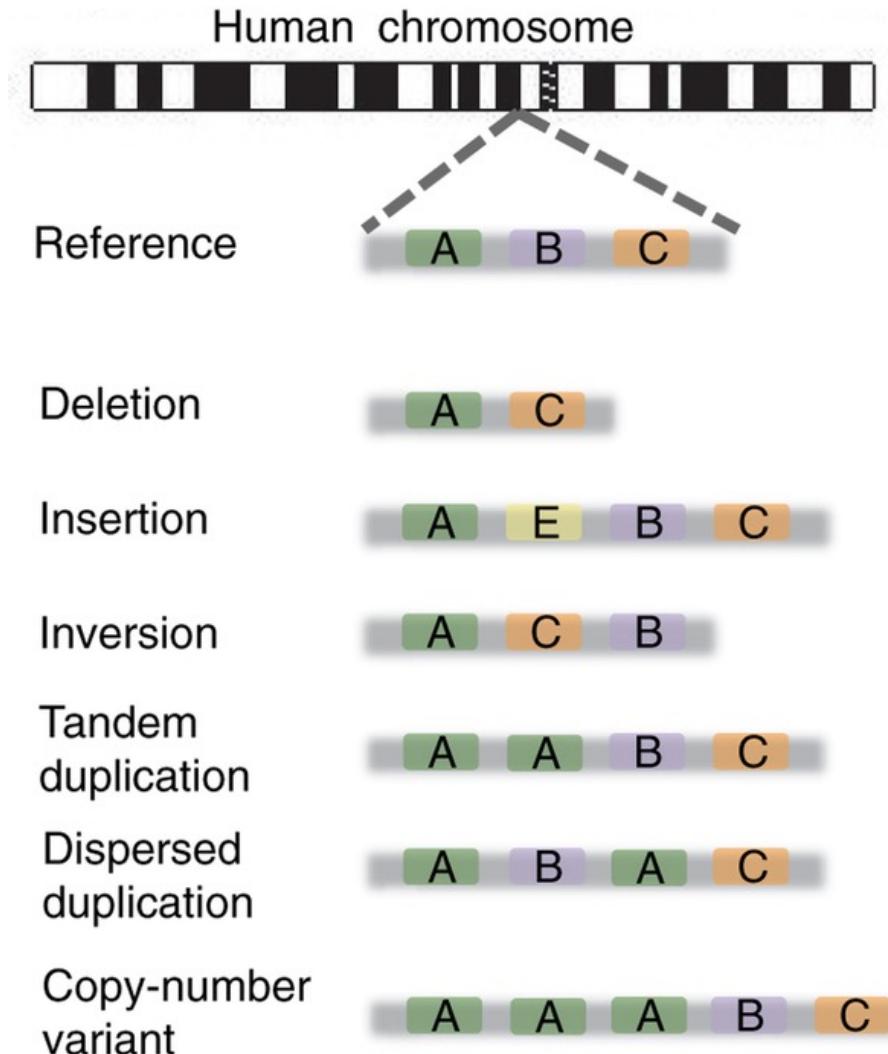
- since most variants that are known to cause these disorders disrupt protein-coding sequences



Terminology

TYPES OF VARIANTS

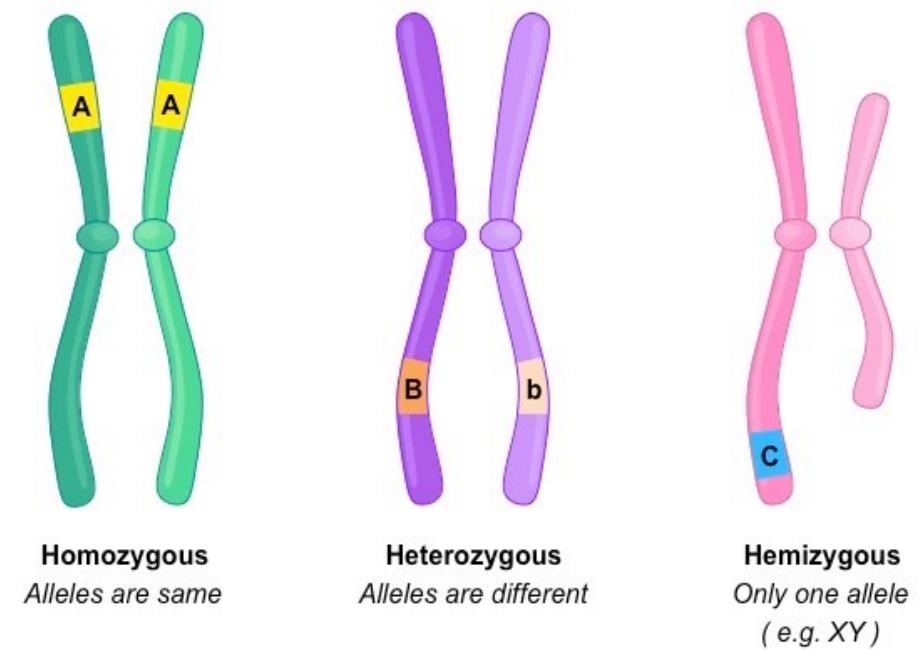
- SNV/SNP (SINGLE NUCLEOTIDE VARIANTS/POLYMORPHISMS)
- STRUCTURAL VARIANTS (>50 bp):
 - DELETION
 - INSERTION
 - INVERSION
 - TRANSLOCATION
 - DUPLICATION
 - CNV(COPY NUMBER VARIANT)



Terminology

VARIANT GENOTYPE

- When a person has a pair of identical alleles at a locus encoded in DNA → **homozygous**
- When the alleles are different and one of the alleles is the wild-type allele → **heterozygous**,
- The term **compound heterozygous** is used to describe a genotype in which two different mutant alleles of a gene are present
- In the special case in which a male has an abnormal allele for a gene located on the X chromosome and there is no other copy of the gene → **hemizygous**.



Terminology

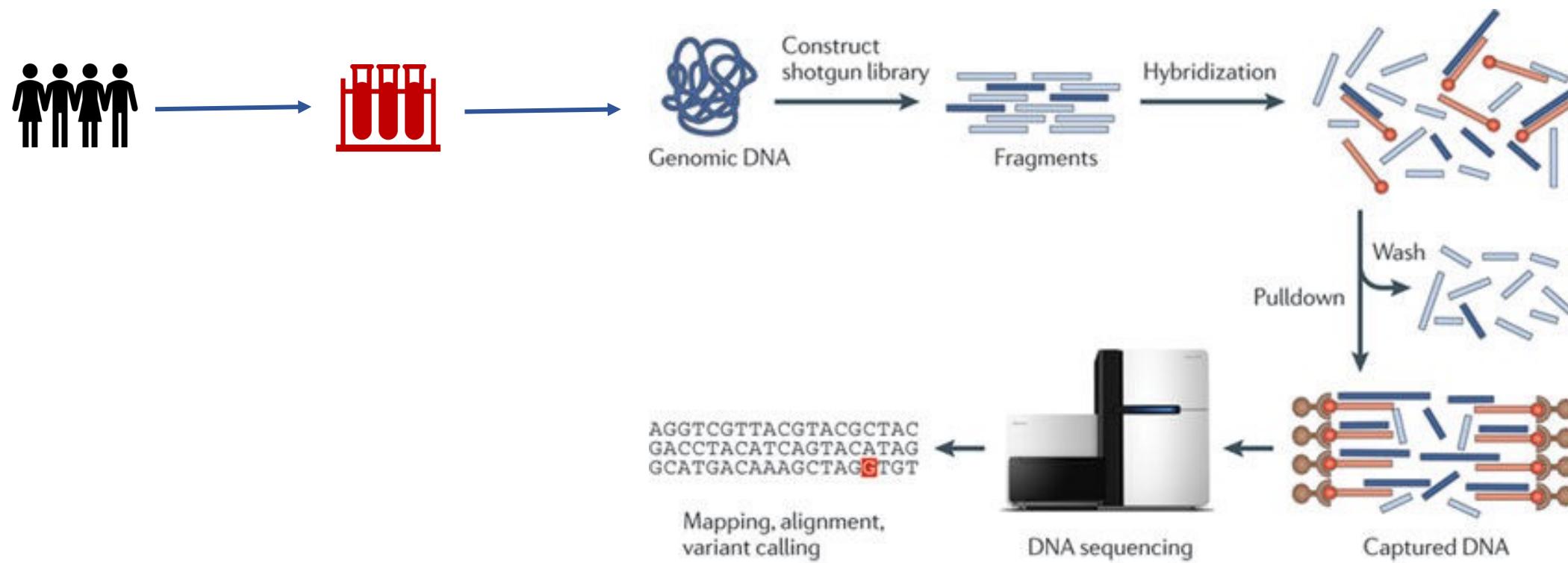
GENOTYPE

Homozygous: - Reference/Reference,
- Variant/Variant (Variant Homozygous)

Heterozygous: - Reference/Variant (Variant Heterozygous),
- Variant A /Variant B (Compound Heterozygous)

WHOLE EXOME SEQUENCING (WES)

The overview of WES



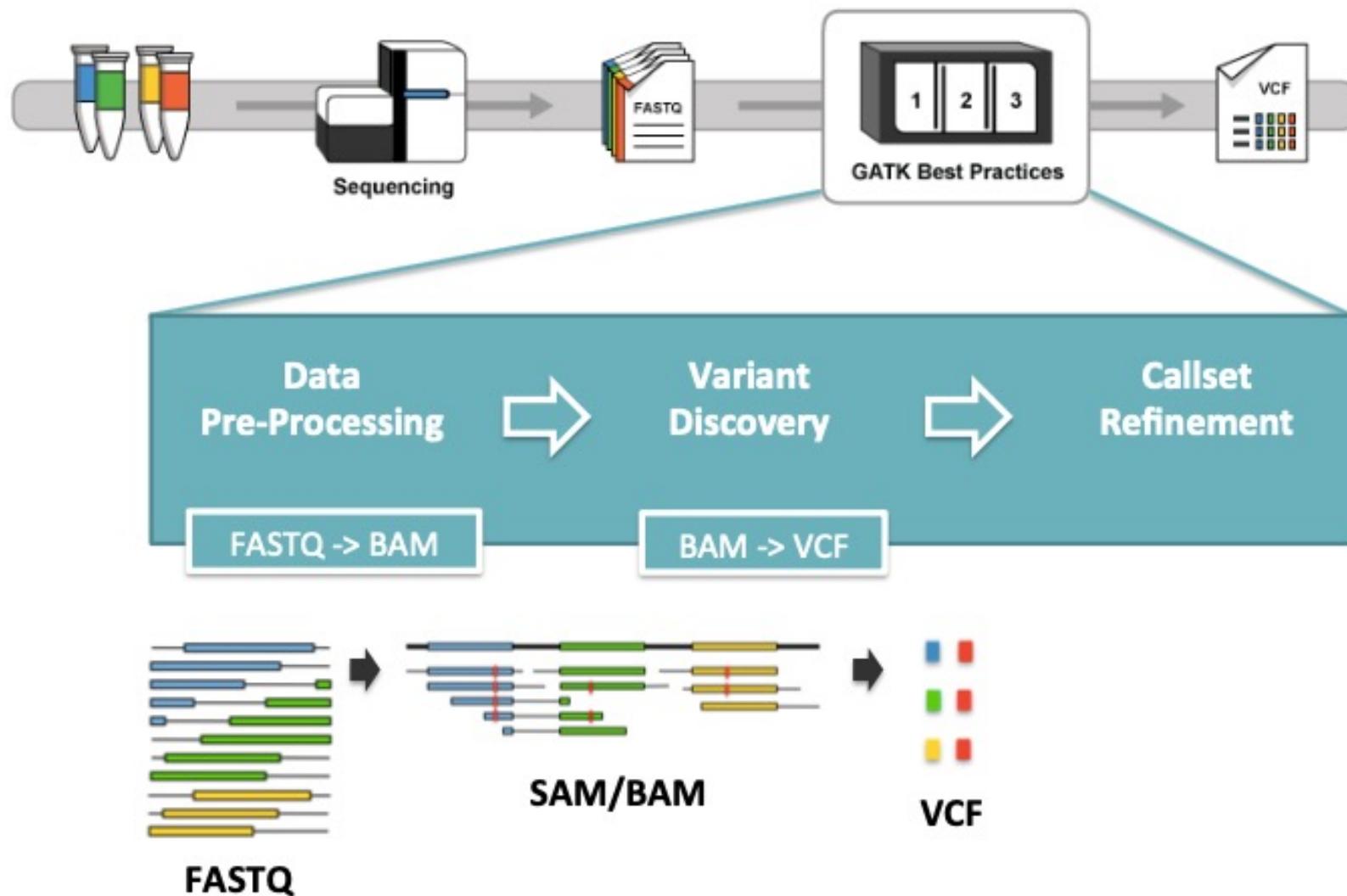
Nature Reviews | Genetics

Bamshad MJ et al. Exome sequencing as a tool for Mendelian disease gene discovery.
Nat Rev Genet. 2011 Nov;12(11):745–55.

You can also watch the following video...

<https://www.youtube.com/watch?v=CZeN-lgjYCo>

File Formats for Sequence Data



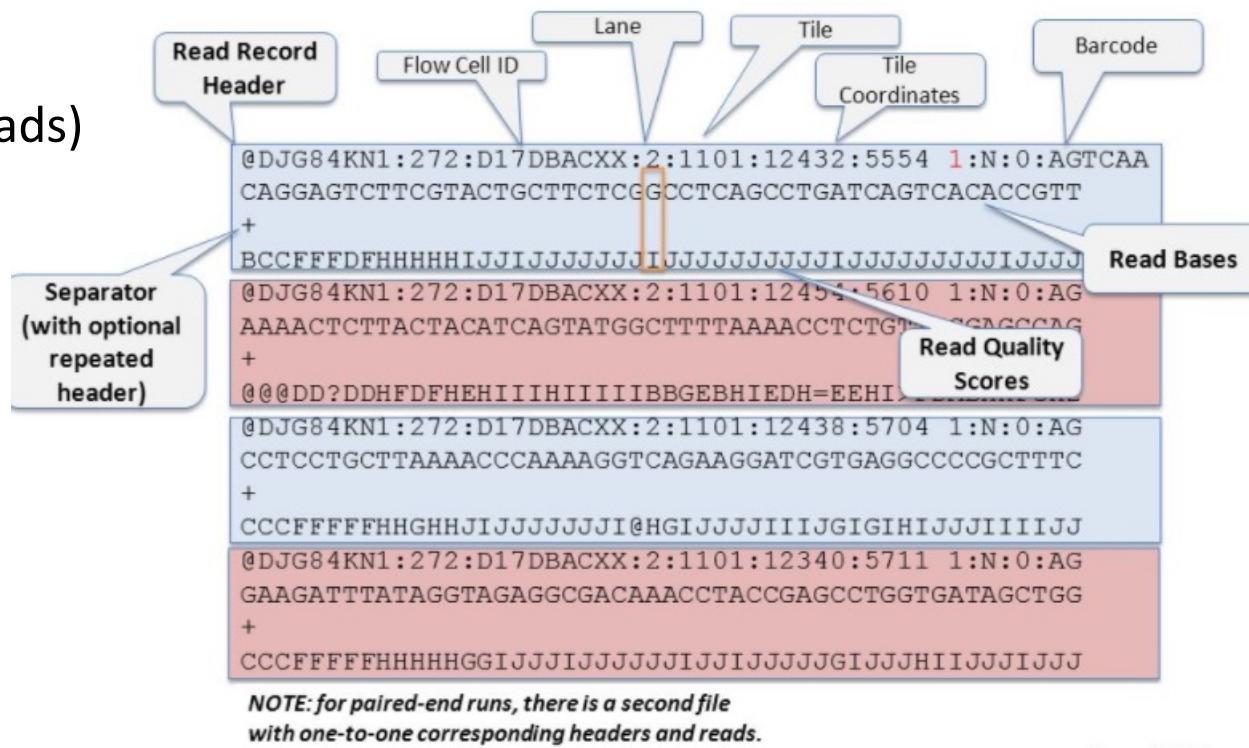
WES Data Analysis



- **Pre-processing - Quality Check**

- FASTQ is widely accepted as the standard file format for NGS raw data.
 - FASTQ = FASTA + base quality scores

FASTQ Format (Illumina Example)



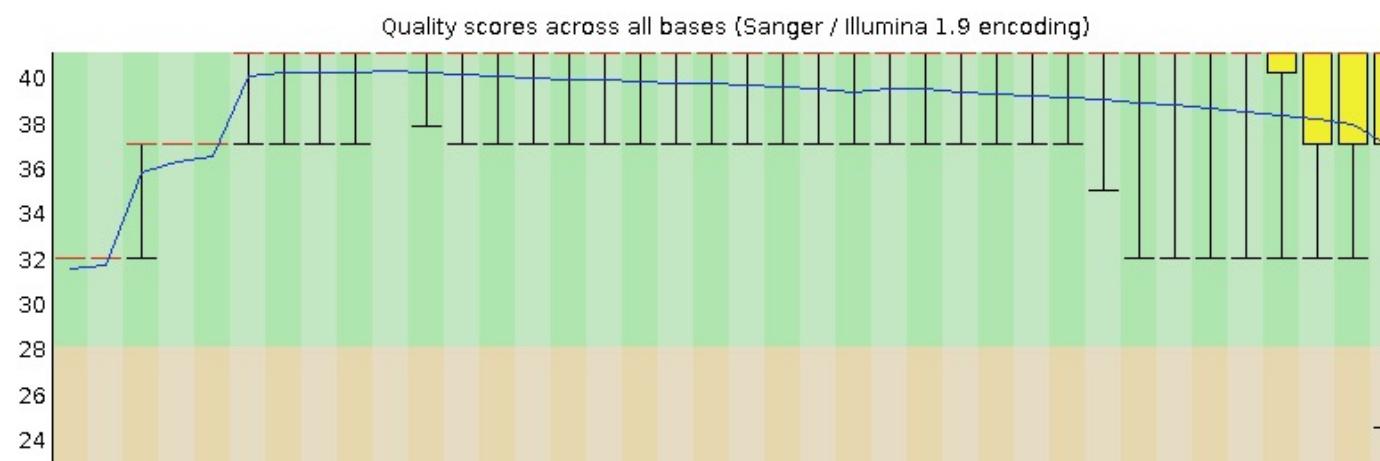
Summary

- [Basic Statistics](#)
- [Per base sequence quality](#)
- [Per tile sequence quality](#)
- [Per sequence quality scores](#)
- [Per base sequence content](#)
- [Per sequence GC content](#)
- [Per base N content](#)
- [Sequence Length Distribution](#)
- [Sequence Duplication Levels](#)
- [Overrepresented sequences](#)
- [Adapter Content](#)

Basic Statistics

| Measure | Value |
|-----------------------------------|-------------------------|
| Filename | 100860_R1.fastq |
| File type | Conventional base calls |
| Encoding | Sanger / Illumina 1.9 |
| Total Sequences | 361077795 |
| Sequences flagged as poor quality | 0 |
| Sequence length | 151 |
| %GC | 40 |

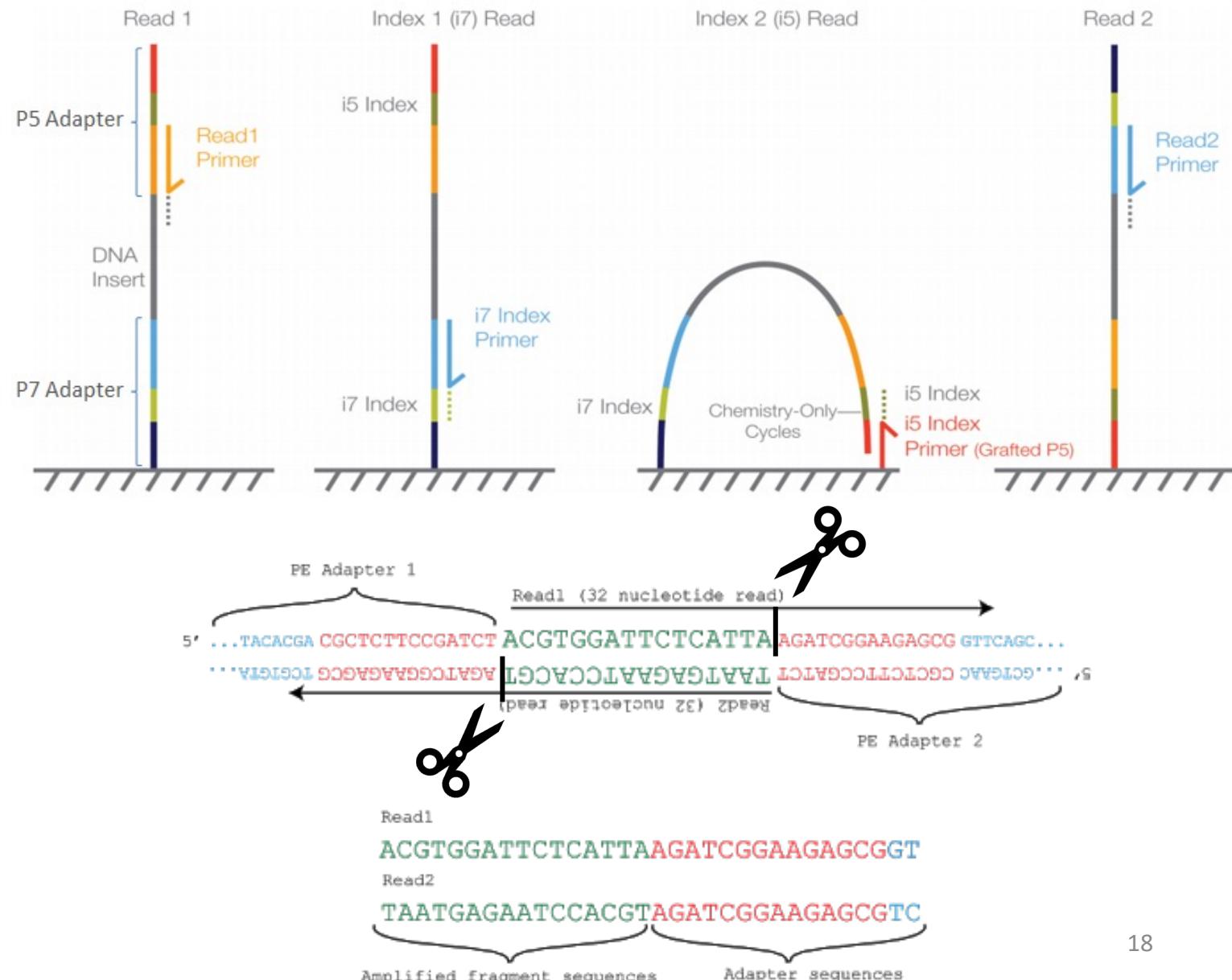
Per base sequence quality



WES Data Analysis

- **Pre-processing - Trimming**

- Standard trimming procedure includes 3' end adapter removal and trimming of low-quality bases at the ends of the reads.
- The adapters contain the sequencing primer binding sites, the index sequences, and the sites that allow library fragments to attach to the flow cell lawn



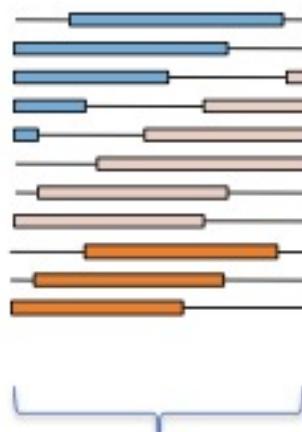
WES Data Analysis



WES Data Analysis

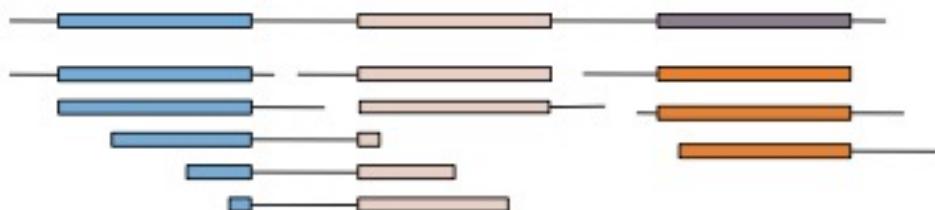
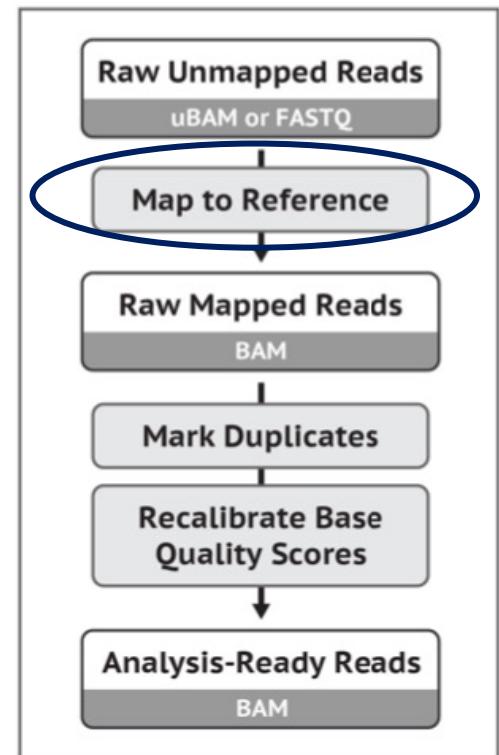
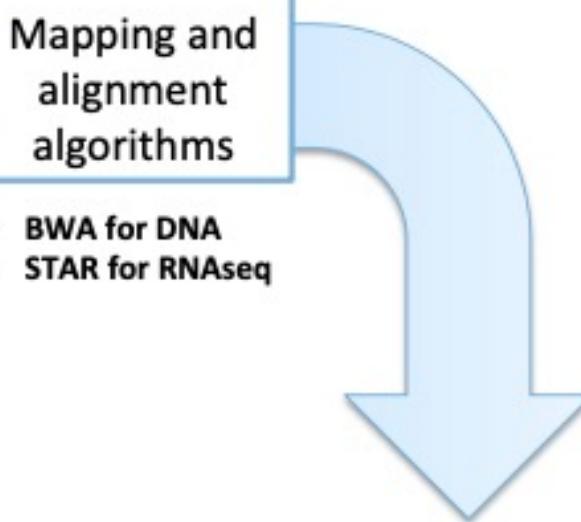
- **Alignment**

- Map the reads to the reference genome and with high efficiency and accuracy.
- Alignment mapping is a classical “string match”

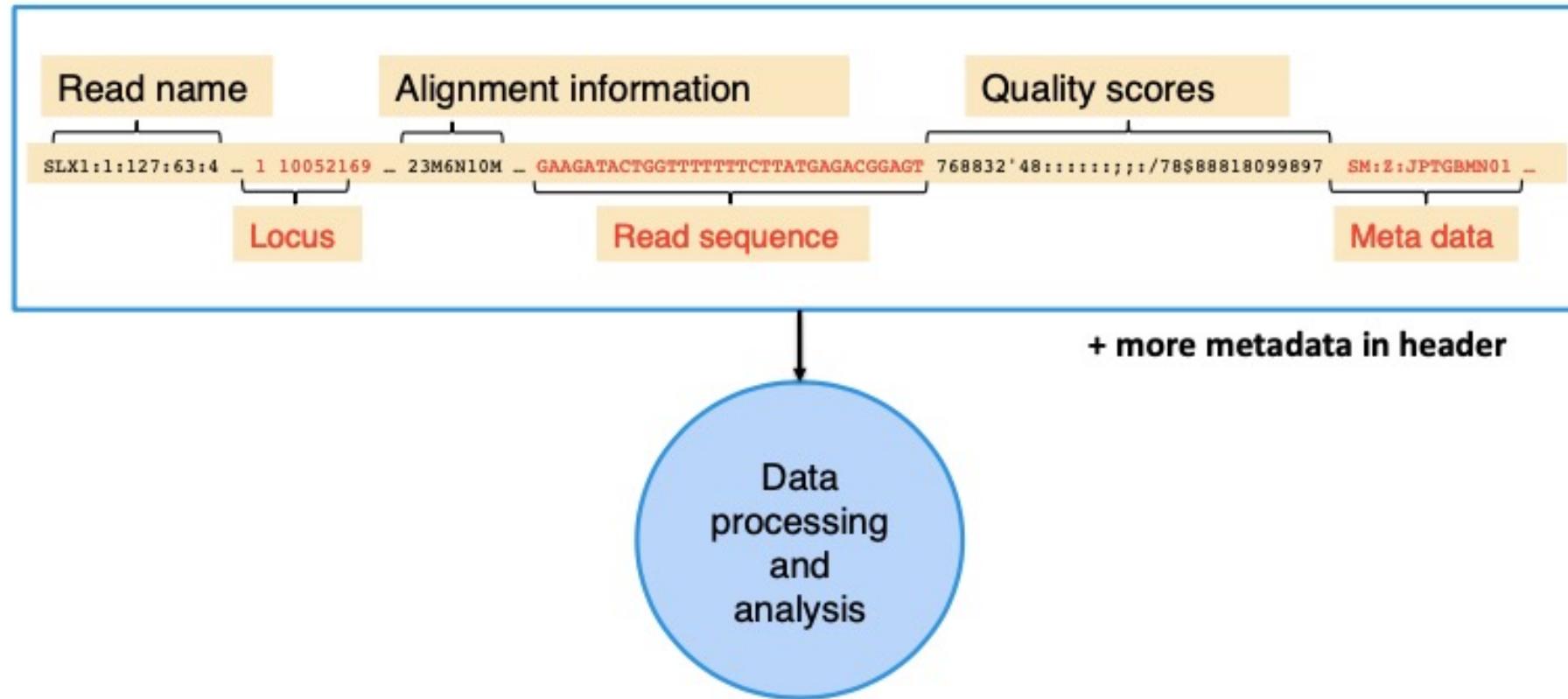


Mapping and
alignment
algorithms

- BWA for DNA
- STAR for RNAseq



The Output of Alignment: Sequence/Binary Alignment Map (SAM/BAM)



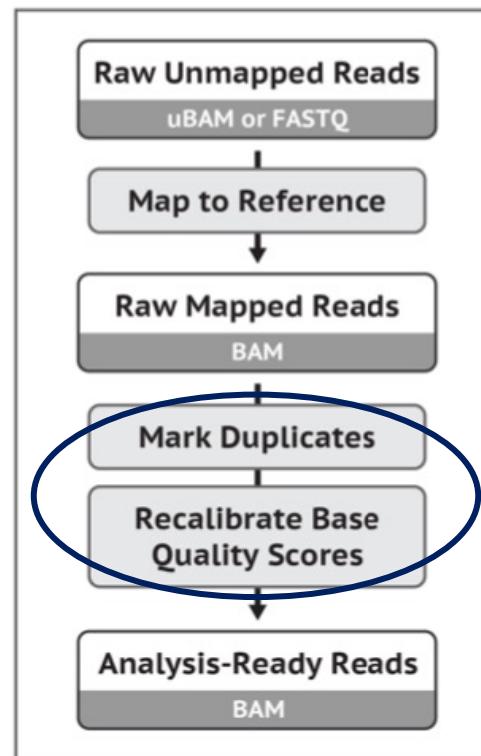
A BAM file can contain data from a single or from several samples

WES Data Analysis



- **Post-alignment Processing**

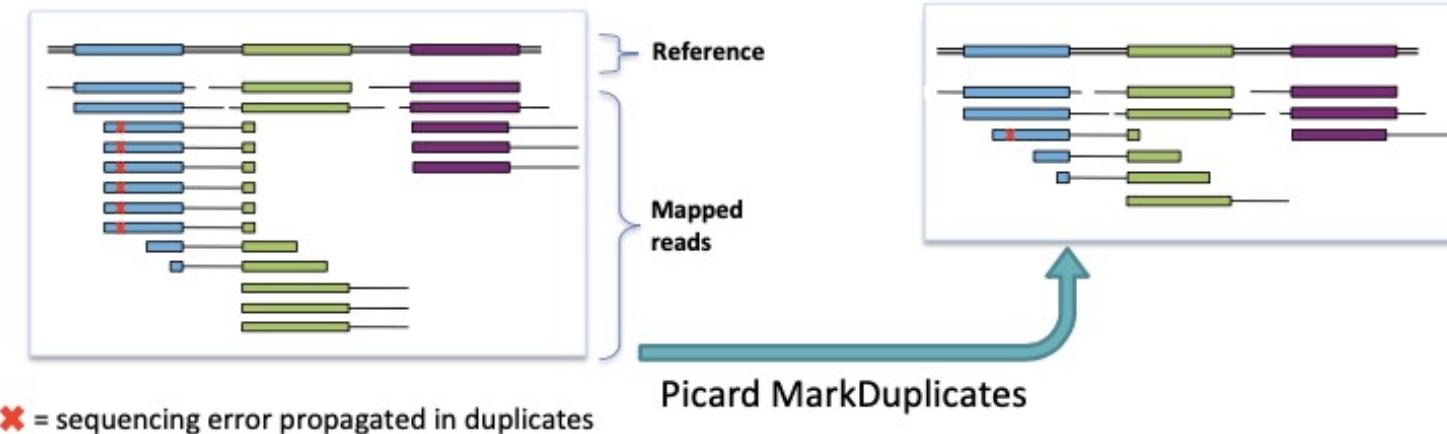
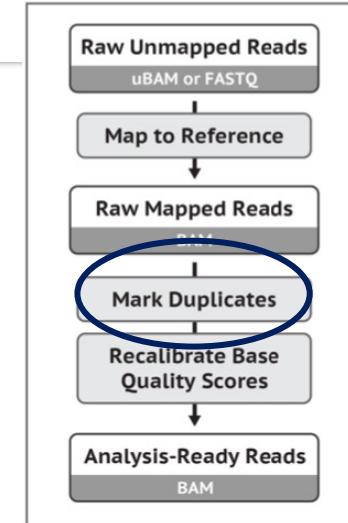
After mapping reads to the reference genome, some post-alignment processing procedure is recommended to minimize the artifacts that may affect the quality of downstream variant calling. It generally consists of read duplicate removal, and base quality score recalibration (BQSR).



WES Data Analysis

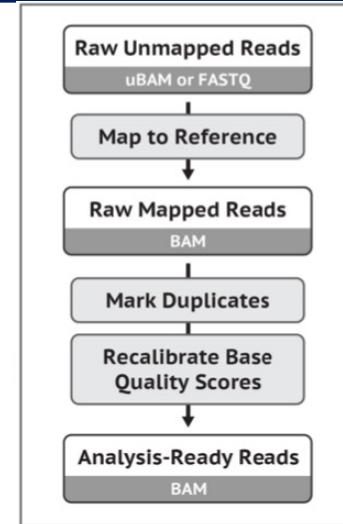
- Post-alignment Processing – Marking Duplicates

- Duplicates = **non-independent measurements** of a sequence
 - Sampled from same template of DNA
 - Violates assumptions of variant calling
 - Errors in sample/library prep will get propagated to *all* the duplicates
- > Just pick the “best” copy – mitigates the effects of errors



WES Data Analysis

- **Post-alignment Processing – Base Quality Score Recalibration (BQSR)**
 - Sequencers make systematic errors in base quality scores
 - BQSR corrects the quality scores (not the bases)

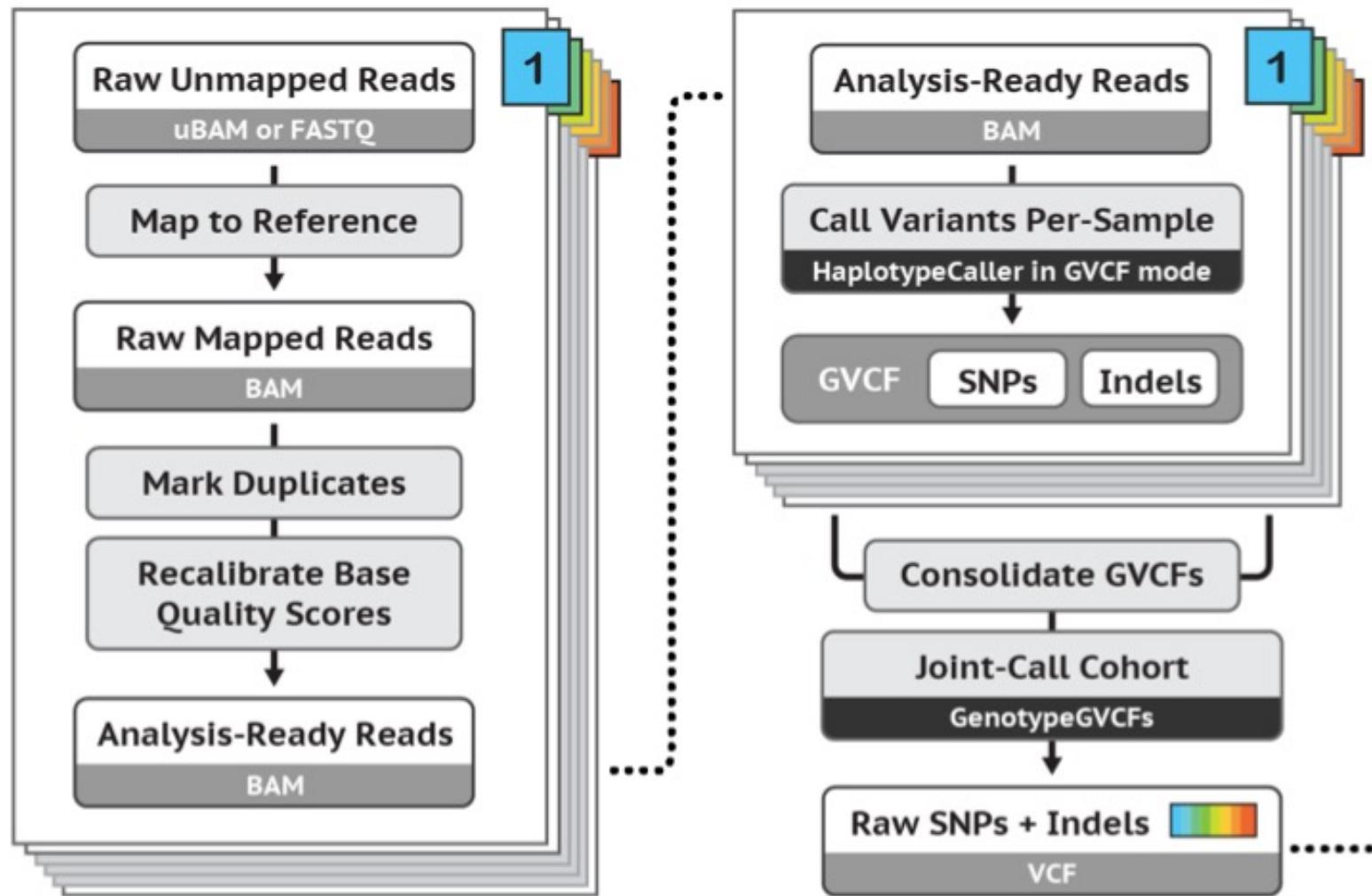


WES Data Analysis



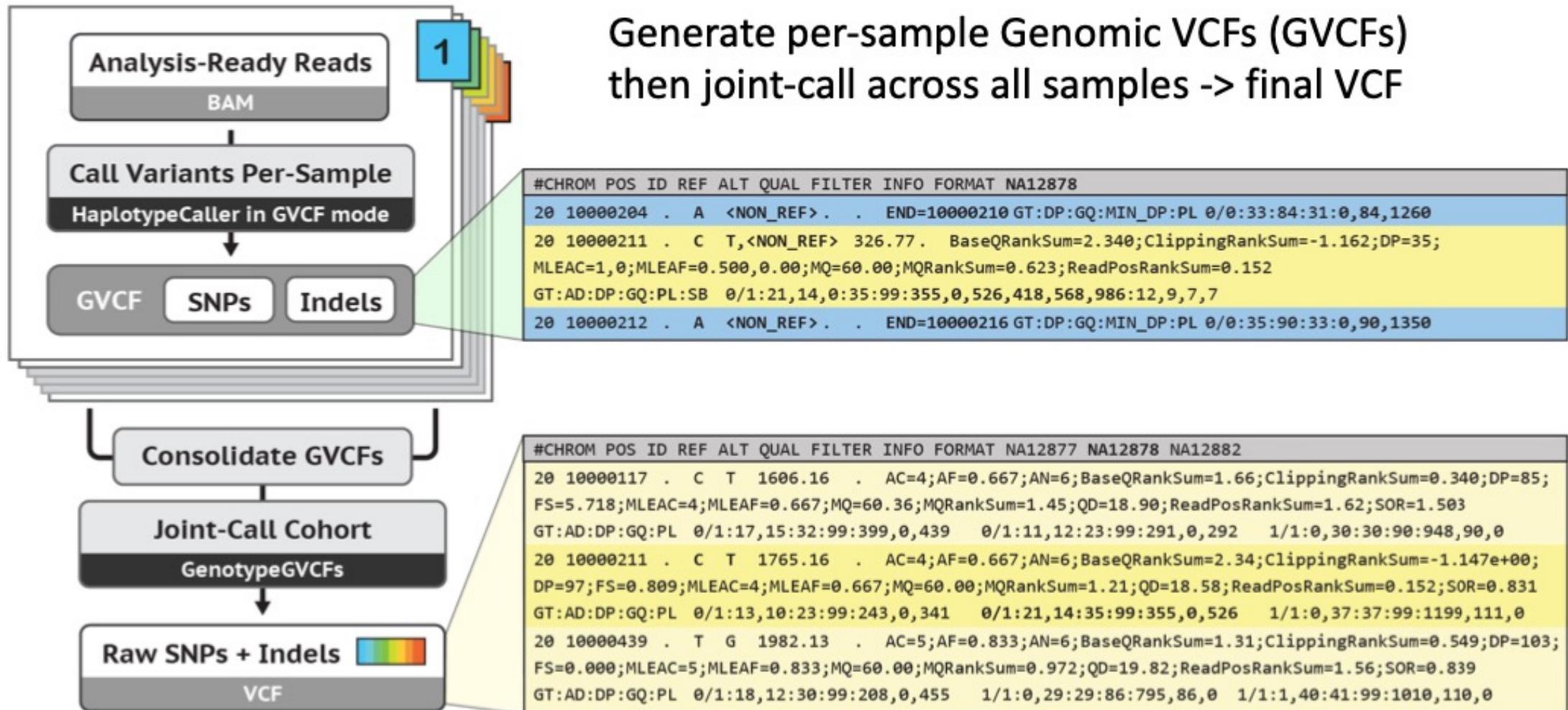
WES Data Analysis

- Germline Variant Calling



WES Data Analysis

- Germline Variant Calling



WES Data Analysis

```
##fileformat=VCFv4.2
##ALT=<ID=NON_REF,Description="Represents any possible alternative allele at this location">
##FILTER=<ID=LowQual,Description="Low quality">
##FORMAT=<ID=AD,Number=R,Type=Integer,Description="Allelic depths for the ref and alt alleles in the order listed">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth (reads with MQ=255 or with bad mates are filtered)">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=MIN_DP,Number=1,Type=Integer,Description="Minimum DP observed within the GVCF block">
##FORMAT=<ID=PGT,Number=1,Type=String,Description="Physical phasing haplotype information, describing how the alternate alleles are phased in relation to one another">
##FORMAT=<ID=PID,Number=1,Type=String,Description="Physical phasing ID information, where each unique ID within a given sample (but not across samples) connects records within a phasing group">
##FORMAT=<ID=PL,Number=G,Type=Integer,Description="Normalized, Phred-scaled likelihoods for genotypes as defined in the VCF specification">
##FORMAT=<ID=PS,Number=1,Type=Integer,Description="Phasing set (typically the position of the first variant in the set)">
##FORMAT=<ID=RGQ,Number=1,Type=Integer,Description="Unconditional reference genotype confidence, encoded as a phred quality -10*log10 p(genotype call is wrong)">
##FORMAT=<ID=SB,Number=4,Type=Integer,Description="Per-sample component statistics which comprise the Fisher's Exact Test to detect strand bias.">
##GATKCommandLine=<ID=CombineGVCFs,CommandLine="CombineGVCFs --output AD1212_vs_AD1213_vs_AD1211.g.vcf.gz --variant /runspace/data/gungor/AD1212/variant_calling/AD1212.g.vcf.gz --variant /runspace/data/gungor/AD1213/variant_calling/AD1213.g.vcf.gz --variant /runspace/data/gungor/AD1211/variant_calling/AD1211.g.vcf.gz"
##GATKCommandLine=<ID=GenotypeGVCFs,CommandLine="GenotypeGVCFs --output AD1212_vs_AD1213_vs_AD1211.vcf.gz --variant AD1212_vs_AD1213_vs_AD1211.g.vcf.gz --intervals /home/resources/kits/hg19/SeqCap_EZ_Exome_v3_hg19_capture_targets.bed --interval-padding 100"
##GATKCommandLine=<ID=HaplotypeCaller,CommandLine="HaplotypeCaller --emit-ref-confidence GVCF --output AD1212.g.vcf.gz --intervals /home/resources/kits/hg19/SeqCap_EZ_Exome_v3_hg19_capture_targets.bed --interval-padding 100 --input /runspace/data/gungor/AD1212/proc"
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes, for each ALT allele, in the same order as listed">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency, for each ALT allele, in the same order as listed">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##INFO=<ID=AS_BaseQRankSum,Number=A,Type=Float,Description="allele specific Z-score from Wilcoxon rank sum test of each Alt Vs. Ref base qualities">
##INFO=<ID=AS_FS,Number=A,Type=Float,Description="allele specific phred-scaled p-value using Fisher's exact test to detect strand bias of each alt allele">
##INFO=<ID=AS_InbreedingCoeff,Number=A,Type=Float,Description="allele specific heterozygosity as estimated from the genotype likelihoods per-sample when compared against the Hardy-Weinberg expectation; relate to inbreeding coefficient">
##INFO=<ID=AS_MQ,Number=A,Type=Float,Description="Allele-specific RMS Mapping Quality">
##INFO=<ID=AS_MQRankSum,Number=A,Type=Float,Description="Allele-specific Mapping Quality Rank Sum">
##INFO=<ID=AS_QD,Number=A,Type=Float,Description="Allele-specific Variant Confidence/Quality by Depth">
##INFO=<ID=AS_RAW_BaseQRankSum,Number=1,Type=String,Description="raw data for allele specific rank sum test of base qualities">
##INFO=<ID=AS_RAW_MQ,Number=1,Type=String,Description="Allele-specific raw data for RMS Mapping Quality">
##INFO=<ID=AS_RAW_MQRankSum,Number=1,Type=String,Description="Allele-specific raw data for Mapping Quality Rank Sum">
##INFO=<ID=AS_RAW_ReadPosRankSum,Number=1,Type=String,Description="allele specific raw data for rank sum test of read position bias">
##INFO=<ID=AS_ReadPosRankSum,Number=A,Type=Float,Description="allele specific Z-score from Wilcoxon rank sum test of each Alt vs. Ref read position bias">
##INFO=<ID=AS_SB_TABLE,Number=1,Type=String,Description="Allele-specific forward/reverse read counts for strand bias tests">
##INFO=<ID=AS_SOR,Number=A,Type=Float,Description="Allele specific strand Odds Ratio of 2x|Alts| contingency table to detect allele specific strand bias">
##INFO=<ID=BaseQRankSum,Number=1,Type=Float,Description="Z-score from Wilcoxon rank sum test of Alt Vs. Ref base qualities">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth; some reads may have been filtered">
##INFO=<ID=DS,Number=0,Type=Flag,Description="Were any of the samples downsampled?">
##INFO=<ID=END,Number=1,Type=Integer,Description="Stop position of the interval">
##INFO=<ID=ExcessHet,Number=1,Type=Float,Description="Phred-scaled p-value for exact test of excess heterozygosity">
##INFO=<ID=FS,Number=1,Type=Float,Description="Phred-scaled p-value using Fisher's exact test to detect strand bias">
##INFO=<ID=InbreedingCoeff,Number=1,Type=Float,Description="Inbreeding coefficient as estimated from the genotype likelihoods per-sample when compared against the Hardy-Weinberg expectation">
```

WES Data Analysis

- **Germline Variant Calling**

| #CHROM | POS | ID | REF | ALT | QUAL | FILTER | INFO | FORMAT | AD1211 | AD1212 | AD1213 |
|--------|-------|----|-----|-----|---------|--------|-------------|------------|------------------|-----------------|-------------------------------|
| chr1 | 14354 | . | C | A | 46.58 | . | AC=2;AF=0.3 | GT:AD:DP:G | 0/0:4,0:4:12 | 0/0:8,0:8:24 | 1/1:0,2:2:6:60,6,0 |
| chr1 | 14464 | . | A | T | 452.13 | . | AC=2;AF=0.3 | GT:AD:DP:G | 0/1:19,9:28:4 | 0/1:24,10:34 | 0/0:10,0:10:24:0,24,360 |
| chr1 | 14542 | . | A | G | 80.29 | . | AC=1;AF=0.1 | GT:AD:DP:G | 0/0:23,2:25:1 | 0/0:24,0:24: | 0/1:4,4:8:89:89,0,103 |
| chr1 | 14653 | . | C | T | 234.92 | . | AC=3;AF=0.5 | GT:AD:DP:G | 0/1:53,11:64:0 | 1/56,7:63:0 | 0/1:36,7:43:74:74,0,922 |
| chr1 | 14673 | . | G | C | 191.27 | . | AC=1;AF=0.1 | GT:AD:DP:G | 0/0:64,0:64:0 | 0/0:70,8:78:0 | 0/1:33,11:44:99:200,0,1018 |
| chr1 | 14677 | . | G | A | 190.06 | . | AC=2;AF=0.3 | GT:AD:DP:G | 0/1:16,6:8:74:0 | 1/0:76,0:76:0 | 0/1:32,12:44:99:189,0,967 |
| chr1 | 14907 | . | A | G | 4084.93 | . | AC=3;AF=0.5 | GT:AD:DP:G | 0/1:179,60:2 | 0/1:207,64:2 | 0/1:117,80:197:99:1857,0,3223 |
| chr1 | 14930 | . | A | G | 4880.93 | . | AC=3;AF=0.5 | GT:AD:DP:G | 0/1:187,54:2 | 0/1:216,65:2 | 0/1:118,90:208:99:2173,0,3515 |
| chr1 | 14976 | . | G | A | 89.25 | . | AC=1;AF=0.1 | GT:AD:DP:G | 0/0:181,0:18:0 | 1/175,23:1 | 0/0:63,0:63:99:0,120,1800 |
| chr1 | 15118 | . | A | G | 577.90 | . | AC=3;AF=0.5 | GT:AD:DP:G | 0/1:31,12:43 | 0/1:49,7:56:0 | 0/1:38,19:57:99:386,0,892 |
| chr1 | 15190 | . | G | A | 505.93 | . | AC=3;AF=0.5 | GT:AD:DP:G | 0/1:17,7:24:0 | 1/0:26,10:36 | 0/1:23,8:31:99:189,0,580 |
| chr1 | 15211 | . | T | G | 1555.93 | . | AC=3;AF=0.5 | GT:AD:DP:G | 0/1:6,13:19:0 | 1/0:8,25:33:0 | 1/0:15,24:29:86:618,0,86 |
| chr1 | 15688 | . | C | T | 147.89 | . | AC=3;AF=0.5 | GT:AD:DP:G | 0/1:12,4:16:0 | 1/0:18,3:21:0 | 1/0:8,4:12:69:69,0,258 |
| chr1 | 15903 | . | G | GC | 421.11 | . | AC=3;AF=0.7 | GT:AD:DP:G | 0/1:1,0:4,4:15:0 | 1/0:1,8:9:9:2 | ./.:1,0:1:0,0,0 |
| chr1 | 16487 | . | T | C | 150.35 | . | AC=1;AF=0.1 | GT:AD:DP:G | 0/0:61,8:69:0 | 1/0:47,11:58 | 0/0:44,0:44:99:0,99,1736 |
| chr1 | 16495 | . | G | C | 133.12 | . | AC=2;AF=0.3 | GT:AD:DP:G | 0/0:158,11:69 | 1/0:46,8:54:0 | 0/0:44,0:44:99:0,99,1736 |
| chr1 | 16497 | . | A | G | 689.93 | . | AC=3;AF=0.5 | GT:AD:DP:G | 0/0:152,16:68 | 1/0:33,20:53 | 0/0:35,8:43:95:95,0,815 |
| chr1 | 16534 | . | C | T | 2048.93 | . | AC=3;AF=0.5 | GT:AD:DP:G | 0/0:125,29:54 | 1/0:19,34:53 | 0/0:11,31:42:99:706,0,174 |
| chr1 | 16571 | . | G | A | 1731.93 | . | AC=3;AF=0.5 | GT:AD:DP:G | 0/0:121,24:45 | 1/0:20,28:48 | 0/0:7,24:31:97:550,0,97 |
| chr1 | 16977 | . | G | A | 1176.93 | . | AC=3;AF=0.5 | GT:AD:DP:G | 0/0:146,19:65 | 1/0:41,26:67 | 0/0:37,19:56:99:369,0,869 |
| chr1 | 17172 | . | G | A | 41.31 | . | AC=1;AF=0.1 | GT:AD:DP:G | 0/0:29,0:29:0 | 1/0:1,30,6:36:0 | 0/0:43,3:49:61:0,61,1110 |
| chr1 | 17385 | . | G | A | 55.25 | . | AC=1;AF=0.1 | GT:AD:DP:G | 0/0:64,0:64:0 | 1/0:36,0:36:0 | 0/0:74,9:83:64:64,0,1971 |
| chr1 | 17538 | . | C | A | 1926.93 | . | AC=3;AF=0.5 | GT:AD:DP:G | 0/0:170,33:10 | 1/0:1,65,25:90 | 0/0:60,23:83:99:570,0,1974 |
| chr1 | 17614 | . | G | A | 431.12 | . | AC=2;AF=0.3 | GT:AD:DP:G | 0/0:135,9:44:0 | 1/0:32,0:32:0 | 0/0:18,8:26:99:249,0,622 |
| chr1 | 17694 | . | C | T | 171.12 | . | AC=2;AF=0.3 | GT:AD:DP:G | 0/0:137,6:43:0 | 1/0:27,4:31:0 | 0/0:33,0:33:99:0,99,1307 |
| chr1 | 17697 | . | G | C | 38.25 | . | AC=1;AF=0.1 | GT:AD:DP:G | 0/0:38,2:40:0 | 1/0:26,5:31:0 | 0/0:33,0:33:99:0,99,1307 |

WES Data Analysis



WES Data Analysis

- Variant Annotation

So many annotation resources...

| 1 | gene | chrom | start | end | ref | alt | type | quality | depth | depth_of_alt_allele | allelic_balance | genotype | rs_ids | is_exonic | is_coding | is_lof | is_splicing | vep_hgvsc | vep_hgvsp |
|----|---------|-------|-----------|-----------|-----|-----|------|---------|-------|---------------------|-----------------|----------|---------|-----------|-----------|--------|-------------|--|-----------|
| 2 | TRIP11 | chr14 | 92441007 | 92441008 | G | A | snp | 99 | 133 | 73 | None | G/A | rs14125 | 1 | 1 | 0 | 0 | ENST000002 ENSP00000267622.4:p.Thr184Ile | |
| 3 | GLI2 | chr2 | 121748047 | 121748048 | G | A | snp | 99 | 212 | 108 | None | G/A | rs11481 | 1 | 1 | 0 | 0 | ENST000003 ENSP00000354586.4:p.Asp1520Asn | |
| 4 | RYR3 | chr15 | 34080627 | 34080628 | C | T | snp | 99 | 169 | 70 | None | C/T | rs18297 | 1 | 1 | 0 | 0 | ENST000003 ENSP00000373884.4:p.Pro3267Ser | |
| 5 | ZFYVE26 | chr14 | 68220860 | 68220861 | G | A | snp | 99 | 194 | 102 | None | G/A | rs15116 | 1 | 1 | 0 | 0 | ENST000003 ENSP00000251119.5:p.Thr2352Ile | |
| 6 | AGL | chr1 | 100349982 | 100349983 | C | T | snp | 99 | 162 | 77 | None | C/T | rs15044 | 1 | 1 | 0 | 0 | ENST000002 ENSP00000294724.4:p.Ser841Phe | |
| 7 | DOCK8 | chr9 | 304627 | 304628 | G | A | snp | 99 | 158 | 77 | None | G/A | rs14991 | 1 | 1 | 0 | 0 | ENST000004 ENSP00000394888.2:p.Arg83Gln | |
| 8 | SIL1 | chr5 | 138378393 | 138378394 | G | A | snp | 99 | 142 | 63 | None | G/A | rs11580 | 1 | 1 | 0 | 0 | ENST000002 ENSP00000265195.5:p.Thr123Ile | |
| 9 | CCDC39 | chr3 | 180379772 | 180379773 | C | T | snp | 99 | 130 | 59 | None | C/T | rs11595 | 1 | 1 | 0 | 0 | ENST000002 ENSP00000273654.4:p.Arg162His | |
| 10 | ANK2 | chr4 | 114288919 | 114288920 | C | A | snp | 99 | 200 | 107 | None | C/A | rs12191 | 1 | 1 | 0 | 0 | ENST000002 ENSP00000264366.6:p.Thr3711Asn | |
| 11 | CFTR | chr7 | 117188796 | 117188797 | A | G | snp | 99 | 213 | 40 | None | A/G | rs20143 | 1 | 1 | 0 | 0 | ENST000000 ENSP00000003084.6:p.Thr438Ala | |
| 12 | SDHC | chr1 | 161332300 | 161332301 | G | A | snp | 99 | 113 | 43 | None | G/A | rs18262 | 1 | 1 | 0 | 0 | ENST000003 ENSP00000356952.3:p.Ala142Thr | |
| 13 | MUTYH | chr1 | 45806052 | 45806053 | G | A | snp | 99 | 223 | 109 | None | G/A | rs32194 | 1 | 0 | 0 | 0 | ENST00000372098.3:c.-127C>T | |
| 14 | HBB | chr11 | 5248049 | 5248050 | C | T | snp | 99 | 123 | 45 | None | C/T | rs35004 | 0 | 0 | 0 | 0 | ENST00000335295.4:c.93-21G>A | |
| 15 | RAG2 | chr11 | 36614082 | 36614083 | A | T | snp | 99 | 57 | 37 | None | A/T | rs54697 | 1 | 0 | 0 | 0 | ENST00000311485.3:c.*52T>A | |
| 16 | TTN | chr2 | 179414126 | 179414127 | C | T | snp | 99 | 194 | 99 | None | C/T | rs75948 | 1 | 1 | 0 | 0 | ENST000003 ENSP00000340554.6:p.Arg21869%3D | |
| 17 | MARS | chr12 | 57909118 | 57909119 | A | C | snp | 99 | 104 | 57 | None | A/C | rs14057 | 0 | 0 | 0 | 1 | ENST000002 ENSP00000262027.5:p.Thr797%3D | |
| 18 | PLA2G6 | chr22 | 38508248 | 38508249 | G | A | snp | 99 | 137 | 64 | None | G/A | rs13868 | 1 | 1 | 0 | 0 | ENST000003 ENSP00000333142.3:p.Asn780%3D | |
| 19 | HSPG2 | chr1 | 22161379 | 22161380 | G | A | snp | 99 | 192 | 96 | None | G/A | rs55875 | 1 | 1 | 0 | 0 | ENST000003 ENSP00000363827.3:p.His3504%3D | |
| 20 | CACNB2 | chr10 | 18816564 | 18816565 | G | A | snp | 99 | 132 | 63 | None | G/A | rs76956 | 1 | 1 | 0 | 0 | ENST000002 ENSP00000282343.8:p.Ser303%3D | |
| 21 | SNTA1 | chr20 | 32000157 | 32000158 | G | A | snp | 99 | 150 | 78 | None | G/A | rs13886 | 1 | 1 | 0 | 0 | ENST000002 ENSP00000217381.2:p.Pro328%3D | |
| 22 | PYGM | chr11 | 64522239 | 64522240 | G | A | snp | 99 | 164 | 51 | None | G/A | rs13972 | 1 | 1 | 0 | 0 | ENST000001 ENSP00000164139.3:p.Ile308%3D | |
| 23 | TPO | chr2 | 1418209 | 1418210 | G | A | snp | 99 | 180 | 84 | None | G/A | rs28909 | 1 | 1 | 0 | 0 | ENST000003 ENSP00000329869.4:p.Thr10%3D | |
| 24 | CFTR | chr7 | 117188849 | 117188850 | G | T | snp | 99 | 156 | 40 | None | G/T | rs79074 | 1 | 1 | 0 | 0 | ENST000000 ENSP00000003084.6:p.Ala455%3D | |
| 25 | FER1L6 | chr8 | 125115419 | 125115420 | G | A | snp | 99 | 157 | 60 | None | G/A | rs56132 | 1 | 1 | 0 | 0 | ENST000003 ENSP00000381982.1:p.Arg1720Gln | |
| 26 | KCNJ12 | chr17 | 21319435 | 21319436 | G | A | snp | 99 | 718 | 150 | None | G/A | rs77270 | 1 | 1 | 0 | 0 | ENST000003 ENSP00000328150.5:p.Arg261His | |
| 27 | LIMS1 | chr2 | 109276098 | 109276099 | G | A | snp | 99 | 229 | 72 | None | G/A | rs74632 | 1 | 1 | 0 | 0 | ENST000003 ENSP00000331775.6:p.Arg12His | |
| 28 | KRR1 | chr12 | 75900381 | 75900382 | C | T | snp | 62 | 21 | 21 | None | T/T | rs11540 | 1 | 1 | 0 | 0 | ENST000002 ENSP00000229214.4:p.Arg134Gln | |
| 29 | SUGP1 | chr19 | 19413091 | 19413092 | C | T | snp | 99 | 118 | 118 | None | T/T | rs17751 | 1 | 1 | 0 | 0 | ENST000002 ENSP00000247001.3:p.Arg290His | |
| 30 | EFCC1 | chr3 | 128755952 | 128755953 | G | A | snp | 99 | 122 | 61 | None | G/A | rs37324 | 1 | 1 | 0 | 0 | ENST000004 ENSP00000414597.2:p.Arg91Gln | |

WES Data Analysis

- Variant Annotation

| 1 | codon_change | aa_change | biotype | impact_so | impact_severity | polyphen_pred | sift_pred | aaf_gnomad_all | aaf_1kg_all | aaf_exac_all | clinvar_sig | clinvar_disease_name | in_dbnsnp | in_omim | ACMG classification | InterVar Evidence | cadd_raw | cadd_scaled | gerp_bp |
|----|--------------|-----------|-----------|---------------|-----------------|-----------------|--------------|----------------|-------------|--------------|---|----------------------|-----------|------------------------|------------------------------|------------------------------|----------|-------------|---------|
| 2 | aCa/aTa | T/I | protein_c | missense_var | MED | probably_damagi | tolerated | 0,00111773 | 0,000599042 | 0,001104 | conflicting_int not_specified not_provide | 1 | TRUE | Uncertain significance | PVS1=0 PS=[0, 0, 0, 0, 0.323 | 24.2 | 4,82000 | | |
| 3 | Gat/Aat | D/N | protein_c | missense_var | MED | probably_damagi | deleterious | 0,0093532 | 0,00459265 | 0,01 | conflicting_int Holoprosencephaly_sequel | 1 | TRUE | Benign | PVS1=0 PS=[0, 0, 0, 0, 0.423 | 31.0 | 4,98000 | | |
| 4 | Ccc/Tcc | P/S | protein_c | missense_var | MED | probably_damagi | deleterious | 0,000313394 | 0,000199681 | 0,0001984 | conflicting_int Epileptic_encephalopathy | 1 | TRUE | Uncertain significance | PVS1=0 PS=[0, 0, 0, 0, 0.394 | 27.4 | 4,40000 | | |
| 5 | aCc/aTc | T/I | protein_c | missense_var | MED | benign | deleterious | 0,00293472 | 0,00179712 | 0,002553 | conflicting_int Spastic_paraplegia Spastic | 1 | TRUE | Benign | PVS1=0 PS=[0, 0, 0, 0, 0.277 | 23.1 | 4,92000 | | |
| 6 | tCt/tTt | S/F | protein_c | missense_var | MED | benign | tolerated | 0,00276834 | 0,00199681 | 0,002762 | conflicting_int Glycogen_storage_disease_ | 1 | TRUE | Likely benign | PVS1=0 PS=[0, 0, 0, 0, 0.292 | 23.4 | 5,80000 | | |
| 7 | cGa/cAa | R/Q | protein_c | missense_var | MED | benign | tolerated | 0,00139228 | 0,000998403 | 0,00145 | conflicting_int Hyperimmunoglobulin_E_r | 1 | TRUE | Benign | PVS1=0 PS=[0, 0, 0, 0, 0.282 | 23.2 | 6,01995 | | |
| 8 | aCc/aTc | T/I | protein_c | missense_var | MED | benign | tolerated | 0,00546435 | 0,00199681 | 0,005757 | conflicting_int Marinesco-Sjogren_syndrom | 1 | TRUE | Likely benign | PVS1=0 PS=[0, 0, 0, 0, 0.181 | 17.5 | 3,16000 | | |
| 9 | cGt/cAt | R/H | protein_c | missense_var | MED | benign | tolerated | 0,00722379 | 0,00599042 | 0,004795 | conflicting_int Ciliary_dyskinesia not_spec | 1 | TRUE | Uncertain significance | PVS1=0 PS=[0, 0, 0, 0, 0.134 | 15.09 | 2,63000 | | |
| 10 | aCt/aAt | T/N | protein_c | missense_var | MED | benign | tolerated_lo | 0,000625 | -1 | 0,0006177 | conflicting_int Arrhythmia Long QT_synd | 1 | TRUE | Uncertain significance | PVS1=0 PS=[0, 0, 0, 0, 0.16 | 16.26 | 2,97000 | | |
| 11 | Act/Gct | T/A | protein_c | missense_var | MED | benign | tolerated | 0,000206544 | -1 | 0,00248 | conflicting_int Cystic_fibrosis not_provide | 1 | TRUE | Benign | PVS1=0 PS=[0, 0, 0, 0, 0.067 | 10.84 | -6,6999 | | |
| 12 | Gcc/Acc | A/T | protein_c | missense_var | MED | benign | tolerated_lo | 0,00169365 | 0,000798722 | 0,001667 | conflicting_int Pheochromocytoma Heron | 1 | TRUE | Uncertain significance | PVS1=0 PS=[0, 0, 0, 0, 0.06 | 3.61 | 1,83000 | | |
| 13 | | | protein_c | 5_prime_UTI | LOW | | | -1 | 0,00938498 | -1 | conflicting_int Hereditary_cancer-predispo | 1 | TRUE | Uncertain significance | PVS1=0 PS=[0, 0, 0, 0, 0.81 | 11.91 | 1,77995 | | |
| 14 | | | protein_c | intron_varia | LOW | | | 0,000155413 | -1 | 0,0001895 | pathogenic beta_Thalassemia Beta tha | 1 | TRUE | Uncertain significance | PVS1=0 PS=[0, 0, 0, 0, 0.11 | 4.4 | -1,70000 | | |
| 15 | | | protein_c | 3_prime_UTI | LOW | | | -1 | 0,000599042 | -1 | conflicting_int Histiocytic_medullary_reti | 1 | TRUE | Uncertain significance | PVS1=0 PS=[0, 0, 0, 0, -0.27 | 0.44 | -3,13000 | | |
| 16 | agG/agA | R | protein_c | synonymous | LOW | | | 8,05652E-06 | -1 | 0,000008273 | conflicting_int Limb-girdle_muscular_dyst | 1 | TRUE | Likely benign | PVS1=0 PS=[0, 0, 0, 0, 0.175 | 17.09 | 3 | | |
| 17 | acA/acC | T | protein_c | splice_region | MED | | | 0,000776191 | 0,000199681 | 0,0008566 | conflicting_int Charcot-Marie-Tooth_disea | 1 | TRUE | Likely benign | PVS1=0 PS=[0, 0, 0, 0, 0.137 | 15.22 | -1,22000 | | |
| 18 | aaC/aaT | N | protein_c | synonymous | LOW | | | 0,00752218 | 0,00399361 | 0,007152 | conflicting_int Infantile_neuroaxonal_dyst | 1 | TRUE | Likely benign | PVS1=0 PS=[0, 0, 0, 0, 0.129 | 14.89 | -1,61000 | | |
| 19 | caC/caT | H | protein_c | synonymous | LOW | | | 0,000784936 | 0,000798722 | 0,0007248 | conflicting_int Schwartz_Jampel_syndrom | 1 | TRUE | Likely benign | PVS1=0 PS=[0, 0, 0, 0, -0.44 | 0.13 | -0,38800 | | |
| 20 | tcG/tcA | S | protein_c | synonymous | LOW | | | 0,00724632 | 0,00359425 | 0,007281 | conflicting_int Brugada_syndrome Brugad | 1 | TRUE | Likely benign | PVS1=0 PS=[0, 0, 0, 0, 0.159 | 16.19 | -1,01995 | | |
| 21 | ccC/ccT | P | protein_c | synonymous | LOW | | | 0,000922866 | -1 | 0,0009397 | conflicting_int Long QT_syndrome Romar | 1 | TRUE | Likely benign | PVS1=0 PS=[0, 0, 0, 0, -0.02 | 2.28 | 0,74400 | | |
| 22 | atC/atT | I | protein_c | synonymous | LOW | | | 0,00144297 | 0,000599042 | 0,001162 | conflicting_int Glycogen_storage_disease_ | 1 | TRUE | Benign | PVS1=0 PS=[0, 0, 0, 0, 0.182 | 17.62 | 3,95000 | | |
| 23 | acG/acA | T | protein_c | synonymous | LOW | | | 0,00312562 | 0,00179712 | 0,002875 | conflicting_int Congenital_hypothyroidism | 1 | TRUE | Benign | PVS1=0 PS=[0, 0, 0, 0, -0.56 | 0.06 | -9,8999 | | |
| 24 | gcG/gcT | A | protein_c | synonymous | LOW | | | 4,78544E-05 | -1 | 0,0006509 | conflicting_int Cystic_fibrosis not_provide | 1 | TRUE | Benign | PVS1=0 PS=[0, 0, 0, 0, 0.004 | 3.29 | -9,6899 | | |
| 25 | cGa/cAa | R/Q | protein_c | missense_var | MED | probably_damagi | deleterious | 0,051045701 | 0,0305511 | 0,05 | None | None | 1 | FALSE | Benign | PVS1=0 PS=[0, 0, 0, 0, 0.402 | 28.2 | 4,71000 | |
| 26 | cGc/cAc | R/H | protein_c | missense_var | MED | probably_damagi | deleterious | 0,0261115 | -1 | 0,163 | None | None | 1 | TRUE | Benign | PVS1=0 PS=[0, 0, 0, 0, 0.441 | 33.0 | 5,42999 | |
| 27 | cGc/cAc | R/H | protein_c | missense_var | MED | probably_damagi | tolerated | 0,229199007 | -1 | 0,236 | None | None | 1 | FALSE | Benign | PVS1=0 PS=[0, 0, 0, 0, 0.413 | 29.4 | 4,63000 | |
| 28 | cGa/cAa | R/Q | protein_c | missense_var | MED | possibly_damagi | deleterious | 0,245585993 | 0,232029 | 0,254 | None | None | 1 | TRUE | Benign | PVS1=0 PS=[0, 0, 0, 0, 0.399 | 27.9 | 4,61999 | |
| 29 | cGt/cAt | R/H | protein_c | missense_var | MED | probably_damagi | deleterious | 0,120514996 | 0,0688898 | 0,12 | None | None | 1 | TRUE | Benign | PVS1=0 PS=[0, 0, 0, 0, 0.437 | 33.0 | 5,03000 | |
| 30 | cGg/cAg | R/Q | protein_c | missense_var | MED | benign | tolerated | 0,365570009 | 0,323083 | 0,362 | None | None | 1 | FALSE | Benign | PVS1=0 PS=[0, 0, 0, 0, 0.277 | 23.1 | 3,48000 | |

WES Data Analysis

- Variant Annotation

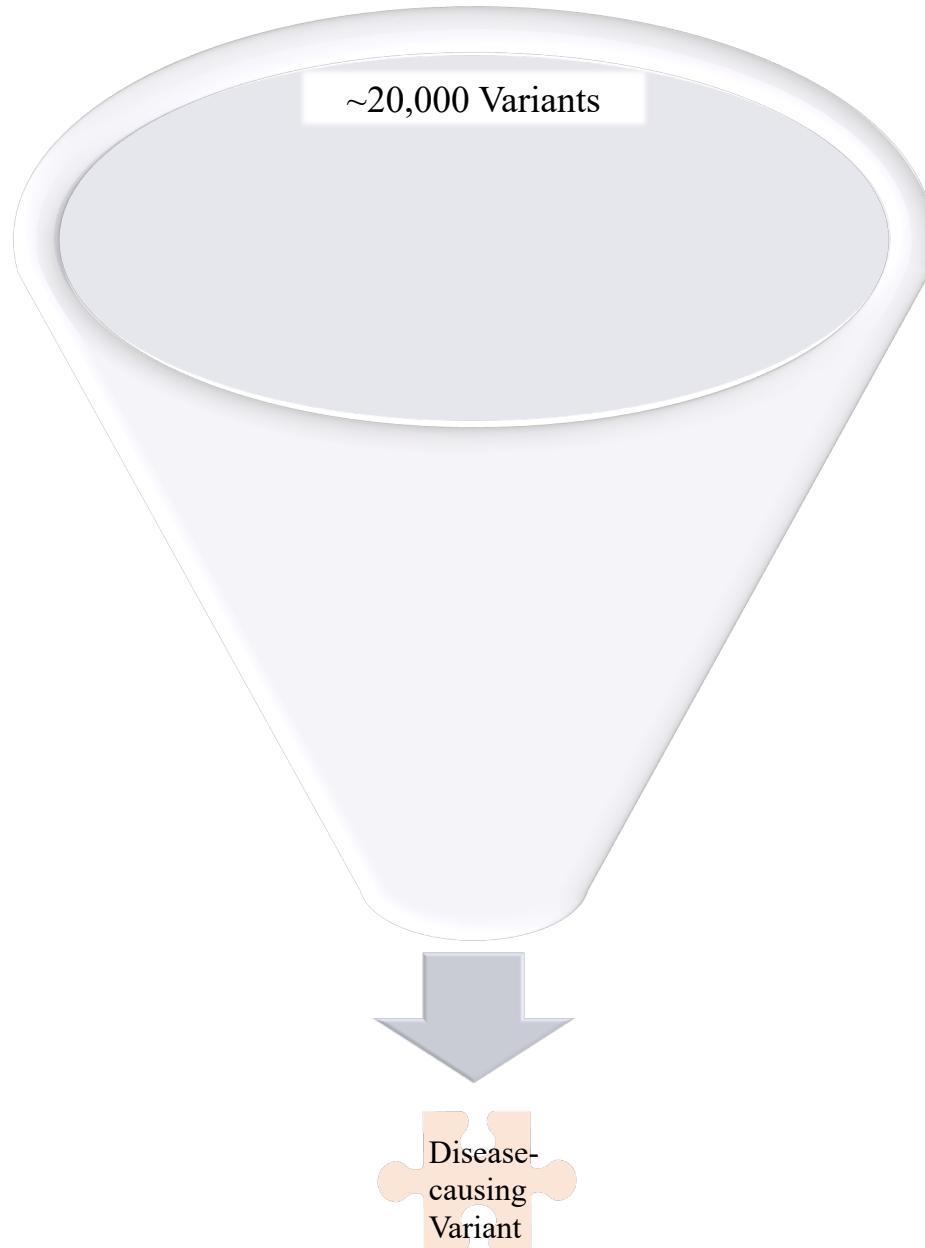
| gerp_element_pval | DANN_score | DANN_rankscore | zygosity | frequency_of_alt_allele | Associated OMIM Phenotype(s) | Associated HPO Phenotype(s) | Associated HPO Disease(s) | ClinVar Reference |
|--------------------|-------------|-----------------|--------------------------------|--|---|---|---|-------------------|
| 2.3.51697e-62 | 0.999000000 | 0.98 | Heterozygous | 0,54887218 | Achondrogenesis, type IA, 200600 (3); Hydrops fetalis; Abnormal enchondral ossification; Achondrogenesis type 1A; ACHONI | No associated phenotype in HPO | Achondrogenesis type 1A; ACHONI | Available |
| 3.0.0 | 0.999000000 | 0.987 | Heterozygous | 0,509433962 | Culler-Jones syndrome, 615849 (3); Axial skeletal anomalies; Strabismus; Holoprosencephaly; Short stature; HOLOPROSENCEPHALY | 9; HPE9; M | Available | |
| 4.4.19156e-115 | 0.999000000 | 0.982 | Heterozygous | 0,414201183 | No associated phenotype in OMIM | No associated phenotype in HPO | No associated disease in HPO | Available |
| 5.7.7712e-110 | 0.997 | 0.8270000000000 | Heterozygous | 0,525773196 | Spastic paraparesis 15, autosomal recessive; Nystagmus; Upper limb spasticity; Hand clumsiness; Autosomal recessive spastic paraparesis | 15 | GLYCOGEN STORAGE DISEASE III; GSD3 | Available |
| 6.1.87209e-212 | 0.996 | 0.764 | Heterozygous | 0,475308642 | Glycogen storage disease IIIa, 232400 | Malar flattening; Cardiomyopathy; Hypertrophic cardiomyopathy; GLYCOGEN STORAGE DISEASE III; GSD3 | Available | |
| 7.3.37427e-166 | 0.995 | 0.6559999999999 | Heterozygous | 0,487341772 | Hyper-IgE recurrent infection syndrome, 614740 (3); Chronic otitis media; Skin ulcer; Verrucous skin lesions; Combined immunodeficiency due to hyperimmunoglobulin E syndrome | Chronic otitis media; Skin ulcer; Verrucous skin lesions | Combined immunodeficiency due to hyperimmunoglobulin E syndrome | Available |
| 8.2.16929e-37 | 0.992 | 0.545 | Heterozygous | 0,443661972 | Marinesco-Sjogren syndrome, 248800 | Skeletal muscle atrophy; Kyphosis; Cervical spine stenosis; MARINESCO-SJOGREN SYNDROME | Available | |
| 9.1.95984e-43 | 0.97 | 0.314 | Heterozygous | 0,453846154 | Ciliary dyskinesia, primary, 14, 613800 | Situs inversus totalis; Cough; Clubbing; Primary ciliary dyskinesia; CILIARY | Available | |
| 10.4.11538e-40 | 0.93 | 0.2189999999999 | Heterozygous | 0,535 | Cardiac arrhythmia, ankyrin-B-related, 604145 | Sinus bradycardia; Autosomal dominant Romano-Ward syndrome; CARDIA | Available | |
| 11.3.65541e-67 | 0.844000000 | 0.1480000000000 | Heterozygous | 0,187793427 | (Bronchiectasis with or without elevation of serum C-reactive protein; Anorexia; Type I diabetes mellitus; Chylomicronemia; Male infertility with azoospermia) | Chylomicronemia; Type I diabetes mellitus; Chylomicronemia; Male infertility with azoospermia | Available | |
| 12.12.0.525 | 0.047 | Heterozygous | 0,380530973 | Gastrointestinal stromal tumor, 606700 | Hoarse voice; Recurrent paroxysmal cough; Carney-Stratakis syndrome; Gastrointestinal stromal tumor | Carney-Stratakis syndrome; Gastrointestinal stromal tumor | Available | |
| 13.13.0.0 | | Heterozygous | 0,488789238 | Adenomas, multiple colorectal, 608450 | Somatic mutation; Pilomatrixoma; GASTRIC CANCER; PILOMATRIXOMA | GASTRIC CANCER; PILOMATRIXOMA | Available | |
| 14.14.0.0 | | Heterozygous | 0,365853659 | Delta-beta thalassemia, 141749 (3); Alpha thalassemia; Genu valgum; Anxiety; Hypothyroidism; Hypothyroidism | Beta-thalassemia intermedia; Beta thalassemia intermedia; Hypothyroidism | Beta-thalassemia intermedia; Beta thalassemia intermedia; Hypothyroidism | Available | |
| 15.15.0.0 | | Heterozygous | 0,649122807 | Combined cellular and humoral immunodeficiencies, 604146 | Aplasia/Hypoplasia of the eyebrow; Hypoplasia of the eyelid; SEVERE COMBINED IMMUNODEFICIENCY | SEVERE COMBINED IMMUNODEFICIENCY | Available | |
| 16.16.7.75191e-299 | | Heterozygous | 0,510309278 | Cardiomyopathy, dilated, 1G, 604145 | Elevated serum creatine kinase; Limitation of physical activity; CARDIOMYOPATHY, DILATED, 1G; CDMYD | CARDIOMYOPATHY, DILATED, 1G; CDMYD | Available | |
| 17.17.1.08474e-61 | | Heterozygous | 0,548076923 | Charcot-Marie-Tooth disease, axonal, 1G, 611876 (3) | No associated phenotype in HPO | No associated disease in HPO | Available | |
| 18.18.1.74482e-50 | | Heterozygous | 0,467153285 | Infantile neuroaxonal dystrophy 1, 257800 | Optic atrophy; Seizures; Neurofibromatosis; NEURODEGENERATION WITH BRAIN ANOMALIES | NEURODEGENERATION WITH BRAIN ANOMALIES | Available | |
| 19.19.4.88373e-106 | | Heterozygous | 0,5 | Dyssegmental dysplasia, Silverman-Harris type, 612955 (3) | Cryptorchidism; Flat face; Overgrowth syndrome; Schwartz-Jampel syndrome; DYSSEGMENTAL DYSPLASIA | Schwartz-Jampel syndrome; DYSSEGMENTAL DYSPLASIA | Available | |
| 20.20.3.24383e-110 | | Heterozygous | 0,477272727 | Brugada syndrome 4, 611876 (3) | Sick sinus syndrome; Right bundle branch block; Brugada syndrome; BRUGADA SYNDROME | BRUGADA SYNDROME | Available | |
| 21.21.1.00007e-53 | | Heterozygous | 0,52 | Long QT syndrome 12, 612955 (3), Autosomal dominant | Torsade de pointes; Long QT syndrome; Autosomal dominant long QT syndrome | Romano-Ward syndrome; LONG QT SYNDROME | Available | |
| 22.22.9.47141e-76 | | Heterozygous | 0,31097561 | McArdle disease, 232600 (3), Autosomal recessive | Myopathy; Elevated serum creatine kinase; Myopathy | GLYCOGEN STORAGE DISEASE V; GSD5 | Available | |
| 23.23.0.0 | | Heterozygous | 0,466666667 | Thyroid dyshormonogenesis 2A, 274500 | Muscular hypotonia; Constipation; Hypothyroidism; Familial thyroid dyshormonogenesis | Familial thyroid dyshormonogenesis | Available | |
| 24.24.3.65541e-67 | | Heterozygous | 0,256410256 | (Bronchiectasis with or without elevation of serum C-reactive protein; Anorexia; Type I diabetes mellitus; Chylomicronemia; Male infertility with azoospermia) | Chylomicronemia; Type I diabetes mellitus; Chylomicronemia; Male infertility with azoospermia | Available | | |
| 25.25.1.79931e-116 | 1.0 | 0.9990000000000 | Heterozygous | 0,382165605 | No associated phenotype in OMIM | No associated phenotype in HPO | No associated disease in HPO | |
| 26.26.0.0 | 1.0 | 0.9990000000000 | Heterozygous | 0,208913649 | No associated phenotype in OMIM | No associated phenotype in HPO | No associated disease in HPO | |
| 27.27.1.26565e-98 | 1.0 | 0.9990000000000 | Heterozygous | 0,31441048 | No associated phenotype in OMIM | No associated phenotype in HPO | No associated disease in HPO | |
| 28.28.1.41339e-214 | 1.0 | 1.0 | Homozygous for the Alternative | 1 | No associated phenotype in OMIM | No associated phenotype in HPO | No associated disease in HPO | |
| 29.29.1.84707e-172 | 1.0 | 1.0 | Homozygous for the Alternative | 1 | No associated phenotype in OMIM | No associated phenotype in HPO | No associated disease in HPO | |
| 30.30.3.45814e-63 | 1.0 | 1.0 | Heterozygous | 0,5 | No associated phenotype in OMIM | No associated phenotype in HPO | No associated disease in HPO | |

WES Data Analysis



WES Data Analysis

- **Variant Prioritization**



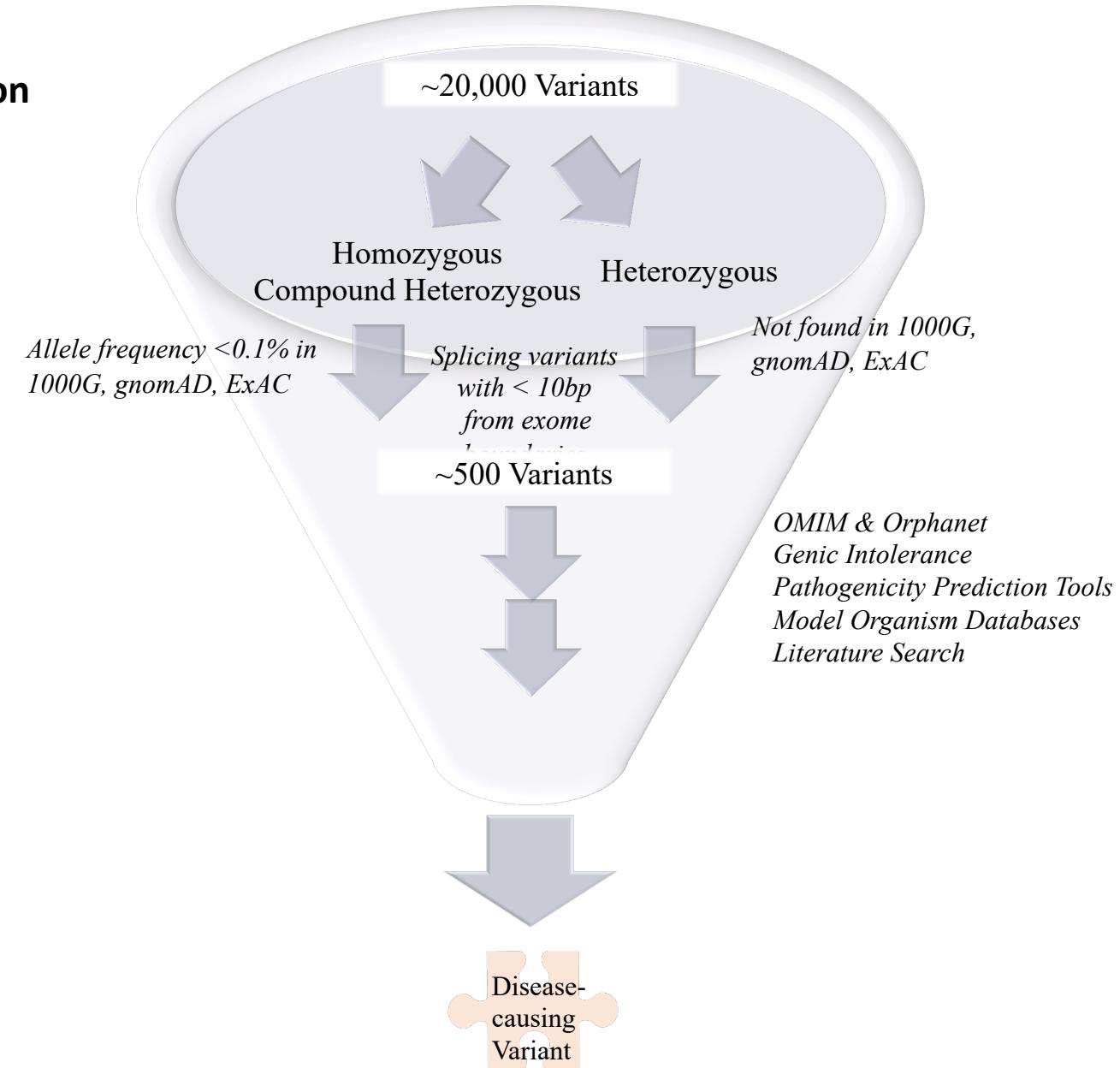
Variant Prioritization Strategies

- **Variant Prioritization**

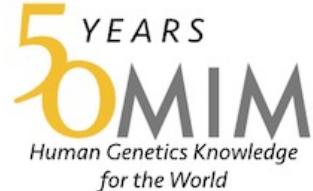
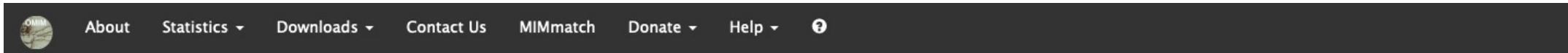
- Consider variants according to their inheritance pattern (Het, Compound Het, Hom)
- Filter variants based on population frequencies
 - GnomAD, ExAC and 1000G allele frequencies are used
 - Homozygous and compound heterozygous variants which are seen < 0.1% are taken
 - Heterozygous variants which are found in at least one database are excluded
- Exclude splicing variants which are far away >10 bp from exome boundaries

WES Data Analysis

- **Variant Prioritization**



Variant Prioritization Strategies



OMIM®

Online Mendelian Inheritance in Man®

An Online Catalog of Human Genes and Genetic Disorders

Updated March 31, 2021

Search OMIM for clinical features, phenotypes, genes, and more...



[Advanced Search : OMIM, Clinical Synopses, Gene Map](#)

[Need help? : Example Searches, OMIM Search Help, OMIM Video Tutorials](#)

[Mirror site : <https://mirror.omim.org>](#)

OMIM is supported by a grant from NHGRI, licensing fees, and [generous contributions from people like you](#).

<https://www.omim.org>

Variant Prioritization Strategies



SLC6A1 solute carrier family 6 (neurotransmitter transporter), member 1

Dataset gnomAD v2.1.1 ▾ gnomAD SVs ⓘ

Ensembl gene ID ENSG00000157103
Ensembl transcript ID ENST00000287766 (canonical)
UCSC Browser 3:11034411-11080934
GeneCards SLC6A1
OMIM 137165

| Gene Constraint ⓘ | | | | | |
|-------------------|---------------|---------------|-----------|-----------------------------|-------|
| Category | Exp. no. SNVs | Obs. no. SNVs | Z = | Constraint metrics | |
| Synonymous | 154.8 | 170 | Z = -0.96 | o/e = 1.1 (0.97 - 1.25) | 0 ⚡ 1 |
| Missense | 370.1 | 144 | Z = 4.18 | o/e = 0.39 (0.34 - 0.45) | 0 ⚡ 1 |
| LoF | 31.7 | 1 | pLI = 1 | o/e = 0.03 (0.01 - 0.15) | 0 ⚡ 1 |

<https://gnomad.broadinstitute.org>

Variant Prioritization Strategies



About Terms Download

Contact

About RVIS (Residual Variation Intolerance Score) is a gene score based module intended to help in the interpretation of human sequence data.

News We now provide a revised RVIS gene score (unpublished) based on the recently released ExAC v2 (release 2.0). The data can be found at (gnomad.broadinstitute.org).

Search for a gene or gene list



Examples - Gene: [SCN1A](#), Gene list: [SCN1A](#), [ATP1A3](#), [HLA-A](#), [MTOR](#), [MUC5B](#) (comma delimited), [All.Gene](#)

Download

| GENE | RVIS | %ExAC RVIS | ExAC LoF FDR | %ExAC v2 RVIS | Edge Case (%OE-ratio) |
|--------|----------------|------------|--------------|--------------------|-----------------------|
| SLC6A1 | -0.36 (29.16%) | 9.06% | 0.000431701 | -1.0796 (11.5347%) | N (1.6789%) |

<http://genic-intolerance.org>

It compares the average number of common functional variants and the number of total coding variants observed in the gene.

Constrained genes have less common functional variation than expected and have lower RVIS % score.
Genes with more common functional variants have higher RVIS % score.

Variant Prioritization Strategies

3812–3814 Nucleic Acids Research, 2003, Vol. 31, No. 13
DOI: 10.1093/nar/gkg509

SIFT: predicting amino acid changes that affect protein function

Pauline C. Ng and Steven Henikoff*

Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N A1-162, Seattle, WA 98109, USA

Received January 4, 2003; Revised and Accepted February 28, 2003

A general framework for estimating the relative pathogenicity of human genetic variants

Martin Kircher^{1,5}, Daniela M Witten^{2,5}, Preti Jain^{3,4}, Brian J O’Roak^{1,4}, Gregory M Cooper³ &

Current methods for annotating and interpreting human genetic variation tend to exploit a single information type (for example, conservation) and/or are restricted in scope (for example, to missense changes). Here we describe Combined Annotation-Dependent Depletion (CADD), a method for objectively integrating many diverse annotations into a single

comparable, making it difficult to evaluate the relative strengths of distinct variant categories or annotation methods trained on known pathogenic mutations. These ascertainment biases and may not be generalizable. A major practical challenge to obtain, let alone integrate, the existing panoply of partially

REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants

Nilah M. Ioannidis,^{1,2,34} Joseph H. Rothstein,^{2,3,4,34} Vikas Pejaver,⁵ Shannon K. McDonnell,⁷ Saurabh Baheti,⁷ Anthony Musolf,⁸ Qing Danielle Karyadi,⁹ Lisa A. Cannon-Albright,¹⁰ Craig C. Teerlink,¹⁰ Janet L. Stanford,¹¹ William B. Isaacs,¹² Jianfeng Xu,¹³ Kathleen A. Cooney,^{10,14} Ethan M. Lange,¹⁵ Johanna Schleutker,^{16,17} John D. Carpten,¹⁸ Isaac J. Powell,¹⁹ Olivier Cussenot,²⁰ Geraldine Cancel-Tassin,²⁰ Graham G. Giles,^{21,22} Robert J. MacInnis,^{21,22} Christiane Maier,^{23,24} Chih-Lin Hsieh,²⁵ Fredrik Wiklund,²⁶ William J. Catalona,²⁷ William D. Foulkes,²⁸ Diptasri Mandal,²⁹ Rosalind A. Eeles,³⁰ Zsofia Kote-Jarai,³⁰ Carlos D. Bustamante,^{1,31} Daniel J. Schaid,⁷ Trevor Hastie,^{31,32} Elaine A. Ostrander,⁹ Joan E. Bailey-Wilson,⁸ Predrag Radivojac,⁵ Stephen N. Thibodeau,³³ Alice S. Whittemore,^{2,31} and Weiva Sieh^{2,3,4,*}

M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity

Karthik A Jagadeesh^{1,5}, Aaron M Wenger^{2,5}, Mark J Berger¹, Harendra Guturu², Peter D Stenson³, David N Cooper³, Jonathan A Bernstein² & Gill Bejerano^{1,2,4}

all benign → eliminate
all pathogenic → keep
no consensus → expert revision

Variant Prioritization Strategies

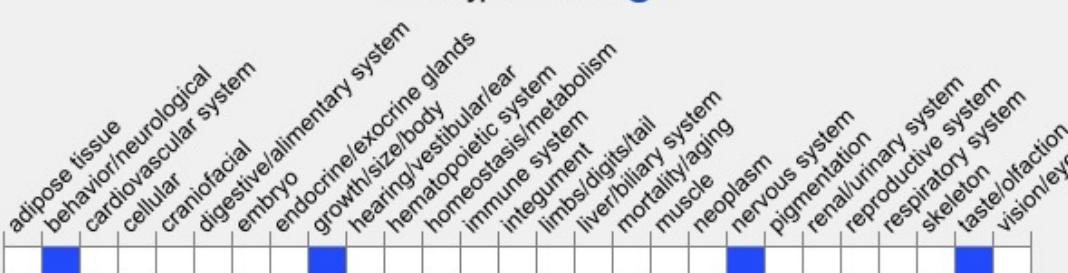
 MG1
[About](#) [Help](#) [FAQ](#)

Mouse Genome Informatics

Mutations, Alleles, and Phenotypes

[less ▾](#) **Phenotype Summary** 29 phenotypes from 3 alleles in 3 genetic backgrounds
17 phenotype references

Phenotype Overview ?



Click cells to view annotations.

Homozygous hypomorphic mice display abnormal inhibitory postsynaptic currents, and abnormal GABA uptake and release. Null mice show hyperactivity and various behavioral abnormalities, as well as an aversion to bitter taste.

All Mutations and Alleles 8

- Gene trapped 2
- Spontaneous 1
- Targeted 5

Genomic Mutations 1 involving *Slc6a1*

Incidental Mutations [Mutagenetix](#), APF

Find Mice (IMSR) 32 strains or lines available

Comparison Matrix Gene Expression + Phenotype

The presence of overlapping symptoms with the patient increases the priority of a gene.

Variant Prioritization Strategies

Literature Search

PubMed.gov

Search PubMed

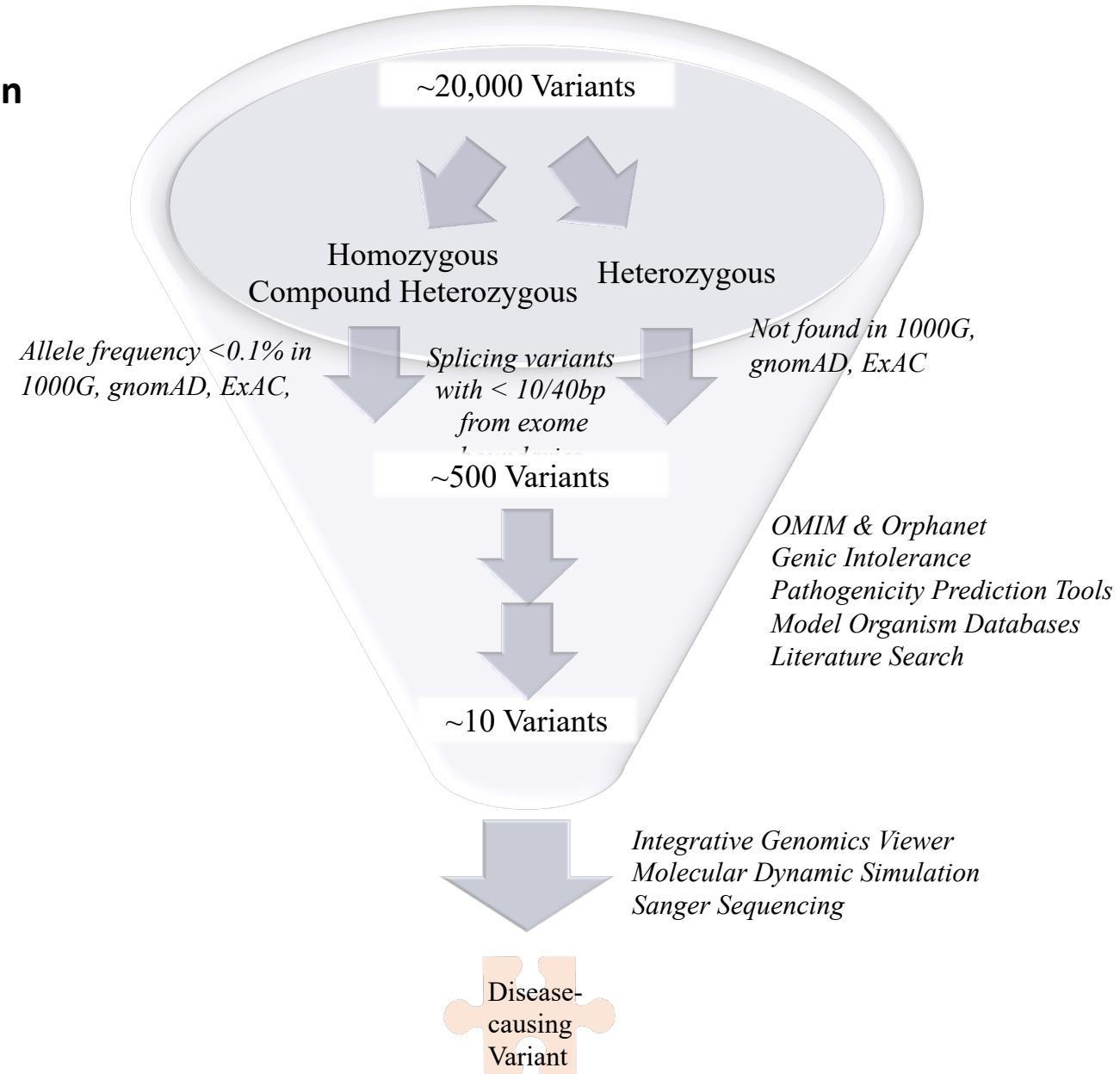
Search

Advanced

PubMed® comprises more than 32 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full text content from PubMed Central and publisher web sites.

WES Data Analysis

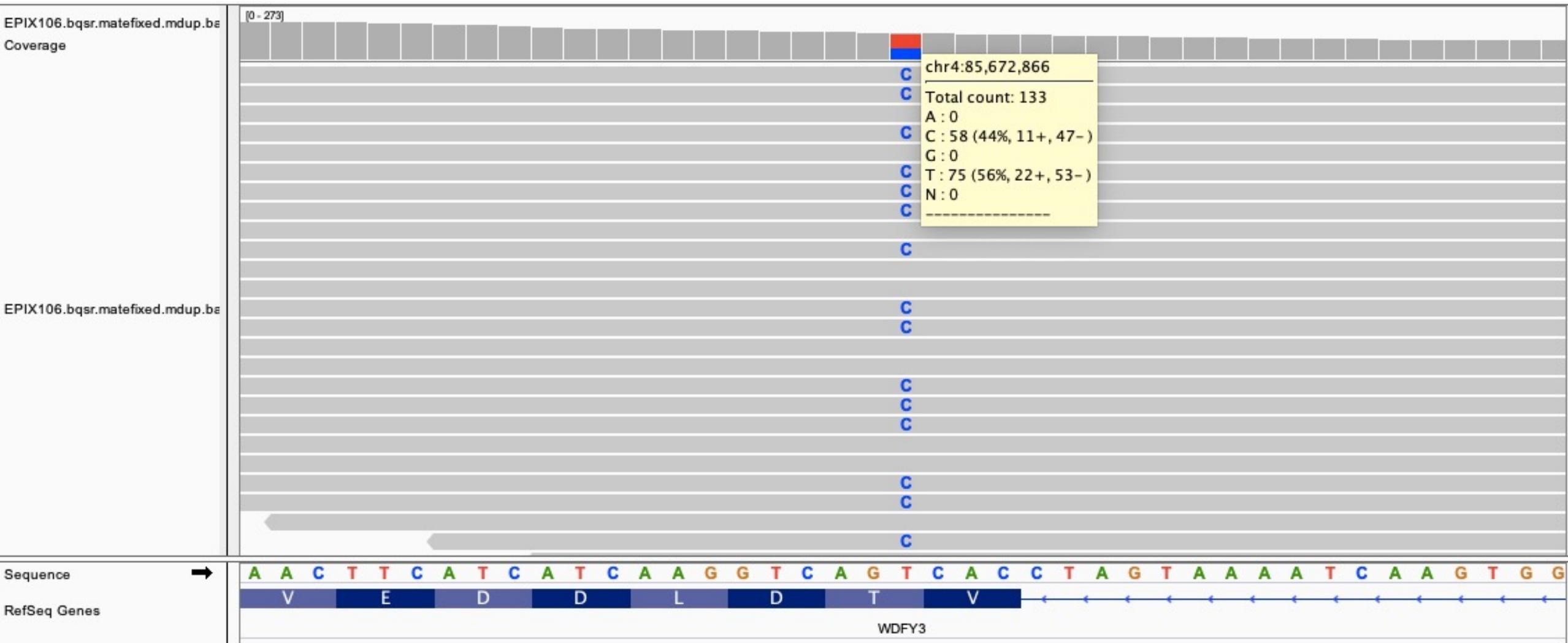
- **Variant Prioritization**



Variant Prioritization Strategies



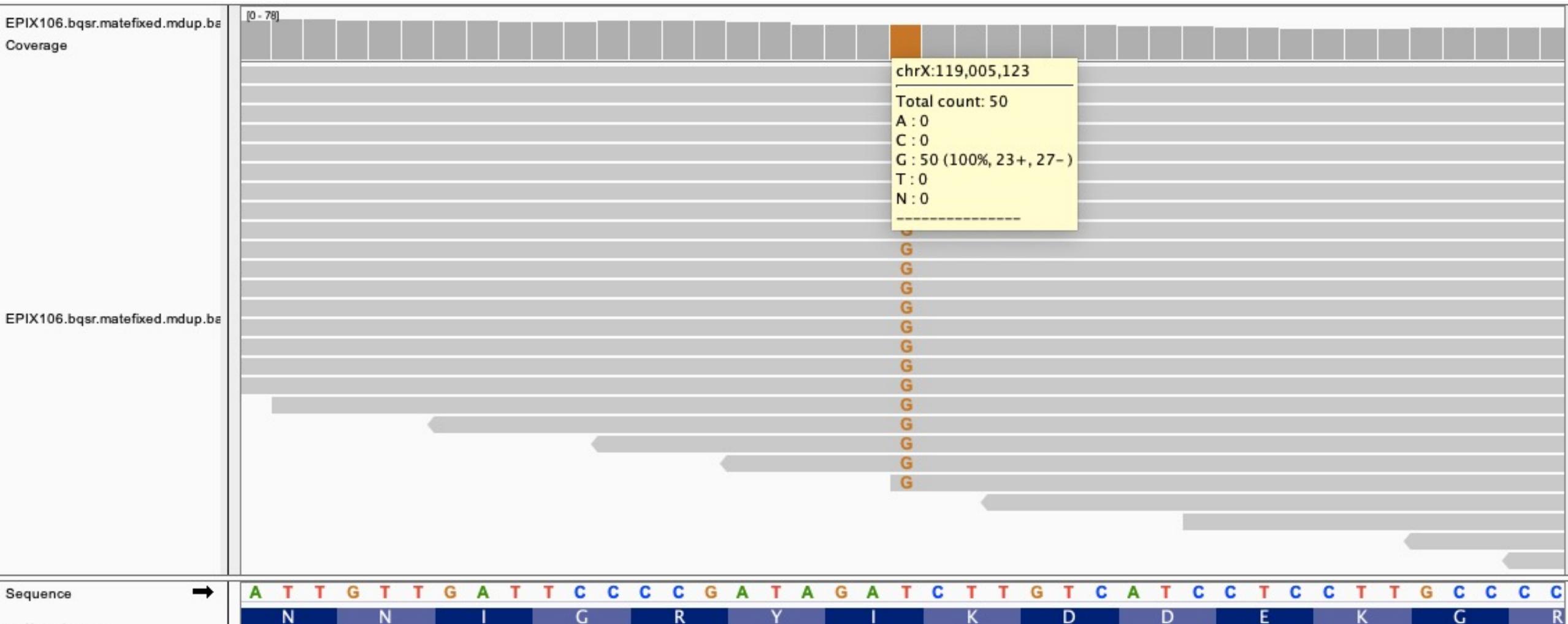
Variant heterozygous:



Variant Prioritization Strategies

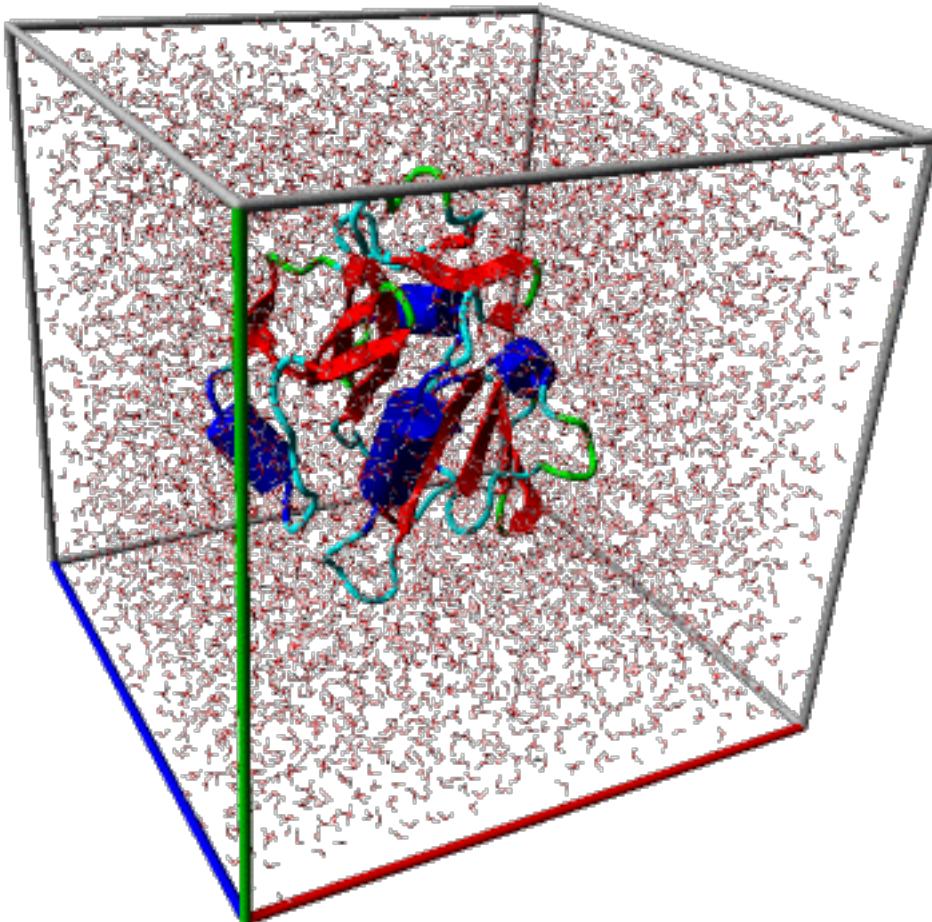


Variant homozygous:



Variant Prioritization Strategies

Molecular Dynamic (MD) Simulations: Predict the functional impact of the selected variant

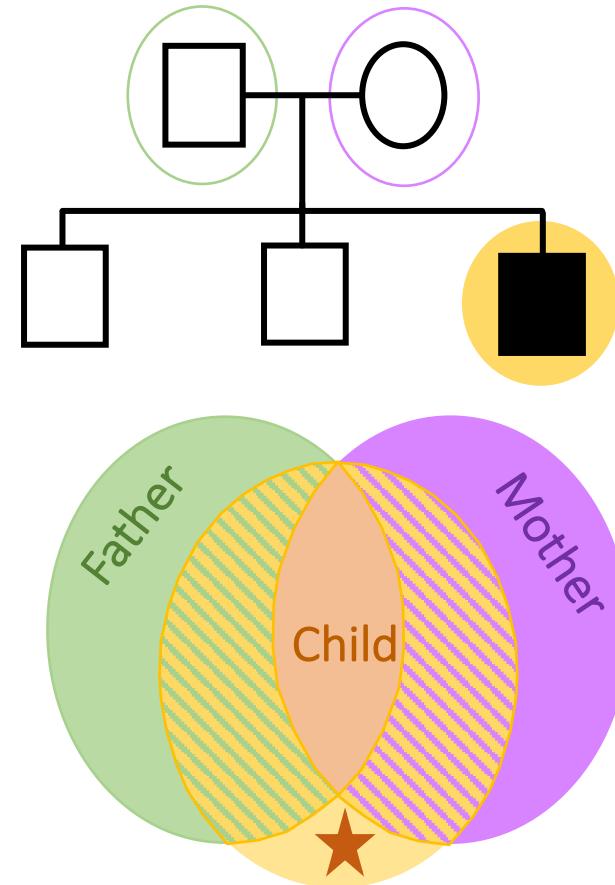


MD is a computer simulation method that monitors the physical movements of several hundreds of atoms and molecules in solution on a femtosecond timescale by calculating the electrostatic charges.

A REAL-LIFE CASE

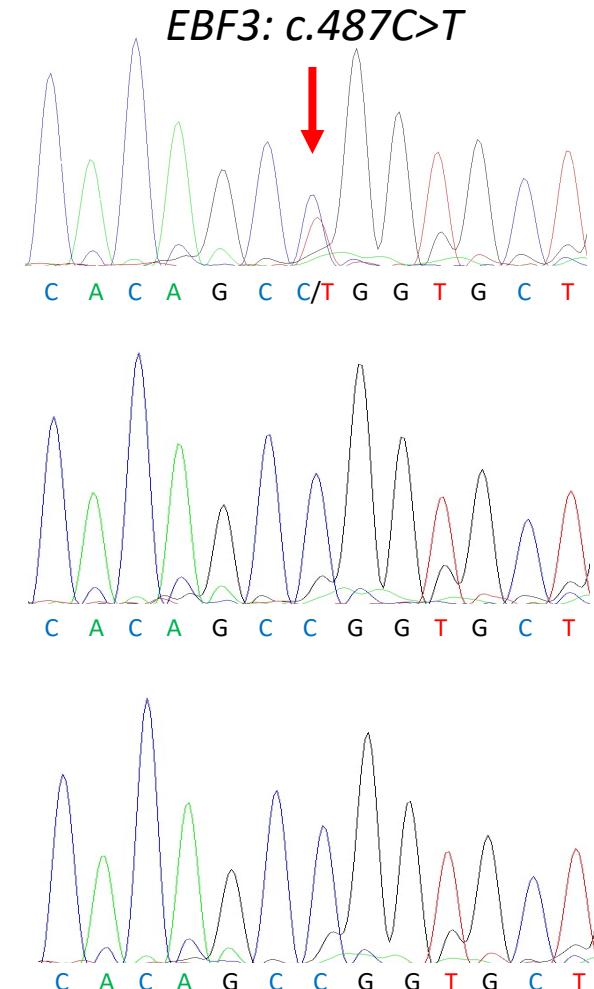
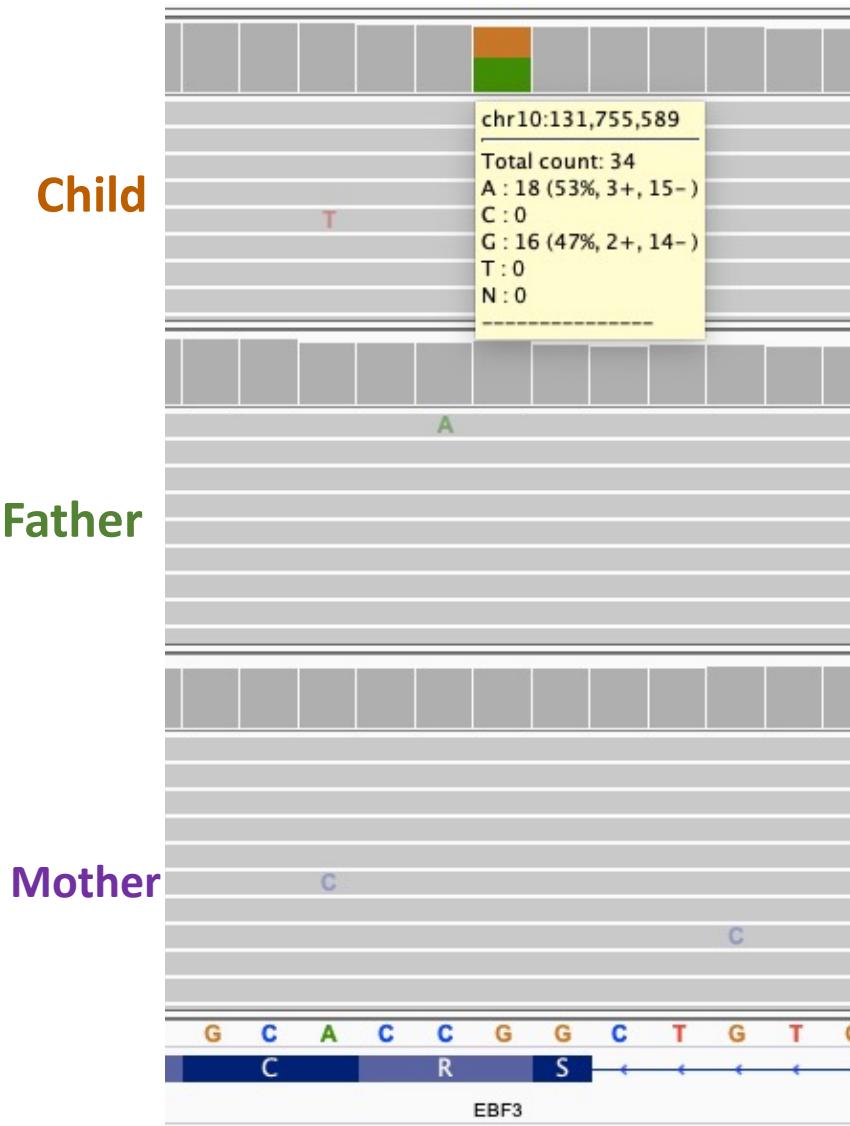


- 6 year-old boy
- Global developmental delay
- Generalized hypotonia
- Speech delay
- Bilateral esotropia
- Decreased pain response
- Hyperactive DTR
- Mild facial dysmorphisms:
frontal bossing, low-set ears



| Heterozygous Variants | | Homozygous and Compound Heterozygous Variants | | | |
|---|--|--|---|---------------------------------|---|
| Gene Variant | Phenotype | • Population databases (GnomAD) | | • Population databases (GnomAD) | |
| | | gnomAD Gene constraint | Intolerance Score (RVIS) | CADD REVEL M-CAP | MGI Phenotype |
| EBF3 (OMIM *607407) NM_0010054 63: exon6: c.C487T: p.R163W | Hypotonia, ataxia, and delayed development syndrome 1 (OMIM #617330) | Missense Z Score = 3.61 Loss of -0.65 Function pLI Score = 1 | | 33 0.575 0.009 | Homozygous mutant mice die perinatally and exhibit impaired olfactory neuron projection. (MGI:894289) |

A REAL-LIFE CASE



A REAL-LIFE CASE

EBF3 and HADDS

* 607407

EARLY B-CELL FACTOR 3; EBF3

Alternative titles; symbols

COLLIER/OLF1/EBF TRANSCRIPTION FACTOR 3;
COE3

EBF3 expression in early post-mitotic neurons during embryogenesis has a crucial role in neuronal differentiation and maturation.

HGNC Approved Gene Symbol: EBF3

*Cytogenetic location: 10q26.3 Genomic coordinates
(GRCh38): 10:129,835,232-129,964,273 (from NCBI)*

No prevalence information about the disease has been reported because of not enough data available.

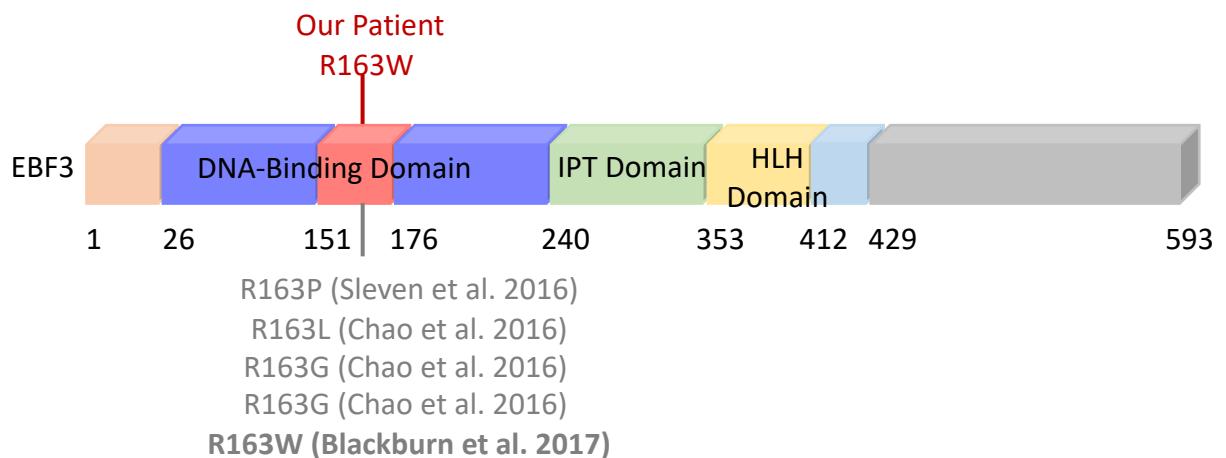
Gene-Phenotype Relationships

| Location | Phenotype | Phenotype MIM number | Inheritance | Phenotype mapping key |
|----------|---|----------------------|-------------|-----------------------|
| 10q26.3 | Hypotonia, ataxia, and delayed development syndrome | 617330 | AD | 3 |

**There are 20 different pathogenic variants with 29 affected individuals (Tanaka et al.2017).*

A REAL-LIFE CASE

EBF3 and HADDS



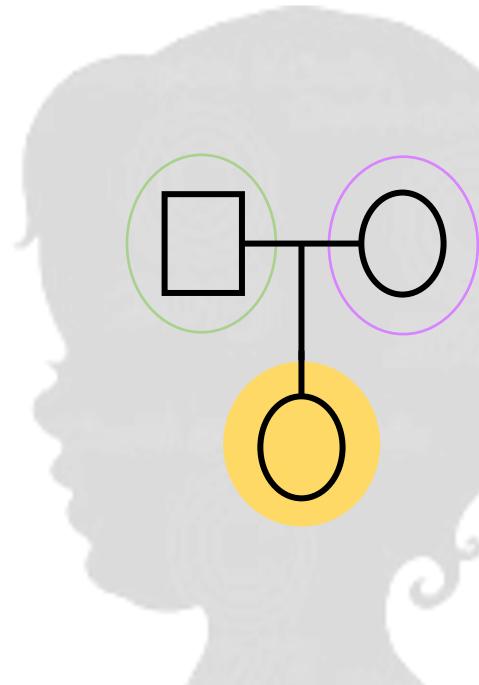
Blackburn et al. 2017, *Cold Spring Harbor Molecular Case Studies*

CONCLUSION

- Collectively RDs are not rare, and there are still challenges to diagnose them.
- WES is a promising way to overcome challenges in the diagnosis.
- There are many software and tools that focus on a particular aspect of the entire analysis but no gold standards.
- The variant filtration is the most critical step and requires a careful interpretation.

CASE-I

It's your turn! The annotated VCF file will be shared with you. Prioritize the variants and diagnose the patient. Prepare a report (.ppt) that includes your study. Please send that to Prof. Sezerman and me via e-mail.



- Intrauterine growth retardation
- Birth measurements: W 2300 g (-2 SD), OFC 29 cm (-3 SD), L 47 cm (-1,5 SD)
- Facial dysmorphism
- Short stature (-6 SD)
- Dystrophy
- Reduced subcutaneous fat and loose skin
- Muscular hypotonia
- Elevated liver enzymes (normal coagulation studies)
- Liver biopsy: normal histology
- Hypogammaglobulinemia (low IgG, A, M, E and reduced B lymphocytes)
- Progressive loss of visual acuity and optic atrophy
- Poor ossification of the calvarium
- Wide open fontanelles
- Multiple occipital wormian bones
- Maxilla hypoplasia
- Mild osteopenia
- Slender ribs and clavicles and slender tubular long bones
- Delayed epiphyseal ossification

Thank You!



Paediatric genomics: diagnosing rare disease in children

Caroline F. Wright¹, David R. FitzPatrick² and Helen V. Firth^{3,4}

<https://www.ncbi.nlm.nih.gov/pubmed/29398702>