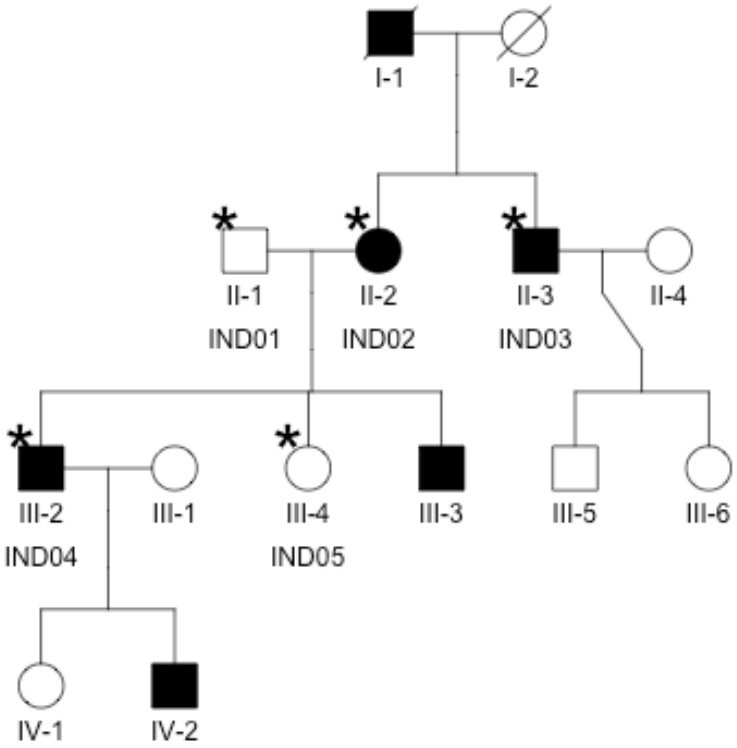
Below is a pedigree of a family with a suspect Mendelian form of hyperlipidemia (very high LDL-C). Shaded individuals have been diagnosed with hyperlipidemia. Five individuals from this pedigree have been sequenced in a 5 MB region surrounding a linkage peak on chromosome 1 and a vcf file has been produced. Individuals that have been sequenced are indicated by a star (*). The vcf has been converted to a tab-delimited file, mendelian_2021.txt, with genotypes for PASS variants seen in at least one of the five individuals.



Part A

Based on the pedigree, what is the likely mode of inheritance?

The most likely mode of inheritance is autosomal dominant because the disease transmits vertically without skipping generations, and approximately half the offspring of affected parents are also affected.

Part B

How many variant records are in the file?

There are 11,268 variants in this file.

Part C

How many of the variants have more than one ALT allele?

63 variants have more than one ALT allele.

Part D

How many of the variants are SNPs? SNPs are the variants with REF and ALT allele that are both one of the four bases.

9,934 of the variants are SNPs, only having a single nucleotide at the reference and alternate allele.

Part E

How many of the SNPs have at least one alternate allele among the 5 individuals? This means the SNP has an allele count over 0.

9,383 SNPs have at least one alternate allele among the 5 individuals.

Part F

For each of the sequenced individuals, calculate the following metrics and create a table with each subject and their values for these QC metrics:

- missingness rate
- # of variant sites
- # of singletons
- Ti/Tv ratio

Individual	Missing Rate	# Variant Sites	#Singletons	Ti/Tv Ratio
1	0.1154%	7735	335	2.2495
2	0.0887%	7687	176	2.2691
3	0.1242%	7337	648	2.2725
4	0.0887%	7506	0	2.2799
5	0.1242%	7450	1	2.3102

Part G

Explain the pattern of number of singletons you observe in these 5 individuals.

Individuals 4 and 5 are the offspring of Individuals 1 and 2, so each half of the children's genes comes from one of their parents. The father and mother have 335 and 176 singletons, respectively, but the children have almost 0 singletons. Singletons are variants where only one individual has the alternate allele. This makes sense because since all the children's alleles are inherited from the parents, so they are not expected to be carrying a different variant from their parents. Individual 5's one singleton must be a mutation.

Individual 2 is shares half her genes with her children, Individuals 4 and 5, and her brother, Individual 3. Since she has the most overlap with every other sequenced individual, she has fewer singletons compared 3. Individual 3 is the least related to every other sequenced member, which is why he has the highest number of singletons.

Part H

Assume complete penetrance, individuals with hyperlipidemia carry at least one alternate allele that causes their high LDL-C, and that those that do not have hyperlipidemia do not have the alternate allele. Use the 5 subjects and the expected mode of inheritance to determine the variants that are segregating in the family. How many variants segregate in the family?

Individuals 1 and 5 are unaffected controls; individuals 2, 3, and 4 are affected. Variants that are associated with hyperlipidemia will be seen in affected individuals but not the controls. From the pedigree, the affected individuals must be heterozygous for the disease-related variant if hyperlipidemia is assumed to be autosomal dominant. Based on that criteria, 26 variants are segregating within the family.

Part I

Do any of the segregating variants alter protein-coding region? Do you think this is the plausible causal variant is in this family, and why?

Disregarding intron and intergenic variants, there are two mutations that alter protein-coding regions. rs1065173 is a synonymous variant, which means the codon substitution doesn't change the encoded amino acid, so this is not a causal variant. rs28942111 is a missense mutation, which would alter the coded amino acid. This variant is a plausible causal variant in the family.