HyperLDL-C Sourcery

Introduction

HyperLDL-C Sourcery should be of interest to anyone concerned with the genetics of hypercholesterolemia (herein referred to as hyperLDL-C). Familial hypercholesterolemia (FH) is a relatively prevalent and under-diagnosed Mendelian genetic disorder that is known to be caused by mutations in the *APOB*, *APOE*, *LDLR*, and *PCSK9* genes. It is usually inherited in an autosomal dominant manner, although rare and more severe forms can be inherited in a recessive manner. However, FH due to mutations in these genes, with a prevalence of approximately 1 in 200-500, only accounts for a relatively small proportion of individuals with hyperLDL-C. HyperLDLC is a very common condition affecting 10-50% of the population. The most frequent cause of hyperLDL-C is multifactorial, that is, due to the combination of lifestyle factors (e.g. diet, exercise, etc) and polygenic factors (i.e. normal genetic variation in genes that each have only a small influence on how the body processes LDL-C). Molecular genetic testing for hyperLDL-C can include sequencing and dosage analysis of the *APOB*, *APOE*, *LDLR*, *LDLRAP1*, and *PCSK9* genes, and may also include genotyping of 6 or 12 single nucleotide polymorphisms (SNPs) that are known to influence LDL-C levels. As outlined in the article by Talmud *et al*. (2013),[1] the genotype profile determined by these SNPs generates an 'LDL-C SNP score', which can be used as measure of an individual's polygenic risk for developing hyperLDL-C, with higher scores conferring a higher risk.

In addition to hyperLDL-C, FH can also present with concomitant tendon xanthoma. In fact, the presentation of hyperLDL-C with tendon xanthoma is a very strong indicator of a pathogenic variant in an FH-related gene, as tendon xanthoma is not generally observed in individuals with multi-factorial (polygenic) hyperLDL-C.

If an individual presents with hyperLDL-C without tendon xanthoma, it can be difficult to determine whether that individual has FH or multi-factorial (polygenic) hypercholesterolemia, even if there is a strong family history, and this is because hyperLDL-C is so very common. Genetic testing for FH is not 100% sensitive; hence, non-detection of a potentially pathogenic variant does not exclude a diagnosis of FH. Alternatively, genetic testing may only detect a variant of uncertain significance. **In either outcome, an important question that remains is 'what is the likelihood that the proband actually has FH?' This is the main question that HyperLDL-C Sourcery is specifically designed to answer.** In this context, FH is defined as the presence of hyperLDL-C due to a pathogenic mutation (whether detected or not). The presentation of hyperLDL-C in the absence of a pathogenic FH-related mutation (detected or not) is considered to be 'multi-factorial hypercholesterolemia', or, some authors have preferred the term 'polygenic hypercholesterolemia.[2]

There are a number of important factors that should be considered in order to answer the aforementioned question. Not least (in fact, arguably the first), is the prior probability that an individual with hyperLDL-C has an FH-causal mutation. Although FH is a relatively common genetic disorder, its prevalence is very significantly lower than multi-factorial (polygenic) hyperLDL-C. Hence, the prior probability of having an FH-causal mutation is generally considered to be low. Another important factor is the individual's polygenic risk, as measured by their LDL-C SNP score. If a proband with hyperLDL-C is found to have either no variant or a variant of uncertain significance, a high LDL-C SNP score makes it less likely that the individual has FH (or in other words: less likely that the VUS is pathogenic). Alternatively, a low LDL-C SNP score would make it more likely that the individual's presentation is due to the detected variant and so could warrant further investigation of the detected VUS, e.g. by cosegregation studies. Other factors to consider include mutation penetrance (which also impacts the prior probability mentioned above), the clinical sensitivity of testing, and the pathogenicity classification assigned to any detected variant.

A mathematical formula to calculate the probability that an individual with hyperLDL-C in whom no pathogenic mutation can be found has multi-factorial (polygenic) hyperLDL-C (i.e. does not have an undetected FH-causing mutation) has previously been described.[3] But what if a VUS or likely pathogenic variant is detected? Or even a variant classified as 'pathogenic'? Unless there is 100% certainty that a detected variant is causal of FH, there still remains a possibility that the variant has *not* caused the hyperLDL-C. According to variant classification guidelines,[4] variants classified as 'likely pathogenic' are required to have at least 90% probability of being pathogenic, and the probability figure is 99% for those classified as 'pathogenic'. That leaves a potential 1% probability that a

'pathogenic' variant in an FH-related gene is not actually causal of FH. For most genetic conditions (where the phenotype and its causes are relatively rare) this is fine. However, in the context of an exceedingly common condition where the huge majority of cases have an alternative cause, that 1% probability of a variant being non-pathogenic equates to a much higher probability of non-causality. The calculation of this probability is where HyperLDL-C Sourcery comes into play. For example, an individual with hyperLDL-C and a SNP score in population decile 5 is found to have a variant in an FH-related gene, which has been classified as 'pathogenic' after being deemed to have 99% probability of pathogenicity. Assuming the default parameters for FH prevalence, mutation penetrance, and test sensitivity are correct (see later for details of these), the probability that the detected variant has actually caused the hyperLDL-C in that individual is calculated to be just 53.55%. For there to be at least 95% probability that a variant has caused hyperLDL-C in such an individual, there needs to be at least 99.94% certainty that the variant is pathogenic. For a 99% probability, the figure is closer to 99.99% certainty. On this basis, it is clearly arguable that the quantity of evidence required to achieve a 'pathogenic' (and perhaps 'likely pathogenic' as well) classification for variants in FH-related genes should be upwardly adjusted from those set out in current general variant classification guidelines.

How to Use HyperLDL-C Sourcery

To use HyperLDL-C Sourcery, users should input / select values into all of the yellow boxes. The 'Default Parameters' at the bottom of the program should be checked / amended first to ensure the user agree with them. The parameters at the top of the program should then be addressed.

Clinical Status: Select whether the individual in question is unaffected, has hyperLDL-C, or hyperLDL-C with tendon xanthoma.

Family History: Choose whether to input or not a ratio of probabilities for a given family history under the assumptions of there being a monogenic mutation that causes the phenotype vs there being no monogenic mutation (i.e. multifactorial or polygenic risks only). For example, 1.5:1 indicates that the family history observed is 1.5× more likely if there is a familial FH-causing mutation than if there is no such mutation. This may be difficult to calculate manually, hence it might be best to choose 'No' for now or insert the ratio as 1:1. I may develop a separate calculator to make it easier to use this input or include some examples in this guide at a later date.

SNP Score population decile: Select 1 – 10. The SNP score decile can be calculated using the LDL-C SNP Scorer program (link) if required.

Genetic Testing Status: Select whether the individual is untested, or tested but no potentially pathogenic variant detected, or tested and a potentially pathogenic variant detected.

Prob Variant is Pathogenic: Insert here a % of >0-100% of how confident you are that the variant detected is pathogenic. This may entail some level of subjectivity. For example a rare variant of uncertain significance with no published cases, segregation, or functional data you might enter as 50%. The new ACGS guidelines suggest that variants that are classed as 'likely pathogenic' should have at least 90% certainty of being pathogenic, and variants classed as 'pathogenic' should have at least 99% certainty of being pathogenic. So, for example, if a variant only just meets criteria for a 'pathogenic' classification, you might enter 99%. If the evidence exceeds the minimum then a figure closer to 100% should be entered.

Default Parameters

I have here selected default values / settings for the certain parameters that the program uses for its calculations, based on what I think are sensible figures from what I have understood from the literature. However, I am no expert on the genetics of FH; hence, users are free to adjust these figures / setting if they disagree with the default values.

Prevalence of FH: 1 in 250 is default. Note that this is the prevalence of the clinical presentation of FH; it should not be interpreted as the proportion of individuals in a population that have an FH mutation.

Clinical Sensitivity of Testing: The default is 90%. In other words, routine genetic testing of the *APOB*, *APOE*, *LDLR*, *LDLRAP1*, and *PCSK9* gene is assumed to detect 90% of FH-causative mutations. A higher figure might be appropriate if all deep intronic regions are sequenced. Or a lower figure might be appropriate if the testing method is not sensitive to copy number variants.

Popn Average Mutation Penetrance: This is set at 50% by default. Although different variants will surely have different penetrance and this will vary depending on the individual, this should be interpreted as the 'average penetrance of all FH-causative variants at a population level'. So if set at 50%, the assumption is that 'half of FH-causative mutation carriers in the population develop hyperLDL-C'. It should be noted that the program determines the prior probability of a random individual having an FH-causative mutation by considering both the 'prevalence of FH' and 'mutation penetrance' such that the prior probability is prevalence / penetrance. Thus, the lower the value of penetrance, the higher the prior probability of an individual having an FH-causative variant. Thus, it may seem counter-intuitive, but if users play with the figures for mutation penetrance they will see that the chance that an individual's hyperLDL-C is caused by an FH-causative mutation is higher if the average mutation penetrance is lower, and this is because of the effect on prior probability. This is only true, however, if the detect variant has <100% certainty of being pathogenic. It should be noted that this program is based on the definition of multi-factorial hyperLDL-C as the presence of hyperLDL-C in an individual that does not have an FH-causing mutation (regardless of its penetrance). And also on the definition of FH as being the presence of hyperLDL-C in an individual with an FH-causing mutation (even if its penetrance is very low). The takeaway message is that users should think about what is the difference between an 'FH-causing mutation' and a 'genetic risk factor'? The difference is entirely down to penetrance, and so users should avoid inputting low penetrance values. Remember, hyperLDL-C is an extremely common condition. If it is assumed that it affects 30% of individuals without an FH-causing mutation, then in theory any mutation that has a penetrance of <30% could be considered protective (rather than causative; this is dependent on how the inputted penetrance figure is used, as described in the next paragraph).

Mutation Penetrance SNP Score Adjustment: The probability that an individual with a particular mutation will develop hyperLDL-C will vary according to their SNP score (or polygenic risk). This parameter determines the method by which that probability is calculated. There is no definitely correct way of calculating this so I have allowed three options as follows. I have also included a table in Appendix 1 to demonstrate how the calculated risk of developing hyperLDL-C differs according to the mutation penetrance value and SNP score adjustment method specified.

- **None:** This is the default option. If this is selected, then the probability is simply equal to the mutation penetrance value (in other words, the probability of being affected is the same regardless of the SNP score decile). Users should think carefully about setting low penetrance values, as this could effectively be stating the mutation to be 'protective' (i.e. mutation penetrance could be less than the risk associated with the specified SNP score decile). It should be noted that this option is selected as default simply because it is the easiest method to understand. However, it may not concord with reality, as research has shown that individuals with FH due to a pathogenic mutation generally have a higher SNP score than control individuals, which suggests that their polygenic risk is also contributing to their hyperLDL-C presentation to some extent, and this option ignores this possibility.
- **PADA:** stands for 'Population Average Decile Adjusted'. If the average mutation penetrance is set at 50% then the PADA method adjusts (by a multiplication factor) the risks of developing hyperLDL-C assigned to each SNP score population decile so that the average across the deciles is 50%. Using this method means the mutation will have a higher apparent penetrance in an individual with SNP score in decile 10 than in an individual with SNP score in decile 1. However, if the mutation penetrance value is set to less than the average polygenic risk across the population deciles (which is ~30%), then the mutation effectively becomes a 'protective variant' (which should be avoided). It should also be noted that the adjusted risk for each decile is capped at 1 (i.e. 100% penetrance, since penetrance cannot be higher). If upper decile risks are capped at 1 (e.g. if a penetrance value of >72% is specified), then the lower deciles are adjusted again to ensure that the average penetrance across the deciles is equal to the value specified. It should also be noted that using the PADA method with lower

penetrance values (e.g. <73%) means that the SNP score decile will not affect the posterior probability of the phenotype being caused by an FH-causal mutation. This is because, under this method, the penetrance of a mutation is defined as $p = rx$ (where p = mutation penetrance, r = risk of hyperLDL-C according to SNP score decile, and x = a constant multiplication factor).

- **IA:** stands for 'Independent Additive'. If this is selected, the chance of an individual with a pathogenic variant developing hyperLDL-C due the mutation (i.e. the penetrance value) is added to the chance they will develop hyperLDL-C due to their polygenic risk given that the mutation is non-penetrant. So if their polygenic risk is, say 30%, and the mutation penetrance is set to 55% then the chance of being affected is calculated to be 0.55 + (0.45x0.3) = 0.685 (or 68.5%). The IA method may intuitively seem less likely (than the PADA method) to reflect the pathophysiology in reality (the chance of being affected is probably a product of the mutation and polygenic risk working together rather than independently). However, it still reflects the reality that an individual with an FH-causing mutation with polygenic risk in a higher decile will have higher chance of being affected than an individual with polygenic risk in a lower decile. It also has the added benefit over the PADA method that even if the mutation penetrance is set very low, the mutation will still increase risk of being affected (i.e. a protective mutation is not possible with this method).

FH Patients with Tendon Xanthoma: Default value is 30%. This is a nominal value to represent the probability that an individual with FH will develop tendon xanthoma. The figure of 30% might not be particularly accurate (it is difficult to choose a figure given age dependence etc.), but the main point is that the probability of developing tendon xanthoma is significantly higher in individuals with FH than individuals that do not have FH.

Prevalence of FH-Unrelated Tendon Xanthoma: This is a nominal value to represent the probability that an individual without FH will develop tendon xanthoma (e.g. the prevalence of cerebrotendinous xanthomatosis). The main point is that the probability of developing tendon xanthoma is significantly lower in individuals without FH. The default value is set at 1 in 25,000.

***P*(LDL>4.9) in WHII Controls Data Values:** This determines whether measured or predicted data is used with respect to risk of developing hyperLDL-C according to SNP score decile as described in Table 3 of Talmud *et al*. (2013). Default is 'predicted' data.

References

[1] Talmud *et al*. (2013). *Lancet*, 381(9874): 1293-1301. PMID: 23433573.
[2] Futema et al. (2018). *Atherosclerosis*, 277: 457-463. PMID: 30270085.
[3] Futema *et al*. (2015). *Clin Chem*, 61(1): 231-238. PMID: 25414277.
[4] Ellard *et al*. (2020). ACGS Best Practice Guidelines for Variant Classification in Rare Disease 2020. URL: https://www.acgs.uk.com/quality/best-practice-guidelines/

Appendix 1

SNP Score Decile	Talmud 2013 (Predicted)	None			PADA			IA		
		10%	50%	90%	10%	50%	90%	10%	50%	90%
1	16.89	10.00	50.00	90.00	5.39	26.95	53.17	25.2	58.44	91.69
2	22.85	10.00	50.00	90.00	7.29	36.47	71.93	30.56	61.42	92.28
3	27.15	10.00	50.00	90.00	8.67	43.34	85.48	34.44	63.58	92.72
4	29.04	10.00	50.00	90.00	9.27	46.35	91.43	36.14	64.52	92.9
5	31.13	10.00	50.00	90.00	9.94	49.68	97.99	38.01	65.56	93.11
6	33.11	10.00	50.00	90.00	10.57	52.85	100.00	39.8	66.56	93.31
7	35.10	10.00	50.00	90.00	11.2	56.02	100.00	41.59	67.55	93.51
8	35.88	10.00	50.00	90.00	11.45	57.27	100.00	42.29	67.94	93.59
9	39.07	10.00	50.00	90.00	12.47	62.36	100.00	45.17	69.54	93.91
10	43.05	10.00	50.00	90.00	13.74	68.71	100.00	48.74	71.52	94.3

This table indicates the probability (expressed as %) that an individual will develop hyperLDL-C that HyperLDL-C Sourcery uses when different methods of 'Mutation Penetrance SNP Score Adjustment' are used and when 'Popn Average Mutation Penetrance' is set to 10%, 50%, or 90%. Also indicated in the second column is the risk associated with each SNP score decile (predicted values in absence of mutation), as determined by Talmud *et al.* (2013).[1]