Prediction of complex multifactorial disease
Comparing family history and genetics

C.B. Do1, J.M. Macpherson1, D.A. Hinds1, B.T. Naughton1, U. Francke1,2, N. Eriksson1.
123andMe, Inc, Mountain View, CA; 2Stanford University School of Medicine, Stanford, CA.

Introduction

Although the use of family history and SNP-based risk assessment is well understood for simple Mendelian disorders, to date, little is known regarding the relative performance of these methods for complex polygenic diseases.

In this study, we used the standard liability-threshold model from quantitative genetic theory to analyze the influence of disease prevalence (K) and heritability ($h^2$) on the predictive accuracy of family history and SNP-based models.

Results

The following graph depicts the relationship between predictive accuracy as measured by AUC (vertical axis) and the proportion of genetic variance explained by known SNP associations (horizontal axis) for a disease of moderate heritability ($h^2 = 0.4$) and prevalence ($K = 0.1$) in a large 3-generation pedigree:

![Graph showing relationship between AUC and proportion of genetic variance explained](image)

Understanding the graph

1. The solid blue line represents the performance of SNP-based risk assessments.
2. The solid black line represents the predictive performance of an ideal family history-based risk assessment.
3. The dotted black line represents the predictive performance of a family history-based model that only distinguishes between 0, 1, or >1 first-degree relatives with disease.
4. The intersection points labeled with percentages show the proportion of genetic variance explained at which SNP-based models outperform family history.

Observations (see Figure 1)

1. Family history is most discriminative for common conditions (as the chance of having an affected relative is higher), whereas SNP-based models maintain high discriminative power for rarer conditions provided that enough of the genetic variance is explained.
2. In most cases, the bulk of predictive accuracy of family history can be captured by a model taking into account only first-degree relatives.
3. For diseases with <1% prevalence, the crossover point occurs between 1-4% of the variance explained. This is well within the detection limits of current GWAS, and in fact, a large fraction of the diseases studied to date have already crossed this line.

Conclusion

Limitations of our model

- No highly penetrant mutations
- No age-of-onset information
- No non-additive effects
- Use of lifetime risk only
- Recall biases for family history in practice
- Difficulty of obtaining heritability estimates

Take-home messages

1. The relative performance of family history and SNP-based models at predicting disease risk depends largely on the characteristics of the disease considered.
2. For diseases of low or moderate frequency (<1% prevalence), current SNP-based risk assessments may be significantly more discriminative than family history.