Genetic Risk Scores

Jessica N. Cooke Bailey[1] and Robert P. Igo Jr[1]

[1]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio

The generation of genome-wide variation data has become commonplace. However, the potential for interpretation and application of these data for clinical assessment of outcomes of interest, and prediction of disease risk, is currently not fully realized. Many common, complex diseases now have numerous, well-established "risk" loci, and likely harbor many genetic determinants with effects too small to be detected at genome-wide levels of statistical significance. A simple and intuitive approach for converting genetic data to a predictive measure of disease susceptibility is to aggregate the risk effects of these loci into a single genetic risk score. Here, some common methods and software packages for calculating genetic risk scores, with focus on studies of common, complex diseases, are described. The basic information needed as well as important considerations for constructing genetic risk scores, including specific requirements for phenotypic and genetic data, and limitations in their application is reviewed. © 2016 by John Wiley & Sons, Inc.

Keywords: auc • complex diseases • complex traits • disease prediction • genetic risk score

How to cite this article:
Cooke Bailey, J.N., and Igo, R.P. Jr., 2016. Genetic risk scores. *Curr. Protoc. Hum. Genet.* 91:1.29.1-1.29.9.
doi: 10.1002/cphg.20

DEFINITIONS

A genetic risk score is an estimate of the cumulative contribution of genetic factors to a specific outcome of interest in an individual that takes into account the reported risk alleles. The score may also take into account the reported effect sizes for those alleles and may be normalized by adjusting for the total number of risk alleles and effect sizes evaluated.

The area under the curve (AUC) measures the predictive ability of a receiver operating characteristic (ROC) generated based on the genetic risk scores for a sample of individuals. The AUC is a function of the ability of the risk score to correctly identify the presence (sensitivity) or absence (specificity) of the outcome of interest.

Disease prediction is the ability to determine disease state given any number of known factors including, but not limited to, genetic, environmental, and lifestyle.

Complex traits and complex diseases (see *UNIT 1.9*; Bush and Haines, 2010) have multiple contributing factors, including both genetic variation and environmental factors. They have a demonstrable genetic component but no simple mode of inheritance.

GENETIC RISK SCORES

The purpose of this unit is to provide an overview of the application of genetic risk scores (GRS) while also presenting guidelines for using GRS for disease prediction. While genome-wide association studies, which have become routine over the past decade, have yielded thousands of genetic loci—mostly single nucleotide polymorphisms (SNPs)—associated with common, complex diseases (Beck et al., 2014; Burdett et al., 2016), a great deal of the heritability of such diseases remains to be dissected (Manolio et al., 2009). Translating the influence of these static loci, which, unlike clinical measures, do not change, to predict disease risk is not always straightforward. As a rule, genetic loci contributing to complex, common diseases each contribute varying degrees of risk (or protection). Connecting these

Current Protocols in Human Genetics 1.29.1-1.29.9, October 2016
Published online October 2016 in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cphg.20

effects, along with demographic, lifestyle, and environmental risk factors, to disease risk is a seemingly impossible task, but through classical and recently developed methods of analysis, progress is being made toward making sense of this vast body of genetic data. Herein, methods for translating genetic information into an assessment of disease risk for complex traits whose genetic component cannot be completely explained by one or several genetic loci of large effect are reviewed. Although calculating GRSs is now common, their potential complications and pitfalls are also emphasized.

The purpose of risk scores is two-fold: (1) to predict the likelihood of an individual developing disease (or a particular outcome of interest) based on some amount of available information, usually genetic, clinical, demographic, or a combination, and (2) to estimate the level of predictive power that is captured by associated variants. The goal is to be able to predict whether a person is likely to develop a disease, a reaction to a drug, etc., based on existing genetic information. Predicting a greater proportion of the "risk" for the outcome of interest indicates the level of success of predictors included in the risk score.

The most common approach to evaluate the cumulative effect of many genetic factors with small effect, with or without non-genetic clinical factors, is by means of genetic risk scores (GRS). A GRS can estimate the overall risk a person has for developing an outcome of interest based on their genotypes at variants that have been statistically associated with risk for that outcome. Because an individual's genetic profile is set at birth, and therefore, risk for disease could theoretically be determined prior to (most) environmental exposures (Wray et al., 2010), a great deal of hope has been put into developing these models as an advancement of precision medicine. However, for many complex diseases, it is debatable whether genomic profiling is ready for clinical use (e.g., Jakobsdottir et al., 2009; Cooke Bailey et al., 2016a). Family history is typically seen as a good proxy for genetic risk as it reflects shared genetic and environmental factors and thus is incorporated into clinical history when possible for genetic diseases (reviewed in Wray et al., 2010). The power of family history is limited by family size, disease prevalence, and information available from relatives; further, a positive family history reflects a certain level of disease risk, while a negative family history does not imply the opposite (Wray et al., 2010). One goal of implementing GRS is to improve upon these factors for a more comprehensive and accurate assessment of disease risk beyond what family history can estimate.

GRS allow for the evaluation of contributions by multiple factors to disease development and outcomes, including, but not limited to, disease susceptibility, progression, and response to treatment. GRS can be based solely on available genetic data or can incorporate environmental, phenotypic, and/or demographic information. Published GRS results tend to vary across studies and are dependent upon the population sample evaluated and the true outcome of interest (i.e., progression from early to late stage disease, disease subtypes, or general disease risk). These are important factors to keep in mind when constructing and evaluating GRS for use in a particular sample.

Prediction accuracy of GRS is most often assessed by measuring the area under the receiver operating characteristic (ROC) curve (AUC), an indicator of model accuracy. The AUC compares the rates of true positives (sensitivity) and false positives (1 – specificity) and indicates the overall performance of predictive models (Janssens et al., 2007). Sensitivity, the probability of correctly classifying an affected individual as affected, indicates the ability of the model to correctly predict individuals with the outcome of interest; specificity, the probability of correctly classifying an unaffected individual as unaffected, indicates the ability of the model to accurately screen out individuals without the outcome of interest. The AUC generally varies between 0.5, indicating a model with a predictive power no better than chance, and 1, indicating a perfect model (Janssens et al., 2007). Models are expected to have an AUC > 0.75 for informative screening of individuals who are at increased disease risk (Janssens et al., 2007). The higher the AUC, the more precise the prediction and thus, the greater the potential clinical utility of the combination of factors included in the model. However, there are limitations to consider. There are two components to predictive accuracy: the potential accuracy if all genetic factors were known, determined by the trait heritability, and the correlation between the GRS and true genetic risk, determined by the quantity of genetic data and selection of genetic variants for prediction. The AUC is limited by both of these factors but cannot distinguish between them (Wray and Goddard, 2010).

Instead of calculating the AUC, the predictive measure emphasized here, one can estimate the proportion of trait variability explained by one or more genetic markers. One measure is the population attributable risk (PAR) or population attributable fraction, which, in general, expresses the fraction of cases attributable to a given exposure (Witte et al., 2014). However, the PAR is not additive over multiple markers. In the case of continuous traits, the resulting multiple R^2 from the linear regression analysis quantifies the trait variance accounted for by the predictors. This may be approximated for binary traits by a pseudo-R^2 measure, derived from the likelihood under the models with and without genetic predictors (Menard, 2000; Witte et al., 2014).

As a practical example of use, GRS have been constructed for many common, complex diseases including age-related macular degeneration (AMD). AMD is one of the few common and complex diseases for which large-effect loci have been identified; the *CFH* and *ARMS2/HTRA1* loci, along with ten other loci, predict disease risk with an AUC of 0.74 and it has been shown that adding other low-effect SNPs to a GRS for AMD does not significantly strengthen the model (Fritsche et al., 2013). More recent analyses support 52 independently associated AMD variants in 34 loci that contribute varying degrees of risk for advanced AMD and together explain 27.2% of overall disease variability and more than half of the heritability (Fritsche et al., 2016).

An alternative measure of predictive accuracy is based on positive predictive value: the probability that an individual labeled "high-risk" is truly affected. One way to estimate this is to subdivide the sample into a small number of bins, or quantiles, of the GRS, and to estimate predictive value by the proportion of cases and controls in each bin. Fritsche et al. (2016) divided samples into deciles of genetic risk based on the calculated GRS and found that individuals in the highest decile have a 44-fold increased risk for advanced AMD versus individuals in the lowest decile; however, only 22.7% of individuals in the highest decile of genetic risk are predicted to actually have AMD. This highlights a challenge of GRS: while ideally a GRS would easily distinguish between disease-susceptible individuals, the extant understanding of disease pathophysiology and available genetic information must expand concurrently with the increase in genetic disease risk knowledge to inform GRS and in turn enlighten the process of genetic disease prediction.

It is important to note that significance and effect size of optimal genetic variants from a risk-based approach, such as a GWAS by logistic regression, and from a classification-based approach, which focuses on accuracy of distinguishing cases from controls, are not always strongly correlated: strong risk markers might be poor classifiers while good classifiers might not show genome-wide levels of significance (Pepe et al., 2004; Jakobsdottir et al., 2009). For example, Jakobsdottir and colleagues (2009) constructed a three-gene model for AMD risk including variants in *CFH* and *LOC387715* (*ARMS2/HTRA1*), each of which had achieved genome-wide significance in a logistic regression model ($P = 9.1 \times 10^{-13}$ and 2.3×10^{-13}, respectively), as well as a variant in *C2* ($P = 1.3 \times 10^{-3}$). This model yielded an AUC of 0.79, exceptionally high for a complex trait. Despite strong association, they calculated that a risk score threshold with a sensitivity of 74% (i.e., correctly identified 74% of cases) also wrongly classified 31% of controls as cases.

GRS METHODS

Many approaches are available to generate GRS. The popular PLINK program for analysis of genome-wide genetic data (Purcell et al., 2007) contains a Profile scoring method for generating a marker-based GRS. PLINK generates risk scores ("profile scoring") by means of the –score function, when provided binary-format (.bed,.bim,.fam) files from a genetic dataset and a myprofile.raw file with the SNP ID, reference allele and score (or weight) for each allele (Fig. 1.29.1), for example:

```
plink --bfile mydata --score
myprofile.raw
```

This command will generate a file named plink.profile with the following information for each individual: family ID, individual ID, phenotype, number of non-missing SNPs used for scoring, number of named alleles, and total risk score for the individual, the sum of the number of reference alleles (0, 1, or 2) at each SNP multiplied by the (user-defined) 'score' for that SNP. A logical choice for the score is the estimated log odds ratio associated with each copy of the minor allele of each tested marker from a logistic regression analysis, as described below.

If a genotype in the score is missing for an individual, the program default is for the

A

CHR	SNP	BP	A1	TEST	NMISS	OR	STAT	P
9	rs1001	1475298	A	ADD	3000	0.6925	-6.566	5.184e-11
9	rs1001	1475298	A	SEX	3000	0.9648	-0.3348	0.7378
11	rs1002	78002095	G	ADD	3000	2.679	8.051	8.226e-16
11	rs1002	78002095	G	SEX	3000	0.9668	-0.3149	0.7528
17	rs1003	32214410	C	ADD	3000	1.945	4.954	7.271e-07
17	rs1003	32214410	C	SEX	3000	0.9717	-0.2673	0.7892

B

rs1001	A	-0.3674
rs1002	G	0.9854
rs1003	C	0.6653

Figure 1.29.1 (**A**) Raw output from the –logistic function in PLINK. The column OR contains the odds ratio per copy of allele A1. (**B**) Corresponding PLINK myprofile.raw file, with marker name, reference allele (A1), and ln(OR) from the test of additive effects (lines labeled ADD under TEST).

value to be imputed based on the sample allele frequency. To change this, the `--score-no-mean-imputation` flag should be used.

A straightforward method to evaluate the GRS is to choose a number k of independent genetic variants with strong (i.e., genome-wide significant in other studies or datasets) association as risk predictors, and to calculate the GRS as the sum of the effect estimates (log odds ratios) β_i from a logistic regression analysis with additive genetic effect, multiplied by the number of risk alleles N_i for each locus:

$$GRS = \sum_{i=1}^{k} \beta_i N_i$$

This equation is directly related to the estimated risk for disease from the logistic model, through the logit (log odds) function, and therefore is statistically intuitive.

PLINK's `--logistic` function for logistic regression covariates reports the odds ratio per copy of the minor allele at each marker (Fig. 1.29.1A, column 7). The natural logarithm of this odds ratio is the value β_i in the *GRS* equation above, and is entered as the score for each marker in the myprofile.raw file for PLINK (Fig. 1.29.1B).

A limitation of PLINK's risk score method is that it calculates the average GRS per non-missing marker, whereas the typical GRS is the sum over all markers. In the case of no missing data, these scores are equivalent; however, a large proportion of missing genotypes may result in atypical values of the average GRS. If this is a concern, it is worth considering removing individuals without complete data at all markers to be evaluated in the GRS.

From this stage, various programs and approaches can be used; many available programs are listed in Table 1.29.1. These require varying degrees of analytical and programming experience. For example, GPRS (Genomic Profile Risk Score analyses version 0.9.0) is an online application belonging to the "Shiny Apps" suite that allows for direct upload of PLINK-generated risk score files (along with covariate files); this program generates the AUC, the variance explained by risk score, and plots results very quickly from an internet browser interface. Various other programs require access to more expansive computational resources.

Focus in this unit is not placed on polygenic risk scores, which evaluate many genetic loci, not only those significantly associated with a disease. The R package, PRSice, generates polygenic risk scores (Euesden et al., 2015) (Table 1.29.1).

CRITICAL PARAMETERS AND COMPLICATING FACTORS

The discriminatory power of the model to determine an individual's "risk" is dependent upon the factors in the model and their contribution to that risk. Several questions regarding factors included in a GRS must be considered. Are these factors useful predictors in only the population in which they were first detected? Are they consistent across ethnic groups? Does the linkage disequilibrium structure vary between the discovery population and the test population and will this cause differences in model predictability? Are the risk alleles coded consistently across datasets? Are the same DNA strands being evaluated? Critical aspects to consider when constructing a GRS are listed in Table 1.29.2.

Choosing an outcome of interest can seem straightforward; however, for common, traits with many genetic determinants, it does

Table 1.29.1 Available Software Programs for Generating Genetic Risk Scores (GRS)

Program/ package[a]	Functions	Remarks	Website or publication
PredictABEL (R)	Calculate GRS	Part of GenABEL package; computes weighted and unweighted GRS	*http://www.genabel.org/PredictABEL/riskScore.html*
PRSice (R)	Optimize polygenic risk scores by p-value threshold	Uses PLINK file types and program code	*http://PRSice.info*
pROC (R)	ROC, AUC, pAUC, compare AUCs	Requires pre-calculated GRS	*https://cran.r-project.org/web/packages/pROC/index.html*
GPRS	ROC, AUC, OR by risk score deciles	Web-based; takes PLINK risk score results as input	*http://cnsgenomics.com/shiny/GPRsS/*
PLINK	Mean genetic risk per marker	Allows missing data	*http://pngu.mgh.harvard.edu/~purcell/plink/profile.shtml*
ShinyApps	Genetic interpretation of AUC; heritability explained by single loci	Web-based	*http://cnsgenomics.com/software.html*
GTX (R)	GRS from meta-analysis results; calculate proportion of variance explained by GRS	grs.summary requires weights for GRS	*http://cran.r-project.org/web/packages/gtx*

[a](R), package or function in the R statistical programming language.

Table 1.29.2 Considerations for Generating a Genetic Risk Score

Factor	Considerations
Outcome of interest	Pleiotropy definition of complex phenotype (component phenotypes); clinical covariates
Loci	P-value threshold (polygenic versus highly significant loci only); imputation type/quality; linkage disequilibrium; mode of inheritance; agreement of DNA strand between training and test data; reference (risk) allele coding
Sample	Population (ethnic background; admixture); demographic makeup (age, sex, socioeconomic status, etc.)
Weighting	log(OR) from logistic regression; total number of risk variants

require understanding the complexity in the genetic component and in the definition of multifactorial traits, and how these two factors interact. As an example, the primary open-angle glaucoma (POAG) phenotype may be defined in many different ways: by cup-to-disc ratio (CDR), intraocular pressure (IOP), visual field (VF) loss, or by some combination of these (Weinreb et al., 2014). Fourteen loci have been identified as genetic risk modifiers of disease (Wiggs, 2015; Cooke Bailey et al., 2016b). The genetic factors identified in studies of POAG and its component phenotypes (such as IOP, CDR, VF changes)

overlap but are not identical. For example, *GAS7* is associated with intraocular pressure as well as overall POAG, *CDKN2B-AS1* is associated with POAG and normal tension glaucoma, and other optic nerve parameters (Wiggs et al., 2012; Wiggs, 2015). Choosing variants relevant to constructing a POAG risk score, for example, would require careful consideration and examination of this information.

Further considerations for choosing loci of interest to include in a GRS are *P*-value threshold, imputation type and quality, and linkage disequilibrium (LD). Goldstein et al. (2015) examined the effect of *P*-value threshold, LD, imputation quality, and type on GRS and determined that for prediction of coronary heart disease risk in 10 years using a GRS, the most relevant factor for choosing an optimally-predictive GRS was SNP *P*-value. While their specific method cannot generally be repeated for other diseases, it supports the use of an empirical approach to determining GRS for particular complex diseases to evaluate which genetic markers are most relevant to that disease.

Mode of inheritance (see *UNIT 1.4*; Dueker and Pericak-Vance, 2014) must also be taken into consideration. The equation described above assumes a multiplicative odds model (additive on the log odds scale, as outlined in the above example) for all risk variants. Whereas the true contribution of genetic variants may not follow this model, some prediction accuracy may be lost, despite dominance effects likely not substantially contributing to sporadic, complex diseases (Hill et al., 2008)

When constructing a GRS, DNA strand and risk allele coding are crucial to keep consistent across studies so as to ensure validity and transferability of GRS (see *UNIT 1.19*; Turner et al., 2011). It is also imperative to note that the risk allele is not necessarily always the minor or reference allele and this is a key aspect to keep consistent between studies, since a protective effect of one allele implies risk at the other allele. Further, effect estimates should be estimated in the largest group possible, but not in the group being tested (to prevent model over-fitting).

Special care should be taken when constructing GRS including SNPs with strand-ambiguous alleles A/T or C/G, because the reference allele is not obviously comparable between samples typed on different DNA strands (see *UNIT 1.19*; Turner et al., 2011). Ensure that the genotypes from new samples are relative to the same DNA strand (e.g., the Illumina TOP or dbSNP + strand), or, if strand information is absent, compare allele frequencies across samples to identify the common minor allele. In the latter case, A/T and C/G SNPs with minor allele frequency >0.4 may not be usable, owing to uncertainty in determining which allele is the minor allele.

Most common variants associated with complex disease confer at best modest effect sizes (e.g., odds ratios in the 1.05 to 1.50 range). Most GWAS have been performed, with increasing frequency in massive datasets, using samples primarily of European ancestry. Application of a GRS derived from these data to smaller or ethnically diverse samples requires careful planning and analysis. The effect sizes and even the risk-associated alleles may be different (e.g., allelic heterogeneity). In similar populations, the true heritability accounts for the possibility of allelic heterogeneity; however, allelic heterogeneity will reduce prediction accuracy if the discovery and testing cohorts are not of the same genetic background. In a sample of Amish individuals, Hoffman et al. (2014) calculated GRS generated from 19 AMD loci reported by Fritsche et al. (2013) and compared the distribution to that in a non-Amish, European-American sample. They found that overall the Amish GRSs were significantly lower in the Amish case and control groups than in the non-Amish sample, despite similar AMD prevalence, thus providing some insight into the genetic architecture of AMD in the Amish. Similarly, studies in African-Americans and Mexican-Americans have only detected association at the *ARMS2* A69S variant (Spencer et al., 2012; Restrepo et al., 2014).

Cooke et al. (2012) previously evaluated a GRS that included 17 known type 2 diabetes (T2D) risk loci in an African American sample and showed that while the risk allele load was higher in African-American cases compared to controls, the high-effect *TCF7L2* rs7903146 risk allele accounted for all increased risk in African-Americans. Expanding on this work, Keaton et al. (2014) genotyped additional T2D SNPs to bring the total to 43 SNPs and confirmed that there are ethnic-specific differences in the genetic architecture of T2D when comparing African-Americans to European-Americans. Unfortunately, many of the GRS scores modeled to date have only been evaluated in European-Caucasian individuals. This is reflective of the widespread genetic data in this population and the (in comparison) relative shortage

of data in individuals of other ancestries (e.g., Bustamante et al., 2011).

ACKNOWLEDGMENTS

J.N.C.B. is supported in part by a PhRMA Informatics Fellowship.

LITERATURE CITED

Beck, T., Hastings, R.K., Gollapudi, S., Free, R.C., and Brookes, A.J. 2014. GWAS Central: A comprehensive resource for the comparison and interrogation of genome-wide association studies. *Eur. J. Hum. Genet.* 22:949-952. doi: 10.1038/ejhg.2013.274.

Burdett, T., Hall, P.N., Hastings, E., Hindorff, L.A., Junkins, H.A., Klemm, A.K., MacArthur, J., Manolio, T.A., Morales, J., Parkinson, H., and Welter, D. 2016. The NHGRI-EBI catalog of published genome-wide association studies. *http://www.ebi.ac.uk/gwas/*

Bush, W.S. and Haines, J. 2010. Overview of linkage analysis in complex traits. *Curr. Protoc. Hum. Genet.* 64:1.9.1-1.9.18. doi: 10.1002/0471142905.hg0109s64.

Bustamante, C.D., Burchard, E.G., and De la Vega, F.M. 2011. Genomics for the world. *Nature* 475:163-165. doi: 10.1038/475163a.

Cooke, J.N., Ng, M.C., Palmer, N.D., An, S.S., Hester, J.M., Freedman, B.I., Langefeld, C.D., and Bowden, D.W. 2012. Genetic risk assessment of type 2 diabetes-associated polymorphisms in African Americans. *Diabetes Care* 35:287-292. doi: 10.2337/dc11-0957.

Cooke Bailey, J.N., Hoffman, J.D., Sardell, R.J., Scott, W.K., Pericak-Vance, M.A., and Haines, J.L. 2016a. The application of genetic risk scores in age-related macular degeneration: A review. *J. Clin. Med.* 5:pii: E31. doi: 10.3390/jcm5030031

Cooke Bailey, J.N., Sobrin, L., Pericak-Vance, M.A., Haines, J.L., Hammond, C.J., and Wiggs, J.L. 2013. Advances in the genomics of common eye diseases. *Hum. Mol. Genet.* 22:R59-R65. doi: 10.1093/hmg/ddt396.

Cooke Bailey, J.N., Loomis, S.J., Kang, J.H., Allingham, R.R., Gharahkhani, P., Khor, C.C., Burdon, K.P., Aschard, H., Chasman, D.I., Igo, R.P., Jr., Hysi, P.G., Glastonbury, C.A., Ashley-Koch, A., Brilliant, M., Brown, A.A., Budenz, D.L., Buil, A., Cheng, C.Y., Choi, H., Christen, W.G., Curhan, G., De, V.I., Fingert, J.H., Foster, P.J., Fuchs, C., Gaasterland, D., Gaasterland, T., Hewitt, A.W., Hu, F., Hunter, D.J., Khawaja, A.P., Lee, R.K., Li, Z., Lichter, P.R., Mackey, D.A., McGuffin, P., Mitchell, P., Moroi, S.E., Perera, S.A., Pepper, K.W., Qi, Q., Realini, T., Richards, J.E., Ridker, P.M., Rimm, E., Ritch, R., Ritchie, M., Schuman, J.S., Scott, W.K., Singh, K., Sit, A.J., Song, Y.E., Tamimi, R.M., Topouzis, F., Viswanathan, A.C., Verma, S.S., Vollrath, D., Wang, J.J., Weisschuh, N., Wissinger, B., Wollstein, G., Wong, T.Y., Yaspan, B.L., Zack, D.J., Zhang, K., Study, E.N., Weinreb, R.N., Pericak-Vance, M.A., Small, K., Hammond, C.J., Aung, T., Liu, Y., Vithana, E.N., MacGregor, S., Craig, J.E., Kraft, P., Howell, G., Hauser, M.A., Pasquale, L.R., Haines, J.L., and Wiggs, J.L. 2016b. Genome-wide association analysis identifies TXNRD2, ATXN2 and FOXC1 as susceptibility loci for primary open-angle glaucoma. *Nat. Genet.* 48:189-194. doi: 10.1038/ng.3482.

Dueker, N. D. and Pericak-Vance, M. A. 2014. Analysis of genetic linkage data for mendelian traits. *Curr. Protoc. Hum. Genet.* 83:1.4.1-1.4.31.doi: 10.1002/0471142905.hg0104s83

Euesden, J., Lewis, C.M., and O'Reilly, P.F. 2015. PRSice: Polygenic risk score software. *Bioinformatics* 31:1466-1468. doi: 10.1093/bioinformatics/btu848.

Fritsche, L.G., Chen, W., Schu, M., Yaspan, B.L., Yu, Y., Thorleifsson, G., Zack, D.J., Arakawa, S., Cipriani, V., Ripke, S., Igo, R.P. Jr., Buitendijk, G.H., Sim, X., Weeks, D.E., Guymer, R.H., Merriam, J.E., Francis, P.J., Hannum, G., Agarwal, A., Armbrecht, A. M., Audo, I., Aung, T., Barile, G.R., Benchaboune, M., Bird, A.C., Bishop, P.N., Branham, K.E., Brooks, M., Brucker, A.J., Cade, W.H., Cain, M.S., Campochiaro, P.A., Chan, C.C., Cheng, C.Y., Chew, E.Y., Chin, K.A., Chowers, I., Clayton, D.G., Cojocaru, R., Conley, Y.P., Cornes, B.K., Daly, M.J., Dhillon, B., Edwards, A.O., Evangelou, E., Fagerness, J., Ferreyra, H.A., Friedman, J.S., Geirsdottir, A., George, R.J., Gieger, C., Gupta, N., Hagstrom, S.A., Harding, S.P., Haritoglou, C., Heckenlively, J.R., Holz, F.G., Hughes, G., Ioannidis, J.P., Ishibashi, T., Joseph, P., Jun, G., Kamatani, Y., Katsanis, N., Keilhauer, N., Khan, J.C., Kim, I.K., Kiyohara, Y., Klein, B.E., Klein, R., Kovach, J.L., Kozak, I., Lee, C.J., Lee, K.E., Lichtner, P., Lotery, A.J., Meitinger, T., Mitchell, P., Mohand-Said, S., Moore, A.T., Morgan, D.J., Morrison, M.A., Myers, C.E., Naj, A.C., Nakamura, Y., Okada, Y., Orlin, A., Ortube, M.C., Othman, M.I., Pappas, C., Park, K.H., Pauer, G.J., Peachey, N.S., Poch, O., Priya, R.R., Reynolds, R., Richardson, A.J., Ripp, R., Rudolph, G., Ryu, E., Sahel, J.A., Schaumberg, D.A., Scholl, H.P., Schwartz, S.G., Scott, W.K., Shahid, H., Sigurdsson, H., Silvestri, G., Sivakumaran, T.A., Smith, R.T., Sobrin, L., Souied, E.H., Stambolian, D.E., Stefansson, H., Sturgill-Short, G.M., Takahashi, A., Tosakulwong, N., Truitt, B.J., Tsironi, E.E., Uitterlinden, A.G., van Duijn, C.M., Vijaya, L., Vingerling, J.R., Vithana, E.N., Webster, A.R., Wichmann, H.E., Winkler, T.W., Wong, T.Y., Wright, A.F., Zelenika, D., Zhang, M., Zhao, L., Zhang, K., Klein, M.L., Hageman, G.S., Lathrop, G.M., Stefansson, K., Allikmets, R., Baird, P.N., Gorin, M.B., Wang, J.J., Klaver, C.C., Seddon, J.M., Pericak-Vance, M.A., Iyengar, S.K., Yates, J.R., Swaroop, A., Weber, B.H., Kubo, M., DeAngelis, M.M., Leveillard, T., Thorsteinsdottir, U., Haines, J.L., Farrer, L.A., Heid, I.M., and Abecasis, G.R. 2013. Seven new loci associated with age-related macular degeneration. *Nat. Genet.* 45:433-439. doi: 10.1038/ng.2578.

Fritsche, L.G., Igl, W., Bailey, J.N., Grassmann, F., Sengupta, S., Bragg-Gresham, J.L., Burdon,

K.P., Hebbring, S.J., Wen, C., Gorski, M., Kim, I.K., Cho, D., Zack, D., Souied, E., Scholl, H.P., Bala, E., Lee, K.E., Hunter, D.J., Sardell, R.J., Mitchell, P., Merriam, J.E., Cipriani, V., Hoffman, J.D., Schick, T., Lechanteur, Y.T., Guymer, R.H., Johnson, M.P., Jiang, Y., Stanton, C.M., Buitendijk, G.H., Zhan, X., Kwong, A.M., Boleda, A., Brooks, M., Gieser, L., Ratnapriya, R., Branham, K.E., Foerster, J.R., Heckenlively, J.R., Othman, M.I., Vote, B.J., Liang, H.H., Souzeau, E., McAllister, I.L., Isaacs, T., Hall, J., Lake, S., Mackey, D.A., Constable, I.J., Craig, J.E., Kitchner, T.E., Yang, Z., Su, Z., Luo, H., Chen, D., Ouyang, H., Flagg, K., Lin, D., Mao, G., Ferreyra, H., Stark, K., von Strachwitz, C.N., Wolf, A., Brandl, C., Rudolph, G., Olden, M., Morrison, M.A., Morgan, D.J., Schu, M., Ahn, J., Silvestri, G., Tsironi, E.E., Park, K.H., Farrer, L.A., Orlin, A., Brucker, A., Li, M., Curcio, C.A., Mohand-Said, S., Sahel, J.A., Audo, I., Benchaboune, M., Cree, A.J., Rennie, C.A., Goverdhan, S.V., Grunin, M., Hagbi-Levi, S., Campochiaro, P., Katsanis, N., Holz, F.G., Blond, F., Blanche, H., Deleuze, J.F., Igo, R.P., Jr., Truitt, B., Peachey, N.S., Meuer, S.M., Myers, C.E., Moore, E.L., Klein, R., Hauser, M.A., Postel, E.A., Courtenay, M.D., Schwartz, S.G., Kovach, J.L., Scott, W.K., Liew, G., Tan, A.G., Gopinath, B., Merriam, J.C., Smith, R.T., Khan, J.C., Shahid, H., Moore, A.T., McGrath, J.A., Laux, R., Brantley, M.A., Jr., Agarwal, A., Ersoy, L., Caramoy, A., Langmann, T., Saksens, N.T., de Jong, E.K., Hoyng, C.B., Cain, M.S., Richardson, A.J., Martin, T.M., Blangero, J., Weeks, D.E., Dhillon, B., van Duijn, C.M., Doheny, K.F., Romm, J., Klaver, C.C., Hayward, C., Gorin, M.B., Klein, M.L., Baird, P.N., den Hollander, A.I., Fauser, S., Yates, J.R., Allikmets, R., Wang, J.J., Schaumberg, D.A., Klein, B.E., Hagstrom, S.A., Chowers, I., Lotery, A.J., Leveillard, T., Zhang, K., Brilliant, M.H., Hewitt, A.W., Swaroop, A., Chew, E.Y., Pericak-Vance, M.A., DeAngelis, M., Stambolian, D., Haines, J.L., Iyengar, S.K., Weber, B.H., Abecasis, G.R., and Heid, I.M. 2016. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat. Genet.* 48:134-143. doi: 10.1038/ng.3448.

Goldstein, B.A., Yang, L., Salfati, E., and Assimes, T.L. 2015. Contemporary considerations for constructing a genetic risk score: An empirical approach. *Genet. Epidemiol.* 39:439-445. doi: 10.1002/gepi.21912.

Hill, W.G., Goddard, M.E., and Visscher, P.M. 2008. Data and theory point to mainly additive genetic variance for complex traits. *PLoS. Genet.* 4:e1000008. doi: 10.1371/journal.pgen.1000008.

Hoffman, J.D., et al. (2014). Rare complement factor H variant associated with age-related macular degeneration in the Amish. *Invest Ophthalmol. Vis. Sci*. 55:4455-4460.

Jakobsdottir, J., Gorin, M.B., Conley, Y.P., Ferrell, R.E., and Weeks, D.E. 2009. Interpretation of genetic association studies: Markers with replicated highly significant odds ratios may be poor classifiers. *PLoS. Genet.* 5:e1000337. doi: 10.1371/journal.pgen.1000337.

Janssens, A.C., Moonesinghe, R., Yang, Q., Steyerberg, E.W., van Duijn, C.M., and Khoury, M.J. 2007. The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. *Genet. Med.* 9:528-535. doi: 10.1097/GIM.0b013e31812eece0.

Johnson, T. 2016. Efficient Calculation for Multi-SNP Genetic Risk Scores. *https://cran.r-project.org/web/packages/gtx/vignettes/ashg2012.pdf*

Keaton, J.M., Cooke Bailey, J.N., Palmer, N.D., Freedman, B.I., Langefeld, C.D., Ng, M.C., and Bowden, D.W. 2014. A comparison of type 2 diabetes risk allele load between African Americans and European Americans. *Hum. Genet.* 133:1487-1495. doi: 10.1007/s00439-014-1486-5.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., latkin, M.S., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F., McCarroll, S.A., and Visscher, P.M. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747-753. doi: 10.1038/nature08494.

Menard, S. 2000. Coefficients of determination for multiple logistic regression analysis. *Amer. Stat.* 54:17-24. doi: 10.1080/00031305.2000.10474502

Pepe, M.S., Janes, H., Longton, G., Leisenring, W., and Newcomb, P. 2004. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am. J. Epidemiol.* 159:882-890. doi: 10.1093/aje/kwh101.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559-575. doi: 10.1086/519795.

Restrepo, N.A., Spencer, K.L., Goodloe, R., Garrett, T.A., Heiss, G., Buzkova, P., Jorgensen, N., Jensen, R.A., Matise, T.C., Hindorff, L.A., Klein, B.E., Klein, R., Wong, T.Y., Cheng, C.Y., Cornes, B.K., Tai, E.S., Ritchie, M.D., Haines, J.L., and Crawford, D.C. 2014. Genetic determinants of age-related macular degeneration in diverse populations from the PAGE study. *Invest. Ophthalmol. Vis. Sci.* 55:6839-6850. doi: 10.1167/iovs.14-14246.

Spencer, K.L., Glenn, K., Brown-Gentry, K., Haines, J.L., and Crawford, D.C. 2012. Population differences in genetic risk for age-related macular degeneration and implications for genetic testing. *Arch. Ophthalmol.* 130:116-117. doi: 10.1001/archopthalmol.2011.1370.

Turner, S., Armstrong, L. L., Bradford, Y., Carlson, C.S., Crawford, D.C., Crenshaw, A.T., de An-

drade, M., Doheny, K.F., Haines, J.L., Hayes, G., Jarvik, G., Jiang, L., Kullo, I.J., Li, R., Ling, H., Manolio, T.A., Matsumoto, M., McCarty, C.A., McDavid, A.N., Mirel, D.B., Paschall, J.E., Pugh, E.W., Rasmussen, L.V., Wilke, R.A., Zuvich, R.L., and Ritchie, M. D. 2011. Quality control procedures for genome-wide association studies. *Curr. Protoc. Hum. Genet.* 68:1.19.1-1.19.18. doi: 10.1002/0471142905.hg0119s68.

Weinreb, R.N., Aung, T., and Medeiros, F.A. 2014. The pathophysiology and treatment of glaucoma: A review. *JAMA* 311:1901-1911. doi: 10.1001/jama.2014.3192.

Wiggs, J.L., Yaspan, B.L., Hauser, M.A., Kang, J.H., Allingham, R.R., Olson, L.M., Abdrabou, W., Fan, B.J., Wang, D.Y., Brodeur, W., Budenz, D.L., Caprioli, J., Crenshaw, A., Crooks, K., Delbono, E., Doheny, K.F., Friedman, D.S., Gaasterland, D., Gaasterland, T., Laurie, C., Lee, R.K., Lichter, P.R., Loomis, S., Liu, Y., Medeiros, F.A., McCarty, C., Mirel, D., Moroi, S.E., Musch, D.C., Realini, A., Rozsa, F.W., Schuman, J.S., Scott, K., Singh, K., Stein, J.D., Trager, E.H., Vanveldhuisen, P., Vollrath, D., Wollstein, G., Yoneyama, S., Zhang, K., Weinreb, R.N., Ernst, J., Kellis, M., Masuda, T., Zack, D., Richards, J.E., Pericak-Vance, M., Pasquale, L.R., and Haines, J.L. 2012. Common variants at 9p21 and 8q22 are associated with increased susceptibility to optic nerve degeneration in glaucoma. *PLoS. Genet.* 8:e1002654. doi: 10.1371/journal.pgen.1002654.

Wiggs, J.L. 2015. Glaucoma Genes and Mechanisms. *Prog. Mol. Biol. Transl. Sci.* 134:315-342. doi: 10.1016/bs.pmbts.2015.04.008.

Witte, J.S., Visscher, P.M., and Wray, N.R. 2014. The contribution of genetic variants to disease depends on the ruler. *Nat. Rev. Genet.* 15:765-776. doi: 10.1038/nrg3786.

Wray, N.R. and Goddard, M.E. 2010. Multi-locus models of genetic risk of disease. *Genome Med*. 2:10. doi: 10.1186/gm131.

Wray, N.R., Yang, J., Goddard, M.E., and Visscher, P.M. 2010. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS. Genet.* 6:e1000864. doi: 10.1371/journal.pgen.1000864.

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. 2011. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88:76-82. doi: 10.1016/j.ajhg.2010.11.011.