



Polygenic Epidemiology

Frank Dudbridge*

Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London, United Kingdom

Received 19 November 2015; Revised 5 February 2016; accepted revised manuscript 5 February 2016.

Published online 7 April 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/gepi.21966

ABSTRACT: Much of the genetic basis of complex traits is present on current genotyping products, but the individual variants that affect the traits have largely not been identified. Several traditional problems in genetic epidemiology have recently been addressed by assuming a polygenic basis for disease and treating it as a single entity. Here I briefly review some of these applications, which collectively may be termed polygenic epidemiology. Methodologies in this area include polygenic scoring, linear mixed models, and linkage disequilibrium scoring. They have been used to establish a polygenic effect, estimate genetic correlation between traits, estimate how many variants affect a trait, stratify cases into subphenotypes, predict individual disease risks, and infer causal effects using Mendelian randomization. Polygenic epidemiology will continue to yield useful applications even while much of the specific variation underlying complex traits remains undiscovered.

Genet Epidemiol 40:268–272, 2016. Published 2016 Wiley Periodicals, Inc.*

KEY WORDS: missing heritability; genetic correlation; genetic risk prediction; Mendelian randomization

The completion of genome-wide association studies (GWAS) for hundreds of complex diseases and traits has created new challenges for genetic epidemiologists as we seek to understand the function of associated loci, using multiple emerging technologies [Ziegler and König, 2014]. At the same time, the realization that most if not all complex traits are polygenic—that is, they are influenced by thousands of genetic variants each having a small effect—has spurred the development of methods that address traditional problems in genetic epidemiology by treating the entire polygenic basis as a single entity. By dealing with the genetic basis *en masse*, these methods access more of the heritable component of complex traits than is possible by single-variant approaches, and thus alleviate much of the missing heritability problem that has recently so exercised genetic epidemiologists [Maher, 2008]. Here I briefly review several such applications, which together form an emerging field of *polygenic epidemiology*.

The concept of a polygenic risk is well established in classical genetics and in humans has been applied for some time in, for example, segregation analysis of complex diseases [Antonou et al., 2002; Risch, 1990]. But only with the development of genome-wide panels of SNP markers has it become possible to treat the polygenic basis explicitly. An early success was in the recalculation of twin-based heritabilities using observed, rather than expected, genetic similarity between dizygous pairs [Visscher et al., 2008]. Two influential papers then introduced complementary methods for demonstrating evidence of a polygenic effect, and have each led to further developments across a range of applications.

The polygenic scoring method was applied by Purcell et al. [2009] to argue that schizophrenia has a polygenic risk. Although their GWAS identified few individually significant SNPs, a small systematic increase in χ^2 statistics was observed across the genome, and could not be reduced by control for population structure or genotyping error. The polygenic scoring method uses a GWAS dataset, called the *training sample*, to estimate effect sizes for each SNP, and to select SNPs according to a *P*-value threshold. In a second dataset, called the *target sample*, a polygenic risk score is calculated for each subject as the weighted sum of risk alleles at the selected SNPs, with the weights being the effects estimated in the training data. A polygenic risk would cause this score to be associated for very liberal selection *P*-values. Purcell et al. observed significant association of the score with a selection threshold as high as $P < 0.5$, and used simulation to argue that this was consistent with a polygenic effect; their argument has since been strengthened by theory [Dudbridge, 2013; Yang et al., 2011b].

Linear mixed models were used by Yang et al. [2010] to show that much of the heritability of height can be explained by current GWAS chips, even though very little could be explained by known associated variants. The approach derives from methods of quantitative genetics used in crop and livestock breeding, in which the genetic value (or breeding value) of an individual can be derived as a random effect by relating the phenotypes of the study subjects to their known pedigree structure. The innovation in human studies is that nominally unrelated subjects can be used, by estimating their (distant) relatedness from genome-wide marker data. By relating genetic to phenotypic similarity across all pairs of subjects, the variance of the genetic values can be estimated and taken as an estimate of narrow-sense heritability. Because this depends

*Correspondence to: Frank Dudbridge, Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. E-mail: frank.dudbridge@lshtm.ac.uk

upon the markers used to estimate relatedness, especially the fact that they are usually selected to be common (e.g., >1% minor allele frequency), this estimate is known as *chip heritability*. A further advantage of using unrelated individuals is that shared environmental effects are minimal and therefore unlikely to bias the chip heritability.

Both polygenic scores and linear mixed models have been used to infer a polygenic basis for a wide range of traits [Bush et al., 2010; Lu et al., 2014; Speliotes et al., 2010; Visscher et al., 2012], to the degree that it is now generally accepted that all complex traits are polygenic. Much attention is now directed at demonstrating and estimating the genetic correlation between pairs of traits. This can be done with polygenic scores, by training the score on one trait and testing it against another: if there is no shared genetic basis, the score will not be associated. A bivariate linear mixed model may also be used, modeling the genetic variance for two traits simultaneously with their covariance. Both approaches have shown a shared basis for schizophrenia and bipolar disorder, as well as for other pairs of psychiatric disorders [Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Lee et al., 2013; Purcell et al., 2009]. This indicates a common molecular etiology for these conditions, which may prove useful for developing novel treatments or identifying individuals at risk, but may also suggest problems with current nosology and diagnosis. A shared basis has also been demonstrated for mammographic breast density and breast cancer, which may in part explain the former as a risk factor for the latter, by mediating some of the genetic risk of disease [Varghese et al., 2012]. On the other hand, genetic correlation has not been observed for some pairs of traits for which it might be expected, especially in neurology [Goris et al., 2014]. Autism appears to occupy a distinct position among psychiatric disorders in that it has no genetic correlation with the other major disorders studied, other than a low (16%) correlation with schizophrenia [Lee et al., 2013].

Linear mixed models give immediate estimates of chip heritability and genetic correlation: inference of the polygenic effect follows by testing these effects against their standard errors [Visscher et al., 2014]. By contrast, polygenic scoring gives a hypothesis test for the polygenic effect, without an immediate estimate of chip heritability. However, methods have been developed to infer chip heritability and genetic correlation from the result of a polygenic score test [Palla and Dudbridge, 2015; Stahl et al., 2012], by estimating under what degree of chip heritability would the observed result be expected. These models also allow for a proportion of variants with no effect on the trait; linear mixed models have also been implemented with this feature [Meuwissen et al., 2001; Moser et al., 2015; Zhou and Stephens, 2012]. A notable finding has been that the proportion of variants affecting a complex trait rarely exceeds 5%, even in heterogeneous samples [Palla and Dudbridge, 2015]. Thus the classical infinitesimal model may not hold in truth, although the traits remain highly polygenic. This finding affects the interpretation of genetic correlation, because the overall genetic correlation is determined by the proportion of variants with effects on both traits and the

correlation of effects among those variants. A given genetic correlation may be concentrated among a few variants with highly concordant effects on both traits, or dispersed among more variants with only weakly concordant effects. Although the genetic correlation is informative at the whole subject level, it is not in itself very informative at the single-variant level. Chip heritability itself is only interpretable in the context of an assumed model, usually (for disease traits) a liability threshold model, and the population-specific environment [Hopper and Mack, 2015]; however it is generally useful to estimate the broad degree of genetic variance and correlation, even if it is challenging to provide or interpret a precise estimate.

A third approach to assessing chip heritability is linkage disequilibrium (LD) scoring [Bulik-Sullivan et al., 2015a]. Its rationale is that the more variants a given marker is in LD with, the higher is its (marginal) association statistic likely to be. Regression of χ^2 statistics for genome-wide markers on their LD “scores” gives an estimate of chip heritability (the slope of the regression) as well as of systematic bias due to population structure (the intercept). An extension allows the estimation of genetic covariance between traits [Bulik-Sullivan et al., 2015b].

A further application is the estimation of chip heritability within sub-groups of markers. Such “genome partitioning” analyses have demonstrated that polygenic effects are spread uniformly across the chromosomes, but variants with known functional effects tend to explain more variation than others [Visscher et al., 2007; Yang et al., 2011a].

Polygenic scores show promise for patient stratification and subphenotyping. Hamshere et al. [2011] showed that, among bipolar disorder cases, polygenic scores for schizophrenia risk could distinguish schizo-affective cases from others, while not distinguishing psychotic cases from nonpsychotic. In inflammatory bowel disease, polygenic scores can distinguish cases with colonic from ileal Crohn’s disease, and from those with ulcerative colitis [Cleynen et al., 2016]. Prostate cancer screening targeted to men with high polygenic risk could reduce the rate of overdiagnosis [Pashayan et al., 2015].

Ultimately we might hope for individual risk prediction from polygenic analysis, which indeed first motivated the polygenic scoring method [Wray et al., 2007]. The performance of genetic prediction is bounded by the heritability [Clayton, 2009], but early attempts fell well short of that limit [Evans et al., 2009]. Furthermore, family history is an informative marker for genetic risk, and any predictor based on measured genotypes should exceed that benchmark to be deemed worthwhile [Aulchenko et al., 2009]. It is now clear that much larger samples are required to attain accurate prediction from polygenic models, of the order of 10^5 subjects [Chatterjee et al., 2013; Daetwyler et al., 2008; Dudbridge, 2013]. The reason is simply that as more variants enter the prediction model (as is necessary to explain the genetic risk), the greater is the sampling error in the total score, so the latter must be kept extremely small at the single-variant level. National biobanks and disease consortia are now beginning to

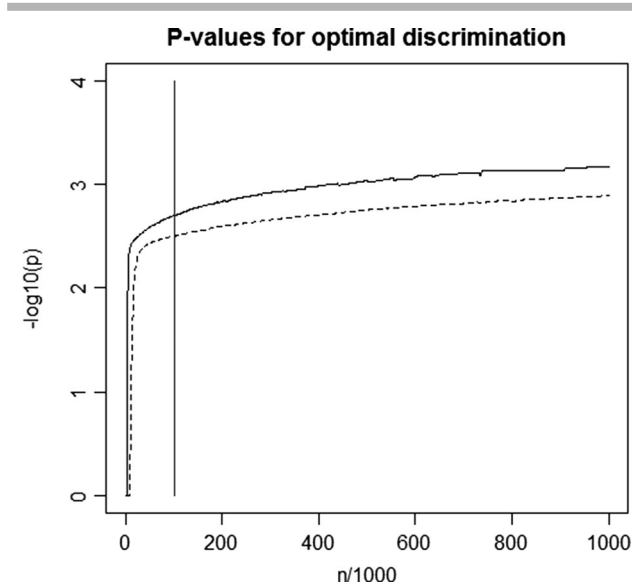


Figure 1. P -values ($-\log_{10}$ scale) for selecting variants into a polygenic score such that the area under the receiver operator characteristic curve (AUC) is maximized. A binary trait with prevalence 10% is assumed, with variants selected from a case/control study with equal number of cases and controls. Chip heritability of 40% (solid line) and 20% (dashed line) is distributed among 100,000 independent variants, of which 5% have normally distributed effects and the rest have no effect. The vertical line is at 50,000 cases and 50,000 controls, at which point over 95% of the maximum AUC is achieved.

approach this scale [CardiogramPlusC4D Consortium, 2015; Locke et al., 2015; Michailidou et al., 2013; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014], and the outlook for genetic prediction is becoming more promising than has recently appeared.

A common practice is to perform genetic prediction using only variants that are robustly associated at genome-wide significance [Mavaddat et al., 2015; Szulkin et al., 2015; Talmud et al., 2015]. This is based on intuition that the risk score contains genuine predictors and no “noise,” and could therefore be easily conveyed to clinicians and policy makers. However, precisely in line with the missing heritability problem, such predictors explain very little variation in disease risk, and therefore have little predictive accuracy. Analytic results [Dudbridge, 2013] suggest that, even at very large sample sizes, prediction is optimized by selecting variants with P -values as high as 0.001 into a risk score (Fig. 1). Perhaps counterintuitively, it is possible to predict with accuracy close to the theoretical maximum even while including many neutral variants in the model. By the same token, as sample sizes continue to grow it should be possible to achieve useful levels of prediction even while many individually associated variants remain to be discovered. In translating such predictors into practice, it will be important to present the polygenic risk as a single entity in order to allay concerns that it contains many irrelevant variants. This is naturally achieved by linear mixed models, which estimate a single genetic value for each individual as a random effect, and these models are now

being extended to allow for subgroups of variants with no or little effect [Moser et al., 2015; Speed and Balding, 2014].

As a final aspect of polygenic epidemiology discussed here, Mendelian randomization studies are increasingly using composite genetic scores to draw causal inferences from observational data [Nuesch et al., 2015; Thrift et al., 2015; Voight et al., 2012; Zhang et al., 2015]. The advantages are similar to those of individual prediction, in that a composite score can predict the intermediate trait to a greater accuracy than can single variants. A particular difficulty however is that Mendelian randomization assumes that the genetic instrument only affects the outcome through the exposure of interest. This assumption is increasingly likely to be violated as more variants enter the model [Burgess and Thompson, 2013; Palmer et al., 2012]. There is a stronger argument for restricting gene scores to robustly associated loci, but pleiotropy can still violate the assumption [Holmes et al., 2015]. In a highly polygenic score, variants with effects on the outcome but not on the exposure can create substantial bias [Evans et al., 2013]. Methods are now available to adjust Mendelian randomization analyses for known [Burgess and Thompson, 2015; Burgess et al., 2015] and unknown [Bowden et al., 2015] pleiotropic effects. Concerns about pleiotropy have so far discouraged the application of linear mixed models to Mendelian randomization, although in principle they could improve precision over polygenic scores.

The small effects of individual variants have caused the value of GWAS to be questioned [Manolio, 2013]. But bearing in mind that most complex traits are strongly heritable, the polygenic risk is highly informative taken *en masse* as a single risk factor. Confirmation is appearing through the applications of polygenic epidemiology discussed here, and further applications—such as identifying interactions, or mediation analysis—are sure to be developed in the near future. So far, the field has mainly developed from psychiatric genetics, in which progress in identifying and following up GWAS associations has been slower than in other areas. However, as seen here, applications are becoming common in cardiovascular, cancer, and immunological genetics. Other areas will add further insights: for example, population geneticists have used GWAS signals to demonstrate polygenic adaptation [Berg and Coop, 2014]. Linear mixed models are in principle the most powerful and accurate class of methods, with the principal challenges being the correct specification of the random effects distribution, and computational issues in large samples [Ge et al., 2015]. Polygenic risk scores, and LD scoring, offer simpler and faster approaches, requiring only summary statistics from completed GWAS, and generally incurring only a moderate loss of precision compared to linear mixed models. The variety of methods and applications now emerging in polygenic epidemiology will make this field a fertile ground for development in the coming years.

Acknowledgments

This work was funded by the MRC (MR/K006215/1). The author declares no conflict of interest.

References

- Antoniou AC, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, Ponder BJ, Easton DF. 2002. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer* 86(1):76–83. Epub 2002/02/22.
- Aulchenko YS, Struchalin MV, Belongova NM, Axenovich TI, Weedon MN, Hofman A, Uitterlinden AG, Kayser M, Oostra BA, van Duijn CM, and others. 2009. Predicting human height by Victorian and genomic methods. *Eur J Hum Genet* 17(8):1070–1075. Epub 2009/02/19.
- Berg JJ, Coop G. 2014. A population genetic signal of polygenic adaptation. *PLoS Genet* 10(8):e1004412. Epub 2014/08/08.
- Bowden J, Davey Smith G, Burgess S. 2015. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 44(2):512–525. Epub 2015/06/08.
- Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM. 2015a. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 47(3):291–295. Epub 2015/02/03.
- Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, Duncan L, Perry JR, Patterson N, Robinson EB, and others. 2015b. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 47(11):1236–1241. Epub 2015/09/29.
- Burgess S, Thompson SG. 2013. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol* 42(4):1134–1144. Epub 2013/09/26.
- Burgess S, Thompson SG. 2015. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 181(4):251–260. Epub 2015/01/30.
- Burgess S, Dudbridge F, Thompson SG. 2015. Re: “Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects.” *Am J Epidemiol* 181(4):290–291. Epub 2015/02/11.
- Bush WS, Sawcer SJ, de Jager PL, Oksenberg JR, McCauley JL, Pericak-Vance MA, Haines JL. 2010. Evidence for polygenic susceptibility to multiple sclerosis—the shape of things to come. *Am J Hum Genet* 86(4):621–625. Epub 2010/04/07.
- CardiogramPlusC4D Consortium. 2015. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 47(10):1121–1130. Epub 2015/09/08.
- Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park JH. 2013. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat Genet* 45(4):400–405, 5e1–5e3. Epub 2013/03/05.
- Clayton DG. 2009. Prediction and interaction in complex disease genetics: experience in type 1 diabetes. *PLoS Genet* 5(7):e1000540. Epub 2009/07/09.
- Cleynen I, Boucher G, Jostins L, Schumm LP, Zeissig S, Ahmad T, Andersen V, Andrews JM, Anness V, Brand S, and others. 2016. Inherited determinants of Crohn’s disease and ulcerative colitis phenotypes: a genetic association study. *Lancet* 387(10014):156–167. Epub 2015/10/23.
- Cross-Disorder Group of the Psychiatric Genomics Consortium. 2013. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 381(9875):1371–1379. Epub 2013/03/05.
- Daetwyler HD, Villanueva B, Woolliams JA. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3(10):e3395. Epub 2008/10/15.
- Dudbridge F. 2013. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 9(3):e1003348. Epub 2013/04/05.
- Evans DM, Visscher PM, Wray NR. 2009. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet* 18(18):3525–3531. Epub 2009/06/26.
- Evans DM, Brion MJ, Paternoster L, Kemp JP, McMahon G, Munafò M, Whitfield JB, Medland SE, Montgomery GW, Timpson NJ, and others. 2013. Mining the human genome using allelic scores that index biological intermediates. *PLoS Genet* 9(10):e1003919. Epub 2013/11/10.
- Ge T, Nichols TE, Lee PH, Holmes AJ, Roffman JL, Buckner RL, Sabuncu MR, Smoller JW. 2015. Massively expedited genome-wide heritability analysis (MEGHA). *Proc Natl Acad Sci USA* 112(8):2479–2484. Epub 2015/02/13.
- Goris A, van Setten J, Diekstra F, Ripke S, Patsopoulos NA, Sawcer SJ, van Es M, Andersen PM, Melki J, Meininger V, and others. 2014. No evidence for shared genetic basis of common variants in multiple sclerosis and amyotrophic lateral sclerosis. *Hum Mol Genet* 23(7):1916–1922. Epub 2013/11/16.
- Hamshere ML, O’Donovan MC, Jones IR, Jones L, Kirov G, Green EK, Moskvina V, Grozeva D, Bass N, McQuillin A, and others. 2011. Polygenic dissection of the bipolar phenotype. *Br J Psychiatry* 198(4):284–288. Epub 2011/10/06.
- Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, Dale CE, Padmanabhan S, Finan C, Swerdlow DI, and others. 2015. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* 36(9):539–550. Epub 2014/01/30.
- Hopper JL, Mack TM. 2015. The heritability of prostate cancer—letter. *Cancer Epidemiol Biomarkers Prev* 24(5):878. Epub 2015/05/03.
- Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, Mowry BJ, Thapar A, Goddard ME, Witte JS, and others. 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 45(9):984–994. Epub 2013/08/13.
- Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, and others. 2015. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518(7538):197–206. Epub 2015/02/13.
- Lu Y, Ek WE, Whiteman D, Vaughan TL, Spurdle AB, Easton DF, Pharoah PD, Thompson DJ, Dunning AM, Hayward NK, and others. 2014. Most common “sporadic” cancers have a significant germline genetic component. *Hum Mol Genet* 23(22):6112–6118. Epub 2014/06/20.
- Maher B. 2008. Personal genomes: the case of the missing heritability. *Nature* 456(7218):18–21. Epub 2008/11/07.
- Manolio TA. 2013. Bringing genome-wide association findings into clinical use. *Nat Rev Genet* 14(8):549–558. Epub 2013/07/10.
- Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M, and others. 2015. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst* 107(5). Epub 2015/04/10.
- Meuwissen TH, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–1829. Epub 2001/04/06.
- Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, and others. 2013. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 45(4):353–361, 61e1–61e2. Epub 2013/03/29.
- Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM. 2015. Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. *PLoS Genet* 11(4):e1004969. Epub 2015/04/08.
- Nuesch E, Dale C, Palmer TM, White J, Keating BJ, van Iperen EP, Goel A, Padmanabhan S, Asselbergs FW, Verschuren WM, and others. 2015. Adult height, coronary heart disease and stroke: a multi-locus Mendelian randomization meta-analysis. *Int J Epidemiol*. Epub 2015/05/17.
- Palla L, Dudbridge F. 2015. A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *Am J Hum Genet* 97(2):250–259. Epub 2015/07/21.
- Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, Davey Smith G, Sterne JA. 2012. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res* 21(3):223–242. Epub 2011/01/11.
- Pashayan N, Duffy SW, Neal DE, Hamdy FC, Donovan JL, Martin RM, and others. 2015. Implications of polygenic risk-stratified screening for prostate cancer on overdiagnosis. *Genet Med* 17(10):789–795. Epub 2015/01/09.
- Purcell SM, Wray NR, Stone JL, Visscher PM, O’Donovan MC, Sullivan PF, Sklar P, International Schizophrenia Consortium. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460(7256):748–752. Epub 2009/07/03.
- Risch N. 1990. Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 46(2):222–228. Epub 1990/02/01.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511(7510):421–427. Epub 2014/07/25.
- Speed D, Balding DJ. 2014. MultiBLUP: improved SNP-based prediction for complex traits. *Genome Res* 24(9):1550–1557. Epub 2014/06/26.
- Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Allen HL, Lindgren CM, Luan J, Magi R, and others. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42(11):937–948. Epub 2010/10/12.
- Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, Voight BF, Kraft P, Chen R, Kallberg HJ, Kurzeeman FA, and others. 2012. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* 44(5):483–489. Epub 2012/03/27.
- Szulkin R, Whittington T, Eklund M, Aly M, Eeles RA, Easton D, Kote-Jarai ZS, Amin Al Olama A, Benlloch S, Muir K, and others. 2015. Prediction of individual genetic risk to prostate cancer using a polygenic score. *Prostate* 75(13):1467–1474. Epub 2015/07/17.
- Talmud PJ, Cooper JA, Morris RW, Dudbridge F, Shah T, Engmann J, Dale C, White J, McLachlan S, Zabaneh D, and others. 2015. Sixty-five common genetic variants and prediction of type 2 diabetes. *Diabetes* 64(5):1830–1840. Epub 2014/12/06.
- Thrift AP, Gong J, Peters U, Chang-Claude J, Rudolph A, Slattery ML, Chan AT, Locke AE, Kahali B, Justice AE, and others. 2015. Mendelian randomization study of body mass index and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* 24(7):1024–1031. Epub 2015/05/16.
- Varghese JS, Thompson DJ, Michailidou K, Lindstrom S, Turnbull C, Brown J, Leyland J, Warren RM, Luben RN, Loos RJ, and others. 2012. Mammographic breast density and breast cancer: evidence of a shared genetic basis. *Cancer Res* 72(6):1478–1484. Epub 2012/01/24.

- Visscher PM, Macgregor S, Benyamin B, Zhu G, Gordon S, Medland S, Hill WG, Hottenga JJ, Willemsen G, Boomsma DI, and others. 2007. Genome partitioning of genetic variation for height from 11,214 sibling pairs. *Am J Hum Genet* 81(5):1104–1110. Epub 2007/10/10.
- Visscher PM, Hill WG, Wray NR. 2008. Heritability in the genomics era—concepts and misconceptions. *Nat Rev Genet* 9(4):255–266. Epub 2008/03/06.
- Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of GWAS discovery. *Am J Hum Genet* 90(1):7–24. Epub 2012/01/17.
- Visscher PM, Hemani G, Vinkhuyzen AA, Chen GB, Lee SH, Wray NR, Goddard ME, Yang J. 2014. Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet* 10(4):e1004269. Epub 2014/04/12.
- Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, Hindy G, Holm H, Ding EL, Johnson T, and others. 2012. Plasma HDL cholesterol and risk of myocardial infarction: a Mendelian randomisation study. *Lancet* 380(9841):572–580. Epub 2012/05/23.
- Wray NR, Goddard ME, Visscher PM. 2007. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 17(10):1520–1528. Epub 2007/09/06.
- Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, and others. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42(7):565–569. Epub 2010/06/22.
- Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, de Andrade M, Feenstra B, Feingold E, Hayes MG, and others. 2011a. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 43(6):519–525. Epub 2011/05/10.
- Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, Smith AV, Ingelsson E, O'Connell JR, Mangino M, and others. 2011b. Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* 19(7):807–812. Epub 2011/03/17.
- Zhang B, Shu XO, Delahanty RJ, Zeng C, Michailidou K, Bolla MK, Wang Q, Dennis J, Wen W, Long J, and others. 2015. Height and breast cancer risk: evidence from prospective studies and Mendelian randomization. *J Natl Cancer Inst* 107(11). Epub 2015/08/25.
- Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44(7):821–824. Epub 2012/06/19.
- Ziegler A, König IR. 2014. Celebrating the 30th anniversary of genetic epidemiology: how to define our scope? *Genet Epidemiol* 38(5):379–380. Epub 2014/06/26.