Editorial

Translating Polygenic Analysis for Prevention
From Who to How

Daniel W. Belsky, PhD

There are no easy answers in complex disease genetics. Common, chronic health conditions such as obesity and heart disease are influenced by many genetic variants scattered across the genome, with each variant making small contributions to risk. Ever-larger genome-wide association studies (GWAS) are uncovering more and more of these variants. For many diseases and disease-related traits, discovered loci now number in triple digits. As GWAS sample sizes grow, the volume of discovered variants is expected to increase further.[1,2] Translation of this growing volume of genetic discoveries to improve population health is needed.

See Articles by Seyednasrollah et al and Nuotio et al

The classic model of translation is to work from the bottom up, from a discovered DNA sequence variant through RNA transcription, protein production, and so on to disease pathogenesis. Such bottom-up translation is critical but can be time and resource intensive. For example, the FTO locus associated with obesity and the 9p21 locus associated with coronary artery disease were discovered a decade ago; understanding their molecular mechanisms remains a work in progress.[3–6] With genetic discoveries already numbering in the hundreds and new findings expected to have ever-smaller individual effects on risk, complementary translational approaches are needed to generate return on investments in GWAS discovery.

An alternative to bottom-up translation of discovered variants one at a time is polygenic score analysis. Polygenic score analysis models genetic influence quantitatively, as the combined effects of many independent variants.[7] Typically, each locus is assigned a weight based on the effect size estimated in GWAS. The count of risk alleles at that locus is then multiplied by the weight, and weighted counts are summed across an individual's genome to compute their polygenic score. That polygenic score provides an overall summary of an individual's genetic liability to manifest the phenotype. In genetic epidemiology, polygenic score analysis is often used to test of how much phenotypic variance can be explained by GWAS results.[8] Translation of polygenic score analysis to the clinic has proved more challenging.

Most efforts toward translation of polygenic score analysis have focused on risk assessment. The simple idea is that a polygenic score can be used as a genetic test that, alongside other risk indicators, is used to stratify patients for prevention and treatment. Because DNA is fixed at conception, readily available from every tissue in the body, and increasingly inexpensive to assay, polygenic risk assessment has conceptual appeal. But for common diseases, the value added from genetic information is limited when considered in the clinical context, where a wealth of other health information is often already available.[9,10] In this issue of the journal, articles by Nuotio et al[11] and Seyednasrollah et al[12] investigate the possibility that polygenic risk assessment for adult dyslipidemia and obesity may provide more value when applied in children.

Applying polygenic risk assessment for adult disease in children is of interest for 2 reasons. First, phenotypic indications of risk may not yet be observable, increasing novelty of risk information from genetic analysis. Second, disease processes may not yet be established, increasing potential for prevention. Testing effectiveness of polygenic risk assessment in children requires data that observe individuals as children, before disease processes are established, and again at an appropriate stage in adulthood, when disease is prevalent in the population. Population-based longitudinal studies with childhood clinical information, adult follow-up, and genome-wide genetic data are still rare. The analyses by Nuotio et al and Seyednasrollah et al of the Cardiovascular Risk in Young Finns study,[13] which enrolled children aged 3 to 18 in 1980 and has to date conducted follow-ups through 2012, are thus of substantial interest.

The study by Nuotio et al[11] of genetic risk assessment for dyslipidemia analyzed data from 2422 Young Finns participants with genetic data and lipid measurements taken during childhood and again in their 30s and 40s. The authors composed lipid-specific genetic risk scores from single-nucleotide polymorphisms (SNPs) identified in the recent Global Lipids Consortium GWAS.[14] Scores were defined separately for low-density lipoprotein cholesterol (58 SNPs), high-density lipoprotein cholesterol (71 SNPs), and triglycerides (40 SNPs). Adult dyslipidemia was determined based on European cut points (low-density lipoprotein cholesterol >3.0 mmol/L or taking lipid-lowering medication, 58% of the cohort; high-density lipoprotein cholesterol <1.0 mmol/L in men and <1.2 mmol/L in women, 24% of the cohort; triglycerides <1.7 mmol/L, 21% of the cohort).[15] Analysis tested whether genetic risk scores improved prediction of adult dyslipidemia over and above information about childhood lipid levels, body mass index (BMI), physical activity, and, for those >12 years of age, smoking.

From the Department of Medicine, Duke University School of Medicine and Center for Population Health Science and Population Research Institute, Duke University.

Correspondence to Daniel W. Belsky PhD, Department of Medicine, Duke University School of Medicine, 2020 W. Main St. Suite 201, Durham NC 27705. E-mail dbelsky@duke.edu

(*Circ Cardiovasc Genet*. 2017;10:e001798. DOI: 10.1161/CIRCGENETICS.117.001798.)

***Circ Cardiovasc Genet* is available at http://circgenetics.ahajournals.org**

DOI: 10.1161/CIRCGENETICS.117.001798

Results were that the addition of genetic information modestly improved risk prediction. Genetic information contributed a ≈0.01 increase in area under the receiver-operating characteristic curve for the model. Put another way, adding genetic information to the prediction model increased by one percentage point the chance that a randomly selected participant with dyslipidemia would have a higher prediction score as compared with a randomly selected participant without dyslipidemia. The authors also computed net reclassification improvement and integrated discrimination index measures.[16] As with the change in area under the receiver-operating characteristic curve, net reclassification improvement and integrated discrimination index metrics indicated modest improvement in predictive accuracy with the addition of genetic information to the models. Results were similar when a machine-learning technique was used to build the prediction models in one half of the data set, and the resulting model was then tested in the other half.

The study by Seyednasrollah et al[12] of genetic risk assessment for obesity analyzed data from 2262 young Finns participants with genetic data and BMI measurements taken during childhood and again in their 30s and 40s, as well as information about maternal BMI and family income from the childhood baseline assessment. The authors composed their genetic risk scores based on the 97 loci identified in The Genetic Investigation of Anthropometric Traits (GIANT) consortium's most recent BMI GWAS.[17] They also composed a second genetic risk score based on the subset of 19 of the GIANT loci for which tag SNPs were associated with adult BMI in the Young Finns Cohort at $P<0.10$. Adult obesity was defined as BMI ≥30 (20% of the cohort). The authors applied a machine-learning approach to build a prediction model in one half of their data and then tested their model in the other half. Analysis tested whether including a genetic risk score in the prediction model improved accuracy over and above information about childhood BMI, maternal BMI, and family income. Results were that addition of genetic information improved risk prediction, but only for those children's whose baseline assessment occurred by the age of 6 years (change in area under the receiver-operating characteristic curve=0.07). For children with baseline BMI measurements taken at later ages, no statistically significant improvement in prediction was observed for either genetic risk score. This finding is consistent with a previous observation that GWAS-discovered genetic risk for adult obesity is mediated by middle-childhood weight gain.[18] Childhood BMI assessments made after middle childhood may, thus, already capture most information contained in the genetic risk score. Also consistent with previous studies,[18–20] the authors observed diverging BMI trajectories in high and low genetic risk children extending from middle childhood into adulthood in both their Young Finns data and a replication cohort drawn from the Bogalusa Heart Study.[21]

These studies, which used unique data following up >2000 Finnish children prospectively into adulthood, identified modest but statistically significant improvements in the predictive accuracy of clinical risk assessments for dyslipidemia and obesity with addition of genetic information. Genetic information on its own is unlikely to provide clinically meaningful risk information for common health conditions like dyslipidemia and obesity. But the Young Finns analyses suggest that integrating genetics with other information can modestly improve prediction models. In future, as ever-larger GWAS fueled by data from national biobanks and health systems uncover new risk loci, value added from genetics could grow.

Optimism about translation of polygenic models for clinical risk assessment should be considered in the context of 2 challenges not easily overcome by ever-larger GWAS samples. One challenge is the population specificity of genetic risk assessments derived from GWAS. Most large-scale GWAS have been conducted in European-descent samples. Polygenic scores based on results from these GWAS are likely to offer much less information when applied in non-European populations.[22,23] Translating polygenic score analysis for risk assessment, thus, raises concerns about equity. As large-scale GWAS of non-European samples become possible,[24] development of parallel polygenic models in non-European populations must be a priority.

A second challenge is potential environmental specificity of genetically informed risk prediction models. Relationships between genotypes and phenotypes can vary across geographic space and historical time, presumably reflecting heterogeneity in nongenetic influences.[25–28] An advantage of the machine-learning approaches to predictive model building applied in the Young Finns analyses is that they incorporate nonlinear and interactive combinations of different risk factors. This approach leverages gene-by-environment and gene-by-phenotype interactions detected in the training data, maximizing potential value of genetic information. A limitation is that models trained in one population may not generalize well to others. Young Finns analyses used split-halves approach to formally separate training and test samples. But both derived from the same cohort—all sampled from the same places at the same time. Building predictive models that can be applied in more diverse clinical settings will require primary research to identify replicable gene–environment and gene–phenotype interactions.

In sum, although the Young Finns data illustrate some potential for polygenic score analysis to improve clinical risk assessment, there are both ethical and technical challenges to implementation. Complementary translational approaches are still needed. Genetically informed prospective longitudinal studies like Young Finns make possible studies of how polygenic risk relates to the natural history of chronic disease. Such developmental polygenic score analysis shifts focus from risk-assessment questions about who we should intervene with to etiologic questions about how we should intervene.[29] The goal is to identify intermediate phenotypes that link DNA differences established at conception with divergent health outcomes across the life span. Such intermediate phenotypes, like rapid early weight gain in the case of obesity,[18] in theory provide targets for prevention and inform models of risk assessment. As complementary approaches, polygenic risk assessment and developmental polygenic score analysis can help advance translation of GWAS discoveries.

Sources of Funding

Dr Belsky is supported by an early career fellowship from the Jacobs Foundation and by US National Institute on Aging grants P30AG028716, R01AG032282, and R21AG054846.

Disclosures

None.

References

1. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012;90:7–24. doi: 10.1016/j.ajhg.2011.11.029.
2. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al; LifeLines Cohort Study. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*. 2016;533:539–542. doi: 10.1038/nature17671.
3. Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. 2014;507:371–375. doi: 10.1038/nature13138.
4. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, et al. FTO obesity variant circuitry and adipocyte browning in humans. *N Engl J Med*. 2015;373:895–907. doi: 10.1056/NEJMoa1502214.
5. Herman MA, Rosen ED. Making biological sense of GWAS data: lessons from the FTO locus. *Cell Metab*. 2015;22:538–539. doi: 10.1016/j.cmet.2015.09.018.
6. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet*. 2017;18:331–344. doi: 10.1038/nrg.2016.160.
7. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res*. 2007;17:1520–1528. doi: 10.1101/gr.6665407.
8. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–52.
9. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet*. 2013;14:507–515. doi: 10.1038/nrg3457.
10. Loos RJ, Janssens AC. Predicting polygenic obesity using genetic information. *Cell Metab*. 2017;25:535–543. doi: 10.1016/j.cmet.2017.02.013.
11. Nuotio J, Pitkänen N, Magnussen CG, Buscot M-J, Venäläinen MS, Elo LL, et al. Prediction of adult dyslipidemia using genetic and childhood clinical risk factors: the Cardiovascular Risk in Young Finns study. *Circ Cardiovasc Genet*. 2017;10:e001604. doi: 10.1161/CIRCGENETICS.116.001604.
12. Seyednasrollah F, Mäkelä J, Pitkänen N, Juonala M, Hutri-Kähönen N, Lehtimäki T, et al. Prediction of adulthood obesity using genetic and childhood clinical risk factors in the Cardiovascular Risk in Young Finns study. *Circ Cardiovasc Genet*. 2017;10:e001554. doi: 10.1161/CIRCGENETICS.116.001554.
13. Raitakari OT, Juonala M, Rönnemaa T, Keltikangas-Järvinen L, Räsänen L, Pietikäinen M, et al. Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol*. 2008;37:1220–1226. doi: 10.1093/ije/dym225.
14. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274–1283.
15. Perk J, De Backer G, Gohlke H, Graham I, Reiner Z, Verschuren M, et al; European Association for Cardiovascular Prevention & Rehabilitation (EACPR); ESC Committee for Practice Guidelines (CPG). European guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). *Eur Heart J*. 2012;33:1635–1701. doi: 10.1093/eurheartj/ehs092.
16. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27:157–72; discussion 207. doi: 10.1002/sim.2929.
17. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al; LifeLines Cohort Study; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; International Endogene Consortium. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518:197–206. doi: 10.1038/nature14177.
18. Belsky DW, Moffitt TE, Houts R, Bennett GG, Biddle AK, Blumenthal JA, et al. Polygenic risk, rapid childhood growth, and the development of obesity: evidence from a 4-decade longitudinal study. *Arch Pediatr Adolesc Med*. 2012;166:515–521. doi: 10.1001/archpediatrics.2012.131.
19. Elks CE, Loos RJ, Hardy R, Wills AK, Wong A, Wareham NJ, et al. Adult obesity susceptibility variants are associated with greater childhood weight gain and a faster tempo of growth: the 1946 British Birth Cohort Study. *Am J Clin Nutr*. 2012;95:1150–1156. doi: 10.3945/ajcn.111.027870.
20. Warrington NM, Howe LD, Wu YY, Timpson NJ, Tilling K, Pennell CE, et al. Association of a body mass index genetic risk score with growth throughout childhood and adolescence. *PLoS One*. 2013;8:e79547. doi: 10.1371/journal.pone.0079547.
21. Smith EN, Chen W, Kähönen M, Kettunen J, Lehtimäki T, Peltonen L, et al. Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. *PLoS Genet*. 2010;6:e1001094. doi: 10.1371/journal.pgen.1001094.
22. Belsky DW, Israel S. Integrating genetics and social science: genetic risk scores. *Biodemography Soc Biol*. 2014;60:137–155. doi: 10.1080/19485565.2014.946591.
23. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet*. 2017;100:635–649. doi: 10.1016/j.ajhg.2017.03.004.
24. The H3Africa Consortium. Enabling the genomic revolution in Africa. *Science*. 2014;344:1346–1348.
25. Boardman JD. State-level moderation of genetic tendencies to smoke. *Am J Public Health*. 2009;99:480–486. doi: 10.2105/AJPH.2008.134932.
26. Demerath EW, Choh AC, Johnson W, Curran JE, Lee M, Bellis C, et al. The positive association of obesity variants with adulthood adiposity strengthens over an 80-year period: a gene-by-birth year interaction. *Hum Hered*. 2013;75:175–185. doi: 10.1159/000351742.
27. Walter S, Mejía-Guevara I, Estrada K, Liu SY, Glymour MM. Association of a genetic risk score with body mass index across different birth cohorts. *JAMA*. 2016;316:63–69. doi: 10.1001/jama.2016.8729.
28. Conley D, Laidley TM, Boardman JD, Domingue BW. Changing polygenic penetrance on phenotypes in the 20(th) century among adults in the US population. *Sci Rep*. 2016;6:30348. doi: 10.1038/srep30348.
29. Belsky DW, Moffitt TE, Caspi A. Genetics in population health science: strategies and opportunities. *Am J Public Health*. 2013;103(suppl 1):S73–S83. doi: 10.2105/AJPH.2012.301139.

KEY WORDS: Editorials ■ alleles ■ epidemiology ■ genetics ■ genomics ■ obesity

Translating Polygenic Analysis for Prevention: From Who to How

Daniel W. Belsky

Circ Cardiovasc Genet. 2017;10:
doi: 10.1161/CIRCGENETICS.117.001798
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1942-325X. Online ISSN: 1942-3268