

Editorial

From Prehistory to Precision Medicine: Are Genetic Risk Scores Our Next Step?

Sarah Reid¹ 



Since the dawn of civilization, man has demonstrated understanding of, and the ability to exploit, the concept of heredity. The earliest known evidence of selective breeding dates to the Mesolithic era, and a 6000-year-old tablet from Babylon shows a horse pedigree containing suggestions of the species' hereditary traits¹. Despite lacking a physical or chemical explanation of the substance transferred from parent to offspring, Gregor Mendel's 1865 paper², which defined the laws of monogenetic heredity, gave an indication of its existence. Around the same time, Francis Galton, a cousin of Charles Darwin, evaluated the effect of nature versus nurture through a considerable number of family and twin studies³. The work of the 2 scientists eventually gave rise to 2 fundamentally different views on heredity, with the critics of Mendelian inheritance arguing that it could not possibly explain traits displaying extreme individual variation in a population, such as stature. The dispute was finally resolved in 1918, when R.A. Fisher⁴ elegantly demonstrated that when a large number of genes with a Mendelian inheritance pattern act on top of each other, the result is a normally distributed variation in the characteristic, such as that seen for stature.

Today, a century after Fisher's publication, we have only just begun to comprehend the magnitude of genetic factors influencing a particular phenotype. For many complex diseases, hundreds or thousands of single-nucleotide polymorphisms (SNP) appear to be involved in the etiology⁵, each typically with an almost negligible effect size. With this realization, the early hopes that genome-wide association studies (GWAS) would lead to the identification of persons at risk of a particular disease, ultimately allowing new preventive and therapeutic strategies, were frustrated⁶. Thus, there remains a critical need for the development of new methods that integrate the results of genetic studies

to predict the individual's genetic predisposition for disease or for specific manifestations of a disease.

Could genetic risk scores (GRS) confer such a method? Recently, 2 extensive papers exploring GRS in systemic lupus erythematosus (SLE) have been published, demonstrating associations between the GRS and more severe manifestations of the disease, including nephritis, accrual of organ damage, and an earlier disease onset. These studies indicate a potential future role for the method in aiding the clinical decision making^{7,8}. To enable its clinical implementation, however, studies refining the method in order to improve their diagnostic accuracy are required.

In their paper entitled, "Relationship between Genetic Risk and Age of Disease Diagnosis in Systemic Lupus Erythematosus," published in this issue of *The Journal*, Dominguez and colleagues⁹ assess whether the cumulative genetic risk for SLE—as defined by several GRS constructed using different approaches—is associated with the age at SLE diagnosis. The study was performed on a large, multiethnic cohort, including both patients with adult- and childhood-onset SLE (cSLE; $n = 865$ and $n = 675$, respectively), and although the authors are not the first to explore the matter, the question is important. Patients with an early onset of SLE typically display a more aggressive disease phenotype with increased disease activity, a higher prevalence of nephritis, and risk of prolonged corticosteroid treatment^{10,11}. To a certain extent, the age at disease onset can thus serve as a proxy for disease severity.

Sometimes referred to as polygenic or genomic risk scores, GRS typically consist of a summarization of the number of risk alleles of known SNP for a particular disease, often weighted by the OR of each SNP⁵. In the Dominguez, *et al* study⁹, the scores that were used included both a simple risk allele count, not weighted for the SLE OR [genetic risk counting scores (GRCS)], and a weighted score [genetic risk weighted scores (GRWS)]. They were further divided into scores including SNP inside or outside of the HLA region. In addition, because the analyses were performed on a multiethnic cohort, with some ancestries having a substantially lower number of previously identified

This work was supported by the Swedish Society of Medicine and Ingegerd Johansson.

¹S. Reid, PhD Student, MD, Rheumatology and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden.

Address correspondence to Dr. S. Reid, Uppsala University, Medical Sciences, Akademiska sjukhuset, Ingång 40, 5 tr, 751 85 Uppsala, Sweden.
Email: sarah.reid@medsci.uu.se

See Genetic risk in SLE, page xxx

SNP than others, the authors also used a standardization of the scores [genetic risk standardized counting score (GRSCS) and genetic risk standardized weighted score (GRSWS)] to ensure a more equitable representation of all patients. The authors have gathered SLE-associated SNP through comprehensive literature review. Their use of OR from previously published studies, minimizing the risk of bias in the GRS, is important in a study not including a replication cohort.

The report includes 2 main findings. First, 3 of the 4 non-HLA GRS that were analyzed displayed significant association with a younger age at SLE diagnosis ($p_{GRCS} = 0.103$, $p_{GRSCS} = 0.007$, $p_{GRWS} = 0.011$, and $p_{GRSWS} = 0.001$). This result is in agreement with previous studies of GRS in SLE^{7,8,12}. Second, the authors demonstrate that conversely, all 4 analyzed HLA-GRS displayed association with a later age at SLE diagnosis ($p_{GRCS} = 0.049$, $p_{GRSCS} = 0.022$, $p_{GRWS} = 0.022$, and $p_{GRSWS} = 0.011$). The association between HLA-GRS and the age at SLE onset has not previously been explored, and this finding is both interesting and surprising. How should it be interpreted? First, it must be pointed out that the HLA region—although it is considered the most important genetic region for SLE development—is a complex region characterized by extreme structural variation and strong haplotype-specific linkage disequilibrium¹³. It is therefore hard to identify causal variants in this region, and although the effect size of SLE risk SNP in this region are typically larger, most risk variants are located outside of the HLA region. For these reasons, HLA-GRS tend to be small and in the present paper, between 1 and 6 SNP were included for each ancestry. Consequently, the contribution to the entire HLA-GRS from a single SNP in the score may be large, particularly if the SNP is common in the population and has a large weight. Thus, one explanation may be that the results involving the HLA-GRS are driven by one particular SNP. One of the classic HLA alleles most strongly associated with SLE is the DRB1*03:01 allele^{14,15}. This allele, in turn, is also associated with the development of Sjögren syndrome as well as a more “Sjögren-like” SLE phenotype, involving the development of SSA and SSB antibodies^{16,17}. Interestingly, several studies have indicated that patients with SLE displaying this particular subphenotype acquire their disease later in life^{11,18}. Could this HLA allele, here represented by the MSH5 risk SNP rs3131379 ($R^2 > 0.6$)¹⁹, be the cause of the association, or is it truly an effect of a cumulative genetic risk? This question requires further, thorough investigation.

Methodologically, the study⁹ is well executed, with rigorous quality control of the genetic data and appropriate selection of SNP and construction of the GRS. A criticism that may be directed at the study is the somewhat obscure presentation and interpretation of statistical data. First, although it is clear that the results demonstrate strong statistical significance, the presence of clinical significance is difficult to ascertain. The patients' age at diagnosis have been logarithmized: a statistical maneuver to allow parametric analysis. This approach is mathematically correct; however, it impedes the interpretation of effect sizes—a matter further complicated by the lack of presentation of the range, mean/median value, and SD/IQR of the GRS. A simple

way to facilitate interpretation, often used in GRS studies^{7,8,12}, is to divide the patients into tertiles, quartiles, or quintiles and compare the outcome between the 2 extreme groups. Second, the authors state that “overall genetic risk scores explained 18% of the variance of age of onset of SLE,” and on multiple occasions refer to the relatively high R^2 values from their multiple regression models including the GRS, ancestry, and sex as covariates. These values, however, represent the combined effect of the independent variables included in the model, rather than that of the GRS itself. How much each of these variables contribute to the result is not presented in the paper, but from the correlation lines in the supplementary figures, it is evident that the inclusion of ancestry as a covariate has a large effect on the model (likely a result of non-European ancestries having a larger proportion of SLE patients compared to the European patients). It thus appears plausible that the association between ancestry and age at diagnosis, rather than that between the GRS and the age at diagnosis, is the largest contributor to the relatively high explained variability. Last, the authors suggest that the GRSWS is the most robust score, and that “future studies use standardized GRS,” indicating that this score was superior to the others. However, as no statistical comparison of the scores was included to show that this is the case, no clear conclusion can be drawn as to whether the standardization actually improved the model.

Currently, one of the largest limitations in the studies of GRS is the history of underrepresentation of many ethnic groups in GWAS²⁰. This is clearly reflected in the number of SNP included for each ancestry in the Dominguez, *et al* report (Supplementary Table 3)⁹. Despite a higher prevalence of SLE in blacks and Hispanics, these populations are represented by 7 SNP and 5 SNP, respectively, out of the included total of 58. This highlights the strong need for future GWAS in SLE to emphasize these minorities. Similarly, previous studies of GRS in SLE have focused mainly on patients with a European ancestry. That makes the contribution of this multiethnic study valuable, although because no subgroup analyses were performed, it does not answer the question of whether the association between the cumulative genetic risk for SLE and the age at onset also apply to the non-European ancestries. In a previous study using a GRS for SLE generated from the top SNP in a Chinese population, the average GRS of an eastern Asian population of SLE patients was—not unexpectedly—higher than that of the European population. However, when using a GRS based on top SNP in the European population, the eastern Asian population still displayed substantially higher scores compared with the Europeans, indicating that their higher prevalence of SLE may be due to an increased genetic load²¹. Thus, it is tempting to assume that the association between the non-HLA GRS and an early age at onset would be more pronounced in non-European ancestries. This is, however, a question that deserves proper evaluation and perhaps something that the authors could investigate next.

Similarly, patients with cSLE have not previously been independently studied in this context, and the impressive cohort of nearly 700 patients who received a diagnosis before age 19 included in the study has a large potential for further

investigation using GRS. It has previously been suggested that GRS could be used in such a cohort in order to identify new genetic variants associated with the disease²². While the hereditary component appears highly polygenic for the majority of patients with SLE, the disease follows a Mendelian inheritance pattern for a small number of patients. These patients carry one of several rare, monogenetic variants conferring both a strong risk of disease and, typically, an early onset²³. As these individuals would be expected to have a polygenic risk matching that of healthy controls, selecting patients with both a low GRS and an early onset of the disease may lead to an enrichment of such variant and increase the chance of their identification. This may contribute both to an increased understanding of the etiology of SLE and to improvements in future polygenic risk assessment.

In conclusion, the Dominguez, *et al* study⁹ is in line with previous studies of GRS in SLE, indicating that patients with a high non-HLA genetic load are younger at diagnosis. As GRS are slowly moving towards clinical implementation in many fields of medicine, this confirmation is valuable. In addition, the authors present a previously unknown association between the cumulative HLA-driven genetic risk of SLE and an onset occurring later in life; however, further studies are required to elucidate the cause of this association.

REFERENCES

1. The Britannica guide to theories and ideas that changed the modern world. 1 ed. Kuiper K, editor. Britannica Educational Publishing; 2010.
2. Gayon J. From Mendel to epigenetics: history of genetics. C R Biol 2016;339:225-30.
3. Galton F. The history of twins, as a criterion of the relative powers of nature and nurture (1,2). Int J Epidemiol 2012;41:905-11.
4. Fisher RA. The causes of human variability. Eugen Rev 1919;10:213-20.
5. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. Hum Mol Genet 2019;28: R133-42.
6. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet 2012;90:7-24.
7. Reid S, Alexsson A, Frodlund M, Morris D, Sandling JK, Bolin K, et al. High genetic risk score is associated with early disease onset, damage accrual and decreased survival in systemic lupus erythematosus. Ann Rheum Dis 2020;79:363-9.
8. Chen L, Wang YF, Liu L, Bielowska A, Ahmed R, Zhang H, et al. Genome-wide assessment of genetic risk for systemic lupus erythematosus and disease severity. Hum Mol Genet 2020; 29:1745-56.
9. Dominguez D, Kamphuis S, Beyene J, Wither J, Harley JB, Blanco, et al. Relationship between genetic risk and age of disease diagnosis in systemic lupus erythematosus. J Rheumatol 2020;47:xxxxxxx.
10. Sousa S, Goncalves MJ, Ines LS, Eugenio G, Jesus D, Fernandes S, et al. Clinical features and long-term outcomes of systemic lupus erythematosus: comparative data of childhood, adult and late-onset disease in a national register. Rheumatol Int 2016;36:955-60.
11. Tomic-Lucic A, Petrovic R, Radak-Perovic M, Milovanovic D, Milovanovic J, Zivanovic S, et al. Late-onset systemic lupus erythematosus: clinical features, course, and prognosis. Clin Rheumatol 2013;32:1053-8.
12. Taylor KE, Chung SA, Graham RR, Ortmann WA, Lee AT, Langefeld CD, et al. Risk alleles for systemic lupus erythematosus in a large case-control collection and associations with clinical subphenotypes. PLoS Genet 2011;7:e1001311.
13. Kennedy AE, Ozbek U, Dorak MT. What has gwas done for hla and disease associations? Int J Immunogenet 2017;44:195-211.
14. Morris DL, Taylor KE, Fernando MM, Nititham J, Alarcon-Riquelme ME, Barcellos LF, et al. Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans. Am J Hum Genet 2012;91:778-93.
15. Kim I, Kim YJ, Kim K, Kang C, Choi CB, Sung YK, et al. Genetic studies of systemic lupus erythematosus in Asia: where are we now? Genes Immun 2009;10:421-32.
16. Imgenberg-Kreuz J, Rasmussen A, Sivits K, Nordmark G. Genetics and epigenetics in primary sjogren's syndrome. Rheumatology 2019 Feb 15 (E-pub ahead of print).
17. Morris DL, Fernando MM, Taylor KE, Chung SA, Nititham J, Alarcon-Riquelme ME, et al. MHC associations with clinical and autoantibody manifestations in European SLE. Genes Immun 2014;15:210-7.
18. Lazaro D. Elderly-onset systemic lupus erythematosus: prevalence, clinical course and treatment. Drugs Aging 2007;24:701-15.
19. Gourraud PA, Khankhanian P, Cereb N, Yang SY, Feolo M, Maier M, et al. Hla diversity in the 1000 genomes dataset. PLoS One 2014;9:e97282.
20. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Med 2020;12:44.
21. Morris DL, Sheng Y, Zhang Y, Wang YF, Zhu Z, Tomblinson P, et al. Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. Nat Genet 2016;48:940-6.
22. Johnson MB, Patel KA, De Franco E, Houghton JAL, McDonald TJ, Ellard S, et al. A type 1 diabetes genetic risk score can discriminate monogenic autoimmunity with diabetes from early-onset clustering of polygenic autoimmunity with diabetes. Diabetologia 2018;61:862-9.
23. Alperin JM, Ortiz-Fernandez L, Sawalha AH. Monogenic lupus: a developing paradigm of disease. Front Immunol 2018;9:2496.