

been proposed to underlie association patterns similar to that of *CREBRF* p.Arg457Gln^{14,15}.

Another question is why the p.Arg457Gln variant was eliminated from almost all other populations. Furthermore, do other genetic variants in *CREBRF* influence obesity and diabetes risk? Publicly available data from large-scale genetic consortia suggest that there is currently little evidence that common *CREBRF* variants are associated with BMI, diabetes risk or glucose levels (see URLs). However, more than 100 low-frequency and rare missense and loss-of-function variants in *CREBRF* have been reported⁴ that may show associations once large enough studies become available.

The increased interest in precision medicine might make one wonder whether *CREBRF* could be a drug target. Given the opposing

effects of at least the p.Arg457Gln variant on obesity and diabetes risk, it might not be an ideal candidate. Nevertheless, by employing a non-traditional study design that leverages the distinct genomic features of unique populations, Minster *et al.* demonstrate that even small samples have the power to discover new disease-causing genes that are not identified by traditional large-scale studies but that may have an impact on the health of populations worldwide.

URLs. T2D-GENES Consortium, GoT2D Consortium and Diagram Consortium, <http://www.type2diabetesgenetics.org/>.

COMPETING FINANCIAL INTERESTS

The author declares no competing financial interests.

1. NCD Risk Factor Collaboration (NCD-RisC). *Lancet* **387**, 1377–1396 (2016).
2. Hawley, N.L. *et al.* *Am. J. Hum. Biol.* **26**, 491–501 (2014).
3. Minster, R.L. *et al.* *Nat. Genet.* **48**, 1049–1054 (2016).
4. Lek, M. *et al.* *Nature* <http://dx.doi.org/10.1038/nature19057> (2016).
5. Frayling, T.M. *et al.* *Science* **316**, 889–894 (2007).
6. Locke, A.E. *et al.* *Nature* **518**, 197–206 (2015).
7. NCD Risk Factor Collaboration (NCD-RisC). *Lancet* **387**, 1513–1530 (2016).
8. Winkler, T.W. *et al.* *PLoS Genet.* **11**, e1005378 (2015).
9. Neel, J.V. *Am. J. Hum. Genet.* **14**, 353–362 (1962).
10. Speakman, J.R. *Int. J. Obes.* **32**, 1611–1617 (2008).
11. Baier, L.J. *et al.* *Diabetes* **64**, 4322–4332 (2015).
12. Moltke, I. *et al.* *Nature* **512**, 190–193 (2014).
13. Styrkarsdottir, U. *et al.* *Nature* **497**, 517–520 (2013).
14. Kilpeläinen, T.O. *et al.* *Nat. Genet.* **43**, 753–760 (2011).
15. Yaghootkar, H. *et al.* *Diabetes* **63**, 4369–4377 (2014).

Genomic landscape of the Greater Middle East

Tayfun Özçelik & Onur Emre Onat

Study of the Greater Middle East (GME), home to approximately 10% of the world's population, has made invaluable contributions to the characterization of rare genetic disease, especially recessive conditions arising from the tradition of consanguinity and large families with multiple children. A new study now reports 1,111 unrelated exomes from the GME and provides a comprehensive view of genetic variation for enhanced discovery of disease-associated genes.

In a study appearing on page 1071, the GME Variome Consortium collected whole-exome data from populations across the GME region, including Northwest and Northeast Africa, the Turkish Peninsula, the Syrian Desert, the Arabian Peninsula, and Persia and Pakistan, and compared these with data from 1000 Genomes Project populations¹. Scott *et al.* detected tight clusters of European and Asian populations, and high levels of divergence among the GME regions. Populations from Northwest Africa, the Arabian Peninsula, and Persia and Pakistan were found to be the least admixed, suggesting that they are founder populations. When patterns of human migration and drift were recapitulated, a west-to-east trajectory was observed, with the GME populations grouped with European populations. Persia and Pakistan to the east and Northwest Africa to the west represented the extremes of the identified subregions and showed the

highest degree of differentiation, comparable to the distance between the Finnish (FIN) and Toscani (TSI) 1000 Genomes Project populations. Data from the Syrian Desert and Turkish Peninsula populations were consistent with higher levels of European admixture.

Consanguinity is a hallmark of the GME

The GME has a roughly 100-fold higher rate of consanguinity than the United States and Europe. Because of historical and contemporary cultural preferences, between 20–50% of all marriages are consanguineous. With the additional burden of endogamy, leading to homozygosity rates greater than those predicted by the degree of kinship alone², the rate of recessive Mendelian disease is roughly doubled in GME populations. The exome-wide view of the GME genomic landscape provided by Scott *et al.*¹ showed that estimated inbreeding coefficients (*F*) ranged from 0.059 to 0.098, around 10- to 20-fold higher than those in European, African and East Asian 1000 Genomes Project populations, where *F* is ~0.005. As expected, family structure determined *F* values, with offspring from first-cousin marriages displaying higher *F* values than those from non-consanguineous marriages. Increased burden and length of runs of homozygosity (ROHs) were also observed.

The increased burden was most striking for ROHs of long length (>1.6 Mb), with runs greater than 4 Mb in length nearly exclusive to GME samples. The authors capitalized on this increased burden of ROHs and searched for homozygous loss-of-function variants in a subgroup of 354 exomes from healthy adults. They identified rare homozygous putative loss-of-function variants in 301 genes, of which a substantial proportion were novel, and demonstrate the value of consanguineous GME populations in cataloging homozygous loss-of-function mutations. However, neither a measurable effect on overall variant burden resulting from consanguinity nor evidence for purging of recessive alleles was detected.

Dawn of reverse genetics

Early efforts to identify disease-related genes relied on finding a genomic region with a chromosomal abnormality³ or linking a phenotype to a polymorphic marker⁴. Then came the painstaking process of mapping the critical region to pinpoint the causal variant so that the underlying physiology and pathological processes could be studied, diagnostic tests offered and targeted treatment strategies developed. In the 1980s, this strategy was celebrated as 'reverse genetics' (ref. 5). This was the revolution that paved the way

Tayfun Özçelik and Onur Emre Onat are in the Department of Molecular Biology and Genetics, Faculty of Science, Bilkent University, Ankara, Turkey. Tayfun Özçelik is also at the Institute of Materials Science and Nanotechnology, National Nanotechnology Research Center (UNAM), Bilkent University, Ankara, Turkey. e-mail: tozcelik@bilkent.edu.tr

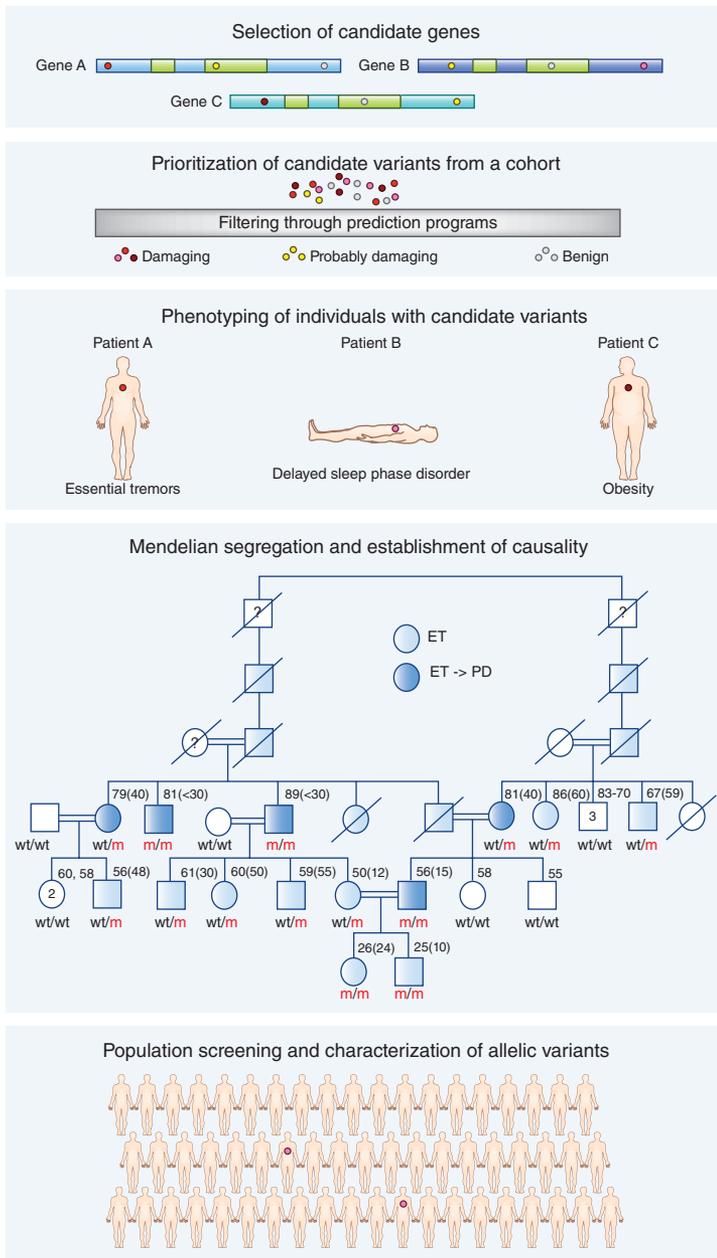


Figure 1 Reverse phenotyping of complex diseases. In reverse phenotyping, the process starts with selection of candidate genes followed by prioritization of variants in these genes identified in a cohort, for which the GME Variome, Exome Aggregation Consortium (ExAC) and Centers for Mendelian Genomics databases constitute excellent resources. Individuals who carry the prioritized mutations are assessed clinically to determine phenotype. A pedigree is constructed for the individual in whom the initial genotype–phenotype correlation was established (the pedigree shown was adapted from ref. 10). Large consanguineous families are very important, if not essential, for segregation analysis. Finally, population screening for the mutant allele is carried out in large and phenotypically well-characterized cohorts, followed by targeted sequencing of the candidate gene in mutation-negative individuals from the same cohort to determine the spectrum of mutations. wt, wild type; m mutant; ET, essential tremor; PD, Parkinson disease. Age (age at onset) are shown above pedigree symbols.

for future genome projects. Now, in the age of next-generation sequencing, identification of causal variants relies on prioritization of alleles that pass filtering criteria along with comparison of their frequencies to the distribution of derived allele frequencies in population databases. To test the potential

of the GME Variome in aiding in the discovery of disease-associated variants, Scott *et al.*¹ concentrated on existing data for recessive hereditary spastic paraplegia, reporting that using the GME Variome data reduced the number of variants requiring further consideration by four- to sevenfold. Although this is

an impressive demonstration of the value of the GME data in expediting the discovery of new genes with rare disease-associated variants, it represents only one small example of the full potential of this database.

Mendel to precision medicine

Continuing along the path set in the 1980s, ‘reverse phenotyping’ of a candidate genomic variant for a complex disease followed by clinical assessments in consanguineous families can be considered the new ‘forward genetics’. This approach, however, can be confounded by the inherent problem of phenocopies among common diseases (Fig. 1).

We are now entering into shades of gray in the black-and-white world of Mendelian inheritance. In the years following the adoption of widespread use of next-generation sequencing, an extreme degree of allelic, locus and phenotypic heterogeneity has emerged as the rule rather than the exception, with this heterogeneity reaching levels suggesting that causality may never be resolved by large-scale association or case–control studies⁶. This possibility is further corroborated by the increased recognition of *de novo*⁷ or early postzygotic⁸ mutations, shifting the focus from the whereabouts of the culprits to the timing of their emergence. It is in this setting that study of consanguineous families becomes all the more important to follow the segregation of an allele through generations while concentrating on its zygosity. Homozygotes may present with an early-onset form of an extreme phenotype, whereas heterozygotes may present with a milder form of the phenotype during later decades of life. Examples of phenotypes with this pattern include lipid metabolism disorders⁹, movement disorders¹⁰ and familial cancers¹¹. The way forward in clinical delivery and informatics is clear in the age of precision medicine. Characterization of GME genetic variation promises to contribute to the discovery of phenotype-associated genes, including for notoriously difficult-to-crack complex phenotypes.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Scott, E.M. *et al.* *Nat. Genet.* **48**, 1071–1076 (2016).
2. Özçelik, T. *et al.* *Nat. Genet.* **42**, 641–645 (2010).
3. Francke, U. *et al.* *Am. J. Hum. Genet.* **37**, 250–267 (1985).
4. Hall, J.M. *et al.* *Science* **250**, 1684–1689 (1990).
5. Orkin, S.H. *Cell* **47**, 845–850 (1986).
6. McClellan, J. & King, M.C. *Cell* **141**, 210–217 (2010).
7. Gulsuner, S. *et al.* *Cell* **154**, 518–529 (2013).
8. Dal, G.M. *et al.* *J. Med. Genet.* **51**, 455–459 (2014).
9. Cohen, J.C. *et al.* *Science* **305**, 869–872 (2004).
10. Unal, Gulsuner, H. *et al.* *Proc. Natl. Acad. Sci. USA* **111**, 18285–18290 (2014).
11. Ricciardone, M.D. *et al.* *Cancer Res.* **59**, 290–293 (1999).