

RESEARCH HIGHLIGHT

Using genetically isolated populations to understand the genomic basis of disease

Eleftheria Zeggini

Abstract

Rare variation has a key role in the genetic etiology of complex traits. Genetically isolated populations have been established as a powerful resource for novel locus discovery and they combine advantageous characteristics that can be leveraged to expedite discovery. Genome-wide genotyping approaches coupled with sequencing efforts have transformed the landscape of disease genomics and highlight the potentially significant contribution of studies in founder populations.

Complex trait locus discovery in isolated populations

Genetically isolated or founder populations have recently returned to the fore of genetic association studies as valuable resources for complex trait gene identification [1]. Population isolates have well-documented characteristics, including reduced phenotypic, environmental and genetic heterogeneity, that can aid in the detection of rare variants associated with complex traits. In isolated populations, where a relatively small number of individuals found a new population, rare variants that were present in the founders can drift up in frequency as the population expands, thus increasing power for genetic association studies. The small effective population size, which remains small over time, leads to increased levels of homozygosity and linkage disequilibrium. In addition, isolated population cohorts often provide the opportunity to recall subjects by genotype, access detailed genealogical records, obtain linkage to health records and follow the cohort longitudinally.

Recent successes in the literature have highlighted how these advantageous characteristics can help with disease gene mapping. Researchers studying the Icelandic population have, in recent years, pioneered the

use of next-generation association studies, a hybrid of genome-wide genotyping and whole genome sequencing (WGS) approaches, for complex disease gene mapping [2,3]. In Iceland, numerous novel loci for complex diseases, such as type 2 diabetes (T2D) and prostate cancer [4,5], have been identified through a combination of WGS and long-range phasing-assisted imputation on a genome-wide genotype scaffold, together with calculation of genotype probabilities in approximately 300,000 untyped individuals by making use of the extended genealogical information available.

More recently, novel insights into the biological pathways underpinning T2D were achieved through the study of a Greenlandic founder population [6]. A nonsense variant in the *TBC1D4* gene was found to be strongly associated with postprandial hyperglycemia, impaired glucose tolerance and T2D. These unique insights into the mechanism conferring muscle insulin resistance for this subset of T2D was afforded by studying the small Greenlandic population, which has experienced a dramatic increase in T2D prevalence, and recalling individuals based on their *TBC1D4* variant status. This polymorphism is common in Greenland (17% minor allele frequency), but vanishingly rare in other global populations (only encountered in one Japanese individual in the 1000 Genomes Project data). This work elegantly demonstrates the value of combining the genetic characteristics of founder populations with the potential to recontact participants for further follow-up of promising results. Studies in extensively phenotyped founder population cohorts, such as the Amish, have also demonstrated the value of combining unique population characteristics with recall of subjects to increase our understanding of disease etiopathology. The Old Order Amish are a cultural isolate and geographically localized, genetically homogeneous population with extensive genealogical records available. This deeply phenotyped cohort has been the subject of long-term genetic studies. For example, in 2008, Pollin *et al.* [7] reported a missense variant (R19X) that abolishes expression of the

Correspondence: Eleftheria@sanger.ac.uk
Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK



© 2014 Zeggini; licensee BioMed Central Ltd. The licensee has exclusive rights to distribute this article, in any medium, for 12 months following its publication. After this time, the article is available under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

APOC3 gene and is strongly associated with a cardioprotective phenotype (higher high-density lipoprotein and lower blood triglyceride levels).

Notably, the same missense cardioprotective variant was also found in an independent isolated population from Greece in the HELIC-MANOLIS study [8]. Residents of the mountainous Mylopotamos villages on Crete have a high fat content diet but anecdotally display lower levels of, for example, T2D complications compared with the general population. The R19X *APOC3* variant was carried by approximately 4% of the individuals studied and reached genome-wide statistical significance with a sample size of fewer than 1,300. Discovery of the same effect in the general population would have required over 50 times the number of subjects. Large-scale studies of over 110,000 individuals of European descent have recently also established an association of rare variants in the *APOC3* locus with protection against high triglyceride levels and coronary artery disease [9]. *APOC3* is now becoming a poster child for the power afforded by founder populations and clearly demonstrates the generalizability of findings in isolates into more cosmopolitan populations.

A prime example of how founder population characteristics coupled with linkage to medical records can accelerate discovery was recently produced by studying the Finnish population [10]. In a whole exome sequencing study of about 3,000 Finns, Lim *et al.* first established that the Finns have fewer variable sites overall but more loss-of-function variants compared with non-Finnish European individuals, and subsequently identified robust associations with key traits of medical relevance. Linkage to national medical records resulted in the demonstration that splice variants in the *LPA* gene that are associated with low levels of plasma lipoprotein(a) confer protection against cardiovascular disease.

Future directions

Going forward, it is clear that founder populations can provide a unique and powerful resource for the identification of low frequency and rare variants of direct medical consequence. Power to detect association is demonstrably boosted for individual sequence variants that have drifted up in frequency. In addition, power to detect a significant accumulation of rare variants at particular loci is further increased in founder populations as neutral rare variation may be lost from the haplotype pool. In this context, meta-analysis at the locus level across different isolates is posited to be important for establishing burden of proof, although this principle requires empirical substantiation. Historically, the transferability of findings in isolates across to more cosmopolitan populations has been a topic of debate. However, there is an accrual of emerging examples of loci

discovered in founder populations that are more widely generalizable, with replication of signals achieved in diverse sample sets [4,5,7-9]. Furthermore, invaluable and unprecedented insights into disease pathogenesis can be afforded by findings restricted to genetically isolated populations, as exemplified by the elegant metabolic trait study in Greenland [6]. Decreasing costs for deep whole genome sequencing and the increasing availability of deeply phenotyped genetically isolated cohorts sets the scene for further success stories in the near future.

Abbreviations

T2D: Type 2 diabetes; WGS: Whole genome sequencing.

Competing interests

The author has no competing interests to declare.

Published online: 17 October 2014

References

1. Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, Daly MJ, Neale BM, Sunyaev SR, Lander ES: **Searching for missing heritability: designing rare variant association studies.** *Proc Natl Acad Sci U S A* 2014, **111**:E455–E464.
2. Holm H, Gudbjartsson DF, Sulem P, Masson G, Helgadóttir HT, Zanon C, Magnusson OT, Helgason A, Saemundsdóttir J, Gylfason A, Stefansdóttir H, Gretarsdóttir S, Matthiasson SE, Thorgeirsson GM, Jonasdóttir A, Sigurdsson A, Stefansson H, Werge T, Rafnar T, Kiemeneý LA, Parvez B, Muhammad R, Roden DM, Darbar D, Thorleifsson G, Walters GB, Kong A, Thorsteinsdóttir U, Arnar DO, Stefansson K: **A rare variant in MYH6 is associated with high risk of sick sinus syndrome.** *Nat Genet* 2011, **43**:316–320.
3. Zeggini E: **Next-generation association studies for complex traits.** *Nat Genet* 2011, **43**:287–288.
4. Steinthorsdóttir V, Thorleifsson G, Sulem P, Helgason H, Grarup N, Sigurdsson A, Helgadóttir HT, Johannsdóttir H, Magnusson OT, Gudjonsson SA, Justesen JM, Harder MN, Jørgensen ME, Christensen C, Brandslund I, Sandbæk A, Lauritzen T, Vestergaard H, Linneberg A, Jørgensen T, Hansen T, Daneshpour MS, Fallah MS, Hreidarsson AB, Sigurdsson G, Azizi F, Benediktsson R, Masson G, Helgason A, Kong A, *et al*: **Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes.** *Nat Genet* 2014, **46**:294–298.
5. Gudmundsson J, Sulem P, Gudbjartsson DF, Masson G, Agnarsson BA, Benediktsson KR, Sigurdsson A, Magnusson OT, Gudjonsson SA, Magnusdóttir DN, Ohannsdóttir H, Helgadóttir HT, Stacey SN, Jonasdóttir A, Olafsdóttir SB, Thorleifsson G, Jonasson JG, Tryggvadóttir L, Navarrete S, Fuertes F, Helfand BT, Hu Q, Csiki IE, Mates IN, Jíngá V, Aben KK, van Oort IM, Vermeulen SH, Donovan JL, Hamdy FC, *et al*: **A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer.** *Nat Genet* 2012, **44**:1326–1329.
6. Moltke I, Grarup N, Jørgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, Korneliusen TS, Andersen MA, Nielsen TS, Krarup NT, Gjesing AP, Zierath JR, Linneberg A, Wu X, Sun G, Jin X, Al-Aama J, Wang J, Borch-Johnsen K, Pedersen O, Nielsen R, Albrechtsen A, Hansen T: **A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes.** *Nature* 2014, **512**:190–193.
7. Pollin TI, Damcott CM, Shen H, Ott SH, Shelton J, Horenstein RB, Post W, McLennan JC, Bielak LF, Peyser PA, Mitchell BD, Miller M, O'Connell JR, Shuldiner AR: **A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection.** *Science* 2008, **322**:1702–1705.
8. Tachmazidou I, Dedoussis G, Southam L, Farmaki AE, Ritchie GR, Xifara DK, Matchan A, Hatzikotoulas K, Rayner NW, Chen Y, Pollin TI, O'Connell JR, Yerges-Armstrong LM, Kiagiadaki C, Panoutsopoulou K, Schwartzentruber J, Moutsianas L, Tsafantakis E, Tyler-Smith C, McVean G, Xue Y, Zeggini E, UK10K consortium: **A rare functional cardioprotective APOC3 variant has risen in frequency in distinct population isolates.** *Nat Commun* 2013, **4**:2872.

9. Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitzel NO, Lange LA, Lu Y, Tang ZZ, Zhang H, Hindy G, Masca N, Stirrups K, Kanoni S, Do R, Jun G, Hu Y, Kang HM, Xue C, Goel A, Farrall M, Duga S, Merlini PA, Asselta R, Girelli D, Olivieri O, Martinelli N, Yin W, Reilly D, Speliotes E, *et al*: **Loss-of-function mutations in APOC3, triglycerides, and coronary disease.** *N Engl J Med* 2014, **371**:22–31.
10. Lim ET, Wurtz P, Havulinna AS, Palta P, Tukiainen T, Rehnstrom K, Esko T, Magi R, Inouye M, Lappalainen T, Chan Y, Salem RM, Lek M, Flannick J, Sim X, Manning A, Ladenvall C, Bumpstead S, Hämmäläinen E, Aalto K, Maksimow M, Salmi M, Blankenberg S, Ardisino D, Shah S, Horne B, McPherson R, Hovingh GK, Reilly MP, Watkins H, *et al*: **Distribution and medical impact of loss-of-function variants in the Finnish founder population.** *PLoS Genet* 2014, **10**:e1004494.

doi:10.1186/s13073-014-0083-5

Cite this article as: Zeggini: Using genetically isolated populations to understand the genomic basis of disease. *Genome Medicine* 2014 **6**:83.