

Published in final edited form as:

Nat Genet. ; 43(4): 287–288. doi:10.1038/ng0411-287.

Next-generation association studies for complex traits

Eleftheria Zeggini

Abstract

A new study successfully applies complementary whole-genome sequencing and imputation approaches to establish robust disease associations in a population isolate. This strategy is poised to help elucidate the genetic architecture of complex traits in the low end of the allele frequency spectrum.

Genome-wide association studies (GWAS) of complex traits have been successful in identifying common variant associations, but a substantial heritability gap remains. The field of complex trait genetics is shifting towards the study of low-frequency (minor allele frequency [MAF] 0.01-0.05) and rare (MAF<0.01) variants, which are hypothesised to have larger effects. The study of these variants can be empowered by focusing on isolated populations. On page xxx of this issue, Hilma Holm and colleagues¹ apply an innovative study design (Figure 1) that combines whole-genome sequence (WGS) and GWAS data from Icelandic individuals to detect a novel sick sinus syndrome (SSS) susceptibility locus (*MYH6*).

Holm et al¹ first identified a signal comprising multiple low frequency SNPs (risk allele frequency [RAF] 0.01-0.26) on chr14 through GWAS and imputation of 1000 Genomes Project² variants into a study sample of 792 cases compared to 37,585 controls. To refine this association 7 cases and 80 controls were whole-genome sequenced at a mean depth of 10x. Eleven million discovered variants were then imputed into the full set of GWAS samples using long-range haplotype phasing approaches³, and strong association was established with a missense mutation, R721W, in *MYH6* (RAF 0.0038, OR 12.53). No significant association remained at chr4 after accounting for R721W. The authors subsequently validated the association by directly-typing this variant in the discovery samples, and by replicating the association in an independent modestly-sized set of 469 Icelandic cases and 1,185 controls (RAF 0.0021, OR 12.59). R721W is not present in the HapMap or 1000 Genomes Project data and was not identified in any of the 1,776 European-descent controls or 135 US cases investigated¹. The lifetime risk of disease is 50% for carriers and 6% for non-carriers and the R721W sibling recurrence risk ratio (λ_s) for SSS is 1.52, substantially higher than most complex disease common-frequency risk loci. This is likely to be an Iceland-specific variant, estimated to have arisen approximately 30 generations ago.

Next-generation study design

This new association is an almost ideal poster child for the type of variants that people hope to discover via sequence-based association, with a low population allele frequency between 0.1% and 1%, and strong effect of the order of ten-fold relative risk.

The work of Holm and colleagues¹ applies a novel and powerful approach for complex trait association studies (Figure 1), representing state-of-the-art at least until WGS costs enable sequencing of the full study sample. For example, simulations⁴ suggest that sequencing 200 individuals from an outbred European-descent population at 6× depth can provide near-comprehensive coverage of variants down to MAF 0.01 and ~40% of variants with MAF

between 0.001 and 0.01. Imputation approaches can then take advantage of variants discovered within the study sample subset, in conjunction with publicly available reference sets, to infer untyped variants in the full dataset using GWAS data as a scaffold. Indeed, this represents one of the key aims of large-scale investments to create repositories of whole genome sequence variation, such as the 1000 Genomes² and UK10K (www.uk10k.org) Projects.

Holm's study also highlights the added value to be gained by focusing on well-characterised sample collections with deep phenotype data. When information on multiple quantitative or dichotomous traits is available, sequencing a representative, carefully-selected subset of the study sample (for example to maximise imputation efficiency by preferentially targeting distantly related individuals for the WGS set) can enable downstream association testing for a wide variety of phenotypes.

Low frequency and rare variants of large effect size, like the R721W mutation in *MYH6*, represent low-hanging fruit amenable to easy detection by applying this powerful strategy. The analytical toolset for straightforward single-variant complex-trait association probing can be borrowed from GWAS and adapted to meet truly full genome scales. The field of statistical genetics is also actively developing more sophisticated methods that consider the aggregation of rare variants within functional units of interest⁵. Incorporation of functional annotation currently represents a challenge for interpreting association at variants, for which putative consequences are not as readily inferred as for missense mutations.

Added value: population isolates

Holm and colleagues¹ have also taken advantage of the increased discovery power afforded by studying a population isolate. Population isolates have well-documented characteristics that can aid in the detection of low frequency and rare variants in complex disorders⁶, namely reduced phenotypic, environmental and genetic heterogeneity. Population bottlenecks and subsequent expansion from a relatively small number of founders can lead to increased genetic homogeneity and a rise in rare variant allele frequency. These characteristics are complemented by elevated levels of linkage disequilibrium, which enable long-range haplotype matching³ and accurate imputation.

A potential drawback is uncertainty about the generalisability of findings in further populations. For example, it remains to be seen whether different variants within *MYH6* are associated with SSS outside of Iceland. Nevertheless, the detected association, even if not directly transferrable beyond the study sample population, provides novel insights into the biology of disease. It is widely accepted that one of the main reasons behind decades of unsuccessful candidate gene studies has been poor understanding of the underlying disease aetiopathology, a point clearly-illustrated by a host of unexpected GWAS findings. Importantly, to establish the biological underpinnings of pathogenicity, functional studies tailored to the phenotype under study are required.

The work of Holm et al¹ provides proof-of-principle for a powerful WGS-based study design paradigm that leverages population genetics characteristics of isolates to reduce the number of samples to be sequenced without compromising power. Using cutting-edge high-throughput sequencing technologies and analytical tools, this work ushers in the new era of next-generation genetic studies that will help fill the missing heritability gap for disease-related complex traits.

References

1. Holm H, et al. *Nat Genet.* 2011; 43:316–320. [PubMed: 21378987]

2. 1000 Genomes Project Consortium. *Nature*. 2010; 467:1061–1073. [PubMed: 20981092]
3. Kong A, et al. *Nat Genet*. 2008; 40:1068–1075. [PubMed: 19165921]
4. Le SQ, Durbin R. *Genome Res*. 2010 PMID: 20980557.
5. Asimit J, Zeggini E. *Annu Rev Genet*. 2010; 44:293–308. [PubMed: 21047260]
6. Peltonen L, et al. *Nat Rev Genet*. 2000; 1:182–190. [PubMed: 11252747]

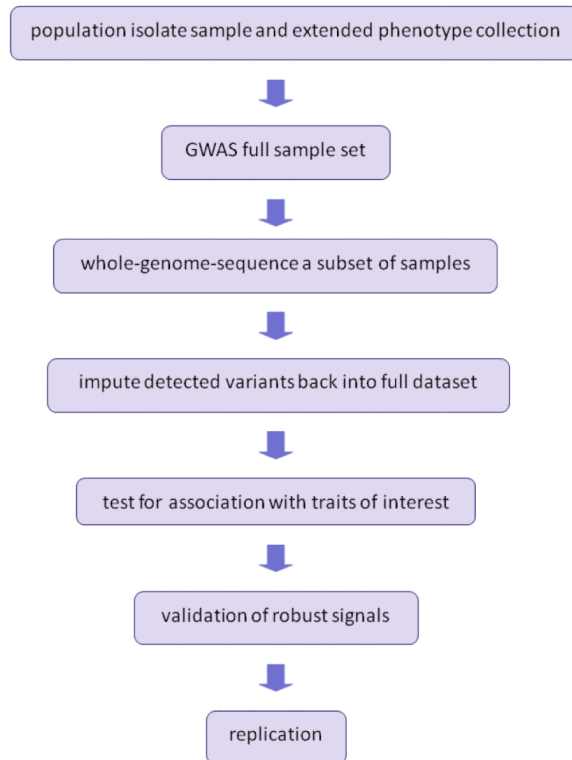


Figure 1. Overview of the steps involved in a next-generation complex trait association study.