Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Review

The genetic architecture of molecular traits

Annique Claringbould[1], Niek de Klein[1] and Lude Franke

Abstract

Most diseases have both an environmental and genetic component. Although many diseases are strongly heritable, individual genetic variants typically confer only a small effect on disease, and thus these diseases are strongly polygenic. Paradoxically, molecular traits, such as gene expression, methylation, protein or metabolite levels, typically have a lower heritability, but sometimes individual genetic variants show much higher effect sizes on these traits. In this review, we discuss the genetic architecture of these molecular traits, and contrast this to the genetic architecture of complex diseases, and provide explanations why strong effects of individual genetic variants on molecular traits do not necessarily need to translate into increased risk of disease.

Addresses

University of Groningen, University Medical Centre Groningen, Department of Genetics, Antonius Deusinglaan 1, 9713 AV, Groningen, The Netherlands

Corresponding author: Franke, Lude (lude@ludesign.nl)

[1] Equal contribution.

Current Opinion in Systems Biology 2017, **1**:25–31

This review comes from a themed issue on **Genomics and epigenomics**

Edited by **Emmanouil Dermitzakis** and **Tuuli Lappalainen**

For a complete overview see the Issue and the Editorial

Available online 6 March 2017

http://dx.doi.org/10.1016/j.coisb.2017.01.002

Introduction

Since most diseases, phenotypes, and molecular traits are heritable, their genetic architecture is a long-standing question. Are most phenotypes caused by a limited number of large effect variants, or are they due to many variants that each have a small effect? In this review, we compare detected effect size and allele frequencies of associated variants from genome-wide association studies (GWAS) on complex traits and diseases with results from expression- and methylation quantitative trait locus (eQTL and meQTL) studies.

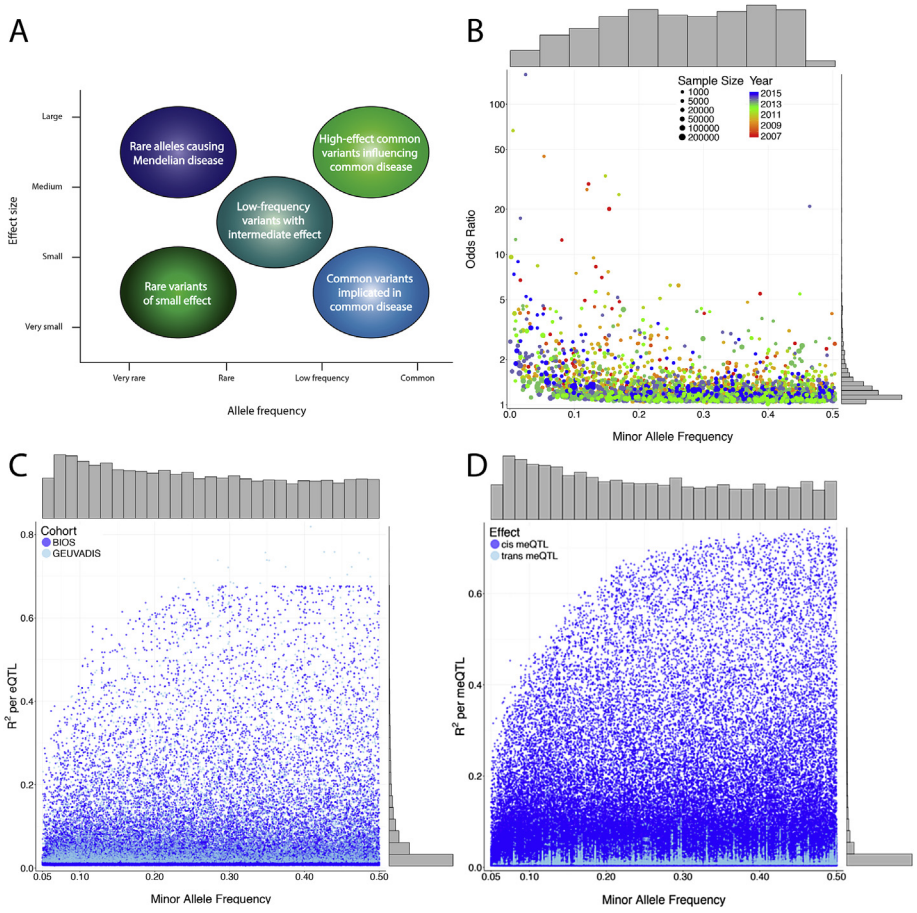
Genetic architecture of common disease: few large effects

Over the last few years, GWAS have shown that associated variants typically only explain a small proportion of the disease variation seen in the majority of common diseases, despite a few examples of common variants with a large effect being found for immune-related traits, particularly in the HLA-region [1]. Currently, variants with an odds ratio (OR) > 10 are considered 'highly unusual' [2]. These observations led to an extensive debate about the genetic architecture of complex diseases [2–6], including issues like the number, frequency and effect size of associated genetic variants, as well as the degree of shared genetic background with other traits [7]. For complex diseases the architecture is by no means uniform: the number of genes and their effect sizes differ widely [5,6,8,9]. However, there is evidence for a shared genetic basis among many diseases [10–12], and the genetic architecture of most complex diseases seems to be highly polygenic.

A much cited interpretation of the overall genetic architecture of diseases places genetic variants in five different groups, based on their allele frequency and effect size or penetrance (Figure 1A, adapted from Refs. [2,13]). In this representation, complex diseases are characterized by many common genetic variants with small effect sizes, whereas Mendelian diseases are caused by rare variants with large effects. Subsequently, methods were developed that can infer the variance explained by using all directly genotyped SNPs, including those that do not attain genome-wide significance. For complex phenotypes such as height it was established that a considerable proportion of the heritability could be explained by common SNPs, suggesting a highly polygenic genetic architecture [14,15].

Indeed, an inventory of the binary traits in the GWAS Catalog (v1.01, r2016-06-12, $p < 5 \times 10^{-8}$, Supplementary Note 1) reveals that most identified SNPs are common (minor allele frequency (MAF) > 0.1) and have small effect sizes (OR between 1.0 and 1.2, Figure 1B). The expectation was that by increasing sample sizes, GWAS would also allow for finding the intermediate frequency variants ($0.005 < \text{MAF} < 0.1$) with larger effects on disease. However, even very large studies have not yet been able to identify many of these hypothesized large effect variants [6] (Figure 1B), suggesting that the missing heritability may rather be explained by large numbers of small effects [15]. On the other hand, recent studies show that relatively rare variants that explain some of the missing heritability can

Figure 1



(A) Proposed genetic architecture of diseases, adapted from Refs. [2,13]. **(B)** Minor allele frequency (MAF) set out against odds ratio (OR) of genome-wide significant GWAS SNPs. The data is downloaded from the GWAS catalog (Supplementary Note 1). Histograms on the right and at the top indicate the frequency distribution of the SNPs, the dot size indicates the total sample size of the GWAS, and the color represents the year of publication. Rare and intermediate frequency variants (MAF < 0.1) have a higher OR on average. **(C)** MAF against variance explained (R^2) of *cis* expression quantitative trait locus (eQTL) SNPs. Light blue SNPs are from the GEUVADIS consortium (N = 373), dark blue SNPs from the BIOS consortium (N = 2116) (Supplementary Note 2). The plots on the top and right of the figure illustrate the density distribution of SNPs from both cohorts. Despite different sample sizes, the distribution of the *cis* eQTL SNPs is similar for both cohorts. **(D)** MAF against variance explained (R^2) of *cis* and *trans* methylation quantitative trait locus (eQTL) SNPs. Dark blue SNPs are *cis* meQTLs and light blue SNPs are *trans* meQTLs (Supplementary Note 3). Histograms on the right and at the top indicate the frequency distribution of the SNPs. Common SNPs often have large effect sizes, and *trans* effects explain much less methylation variation on average.

successfully be detected by targeted or whole-genome sequencing [16,17].

Genetic architecture of molecular traits

Surprisingly, the genetic architecture of complex phenotypic (disease) traits differs substantially from the genetic architecture of molecular traits such as gene expression, methylation, or protein levels. Although the genetic architecture of molecular traits can also be polygenic, a single SNP can often explain a considerable part of the heritability compared to disease phenotypes, whereas the heritability of gene expression, methylation or protein levels is typically lower than complex diseases [18,19].

Large effect-sizes of SNPs affect molecular traits

Genetic variation influences the risk of developing a complex disease through several molecular traits such as gene expression [20] and methylation [21]. Investigating eQTLs [22–27] and allele-specific expression (ASE) [28–31] can characterize the effect of both common and rare genetic variants on (individual) gene expression patterns. Similarly, intermediate frequency and common SNPs that affect methylation levels at CpG sites can be detected by meQTL mapping [32].

Analogous to the genetic architecture of common diseases presented in Figure 1B, the genetic architecture of gene expression and methylation may be represented by