How Many Genes Does It Take?

Clinical genetics examine the pedigrees of families with a patient according to the mendelian framework of autosomal recessive, dominant, or X-linked inheritance. In case no evidence for inheritance can be found, the patient is considered to be sporadic and a search for de novo copy number variants (CNVs) or single nucleotide variants (SNVs) is initiated. Implicitly, clinical genetics follows the paradigm that one mutated gene must be involved in one disorder. Thus, clinical geneticists search for “the” gene to be held responsible for “this” specific disorder. To identify this specific gene, genome-wide screening for CNVs or whole-exome screening (WES), and even whole-genome sequencing are used. In addition, novel disease loci are pinpointed by genome-wide association studies (GWAS) of cohorts of patients. While family-based studies of individuals or a few patients uncovered rare, often unique, CNVs and SNVs, GWAS rely on common variants among the SNPs studied. To aid in clinical decision-making, the hundreds of genes identified in studies of cohorts of patients with autism spectrum disorders (ASDs) were classified according to chromosomal location, the way by which they were discovered, and a score for their strength of association with ASD (https://gene.sfari.org/database/human-gene/).

With these efforts, 2 unexpected “by-products” arose. Using SNP arrays, designed to be heterozygous for all SNPs, to search for CNVs in the genome of a patient, incidentally regions of homozygous alleles were encountered. Such regions of homozygous SNPs reflect a close genetic relationship between the parents. These emerge in small populations on islands or other geographically or culturally isolated regions. In such populations, consanguineous marriages occur more frequently, which result in a higher burden of single-gene recessive and complex diseases [Bittles and Black, 2010]. ASDs are a quintessential form of such inherited complex neurodevelopmental disorders [Lichtenstein et al., 2010; Lyall et al., 2017]. Numerous studies of ASD patients have identified mutations affecting genes involved in regulating transcription, mRNA translation, protein synthesis and degradation as well as synaptic function, which apparently converge on cellular pathways affecting neuronal and synaptic homeostasis [Huguet and Bourgeron, 2013; Pinto et al., 2014; Bourgeron, 2015]. Yet, it is often difficult to pinpoint a single locus or gene that may be responsible for ASD in a single or several patients within a family [Salyakina et al., 2011; van Daalen et al., 2011]. In examining the pedigrees of families with ASD patients, one is often left with the impression that in addition to a de novo CNV, some undiscovered “susceptibility factor” may have been “missed.” Concurrent searches for SNVs and CNVs in genome-sequencing data of ASD patients showed that an appreciable number of them carried 3 or even more de novo variants [Turner et al., 2017]. However, most pedigrees are too small to allow for rigorous linkage analysis. By combining WES with homozygosity mapping in families selected for likely recessive inheritance of ASD, the number of genomic segments were narrowed down to those that cosegregated with disease, which allowed to pinpoint biallelic mutations in plausible candidate genes [Yu et al., 2013].

In a family with 2 boys with ASD and facial and skeletal anomalies, 3 phenotypically relevant loci have been
identified by WES combined with segregation analysis and homozygosity mapping [Valilou et al., this issue]. In this way, in both patients cosegregating, homozygous variants in 3 genes in regions of homozygosity were uncovered. First, FBN1, the gene encoding the extracellular microfibril protein fibrillin. Variants in FBN1 have been found in patients with Marfan syndrome, familial ectopia lentis, acromicric dysplasia, and Weill-Marchesani syndrome 2. The 2 patients presented with marfanoid features such as arachnodactyly and micromastria. In ASD patients, de novo missense variants of FBN1 were previously reported [Iossifov et al., 2012, 2014; De Rubeis et al., 2014; Krumm et al., 2015].

Second, both patients carried homozygous variants in the PLOD2 gene that encodes the procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 lysyl hydroxylase 2, which hydroxylates the lysine residues at the collagen type 1 carboxy-terminal telopeptide, allowing intermolecular cross-links between collagen fibrils [Uzawa et al., 1999]. Pathogenic variants in PLOD2 cause an autosomal recessive disease reminiscent of the osteogenesis imperfecta such as Bruck syndrome type 2. The 2 patients did not present with the full clinical features of Bruck syndrome, yet the variants in PLOD2 may contribute to their bone fragility.

Third, both patients had homozygous variants in the transferrin (TF) gene. The latter has not yet been implicated in ASD, but is pivotal to cellular antioxidant defences, which if impaired, may lead to lipid peroxidation and other pathological effects. This study illustrates the utility of combining WES with segregation analysis for patients, who do not carry diagnostic features fitting to a single mendelian disorder, but in contrast, may possibly carry a “double disorder” [Yu et al., 2013; Posey et al., 2017]. Moreover, this study points toward a novel mechanism of double, or even triple, disorders in addition to the well-known “unmasking” of a recessive allele for one disorder by a deletion encompassing, among others, a gene for a distinct disorder on the other chromosome [Poot and Haaf, 2015]. The mutation in the remaining single allele fits to an autosomal recessive disorder, while the de novo deletion follows a dominant pattern. This constellation blurs the distinction between recessive and dominant disorders.

This case report also points to polygenic mechanisms of disease, a concept that has emerged as a secondary outcome of GWAS studies, and assumes that multiple loci and alleles are associated with a significant risk for a disorder. Thus, a direct assessment of an individual’s genetic predisposition for disease risk can be expressed as the cumulative effect of common alleles identified in GWAS via polygenic risk scores. Together with the involved loci, also the risk alleles for specific neurodevelopmental disorders have been identified [Weiner et al., 2017]. This more systematic approach goes beyond anecdotal pinpointing of a risk allele after SNP genotyping of an individual patient [Poot et al., 2010].

Recent studies have allowed to better estimate the impact of polygenic risks and revealed features that may become of clinical utility in the near future [Chatterjee et al., 2016; Weiner et al., 2017; Torkamani et al., 2018; Grove et al., 2019]. First, polygenic variation appears to contribute additively to risk in ASD cases that carry a strong acting de novo variant [Weiner et al., 2017]. Second, elements of polygenic risk are independent of each other and differ in their relationship with the ASD phenotype [Weiner et al., 2017; McKenna et al., 2018]. Third, genetic influences on ASD are additive and may create risk through partially distinct etiologic pathways [Weiner et al., 2017; McKenna et al., 2018]. Fourth, in 2 independent cohorts of ASD patients, polygenic risk for ASD positively correlated with general cognitive ability and academic achievement [Clarke et al., 2016; Grove et al., 2019]. Thus, ASD patients can be divided in subtypes according to their polygenic risk profile. Although analyses of polygenic risk are novel and ongoing, they carry a great promise to resolve the genomic risk profile of individual ASD patients and their families. Based on such risk profiles, it will become possible to estimate the recurrence risk for ASD in very young children, even babies [D’Abate et al., 2019]. Eventually, genome-wide screening for polygenic risk alleles will complement the classical mendelian framework of genetics.

**Martin Poot**

---

**References**


