RESEARCH HIGHLIGHT

Molecular basis for phenotypic similarity of genetic disorders

Vijay Kumar Pounraja[1] and Santhosh Girirajan[1,2,3*]

Abstract

The contribution of distinct genes to overlapping phenotypes suggests that such genes share ancestral origins, membership of disease pathways, or molecular functions. A recent study by Liu and colleagues identified mutations in *TCF20*, a paralog of *RAI1*, among individuals manifesting a novel syndrome that has phenotypes similar to those of Smith-Magenis syndrome (a disorder caused by disruption of *RAI1*). This study highlights how structural similarity among genes contributes to shared phenotypes, and shows how this relationship can contribute to our understanding of the genetic basis of complex disorders.

Paradigms for establishing shared genetic etiology

Establishing the association between genotype and phenotype is the central element for most genetic analyses of complex disorders. Several of these disorders are characterized by genetic heterogeneity, where disruption of a variety of distinct genes can cause similar phenotypes. This heterogeneity can occur as the result of several common evolutionary and functional properties, including shared ancestral origins, protein sequence similarity, overlapping molecular functions, or membership of the same pathways. For example, disruption of *PKD1* and *PKD2*, two polycystin protein genes that share four transmembrane protein domains and interact with each other, can independently lead to polycystic kidney disease [1]. Another classic example is the presence of a common constellation of clinical features in Bardet-Biedl syndrome. These features include rod-cone

dystrophy, obesity, hypogonadism, and renal anomalies, among individuals with recessive mutations in any of the cilia-formation genes, such as *BBS1* and *BBS2* [2]. Similarly, *EHMT1* and *MBD5* cause specific intellectual disability-associated disorders and interact within a chromatin-modification network module, highlighting how epigenetic defects may underlie cognitive deficits [3]. Further, mutations in gap junction proteins *GJB2* and *GJB6*, which interact to form heteromeric complexes, both result in deafness [4]. Therefore, investigating functional and molecular similarities among genes that contribute to related phenotypes can provide a broader framework to help establish the etiologies of complex disorders.

Functional relatedness translating to shared phenotypes

Genomic studies on large cohorts of affected individuals have identified hundreds of genes that may be involved in the etiology of developmental disorders. Several of these studies have unraveled specific genes that are associated with rare Mendelian disorders, whereas others have identified numerous genes that contribute to similar disorders with shared phenotypes. In a recent study, Liu and colleagues analyzed exome sequencing and chromosomal microarray data and identified pathogenic mutations in *TCF20* in 32 affected individuals from 31 unrelated families [5]. *TCF20* encodes an SPRE-binding transcription factor that is strongly expressed in pre-migratory neural crest cells and is known to influence other transcription factors [6]. Although *TCF20* has been previously associated with autism, intellectual disability, and related phenotypes, the authors of this study performed a deeper assessment of phenotypes and identified a pattern of features that were reminiscent of Smith-Magenis syndrome (SMS), a rare disorder caused by disruption of *RAI1* (encoding the retinoic acid induced 1 protein). Like children with SMS, patients with *TCF20* mutations presented a core set of features
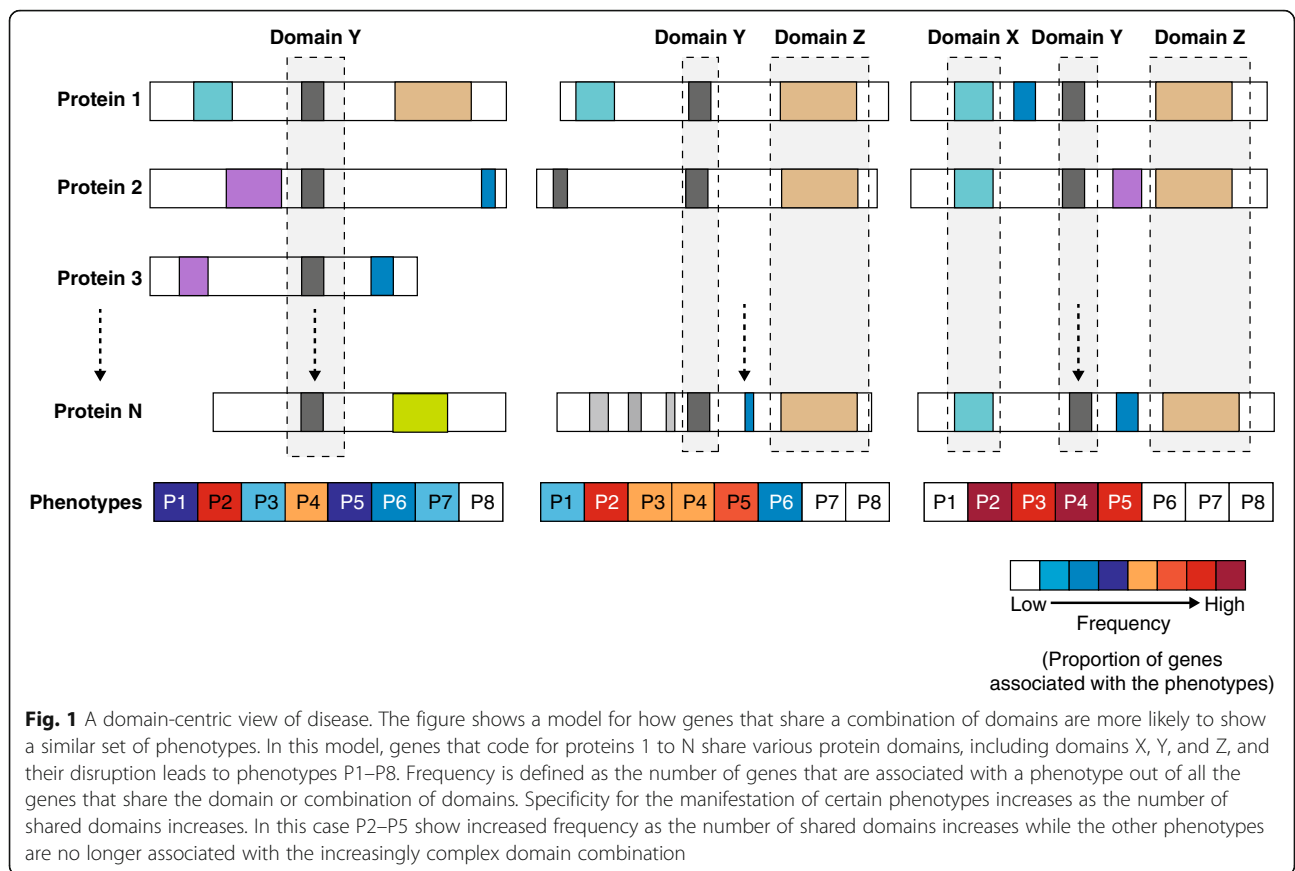
* Correspondence: sxg47@psu.edu
[1]Bioinformatics and Genomics Program, The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802, USA
[2]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA
Full list of author information is available at the end of the article

Fig. 1 A domain-centric view of disease. The figure shows a model for how genes that share a combination of domains are more likely to show a similar set of phenotypes. In this model, genes that code for proteins 1 to N share various protein domains, including domains X, Y, and Z, and their disruption leads to phenotypes P1–P8. Frequency is defined as the number of genes that are associated with a phenotype out of all the genes that share the domain or combination of domains. Specificity for the manifestation of certain phenotypes increases as the number of shared domains increases. In this case P2–P5 show increased frequency as the number of shared domains increases while the other phenotypes are no longer associated with the increasingly complex domain combination

including facial dysmorphology, hypotonia, seizures, and sleep disturbance.

Liu and colleagues found that commonalities in gene structure and function between *TCF20* and *RAI1* could explain the shared core clinical features and molecular effects [5]. In fact, *TCF20* shares several essential protein domains with *RAI1*, including N-terminal transactivation domains, zinc-finger-like plant homeodomains (PHD), and nuclear localization signal domains [6]. The high sequence homology and conservation of specific domain combinations between *TCF20* and *RAI1* is attributed to a gene duplication event that occurred during early vertebrate evolution [5]. For example, the chromatin-binding PHD domain is highly conserved in both *TCF20* and *RAI1*, and a patient with a missense mutation in the PHD domain of *TCF20* presented with strong SMS-like features [5]. In fact, several PHD domain-containing genes are involved in chromatin modification and transcriptional regulation functions, and are therefore relevant not only to SMS and *TCF20*-associated disorders but also to several other disorders, including *NSD1* and Sotos syndrome, *CREBBP* and Rubinstein-Taybi syndrome, *DPF2* and Coffin-Siris syndrome, and *KMT2D* and Kabuki syndrome [7].

This study adds *TCF20* to a growing list of genes that cause SMS-like phenotypes and to the list of disorders

that should be considered in differential diagnosis. In fact, previous studies on individuals who presented with features typical of SMS but did not carry *RAI1* mutations found that these individuals had mutations in: *MBD5*, which were associated with a set of SMS-like neurodevelopmental features and with autism; *EHMT1*, the causative gene for Kleefstra syndrome; *PHF21A*, which is associated with Potocki-Shaffer syndrome; or *TCF4*, which is associated with Pitt-Hopkins syndrome. Furthermore, Loviglio and colleagues [8] identified *POGZ*, *BRD2*, *KDM5C*, and *ZBTB17* within an *RAI1*-associated network, and Berger and colleagues [9] identified *DEAF1* and *IQSEC2*, whose disruption resulted in phenotypes overlapping with those associated with SMS. As the network of genes associated with SMS-like phenotypes grows, it is likely that some of these genes traverse pathways related to common disorders including autism, cognitive defects, and sleep abnormalities.

A protein domain-centric view of disease

With the emergence of the deep phenotyping and genome sequencing of large cohorts of individuals as part of clinical care and precision-medicine initiatives, future studies will expand upon the approach outlined by Liu and colleagues to identify novel instances of phenotypic

convergence among individuals carrying mutations in functionally related genes [5]. One approach would be to interrogate whether genes that share common protein domains confer risks for similar phenotypes (Fig. 1). For example, the PDZ-domain-containing *SHANK* and *NLGN* family of genes are involved in synaptic signaling and are associated with autism [10]. Nevertheless, the presence of a single domain in a gene may not always be predictive of a specific phenotype, because the ultimate biological effects of that gene could also depend on the presence of other functional domains. As observed for *RAI1* and *TCF20*, genes that share combinations of domains could confer greater specificity for a particular set of phenotypes. This could potentially explain why other genes that both encode proteins that contain PHD domains, such as *NSD1* and *KMT2B*, and contribute to neurodevelopmental disorders do not share a full set of phenotypic associations with SMS [7]. Further studies could also search for an over- or under-representation of genes with a specific composition of conserved protein domains in one or more phenotypic categories. Liu and colleagues carve an exciting paradigm for identifying the functional relatedness of genes on the basis of shared phenotypes, which could potentially help to refine common networks and pathways for neurodevelopmental disorders and other complex genetic diseases [5].

Conclusions

Studies on larger populations of affected individuals will continue to identify associations between diseases and genes that are associated with distinct categories of biological pathways, genetic networks, and molecular mechanisms. The discovery of disease-associated genes on the basis of shared domains and evolutionary history, as described by Liu and colleagues [5], could be used to further refine connections between genes that contribute to related disorders and to provide mechanistic specificities for genes within these broader functional categories.

Abbreviations
PHD: Plant homeodomain; SMS: Smith-Magenis syndrome

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Bioinformatics and Genomics Program, The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802, USA. [2]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA. [3]Department of Anthropology, Pennsylvania State University, University Park, PA 16802, USA.

Published online: 23 April 2019

References
1. Tsiokas L, Kim E, Arnould T, Sukhatme VP, Walz G, Germino G, et al. Homo- and heterodimeric interactions between the gene products of PKD1 and PKD2. Proc Natl Acad Sci U S A. 1997;94:6965–70.
2. Badano JL, Mitsuma N, Beales PL, Katsanis N. The ciliopathies: an emerging class of human genetic disorders. Annu Rev Genomics Hum Genet. 2006;7:125–48.
3. Kleefstra T, Kramer JM, Neveling K, Willemsen MH, Koemans TS, Vissers LELM, et al. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. Am J Hum Genet. 2012;91:73–82.
4. Nickel R, Forge A. Gap junctions and connexins in the inner ear: their roles in homeostasis and deafness. Curr Opin Otolaryngol Head Neck Surg. 2008;16:452–7.
5. Vetrini F, McKee S, Rosenfeld JA, Suri M, Lewis AM, Nugent KM, et al. De novo and inherited TCF20 pathogenic variants are associated with intellectual disability, dysmorphic features, hypotonia, and neurological impairments with similarities to smith–Magenis syndrome. Genome Med. 2019;11:12.
6. Schäfgen J, Cremer K, Becker J, Wieland T, Zink AM, Kim S, et al. De novo nonsense and frameshift variants of TCF20 in individuals with intellectual disability and postnatal overgrowth. Eur J Hum Genet. 2016;24:1739–45.
7. Vasileiou G, Vergarajauregui S, Endele S, Popp B, Büttner C, Ekici AB, et al. Mutations in the BAF-complex subunit DPF2 are associated with coffin-Siris syndrome. Am J Hum Genet. 2018;102:468–79.
8. Loviglio MN, Beck CR, White JJ, Leleu M, Harel T, Guex N, et al. Identification of a RAI1-associated disease network through integration of exome sequencing, transcriptomics, and 3D genomics. Genome Med. 2016;8:105.
9. Berger SI, Ciccone C, Simon KL, Malicdan MC, Vilboux T, Billington C, et al. Exome analysis of smith-Magenis-like syndrome cohort identifies de novo likely pathogenic variants. Hum Genet. 2017;136:409–20.
10. Südhof TC. Neuroligins and neurexins link synaptic function to cognitive disease. Nature. 2008;455:903–11.