

GWAS signals revisited using human knockouts

Sateesh Maddirevula, PhD¹, Fatema AlZahrani, BSc¹, Shams Anazi, MSc¹, Mariam Almureikhi, MD², Tawfeg Ben-Omran, MD², Ghada M.H. Abdel-Salam, MD, PhD³, Mais Hashem, BSc¹, Niema Ibrahim, BSc¹, Firdous M. Abdulwahab, BSc¹, Neama Meriki, MD⁴, Fahad A. Bashiri, MD⁵, Meow-Keong Thong, MD⁶, Premala Muthukumarasamy, MBBS⁶, Rifhan Azwani Mazlan, MBBS⁶, Ranad Shaheen, PhD¹ and Fowzan S. Alkuraya, MD^{1,7,8}

Purpose: Genome-wide association studies (GWAS) have been instrumental to our understanding of the genetic risk determinants of complex traits. A common challenge in GWAS is the interpretation of signals, which are usually attributed to the genes closest to the polymorphic markers that display the strongest statistical association. Naturally occurring complete loss of function (knockout) of these genes in humans can inform GWAS interpretation by unmasking their deficiency state in a clinical context.

Methods: We exploited the unique population structure of Saudi Arabia to identify novel knockout events in genes previously highlighted in GWAS using combined autozygome/exome analysis.

Results: We report five families with homozygous truncating mutations in genes that had only been linked to human disease through GWAS. The phenotypes observed in the natural knockouts for these genes (*TRAF3IP2*, *FRMD3*, *RSRC1*, *BTBD9*, and *PXDNL*) range from consistent with, to unrelated to, the previously reported GWAS phenotype.

Conclusion: We expand the role of human knockouts in the medical annotation of the human genome, and show their potential value in informing the interpretation of GWAS of complex traits.

Genet Med advance online publication 22 June 2017

Key Words: autozygome; consanguinity; GWAS; loss of function

INTRODUCTION

Genome-wide association studies (GWAS) have emerged as a cornerstone in the delivery of the Human Genome Project's promise to revolutionize medicine.¹ Far from being caused by individual mutations, complex traits that account for the overwhelming majority of illnesses have proven very difficult to study using classic positional mapping or hypothesis-driven case-control studies. Under the "common variant-common disease" model, a new approach was needed to address the polygenicity of these traits by taking into account the additive/synergistic interaction of multiple common variants and the heterogeneity of these combinatorial effects across individuals.² To date, 1,751 studies have been published that collectively examined millions of polymorphic markers in hundreds of thousands of patients and controls to extract signals of association according to the National Human Genome Research Institute GWAS Catalog.³ The identification of 11,912 such signals may not have yet penetrated clinical practice for the purpose of prediction and prevention of common diseases, but it has greatly informed our understanding of their pathophysiology, and we are starting to see their potential to inform novel drug development.⁴⁻⁷

The power of GWAS stems from their hypothesis-free design, which makes it possible to uncover unexpected genetic, and in turn, biological mechanisms. The fact that the medical relevance, historically inferred from Mendelian mutations, is unknown for most of the ~20,000 protein-coding human genes means that a GWAS signal is often detected in a gene with no established medical context. This challenge is further compounded by the finding that the majority of variants are not within the 1.5% coding part of the human genome.⁸ As a result, alleles that produce significant signals are often interpreted in light of the physically closest genes even though they may not be the source of the biological signal.⁹ Extensive follow-up functional analysis of the genes that are the putative source of the signal is often required especially when little had been published about the function of these genes.¹⁰

Mendelian mutations have long been recognized as the best source of information on gene function in humans.^{11,12} Although gain of function and loss of function can both be informative in this regard, loss of function mutations are often more faithful in exposing the physiological context of genes. Biallelic truncating mutations are the most conspicuous forms

¹Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; ²Section of Clinical and Metabolic Genetics, Department of Pediatrics, Hamad Medical Corporation, Doha, Qatar; ³Clinical Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt; ⁴Department of Obstetrics and Gynecology, College of Medicine, King Saud University, Riyadh, Saudi Arabia; ⁵Department of Pediatrics, College of Medicine & King Khalid University Hospital, King Saud University, Riyadh, Saudi Arabia; ⁶Genetics and Metabolism Unit, Department of Paediatrics, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia; ⁷Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia; ⁸Saudi Human Genome Program, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia. Correspondence: Fowzan S. Alkuraya (falkuraya@kfshrc.edu.sa)

Submitted 23 November 2016; accepted 12 April 2017; advance online publication 22 June 2017. doi:10.1038/gim.2017.78

of loss of function, especially those that occur early in the reading frame and across all known isoforms, essentially representing knockouts of these genes.¹³ We have previously shown that enhanced homozygosity as a function of autozygosity, a signature genomic feature of inbreeding, can facilitate the occurrence of naturally occurring human knockouts.¹⁴ These knockouts span the entire phenotypic spectrum from the apparently healthy to the very early embryonic lethal.¹⁵

In this paper, we expand our work on human knockouts by reporting their relevance to the study of complex traits. Specifically, we report five genes that had not been reported in a Mendelian context but had been suggested to influence complex trait genetic risk through GWAS signals. By comparing the phenotypes of patients who are knocked out for these genes to the purported complex trait association, we show the potential value of this approach in improving the specificity of GWAS interpretation.

MATERIALS AND METHODS

Human subjects

Human subjects are individuals who presented with phenotypes that matched existing institutional review board-approved research protocols (KFSHRC RAC 2070023, 2080006, 2121053). After informed consent was obtained, blood was collected from index and available family members as appropriate for DNA extraction and subsequent analysis. A separate consent to publish identifiable photographs was also obtained.

Autozygome analysis

Mapping of all autozygous segments per genome (autozygome) was as described before. Briefly, regions of homozygosity > 2 Mb were used as surrogates of autozygosity as determined by AutoSNPa. The genotyping platform was Axiom single-nucleotide polymorphism (SNP) chip from Affymetrix (Santa Clara, CA, USA).

Exome analysis and variant interpretation

Exome capture was performed using TruSeq Exome Enrichment kit (Illumina, San Diego, CA, USA) following the manufacturer's protocol. Samples were prepared as an Illumina sequencing library, and in the second step, the sequencing libraries were enriched for the desired target using the Illumina Exome Enrichment protocol. The captured libraries were sequenced using Illumina HiSeq 2000 Sequencer. The reads were mapped against UCSC hg19 (<http://genome.ucsc.edu/>) by BWA (<http://bio-bwa.sourceforge.net/>). The SNPs and indels were detected by SAMTOOLS (<http://samtools.sourceforge.net/>). For this study, we included homozygous truncating variants only if they met the following criteria: (i) within the autozygome, (ii) novel or rare (< 0.001) in ExAC and 2,379 Saudi exomes, (iii) in a gene with no established Mendelian phenotype, and (iv) in a gene with suggested association to a complex trait in a previously peer-reviewed published GWAS.

RESULTS

We report in this study the identification of five knockout events that met our eligibility criteria and are described below (Table 1, Figure 1 and Supplementary Figure S1 online). Details of the clinical phenotype are provided in the accompanying supplemental Supplementary Note S1 (the file called "GWAS paper clinical Notes:v5").

BTBD9

Several GWAS studies have implicated this poorly characterized gene in the pathogenesis of restless leg syndrome and periodic limb movement of sleep although the mechanism remains unclear.¹⁶ In a consanguineous family from Yemen with multiple deaths due to severe unexplained myopathy (normal creatine kinase) and normal brain magnetic resonance image, we identified a homozygous truncating mutation NM_001172418.1:c.1015C>T:p.(Arg339*) in the two affected siblings we had access to.

TRAF3IP2

This gene encodes an interactor with *IL17RA* and this interaction is thought to mediate the suggested role of *TRAF3IP2* in psoriasis susceptibility, which was reported by several GWAS.^{17,18} We have identified three siblings with severe eczema and elevated IgE who are all homozygous for the truncating mutation NM_147686.3:c.488_492delTACCT:p.(Leu163*).

RSRC1

The encoded protein is thought to play a role in RNA splicing and has been implicated in height and subjective wellbeing based on GWAS approach.^{19,20} We show that in a family of three siblings with nonsyndromic intellectual disability and normal brain magnetic resonance image, a homozygous truncating mutation NM_016625.3:c.268C>T:p.(Arg90*) fully segregated with the phenotype.

FRMD3

The encoded protein belongs to the 4.1 protein family and is highly enriched in the kidney. Several GWAS have implicated *FRMD3* in the genetic risk of diabetic kidney disease.^{21–23} However, only its association with Lewy body disease is listed in GWAS Catalog based on the study by Beecham *et al.*²⁴ Remarkably, we have identified a homozygous *FRMD3* truncation NM_001244962.1:c.70C>T:p.(Arg24*) in a female patient who died shortly after birth with Potter sequence and severely echogenic kidneys with no identifiable cysts.

PXDNL

This gene encodes a protein thought to play a role in cytoskeleton remodeling and cell motility.²⁴ An intragenic SNP was reported among the top association signals for neuritic plaques, which in turn correlated positively with Alzheimer disease risk.²⁴ We have observed a homozygous splicing mutation NM_144651.4:c.695-2 A>T in a patient

Table 1 Summary of clinical findings and mutation information in the study cohort

Gene	GWAS phenotype	GWAS SNP (P value)	GWAS reference	Index ID	Mutation	MAF ExAC	MAF internal	KO phenotype
TRAF3IP2	Psoriasis vulgaris	rs33980500 (1x10 ⁻²³)	PMID:26626624	16DG0333	NM_147686.3:	0.0	0.0	Mendelian severe form of eczema, high IgE and macrocephaly
		rs33980500 (1x10 ⁻²⁰)	PMID:20953186		c.488_492delTACCT:p.(Leu163*)			
Psoriatic arthritis		rs33980500 (1x10 ⁻¹⁷)	PMID:26626624					
	Psoriasis	rs33980500 (1x10 ⁻¹⁶)	PMID:20953188					
Ulcerative colitis		rs458017 (2x10 ⁻¹⁶)	PMID:20953190					
	Crohn disease	rs3851228 (1 x10 ⁻¹⁴)	PMID:26192919					
Schizophrenia		rs3851228 (3 x10 ⁻⁸)	PMID:26192919					
	Lewy body disease	rs6568685 (7x10 ⁻⁷)	PMID:26198764					
FRMD3		rs1570582 (2x10 ⁻⁶)	PMID: 25188341	15DG2538	NM_001244962.1:	0.00004160	0.0	Echogenic nonfunctioning kidneys, anhydramnios, early neonatal death
		rs1570582 (9x10 ⁻⁶)			c.70C > T;p.(Arg24*)	(6 HET)		
RSRC1	Height	rs2362965 (2x10 ⁻⁹)	PMID: 23563607	16DG1196	NM_016625.3:	0.0	0.0	Intellectual disability
					c.268C > T;p.(Arg90*)			
BTBD9	Subjective well-being	rs6787172 (4x10 ⁻⁶)	PMID: 27089181					
		rs1714507 (9x10 ⁻⁶)						
Lethal myopathy	Restless legs syndrome	rs9357271 (8x10 ⁻²²)	PMID:21779176	14DG0470	NM_001172418.1:	0.00005141	0.0004	Lethal myopathy
		rs9296249 (4x10 ⁻¹⁸)	PMID:21779176		c.1015C > T;p.(Arg339*)	(8HET)	(1 HET)	
Postbronchodilator FEV ₁ /FVC ratio in COPD		rs3923809 (1x10 ⁻¹⁷)	PMID:17634447					
		rs114407411 (2x10 ⁻⁷)	PMID:26634245					
Mammographic density		rs192452148 (2x10 ⁻⁷)						
		rs150713181 (4x10 ⁻⁷)						
Mild influenza (H1N1) infection		rs77928651 (1x10 ⁻⁶)						
		rs139632137 (2x10 ⁻⁶)						
Neuritic plaque		rs4565302 (5x10 ⁻⁶)	PMID:25353672					
		rs201466262 (2x10 ⁻⁸)	PMID: 26379185	15DG2011	NM_144651.4:	0.00000819	0.0	Intellectual disability, nystagmus, subtle facial dysmorphism, and interhemispheric cyst
Tuberculosis					c.695-2A > T	(2 HET)		
		rs62506864 (5x10 ⁻⁶)	PMID: 25188341					
		rs7821565 (8x10 ⁻⁶)	PMID: 20694014					

ExAC, Exome Aggregation Consortium; COPD, chronic obstructive pulmonary disease; GWAS, genome-wide association studies; KO, knockout; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.

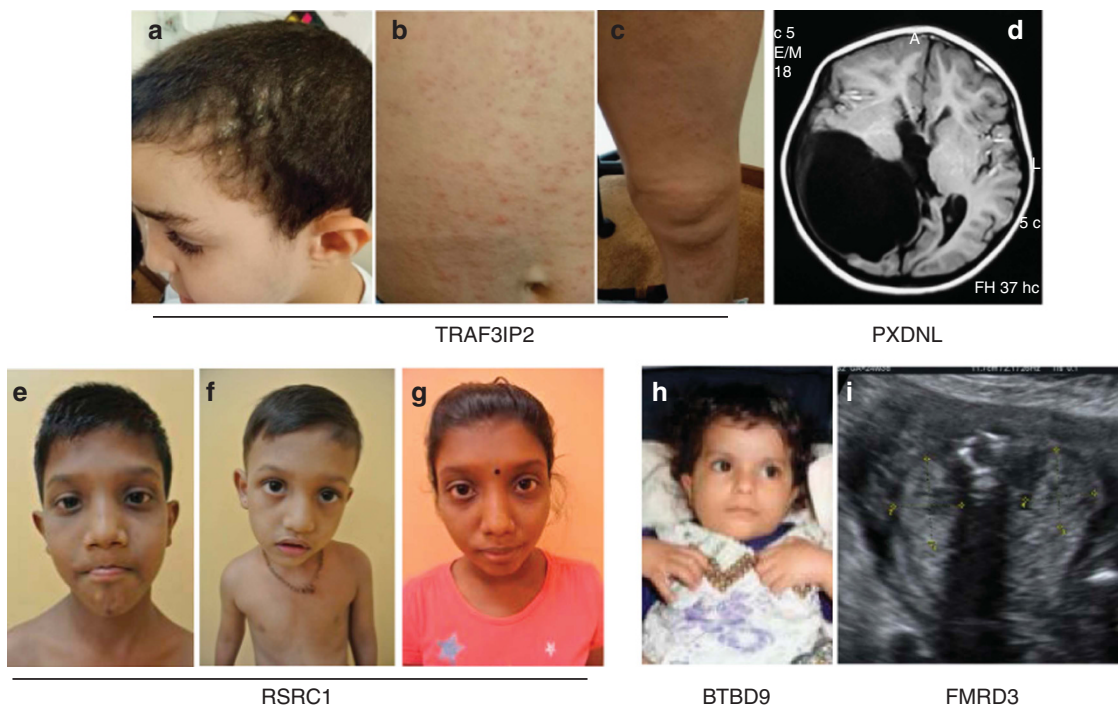


Figure 1 Clinical features of human knockouts studied in this cohort. (a–c) Clinical images of the index with *TRAF3IP2* biallelic truncation showing extensive skin disease including the scalp. (d) Magnetic resonance image of the index with *PXDNL* biallelic splicing variant showing interhemispheric cyst. (e–g) Lack of obvious facial dysmorphism in the three siblings with *RSRC1* biallelic truncation. (h) Hypotonic posture of the index with biallelic truncation of *BTBD9*. (i) Antenatal ultrasonography of the index with biallelic truncation of *FRMD3* showing anhydramnios and echogenic kidneys of normal size.

with intellectual disability, partial agenesis of corpus callosum, and hypoplastic cerebellum and brainstem.

DISCUSSION

The contribution of Mendelian genes to the genetic risk of complex traits has long intrigued the human genetics community.²⁵ While early case–control studies that were designed to investigate the contribution of specific Mendelian genes did not generate rigorous associations, subsequent GWAS have indeed confirmed numerous associations between common and rare variants in Mendelian genes and the complex counterpart.²⁵ Genes with Mendelian links can serve as landmarks to guide GWAS investigators as they attempt to interpret the likely source of biological signal when the genetic signal is not in a gene. For example, the study by Harley *et al.* on the genetics of systemic lupus erythematosus identified a nongenic SNP, and the authors attributed the signal to the closest gene, *PXK*, even though the gene has no obvious connection to autoimmunity.²⁶ However, our subsequent finding that human knockouts for *DNASE1L3*, another gene in the same vicinity of the SNP as *PXK*, prompted a re-analysis that confirmed that *DNASE1L3* was indeed the likely source of the signal.^{27,28} On the other hand, the finding of Mendelian forms in the same gene implicated in GWAS provides additional and independent support that the signal identified by the latter is likely causal, as demonstrated by our finding that the GWAS-implicated *LACCI* in inflammatory bowel

disease can cause a Mendelian form of the disorder, and the GWAS-implicated *ARL6IP6* in stroke can cause a Mendelian disease with severe cerebrovascular disease.^{29,30}

Human knockouts have historically contributed to the mapping of novel disease genes. However, the recent appreciation of the widespread occurrence of knockouts has prompted reconsideration of their phenotypic consequences especially since some are clearly not associated with any conspicuous phenotype.¹³ Therefore, we caution against the conclusion that the phenotypes we report in this study are necessarily caused by the knockout events before additional patients are described. However, they can still be helpful in interpreting the GWAS signals that had been attributed to these genes. We note that in certain instances, the knockout phenotype likely represents the severe Mendelian counterpart of the complex trait with reported association. This is most evident in *TRAF3IP2* (the severe skin phenotype versus psoriasis susceptibility) although we also note significant overlap in the case of *FRMD3* (severely echogenic kidneys versus diabetic nephropathy susceptibility), and *BTBD9* (severe myopathy versus restless leg syndrome susceptibility). In contrast, the observed knockout phenotypes for *PXDNL* and *RSC1* appear to be unrelated to the phenotypes with which variants in these genes have been associated in GWAS. It is worth highlighting that the essentially loss-of-function consequence of human knockout events leaves open the possibility that the observed discrepancy with GWAS is due to

potential gene upregulation caused by the reported variants in those studies.

In conclusion, we suggest that naturally occurring knockout events in humans can aid in the interpretation of GWAS signals depending on the resulting phenotype. An expanded catalog of these events and their associated phenotypes is under way to inform research into Mendelian and complex traits.

ACKNOWLEDGMENTS

We thank the Genotyping and Sequencing Core Facilities at King Faisal Specialist Hospital & Research Centre for their technical help. This work was supported in part by King Salman Center for Disability Research grant (FSA) and King Abdulaziz City for Science and Technology grant 15-BIO3688-20 (FSA). We also thank the study families for their enthusiastic participation.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

- Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* 2012;90:7–24.
- Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461:747–753.
- Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014;42(D1):D1001–D1006.
- Kamb A, Harper S, Stefansson K. Human genetics as a foundation for innovative drug development. *Nat Biotechnol* 2013;31:975–978.
- Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov* 2013;12:581–594.
- Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015;372:793–795.
- Vrieze SI, McGue M, Iacono WG. The interplay of genes and adolescent development in substance use disorders: leveraging findings from GWAS meta-analyses to test developmental hypotheses about nicotine consumption. *Hum Genet* 2012;131:791–801.
- Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 2009;106:9362–9367.
- Hou L, Zhao H. A review of post-GWAS prioritization approaches. *Front Genet* 2013;4:280.
- Ritchie GR, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. *Nat Methods* 2014;11:294–296.
- Alkuraya FS. Discovery of mutations for Mendelian disorders. *Hum Genet* 2016;135:615–623.
- Burke W. Genomics as a probe for disease biology. *N Engl J Med* 2003;349:969–974.
- Alkuraya FS. Human knockout research: new horizons and opportunities. *Trends Genet* 2015;31:108–115.
- Alsalem AB, Halees AS, Anazi S, Alshamekh S, Alkuraya FS. Autozygome sequencing expands the horizon of human knockout research and provides novel insights into human phenotypic variation. *PLoS Genet* 2013;9:e1004030.
- Shamseldin HE, Tulbah M, Kurdi W, et al. Identification of embryonic lethal genes in humans by autozygosity mapping and exome sequencing in consanguineous families. *Genome Biol* 2015;16:116.
- Winkelmann J, Schormair B, Lichtner P, et al. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nat Genet* 2007;39:1000–1006.
- Hüffmeier U, Uebe S, Ekici AB, et al. Common variants at TRAF3IP2 are associated with susceptibility to psoriatic arthritis and psoriasis. *Nat Genet* 2010;42:996–999.
- Ellinghaus E, Ellinghaus D, Stuart PE, et al. Genome-wide association study identifies a psoriasis susceptibility locus at TRAF3IP2. *Nat Genet* 2010;42:991–995.
- Berndt SI, Gustafsson S, Mägi R, et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* 2013;45:501–512.
- Okbay A, Baselmans BM, De Neve J-E, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* 2016;48:624–33.
- Pezzolesi MG, Poznik GD, Mychaleckyj JC, et al. Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes. *Diabetes* 2009;58:1403–1410.
- Mooyaart A, Valk E, Van Es L, et al. Genetic associations in diabetic nephropathy: a meta-analysis. *Diabetologia* 2011;54:544–553.
- Martini S, Nair V, Patel SR, et al. From single nucleotide polymorphism to transcriptional mechanism: a model for FRMD3 in diabetic nephropathy. *Diabetes* 2013;62:2605–2612.
- Beecham GW, Hamilton K, Naj AC, et al. Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet* 2014;10:e1004606.
- Jin W, Qin P, Lou H, Jin L, Xu S. A systematic characterization of genes underlying both complex and Mendelian diseases. *Hum Mol Genet* 2011;21:1611–1624.
- Harley JB, Alarcón-Riquelme ME, Criswell LA, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PTK, KIAA1542 and other loci. *Nat Genet* 2008;40:204–210.
- Al-Mayouf SM, Sunker A, Abdwani R, et al. Loss-of-function variant in DNASE1L3 causes a familial form of systemic lupus erythematosus. *Nat Genet* 2011;43:1186–1188.
- Mayes MD, Bossini-Castillo L, Gorlova O, et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. *Am J Hum Genet* 2014;94:47–61.
- Patel N, El Mouzan MI, Al-Mayouf SM, et al. Study of Mendelian forms of Crohn's disease in Saudi Arabia reveals novel risk loci and alleles. *Gut* 2014;63:1831–1832.
- Abumansour IS, Hijazi H, Alazmi A, et al. ARL6IP6, a susceptibility locus for ischemic stroke, is mutated in a patient with syndromic cutis marmorata telangiectatica congenita. *Hum Genet* 2015;134:815–822.