

SPOTLIGHT

Embracing human genetics: a primer for developmental biologists

Elizabeth J. Leslie*

ABSTRACT

Understanding the etiology of congenital disorders requires interdisciplinary research and close collaborations between clinicians, geneticists and developmental biologists. The pace of gene discovery has quickened due to advances in sequencing technology, resulting in a wealth of publicly available sequence data but also a gap between gene discovery and crucial mechanistic insights provided by studies in model systems. In this Spotlight, I highlight the opportunities for developmental biologists to engage with human geneticists and genetic resources to advance the study of congenital disorders.

KEY WORDS: Congenital disorders, Birth defects, Human genetics, Genetics databases

Introduction

As a human geneticist studying a congenital developmental disorder, I frequently receive emails from my developmental biologist colleagues asking if we have found mutations in their new favorite gene. I have a growing stack of post-it notes waiting for the day when we find a genetic variant in the DNA samples from our research participants that fill our freezers. The emails and post-it notes are evidence of collaborative partnerships between my human genetics lab and the mouse and zebrafish labs of my colleagues. Our collaboration works both ways; for every incoming query about human sequence variants there is an outgoing email from my lab asking about functional validation of sequence variants for one of our new genes.

Understanding disorders of development requires a close relationship between clinicians, geneticists, developmental biologists and others to understand the molecular mechanisms of disease pathogenesis. Ultimately, both geneticists and developmental biologists seek to understand the factors that cause cells to organize into a specific form, whether ordered or disordered. Our common goals are also mirrored in our scientific approaches. Forward genetic screens with spontaneous or mutagen-induced mutations are foundational approaches in developmental biology, while rapid reverse genetic screens are enabled by genetic engineering technologies such as CRISPR. And although most human geneticists would not describe their work as a ‘screen’ per se, advances in sequencing technology mean there are now enormous amounts of sequence and phenotype data that are often publicly available. Large genomic resources in cohorts with a particular disease have made forward human genetics more efficient, but there is also sufficient data to identify human ‘knockouts’ from large-scale genetic databases with linked health records, bringing reverse-genetics approaches to the study of human phenotypes (Alkuraya, 2015;

Narasimhan et al., 2016b). Although this is still an inefficient endeavor because of the low rate of homozygous mutations in randomly-mating populations, deep phenotyping of individual ‘knockouts’ identified by population-based sequencing of consanguineous or bottlenecked populations shows the promise of this approach as databases expand (Lim et al., 2014; Narasimhan et al., 2016a; Saleheen et al., 2017). Despite the inherent similarities in the two fields and the necessity for collaboration, we are too often siloed and the potential for model organism research to advance diagnosis, mechanism and treatment of these disorders goes unrealized. The goal of this Spotlight is to highlight the opportunities, tools and resources (summarized in Table 1) for developmental biologists to engage with human genetics databases, researchers and patients in the study of congenital disorders.

Congenital disorders are collectively common and one of the leading causes of childhood mortality (Wallingford, 2019). In recent years, there has been a concerted effort through clinical and research based sequencing programs to identify the genetic causes of these disorders. For example, the Gabriella Miller Kids First Pediatric Research Program (GMKF; <http://commonfund.nih.gov/kidsfirst>) is a pan-National Institutes of Health program to develop a large-scale publicly-available data resource to identify the genetic causes of structural birth defects and childhood cancer (and to understand the strong association between them; Botto et al., 2013; Carozza et al., 2012). Since its launch in 2015, GMKF has sequenced almost 20,000 DNA samples from affected individuals and their parents and generated 1.3 petabytes of whole-genome sequencing data on congenital heart disease, diaphragmatic hernia, orofacial clefts, craniosynostosis, disorders of sex development and many others. Because of programs such as GMKF and the Centers for Mendelian Genomics (<http://mendelian.org/>), discoveries of new genes associated with human disease phenotypes are being made at an average rate of 263–281 new discoveries per year (Boycott et al., 2017; Posey et al., 2019) and rapidly outpace functional studies. Consequently, little is generally known about the effects of these variants or the roles of implicated genes. The number of known genes without an associated human phenotype decreases with each discovery, but examples of phenotype expansion (multiple phenotypes associated with mutation in the same gene) and multiple modes of inheritance at a single locus are increasing (Posey et al., 2019), underscoring the complexity of human genetics. The gap between disease-gene associations and a mechanistic understanding is likely to widen: recent bibliometric analysis suggest that current research is heavily influenced by previous studies causing a ‘rich get richer’ loop that hinders translational research (Stoeger et al., 2018). The good news is that there is enormous potential and a crucial need for new research directions to bridge this gap.

The N=1 problem

Although some congenital disorders are quite common (i.e. congenital heart disease, orofacial clefts), others are rare or ultrarare in the population, presenting their own challenges for

Department of Human Genetics, Emory University, Atlanta, GA 30322, USA.

*Author for correspondence (ejlesli@emory.edu)

 E.J.L., 0000-0002-5735-1712

Table 1. Tools and resources for collaboration and variant interpretation

Tool	Description
Kids First kidsfirstdrc.org	The Gabriella Miller Kids First Pediatric Research program has created a data resource portal to bring together whole-genome sequencing from childhood cancer and structural birth defect research communities. The portal provides integrated data resources and cloud-based computation for researchers.
Centers for Mendelian Genomics Mendelian.org	The goal of the Centers for Mendelian Genomics is to use exome and genome sequencing and other genomic approaches to discover the genetic basis underlying as many Mendelian traits as possible, and accelerate discoveries by disseminating the obtained knowledge and outreach to investigators and rare disease consortia worldwide.
GeneMatcher genematcher.org	The GeneMatcher website is designed to enable connections between patients, clinicians and researchers who share an interest in the same gene or genes in order to solve 'unsolved' exomes/genomes.
DECIPHER decipher.sanger.ac.uk	DECIPHER is an interactive online database that reports participant phenotype and genetic variants and incorporates a variety of bioinformatic resources to aid the interpretation of genomic variants.
MyGene2 mygene2.org	The MyGene2 website is a family registry through which families with rare genetic conditions can share health and genetic information and connect with other families, clinicians and researchers. Researchers studying a gene or families with a rare condition can use MyGene2 to identify and contact other families who have the same rare condition or candidate gene.
ClinVar www.ncbi.nlm.nih.gov/clinvar	ClinVar is a freely accessible, public archive of reports of the relationships between human genetic variation and phenotypes, with supporting evidence according to American College of Medical Genetics and Genomics guidelines.
gnomAD gnomad.broadinstitute.org	gnomAD (formally ExAC) is an aggregation of large-scale whole-genome and exome sequencing data that can serve as a reference database for the interpretation of genetic variants. Individuals with known severe pediatric disease and their first-degree relatives were removed, so these datasets can serve as useful reference sets of allele frequencies for severe pediatric disease studies.

gene discovery. Because even seemingly healthy humans carry a remarkably large number of deleterious variants, usually in the heterozygous state (Lek et al., 2016), confidently defining causal variants requires multiple unrelated individuals with a similar phenotype who have mutations in the same gene. Theoretically, three unrelated cases with homozygous or compound heterozygous variants are sufficient to confirm a link between gene and disease for autosomal recessive disorders and five affected cases with heterozygous variants are necessary for autosomal dominant disorders (Gilissen et al., 2012). This is a difficult task in rare and ultrarare disorders, for which the 'N=1' problem makes it difficult to establish causality when only a single individual or family has been identified. It is not impossible to publish studies on single affected individuals or families; careful assessment of the genetic evidence is required, and it is insufficient to assume a variant is disease causing simply because it is rare (Whiffin et al., 2019). In fact, it is now accepted that the previous thresholds for calling a variant rare (and therefore potentially deleterious) were too lenient and may account for many published false associations between variants and human disease (Lek et al., 2016). In the absence of independent cases, detailed functional studies in model systems are essential for providing the evidence of causality (Liegel et al., 2019; Lukacs et al., 2019).

To overcome the 'N=1' problem, programs such as Matchmaker Exchange (<http://matchmakerexchange.org>) have been created to enable sharing of phenotypic and genotypic information to connect cases from around the world (Sobreira et al., 2017). Whether matches are made between cases or not, functional studies can provide crucial insights on gene function and the consequences of genetic variants.

But how do the clinicians or researchers with these cases get matched to the basic scientists most equipped to carry out these studies? In Matchmaker Exchange, the same sharing mechanisms that connect clinicians to each other also connect clinicians with the basic scientists working on a particular gene, pathways or phenotypes in model organisms. One of the many projects connected to Matchmaker Exchange is GeneMatcher (see Table 1), a web-based tool to which clinicians or basic science researchers can submit genetic or phenotypic information (Au et al., 2015; Sobreira et al., 2015). When a match is made, submitters receive an email notification with contact information of their match

to follow up on the submitted information. As of February 2020, GeneMatcher has 40,450 submissions from 8018 submitters in 89 countries. The submissions cover 12,065 unique genes but only 6834 have matches. The numbers of users and submissions are rapidly increasing and many new gene discoveries have been made (Au et al., 2015; Yigit et al., 2020).

A similar resource connected to GeneMatcher, but searchable through its own website, is DECIPHER (Database of Genomic Variation and Phenotype in Humans Using Ensemble Resources) (Table 1) (Firth et al., 2009). The DECIPHER database is a repository of genetic variants (copy-number variants and exomes) and associated phenotypes from patients with neurocognitive developmental disorders, many of whom have additional structural defects.

Another less conventional option for identifying novel genes in need of follow-up is to connect with genetic testing laboratories. Genetic testing that used to be carried out as a several-hundred gene panel is increasingly done through targeted sequencing panels capturing thousands of genes, or as whole-exome sequencing, for technical efficiency. Only a portion of clinical tests are positive, but the genetic data for the remaining 'unsolved' patients are static and easily stored. Reanalysis of the existing, deidentified data from clinical laboratories has the potential to reveal novel genes and new diagnoses (Butler et al., 2018; Mattison et al., 2018).

Interpreting variants

As sequencing technologies have evolved so has the thinking about access to one's own genetic information. Patients and their families now have unprecedented access to their genetic information through traditional providers and direct-to-consumer genotyping and sequencing companies. Families are organizing websites, blog posts and social media support groups to share genetic information and observations, and to catalyze discoveries (Might and Wilsey, 2014). MyGene2 (mygene2.org) (Table 1) is a Matchmaker Exchange-associated portal on which families can create their own profiles to share their genetic information and offer support. Like GeneMatcher, MyGene2 also allows clinicians and researchers studying a rare disorder or candidate gene to create profiles for matching.

We are witnessing a democratization of the gene discovery process. Families, armed with genetic information and a desire for

Variant: ZFH4 c.C7990T (p.Q2644X) identified as a *de novo* mutation in a child with cleft lip and palate

A ClinVar

Showing for results for variants in the ZFH4 gene. [Search instead for all ClinVar records that mention ZFH4](#)

Search result

Filters activated: Frameshift, Nonsense, Splice site. [Clear all](#) to show 66 items.

The following terms were not found in ClinVar: splice 3[Molecular consequence], splice 5[Molecular consequence], splice site[Molecular consequence].

Variation Location	Gene(s)	Protein change	Condition(s)	Clinical significance (Last reviewed)	Review status	Accession
1. NM_024721.5(ZFH4):c.3341del (p.Gly1114fs) GRCh37: Chr8:77745587 GRCh38: Chr8:76833351	ZFH4	G1114fs	not provided	Uncertain significance (Mar 6, 2018)	criteria provided, single submitter	VCV000504414

B gnomAD

gnomAD browser gnomAD v2.1.1 About News Downloads Terms Publications Contact FAQ

gnomAD v3 released! 71,702 genomes aligned on GRCh38.

ZFH4 zinc finger homeobox 4

Dataset: gnomAD v2.1.1 | gnomAD SVs v2.1

Genome build: GRCh37 / hg19
 Ensembl gene ID: ENSG00000091656.11
 Canonical transcript: ENST00000521891.2
 Region: 8:77593454-77779521
 References: [Ensembl](#), [UCSC Browser](#), and [more](#)

Constraint

Category	Exp. SNVs	Obs. SNVs	Constraint metrics
Synonymous	781.2	812	Z = -0.87 o/e = 1.04 (0.98 - 1.1)
Missense	1956.2	1831	Z = 1.01 o/e = 0.94 (0.9 - 0.97)
pLoF	117.6	10	pLI = 1 o/e = 0.09 (0.05 - 0.14)

Variant ID	Source	Consequence	Annotation	Flags	Allele Count	Allele Number	Allele Frequency
8-77767140-G-A	E	p.Arg2661Arg	synonymous		3	248344	1.21e-5
8-77767146-C-G	E	p.Gly2663Gly	synonymous		1	248302	4.03e-6
8-77767148-A-G	E	p.Gln2664Arg	missense		1	248414	4.03e-6
8-77767149-G-A	E	p.Gln2664Gln	synonymous		1	248448	4.02e-6
8-77767153-C-T	E	p.Arg2666Trp	missense		1	248408	4.03e-6
8-77767157-C-T	E	p.Ala2667Val	missense		10	248454	4.02e-5
8-77767158-G-A	E	p.Ala2667Ala	synonymous		4	279908	1.43e-5
8-77767158-G-T	E	p.Ala2667Ala	synonymous		1	248510	4.02e-6

C In-silico predictions

Uploaded variant	Consequence	Symbol	Protein position	Amino acids	MANE	SIFT	PolyPhen	MutationTaster pred	CADD PHRED	LoFtool
8_76854911_C/T	stop_gained	ZFH4	2638	Q/*	-	-	-	A, A, A, A	44	0.421

D Matchmaker

GENE MATCHER Home About Publications Statistics Contact Us Help My Account Log Out

Submissions **Gene Search** [Diagnosis Search](#) [Reports](#) [Match Outcomes](#)

Quick Search

Identifier...

(Submission ID, identifier)

Submissions

- Create a new submission
- List my active submissions (four entries)
- List my suspended submissions (no entries)
- List all my submissions (four entries)

Fig. 1. See next page for legend.

Fig. 1. Variant interpretation. (A-D) An example of how to interpret variants using some of the described tools. The example variant, a stop-gain mutation in *ZFHX4* (p.Q2446X) recently reported from whole-genome sequencing in trios with cleft lip and palate (Bishop et al., 2020 preprint). (A) ClinVar shows no known pathogenic variants deposited. The only variant is a frameshift mutation classified clinically as a ‘variant of unknown significance’. By clicking on the variant, you can read about the evidence. (B) gnomAD shows that *ZFHX4* is a constrained gene because only 10 predicted loss-of-functions were identified but 117 were expected ($o/e=0.14$). The Q2644X variant is novel and was absent from >124,000 individuals. (C) As a nonsense variant, Q2644X is automatically predicted to be damaging (Ensembl’s Variant Effect Predictor shown here). For missense variants, agreement across multiple prediction tools can be used as contributing evidence for pathogenicity. Given all this information, the variant could be pathogenic. Submitting to Matchmaker (D) could identify more mutations or find investigators interested in research on this gene.

answers, may directly contact basic scientists for help: ‘My son has a mutation in this gene. I read your papers, does this mutation cause his condition?’ Before hanging up the phone, consider the resources at your disposal. Of course, an appropriate response will always be to refer them to a clinical geneticist at your institution. But there is value in developmental biology researchers knowing how to interpret genetic variants (Fig. 1). A first stop should be ClinVar (Table 1), a database in which genetic variants from clinical labs are catalogued and annotated using pathogenicity guidelines from the American College of Medical Genetics and Genomics (ACMG) (Richards et al., 2015).

The second stop should be gnomAD (Table 1), a database of 71,702 genomes (gnomAD v3) or 125,748 exomes and 15,708 genomes (gnomAD v2) from anonymous individuals from around the world (Karczewski et al., 2020 preprint). There is no phenotype information available, but individuals known to be affected by severe pediatric conditions have been removed, so these databases allow us to define ‘normal’ genetic variation in adult human populations. For rare disorders, any variants found with an allele frequency greater than 0.001–0.1% and in an affected individual are unlikely to be disease causing. As indicated earlier, analyzing allele frequencies allows us to remove variants that are too common to cause disease, but every individual carries hundreds of rare genetic variants so rarity is insufficient to classify a variant as disease-causing. *In silico* prediction tools (i.e. Polyphen, SIFT, Mutation Taster) lack sensitivity and specificity, but agreement across multiple tools is an additional piece of evidence considered under ACMGG pathogenicity guidelines.

Another useful metric for determining if a variant is potentially causal is to determine whether the gene is under constraint (the intolerance of a gene to heterozygous variants due to natural selection). In gnomAD, this is measured as ratios of observed counts of synonymous, missense or loss-of-function variants to expected counts based on a mutational model that considers sequence context and coverage (Karczewski et al., 2020 preprint; Samocha et al., 2014). For example, haploinsufficient genes, such as those that cause dominant Mendelian syndromes, are depleted for loss-of-function variants in gnomAD, whereas genes that are dispensable, such as olfactory receptors, are unconstrained. What this means for establishing pathogenicity of a variant for a congenital disorder is that identifying a rare loss-of-function variant in a gene with high constraint is supportive evidence of pathogenicity. To give some benchmarks for these constraint scores or observed/expected ratios (o/e): a gene with a loss-of-function o/e of 0.4 indicates that only 40% of the expected loss-of-function variants were observed and therefore is likely under selection against loss-of-function variants. In clinical and research interpretation of Mendelian and

rare disorders, it is suggested to use the $o/e < 0.35$ if a hard threshold is needed to identify a constrained gene (lower o/e scores are even better).

Variant interpretation requires careful assessment of multiple lines of evidence on allele frequency in large population databases, properties of the mutated gene and *in silico* pathogenicity predictions. The computational tools and resources described here can be helpful in prioritizing variants but, ultimately, functional studies in model systems are a crucial piece of evidence in this endeavor.

What is not solved

Since the advent of exome sequencing, clinical genetics prognosticators have discussed the day when all Mendelian disease genes will be identified. The identification of disease genes and recapitulation of the human phenotype in a model system should not be the end point for disease gene discovery. In addition to the still unknown fraction of disease attributable to the noncoding portion of the human genome (Smedley et al., 2016), we do not yet understand the extent to which allelic and locus heterogeneity, the presence of pathogenic variants at multiple loci and common variants influence Mendelian disorders. The same is true for non-Mendelian disorders, which are even more complex. A more precise understanding of variable expressivity and penetrance will be necessary to completely realize the goals of genomic medicine. There is therefore a need for the development of high-throughput assays to test human variants for individual genes and non-coding regulatory elements (Findlay et al., 2018; Kvon et al., 2020). Not only could these assays distinguish benign from pathogenic variation, quantitative assays may sort out differences in disease severity. The implementation of these assays is predicated on knowing about how these genes are regulated, the function of their encoded proteins and the mechanisms by which they cause human disease, insights that are best gleaned from bench science and model organisms. Large-scale efforts from the International Mouse Phenotyping Consortium (Meehan et al., 2017) and the Deep Genome Project (Lloyd et al., 2020) to annotate and catalog the function of human gene orthologs in the mouse genome form a foundation for these studies. Human disease is also causally influenced and modified by environmental exposures. For congenital disorders, the interplay of gene and environmental exposures necessarily must be studied in model systems, which may include a combination of traditional animal models and advanced *in vitro* models such as organoids (Lancaster and Knoblich, 2014) and microphysiological models (Wikswow, 2014).

Sequencing and gene discovery in human disease and congenital disorders will continue at a rapid pace for the foreseeable future. An interdisciplinary effort is essential to understand the etiology of congenital disorders and elucidate pathogenic mechanisms. Functional follow-up of variants is necessary, time-consuming and costly, making it even more important that geneticists and developmental biologists work together to prioritize variants for the most impact. Making the most of these publicly available sequence resources will limit duplication of effort and, as the Matchmaker model shows, the path from discovery to functional studies need not be unidirectional.

Acknowledgements

Many thanks to Tamara Caspary and Rob Lipinski for helpful discussions in preparing this manuscript, and to the many study participants and colleagues from all disciplines that make our research on craniofacial birth defects meaningful.

Competing interests

The authors declare no competing or financial interests.

Funding

The authors' research is supported by the National Institutes of Health (DE025060, DE027193). Deposited in PMC for release after 12 months.

References

- Alkuraya, F. S. (2015). Natural human knockouts and the era of genotype to phenotype. *Genome Med.* **7**, 48. doi:10.1186/s13073-015-0173-z
- Au, P. Y. B., You, J., Caluseriu, O., Schwartzentruber, J., Majewski, J., Bernier, F. P., Marcia Ferguson Care for Rare Canada Consortium, Valle, D., Parboosingh, J. S., Sobreira, N. et al. (2015). GeneMatcher aids in the identification of a new malformation syndrome with intellectual disability, unique facial dysmorphisms, and skeletal and connective tissue abnormalities caused by de novo variants in HNRNP. *Hum. Mutat.* **36**, 1009-1014. doi:10.1002/humu.22837
- Bishop, M., Diaz Perez, K., Sun, M., Ho, S., Chopra, P., Mukhopadhyay, N., Hetmanski, J., Taub, M., Moreno-Urbe, L., Valencia-Ramirez, L. C. et al. (2020). Genome-wide enrichment of de novo coding mutations in orofacial cleft trios. *Am. J. Hum. Genet.* (in press). doi:10.1016/j.ajhg.2020.05.018
- Botto, L. D., Flood, T., Little, J., Fluchel, M. N., Krikov, S., Feldkamp, M. L., Wu, Y., Goedken, R., Puzhankara, S. and Romitti, P. A. (2013). Cancer risk in children and adolescents with birth defects: a population-based cohort study. *PLoS ONE* **8**, e69077. doi:10.1371/journal.pone.0069077
- Boycott, K. M., Rath, A., Chong, J. X., Hartley, T., Alkuraya, F. S., Baynam, G., Brookes, A. J., Brudno, M., Carracedo, A., den Dunnen, J. T. et al. (2017). International cooperation to enable the diagnosis of all rare genetic diseases. *Am. J. Hum. Genet.* **100**, 695-705. doi:10.1016/j.ajhg.2017.04.003
- Butler, K. M., Moody, O. A., Schuler, E., Coryell, J., Alexander, J. J., Jenkins, A. and Escayg, A. (2018). De novo variants in GABRA2 and GABRA5 alter receptor function and contribute to early-onset epilepsy. *Brain* **141**, 2392-2405. doi:10.1093/brain/awy171
- Carozza, S. E., Langlois, P. H., Miller, E. A. and Canfield, M. (2012). Are children with birth defects at higher risk of childhood cancers? *Am. J. Epidemiol.* **175**, 1217-1224. doi:10.1093/aje/kwr470
- Findlay, G. M., Daza, R. M., Martin, B., Zhang, M. D., Leith, A. P., Gasperini, M., Janizek, J. D., Huang, X., Starita, L. M. and Shendure, J. (2018). Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **562**, 217-222. doi:10.1038/s41586-018-0461-z
- Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R. M. and Carter, N. P. (2009). DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am. J. Hum. Genet.* **84**, 524-533. doi:10.1016/j.ajhg.2009.03.010
- Gilissen, C., Hoischen, A., Brunner, H. G. and Veltman, J. A. (2012). Disease gene identification strategies for exome sequencing. *Eur. J. Hum. Genet.* **20**, 490-497. doi:10.1038/ejhg.2011.258
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P. et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443. doi:10.1038/s41586-020-2308-7
- Kvon, E. Z., Zhu, Y., Kelman, G., Novak, C. S., Plajzer-Frick, I., Kato, M., Garvin, T. H., Pham, Q., Harrington, A. N., Hunter, R. D. et al. (2020). Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. *Cell* **180**, 1262-1271.e1215. doi:10.1016/j.cell.2020.02.031
- Lancaster, M. A. and Knoblich, J. A. (2014). Organogenesis in a dish: modeling development and disease using organoid technologies. *Science* **345**, 1247125. doi:10.1126/science.1247125
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B. et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291. doi:10.1038/nature19057
- Liegel, R. P., Finnerty, E., Blizzard, L., DiStasio, A., Hufnagel, R. B., Saal, H. M., Sund, K. L., Prows, C. A. and Stottmann, R. W. (2019). Using human sequencing to guide craniofacial research. *Genesis* **57**, e23259. doi:10.1002/dvg.23259
- Lim, E. T., Würtz, P., Havulinna, A. S., Palta, P., Tukiainen, T., Rehnström, K., Esko, T., Mägi, R., Inouye, M., Lappalainen, T. et al. (2014). Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* **10**, e1004494. doi:10.1371/journal.pgen.1004494
- Lloyd, K. C. K., Adams, D. J., Baynam, G., Beaudet, A. L., Bosch, F., Boycott, K. M., Braun, R. E., Caulfield, M., Cohn, R., Dickinson, M. E. et al. (2020). The Deep Genome Project. *Genome Biol.* **21**, 18. doi:10.1186/s13059-020-1931-9
- Lukacs, M., Gilley, J., Zhu, Y., Orsomando, G., Angeletti, C., Liu, J., Yang, X., Park, J., Hopkin, R. J., Coleman, M. P. et al. (2019). Severe biallelic loss-of-function mutations in nicotinamide mononucleotide adenylyltransferase 2 (NMNAT2) in two fetuses with fetal akinesia deformation sequence. *Exp. Neurol.* **320**, 112961. doi:10.1016/j.expneurol.2019.112961
- Mattison, K. A., Butler, K. M., Inglis, G. A. S., Dayan, O., Boussidan, H., Bhamhani, V., Philbrook, B., da Silva, C., Alexander, J. J., Kanner, B. I. et al. (2018). SLC6A1 variants identified in epilepsy patients reduce γ -aminobutyric acid transport. *Epilepsia* **59**, e135-e141. doi:10.1111/epi.14531
- Meehan, T. F., Conte, N., West, D. B., Jacobsen, J. O., Mason, J., Warren, J., Chen, C.-K., Tudose, I., Relac, M., Matthews, P. et al. (2017). Disease model discovery from 3,328 gene knockouts by The International Mouse Phenotyping Consortium. *Nat. Genet.* **49**, 1231-1238. doi:10.1038/ng.3901
- Might, M. and Wilsey, M. (2014). The shifting model in clinical diagnostics: how next-generation sequencing and families are altering the way rare diseases are discovered, studied, and treated. *Genet. Med.* **16**, 736-737. doi:10.1038/gim.2014.23
- Narasimhan, V. M., Hunt, K. A., Mason, D., Baker, C. L., Karczewski, K. J., Barnes, M. R., Barnett, A. H., Bates, C., Bellary, S., Bockett, N. A. et al. (2016a). Health and population effects of rare gene knockouts in adult humans with related parents. *Science* **352**, 474-477. doi:10.1126/science.aac8624
- Narasimhan, V. M., Xue, Y. and Tyler-Smith, C. (2016b). Human knockout carriers: dead, diseased, healthy, or improved? *Trends Mol. Med.* **22**, 341-351. doi:10.1016/j.molmed.2016.02.006
- Posey, J. E., O'Donnell-Luria, A. H., Chong, J. X., Harel, T., Jhangiani, S. N., Coban Akdemir, Z. H., Buyske, S., Pehlivan, D., Carvalho, C. M. B., Baxter, S. et al. (2019). Insights into genetics, human biology, and disease gleaned from family based genomic studies. *Genet. Med.* **21**, 798-812. doi:10.1038/s41436-018-0408-7
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E. et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405-423. doi:10.1038/gim.2015.30
- Saleheen, D., Natarajan, P., Armean, I. M., Zhao, W., Rasheed, A., Khetarpal, S. A., Won, H.-H., Karczewski, K. J., O'Donnell-Luria, A. H., Samocha, K. E. et al. (2017). Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* **544**, 235-239. doi:10.1038/nature22034
- Samocha, K. E., Robinson, E. B., Sanders, S. J., Stevens, C., Sabo, A., McGrath, L. M., Kosmicki, J. A., Rehnström, K., Mallick, S., Kirby, A. et al. (2014). A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46**, 944-950. doi:10.1038/ng.3050
- Smedley, D., Schubach, M., Jacobsen, J. O. B., Köhler, S., Zemojtel, T., Spielmann, M., Jäger, M., Hochheiser, H., Washington, N. L., McMurry, J. A. et al. (2016). A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. *Am. J. Hum. Genet.* **99**, 595-606. doi:10.1016/j.ajhg.2016.07.005
- Sobreira, N., Schiettecatte, F., Valle, D. and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* **36**, 928-930. doi:10.1002/humu.22844
- Sobreira, N. L. M., Arachchi, H., Buske, O. J., Chong, J. X., Hutton, B., Foreman, J., Schiettecatte, F., Groza, T., Jacobsen, J. O. B., Haendel, M. A. et al. (2017). Matchmaker exchange. *Curr. Protoc. Hum. Genet.* **95**, 9.31.1-9.31.15. doi:10.1002/cphg.50
- Stoeger, T., Gerlach, M., Morimoto, R. I. and Nunes Amaral, L. A. (2018). Large-scale investigation of the reasons why potentially important genes are ignored. *PLoS Biol.* **16**, e2006643. doi:10.1371/journal.pbio.2006643
- Wallingford, J. B. (2019). We are all developmental biologists. *Dev. Cell* **50**, 132-137. doi:10.1016/j.devcel.2019.07.006
- Whiffin, N., Ware, J. S. and O'Donnell-Luria, A. (2019). Improving the understanding of genetic variants in rare disease with large-scale reference populations. *JAMA* **322**, 1305-1306. doi:10.1001/jama.2019.12891
- Wikswø, J. P. (2014). The relevance and potential roles of microphysiological systems in biology and medicine. *Exp. Biol. Med. (Maywood)* **239**, 1061-1072. doi:10.1177/1535370214542068
- Yigit, G., Saida, K., DeMarzo, D., Miyake, N., Fujita, A., Yang Tan, T., White, S. M., Wadley, A., Toliat, M. R., Motameny, S. et al. (2020). The recurrent postzygotic pathogenic variant p.Glu47Lys in RHOA causes a novel recognizable neuroectodermal phenotype. *Hum. Mutat.* **41**, 591-599. doi:10.1002/humu.23964