



Insights into genetics, human biology and disease gleaned from family based genomic studies

Jennifer E. Posey, MD, PhD¹, Anne H. O'Donnell-Luria, MD, PhD^{2,3,4}, Jessica X. Chong, PhD⁵, Tamar Harel, MD, PhD⁶, Shalini N. Jhangiani, MS⁷, Zeynep H. Coban Akdemir, PhD¹, Steven Buyske, PhD^{8,9}, Davut Pehlivan, MD¹, Claudia M. B. Carvalho, PhD¹, Samantha Baxter, MS, CGC³, Nara Sobreira, MD, PhD¹⁰, Pengfei Liu, PhD^{1,11}, Nan Wu, MD^{1,12}, Jill A. Rosenfeld, MS, CGC¹, Sushant Kumar, PhD¹³, Dimitri Avramopoulos, MD, PhD¹⁰, Janson J. White, PhD^{1,5}, Kimberly F. Doheny, PhD^{10,14}, P. Dane Witmer, PhD^{10,14}, Corinne Boehm, MS, CGC¹⁰, V. Reid Sutton, MD¹, Donna M. Muzny, MS⁷, Eric Boerwinkle, PhD^{7,15}, Murat Günel, MD^{16,17}, Deborah A. Nickerson, PhD¹⁸, Shrikant Mane, PhD¹⁹, Daniel G. MacArthur, PhD^{2,3}, Richard A. Gibbs, PhD^{1,7}, Ada Hamosh, MD¹⁰, Richard P. Lifton, MD, PhD^{16,20,21}, Tara C. Matise, PhD⁸, Heidi L. Rehm, PhD^{2,3}, Mark Gerstein, PhD¹³, Michael J. Bamshad, MD^{5,18}, David Valle, MD¹⁰ and James R. Lupski, MD, PhD, DSc (hon)^{1,7,22,23} Centers for Mendelian Genomics

Identifying genes and variants contributing to rare disease phenotypes and Mendelian conditions informs biology and medicine, yet potential phenotypic consequences for variation of >75% of the ~20,000 annotated genes in the human genome are lacking. Technical advances to assess rare variation genome-wide, particularly exome sequencing (ES), enabled establishment in the United States of the National Institutes of Health (NIH)-supported Centers for Mendelian Genomics (CMGs) and have facilitated collaborative studies resulting in novel “disease gene” discoveries. Pedigree-based genomic studies and rare variant analyses in families with suspected Mendelian conditions have led to the elucidation of hundreds of novel disease genes and highlighted the impact of de novo mutational events, somatic variation underlying nononcologic traits, incompletely penetrant alleles, phenotypes with high locus heterogeneity, and multilocus pathogenic variation. Herein, we highlight CMG collaborative discoveries that have contributed to understanding both rare and common diseases and discuss opportunities for future discovery in single-locus Mendelian

disorder genomics. Phenotypic annotation of all human genes; development of bioinformatic tools and analytic methods; exploration of non-Mendelian modes of inheritance including reduced penetrance, multilocus variation, and oligogenic inheritance; construction of allelic series at a locus; enhanced data sharing worldwide; and integration with clinical genomics are explored. Realizing the full contribution of rare disease research to functional annotation of the human genome, and further illuminating human biology and health, will lay the foundation for the Precision Medicine Initiative.

Genetics in Medicine (2019) 21:798–812; <https://doi.org/10.1038/s41436-018-0408-7>

Keywords: rare variant phenotypes; Mendelian conditions; Centers for Mendelian Genomics (CMG); genetic models for disease; disease traits

INTRODUCTION

Mendelian conditions are individually rare, but collectively contribute to disease in ~0.4% of children and young adults, and 8% of live births if all congenital anomalies are considered.¹ These findings likely underestimate the true burden of Mendelian conditions; the estimates focus on the severe end of the phenotypic spectrum and often fail to capture disorders caused by de novo pathogenic variant alleles or characterized by adult onset. Prior to the broader availability of genome-wide assays, discovery of loci

underlying Mendelian conditions relied heavily on traditional genetic mapping and positional cloning approaches that had little power to detect disorders characterized by de novo variation, incomplete penetrance, and locus heterogeneity. Chromosome microarray analysis (CMA) and next-generation sequencing (NGS), applied to well-phenotyped individuals,² have provided substantial technological advances toward clinical genomics and identifying a more complete variant spectrum (single-nucleotide variants [SNVs], indels, and copy-number variants [CNVs]) and molecular basis for

Correspondence: Jennifer E. Posey (jennifer.posey@bcm.edu) or James R. Lupski (jlupski@bcm.edu)

Submitted 12 June 2018; accepted: 5 December 2018

Published online: 18 January 2019

human Mendelian conditions.^{3–10} Despite substantial progress in variant detection genome-wide, the overwhelming majority of annotated genes have yet to be assigned function in the context of human disease traits. Thus, a comprehensive molecular understanding of disease biology and disease gene function remains to be achieved.

Several national and international programs have been developed to both stimulate and support the study of Mendelian conditions. In Canada, Finding of Rare Disease Genes (FORGE)¹¹ and Care4Rare Canada Consortium¹² have contributed to this global effort, leading to development of the Canadian Rare Diseases Models and Mechanisms Network (RDMM), which supports collaboration among clinical and human geneticists and model organism researchers in the study of rare variants and their functional impact. In the UK, the Deciphering Developmental Disorders (DDD)^{13–17} study has for over a decade made significant contributions to the understanding of the molecular etiologies of neurodevelopmental delay and the roles of different variant types, mutational mechanisms, and new pathogenic variants in disease traits. Additional international efforts in rare disease gene discovery include the Undiagnosed Diseases Network International (UDNI)¹⁸ and the International Rare Diseases Research Consortium (IRDiRC).¹⁹

In the United States, the Centers for Mendelian Genomics (CMGs)^{20,21} and Undiagnosed Diseases Network (UDN)²² use complementary approaches to investigate the molecular etiology of Mendelian conditions. The CMGs comprise four Centers: a joint Baylor College of Medicine–Johns Hopkins University Center (BHCMG), the Broad Institute/Harvard University (BIHCMG), University of Washington (UWCMG), and Yale University (YCMG) (www.mendelian.org). The Centers are supported by the National Human Genome Research Institute (NHGRI); the National Heart, Lung, and Blood Institute (NHLBI); and the National Eye Institute (NEI); are leveraged by local resources; and are focused on shared goals of novel disease gene discovery using exome and genome sequencing (ES/GS), and rare variant, family based genomics approaches. Knowledge dissemination is facilitated through publication (both in scientific journals and online at www.mendelian.org), resource and data sharing, education of the scientific and medical communities, and collaboration with clinicians, families, and researchers worldwide.

ES coupled with the power inherent in a rare variant, family based analysis enables the identification of rare, de novo, and cosegregating variants with large phenotypic effect, i.e., disease traits tied to a specific locus, yielding results of immediate clinical utility and driving novel disease gene discovery. Gene-first approaches, in which a cohort of individuals with rare variation at a particular locus undergo careful phenotyping, can elucidate the full phenotypic spectrum associated with an allelic series at a disease gene locus.^{23,24} For example, analysis of variation at the *POGZ* locus led to the delineation of White–Sutton syndrome (WHSUS; MIM 616364), after phenotype-focused cohort

studies identified only a subset (developmental delay/intellectual disability [DD/ID], autism spectrum disorder, schizophrenia) of the cognitive phenotypes associated with rare variation in *POGZ*.^{14,25–29} Further examples of this CMG and collaborator-facilitated gene-first approach include studies of *ASXL3* (refs.^{30,31}), *CDK13* (ref.³²), *PNPLA1* (ref.³³), *POLE*,³⁴ *IFT81* (ref.³⁵), *HDAC8* (ref.³⁶), *AHDC1* (ref.³⁷), and *CDC42* (ref.³⁸).

Disease gene discovery and genetic diagnosis informs clinical care

Rare disease research and gene discovery inform and enhance molecular diagnosis and can impact patient management. Molecular diagnostic assays provide potential precise genetic contributors to clinical diagnoses, important prognostic information, and guidance for clinical management and disease surveillance, and enable more accurate recurrence risk estimates for families. In turn, this individualized “precision” information provides an entry for illuminating disease biology and insight, enabling development and implementation of rational and targeted therapeutics. For example, the discovery of loss-of-function (LoF) *PCSK9* variants causing hypocholesterolemia led to the rapid development of monoclonal antibodies targeting PCSK9, to treat cardiovascular disease and familial hypercholesterolemia.^{39–41}

At the initiation of the CMG program, we and others predicted that opportunities provided by NGS technologies, novel computational analytic approaches, and the access to these technologies for clinicians and families from populations around the world would transform the field of Mendelian genomics, and our understanding of both human biology and perturbations to homeostasis resulting in disease.^{10,20,42,43} However, it was not anticipated that such studies might potentially enable building testable models for the genetics of disease from the bottom up. In the next section, we highlight CMG accomplishments that are driving this transformation.²¹

CMG-FACILITATED DISCOVERIES

Disease gene discovery and functional annotation of the human genome

A primary goal of the CMGs is to identify novel disease genes responsible for human Mendelian conditions.^{20,21} The CMGs have reported a total of 3617 disease gene–phenotype pairs (<http://mendelian.org/phenotypes-genes>), categorized as novel, phenotypic expansion (phenotypic features extending beyond those previously reported for a Mendelian condition),²¹ or known (Fig. 1, Supplemental Figure 1). The CMGs are well positioned to achieve the overall goal of connecting phenotypes to high penetrance variants in a substantial fraction of all ~20,000 annotated human genes, and the current pace of discovery within the CMGs does not show evidence of slowing (Fig. 2). This simple accounting or “tally” of gene discovery does not fully represent the genetic and genomic insights and new understanding generated by CMG international collaborative efforts regarding disease traits,

human biology, and human developmental and homeostatic processes.

Dissemination of knowledge

As of 31 August 2018, the CMGs have contributed to a total of 522 manuscripts with collaborators worldwide (Table 1). These efforts have supported the establishment of tenure-track positions for 16 junior faculty, and the successful preparation of 9 K- or R-level NIH-funded grants (Appendix, Supplemental Tables 1 and 2), and aided in the training of numerous graduate students. The CMGs have also taken steps to disseminate prepublished data to the scientific community. The Genomic Sequencing Program Coordinating Center (GSPCC)-managed CMG website provides public access to a searchable phenotypes and genes database of CMG disease gene discoveries; depositions to ClinVar and dbGaP further support knowledge dissemination.

Contribution to clinical diagnosis and patient management

The CMGs have engaged diagnostic laboratories as an extension of the research laboratory efforts, increasing gene discovery through analysis of nondiagnostic clinical exomes. This interaction facilitates rapid transition from novel disease gene discovery to patient report, with direct involvement of additional stakeholders in the discovery efforts.⁴⁴ Review of 12,577 sequential noncancer cases referred to the Baylor Genetics diagnostic laboratory yielded 4075 cases for which ES was diagnostic. Of these, 333 molecular diagnoses explaining part or all of the reported clinical phenotype involved CMG discovery genes (Supplemental Figure 2A). A precise molecular diagnosis (*PGM3* [ref. ⁴⁵], *TANGO2* [ref. ⁴⁶] and *ABL1* [ref. ⁴⁷]) informed medical management of 21 individuals, with several additional clinically impactful CMG disease gene discoveries beyond this clinical cohort (Supplemental Table 3). Other collaborations between the CMGs and worldwide diagnostic and research laboratories make extensive use of the Matchmaker Exchange network^{48,49} which includes CMG-developed nodes GeneMatcher,^{50,51} MyGene2 (ref. ⁵²), and *matchbox*, facilitating novel disease gene discoveries worldwide (Supplemental Table 4). The impact of CMG disease gene discoveries on molecular diagnostics is further reflected in the number of pathogenic or likely pathogenic variant entries in discovery genes in ClinVar and the Genetic Testing Registry database (GTR, <https://www.ncbi.nlm.nih.gov/gtr/>; Supplemental Figure 2B–D). These findings illustrate the substantial impact of the CMGs on both clinical diagnostics and medical management, demonstrating unequivocally a successful “bedside-to-bench-to-bedside” approach.

Molecular mechanisms underlying Mendelian conditions

Over the past decade there has been tremendous progress in understanding the molecular basis of and mechanisms underlying Mendelian conditions. De novo pathogenic variants have been increasingly recognized as a major source of rare conditions, particularly those that reduce reproductive

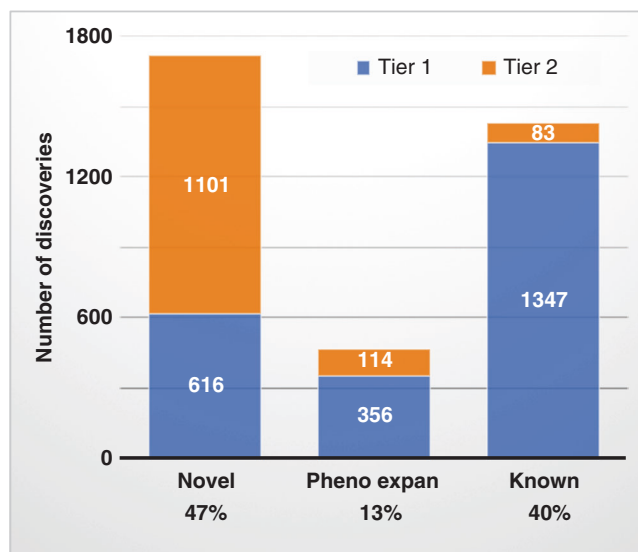


Fig. 1 Centers for Mendelian Genomics (CMG) disease gene discovery through 31 May 2018 (year 7, quarter 2) by the four centers.

Discoveries are defined as “novel” if (1) the causal variant was identified in a gene not previously associated with a Mendelian phenotype at the time of case acceptance to the study (i.e., novel gene), or (2) the causal variant was identified in association with a Mendelian phenotype with a MIM number (a known phenotype) and for which no causal variants had previously been reported (i.e., novel gene, unexplained known phenotype), or (3) the causal variant was identified in association with a Mendelian phenotype with no MIM number and for which no variants in the identified gene had been previously reported as causal of a Mendelian phenotype (i.e., novel gene, new phenotype). Graph of discoveries (genotype–phenotype pairs) categorized as novel, phenotypic expansion, or known. Discoveries are classified as either tier 1 (blue bars) or tier 2 (orange bars, not meeting tier 1 definition). Tier 1 genes include high confidence genes reported by individual centers as tier 1, defined as having been identified in either (1) multiple kindreds with shared phenotypic features and likely pathogenic variants in the same gene, or (2) a single family plus a model organism with orthologous phenotypic features, or (3) a single family with supportive functional and mapping data. *Pheno expan* phenotypic expansion.

fitness.^{53–55} This has been borne out in clinical referral cohorts across all ages,^{56,57} as well as across numerous phenotypes, such as neurodevelopmental disorders,^{58,59} Meier–Gorlin syndrome (MIM 616835) (ref. ⁶⁰), visceral myopathy (megacystis–microcolon–intestinal hypoperistalsis syndrome [MIM 155310]) (ref. ⁶¹), and nasopalpebral lipoma–coloboma syndrome.⁶² Somatic mosaicism has been demonstrated to be an important contributor to rare disease.^{63–65} Parental mosaicism can impact recurrence risk counseling for families with apparently sporadic disease, and the likelihood of parental germline mosaicism is dependent on whether the new variant arises on the maternal or paternal allele.^{66–68} Proband mosaicism has been found to underlie many conditions, including Cornelia de Lange syndrome (CdLS), for which both genetic heterogeneity and mosaicism can impact clinical expressivity of disease.^{69–72} Mosaic reversion of pathogenic variants to wild type has also been described in ichthyosis with confetti lesions caused by variants in *KRT10* or *KRT1* (refs. ^{73,74}), and in immunodeficiency

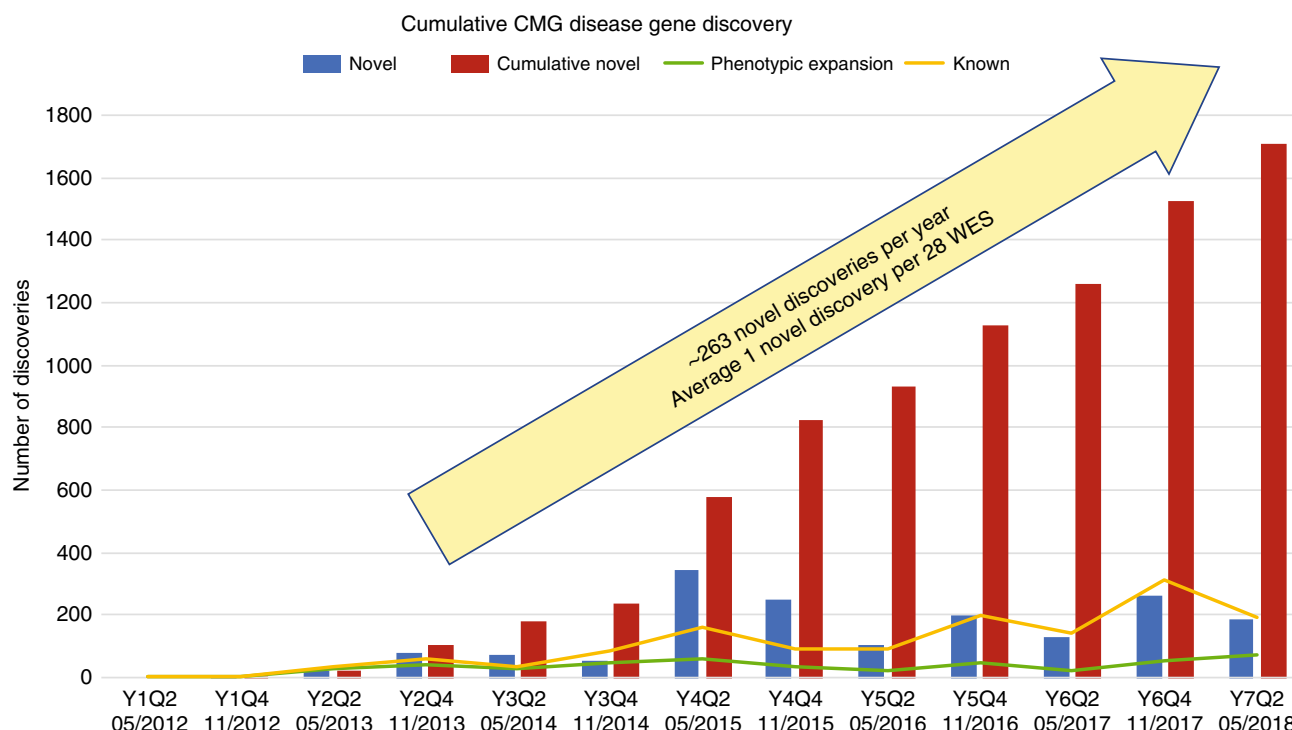


Fig. 2 Cumulative Centers for Mendelian Genomics (CMG) disease gene discovery. The number of novel gene–phenotype discoveries as reported by all four centers is graphed by progress reporting period (blue bars) and cumulatively (red bars). Biannual phenotypic expansion discoveries involving previously known disease genes (green line) and biannual known disease gene discoveries (yellow line) are also graphed. The yellow arrow indicates the pace of novel gene–phenotype discovery, and demonstrates a pace of 263 novel gene–phenotype discoveries per year, or 1 novel gene–phenotype discovery for every 28 exome sequences (ES) performed.

syndromes for which the affected cell populations are under strong negative selection.⁷⁵

Intragenic CNVs, most notably exon deletion or “dropout” alleles sometimes affecting only a single exon, have been identified as a difficult-to-detect cause of many Mendelian conditions (Supplemental Table 5). Additionally, exonic deletions from clinical diagnostic CMA have fostered gene discovery efforts.⁷⁶ Mosaic and copy-number variants are underdiagnosed by current NGS technologies,⁷⁷ and these examples illustrate the clinical relevance of such discovery and the need to develop and apply dedicated computational pipelines to their identification.

Variation in patterns of disease inheritance

For certain Mendelian conditions, empirical observations suggest more than one pattern of disease trait transmission associated with variation at a single locus (e.g., autosomal recessive, AR; autosomal dominant, AD; in some instances, X-linked, XL; and even common, complex), which may confound genomic mapping studies for a particular trait.^{58,78–82} These observations can be explained by allelic heterogeneity, with different consequences of the pathogenic variants (i.e., LoF, GoF, dominant negative) at a given locus or variable magnitude of the pathogenic variant effect.^{58,79} This is exemplified by *SMCHD1*, for which missense variants located within the ATPase domain are associated with Bosma arhinia microphthalmia syndrome (MIM 603457), whereas

LoF is associated with facioscapulohumeral muscular dystrophy type 2 (FSHD2, MIM 158901) and digenic inheritance.⁸³ Collectively, the CMGs have identified over 30 loci (<http://mendelian.org/phenotypes-genes>) with known or proposed human disease phenotypes for which elucidation of the responsible gene and causative variants explains the clinical observation of both dominant (monoallelic) and recessive (biallelic) inheritance of the corresponding disease traits with either similar or dissimilar phenotypic features (Supplemental Table 6).

There are increasing examples of variants that escape nonsense-mediated decay (NMD) and result in expression of a phenotype due to GoF.^{84–86} An NMD escape intolerance score metric based on the depletion of protein-truncating variants within gene regions predicted to escape NMD may facilitate the identification of variants that function through a GoF mechanism.⁸⁷ Such variants may be present in genes with low probability of loss-of-function intolerance (pLI) scores predicting tolerance to LoF variants.⁸⁷

The CMGs have unraveled the biology of genetic heterogeneity in analyses of several cohorts with apparently homogeneous phenotypes. The identification of novel disease genes (*DVL1*, *DVL3*, *FZD2*, and *NXN*) in Robinow syndrome (MIM 268310, 180700, 616331, 616894) has provided a molecular diagnosis in potentially 95% of the studied disease cohort and highlighted a shared role in the noncanonical Wnt pathway for this phenotype.^{88–90} Similarly, in Noonan

Table 1 Centers for Mendelian Genomics (CMG)-wide recruitment, production, and knowledge dissemination and collaboration achievements

Recruitment		Production		Dissemination & collaboration	
	Number		Number		Number
Samples	61,286	Phenotypes	2248	Publications	522
Families	22,742	ES	47,129	Authors	5832
Collaborators	3928	GS	1169	Organizations	1540
Countries	80	RNAseq	144	Countries	77

CMG-wide recruitment and production achievements as of 31 May 2018. CMG-wide knowledge dissemination, and collaboration achievements as of 31 August 2018. ES exome sequencing, GS genome sequencing, RNAseq RNA sequencing.

syndrome, the CMGs and others have identified novel disease genes with a role in the RAS/MAPK pathway.^{91–100} Rare variation in genes encoding the cohesin complex have now been described to underlie Mendelian conditions termed “cohesinopathies,” which demonstrate clinical features that are similar to those observed in CdLS.^{72,101} The frequency of the cohesin complex subunit protein/gene contribution depends on how the phenotype is ascertained: specifically as CdLS-like phenotypes, or more broadly as DD/ID.⁷² Increasingly, Mendelian inheritance is appreciated as vastly more complicated and nuanced than the simple binary patterns Gregor Mendel described in 1864, and ES now enables delineation of this complexity through identification of allelic and locus heterogeneity in human Mendelian disorders.

CMG-facilitated studies contributed to elucidation of multilocus pathogenic variant effects on disease trait manifestations:

- **Digenic inheritance** has been described in facioscapulo-humeral dystrophy type 2 (FSHD2 [MIM 158901]), involving rare variation in *SMCHD1* and a permissive *DUX4* allele, both required for expression of disease.^{102,103} Digenic inheritance of a rare *SMAD6* variant in association with a common variant downstream of *BMP2* was described in association with midline craniosynostosis.¹⁰⁴ In both examples, the observation of reduced penetrance drove discovery of the second locus required for disease expression.
- **Dual/multiple molecular diagnoses or multilocus pathogenic variation** involving CNVs and/or SNVs result in blended phenotypes estimated to comprise at least 4.9% of all diagnostic clinical exome cases.^{56,57,105–109} Presenting phenotypes may be distinct or overlapping, and may obscure clinical ascertainment, and parental mosaicism can impact recurrence risk.^{56,110}
- **Mutational burden and modifiers** can modulate the phenotypic severity of the observed trait, and may explain intrafamilial phenotypic variability, as has been observed in peripheral neuropathy.¹¹¹ Similarly, an aggregation of rare variants has been shown to influence susceptibility to Parkinson disease,¹¹² and the age of onset of amyotrophic lateral sclerosis (ALS).¹¹³

- **Phenotypic expansion**²¹ is often observed with recently discovered disease genes, for which the full phenotypic spectrum of disease has not yet been appreciated. Multilocus variation can explain some cases of apparent phenotypic expansion,¹¹⁴ resulting in the observation of additional phenotypic features (multiple molecular diagnoses) or modifying the severity or characteristics of the primary observed phenotype (as multiple molecular diagnoses, or as modifiers).

Bioinformatic tool development

CMG investigators have developed tools for gene matching, data sharing, phenotype analysis, and exome variant data analysis (Table 2). Gene-matching tools connecting clinicians and human and model organism genetics investigators include GeneMatcher,^{50,51} MyGene2 (which includes a patient-facing portal for data sharing),⁵² and *matchbox* (Fig. 3). These tools each communicate through the MME (www.matchmakerexchange.org/), enabling gene and phenotype matching both within and across matching tools in the United States and internationally.⁴⁹ Members of the CMGs have essential roles in developing and maintaining the MME.^{48,49}

The CMGs have also developed software to record and compare phenotype data and analyze sequence data with the aim of identifying responsible genes and variants. These include PhenoDB^{50,115} and *seqr* (<https://seqr.broadinstitute.org>), which, in addition to recording phenotypic data in a standard structured ontology (e.g., Human Phenotype Ontology, HPO),¹¹⁶ also enable variant prioritization utilizing patterns of Mendelian inheritance, minor allele frequencies from reference population databases, and annotation of genes and variants by OMIM, ClinVar, and other resources (Table 2). ALoFT (annotation of loss-of-function transcripts) annotates and predicts putative disease-causing LoF pathogenic variants. It can further distinguish between disease-causing LoFs, which are heterozygous, compared with those in a homozygous state.¹¹⁷ Quantification of missense variant-induced local perturbation on a protein structure can identify putative disease-causing missense pathogenic variants.¹¹⁸ The localized frustration metric can identify variants that disrupt protein function without severely affecting the global stability of proteins.¹¹⁸ Additional analysis software developed by the CMGs is listed in Table 2.

Table 2 Bioinformatic analysis tools developed by the CMGs

Type of tool	Name	Description	URL
Phenotyping, genotyping, and sequence analysis	PhenoDB	Phenotype and genotype storage & analysis	https://phenodb.org
Variant analysis	seqr	Exome and genome analysis software for variant prioritization	https://seqr.broadinstitute.org/ https://github.com/macarthur-lab/seqr/
	HMZDelFinder	Homozygous & hemizygous CNV caller	https://github.com/BCM-Lupskilab/HMZDelFinder
	CoNIFER	Homozygous & hemizygous CNV caller	https://github.com/nkrumm/CoNIFER
	BafAOHcalculator	Absence of heterozygosity caller	https://github.com/BCM-Lupskilab/BafCalculator
	NMDescPredictor	Prediction of variants subject to NMD	https://nmdprediction.shinyapps.io/nmdescpredictor/
	NMDEscIntoleranceScore	Calculation of a gene's NMD escape intolerance score	https://nmdprediction.shinyapps.io/nmdescintolerancescore/
	DNM-Finder	De novo variant caller	https://github.com/BCM-Lupskilab/DNM-Finder
	AluAluCNVpredictor	<i>Alu/Alu</i> -mediated recombination predictor	https://github.com/BCM-Lupskilab/AAMR-Predictor
	OMIMExplorer	Variant prioritization based on phenotypic similarity	https://omimexplorer.research.bcm.edu/
	ALoFT	Annotation of putative loss-of-function variants	http://aloft.gersteinlab.org
Gene and phenotype matching and data sharing	GeneMatcher	Matching with researchers/clinicians with shared gene of interest	https://www.genematcher.org/
	Geno2MP	Bulk sharing of rare variants tied to broad phenotypic descriptions	http://geno2mp.gs.washington.edu
	MyGene2	Matching with patients in addition to researchers/clinicians with shared gene of interest	https://mygene2.org/
	matchbox	Matching with researchers/clinicians with shared gene of interest	https://github.com/macarthur-lab/matchbox

CMG Centers for Mendelian Genomics, CNV copy-number variant, NMD nonsense-mediated decay.

WHAT REMAINS TO BE DONE

Phenotypic annotation of variant effects in all ~20,000 human genes will provide the necessary evidence base to study the biologic relevance of each locus in the human genome. Some of the key challenges in meeting this long-term goal are elucidated below.

Disease gene discovery

Despite the progress of the CMGs, thousands of disease genes remain to be discovered. Currently, OMIM lists 3961 genes known to have high penetrance variants (~19% of the total annotated protein coding genes) underlying Mendelian conditions (www.OMIM.org; 4 October 2018). The early years of the CMGs saw rapid-paced gene discovery, including much of the “low-hanging fruit” available for study. Moving forward, the CMGs plan to explore new strategies for identification and engagement of families and clinicians worldwide and mainstreaming use of more complex, biologically based analysis strategies to identify novel genes for Mendelian conditions. Discoveries will stimulate new

biological questions about the relationship between rare variation and human Mendelian conditions, including the impact of different mutational mechanisms, the consequences of pathogenic variants on RNA and protein function, and the extent and consequences of mosaicism. Elucidation of the phenotypes resulting from allelic heterogeneity (LoF, GoF) and the impact of variation on protein function (LoF [amorph], partial LoF [hypomorph], increase in function [hypermorph], novel function [neomorph], and dominant negative [antemorph])¹¹⁹ are incompletely explored for most loci, underscoring the important role of allelic series. We will expand allelic series for newly identified disease genes and further explore gene-first approaches to develop more sophisticated genotype–phenotype relationships. This rapidly expanding application of genomic sequencing to identify novel genotype–phenotype relationships across the world's population will expand the utility of clinical genomics, enable precision medicine in all countries, and fuel the rapidly increasing trajectory of biological insights into perturbed homeostasis and disease.

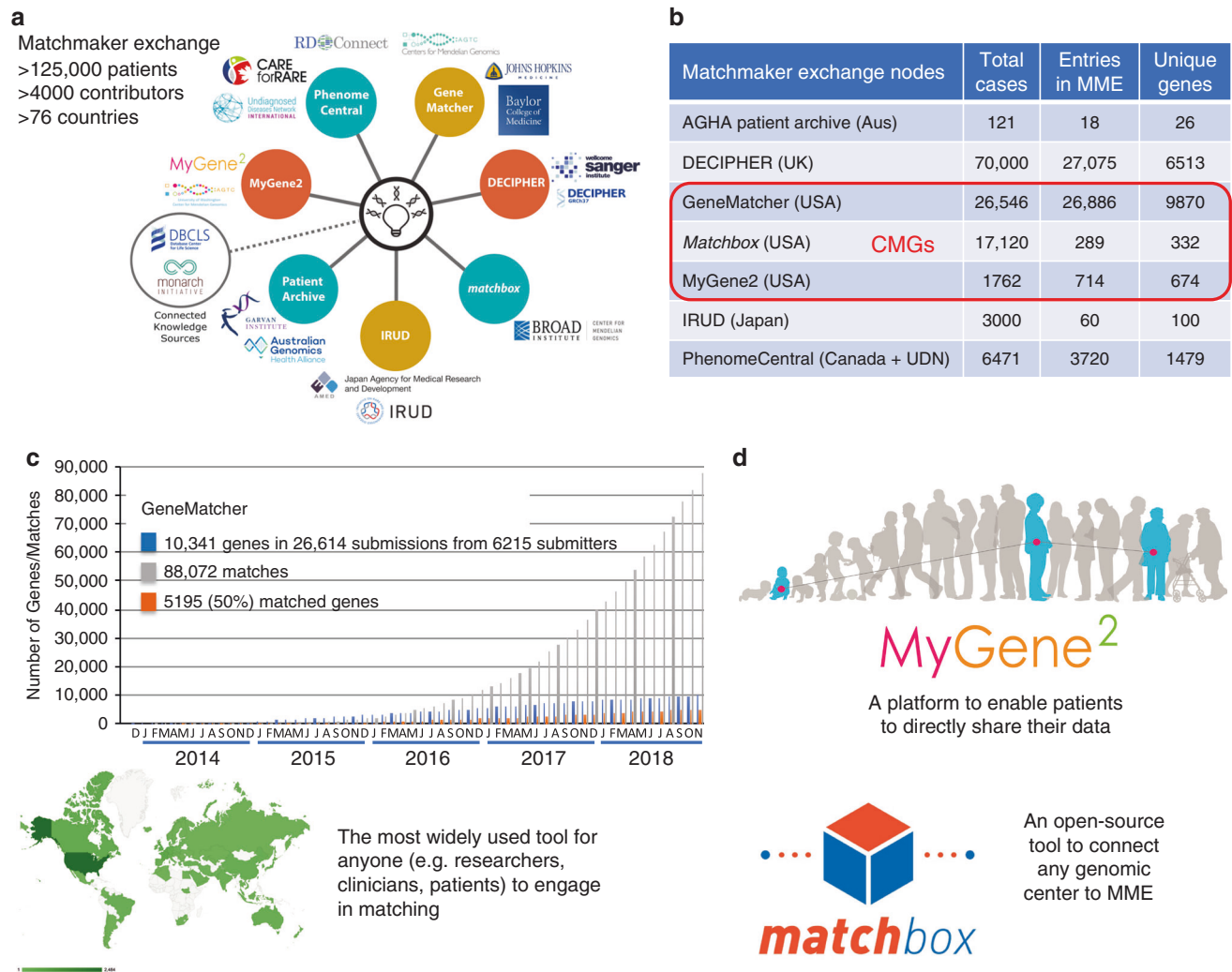


Fig. 3 Matchmaking tools developed through the Centers for Mendelian Genomics (CMGs). (a) The Matchmaker Exchange (MME) facilitates communication among multiple databases of human genomic and phenotypic data, each unique in focus and design. Each database functions as a node within the MME. (b) Total number of entries in each MME node, as well as total number of entries per node shared within the MME, are indicated. Also listed is the total number of unique genes per node. Note that a given unique gene may be present in more than one node. (c) Cumulative GeneMatcher statistics demonstrate 26,614 submissions of 10,341 genes through 1 November 2018. This has resulted in 5195 matched genes. GeneMatcher submitters in 77 countries today, demonstrating worldwide democratization of disease gene discovery. (d) MyGene2 is a database through which patients and families can directly share their genomic data. Matchbox is an open-source tool through which institutions or groups with genomic data can connect to the MME.

Methods and approach

Continued evaluation and integration of appropriate technological advances will provide the greatest likelihood of success in reaching our goals. To date, ES has been the predominant platform contributing to discovery of new disease genes, owing to its markedly lower cost compared with GS and enrichment of rare variants with large phenotypic effect in coding regions. In support of ES reanalysis to increase molecular diagnostic rates, one pilot study of ES reanalysis in 74 nondiagnostic clinical cases, with expansion to available relatives for trio- and multiplex-ES, led to the identification of a likely or potential molecular diagnosis in 51% (38/74) of previously unsolved clinical cases.⁴⁶

On contemporary capture platforms, sensitivity of detection of SNVs is similar between ES and GS, however GS enables

more sensitive structural variant calling, particularly for copy-neutral variation (e.g., inversions) and smaller sized (<10 Kb) CNVs. GS also promotes integration of CMG data with those generated by ENCODE, GTEx, and GENCODE. Caveats to the use of GS include limited annotation of noncoding variants and lower-fold coverage, which reduces power to detect low allele fraction mosaic variants and increases false positive de novo variant calls. Reports of improved coverage of coding regions by GS compared with ES are complicated by lack of an appropriately powered head-to-head study comparing contemporaneous versions of both technologies. Despite several large-scale investments in GS, the paucity of new disease gene discoveries reported from GS that would not have been made at much lower cost by ES challenges GS as a cost-effective strategy. Some studies suggest, however, that the

addition of RNAseq and/or GS identifies causative variants or adds functional support for the interpretation of variants discovered by ES. For example:

- **Muscle-related and mitochondrial phenotypes** for which highly penetrant noncoding variants and/or tissue-specific transcript-level changes can cause Mendelian conditions^{120–122}
- **Recessive conditions for which a single (i.e., mono-allelic) rare coding variant has been identified in a candidate gene**, increasing the likelihood of having a noncoding SNV or CNV impacting splicing on the second allele (in *trans*)^{123–125}

Further development of analytic methods and bioinformatics tools

Several variant types remain poorly (or at least not routinely) recognized by current variant-calling methods. The sensitivity for indel calling is suboptimal by standard currently utilized analytic tools (Atlas2, GATK). Analytic methods need to incorporate information about imprinted genomic areas, X-linked pseudoautosomal regions, and uniparental disomy.^{126,127} Structural variant identification also remains a challenge, particularly single-exon dropout alleles, small CNVs 50–1000 bp in size, mobile element insertions (MEIs), and copy number neutral structural variants (e.g., inversions and translocations), as well as trinucleotide repeat expansions. Improved methods design needs to consider family structure, modes of inheritance, and the contribution of rare, or even private variants to Mendelian conditions.¹²⁸ Methods such as Combined Annotation Dependent Depletion (CADD),^{129,130} which predict missense variant impact on protein function, are needed to support resolution of clinically reported variants of uncertain clinical significance.

We need a better understanding of the contribution of synonymous and noncoding variants to altered function through transfer RNA (tRNA) abundance and splicing effects.^{131–133} There is also a clear need for development of programs, such as NMDEscPredictor, to predict a variant's effect on NMD.⁸⁷ Simultaneous computational integration of rare and common variant analyses should be evaluated for enabling identification of conditions resulting from a combination of variants of these types (e.g., compound inheritance underlying 10% of congenital scoliosis).¹³⁴ Population-specific databases are increasingly helpful tools in identifying rare variants within a given population, and continued growth and diversification of these resources are needed. The CMGs have studied multiple non-European populations including large Turkish, Middle Eastern, and African cohorts of over 1000 individuals.¹³⁵

The challenge of non-Mendelian inheritance

The insights and discoveries described in previous sections define a shift that moves beyond the boundaries of one-disease-one-gene models. Mechanistically, more work is needed to explore the molecular basis of penetrance. Two

CMG-studied conditions illustrate compound inheritance of both rare and common variant models for incomplete penetrance involving a single or more than one locus. The observation of incomplete penetrance after identifying variants in *SMAD6* in nonsyndromic midline craniosynostosis cohort prompted the additional discovery of a common variant (minor allele frequency [MAF] 0.41) downstream of *BMP2*, which explained the incomplete penetrance.¹⁰⁴ In Han Chinese individuals with congenital scoliosis, incomplete penetrance observed in relatives sharing a 16p11.2 deletion or *TBX6* LoF variant led to the discovery of a common *TBX6* hypomorphic allele (MAF 0.44 in Chinese population, 0.33 in Caucasians, <0.01 in individuals of African descent) in *trans* with the rare *TBX6* null allele (MAF of 16p11.2 deletion is 0.0003 worldwide).¹³⁴ Individuals with biallelic LoF + hypomorphic *TBX6* variants have a distinct *TBX6*-associated congenital scoliosis (TACS) phenotype characterized by hemivertebrae and/or butterfly vertebrae involving the lower spine.¹³⁶ Mouse models of biallelic LoF + hypomorphic *TBX6* alleles demonstrated reduced *Tbx6* expression from hypomorphic alleles, leading to a vertebral malformation phenotype;¹³⁷ homozygosity for the null allele leads to distortion of Mendelian ratios through embryonic lethality. These studies demonstrate that a rare null and a noncoding common hypomorphic allele can influence gene dosage and expression at a locus and thereby impact phenotypic expression of human disease traits.

We need to understand the contributions to penetrance of variation in environmental exposures,^{138,139} variation at various modifier loci, and epigenetic effects. From a computational perspective, we also need a more refined definition of LoF, with distinction between nonsense and frameshifting variants that are likely to escape—or be subject to—NMD. This distinction will be increasingly important for understanding the pathogenesis underlying variants that result in premature translation termination, a class of variants for which premature truncation readthrough-based therapeutics may become available.¹⁴⁰

Bridging the gap between rare and common variation and disease

The genetic architecture of rare and common disease is often conceptualized as a continuum based primarily on the frequencies of the relevant variant alleles: rare variant alleles causative for Mendelian conditions and common alleles contributing risk for common disease phenotypes.¹⁴¹ Rare diseases, defined in the United States as any condition affecting fewer than 200,000 individuals, typically have etiologic or causal variants of large effect and a population frequency of far less than 0.1%. Common diseases of adult life often have a mixed genetic and environmental etiology, with susceptibility variants that are more common (>1%) and have markedly smaller effect sizes.¹⁴²

Discoveries in Mendelian conditions have refined our understanding of the genetic architecture of common disease, with contributions of both rare and common variants to

common disease.^{14,111,143–148} An analysis of genes associated with rare disease revealed that almost 20% were nearest to, or contain, a variant that had been associated with common disease.²¹ Moreover, rare de novo SNVs with large phenotypic effect contribute to common childhood traits including neurodevelopmental and congenital heart conditions.^{14–16,149–156} Abnormalities of gene dosage mediated by rare CNVs have further been recognized as an etiology of both Mendelian conditions and risk for common diseases such as neuropathy, dementia, depression, bipolar disease, schizophrenia, autism, and intellectual disability.^{53,151,157–166}

The ongoing exploration of Mendelian conditions by the CMGs and others has increased appreciation for the extent of allelic and locus heterogeneity, variability of expression and penetrance, the role of new mutation, and mosaicism in disease and the phenotypic complexity that can arise from combinatorial effects of rare alleles at a locus (biallelic versus monoallelic) or at different loci (i.e., multilocus pathogenic variation)—characteristics shared by both rare and common disorders. We explore these concepts using four examples, and discuss the impact of genomics informed by pedigree structure and mode of inheritance on the human genetics field's understanding of the architecture of common disease:

1. Rare variation may present phenotypically as a common disease, obscuring recognition of a distinct monogenic disorder. Monogenic forms of steroid-resistant nephrotic syndrome due to rare variation in *NUP93*, *NUP205*, *XPO5*, and *FAT1* illustrate this concept: these monogenic conditions implicated a role for BMP7-induced SMAD signaling and Rho-like small GTPase signaling pathways in defective podocyte migration, providing therapeutic targets for drug development.^{143,144} A recent analysis of electronic health records for correlations between phenotypes overlapping with a Mendelian condition using a phenotypic risk score (PheRS) and genotype data in individuals with presumed common disease revealed 18 previously unrecognized Mendelian diagnoses.¹⁴⁶
2. Rare variants causing dominant traits may present as a phenotypically milder common trait, such as the dominant carpal tunnel syndrome that may be observed in *PMP22* deletion heterozygotes, who typically are expected to develop hereditary neuropathy with liability to pressure palsies (HNPP [MIM 162500]) (refs.^{167,168}). Allelic series have elucidated loci harboring rare, highly penetrant variants leading to Mendelian conditions, and more common variants contributing risk for common disease; for example, rare and common variants in *SNCA*, including duplication and triplication CNV of the locus, have been described in association with familial and sporadic forms of Parkinson disease, respectively.^{169–173}
3. Several genes identified because of their association with recessive Mendelian conditions were later discovered to contribute to risk for common complex disease in heterozygous individuals, representing an expansion of

the originally defined phenotypes (Supplemental Table 7). Notably, heterozygosity for alleles that cause severe recessive disease may be associated with reduced risk for common disease. For example, heterozygosity for pathogenic alleles in *SLC12A1*, *KCNJ1*, and *SLC12A3*, associated with Bartter and Gitelman syndromes, reduces blood pressure and protects against adult-onset hypertension.¹⁴⁷ Population cohorts with a high rate of consanguinity and carrier frequency for recessive conditions represent an opportunity to analyze phenotypic effects of heterozygous LoF.^{148,174}

4. Multilocus mutational burden can impact expression of common disease. Genomic studies of neuropathy and Parkinson disease have suggested a model in which an aggregation of rare variants in disease-associated genes can influence clinical severity and can contribute to common complex traits.^{111,112}

These discoveries at the intersection of rare and common disease will facilitate further development of precision medicine through elucidation of targetable pathways underlying disease.

Data sharing

A worldwide effort to share individual-level exome variant and phenotype databases can be highly beneficial for rare disease research as well as other genetic studies. CMG data are deposited to dbGaP and ClinVar. Additionally, access to the Broad CMG data can be applied for through the Broad's Data Use Oversight System (DUOS). DUOS (<https://duos.broadinstitute.org/#/home>) is a novel framework for automating the data use oversight process that is overseen by a Data Access Committee. DUOS provides de-identified genotype and phenotype data to authorized researchers in a substantially more usable fashion than the currently cumbersome dbGaP platform. De-identified rare variants tied to broad phenotype data for all cases sequenced by the UWCMG are shared publicly through Geno2MP (<http://geno2mp.gs.washington.edu>) and deposited in MyGene2. Variant data for candidate genes can be directly requested from both BHCMB and Baylor Genetics clinical diagnostic laboratories. Submission of candidate genes to the MME further fosters global involvement in discovery. Continued integration of data with patient-facing portals, for example MyGene2 (ref.⁵²), may facilitate further engagement of stakeholders to support patient involvement in research studies. The additional development of publicly accessible online tools for direct interrogation of exome data would be useful to further patient and physician engagement.

Integration with other genome sequencing programs

Partnership with the Centers for Common Disease Genomics (CCDGs) will continue to be an important strategy for the CMGs as rare disease discoveries are likely to increasingly impact common disease discoveries and both programs implement genome-wide approaches. Collaborations with

the CCDGs have already been instrumental in development, improvement, and implementation of sequencing methods and variant data processing, annotation, and analysis pipelines (Farek et al., <https://github.com/jfarek/xatlas/blob/master/README.md>).^{175,176} The genomics community and CMGs in particular have benefitted tremendously from the development of the ExAC and gnomAD databases as well as the ARIC database.^{87,129,177} The ARIC database is represented by a more general population and not a disease cohort. Using these former resources, the study of constrained genes that show fewer than expected LoF or missense variants in general populations has refined prioritization of candidate disease genes. Likewise, GoF variant allele prioritization has been assisted by the ARIC database and NMDEscPredictor.⁸⁷

A recent formalized collaboration between the CMGs and the Knock-Out Mouse Project (KOMP, <https://www.komp.org>) centers promotes rapid sharing of CMG discovery gene lists with the KOMP centers. Mutant mouse strains generated through these collaborations will be available to researchers worldwide. Similarly, an enhanced interface with the UDN clinical and model organism screening centers (MOSCs) should enable in-depth characterization of allelic series for disease genes.¹⁷⁸ As many as 45% of *Drosophila* genes important for neurodevelopment have a human disease ortholog, and *Drosophila* genes with more than one human ortholog are enriched eightfold for human disease genes.^{178,179} One such gene, *ANKLE2*, had been identified as a CMG tier 2 gene in a family with severe microcephaly; recent studies implicate *ANKLE2* as a target of the Zika virus.¹⁸⁰ Collaborative efforts have included the UDN,^{32,78,79,179–183} the DDD,^{59,72,184} and the UK10K Project.⁸⁰ Expansion of such collaborations to similar clinically oriented discovery programs, such as the Gabriella Miller Kids First (GMKF) program, and deeper integration with international programs like FORGE Canada Consortium,¹¹ DECIPHER (<https://decipher.sanger.ac.uk>), Care4Rare Canada Consortium (<http://care4rare.ca>) and rare disease programs affiliated with IRDiRC (<http://www.irdirc.org>) and in Asia^{134,136} could further foster international collaboration and stakeholder impact for CMG discovery.

Integration with clinical testing programs

With the goal of rapidly translating novel CMG discoveries to maximally impact patient care, expanded partnerships with clinical diagnostic laboratories and engaging clinicians worldwide will be important. Clinicians provide perhaps the most important role in discovery and are truly at the forefront of efforts to engage in detailed phenotyping. Collaboration with diagnostic laboratories provides several advantages: (1) availability of thousands of cases for which ES has been nondiagnostic, maximizing the likelihood of novel disease gene discovery; (2) potential for enrollment of individuals with pre-existing exome data into research; (3) contact with referring physicians, allowing access to phenotypic information and the opportunity for clinical reassessment; (4) collaborative research; and (5) clinical (College of American

Pathologists [CAP], CLIA-accredited) reporting of novel discoveries to the referring physician, facilitating rapid dissemination of information from bench to bedside.

CONCLUSION

CMG-facilitated collaborative research efforts have provided clear deliverables including, most notably, >1000 new disease genes and >500 peer-reviewed publications. However, much work remains. Extending gene discoveries to the interrogation of LoF, GoF, and dominant negative variants on disease expression, and modeling allelic series in mice, underscores the need for analysis of multiple variants per gene. The relationship between rare and common disease is real but complex and can include the intersection of both rare variant and common variant alleles at one or more loci. The extent to which multilocus pathogenic variation contributes to blended phenotypes, phenotypic severity, and phenotypic expansion remains to be explored.

As we, the CMG and world collaborators, investigate personal genome variation in the context of an individual's phenotype, computational methods for analysis of observed clinical phenotypes using structured phenotypic ontologies¹⁰⁵ will enable the field to fully explore genotype–phenotype relationships and to potentially achieve individualized care. Expansion of recruitment efforts to understudied countries, ethnicities, and phenotypes will further expand disease gene discovery and improve clinical utility. Continued development of sequencing technology and bioinformatic tools for genomic data analysis will also increase the effectiveness and efficiency of the CMG collaborative efforts. Finally, and perhaps most importantly, increased integration with clinical genomics,² extending the reach of the research laboratories and enabling novel discoveries that benefit patient care as expeditiously as possible, is essential for realizing the maximal benefit to world populations.

SUPPLEMENTARY INFORMATION

The online version of this article (<https://doi.org/10.1038/s41436-018-0408-7>) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

The Baylor Hopkins Center for Mendelian Genomics, Broad Institute Harvard Center for Mendelian Genomics, University of Washington Center for Mendelian Genomics, and Yale Center for Mendelian Genomics were funded by the National Human Genome Research Institute (NHGRI) awards UM1 HG006542, UM1 HG008900, UM1 HG006493, and UM1 HG006504, respectively. Funds were also provided under the National Heart, Lung, and Blood Institute (NHLBI) under the Trans-Omics for Precision Medicine Program (TOPMed), and the National Eye Institute (NEI). The GSP Coordinating Center (U24 HG008956) contributed to cross-program scientific initiatives and provided logistical and general study coordination. Funds were also provided under the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD R03 HD092569)

to C.M.B.C. and V.R.S. J.E.P. was supported by NHGRI K08 HG008986. A.H.O.-L. was supported by National Institute of Child Health and Human Development (NICHD) K12 HD052896. N.W. was supported by National Natural Science Foundation of China (81501852), Beijing Natural Science Foundation (7172175), and 2016 Milstein Medical Asian American Partnership Foundation Fellowship Award in Translational Medicine. This work was also supported by the National Institute of Neurological Disorders and Stroke (NINDS) R35 NS105078 to J.R.L.

DISCLOSURE

Baylor College of Medicine (BCM) and Miraca Holdings Inc. have formed a joint venture with shared ownership and governance of Baylor Genetics (BG), formerly the Baylor Miraca Genetics Laboratories (BMGL), which performs clinical exome sequencing and chromosomal microarray analysis for genome-wide detection of CNV. J.R.L. serves on the Scientific Advisory Board of BG. J.R.L. has stock ownership in 23andMe, is a paid consultant for Regeneron Pharmaceuticals, and is a coinventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases and bacterial genomic fingerprinting. The other authors declare no conflicts of interest.

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

REFERENCES

- Baird PA, Anderson TW, Newcombe HB, Lowry RB. Genetic disorders in children and young adults: a population study. *Am J Hum Genet*. 1988;42:677–693.
- Lupski JR. Clinical genomics: from a truly personal genome viewpoint. *Hum Genet*. 2016;135:591–601.
- Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008;452:872–876.
- Levy S, Sutton G, Ng PC, et al. The diploid genome sequence of an individual human. *PLoS Biol*. 2007;5:e254.
- Albert TJ, Molla MN, Muzny DM, et al. Direct selection of human genomic loci by microarray hybridization. *Nat Methods*. 2007;4:903–905.
- Bainbridge MN, Wang M, Burgess DL, et al. Whole exome capture in solution with 3 Gbp of data. *Genome Biol*. 2010;11:R62.
- Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. *Annu Rev Med*. 2012;63:35–61.
- Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009;461:272–276.
- Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA*. 2009;106:19096–19101.
- Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*. 2010;42:30–35.
- Beaulieu CL, Majewski J, Schwartzentruber J, et al. FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. *Am J Hum Genet*. 2014;94:809–817.
- Sawyer SL, Hartley T, Dymont DA, et al. Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clin Genet*. 2016;89:275–284.
- Firth HV, Wright CF, Study DDD. The Deciphering Developmental Disorders (DDD) study. *Dev Med Child Neurol*. 2011;53:702–703.
- Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature*. 2015;519:223–228.
- Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature*. 2017;542:433–438.
- Wright CF, Fitzgerald TW, Jones WD, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet*. 2015;385:1305–1314.
- Wright CF, McRae JF, Clayton S, et al. Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genet Med*. 2018;20:1216–1223.
- Taruscio D, Groft SC, Cederroth H, et al. Undiagnosed Diseases Network International (UDNI): white paper for global actions to meet patient needs. *Mol Genet Metab*. 2015;116:223–225.
- Gasser SM, Lupski JR, Le Cam Y, Menzel O. Leap year: Rare day to highlight rare diseases. *Nature*. 2012;481:265.
- Bamshad MJ, Shendure JA, Valle D, et al. The Centers for Mendelian Genomics: a new large-scale initiative to identify the genes underlying rare Mendelian conditions. *Am J Med Genet A*. 2012;158A:1523–1525.
- Chong JX, Buckingham KJ, Jhangiani SN, et al. The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *Am J Hum Genet*. 2015;97:199–215.
- Gahl WA, Mulvihill JJ, Toro C, et al. The NIH Undiagnosed Diseases Program and Network: applications to modern medicine. *Mol Genet Metab*. 2016;117:393–400.
- White J, Beck CR, Harel T, et al. *POGZ* truncating alleles cause syndromic intellectual disability. *Genome Med*. 2016;8:3.
- Stessman HAF, Willemsen MH, Fencikova M, et al. Disruption of *POGZ* is associated with intellectual disability and autism spectrum disorders. *Am J Hum Genet*. 2016;98:541–552.
- Iossifov I, O'Roak BJ, Sanders SJ, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*. 2014;515:216–221.
- Iossifov I, Ronemus M, Levy D, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron*. 2012;74:285–299.
- Gilissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature*. 2014;511:344–347.
- Neale BM, Kou Y, Liu L, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*. 2012;485:242–245.
- Fromer M, Pocklington AJ, Kavanagh DH, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature*. 2014;506:179–184.
- Bainbridge MN, Hu H, Muzny DM, et al. De novo truncating mutations in *ASXL3* are associated with a novel clinical phenotype with similarities to Bohring-Opitz syndrome. *Genome Med*. 2013;5:11.
- Russell B, Graham JM Jr. Expanding our knowledge of conditions associated with the *ASXL* gene family. *Genome Med*. 2013;5:16.
- Bostwick BL, McLean S, Posey JE, et al. Phenotypic and molecular characterisation of *CDK13*-related congenital heart defects, dysmorphic facial features and intellectual developmental disorders. *Genome Med*. 2017;9:73.
- Boyden LM, Craiglow BG, Hu RH, et al. Phenotypic spectrum of autosomal recessive congenital ichthyosis due to *PNPLA1* mutation. *Br J Dermatol*. 2017;177:319–322.
- Spier I, Holzapfel S, Altmüller J, et al. Frequency and phenotypic spectrum of germline mutations in *POLE* and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. *Int J Cancer*. 2015;137:320–331.
- Duran I, Taylor SP, Zhang W, et al. Destabilization of the IFT-B cilia core complex due to mutations in *IFT81* causes a spectrum of short-rib polydactyly syndrome. *Sci Rep*. 2016;6:34232.
- Kaiser FJ, Ansari M, Braunholz D, et al. Loss-of-function *HDAC8* mutations cause a phenotypic spectrum of Cornelia de Lange syndrome-like features, ocular hypertelorism, large fontanelle and X-linked inheritance. *Hum Mol Genet*. 2014;23:2888–2900.
- Jiang Y, Wangler MF, McGuire AL, et al. The phenotypic spectrum of Xia-Gibbs syndrome. *Am J Med Genet A*. 2018;176:1315–1326.
- Martinelli S, Krumbach OHF, Pantaleoni F, et al. Functional dysregulation of *CDC42* causes diverse developmental phenotypes. *Am J Hum Genet*. 2018;102:309–320.
- Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in *PCSK9*. *Nat Genet*. 2005;37:161–165.
- Kotowski IK, Pertsemlidis A, Luke A, et al. A spectrum of *PCSK9* alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am J Hum Genet*. 2006;78:410–422.

41. Johansen CT, Hegele RA. Using Mendelian randomization to determine causative factors in cardiovascular disease. *J Intern Med*. 2013;273:44–47.
42. Biesecker LG. Exome sequencing makes medical genomics a reality. *Nat Genet*. 2010;42:13–14.
43. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*. 2011;12:745–755.
44. Eldomery MK, Coban-Akdemir Z, Harel T, et al. Lessons learned from additional research analyses of unsolved clinical exome cases. *Genome Med*. 2017;9:26.
45. Stray-Pedersen A, Backe PH, Sorte HS, et al. *PGM3* mutations cause a congenital disorder of glycosylation with severe immunodeficiency and skeletal dysplasia. *Am J Hum Genet*. 2014;95:96–107.
46. Lalani SR, Liu P, Rosenfeld JA, et al. Recurrent muscle weakness with rhabdomyolysis, metabolic crises, and cardiac arrhythmia due to bi-allelic *TANGO2* mutations. *Am J Hum Genet*. 2016;98:347–357.
47. Wang X, Charnig WL, Chen CA, et al. Germline mutations in *ABL1* cause an autosomal dominant syndrome characterized by congenital heart defects and skeletal malformations. *Nat Genet*. 2017;49:613–617.
48. Philippakis AA, Azzariti DR, Beltran S, et al. The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum Mutat*. 2015;36:915–921.
49. Sobreira NLM, Arachchi H, Buske OJ, et al. Matchmaker Exchange. *Curr Protoc Hum Genet*. 2017;95:9.31. 31–39.31.15.
50. Sobreira N, Schiettecatte F, Boehm C, Valle D, Hamosh A. New tools for Mendelian disease gene identification: PhenoDB variant analysis module; and GeneMatcher, a web-based tool for linking investigators with an interest in the same gene. *Hum Mutat*. 2015;36:425–431.
51. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat*. 2015;36:928–930.
52. The AJMG Sequence: Decoding news and trends for the medical genetics community. Website aims to accelerate gene discovery, diagnosis, treatment: MyGene2.org fosters open sharing among families, researchers, and clinicians. *Am J Med Genet A*. 2016;170:1388–1389.
53. Lupski JR. Genomic rearrangements and sporadic disease. *Nat Genet*. 2007;39 7 suppl:S43–47.
54. Acuna-Hidalgo R, Veltman JA, Hoischen A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol*. 2016;17:241.
55. Veltman JA, Brunner HG. De novo mutations in human genetic disease. *Nat Rev Genet*. 2012;13:565–575.
56. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*. 2014;312:1870–1879.
57. Posey JE, Rosenfeld JA, James RA, et al. Molecular diagnostic experience of whole-exome sequencing in adult patients. *Genet Med*. 2016;18:678–685.
58. Chong JX, McMillin MJ, Shively KM, et al. De novo mutations in *NALCN* cause a syndrome characterized by congenital contractures of the limbs and face, hypotonia, and developmental delay. *Am J Hum Genet*. 2015;96:462–473.
59. Lessel D, Schob C, Kury S, et al. De novo missense mutations in *DHX30* impair global translation and cause a neurodevelopmental disorder. *Am J Hum Genet*. 2017;101:716–724.
60. Burrage LC, Charnig WL, Eldomery MK, et al. De novo *GMNN* mutations cause autosomal-dominant primordial dwarfism associated with Meier-Gorlin syndrome. *Am J Hum Genet*. 2015;97:904–913.
61. Wangler MF, Gonzaga-Jauregui C, Gambin T, et al. Heterozygous de novo and inherited mutations in the smooth muscle actin (*ACTG2*) gene underlie megacystis-microcolon-intestinal hypoperistalsis syndrome. *PLoS Genet*. 2014;10:e1004258.
62. Chacon-Camacho OF, Sobreira N, You J, Pina-Aguilar RE, Villegas-Ruiz V, Zenteno JC. Exome sequencing identifies a de novo frameshift mutation in the imprinted gene *ZDBF2* in a sporadic patient with nasopalpebral lipoma-coloboma syndrome. *Am J Med Genet A*. 2016;170:1934–1937.
63. Lim YH, Ovejero D, Derrick KM, Yale Center for Mendelian Genetics, Collins MT, Choate KA. Cutaneous skeletal hypophosphatemia syndrome (CSHS) is a multilineage somatic mosaic RASopathy. *J Am Acad Dermatol*. 2016;75:420–427.
64. Lim YH, Ovejero D, Sugarman JS, et al. Multilineage somatic activating mutations in *HRAS* and *NRAS* cause mosaic cutaneous and skeletal lesions, elevated FGF23 and hypophosphatemia. *Hum Mol Genet*. 2014;23:397–407.
65. Lim YH, Bacchiocchi A, Qiu J, et al. *GNA14* somatic mutation causes congenital and sporadic vascular tumors by MAPK activation. *Am J Hum Genet*. 2016;99:443–450.
66. Campbell IM, Yuan B, Robberecht C, et al. Parental somatic mosaicism is underrecognized and influences recurrence risk of genomic disorders. *Am J Hum Genet*. 2014;95:173–182.
67. Campbell IM, Stewart JR, James RA, et al. Parent of origin, mosaicism, and recurrence risk: probabilistic modeling explains the broken symmetry of transmission genetics. *Am J Hum Genet*. 2014;95:345–359.
68. Campbell IM, Shaw CA, Stankiewicz P, Lupski JR. Somatic mosaicism: implications for disease and transmission genetics. *Trends Genet*. 2015;31:382–392.
69. Ansari M, Poke G, Ferry Q, et al. Genetic heterogeneity in Cornelia de Lange syndrome (CdLS) and CdLS-like phenotypes with observed and predicted levels of mosaicism. *J Med Genet*. 2014;51:659–668.
70. Cheng YW, Tan CA, Minor A, et al. Copy number analysis of *NIPBL* in a cohort of 510 patients reveals rare copy number variants and a mosaic deletion. *Mol Genet Genomic Med*. 2014;2:115–123.
71. Huisman SA, Redeker EJ, Maas SM, Mannens MM, Hennekam RC. High rate of mosaicism in individuals with Cornelia de Lange syndrome. *J Med Genet*. 2013;50:339–344.
72. Yuan B, Neira J, Pehlivan D, et al. Clinical exome sequencing reveals locus heterogeneity and phenotypic variability of cohesinopathies. *Genet Med*. 2018 Aug 30; doi:10.1038/s41436-018-0085-6 [Epub ahead of print].
73. Choate KA, Lu Y, Zhou J, et al. Mitotic recombination in patients with ichthyosis causes reversion of dominant mutations in *KRT10*. *Science*. 2010;330:94–97.
74. Choate KA, Lu Y, Zhou J, et al. Frequent somatic reversion of *KRT1* mutations in ichthyosis with confetti. *J Clin Invest*. 2015;125:1703–1707.
75. Stray-Pedersen A, Sorte HS, Samarakoon P, et al. Primary immunodeficiency diseases: genomic approaches delineate heterogeneous Mendelian disorders. *J Allergy Clin Immunol*. 2017;139:232–245.
76. Gambin T, Yuan B, Bi W, et al. Identification of novel candidate disease genes from de novo exonic copy number variants. *Genome Med*. 2017;9:83.
77. Boone PM, Bacino CA, Shaw CA, et al. Detection of clinically relevant exonic copy-number changes by array CGH. *Hum Mutat*. 2010;31:1326–1342.
78. Harel T, Yesil G, Bayram Y, et al. Monoallelic and biallelic variants in *EMC1* identified in individuals with global developmental delay, hypotonia, scoliosis, and cerebellar atrophy. *Am J Hum Genet*. 2016;98:562–570.
79. Harel T, Yoon WH, Garone C, et al. Recurrent de novo and biallelic variation of *ATAD3A*, encoding a mitochondrial membrane protein, results in distinct neurological syndromes. *Am J Hum Genet*. 2016;99:831–845.
80. Rainger J, Pehlivan D, Johansson S, et al. Monoallelic and biallelic mutations in *MAB21L2* cause a spectrum of major eye malformations. *Am J Hum Genet*. 2014;94:915–923.
81. Koch MC, Steinmeyer K, Lorenz C, et al. The skeletal muscle chloride channel in dominant and recessive human myotonia. *Science*. 1992;257:797–800.
82. George AL Jr, Crackower MA, Abdalla JA, Hudson AJ, Ebers GC. Molecular basis of Thomsen's disease (autosomal dominant myotonia congenita). *Nat Genet*. 1993;3:305–310.
83. Shaw ND, Brand H, Kupchinsky ZA, et al. *SMCHD1* mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. *Nat Genet*. 2017;49:238–248.
84. Bayram Y, White JJ, Elcioglu N, et al. *REST* final-exon-truncating mutations cause hereditary gingival fibromatosis. *Am J Hum Genet*. 2017;101:149–156.
85. Khajavi M, Inoue K, Lupski JR. Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. *Eur J Hum Genet*. 2006;14:1074–1081.
86. Poli MC, Ebstein F, Nicholas SK, et al. Heterozygous truncating variants in *POMP* escape nonsense-mediated decay and cause a unique immune dysregulatory syndrome. *Am J Hum Genet*. 2018;102:1126–1142.
87. Coban-Akdemir Z, White JJ, Song X, et al. Identifying genes whose mutant transcripts cause dominant disease traits by potential gain-of-function alleles. *Am J Hum Genet*. 2018;103:171–187.
88. White JJ, Mazzeu JF, Hoischen A, et al. *DVL3* alleles resulting in a -1 frameshift of the last exon mediate autosomal-dominant Robinowsky syndrome. *Am J Hum Genet*. 2016;98:553–561.

89. White JJ, Mazzeu JF, Coban-Akdemir Z, et al. WNT signaling perturbations underlie the genetic heterogeneity of Robinow syndrome. *Am J Hum Genet.* 2018;102:27–43.
90. White J, Mazzeu JF, Hoischen A, et al. *DVL1* frameshift mutations clustering in the penultimate exon cause autosomal-dominant Robinow syndrome. *Am J Hum Genet.* 2015;96:612–622.
91. Yamamoto GL, Aguen M, Gos M, et al. Rare variants in *SOS2* and *LZTR1* are associated with Noonan syndrome. *J Med Genet.* 2015;52:413–421.
92. Cirstea IC, Kutsche K, Dvorsky R, et al. A restricted spectrum of *NRAS* mutations causes Noonan syndrome. *Nat Genet.* 2010;42:27–29.
93. Pandit B, Sarkozy A, Pennacchio LA, et al. Gain-of-function *RAF1* mutations cause Noonan and LEOPARD syndromes with hypertrophic cardiomyopathy. *Nat Genet.* 2007;39:1007–1012.
94. Tartaglia M, Mehler EL, Goldberg R, et al. Mutations in *PTPN11*, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nat Genet.* 2001;29:465–468.
95. Tartaglia M, Pennacchio LA, Zhao C, et al. Gain-of-function *SOS1* mutations cause a distinctive form of Noonan syndrome. *Nat Genet.* 2007;39:75–79.
96. Roberts AE, Araki T, Swanson KD, et al. Germline gain-of-function mutations in *SOS1* cause Noonan syndrome. *Nat Genet.* 2007;39:70–74.
97. Aoki Y, Niihori T, Banjo T, et al. Gain-of-function mutations in *RIT1* cause Noonan syndrome, a RAS/MAPK pathway syndrome. *Am J Hum Genet.* 2013;93:173–180.
98. Niihori T, Aoki Y, Narumi Y, et al. Germline *KRAS* and *BRAF* mutations in cardio-facio-cutaneous syndrome. *Nat Genet.* 2006;38:294–296.
99. Razzaque MA, Nishizawa T, Komoike Y, et al. Germline gain-of-function mutations in *RAF1* cause Noonan syndrome. *Nat Genet.* 2007;39:1013–1017.
100. Schubert S, Zenker M, Rowe SL, et al. Germline *KRAS* mutations cause Noonan syndrome. *Nat Genet.* 2006;38:331–336.
101. Yuan B, Pehlivan D, Karaca E, et al. Global transcriptional disturbances underlie Cornelia de Lange syndrome and related phenotypes. *J Clin Invest.* 2015;125:636–651.
102. Lemmers RJ, Tawil R, Petek LM, et al. Digenic inheritance of an *SMCHD1* mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat Genet.* 2012;44:1370–1374.
103. Lupski JR. Digenic inheritance and Mendelian disease. *Nat Genet.* 2012;44:1291–1292.
104. Timberlake AT, Choi J, Zaidi S, et al. Two locus inheritance of non-syndromic midline craniosynostosis via rare *SMAD6* and common *BMP2* alleles. *Elife.* 2016;5:e20125.
105. Posey JE, Harel T, Liu P, et al. Resolution of disease phenotypes resulting from multilocus genomic variation. *N Engl J Med.* 2017;376:21–31.
106. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med.* 2013;369:1502–1511.
107. Farwell KD, Shahmirzadi L, El-Khechen D, et al. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet Med.* 2015;17:578–586.
108. Retterer K, Juusola J, Cho MT, et al. Clinical application of whole-exome sequencing across clinical indications. *Genet Med.* 2016;18:696–704.
109. Balci TB, Hartley T, Xi Y, et al. Debunking Occam's razor: diagnosing multiple genetic diseases in families by whole-exome sequencing. *Clin Genet.* 2017;92:281–289.
110. Jehue FS, de Oliveira VT, Gurgel-Giannetti J, et al. Dual molecular diagnosis contributes to atypical Prader-Willi phenotype in monozygotic twins. *Am J Med Genet A.* 2017;173:2451–2455.
111. Gonzaga-Jauregui C, Harel T, Gambin T, et al. Exome sequence analysis suggests that genetic burden contributes to phenotypic variability and complex neuropathy. *Cell Rep.* 2015;12:1169–1183.
112. Robak LA, Jansen IE, van Rooij J, et al. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. *Brain.* 2017;140:3191–3203.
113. Cady J, Allred P, Bali T, et al. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. *Ann Neurol.* 2015;77:100–113.
114. Karaca E, Posey JE, Coban Akdemir Z, et al. Phenotypic expansion illuminates multilocus pathogenic variation. *Genet Med.* 2018 Apr 26; doi:10.1038/gim.2018.33 [Epub ahead of print].
115. Hamosh A, Sobreira N, Hoover-Fong J, et al. PhenoDB: a new web-based tool for the collection, storage, and analysis of phenotypic features. *Hum Mutat.* 2013;34:566–571.
116. Kohler S, Doelken SC, Mungall CJ, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* 2014;42 database issue:D966–974.
117. Balasubramanian S, Fu Y, Pawashe M, et al. Using ALOFT to determine the impact of putative loss-of-function variants in protein-coding genes. *Nat Commun.* 2017;8:382.
118. Kumar S, Clarke D, Gerstein M. Localized structural frustration for evaluating the impact of sequence variants. *Nucleic Acids Res.* 2016;44:10062–10073.
119. Muller HJ. Further studies on the nature and causes of gene mutations. *Proc 6th Int Congr Genet.* 1932;1:213–255.
120. Smedley D, Schubach M, Jacobsen JOB, et al. A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. *Am J Hum Genet.* 2016;99:595–606.
121. Cummings BB, Marshall JL, Tukiainen T, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med.* 2017;9:386.
122. Kremer LS, Bader DM, Mertes C, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun.* 2017;8:15824.
123. Poncz M, Ballantine M, Solowiejczyk D, Barak I, Schwartz E, Surrey S. beta-Thalassemia in a Kurdish Jew. Single base changes in the T-A-T-A box. *J Biol Chem.* 1982;257:5994–5996.
124. Ludlow LB, Schick BP, Budarf ML, et al. Identification of a mutation in a GATA binding site of the platelet glycoprotein Ibbeta promoter resulting in the Bernard-Soulier syndrome. *J Biol Chem.* 1996;271:22076–22080.
125. Giardine B, van Baal S, Kaimakis P, et al. HbVar database of human hemoglobin variants and thalassemia mutations: 2007 update. *Hum Mutat.* 2007;28:206.
126. King DA, Fitzgerald TW, Miller R, et al. A novel method for detecting uniparental disomy from trio genotypes identifies a significant excess in children with developmental disorders. *Genome Res.* 2014;24:673–687.
127. Sobreira N. Novel analytic approaches used to solve unsolved whole exome sequencing data. Paper presented at: 67th Annual Meeting of the American Society of Human Genetics; October 23, 2017; Orlando, FL.
128. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. *Cell.* 2011;147:32–43.
129. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536:285–291.
130. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46:310–315.
131. Gelfman S, Wang Q, McSweeney KM, et al. Annotating pathogenic non-coding variants in genic regions. *Nat Commun.* 2017;8:236.
132. Lee M, Roos P, Sharma N, et al. Systematic computational identification of variants that activate exonic and intronic cryptic splice sites. *Am J Hum Genet.* 2017;100:751–765.
133. Park E, Pan Z, Zhang Z, Lin L, Xing Y. The expanding landscape of alternative splicing variation in human populations. *Am J Hum Genet.* 2018;102:11–26.
134. Wu N, Ming X, Xiao J, et al. *TBX6* null variants and a common hypomorphic allele in congenital scoliosis. *N Engl J Med.* 2015;372:341–350.
135. Scott EM, Halees A, Itan Y, et al. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet.* 2016;48:1071–1076.
136. Liu J, Wu N, Deciphering Disorders Involving Scoliosis and Comorbidities (DISCO) study, et al. *TBX6*-associated congenital scoliosis (TACS) as a clinically distinguishable subtype of congenital scoliosis: further evidence supporting the compound inheritance and *TBX6* gene dosage model. *Genet Med* 2018 <https://doi.org/10.1038/s41436-018-0377-x>
137. Yang N, Wu N, Zhang L, et al. *TBX6* compound inheritance leads to congenital vertebral malformations in humans and mice. *Hum Mol Genet.* 2018. <https://doi.org/10.1093/hmg/ddy358> [Epub ahead of print].
138. Sparrow DB, Chapman G, Smith AJ, et al. A mechanism for gene-environment interaction in the etiology of congenital scoliosis. *Cell.* 2012;149:295–306.

139. Stuart BD, Choi J, Zaidi S, et al. Exome sequencing links mutations in *PARN* and *RTEL1* with familial pulmonary fibrosis and telomere shortening. *Nat Genet.* 2015;47:512–517.
140. McDonald CM, Campbell C, Torricelli RE, et al. Ataluren in patients with nonsense mutation Duchenne muscular dystrophy (ACT DMD): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet.* 2017;390:1489–1498.
141. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* 2008;9:356–369.
142. Manolio TA. Bringing genome-wide association findings into clinical use. *Nat Rev Genet.* 2013;14:549–558.
143. Braun DA, Sadowski CE, Kohl S, et al. Mutations in nuclear pore genes *NUP93*, *NUP205* and *XPO5* cause steroid-resistant nephrotic syndrome. *Nat Genet.* 2016;48:457–465.
144. Gee HY, Sadowski CE, Aggarwal PK, et al. *FAT1* mutations cause a glomerulotubular nephropathy. *Nat Commun.* 2016;7:10822.
145. Takahashi S, Andreoletti G, Chen R, et al. De novo and rare mutations in the *HSPA1L* heat shock gene associated with inflammatory bowel disease. *Genome Med.* 2017;9:8.
146. Bastarache L, Hughey JJ, Hebring S, et al. Phenotype risk scores identify patients with unrecognized Mendelian disease patterns. *Science.* 2018;359:1233–1239.
147. Ji W, Foo JN, O'Roak BJ, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet.* 2008;40:592–599.
148. Saleheen D, Natarajan P, Armean IM, et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature.* 2017;544:235–239.
149. O'Roak BJ, Stessman HA, Boyle EA, et al. Recurrent de novo mutations implicate novel genes underlying simplex autism risk. *Nat Commun.* 2014;5:5595.
150. de Ligt J, Willemsen MH, van Bon BW, et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med.* 2012;367:1921–1929.
151. Lupski JR. New mutations and intellectual function. *Nat Genet.* 2010;42:1036–1038.
152. Vissers LE, de Ligt J, Gilissen C, et al. A de novo paradigm for mental retardation. *Nat Genet.* 2010;42:1109–1112.
153. Coe BP, Witherspoon K, Rosenfeld JA, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet.* 2014;46:1063–1071.
154. Turner TN, Coe BP, Dickel DE, et al. Genomic patterns of de novo mutation in simplex autism. *Cell.* 2017;171:710–722 e712.
155. Turner TN, Hormozdiari F, Duyzend MH, et al. Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA. *Am J Hum Genet.* 2016;98:58–74.
156. Zaidi S, Choi M, Wakimoto H, et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature.* 2013;498:220–223.
157. Weiss LA, Shen Y, Korn JM, et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med.* 2008;358:667–675.
158. McCarthy SE, Makarov V, Kirov G, et al. Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet.* 2009;41:1223–1227.
159. Kumar RA, KaraMohamed S, Sudi J, et al. Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet.* 2008;17:628–638.
160. Kotlar AV, Mercer KB, Zwick ME, Mulle JG. New discoveries in schizophrenia genetics reveal neurobiological pathways: a review of recent findings. *Eur J Med Genet.* 2015;58:704–714.
161. Crespi B, Stead P, Elliot M. Evolution in health and medicine Sackler colloquium: comparative genomics of autism and schizophrenia. *Proc Natl Acad Sci USA* 2010;107 suppl 1:1736–1741.
162. Brunetti-Pierri N, Berg JS, Scaglia F, et al. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet.* 2008;40:1466–1471.
163. Lu XY, Phung MT, Shaw CA, et al. Genomic imbalances in neonates with birth defects: high detection rates by using chromosomal microarray analysis. *Pediatrics.* 2008;122:1310–1318.
164. Lupski JR, de Oca-Luna RM, Slaugenhaupt S, et al. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell.* 1991;66:219–232.
165. Cabrejo L, Guyant-Marechal L, Laquerriere A, et al. Phenotype associated with *APP* duplication in five families. *Brain.* 2006;129 pt 11:2966–2976.
166. Rovelet-Lecrux A, Hannequin D, Raux G, et al. *APP* locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet.* 2006;38:24–26.
167. Del Colle R, Fabrizi GM, Turazzini M, Cavallaro T, Silvestri M, Rizzuto N. Hereditary neuropathy with liability to pressure palsies: electrophysiological and genetic study of a family with carpal tunnel syndrome as only clinical manifestation. *Neurol Sci.* 2003;24:57–60.
168. Potocki L, Chen KS, Koeuth T, et al. DNA rearrangements on both homologues of chromosome 17 in a mildly delayed individual with a family history of autosomal dominant carpal tunnel syndrome. *Am J Hum Genet.* 1999;64:471–478.
169. Athanassiadou A, Voutsinas G, Psiouri L, et al. Genetic analysis of families with Parkinson disease that carry the Ala53Thr mutation in the gene encoding alpha-synuclein. *Am J Hum Genet.* 1999;65:555–558.
170. Polymeropoulos MH, Lavedan C, Leroy E, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science.* 1997;276:2045–2047.
171. Singleton AB, Farrer M, Johnson J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science.* 2003;302:841.
172. Simon-Sanchez J, Schulte C, Bras JM, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet.* 2009;41:1308–1312.
173. Satake W, Nakabayashi Y, Mizuta I, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet.* 2009;41:1303–1307.
174. Li AH, Morrison AC, Kovar C, et al. Analysis of loss-of-function variants and 20 risk factor phenotypes in 8,554 individuals identifies loci influencing chronic disease. *Nat Genet.* 2015;47:640–642.
175. Reid JG, Carroll A, Veeraraghavan N, et al. Launching genomics into the cloud: deployment of Mercury, a next generation sequence analysis pipeline. *BMC Bioinformatics.* 2014;15:30.
176. Wang M, Beck CR, English AC, et al. PacBio-LITS: a large-insert targeted sequencing method for characterization of human disease-associated chromosomal structural variations. *BMC Genomics.* 2015;16:214.
177. Gambin T, Jhangiani SN, Below JE, et al. Secondary findings and carrier test frequencies in a large multiethnic sample. *Genome Med.* 2015;7:54.
178. Wangler MF, Yamamoto S, Chao HT, et al. Model organisms facilitate rare disease diagnosis and therapeutic research. *Genetics.* 2017;207:9–27.
179. Yamamoto S, Jaiswal M, Charnig WL, et al. A drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. *Cell.* 2014;159:200–214.
180. Shah PS, Link N, Jang GM, et al. Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. *Cell.* 2018;175:1931–1945.
181. Yoon WH, Sandoval H, Nagarkar-Jaiswal S, et al. Loss of nardilysin, a mitochondrial co-chaperone for alpha-ketoglutarate dehydrogenase, promotes mTORC1 activation and neurodegeneration. *Neuron.* 2017;93:115–131.
182. Luo X, Rosenfeld JA, Yamamoto S, et al. Clinically severe *CACNA1A* alleles affect synaptic function and neurodegeneration differentially. *PLoS Genet.* 2017;13:e1006905.
183. Schoch K, Meng L, Szelinger S, et al. A recurrent de novo variant in *NACC1* causes a syndrome characterized by infantile epilepsy, cataracts, and profound developmental delay. *Am J Hum Genet.* 2017;100:343–351.
184. Carvalho CM, Pfundt R, King DA, et al. Absence of heterozygosity due to template switching during replicative rearrangements. *Am J Hum Genet.* 2015;96:555–564.

Jennifer E. Posey, MD, PhD¹, Anne H. O'Donnell-Luria, MD, PhD^{2,3,4}, Jessica X. Chong, PhD⁵, Tamar Harel, MD, PhD⁶, Shalini N. Jhangiani, MS⁷, Zeynep H. Coban Akdemir, PhD¹, Steven Buyske, PhD^{8,9}, Davut Pehlivan, MD¹, Claudia M. B. Carvalho, PhD¹, Samantha Baxter, MS, CGC³, Nara Sobreira, MD, PhD¹⁰, Pengfei Liu, PhD^{1,11}, Nan Wu, MD^{1,12}, Jill A. Rosenfeld, MS, CGC¹, Sushant Kumar, PhD¹³, Dimitri Avramopoulos, MD, PhD¹⁰, Janson J. White, PhD^{1,5}, Kimberly F. Doherty, PhD^{10,14}, P. Dane Witmer, PhD^{10,14}, Corinne Boehm, MS, CGC¹⁰, V. Reid Sutton, MD¹, Donna M. Muzny, MS⁷, Eric Boerwinkle, PhD^{7,15}, Murat Günel, MD^{16,17}, Deborah A. Nickerson, PhD¹⁸, Shrikant Mane, PhD¹⁹, Daniel G. MacArthur, PhD^{2,3}, Richard A. Gibbs, PhD^{1,7}, Ada Hamosh, MD¹⁰, Richard P. Lifton, MD, PhD^{16,20,21}, Tara C. Matise, PhD⁸, Heidi L. Rehm, PhD^{2,3}, Mark Gerstein, PhD¹³, Michael J. Bamshad, MD^{5,18}, David Valle, MD¹⁰ and James R. Lupski, MD, PhD, DSc (hon)^{1,7,22,23}

Centers for Mendelian Genomics

¹Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. ²Analytic and Translational Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. ³Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁴Boston Children's Hospital, Boston, MA, USA. ⁵Department of Pediatrics, University of Washington, Seattle, WA, USA. ⁶Department of Genetic and Metabolic Diseases, Hadassah-Hebrew University Medical Center, Jerusalem, Israel. ⁷The Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. ⁸Department of Genetics, Rutgers University, Piscataway, NJ, USA. ⁹Department of Statistics, Rutgers University, Piscataway, NJ, USA. ¹⁰McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD, USA. ¹¹Baylor Genetics Laboratory, Houston, TX, USA. ¹²Department of Orthopedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, China. ¹³Computational Biology and Bioinformatics Program, Yale University Medical School, New Haven, CT, USA. ¹⁴Center for Inherited Disease Research, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ¹⁵Human Genetics Center, University of Texas Health Science Center, Houston, TX, USA. ¹⁶Department of Neurosurgery, Yale School of Medicine, New Haven, CT, USA. ¹⁷Department of Genetics, Yale School of Medicine, New Haven, CT, USA. ¹⁸Department of Genome Sciences, University of Washington, Seattle, WA, USA. ¹⁹Yale Center for Genome Analysis, Yale School of Medicine, Yale University, New Haven, CT, USA. ²⁰Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA. ²¹Laboratory of Human Genetics and Genomics, The Rockefeller University, New York, NY, USA. ²²Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA. ²³Texas Children's Hospital, Baylor College of Medicine, Houston, TX, USA