COMMENTARY

FORGE Canada Consortium: Outcomes of a 2-Year National Rare-Disease Gene-Discovery Project

Chandree L. Beaulieu,[1] Jacek Majewski,[2] Jeremy Schwartzentruber,[3] Mark E. Samuels,[4] Bridget A. Fernandez,[5] Francois P. Bernier,[6] Michael Brudno,[7,12] Bartha Knoppers,[8] Janet Marcadier,[1] David Dyment,[1] Shelin Adam,[9] Dennis E. Bulman,[1] Steve J.M. Jones,[10] Denise Avard,[8] Minh Thu Nguyen,[8] Francois Rousseau,[11] Christian Marshall,[12] Richard F. Wintle,[12] Yaoqing Shen,[10] Stephen W. Scherer,[12,13] FORGE Canada Consortium,[1] Jan M. Friedman,[9] Jacques L. Michaud,[4] and Kym M. Boycott[1,*]

Inherited monogenic disease has an enormous impact on the well-being of children and their families. Over half of the children living with one of these conditions are without a molecular diagnosis because of the rarity of the disease, the marked clinical heterogeneity, and the reality that there are thousands of rare diseases for which causative mutations have yet to be identified. It is in this context that in 2010 a Canadian consortium was formed to rapidly identify mutations causing a wide spectrum of pediatric-onset rare diseases by using whole-exome sequencing. The FORGE (Finding of Rare Disease Genes) Canada Consortium brought together clinicians and scientists from 21 genetics centers and three science and technology innovation centers from across Canada. From nation-wide requests for proposals, 264 disorders were selected for study from the 371 submitted; disease-causing variants (including in 67 genes not previously associated with human disease; 41 of these have been genetically or functionally validated, and 26 are currently under study) were identified for 146 disorders over a 2-year period. Here, we present our experience with four strategies employed for gene discovery and discuss FORGE's impact in a number of realms, from clinical diagnostics to the broadening of the phenotypic spectrum of many diseases to the biological insight gained into both disease states and normal human development. Lastly, on the basis of this experience, we discuss the way forward for rare-disease genetic discovery both in Canada and internationally.

Introduction

Seventy-five percent of rare diseases affect children and thus have an enormous impact on the well-being of families.[1] A rare disease is defined as one that affects fewer than 200,000 people in the United States or fewer than 1 in 2,000 people in Europe; although individually rare, collectively these conditions affect millions of children worldwide. Most rare diseases are genetic in origin; the precise number is unknown, but best estimates suggest that there are at least 7,000, and possibly many more, rare genetic diseases.[2,3] An early and accurate genetic diagnosis is critical to the optimal care for a child with a rare genetic disease and their family. However, diagnosis of a rare genetic disease can be a challenge and is clearly contingent upon understanding the molecular etiology of the disease.

The number of genes known to harbor pathogenic variants, which currently account for approximately half of the estimated 7,000 rare genetic diseases,[2] is rapidly increasing with the application of next-generation sequencing (NGS) technologies to rare-disease research.[4] In 2010, Canadian funding agencies Genome Canada and the Canadian Institutes for Health Research partnered in a call for a collaborative national consortium to study rare diseases by using NGS technology. From this funding opportunity, the FORGE (Finding of Rare Disease Genes) Canada Consortium was launched with the objective of rapidly identifying genes associated with a wide spectrum of rare pediatric-onset single-gene disorders present in the Canadian population over a 2-year period (April 2011 to March 2013). Here, we present the FORGE network infrastructure, clinical and gene-discovery pipelines, results, and insight gained from the study of over 250 rare childhood genetic diseases.

FORGE Canada

The FORGE Canada Consortium was developed with the concept that cooperation and collaboration, on both national and international levels, are critical factors for success in the study of rare disease. Canada, the world's second-largest country by area, has a

[1]Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, ON K1H 8L1, Canada; [2]Department of Human Genetics, McGill University, Montreal, QC H3A 1B1, Canada; [3]McGill University and Genome Quebec Innovation Centre, Montreal, QC H3A 1A4, Canada; [4]Centre de Recherche du Centre Hospitalier Universitaire Sainte-Justine, Department of Medicine, Université de Montréal, Montreal, QC H3T 1C5, Canada; [5]Disciplines of Genetics and Medicine, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL A1B 3V6, Canada; [6]Department of Medical Genetics, University of Calgary, Calgary, AB T2N 4N1, Canada; [7]Department of Computer Science, Donnelly Centre and Banting and Best Department of Medical Research, University of Toronto and Centre for Computational Medicine, Hospital for Sick Children, Toronto, ON M5S 3G4, Canada; [8]McGill University and Centre of Genomics and Policy, Montreal, QC H3A 1A4, Canada; [9]Department of Medical Genetics, University of British Columbia, Vancouver, BC V6H 3N1, Canada; [10]Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, BC V5Z4S6, Canada; [11]Centre de Recherche du Centre Hospitalier Universitaire de Québec, Department of Molecular Biology, Medical Biochemistry, and Pathology, Université Laval, Quebec City, QC G1L 3L5, Canada; [12]The Centre for Applied Genomics and Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON M5G 0A4, Canada; [13]McLaughlin Centre and the Department of Molecular Genetics, University of Toronto, Toronto, ON M5G 0A4, Canada
*Correspondence: kboycott@cheo.on.ca
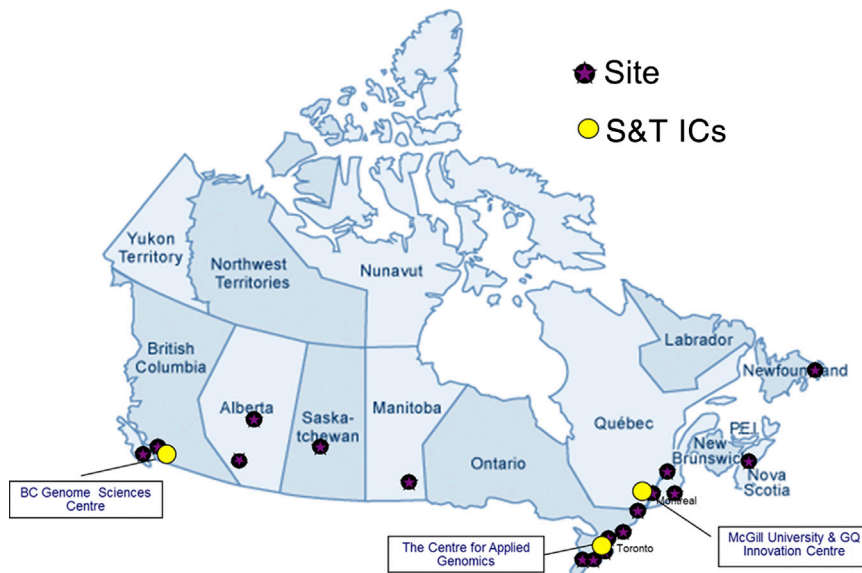
CrossMark

Figure 1. A Map of Canada Depicts the Location of Participating Clinical Sites and S&T ICs
Abbreviations are as follows: GQ, Genome Quebec; and S&T ICs, science and technology innovation centers.

population of approximately 35 million people living in ten provinces and three territories. Many of the Canadians living with rare genetic diseases are evaluated by one of the ~95 clinical geneticists working from one of 21 different genetics centers spread across the ten provinces. Despite this wide geographic dispersion, almost all clinical geneticists belong to the Canadian College of Medical Geneticists, resulting in a tightly knit community. In addition to including clinical geneticists, the 170 FORGE members consist of pediatric subspecialists, bioinformaticians, and molecular biologists with expertise in rare genetic diseases. International collaborations with clinicians from 17 different countries were established on an ad hoc basis. Each of the major clinical genetics centers identified a site lead (Figure 1; Table S1, available online) to ensure national engagement, and a steering committee of nine individuals was appointed to direct all administrative and operational aspects of the project. The Children's Hospital of Eastern Ontario Research Institute was established as the lead institution.

Clinical Pipeline
National calls for disorders to be studied were emailed to the Canadian clinical genetics and genomics community with a membership listserv. Submitted conditions had to be congenital or of pediatric onset, be most likely monogenic and have a molecular etiology not understood, and affect at least one Canadian person; appropriate investigations (the standard of care, including chromosomal microarray, for the respective province) had to have been performed to exclude known causes. Two-page applications for each disorder were evaluated by the steering committee for likelihood of successful gene discovery by one of four strategies (Table 1). When necessary, the clinical network was used for identifying additional affected individuals before the disorder entered the pipeline. Our GE[3]LS (genomics and its ethical, environmental, economic, legal, and social aspects) team developed a model consent form reflecting core principles based on best practices, particularly with regard to the disclosure of research results and the sharing of data and biological samples within Canada and internationally (consent templates and supporting documents are provided in Figure S1). A total of 371 disorders qualified for entry into the pipeline; 264 disorders representing more than 1,000 Canadian samples and an additional 300 international samples were studied during the project. The types of disorders studied represented a broad range of pediatric-onset rare diseases but were enriched with multiple-malformation syndromes and neurodegenerative disorders (Table S2). The group of affected individuals and their families assembled for study was a remarkable Canadian resource for gene discovery and was a critical aspect of the success of FORGE.

Gene-Discovery Pipeline
Once a sufficiently sized set of affected individuals for gene discovery was identified, a project team was assembled from investigators who had expressed an interest in studying the disorder, and samples were sent for whole-exome sequencing (WES) to one of three Genome Canada science and technology innovation centers (S&T ICs; Figure 1): McGill University and Genome Quebec Innovation Centre (Montreal), The Centre for Applied Genomics (Toronto), and The Genome Sciences Centre (Vancouver). To facilitate communication, the sharing of data, and the development of analysis tools, a national data coordination center (NDCC) was established at The Hospital for Sick Children and the University of Toronto. The 264 rare disorders were studied by WES analysis of 783 samples.

Exome target enrichment was performed with the Agilent SureSelect 50Mb (V3) All Exon Kit; thereafter, the majority of sequencing was performed on the Illumina HiSeq 2000, multiplexing three samples per lane. After duplicate reads were accounted for, the mean coverage of coding-sequence regions ranged from 70× to 200×. The FORGE informatics team analyzed WES data at each S&T IC by using very similar pipelines based on alignment with the Burrows-Wheeler Aligner,[5] removal of duplicate reads with Picard, local indel realignment with the Genome Analysis Toolkit,[6] variant calling with SAMtools,[7] and annotation with ANNOVAR[8] and custom scripts. Annotations used regularly included variant functional

Table 1. FORGE Strategies for Gene Discovery

Strategy	Sample Characteristics	Approach to Analysis	No. of Disorders
1	multiple unrelated individuals or families affected by the same very rare but highly recognizable clinical condition	identify a common disease-associated gene or pathway shared between unrelated affected individuals	32
2a	consanguineous families	map on the basis of homozygosity to exclude the majority of the genome	60
2b	autosomal-dominant families	map to exclude the majority of the genome	19
3	nonconsanguineous families with two or more affected siblings	identify compound-heterozygous variants (in the same gene) shared between affected siblings	62
4	single affected individuals with no family history	identify deleterious variants in genes with disease associations	91

effect, gene OMIM associations, the number of internal exomes in which the variant was seen, the frequency of gene mutations in control exomes (for identifying dispensable genes), genotype counts from the NHLBI exomes (for determining whether a variant had been seen in homozygous form), Genomic Evolutionary Rate Profiling conservation scores, and metrics for missense single-nucleotide-variant pathogenicity (SIFT, Poly-Phen, Mutation Taster, likelihood-ratio test). Copy-number variants (CNVs) were called from exome data with FishingCNV[9] and XHMM,[10] which was important for identifying a small number of pathogenic CNVs. For a given disorder, internal exomes from the other disorders studied to that point were used as controls for CNV calling and for filtering out variants likely to be technical artifacts. Only rare variants were considered candidates for any of the disorders studied; therefore, variants were removed from candidate lists if they were seen at greater than 3% allele frequency in our control exomes or greater than 1% allele frequency in the 1000 Genomes Project or NHLBI Exome Sequencing Project Exome Variant Server exomes. Close collaboration between the project teams and bioinformaticians processing the WES data facilitated variant filtering and the gene-discovery process.

The project teams performed experiments necessary to validate the candidate variants as disease causing within a particular gene and thus render the disorder explained or "solved." A definitive outcome for a particular disorder occurred when the variant under consideration was in a gene previously known to cause disease (known gene), had been previously identified and assessed for pathogenicity, or had attributes that allowed for relatively accurate clinical interpretation and when the referring clinician provided feedback that this explained the affected individual's phenotype. It was more challenging to interpret uncharacterized variants in a gene not previously associated with disease (novel gene) as disease causing, and thus causality was supported with either genetic or functional data. Deleterious-appearing variants in the same novel gene in unrelated affected individuals with the same or overlapping phenotype typically constituted confirmation of causality. However, biochemical validation of a putative pathological variant in a novel gene to confirm the functional significance was considered confirmatory in some cases, for example, when a deleterious-appearing variant in a well-defined molecular pathway resulted in a metabolite that was easily measured with a clinical laboratory test. In other instances, additional in vitro and in vivo functional studies were required. If the identification of disease-causing variant(s) in a novel gene was successful, the project team was responsible for publishing the findings.

Gene Identification

Of the 264 disorders studied during the FORGE project, definitively dis-ease-causing mutations were identified to explain 120 disorders, and for 26 disorders, highly likely disease-causing variants were identified in novel candidate genes; 118 remain unexplained. These 146 disorders represented 67 novel genes and 95 known genes. This total of mutations in 162 genes for 146 disorders reflects that on several occasions, mutations in different genes were found to be causal in different affected individuals in a cohort with ostensibly the same disorder; alternatively, what was initially seen as one disease in an affected individual or family was found to be two or more. Of the 67 identified novel genes with disease-causing variants (20 have been published as of January 2014, see Table 2), 30 have mutations in unrelated affected individuals with the same phenotype, 11 have functional evidence in support of disease causation, and 26 have rare predicted pathogenic variants that segregate with the disease and a credible functional link (based on disease pathway or model organism data) but are in the process of validation. The rate of novel-gene discovery and the overall success rate differ for the four strategies utilized (Figure 2) and inform the way to successfully investigate the remaining unsolved rare diseases.

Strategy 1: Multiple Unrelated Individuals or Families Affected by the Same Disease
Strategy 1 was the most successful approach for novel-gene discovery and highlighted the importance of comprehensive and detailed phenotyping, as well as the challenge of

Table 2. Gene-Discovery Publications by FORGE Canada as of January 2014

Disorder	MIM	Category	Gene	Year
CAPOS syndrome[11]	601338	1	ATP1A3[a]	2014
Chudley-McCullough syndrome[12]	604213	2a	GPSM2[a]	2012
Floating-Harbor syndrome[13]	136140	1	SRCAP	2012
Hajdu-Cheney syndrome[14]	102500	1	NOTCH2[a]	2011
Hereditary spastic paraplegia (complex)[15]	615033	3	DDHD2	2012
Infantile mitochondrial complex II and III deficiency[16]	603485	2a	NFS1	2014
Jeune syndrome[17]	615630	2a	IFT172	2013
Infantile myofibromatosis[18]	228550	1	PDGFRB	2013
Intellectual disability[19]	613680	2a	THOC6	2013
Joubert syndrome[20]	614970	1	TMEM231	2012
Joubert syndrome[21]	614615	1	C5ORF42	2012
Leber congenital amaurosis[22]	608553	2a	NMNAT1	2012
Mandibulofacial dysostosis with microcephaly[23]	610536	1	EFTUD2	2012
Metaphyseal dysplasia with maxillary hypoplasia and brachydactyly[24]	156510	2b	RUNX2[a]	2013
Megalencephaly syndromes[25]	603387	1	AKT3, PIK3R2, and PIK3CA	2012
Microcephaly-capillary malformation syndrome[26]	614261	1	STAMBP	2013
Multiple intestinal atresia[27]	243150	2a	TTC7A	2013
Nager syndrome[28]	154400	1	SF3B4	2012
SHORT syndrome[29]	269880	1	PIK3R1	2013
Weaver syndrome[30]	277590	1	EZH2	2012

[a]Already disease-associated gene found to be associated with a previously undescribed mechanism or phenotype.

rare-disease genetic heterogeneity. WES from multiple singletons or trios was analyzed for pathogenic variants in a gene common to unrelated affected individuals (two to eight affected individuals underwent WES per disorder). Thirty-two disorder cohorts, both autosomal recessive and dominant, were studied with this strategy. Overall, disease-causing variants were identified in 25 novel genes (both published and unpublished; 23 validated and two under study) and 20 known genes for 30 disorder cohorts; for two disorders, no causative variants were revealed. Fourteen of the 32 disorders studied (e.g., Floating-Harbor syndrome[13] [MIM 136140], mandibulofacial dysostosis with microcephaly[23] [MIM 610536], microcephaly-capillary malformation syndrome[26] [MIM 614261], multiple intestinal atresia[27] [MIM 243150], Nager syndrome[28] [MIM 154400], SHORT syndrome[29] [MIM 269880], and Weaver syndrome[30] [MIM 277590]) had mutations identified in a single gene in the majority of affected individuals. Several ostensibly recognizable disorders presumed to have a single cause had causative variants identified in multiple genes. For example, mutations (de novo germline and postzygotic) were detected in three genes encoding PI3K-AKT pathway proteins in the megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome (MIM 603387) cohort.[25] Elsewhere, an unexpected level of genetic heterogeneity was encountered in a French Canadian Joubert syndrome (MIM 614970 and 614615) cohort coming from the same region of Quebec; five novel genes (including TMEM231[20] [MIM 614949] and C5ORF42[21] [MIM 614571]) and three known genes were identified. In contrast, our analyses of all available previously reported individuals with Fitzsimmons syndrome (MIM 270710) revealed that in this instance, the syndrome was a combination of two or more rare diseases and thus unlikely to exist as a single entity (data not shown). In this regard, given the level of genetic heterogeneity combined with challenging clinical recognition, it is likely that gene identification for some and perhaps many of the remaining recognizable syndromes (e.g., Dubowitz syndrome [MIM 223370]) will require significant investment in resources and broad collaboration to move forward.

Strategy 2: Single Families with Mapping Data

A significant proportion of the genome was excluded from analysis with the use of mapping data for both consanguineous and autosomal-dominant families (with more than five members available for analysis). One or two affected individuals per family underwent WES and subsequent primary analysis focused on genetic variants within the mapped regions established with either SNP arrays or the WES data (strategy 2a only). Homozygous mutations were identified (strategy 2a, Figure 2) in 37 out of 60 (62%) disorders; 17 of these were mutations in novel genes (nine validated and eight under study). For example, homozygous mutations were identified in IFT172 (MIM 607386), associated with Jeune syndrome[17] (MIM 615630), which was subsequently validated through the identification of additional affected individuals as part of an international collaboration; in NFS1 (MIM 603485), associated with infantile mitochondrial complex II and III deficiency;[16] and in THOC6 (MIM 615403), associated with intellectual disability[19] (MIM 613680). The latter two genes are supported as disease

causing by functional studies. Twenty of the disorders in strategy 2a had homozygous mutations identified in known genes. Some of these findings expanded the phenotype beyond that originally recognized, whereas others were mutations in recently discovered genes lacking clinical testing or, as in the case of *SLC52A2* (MIM 607882), associated with sensory neuropathy[31] (MIM 614707), published while our study was underway. Many of the known genes detected were for genetically heterogeneous diseases where clinical testing had only ruled out the most commonly involved genes. Five disorders manifesting no strong candidate homozygous variant had compound-heterozygous mutations detected (in known disease-associated genes *CRB1* [MIM 604210] and *SYNE1* [MIM 608441], one validated novel gene, and two potentially novel genes). Eighteen of the 60 disorders remained unsolved (30%); these had no strong or unambiguous homozygous candidate variant detected.

The study of moderately sized autosomal-dominant families with available mapping data (we did not require a LOD score greater than 3 for families to be included, but usually at least five informative meiotic events were required) proved to be challenging. Out of 19 families, disease-causing variants in seven (37%) genes were identified; only one was a potentially novel gene (and is currently being functionally validated), and another was a novel mechanism in a known disease-associated gene, *RUNX2* (MIM 600211).[24] Depending on the resolution of the mapping data, some disorders were left with more heterozygous candidate variants than could be prioritized, whereas others had no candidate variants identified within a shared haplotype. Those
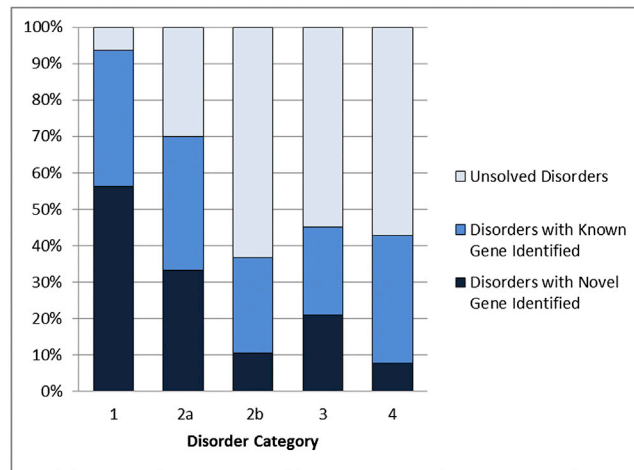

Figure 2. Outcomes of the 264 Disorders Studied with Each Strategy
Strategy 1 was used for multiple unrelated individuals or families affected by the same very rare but highly recognizable clinical condition (32 disorders); strategy 2a was for consanguineous families (60 disorders); strategy 2b was for autosomal-dominant families (19 disorders); strategy 3 was for nonconsanguineous families with two or more affected siblings (62 disorders); and strategy 4 was for single affected individuals with no family history (91 disorders).

with a well-defined single mapped region and no candidate variant are being pursued further by whole-genome sequencing in the search for a noncoding mutation.

Strategy 3: Autosomal-Recessive, Nonconsanguineous Families

This approach focused on the identification of compound-heterozygous variants shared between two or more affected siblings; in some instances, one parent was also sequenced for facilitating rapid phasing of variants in the same gene. Of 62 disorders studied with this pedigree structure, disease-causing variants were identified in 13 novel genes (two validated and 11 under study) and 15 genes in which mutations are known to be associated with human disease (success rate of 45%). The majority of the variants identified were compound heterozygous (22 out of the 28 disorders), as expected. Other inheritance patterns included homozygous recessive (3), X-linked (2), and dominant with suspected parental gonadal mosaicism (1). Although often a small number of very rare recessive variants predicted to affect the protein (nonsynonymous, frameshift,

splicing) and shared between two siblings (in two to ten genes depending on the number of siblings and ethnicity in comparison to control samples) were identified, the validation of novel genes found in only a single family remains an ongoing challenge. For example, analysis of two siblings with complex hereditary spastic paraplegia (MIM 615033) yielded two genes with multiple heterozygous variants (potentially compound heterozygous) and two genes with homozygous variants. Ultimately, the disease-associated gene was tentatively identified as *DDHD2* (MIM 615003) on the basis of variant type, rarity, and gene function and subsequently validated by the identification of additional families through international collaboration.[15]

Strategy 4: Single Affected Individual with No Family History

We believe that the scenario of a single affected individual with no family history or mapping information will be encountered with greater frequency as increasingly rare disorders are analyzed and clinical exome sequencing becomes broadly available. The diagnostic yield of this approach was explored in the final year of FORGE, particularly for genetically heterogeneous conditions. We used WES to study 91 affected individuals as singletons (67) or trios (24) and identified seven novel genes (five validated and two under study) and 32 known genes as having disease-causing variants. The success rate (43%) was similar to that of strategy 3; however, a much larger proportion of known genes was identified, which is not unexpected given our inclusion criteria and the fact that this strategy is not ideal for discovery. Compound-heterozygous or homozygous (24 genes), heterozygous (14 genes), and X-linked (one gene) mutations were all identified. Thirteen

known and two novel genes (both validated) were identified in the 24 trios (success rate of 62%), and 19 known genes and five novel genes (three validated and two under study) were identified in the 67 singletons (success rate of 36%). Despite the observed difference in success rate between the trio and singleton approaches, our experience indicated that trio analysis did not appear necessary for the detection of known disease-causing genes for genetically heterogeneous conditions because a single candidate was evident in the majority of cases before analysis of the parental WES data. However, the identification of the two novel genes with de novo mutations would not have been readily identified by singleton sequencing only.

Return of Results to Families and Impact on Patient Care

An important aspect of this study was to return the results to participating families; as front-line care givers, our team can attest to the significant benefits that a molecular diagnosis can provide families. We were able to return a definitive molecular etiology, including mutations in both novel and known genes, to hundreds of families. Validation of suspected pathogenic variants in a known disease-associated gene was often as straightforward as having the contributing clinician review the affected individual's clinical presentation in light of the new genotypic information. Mutations in genes not previously associated with human disease were generally shared with families during manuscript preparation. For mutations in both known and novel genes, the contributing clinician discussed the findings with the family and confirmed the mutation(s) in a clinically certified molecular diagnostic laboratory before they were used in making patient management decisions.

Our consent process for FORGE informed families that we would not systematically look at WES data for incidental (or secondary) findings but that if we observed a convincing mutation in a medically actionable gene in childhood, we would share this result with the family. Of the >700 exomes analyzed as part of this project, we had only one instance of a medically actionable incidental finding in a child: a previously reported RYR1 mutation causing susceptibility to malignant hyperthermia. As part of the FORGE project, we explored the larger issue of the return of incidental findings from genomic sequencing in pediatric research. Members of our network, as well as the affected individuals and families, were participants in various studies. Perspectives on the professional duty to disclose incidental findings were explored with the use of a questionnaire survey both for parents of children involved in pediatric research and for health professionals.[32] The study revealed that, in general, parents want to receive as much information about their child's health as possible. Concurrently, we engaged parents in focus groups to explore perceptions of genetic risks (incidental findings) from genomic data in a qualitative study.[33] As in the written survey results, parents believed that they should be made aware of all results pertaining to their child's health status irrespective of the potential severity and that they would assume responsibility for communicating this information to their child. Thus, despite the potential negative consequences, respondents perceived the benefits of receiving all incidental findings as outweighing the potential harm. The results of these studies informed, in part, the first draft of the statement of "Principles on the Return of Research Results and Incidental Findings," now endorsed by the Quebec Network of Applied Genetic Medicine Board of Directors and officially adopted in May 2013.[34]

The effects on patient care of achieving a definitive molecular diagnoses during FORGE were often both broad and far reaching. In a few instances, an available effective therapy (e.g., enzyme replacement therapy for a family affected by an atypical form of Hunter syndrome[35] [MIM 309900]) was identified. In several cases, medications were tailored on the basis of the molecular insight (e.g., a change in antiepileptic medication for a child with intractable seizures [MIM 245570] secondary to a GRIN2A [MIM 138253] mutation[36]). The end of the diagnostic odyssey was an important end point for all families who were given a diagnosis during FORGE (many of these families had been investigated for well over a decade). A clear diagnosis facilitated access to services for some children, both in the school and in the community. Screening for complications could also be tailored for many affected individuals, and insight into the clinical ends of the spectrum of a particular disease informed natural history. Families were provided with accurate reproductive counseling, and prenatal diagnosis could be offered. Finally, we found that the psychosocial benefits of diagnostic clarity for families seeking a reason for their child's problems were often dramatic.

Insights Gained from the Study of 264 Disorders

FORGE set out to identify novel genes with mutations causing rare diseases of pediatric onset. Of the 264 disorders studied, 146 have been solved to date, and these represent 67 genes not previously associated with human disease. Four strategies (Table 1) were used for gene discovery, and the most successful were for multiple unrelated individuals or families affected by the same recognizable condition (strategy 1) and consanguineous families (strategy 2a). The distinct advantage in the former case was the validation of the novel gene as associated with disease within the cohort under study, and in the latter case, the number of genes with deleterious homozygous variants was limited by the degree of consanguinity. However, in the latter instance, as with the remaining strategies, identification of additional affected individuals or functional data was needed to support causality.

Overall, the 67 identified novel genes (41 validated and 26 candidates

Table 3. 95 Known Disease-Associated Genes Identified during the Study of 264 Disorders

Inheritance	Genes
Autosomal dominant	*ACTC1*,[37] *ACVR1*,[a] *ARID1B*,[a] *COL11A1*, *EFNB1*, *EFTUD2*[a] (two proposals[b]), *EP300*,[a] *GRIN2A*,[a] *IDS*,[35] *KAT6B*,[a] *MLL2*,[a] *MYOC* (two proposals[b]), *NR5A1*, *NTF4*,[a] *OTX2*, *RPE65*, *SPTAN1*,[a] *SYNGAP1*,[38,a] *TERT*, *WDR36*,[a] and *WNT5A*
Autosomal recessive	*ACSF3*, *AGL*, *AICDA*, *AIMP1*, *ALDH6A1*,[39] *ALG3*, *ASAH1*,[40] *ATM*, *B4GALNT1*, *BRCA1*, *C12orf65*, *CC2D2A* (two proposals[b]), *CCBE1*, *CCDC39*, *CEP290*, *COQ9*, *CORO1A*, *COX10*, *CRB1*, *CYP26C1*, *DHRS3*, *FBXL4*, *FRAS1*, *GNE*, *HSD17B4*[41] (two proposals[b]), *IGHMBP2*, *KCTD7*, *LRP5*, *LRRC6*, *MERTK*, *MTO1*, *MUSK* (two proposals[b]), *NDUFS2*, *NGLY1*, *NNT*,[42] *OFD1*, *PLA2G6*, *PLCB4*, *PMM2*, *POMC*,[43] *PYCR1*, *RAB3GAP1*, *RARS2*, *RLBP1*, *RNF216*, *RTTN*, *SACS* (four proposals[b]), *SEPN1*, *SETX*, *SIL1* (two proposals[b]), *SLC25A1*, *SLC45A2* and *G6PC3* (found together in one family),[44] *SLC52A2*, *SPAG1*, *STAR*, *SYNE1*, *TMPRSS6*,[45] *TRPV4*, *WDR62*, and *ZMYND10*
X-linked	*ABCD1*, *AR*, *CHM*, and *PRPS1*

[a]De novo dominant mutations.
[b]Proposal number indicates separate families submitted by different consortium members.

from the 264 disorders that entered the pipeline) represent a success rate of 16%–25% for novel-gene discovery for the 2-year period. The rate of novel-gene discovery decreased over the period of the project, most likely because the most promising projects entered the pipeline earlier. The novel genes published to date have contributed significantly to our understanding of the biological basis of rare disease and normal human development. Two interesting outcomes from our relatively small subset of novel genes were (1) the observation of convergence to a common underlying biological mechanism for overlapping disease phenotypes and (2) that alterations in the same pathway lead to very distinct diseases. An example of the former is the discovery of mutations in novel genes for two malformation syndromes with overlapping craniofacial features: mandibulofacial dysostosis with microcephaly (*EFTUD2*)[23] and Nager syndrome (*SF3B4*).[28] Both genes are implicated in RNA splicing; *SF3B4* (MIM 605593) encodes splicing factor 3B subunit 4, a component of the U2 pre-mRNA spliceosomal complex, and *EFTUD2* (MIM 603892) encodes U5-116kD, a highly conserved GTPase with a central regulatory role in catalytic RNA splicing and post-splicing-complex disassembly. We also observed within our modest subset of genes how different alterations of a biological pathway can result in distinct diseases. For example, SHORT syndrome, characterized by dysmorphic facies, lipodystrophy, and short stature, is caused by mutations in *PIK3R1* (MIM

171833), and using cell lines derived from affected individuals, we demonstrated downregulation of the AKT-mTOR pathway with the diminished phosphorylation of downstream targets.[29] In marked distinction to the SHORT syndrome phenotype, megalencephaly-capillary malformation syndrome, characterized by overgrowth and cellular proliferation, was found to be secondary to germline or tissue-specific mosaic mutations disrupting other components of this pathway and thus resulting in hyperactivation of the AKT-mTOR pathway.[25] Such discoveries will form the basis of future investigation into new biological processes and might one day enable the configuration of novel therapies.

One of the most surprising insights during FORGE was the large proportion of causative variants identified in genes already known to be associated with human disease despite standard-of-care investigations by our clinical network (Table 3); these mutations were often associated with the broadening of the known disease phenotype. These observations contribute to the growing body of literature supporting the diagnostic utility of WES. The other unexpected finding was the number of affected individuals whose disease presentation represents a conflation of more than one rare disease,[44] suggesting that a subset of affected individuals in genetics clinics appear to have a novel and previously undescribed disease secondary to this phenomenon. Using the lessons learned through FORGE Canada, we are now working

to facilitate translation of this genomic technology into our Canadian genetics clinics.

Finally, how do the 118 (of 264) disorders that remain unsolved (without a clear candidate gene) inform us regarding next steps? A subset of these will most likely include mutations not captured by exome sequencing technologies, e.g., poorly covered coding regions, noncoding mutations, and other disease mechanisms; however, we believe that a significant portion represents an insufficient number of affected individuals to establish disease causation. To facilitate the comparison of phenotypes and genotypes, we established two important and connected resources. The first, PhenoTips,[46] is a standardized phenotyping tool based on the internationally recognized Human Phenotype Ontology.[47] This tool allows clinicians to describe an affected individual with standard terms in less than 5 min by using a web interface with a predictive terminology search. These data are then linked with genomic data in PhenomeCentral, an integrated portal developed to facilitate collaboration and gene discovery. PhenomeCentral is a centralized repository for unsolved rare diseases and uses an automated matching system that connects users who contribute data with strong genotypic and phenotypic similarity. Large-scale collaboration within PhenomeCentral should enable WES results for single affected individuals or families to become more readily interpretable and variants in genes to become more easily validated through

access to data on thousands of other affected individuals contributed by international partners; in this way, it converts affected individuals studied via strategy 4 to a virtual strategy 1. We believe that this freely available resource will assume a central importance as the rarity of the remaining unsolved diseases increases.

Moving Forward

The successful completion of the activities of the FORGE Canada project has provided a coordinated and sustainable national consortium focused on the investigation of the genetic basis of rare human diseases. Moving forward under a new collaborative project, Care4Rare, we expect to continue to have a substantial impact on the diagnostic journey of families living with rare disease in Canada. Care4Rare is focused on delivering two benefits for all Canadians affected by rare disease: (1) efficient and cost-effective molecular diagnoses and (2) a platform for identifying therapeutic opportunities for rare disease. To achieve these benefits, we will continue our gene-discovery pipeline to screen for causal mutations underlying an additional 350 rare genetic disorders affecting Canadian families and facilitate the integration of NGS into the clinic. Care4Rare will develop and validate a pipeline to identify therapeutic opportunities based on repurposing clinic-ready compounds. The disorders submitted to FORGE, but not yet studied, provided the first samples for study within Care4Rare. The expertise in WES analysis gained during FORGE provides a framework for Care4Rare. The National Data Coordination Centre infrastructure, including PhenomeCentral, is in place and ready for use at the larger scale necessary for Care4Rare. Importantly, a number of diseases solved in FORGE are entering the therapeutic discovery pipeline of Care4Rare. Finally, the FORGE Canada Consortium has facilitated our opportunity to become a contributing project to efforts of the International Rare Diseases Research Consortium, thereby ensuring Canada's contribution to this world-wide collaboration going forward.

Supplemental Data

Supplemental Data include one figure and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2014.05.003.

Web Resources

The URLs for data presented herein are as follows:

Canadian College of Medical Geneticists, http://www.ccmg-ccgm.org

Care4Rare, http://care4rare.ca

FASTX-Toolkit, http://hannonlab.cshl.edu/fastx_toolkit/

International Rare Disease Research Consortium, http://www.irdirc.org

NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/

Online Mendelian Inheritance in Man (OMIM), http://www.omim.org

PhenomeCentral, http://phenomecentral.org

Picard, http://picard.sourceforge.net/

SAMtools, http://samtools.sourceforge.net/

References

1. Dodge, J.A., Chigladze, T., Donadieu, J., Grossman, Z., Ramos, F., Serlicorni, A., Siderius, L., Stefanidis, C.J., Tasic, V., Valiulis, A., and Wierzba, J. (2011). Arch. Dis. Child. 96, 791–792.

2. McKusick, V.A. (2007). Am. J. Hum. Genet. 80, 588–604.

3. Samuels, M.E. (2010). Curr. Genomics 11, 482–499.

4. Boycott, K.M., Vanstone, M.R., Bulman, D.E., and MacKenzie, A.E. (2013). Nat. Rev. Genet. 14, 681–691.

5. Li, H., and Durbin, R. (2009). Bioinformatics 25, 1754–1760.

6. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). Genome Res. 20, 1297–1303.

7. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). Bioinformatics 25, 2078–2079.

8. Wang, K., Li, M., and Hakonarson, H. (2010). Nucleic Acids Res. 38, e164.

9. Shi, Y., and Majewski, J. (2013). Bioinformatics 29, 1461–1462.

10. Fromer, M., Moran, J.L., Chambert, K., Banks, E., Bergen, S.E., Ruderfer, D.M., Handsaker, R.E., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., et al. (2012). Am. J. Hum. Genet. 91, 597–607.

11. Demos, M.K., van Karnebeek, C.D., Ross, C.J., Adam, S., Shen, Y., Zhan, S.H., Shyr, C., Horvath, G., Suri, M., Fryer, A., et al.; FORGE Canada Consortium (2014). Orphanet J. Rare Dis. 9, 15.

12. Doherty, D., Chudley, A.E., Coghlan, G., Ishak, G.E., Innes, A.M., Lemire, E.G., Rogers, R.C., Mhanni, A.A., Phelps, I.G., Jones, S.J., et al.; FORGE Canada Consortium (2012). Am. J. Hum. Genet. 90, 1088–1093.

13. Hood, R.L., Lines, M.A., Nikkel, S.M., Schwartzentruber, J., Beaulieu, C., Nowaczyk, M.J., Allanson, J., Kim, C.A., Wieczorek, D., Moilanen, J.S., et al.; FORGE Canada Consortium (2012). Am. J. Hum. Genet. 90, 308–313.

14. Majewski, J., Schwartzentruber, J.A., Caqueret, A., Patry, L., Marcadier, J., Fryns, J.-P., Boycott, K.M., Ste-Marie, L.-G., McKiernan, F.E., Marik, I., et al.; FORGE Canada Consortium (2011). Hum. Mutat. 32, 1114–1117.

15. Schuurs-Hoeijmakers, J.H., Geraghty, M.T., Kamsteeg, E.-J., Ben-Salem, S., de Bot, S.T., Nijhof, B., van de Vondervoort, I.I., van der Graaf, M., Nobau, A.C., Otte-Höller, I., et al.; FORGE Canada Consortium (2012). Am. J. Hum. Genet. 91, 1073–1081.

16. Farhan, S.M., Wang, J., Robinson, J.F., Lahiry, P., Siu, V.M., Prasad, C.,

Kronick, J.B., Ramsay, D.A., Rupar, C.A., and Hegele, R.A. (2014). Mol. Genet. Genomic Med. *2*, 73–80.

17. Halbritter, J., Bizet, A.A., Schmidts, M., Porath, J.D., Braun, D.A., Gee, H.Y., McInerney-Leo, A.M., Krug, P., Filhol, E., Davis, E.E., et al.; UK10K Consortium (2013). Am. J. Hum. Genet. *93*, 915–925.

18. Cheung, Y.H., Gayden, T., Campeau, P.M., LeDuc, C.A., Russo, D., Nguyen, V.-H., Guo, J., Qi, M., Guan, Y., Albrecht, S., et al. (2013). Am. J. Hum. Genet. *92*, 996–1000.

19. Beaulieu, C.L., Huang, L., Innes, A.M., Akimenko, M.-A., Puffenberger, E.G., Schwartz, C., Jerry, P., Ober, C., Hegele, R.A., McLeod, D.R., et al. (2013). Orphanet J. Rare Dis. *8*, 62.

20. Srour, M., Hamdan, F.F., Schwartzentruber, J.A., Patry, L., Ospina, L.H., Shevell, M.I., Désilets, V., Dobrzeniecka, S., Mathonnet, G., Lemyre, E., et al.; FORGE Canada Consortium (2012). J. Med. Genet. *49*, 636–641.

21. Srour, M., Schwartzentruber, J., Hamdan, F.F., Ospina, L.H., Patry, L., Labuda, D., Massicotte, C., Dobrzeniecka, S., Capo-Chichi, J.-M., Papillon-Cavanagh, S., et al.; FORGE Canada Consortium (2012). Am. J. Hum. Genet. *90*, 693–700.

22. Koenekoop, R.K., Wang, H., Majewski, J., Wang, X., Lopez, I., Ren, H., Chen, Y., Li, Y., Fishman, G.A., Genead, M., et al.; Finding of Rare Disease Genes (FORGE) Canada Consortium (2012). Nat. Genet. *44*, 1035–1039.

23. Lines, M.A., Huang, L., Schwartzentruber, J., Douglas, S.L., Lynch, D.C., Beaulieu, C., Guion-Almeida, M.L., Zechi-Ceide, R.M., Gener, B., Gillessen-Kaesbach, G., et al.; FORGE Canada Consortium (2012). Am. J. Hum. Genet. *90*, 369–377.

24. Moffatt, P., Ben Amor, M., Glorieux, F.H., Roschger, P., Klaushofer, K., Schwartzentruber, J.A., Paterson, A.D., Hu, P., Marshall, C., Fahiminiya, S., et al.; FORGE Canada Consortium (2013). Am. J. Hum. Genet. *92*, 252–258.

25. Rivière, J.-B., Mirzaa, G.M., O'Roak, B.J., Beddaoui, M., Alcantara, D., Conway, R.L., St-Onge, J., Schwartzentruber, J.A., Gripp, K.W., Nikkel, S.M., et al.; Finding of Rare Disease Genes (FORGE) Canada Consortium (2012). Nat. Genet. *44*, 934–940.

26. McDonell, L.M., Mirzaa, G.M., Alcantara, D., Schwartzentruber, J., Carter, M.T., Lee, L.J., Clericuzio, C.L., Graham, J.M. Jr., Morris-Rosendahl, D.J., Polster, T., et al.; FORGE Canada Consortium (2013). Nat. Genet. *45*, 556–562.

27. Samuels, M.E., Majewski, J., Alirezaie, N., Fernandez, I., Casals, F., Patey, N., Decaluwe, H., Gosselin, I., Haddad, E., Hodgkinson, A., et al. (2013). J. Med. Genet. *50*, 324–329.

28. Bernier, F.P., Caluseriu, O., Ng, S., Schwartzentruber, J., Buckingham, K.J., Innes, A.M., Jabs, E.W., Innis, J.W., Schuette, J.L., Gorski, J.L., et al.; FORGE Canada Consortium (2012). Am. J. Hum. Genet. *90*, 925–933.

29. Dyment, D.A., Smith, A.C., Alcantara, D., Schwartzentruber, J.A., Basel-Vanagaite, L., Curry, C.J., Temple, I.K., Reardon, W., Mansour, S., Haq, M.R., et al.; FORGE Canada Consortium (2013). Am. J. Hum. Genet. *93*, 158–166.

30. Gibson, W.T., Hood, R.L., Zhan, S.H., Bulman, D.E., Fejes, A.P., Moore, R., Mungall, A.J., Eydoux, P., Babul-Hirji, R., An, J., et al.; FORGE Canada Consortium (2012). Am. J. Hum. Genet. *90*, 110–118.

31. Johnson, J.O., Gibbs, J.R., Megarbane, A., Urtizberea, J.A., Hernandez, D.G., Foley, A.R., Arepalli, S., Pandraud, A., Simón-Sánchez, J., Clayton, P., et al. (2012). Brain *135*, 2875–2882.

32. Fernandez, C.V., Strahlendorf, C., Avard, D., Knoppers, B.M., O'Connell, C., Bouffet, E., Malkin, D., Jabado, N., Boycott, K., and Sorensen, P.H. (2013). Genet. Med. *15*, 558–564.

33. Kleiderman, E., Knoppers, B.M., Fernandez, C.V., Boycott, K.M., Ouellette, G., Wong-Rieger, D., Adam, S., Richer, J., and Avard, D. (2013). J. Med. Ethics. Published online December 19, 2013. http://dx.doi.org/10.1136/medethics-2013-101648.

34. Senecal, K., Levesque, E., Fernandez, C., Tasse, A.M., Zawati, M., Knoppers, B.M., and Avard, D. (2013). Statement of principles on the return of research results and incidental findings. http://www.rmga.qc.ca/en/documents/RMGAStatement_Principles_English_May272013.pdf.

35. Nikkel, S.M., Huang, L., Lachman, R., Beaulieu, C.L., Schwartzentruber, J., Majewski, J., Geraghty, M.T., and Boycott, K.M.; FORGE Canada Consortium (2013). Clin. Genet. Published online July 11, 2013. http://dx.doi.org/10.1111/cge.12236.

36. Venkateswaran, S., Myers, K.A., Smith, A.C., Beaulieu, C.L., Schwartzentruber, J.A., Consortium, F.C., Majewski, J., Bulman, D., Boycott, K.M., and Dyment, D.A. (2014). Epilepsia http://dx.doi.org/10.1111/epi.12663.

37. Greenway, S.C., McLeod, R., Hume, S., Roslin, N.M., Alvarez, N., Giuffre, M., Zhan, S.H., Shen, Y., Preuss, C., Andelfinger, G., et al.; FORGE Canada Consortium (2014). Can. J. Cardiol. *30*, 181–187.

38. Berryer, M.H., Hamdan, F.F., Klitten, L.L., Møller, R.S., Carmant, L., Schwartzentruber, J., Patry, L., Dobrzeniecka, S., Rochefort, D., Neugnot-Cerioli, M., et al. (2013). Hum. Mutat. *34*, 385–394.

39. Marcadier, J.L., Smith, A.M., Pohl, D., Schwartzentruber, J., Al-Dirbashi, O.Y., Majewski, J., Ferdinandusse, S., Wanders, R.J., Bulman, D.E., Boycott, K.M., et al. (2013). Orphanet J. Rare Dis. *8*, 98.

40. Dyment, D., Sell, E., Vanstone, M., Smith, A., Garandeau, D., Garcia, V., Carpentier, S., Le Trionnaire, E., Sabourdy, F., Beaulieu, C., et al.; FORGE Canada Consortium (2013). Clin. Genet. Published online October 25, 2013. http://dx.doi.org/10.1111/cge.12307.

41. McMillan, H.J., Worthylake, T., Schwartzentruber, J., Gottlieb, C.C., Lawrence, S.E., Mackenzie, A., Beaulieu, C.L., Mooyer, P.A., Wanders, R.J., Majewski, J., et al.; FORGE Canada Consortium (2012). Orphanet J. Rare Dis. *7*, 90.

42. Hasselmann, C., Deladoëy, J., Vuissoz, J.-M., Patry, L., Alirezaie, N., Schwartzentruber, J., Consortium, F.C., Deal, C.L., Vliet, G.V., Majewski, J., et al. (2013). Journal of Genomes and Exomes *2*, 19–30.

43. Samuels, M.E., Gallo-Payet, N., Pinard, S., Hasselmann, C., Magne, F., Patry, L., Chouinard, L., Schwartzentruber, J., René, P., Sawyer, N., et al.; FORGE Canada Consortium (2013). J. Clin. Endocrinol. Metab. *98*, 736–742.

44. Fernandez, B.A., Green, J.S., Bursey, F., Barrett, B., MacMillan, A., McColl, S., Fernandez, S., Rahman, P., Mahoney, K., Pereira, S.L., et al.; FORGE Canada Consortium (2012). BMC Med. Genet. *13*, 111.

45. Khuong-Quang, D.-A., Schwartzentruber, J., Westerman, M., Lepage, P., Finberg, K.E., Majewski, J., and Jabado, N. (2013). Pediatrics *131*, e620–e625.

46. Girdea, M., Dumitriu, S., Fiume, M., Bowdin, S., Boycott, K.M., Chénier, S., Chitayat, D., Faghfoury, H., Meyn, M.S., Ray, P.N., et al. (2013). Hum. Mutat. *34*, 1057–1065.

47. Köhler, S., Doelken, S.C., Mungall, C.J., Bauer, S., Firth, H.V., Bailleul-Forestier, I., Black, G.C., Brown, D.L., Brudno, M., Campbell, J., et al. (2014). Nucleic Acids Res. *42* (Database issue), D966–D974.