



Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing

Gregory Costain¹ · Rebekah Jobling^{1,2} · Susan Walker^{3,4} · Miriam S. Reuter^{3,4} · Meaghan Snell¹ · Sarah Bowdin^{1,5,6} · Ronald D. Cohn^{1,4,5,6} · Lucie Dupuis¹ · Stacy Hewson¹ · Saadet Mercimek-Andrews^{1,4,6} · Cheryl Shuman^{1,7} · Neal Sondheimer^{1,4,6} · Rosanna Weksberg^{1,4,6} · Grace Yoon^{1,6,8} · M. Stephen Meyn^{1,5,6,7} · Dimitri J. Stavropoulos^{2,9} · Stephen W. Scherer^{3,4,5,7} · Roberto Mendoza-Londono^{1,5,6} · Christian R. Marshall^{1,2,3,5,9}

Received: 24 September 2017 / Revised: 11 December 2017 / Accepted: 23 January 2018 / Published online: 16 February 2018
© European Society of Human Genetics 2018

Abstract

Whole-genome sequencing (WGS) as a first-tier diagnostic test could transform medical genetic assessments, but there are limited data regarding its clinical use. We previously showed that WGS could feasibly be deployed as a single molecular test capable of a higher diagnostic rate than current practices, in a prospectively recruited cohort of 100 children meeting criteria for chromosomal microarray analysis. In this study, we report on the added diagnostic yield with re-annotation and reanalysis of these WGS data ~2 years later. Explanatory variants have been discovered in seven (10.9%) of 64 previously undiagnosed cases, in emerging disease genes like *HMGA2*. No new genetic diagnoses were made by any other method in the interval period as part of ongoing clinical care. The results increase the cumulative diagnostic yield of WGS in the study cohort to 41%. This represents a greater than 5-fold increase over the chromosomal microarrays, and a greater than 3-fold increase over all the clinical genetic testing ordered in practice. These findings highlight periodic reanalysis as yet another advantage of genomic sequencing in heterogeneous disorders. We recommend reanalysis of an individual's genome-wide sequencing data every 1–2 years until diagnosis, or sooner if their phenotype evolves.

Introduction

Whole-genome sequencing (WGS) has the potential to revolutionize our approach to clinical genetic diagnostics

[1–5]. One proposed advantage of whole-exome sequencing (WES) and WGS is the opportunity for periodic reanalysis of the data in individuals not diagnosed on initial testing [2, 6–10]. The Genome Clinic at The Hospital for Sick Children (Toronto, Canada) is a longitudinal multifaceted research project designed to integrate WGS into mainstream clinical practice [1, 2]. In a previous study, we prospectively recruited 100 paediatric patients referred for a clinical genetics assessment and meeting criteria for chromosomal microarray analysis (CMA) [1]. We found that singleton

These authors contributed equally: Gregory Costain, Rebekah Jobling.

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-018-0114-6>) contains supplementary material, which is available to authorized users.

✉ Christian R. Marshall
crm@sickkids.ca

- 1 Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada
- 2 Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada
- 3 The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada
- 4 Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada

- 5 Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, ON, Canada
- 6 Department of Paediatrics, University of Toronto, Toronto, ON, Canada
- 7 Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada
- 8 Division of Neurology, The Hospital for Sick Children, Toronto, ON, Canada
- 9 Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada

Table 1 Seven diagnostic variants identified after reanalysis of whole-genome sequencing data

Case ID	Phenotype ^a	Sex	Gene	IP	Genomic variant (Zygosity) [transcript]	Origin	MIM Gene #/Phenotype #	Reason not detected in initial analysis
1036	Autism spectrum disorder, GDD	M	<i>ZNF711</i>	XL	c.430G>A p.(Val144Met) (hem) [NM_021998.4]	Mat ^b	314990/300803	Insufficient evidence for genotype-phenotype association
1058	GDD, abnormality of brain morphology, dysmorphic facial features	M	<i>SON</i>	AD	c.3476delC p.(Pro1159Argfs*9) (het) [NM_138927.2]	DN	182465/617140	Gene not recognized as a disease gene
1076	GDD, microcephaly, abnormality of brain morphology, visual impairment	F	<i>AP3B2</i>	AR	c.199C>T p.(Arg67*) (hom) ^c [NM_004644.4]	Mat/Pat	602166/617276	Gene not recognized as a disease gene
1092	Sagittal craniosynostosis, intrauterine growth retardation ^d	M	<i>SMAD6</i>	AD	c.821C>G p.(Ser274Cys) (het) ^e [NM_005585.4]	Mat ^f	602931/617439	Gene not associated with this phenotype
1096	Small for gestational age, short stature, learning difficulties	M	<i>HMGAA2</i>	AD	c.3034delC p.(Ser102Hisfs*64) (het) [NM_003483.4]	Mat ^g	600698/NA	Gene not recognized as a disease gene
1099	GDD, intellectual disability, mild hypotonia, dysmorphic facial features	F	<i>WAC</i>	AD	c.1622delT p.(Leu541Tyrfs*20) (het) [NM_100264.2]	UK ^h	615049/616708	Gene not recognized as a disease gene
1106	Seizures, GDD, ataxia, dystonia	M	<i>KCNBI</i>	AD	c.623T>C* p.(Leu208Pro) (het) ⁱ [NM_004975.3]	DN	600397/616056	Gene not recognized as a disease gene

AD autosomal dominant, DN de novo, F female, GDD global developmental delay, hem hemizygous, het heterozygous, IP inheritance pattern, M male, Mat maternal, NA not available, Pat paternal, XL X-linked, UK unknown

^aOnly primary Human Phenotype Ontology (HPO) terms are displayed. Additional phenotypic details are available upon request

^bMonozygotic twin with a similar phenotype confirmed to carry the same variant

^cPreviously reported as Family 5 in ref. [15]; her similarly affected sister is also homozygous for the variant

^dHe demonstrated appropriate catch-up growth postnatally, and the intrauterine growth retardation is now attributed to being carried in a twin pregnancy

^eMonozygotic twin with a similar phenotype confirmed to carry the same variant. The twins are also homozygous alternate for a putative modifier SNP located upstream of *BMP2* (rs1884302) [23]

^fThe mother is not known to have craniosynostosis, consistent with incomplete penetrance [23]

^gThe mother is similarly affected (Supplemental file). Testing of other affected family members is in progress

^hParental samples were unavailable for testing; however, the variant nonetheless met ACMG criteria for likely pathogenic

ⁱPreviously reported in ref. [14]

WGS identified diagnostic variants in 34 participants. This represented a 4-fold increase in diagnostic rate over CMA alone (8%), and a >2-fold increase over all genetic tests ordered by the clinicians (13%). In only two cases did targeted genetic tests lead to diagnoses not detectable by WGS: microsatellite analysis of parents and offspring for UPD14 (heterodisomy) and a methylation test for Silver–Russell syndrome. We have now systematically re-annotated and reanalyzed the WGS data from our original study, 3 years after the initial annotation. Explanatory variants have been discovered in seven (10.9%) of 64 previously undiagnosed cases, thereby increasing the cumulative diagnostic yield of WGS in the study cohort to 41%. These results provide further support for WGS as a first-tier genetic test.

Subjects and methods

The prospective recruitment and phenotyping of the study participants is described in detail elsewhere [1]. Families were eligible for this study if the proband met clinical criteria for CMA. The study was approved by the Research Ethics Board at The Hospital for Sick Children, and informed written consent was obtained for each participant. WGS was done as a singleton (not trio) experiment, using standard methods [1]. The WGS data were initially annotated in 2014, with all analyses completed by the end of 2015. These data were deposited in the European Genome–Phenome Archive (www.ebi.ac.uk/ega/) under accession number EGAS00001001623. The primary aims of the study were to compare the diagnostic rate of WGS with that of CMA alone, and with that of all genetic testing ordered in the course of routine clinical practice.

WGS variant calls were re-annotated in February 2017 at The Centre for Applied Genomics (Toronto, Canada) using a custom pipeline [1, 2]. This used recent downloads from publicly available databases for allele frequency, gene function, and human disease association. Molecular and clinical geneticists examined variant files and prioritized clinically relevant nuclear DNA variants using the following parameters: (i) sequence quality, (ii) allele frequency, (iii) conservation and predicted impact on coding and non-coding sequence, (iv) presence in ClinVar [11] or Human Gene Mutation Database (HGMD) [12], (v) genic phenotype in Online Mendelian Inheritance in Man (OMIM) and Clinical Genomic Database (CGD) [13], (vi) zygosity and genetic mode of inheritance, and (vii) relevance to clinical phenotype provided. One variant had initially been identified using an alternative analysis method [14], and another was previously identified and included in a case series describing a novel disease gene [15]. Updated phenotype data were extracted from the medical record. Candidate

variants were classified according to the American College of Medical Genetics and Genomics (ACMG) guidelines [16], discussed with the referring clinician, and designated as diagnostic by consensus. These variants were then confirmed by Sanger sequencing in a laboratory with Clinical Laboratory Improvement Amendments (CLIA)/College of American Pathologists (CAP) certification. Inheritance of variants was determined via targeted analysis of parental DNA samples.

Results

New diagnostic variants were identified in seven (10.9%) of the 64 cases after reassessment of all sequence and structural variation in the WGS data (Table 1). All were single nucleotide variants (SNVs), and were successfully confirmed by Sanger sequencing. Five were designated as likely pathogenic or pathogenic using ACMG criteria [16]. The remaining two (in *SMAD6* and *ZNF711*) were designated as variants of uncertain significance and returned to the families by the clinician as probable contributors to the respective proband's phenotype. No diagnoses were made in these 64 study participants by any clinical genetic testing arranged in the interval period. No diagnoses were made by systematic reanalysis of the existing CMA data (data not shown). Thus, the seven new diagnoses increased the cumulative diagnostic yield of WGS in the entire study cohort to 41%, which represents a >5-fold increase over CMA and a >3-fold increase over all testing arranged in the course of routine clinical practice (Fig. 1).

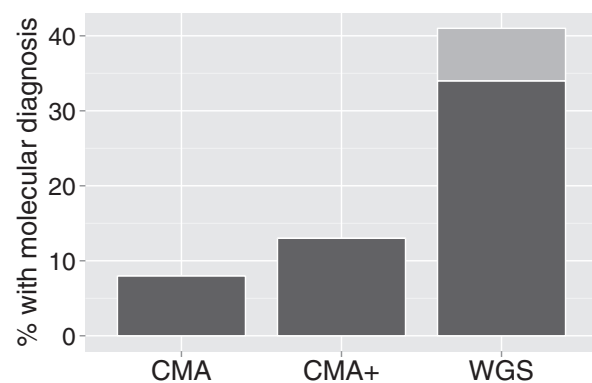


Fig. 1 Diagnostic yield in a prospective cohort study after systematic reanalysis of whole-genome sequencing data. Bar plot showing percentage of study participants ($n = 100$) with molecular diagnoses via chromosomal microarray analysis (CMA), all clinical genetic testing performed in this cohort (CMA+), and whole-genome sequencing (WGS). The CMA and CMA+ diagnostic yields are significantly different ($p < 0.0001$) from the WGS diagnostic yield using a chi-square proportion test. Lighter blue colouring represents the new diagnoses made upon reanalysis of the WGS data

All seven variants were detected by the initial WGS experiments but not recognized as pathogenic. At the time of the first data annotation in 2014, five of the seven genes (*AP3B2*, *HMGA2*, *KCNB1*, *SON*, and *WAC*) were not recognized in OMIM to cause human disease [17–22]. In one case (*SMAD6*), the phenotypes of the individuals reported in the literature did not overlap the clinical presentation of our patient. Variants in *SMAD6* have since been associated with craniosynostosis, in conjunction with incomplete penetrance [23]. For the SNV in *ZNF711*, there was felt to be insufficient evidence in support of pathogenicity at the time of the initial review. The identification of additional cases has now bolstered the argument for causality [24].

Discussion

A diagnostic rate of ~10% after reanalysis is consistent with that of a previous study that reanalyzed singleton WES data after a 1–3 year period (10%; 4 of 40) [7]. Reassessment of pre-existing data can be performed rapidly relative to performing new genetic testing. Currently, a main advantage is the ability to immediately capitalize on the discovery of new disease genes. For example, since 2015 the first three probands have been reported in the literature with *HMGA2* sequence variants and a phenotype resembling Silver–Russell syndrome [17, 18]. The phenotype of Case 1096 was notable for intrauterine growth restriction, short stature, and other features (Supplemental file). Clinical genetic testing included CMA, methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) for 11p15.5 gene dosage and H19 hypomethylation, short tandem repeat analysis with DNA markers on chromosome 7 for uniparental disomy, and sequencing of *PIK3R1*. All results were negative or normal. Reanalysis of his WGS data identified the novel loss-of-function variant in exon 5 of *HMGA2* (Table 1). Databases used for clinical annotation lag behind the fast pace of the published literature, and many diagnostic laboratories cannot afford to frequently validate new pipelines with updated downloads from these databases. The variant could have been missed, for example, because as of December 2017 there is no phenotype associated with *HMGA2* in either OMIM or CDG. This further emphasizes the importance of periodic reanalysis.

This study was not designed to compare WGS with WES. Although the SNVs in Table 1 could potentially have been found with WES, other diagnoses we have made with WGS were (or would have been) missed [1, 2]. In one study that re-annotated and reanalyzed six undiagnosed WES trios, it was necessary in some cases to add coverage to detect the causal variant [8]. There are several reasons why

periodic reanalysis of WGS data may result in more diagnoses over time than WES, such as: (i) more uniform and more comprehensive coverage, including within the exome; (ii) our improving ability to interpret variation in regulatory regions, deep intronic regions, and non-coding DNA; and, (iii) the superior detection of structural variation. More generally, advances in clinical annotation of genome-wide sequencing data [25], a trio (as opposed to singleton) design, and pairing WGS with ancillary RNA sequencing, may all further increase the diagnostic yield in our cohort.

These findings highlight periodic reanalysis as yet another advantage of genomic sequencing in the diverse paediatric population meeting criteria for CMA. The revised diagnostic yield of 41% in this cohort is similar to that observed in the second WGS study from the Genome Clinic (42%), which involved a heterogeneous group of 103 patients recruited from non-genetic paediatric subspecialty clinics and where data were annotated in 2016 [2]. We recommend reanalysis of an individual's genome-wide sequencing data every 1–2 years until diagnosis, or sooner if their phenotype evolves. This should be part of pre-test counselling. Detailed phenotyping and the opportunity for reverse-phenotyping are essential, as WGS is both hypothesis-free and hypothesis generating. One limitation of clinical WGS is the relative shortage of those trained to medically interpret a genome. Another major practical and financial consideration is long-term data storage. With decreasing costs of sequencing and advancements in sequencing technology, it may become cost effective in time to periodically re-sequence a banked DNA sample rather than store and reanalyze pre-existing WGS data. Regardless of these factors, the data suggest that utilization of WGS early in the diagnostic odyssey warrants further consideration in routine clinical genetics practice.

Acknowledgements We thank the patients and families whose participation made this project possible, the many healthcare providers involved in their care, and the staff at The Centre for Applied Genomics. We also thank the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. A full list of contributing groups can be found at <http://gnomad.broadinstitute.org/about>. This study was funded by the Centre for Genetic Medicine, The Centre for Applied Genomics, The Hospital for Sick Children, Genome Canada, and the University of Toronto McLaughlin Centre. R.D.C. holds the Women's Auxiliary Chair in Clinical and Metabolic Genetics at The Hospital for Sick Children. S. W.S. holds the Canadian Institutes for Health Research (CIHR) GlaxoSmithKline Endowed Chair in Genome Sciences at The Hospital for Sick Children and the University of Toronto.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Stavropoulos DJ, Merico D, Jobling R, et al. Whole genome sequencing expands diagnostic utility and improves clinical management in pediatric medicine. *NPJ Genom Med.* 2016;1:15012.
2. Lionel AC, Costain G, Monfared N, et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet Med.* E-published 2017 Aug 3. doi: 10.1038/gim.2017.119.
3. Gilissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature.* 2014;511:344–7.
4. Taylor JC, Martin HC, Lise S, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet.* 2015;47:717–26.
5. Willig LK, Petrikin JE, Smith LD, et al. Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. *Lancet Respir Med.* 2015;3:377–87.
6. Costain G, Shugar A, Krishnan P, et al. Homozygous mutation in PRUNE1 in an Oji-Cree male with a complex neurological phenotype. *Am J Med Genet A.* 2017;173:740–3.
7. Wenger AM, Guturu H, Bernstein JA, et al. Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. *Genet Med.* 2017;19:209–14.
8. Need AC, Shashi V, Schoch K, et al. The importance of dynamic re-analysis in diagnostic whole exome sequencing. *J Med Genet.* 2017;54:155–6.
9. 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;89:275–84.
10. Tan TY, Dillon OJ, Stark Z, et al. Diagnostic impact and cost-effectiveness of whole-exome sequencing for ambulant children with suspected monogenic conditions. *JAMA Pediatr.* 2017;171:855–62.
11. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 2016;44:D862–868.
12. Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet.* 2017;136:665–77.
13. Solomon BD, Nguyen AD, Bear KA, et al. Clinical genomic database. *Proc Natl Acad Sci USA.* 2013;110:9851–5.
14. Pal LR, Kundu K, Yin Y, et al. CAGI4 SickKids clinical genomes challenge: a pipeline for identifying pathogenic variants. *Hum Mutat.* 2017;38:1169–81.
15. Assoum M, Philippe C, Isidor B, et al. Autosomal-recessive mutations in AP3B2, adaptor-related protein complex 3 beta 2 subunit, cause an early-onset epileptic encephalopathy with optic atrophy. *Am J Hum Genet.* 2016;99:1368–76.
16. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17:405–24.
17. De Crescenzo A, Citro V, Freschi A, et al. A splicing mutation of the HMGGA2 gene is associated with Silver-Russell syndrome phenotype. *J Hum Genet.* 2015;60:287–93.
18. Abi Habib W, Brioude F, Edouard T, et al. Genetic disruption of the oncogenic HMGGA2-PLAG1-IGF2 pathway causes fetal growth restriction. *Genet Med.* E-published 2017 Aug 10. doi: 10.1038/gim.2017.105.
19. Torkamani A, Bersell K, Jorge BS, et al. De novo KCNB1 mutations in epileptic encephalopathy. *Ann Neurol.* 2014;76:529–40.
20. Kim JH, Shinde DN, Reijnders MR, et al. De novo mutations in SON disrupt RNA splicing of genes essential for brain development and metabolism, causing an intellectual-disability syndrome. *Am J Hum Genet.* 2016;99:711–9.
21. DeSanto C, D'Aco K, Araujo GC, et al. WAC loss-of-function mutations cause a recognisable syndrome characterised by dysmorphic features, developmental delay and hypotonia and recapitulate 10p11.23 microdeletion syndrome. *J Med Genet.* 2015;52:754–61.
22. Lugtenberg D, Reijnders MR, Fenckova M, et al. De novo loss-of-function mutations in WAC cause a recognizable intellectual disability syndrome and learning deficits in *Drosophila*. *Eur J Hum Genet.* 2016;24:1145–53.
23. 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.
24. van der Werf IM, Van Dijk A, Reyniers E, et al. Mutations in two large pedigrees highlight the role of ZNF711 in X-linked intellectual disability. *Gene.* 2017;605:92–98.
25. Steward CA, Parker APJ, Minassian BA, et al. Genome annotation for clinical genomic diagnostics: strengths and weaknesses. *Genome Med.* 2017;9:49.
26. Nambot S, Thevenon J, Kuentz P, et al. Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: substantial interest of prospective annual reanalysis. *Genet Med.* E-published 2017 Nov 2. doi: 10.1038/gim.2017.162.
27. 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.* E-published 2018 Jan 11. doi: 10.1038/gim.2017.246.