**Prospective evaluation of genome sequencing versus standard-of-care as a first molecular diagnostic test**

Brockman, Austin-Tse et al.

Supplementary Information

**Figure S1.** cWGS analysis strategy

**Figure S2.** Study exclusion reasons

**Figure S3.** SOC genetic tests ordered for all patients enrolled

**Table S1.** Eligibility criteria

**Table S2.** HPO terms by clinic

**Table S3.** Diagnostic variants reported on SOC – Technical comparison to cWGS

 *Separate excel file*

**Table S4.** Retrospective analysis of SOC-reported CNVs in cWGS data

**Table S5.** Diagnostic variants reported on cWGS – Technical comparison to SOC

 *Separate excel file*

**Table S6.** cWGS and SOC Results

 *Separate excel file*

**Figure S1. cWGS analysis strategy.**



Each case underwent a two-tiered filtration strategy composed of both genotype-driven and phenotype-driven analyses. A) Phenotype-driven analyses are designed to capture all rare variants in genes relevant to the patient phenotype. Relevant gene lists were manually curated for each indication from literature and database searches, and varied in size depending on the phenotype. B) Genotype-driven analyses are designed to capture all highly suspicious variation in a patient’s genome, including previously published disease-causing variants found in HGMD and ClinVar, loss-of-function variants in ~5000 medically actionable genes (a custom-generated list designed to capture all genes that have been reported in association with human disease), and all internally classified pathogenic and likely pathogenic variants. All variants returned by these filtration criteria were reviewed for disease causality and relevance to the patient phenotype. Abbreviations: DM – Disease-causing mutation; HGMD – Human Gene Mutation Database; LOF – loss of function; MAF – maximum minor allele frequency in gnomAD: Path. – Pathogenic; Likely Path. – Likely pathogenic

*cWGS Analysis: Sequence Variants*

cWGS sequence variant analysis methods are described in Figure S1. For patients who elected to receive secondary findings, we additionally screened for previously reported and novel variants in 59 genes of medical significance that may be unrelated to the patient phenotype1. Furthermore, when patients were enrolled as trios, an additional analysis was performed to identify any variants that were identified in the proband but absent from both parents (de novo variation).

The evidence for gene-disease validity and relevance to the patient phenotype was evaluated for each variant resulting from the filtering strategies above and variants were classified based on ACMG/AMP criteria with ClinGen rule specifications ([http://www.clinicalgenome.org/working-groups/sequence-variant-interpretation](https://www.clinicalgenome.org/working-groups/sequence-variant-interpretation)) 2,3.

Variants were included on the cWGS report if they met one of the following criteria: (1) VUS/LP/P in a dominant gene related to the patient’s phenotype, (2) VUS/LP/P biallelic variants in a recessive gene related to the patient’s phenotype, (3) monoallelic VUS – Favor Pathogenic/LP/P in a recessive gene related to the patient’s phenotype, (4) LP/P variants in a gene related to a documented family history of disease, or (5) LP/P variants in one of the 59 secondary findings genes recommended by the American College of Medical Genetics and Genomics (Kalia 2017). All clinically reported variants were confirmed via an orthogonal technology (Sanger sequencing for short length sequence variants.

*cWGS Analysis: Structural variants*

The homozygous STRC/CATSPER2 deletion identified in participant 1170 was suspected based on poor coverage of the STRC gene by cWGS and manual review of the STRC gene region using the Integrative Genomics Viewer4. This variant was orthogonally confirmed via droplet digital PCR prior to reporting5.

For participants whose SOC results included a CNV, retrospective analysis was performed on the cWGS data using a research-grade structural variant (SV) calling pipeline as previously described6. The SV pipeline outputs were screened exclusively for the presence of the SOC-reported variant. Given that precise CNV boundaries were not defined by the SOC testing methods, successful WGS detection of the SOC-reported variant was defined as the presence of an overlapping CNV of similar size.

**Figure S2. Study exclusion reasons.**

**Figure S3. SOC genetic tests ordered for all patients enrolled.**

**Table S1.** Eligibility criteria

|  |  |
| --- | --- |
| **Inclusion Criteria** | **Exclusion Criteria** |
| Pediatric and adult (≥ 3 months of age) | Non-English speaking |
| Genetics evaluation and genetic testing ordered at MGH | Prior genetic testing for current referral indication |
| Have a *suspected genetic disorder* in which the genetic cause is unknown |  |
| < 18 years of age \* |  |

*\*additional eligibility criteria to be enrolled as a trio*

**Table S2. HPO terms by clinic**

|  |  |  |  |
| --- | --- | --- | --- |
|   | **Mean Number of HPO Terms** |   |   |
| **Clinic** | **Total** | **Primary - phenotype** | **Non-primary - phenotype** | **No. Body Systems** | **No. Genes** |
| ATX | 9.40 | 6.50 | 2.90 | 4.50 | 228.30 |
| CGP | 4.56 | 1.79 | 2.76 | 2.56 | 214.97 |
| ETG | 6.63 | 2.63 | 4.00 | 3.38 | 154.75 |
| GIC | 2.92 | 1.58 | 1.50 | 2.25 | 401.64 |
| MGP | 7.50 | 4.83 | 2.93 | 4.37 | 219.04 |
| PUL | 7.14 | 6.71 | 0.43 | 4.14 | 331.29 |

Across different clinics including both study arms (n=), there were significantly different mean numbers of total HPO terms, primary HPO terms and body systems (P<0.001), but the mean number of non-primary HPO terms was not significantly different (P=0.2). This analysis was done using ANOVA.

**Table S4. Retrospective analysis of SOC-reported CNVs in cWGS data**

|  |  |  |
| --- | --- | --- |
| **Study ID** | **CNV reported by SOC** | **Corresponding CNV called by WGS SV algorithm** |
| 152CGS | 22q11.2 duplication (2Mb)\* | 22q11.21 duplication (2.45Mb, chr22:18,698,501-21,514,289x3) |
| 170CGS | *STRC/CATSPER2* Chr15:g.(?\_43,892,948)\_(43,931,260\_?)del, homozygous\* | Deletion encompassing the PPIP5K1, CKMT1B, STRC, CATSPER2, and CKMT1A genes (chr15:43,853,072-43,988,198x0) |
| 204CGS | 15q11.2 duplication (648.23 kb)\* | 15q11.2 duplication (519.76kb, chr15:22,750,068-23,268,854x3) |
| 204CGS | 5p15.33-p15.32 duplication (1.102 Mb)  | 5p15.33-p15.32 duplication (1.122Mb, chr5:4,313,487-5,435,139x3) |
| 156CGS | 15q11.2 duplication (min: arr 15q11.2(22,652,061 - 23,586,276)x3 (934.216 kb), max: arr 15q11.2(22,587,189 - 23,656,935)x3 (1.070 Mb)) | 15q11.2 duplication (664.76kb, chr15:22,748,962-23,407,341x3) |
| 70CGS | *DOCK8*, NM\_203447.3: Gain (Exons 8-48) | 9p24.3 duplication (*DOCK8* and *KANK1*, chr9:317,687-518,548x3) |

\*Variant considered to be diagnostic by SOC.

All genomic coordinates are GRCh37

1. Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genetics in Medicine*. 2017;19(2):249-255. doi:10.1038/gim.2016.190

2. Strande NT, Riggs ER, Buchanan AH, et al. Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am J Hum Genet*. 2017;100(6):895-906. doi:10.1016/j.ajhg.2017.04.015

3. 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(5):405-424. doi:10.1038/gim.2015.30

4. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26. doi:10.1038/nbt.1754

5. Mandelker D, Amr SS, Pugh T, et al. Comprehensive diagnostic testing for stereocilin: an approach for analyzing medically important genes with high homology. *J Mol Diagn*. 2014;16(6):639-647. doi:10.1016/j.jmoldx.2014.06.003

6. Collins RL, Brand H, Karczewski KJ, et al. A structural variation reference for medical and population genetics. *Nature*. 2020;581(7809):444-451. doi:10.1038/s41586-020-2287-8