UC Irvine
UC Irvine Previously Published Works

Title

Increased diagnostic yield from negative whole genome-slice panels using automated reanalysis.

Permalink

https://escholarship.org/uc/item/1sq768cr

Journal

Clinical Genetics, 104(3)

Authors

Berger, Seth
Pitsava, Georgia
Cohen, Andrea
et al.

Publication Date

2023-09-01

DOI

10.1111/cge.14360

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Clin Genet. Author manuscript; available in PMC 2024 September 01.

Published in final edited form as:

Clin Genet. 2023 September ; 104(3): 377–383. doi:10.1111/cge.14360.

Increased diagnostic yield from negative whole genome-slice panels using automated reanalysis

Seth I. Berger, MD, PhD[1,2], **Georgia Pitsava**[2], **Andrea J. Cohen**[1,2,3], **Emmanuèle C. Délot**[2,4], **Jonathan LoTempio**[2,4], **Erin Hallie Andrew**[1,2], **Gloria Mas Martin**[5], **Sofia Marmolejos**[2], **Jessica Albert**[6], **Beatrix Meltzer**[6], **Jamie Fraser**[1,2], **Debra S. Regier**[1,2], **Amanda H. Kahn-Kirby**[5], **Erica Smith**[5], **Susan Knoblach**[2], **Arthur Ko**[2], **Vincent A. Fusaro**[5], **Eric Vilain**[2,4,7]

[1]Children's National Rare Disease Institute, Division of Genetics and Metabolism, Washington, DC, USA

[2]Center for Genetic Medicine Research, Children's National Research Institute, Washington, DC, USA

[3]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

[4]Department of Genomics and Precision Medicine, George Washington University, Washington, DC, USA

[5]Invitae Corporation, San Francisco, CA, USA

[6]Molecular Diagnostics Laboratories, Children's National Hospital, Washington, DC, USA

[7]Institute for Clinical and Translational Science, University of California, Irvine, CA, USA

Abstract

We evaluated the diagnostic yield using genome-slice panel reanalysis in the clinical setting using an automated phenotype/gene ranking system.

We analyzed whole genome sequencing (WGS) data produced from clinically ordered panels built as bioinformatic slices for 16 clinically diverse, undiagnosed cases referred to the Pediatric Mendelian Genomics Research Center, an NHGRI-funded GREGoR Consortium site. Genome-wide reanalysis was performed using Moon™, a machine-learning-based tool for variant prioritization.

In five out of 16 cases, we discovered a potentially clinically significant variant. In four of these cases, the variant was found in a gene not included in the original panel due to phenotypic

Corresponding author: Seth I. Berger, MD, PhD, Rare Disease Institute at the Research & Innovation Campus, Washington, DC, 20012, 202-545-2500, sberger@childrensnational.org.

Author Contributions:

SIB and GP wrote inital draft. SIB, GP, AJC, JL, GMM, ES and ECD analyzed the genomic data. EHA, GP, and SM oversaw enrollment and participant data collection. SIB, JLF, and DSR provided clinical evaluation of cases. JA and BM evaluated clinical laboratory data and developed clinical lab assays. SK and AHK developed RNA assay for case 3 and AK analyzed the RNA data. SIB, VAF, and EV conceived of and supervised the project. All authors edited and approved the final manuscript.

Ethics Declaration

Informed consent was obtained from every individual (or legal representative) whose data is included.

Conflict of interest

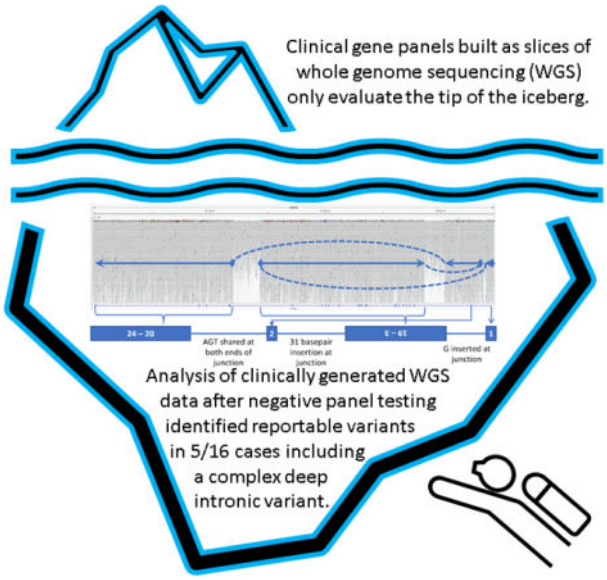
A.H.K.K., V.A.F., G.M. and E.S. worked for Invitae. Other authors have no conflict of interest to report.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

expansion of a disorder or incomplete initial phenotyping of the patient. In the fifth case, the gene containing the variant was included in the original panel, but being a complex structural rearrangement with intronic breakpoints outside the clinically analyzed regions, it was not initially identified.

Automated genome-wide reanalysis of clinical WGS data generated during targeted panels testing yielded a 25% increase in diagnostic findings and a possibly clinically relevant finding in one additional case, underscoring the added value of analyses versus those routinely performed in the clinical setting.

Graphical Abstract



Whole genome sequencing (WGS) data produced from clinically ordered panels built as bioinformatic slices for 16 clinically diverse, undiagnosed cases was reanalyzed using a machine-learning-based tool for variant prioritization yielded additional potentially clinically significant variants in five out of 16 cases including a complex variant not detectable by exome.

Keywords

whole genome sequencing; panel testing

Introduction

For rare disease diagnosis, phenotype-driven analysis of whole genome sequencing (WGS) requires prioritization of likely clinically relevant genetic variants. Variant review is time-consuming and relies on accurate curation and phenotypic interpretation. Alternative testing strategies involve gene panels where analysis focuses on predefined lists of genes with established clinical validity. Panels are powerful clinical diagnostic tools due to cost effectiveness, straightforward pretest counseling, and relative ease of analysis [1]. However, they provide limited assessment leading to missed diagnostic opportunities.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Many panels are built as bioinformatic slices of larger exome- or genome-based assays. Benchwork is test-agnostic, but analysis focuses on specific genes. Sequencing data remains available for reanalysis or evaluation of additional genes [2].

We demonstrate the potential of automated reanalysis with variant calling for both small and structural variants coupled with sophisticated machine-learning-based variant annotation and prioritization for patients with complex phenotypes undiagnosed after clinical genome-slice panels.

Methods

Clinical Samples

Sixteen genetically undiagnosed cases after genome-slice panels performed at Children's National Hospital Clinical (CNH) Molecular Laboratory between 2021–2022 (genes originally analyzed are shown in Supplemental Table 1) were referred to the Pediatric Mendelian Genomics Research Center (PMGRC) (Table 1). Phenotypic data were summarized using Human Phenotype Ontology (HPO) terms.

Variant prioritization and analysis with Moon™

WGS bam files were processed to fastq files and re-aligned to hg19 using bwa-mem2. Variants were called using octopus [3]. Variant files, HPO terms, ages of onset and sex were uploaded to Moon™ (Invitae, San Francisco, CA) [4, 5]. Variants were classified according to ACMG guidelines [6].

Variant validation by high-coverage RNA sequencing was performed for Case 3 (Methods in supplemental material).

Results

Automated reanalysis of WGS using MOON™ after negative clinical panels yielded potentially diagnostic findings in five cases (Eleven remaining cases are described in Supplemental File).

Case 1

A five-year-old male presented with dermatitis. Oral steroids, not topical steroids or antihistamines, slowed progression and prevented new eruptions. He was otherwise healthy, without exposures or acute illnesses around onset of symptoms.

Testing for two genes related to erythrokeratodermia variabilis was negative. Reanalysis of WGS data revealed two pathogenic loss-of-function (LoF) variants in the *FLG* gene associated with autosomal recessive ichthyosis vulgaris, explaining the xerosis and hyperlinear palms (Table 1). This gene was not included on the original panel as a narrower phenotype was considered but was the highest ranked candidate by Moon™ due to the HPO term for dermatitis. This gene is often excluded on panel testing due to highly homologous regions which confound exome-based variant calling. Variants were confirmed by Sanger sequencing in a CLIA lab.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Case 2

A two-month-old female presented with left ventricular posterior wall hypertrophy and failure to thrive. A 100-gene cardiomyopathy and arrhythmia panel reported carrier status for variants in *ACADVL* and *TRDN*, inconsistent with clinical and biochemical testing. At four months, a 19-gene interstitial lung disease panel ordered for new phenotypic findings returned negative.

Reanalysis of the WGS data revealed a maternally inherited heterozygous pathogenic variant in *NARS2* (deletion, exons 8–9) and a paternally inherited heterozygous VUS in *NARS2* (p.Arg250Gln). *NARS2* encodes an aminoacyl-tRNA-synthetase; biallelic variants cause combined oxidative phosphorylation deficiency 24 (COXPD24; OMIM #616239). *NARS2* was not evaluated by previous genetic testing as the panels targeted specific phenotypes, which presented as components of the complex presentation. Moon ranked it high due to phenotypic terms generalized hypotonia, laryngomalacia, supraventricular tachycardiac, respiratory distress, and status epilepticus and was considered consistent with the clinical presentation.

Case 3

A nine-year-old male presented with acute liver failure, acute kidney injury, and history of autism spectrum disorder. While recovering in an intensive care unit, he became agitated and received haloperidol; he developed torsades de pointes (TdP) leading to ventricular tachycardic arrest.

An initial 70-gene cholestasis panel, specifically to evaluate for *POLG* variants, had not revealed any causative variants. An 81-gene cardiac arrhythmia panel reported two VUS, neither of which were considered causative based on population frequencies and review in ClinVar.

Genome reanalysis identified a hemizygous VUS in *SLC6A8*, a gene associated with X-linked cerebral creatine deficiency syndrome 1 (CCDS1, OMIM #300352). This intronic 2-nucleotide deletion maintains the canonical splice site. SpliceAI scores the variant at 0.18, inconclusive for splicing effects. Magnetic resonance spectroscopy demonstrated decreased creatine peak, pathognomonic for cerebral creatine deficiency, and elevated urinary creatine to creatine ratio and guanidinoacetate, confirming the biochemical diagnosis (Figure 1a). RNAseq analysis (Figure 1b/c) revealed aberrant splicing in 10% of *SCL6A8* transcripts and expression level z-score of −2.3 suggestive of nonsense mediated decay of the transcript and clinically validated (MNG labs). This was considered uncertain significance and suggested there might be some residual functional creatine transporter expression. Supplementation with creatine, arginine, and glycine led to parental report of marked improvement in concentration, planning behaviors, and decreased aggressive behaviors at school. While most individuals with *SLC6A8* loss-of-function do not respond to treatment due to inability to transport the creatine into the brain, residual expression of the typical transcript from the leaky splice variant may permit response to therapy.

The variant was identified based on the HPO term for autism. However, recent reports suggest that CCDS1 is associated with QT prolongation [9]. This gene is not routinely

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

on cardiac arrhythmia panels and thus not reported in the original testing. By our request, *SLC6A8* was associated with QT Interval Prolongation (HP:0001657) in the April 2022 HPO release. This is the first case, to our knowledge, of drug-induced TdP in a patient with CCDS1.

Case 4

A newborn female presented with clinical signs of oculocutaneous albinism. A 22-gene molecular albinism panel identified a single heterozygous pathogenic variant in the *OCA2* gene consistent with carrier status, but not diagnostic for this recessive condition. Single gene sequencing and deletion duplication analysis by a different clinical laboratory did not identify any additional changes.

Reanalysis of WGS data confirmed the previously reported variant, but also revealed a complex rearrangement of the other *OCA2* allele with deep intronic breakpoints (Figure 2). This pathogenic rearrangement and was missed by clinical testing because it is exon-copy-neutral. The evidence for the variant was breakpoints deep in introns 1, 2, and 19, which was listed on Moon's structural variant short list. It was called by MANTA [7] and SvABA [8] and apparent on direct review in IGV [9], demonstrating deletions within these introns and split reads allowing reconstruction of the rearrangement. Junctional PCR over the breakpoints confirmed the rearrangement and its maternal inheritance, *in trans* with the paternally inherited missense variant, providing a definitive diagnosis (Supplemental Figure 1).

Case 5

A 14-year-old male presented with digital clubbing, idiopathic liver failure (requiring liver transplantation), and restrictive lung disease. Copper/ceruloplasmin levels to screen for Wilson's disease were normal, as was alpha-1 antitrypsin genotyping. A sequencing panel of 15 genes associated with early cirrhosis and lung disease resulted negative.

Moon™ ranked highly a VUS in *UROD*, a gene for porphyria cutanea tarda type 2 (PCT2). There are 9 heterozygotes in gnomAD *v2.1.1* and it affects the same amino acid residue as a known *likely pathogenic* PCT2 variant. Although early cirrhosis may seen in PCT2, age was much younger than typically presentation. Urine porphyrin fractionation was inconclusive, potentially due to transplanted liver function. Red blood cell uroporphyrinogen decarboxylase (UROD) activity, a test unaffected by liver transplant, demonstrated decreased UROD activity. While this confirms that the patient's variant has functional implications, it does not provide a definitive diagnosis. Most individuals with PCT2 remain asymptomatic into late adulthood and require additional disease-causing factors (e.g. alcoholism, iron overload, or infections). It is unclear if the variant is related to the patient's liver failure, a risk factor exacerbating another process, or an unrelated finding. It was paternally inherited from an asymptomatic father, but paternal urine and enzyme testing could not be performed.

Discussion

We reanalyzed WGS data generated during clinical gene panel testing of 16 patients with suspected genetic disorders using an automated, machine-learning-based variant-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

prioritization tool[5, 10]. In five of the 16 patients (31%), reanalysis yielded additional reportable findings. Our study serves as proof-of-concept that genome analysis after negative panels, supported by sophisticated variant annotation tools, can increase diagnostic yield in clinical settings and argues for increased use of WGS as a backbone test for genetic panels.

In four of the five cases where our reanalysis uncovered reportable variants, variants were present in genes not included in the originally tested panel, either due to phenotypic expansion of the disorder or incomplete initial phenotyping, highlighting the importance of maintaining broad differentials from inception of genetic evaluations. In the remaining case, the causative gene was on the original panel but the variant was not detected because it was a complex structural rearrangement with deep intronic breakpoints falling outside the clinical reporting regions. Loftus *et al.* recently identified this complex rearrangement in approximately 3% of OCA2 cases based on a custom sequencing approach [11]. We demonstrated that WGS with appropriate bioinformatics can detect this important variant not detectable by exome sequencing.

We also identified a treatable diagnosis and important phenotypic expansion in case 3. CCDS1 is a known cause of autism and neurodevelopmental delay [12], but this patient also developed TdP, of which QT prolongation is a known risk factor [13]. At the time of analysis, CCDS1 did not have a known arrhythmia association, which has since been published [14]. RNA analysis provided evidence for low level expression of typical *SLC6A8*, which explains the favorable response to creatine supplementation.

The re-analyses presented in our study are based on WGS. This overcomes known limitations of exome sequencing such as inability to detect variants in regulatory or intronic regions. Many clinical panels are based on exome slices [15], but our work shows that building clinical panels as slices of WGS provides added value. If clinical suspicion for specific diagnoses remains high after WGS analysis, specific targeted testing could be considered to address analytical limitations of genomes, such as mosaic variants, complex genomic regions, and short tandem repeats. Our approach of systematically providing high-fidelity panel-based testing built on a genome platform allows a rapid initial clinical screen of a predefined gene list while facilitating improved diagnostic rates through phenotype-driven unblinding of clinically generated WGS data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

Study was supported in part by the National Human Genome Research Institute of the National Institutes of Health grant U01HG011745, as part of the GREGoR Consortium. The content is solely the responsibility of the authors and does not represent official views of the National Institutes of Health.

Data availability

The genomic data sets are publicly available on AnVIL through the GREGoR Consortium data release.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

References

1. Li Y and Luo Y, Optimizing the evaluation of gene-targeted panels for tumor mutational burden estimation. Sci Rep, 2021. 11(1): p. 21072. [PubMed: 34702927]
2. Martin AR, et al. , PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat Genet, 2019. 51(11): p. 1560–1565. [PubMed: 31676867]
3. Cooke DP, Wedge DC, and Lunter G, A unified haplotype-based method for accurate and comprehensive variant calling. Nat Biotechnol, 2021. 39(7): p. 885–892. [PubMed: 33782612]
4. O'Brien TD, et al. , Artificial intelligence (AI)-assisted exome reanalysis greatly aids in the identification of new positive cases and reduces analysis time in a clinical diagnostic laboratory. Genet Med, 2022. 24(1): p. 192–200. [PubMed: 34906498]
5. Clark MM, et al. , Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. Sci Transl Med, 2019. 11(489).
6. Richards 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): p. 405–24. [PubMed: 25741868]
7. Chen X, et al. , Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics, 2016. 32(8): p. 1220–2. [PubMed: 26647377]
8. Wala JA, et al. , SvABA: genome-wide detection of structural variants and indels by local assembly. Genome Res, 2018. 28(4): p. 581–591. [PubMed: 29535149]
9. Robinson JT, et al. , Integrative genomics viewer. Nat Biotechnol, 2011. 29(1): p. 24–6. [PubMed: 21221095]
10. James KN, et al. , Partially automated whole-genome sequencing reanalysis of previously undiagnosed pediatric patients can efficiently yield new diagnoses. NPJ Genom Med, 2020. 5: p. 33.
11. Loftus SK, et al. , A custom capture sequence approach for oculocutaneous albinism identifies structural variant alleles at the OCA2 locus. Hum Mutat, 2021. 42(10): p. 1239–1253. [PubMed: 34246199]
12. Miller JS, et al. , Early Indicators of Creatine Transporter Deficiency. J Pediatr, 2019. 206: p. 283–285. [PubMed: 30579583]
13. Tisdale JE, et al. , Development and validation of a risk score to predict QT interval prolongation in hospitalized patients. Circ Cardiovasc Qual Outcomes, 2013. 6(4): p. 479–87. [PubMed: 23716032]
14. Levin MD, et al. , X-linked creatine transporter deficiency results in prolonged QTc and increased sudden death risk in humans and disease model. Genet Med, 2021. 23(10): p. 1864–1872. [PubMed: 34050321]
15. van Wijngaarden AL, et al. , Identification of known and unknown genes associated with mitral valve prolapse using an exome slice methodology. J Med Genet, 2020. 57(12): p. 843–850. [PubMed: 32277046]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

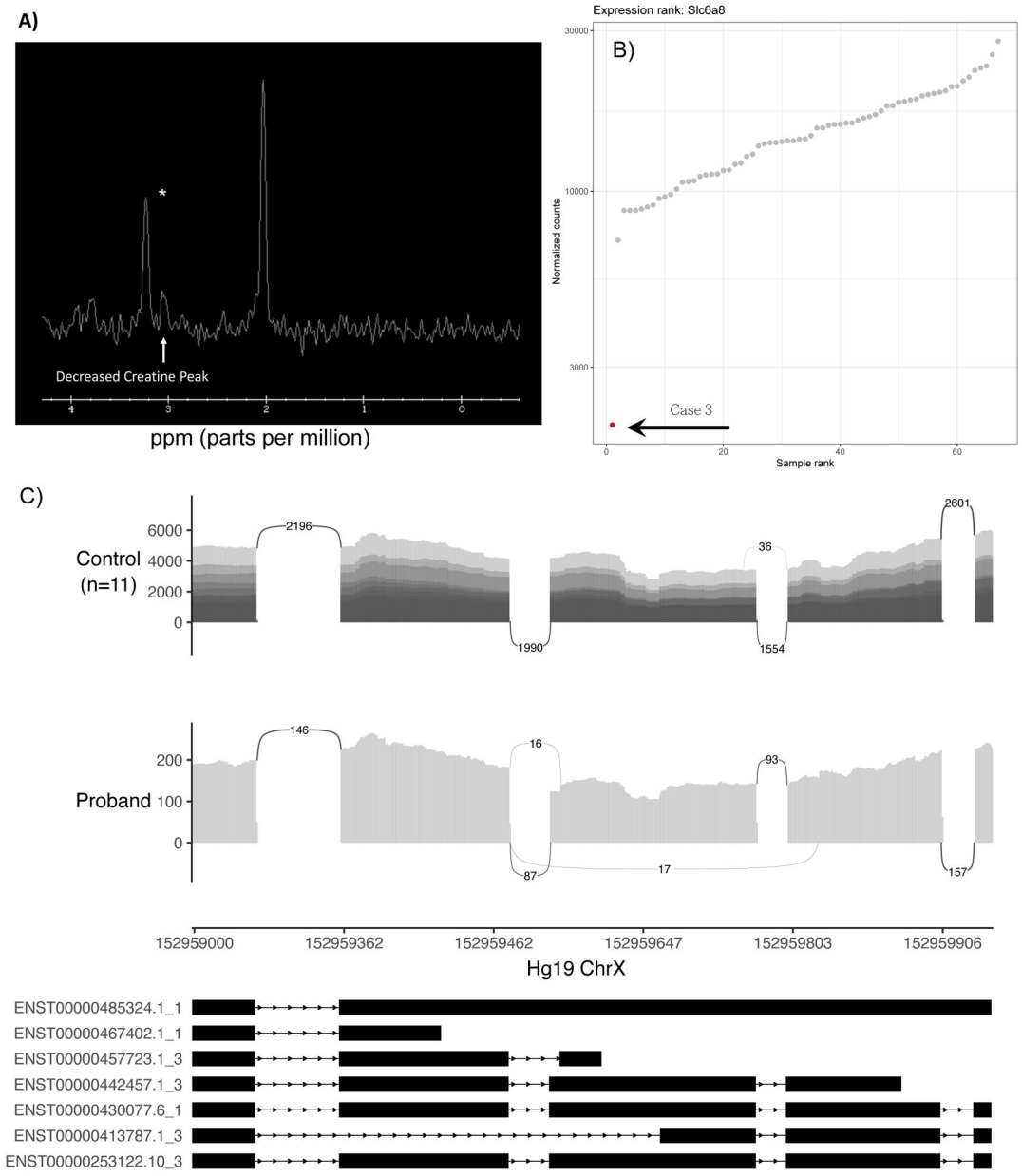


Figure 1.

A) Magnetic Resonance Spectroscopy over the basal ganglia in case 3 demonstrates markedly decreased creatine peak at 3 ppm (arrow) compared to the expected height labeled with an asterisk(*) pathognomonic for a cerebral creatine deficiency. B) Outlier expression analysis show significant decreased expression of *SLC6A8* in case 3 compared to control panel. C) Sashimi plot demonstrating skipping of exon 9 and 10-fold decreased expression of *SLC6A8* compared to counts from panel of 11 male controls.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

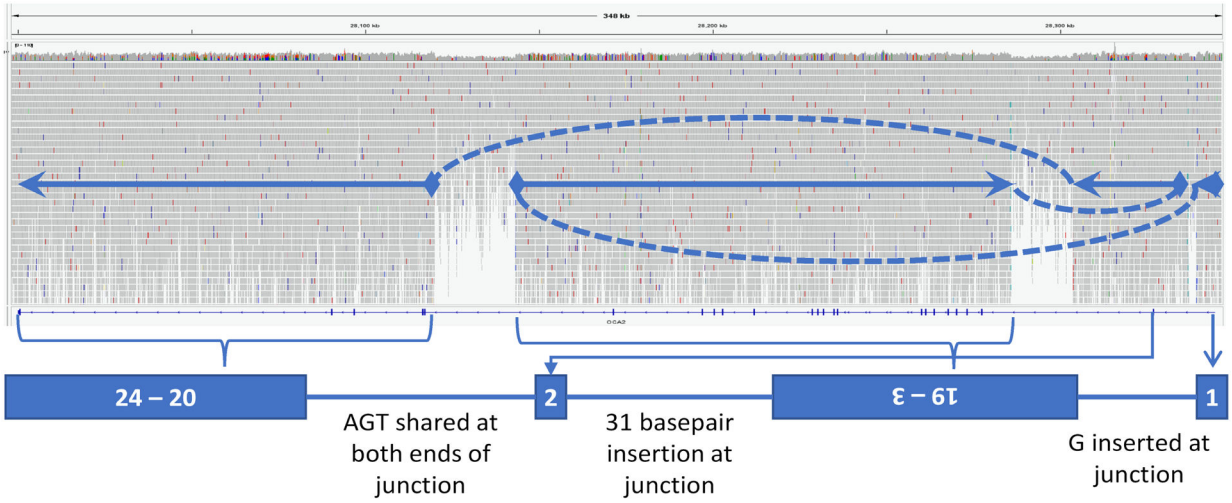


Figure 2.

In case 4, a complex rearrangement of *OCA2*, with multiple deep intronic breakpoints, was found *in trans* with a clinically reported pathogenic missense variant diagnostic for oculocutaneous albinism. 348 kb region of chr15 (hg19:chr15:27,998,021–28,346,461) is visualized in IGV, with schematic arrows showing the pattern of the rearrangement and dashed lines connecting split reads. Diagram below shows inversion of the segment containing exons 3–19 followed by subsequent insertion within the first intron.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Phenotypic description, previous genetic panel testing, and reanalysis outcome of 16 patients with negative clinical panel testing

Case #	PMGRC ID	Clinical description	Panel ordered (# of genes)	Reanalysis outcome
1	32-32-0	Eczematoid dermatitis on the left arm, which subsequently spread to the entire body, sparing face and scalp. Numerous scattered thin arcuate plaques with erythematous border (1–20 cm) and perivascular and interstitial dermatitis with neutrophils on biopsy. Ichthyosiform xerosis involving all extremities with prominent lines on the dorsal hands and fine scaling on the palms and soles. Noncontributory family history.	Erythrokeratodermia variabilis panel (2)	• Gene: *FLG* • Variants: NM_002016.2, c.1297_1298delGA, p.Asp433fs and c.2282_2285delCAGT, p.Ser761fs • Interpretation: Diagnostic
2	88-88-0	Critically ill neonate with multisystemic disease. Noncontributory family history. 4 month old: new findings of interstitial lung disease and refractory seizures.	• Cardiomyopathy & arrhythmia panel (100) Carrier status for pathogenic variant in *ACADVL* (NM_000018.4, c.1385dupG, p.Met463fs) and VUS in *TRDN* (NM_006073.4, c.360 A>T, p.Glu120Asp) • Interstitial lung disease panel (19)	• Gene: *NARS2* • Variants: Deletion, Exons 8–9, NM_024678.6, c.749G>A, p.Arg250Gln • Interpretation: Diagnostic
3	205-205-0	Acute liver failure, acute kidney injury, history of autism spectrum disorder	Arrhythmia panel (81) Cholestasis panel (70)	• Gene: *SLC6A8* • Variant: NM_005629.4, c.1255–3_1255–2delCA, Intron 8 • Interpretation: Diagnostic
4	212-212-0	Oculocutaneous albinism	Albinism gene panel (22)	• Gene: *OCA2* • Variants: Missense variant NM_000275.3, c.1465A>G, p.Asn489Asp and complex rearrangement with deep intronic breakpoints • Interpretation: Diagnostic
5	209-209-0	Digital clubbing, idiopathic liver failure (s/p transplantation), restrictive lung disease, coagulopathy	Custom liver failure panel (15)	• Gene: *UROD* • Variant: NM_000374.5, c.577C>T, p.Arg193Cys • Interpretation: Possible risk factor
6	36-36-0	Global developmental delay, cutis aplasia	Single gene panel (1)	No new variants reported

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Case #	PMGRC ID	Clinical description	Panel ordered (# of genes)	Reanalysis outcome
7	101-101-0	Hemihypertrophy, nystagmus, R club foot	Beckwith Wiedemann syndrome panel (1)	No new variants reported
8	8-8-0	Massive splenomegaly and worsening liver failure, pre-B ALL in remission and Neurofibromatosis 1	Splenomegaly panel (220)	No new variants reported
9	79-79-0	Photosensitive dermatitis	Cutaneous porphyria panel (3)	No new variants reported
10	106-106-0	Febrile infection-related epilepsy syndrome	• Nuclear mitochondrial gene panel (349) • HLH panel (14)	No new variants reported
11	108-108-0	Autism, agenesis of the corpus callosum, anosmia, obesity, primary amenorrhea, OCD	Kallmann and related disorders panel (33)	No new variants reported
12	246-246-0	Infantile spasms, hypotonia, motor delays, exotropia	• Epilepsy panel (147) • Encephalopathy panel (214)	No new variants reported
13	239-239-0	Multiple cardiac defects, symmetrically short stature, heart vs heart/lung transplant	Panel for pediatric pulmonary hypertension (38)	No new variants reported
14	167-167-0	Congenital thumb aplasia, wide 1–2 gap, ASD	Panel for congenital limb anomalies (33)	No new variants reported
15	255-255-0	Hyperextensible skin, hypermobile joints, easy bruising, speech delay, autism with high cognitive abilities	Connective tissue panel (102)	No new variants reported
16	258-258-0	Neonate with multiple congenital defects	Panel including limb malformation genes, genes associated with VACTER-like syndrome and cleft palate genes (110)	No new variants reported

ASD atrial septal defect, *HLH* hemophagocytic lymphohistiocytosis, *OCD* obsessive compulsive disorder, *R* right, *VUS*, variant of uncertain significance