Phenotypic Expansion Illuminates Multilocus Pathogenic Variation

Ender Karaca, M.D.[1,*,#], **Jennifer E. Posey, M.D., Ph.D.**[1,*], **Zeynep Coban Akdemir, Ph.D.**[1], **Davut Pehlivan, M.D.**[1], **Tamar Harel, M.D., Ph.D.**[2], **Shalini N. Jhangiani, M.S.**[3], **Yavuz Bayram, M.D.**[1,#], **Xiaofei Song, B.S.**[1], **Vahid Bahrambeigi, M.S.**[1,4], **Ozge Ozalp Yuregir, M.D., Ph.D.**[5], **Sevcan Bozdogan, M.D.**[6], **Gozde Yesil, M.D.**[7], **Sedat Isikay, M.D.**[8], **Donna Muzny, M.S.**[3], **Richard A. Gibbs, Ph.D.**[1,3], and **James R. Lupski, M.D., Ph.D., D.Sc.(hon)**[1,3,9,10]

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston 77030, TX, USA

[2]Department of Genetic and Metabolic Diseases, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel

[3]Human Genome Sequencing Center, Baylor College of Medicine, Houston 77030, TX, USA

[4]Graduate Program in Diagnostic Genetics, School of Health Professions, University of Texas MD Anderson Cancer Center, Houston 77030, TX, USA

[5]Genetic Diagnosis Center, University of Health Sciences, City Hospital, Adana 01370, Turkey

[6]Department of Medical Genetics, Cukurova University Faculty of Medicine, Adana 01330, Turkey

[7]Department of Medical Genetics, Bezmialem University, Istanbul 34093, Turkey

[8]Department of Physiotherapy and Rehabilitation, Hasan Kalyoncu University, School of Health Sciences, Gaziantep 27410, Turkey

[9]Department of Pediatrics, Baylor College of Medicine, Houston 77030, TX, USA

[10]Texas Children's Hospital, Houston 77030, TX, USA

Corresponding Authors: Jennifer E. Posey, M.D., Ph.D., Assistant Professor, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, 604B, Houston, TX 77030, (713) 798-6873 (Lab), Jennifer.Posey@bcm.edu. James R. Lupski, M.D., Ph.D., D.Sc.(hon), Cullen Professor, Department of Molecular and Human Genetics and Professor of Pediatrics, Baylor College of Medicine, One Baylor Plaza, 604B, Houston, TX 77030, (713) 798-6530 (Office), (713) 798-3727 (Betty Fernandini), Fax: (713) 798-5073, jlupski@bcm.edu.
*co-first authors
#Current Addresses: **Ender Karaca M.D.** Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294, USA, **Yavuz Bayram M.D.** Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

Please see Supplementary Materials and Methods for details of clinical cohort, analysis approach and exome coverage metrics.

Abstract

Purpose: Multilocus variation, pathogenic variants in two or more disease genes, can potentially explain the underlying genetic basis for apparent phenotypic expansion in cases for which the observed clinical features extend beyond those reported in association with a "known" disease gene.

Methods: Analyses focused on 106 patients, 19 for which apparent phenotypic expansion was previously attributed to variation at known disease genes. We performed a retrospective computational re-analysis of whole exome sequencing data using stringent Variant Call File filtering criteria to determine whether molecular diagnoses involving additional disease loci might explain the observed expanded phenotypes.

Results: Multilocus variation was identified in 31.6% (6/19) of families with phenotypic expansion and 2.3% (2/87) without phenotypic expansion. Intrafamilial clinical variability within 2 families was explained by multilocus variation identified in the more severely affected sibling.

Conclusions: Our findings underscore the role of multiple rare variants at different loci in the etiology of genetically and clinically heterogeneous cohorts. Intrafamilial phenotypic and genotypic variability allowed a dissection of genotype-phenotype relationships in 2 families. Our data emphasize the critical role of the clinician in diagnostic genomic analyses and demonstrate that apparent phenotypic expansion may represent blended phenotypes resulting from pathogenic variation at more than one locus.

Keywords

INTRODUCTION

The diagnostic path leading from clinical evaluation to genetic diagnosis is not always direct, and physicians have historically relied on hallmark patterns of history and physical examination findings, or phenotypic features, to formulate a differential diagnosis and guide clinical investigation. However, even 'typical' clinical presentations can be misleading, as illustrated by early reports of Fitzsimmons syndrome describing 6 individuals with a seemingly unique combination of spastic paraplegia, skeletal anomalies, dysarthria, and mild developmental delay.[1–3] Whole exome sequencing (WES) ultimately demonstrated that despite the consistent set of shared phenotypic features, a single genetic locus did not explain the full phenotypic spectrum of the condition.[4] Multiple segregating pathogenic alleles at single[5,6] or multiple[7,8] loci in a family or clan can challenge both clinical and molecular diagnosis.

The occurrence of dual, or multiple, molecular diagnoses in an individual is not new, and reports dating back to the 1960s describe individuals with both hemolytic anemia due to glucose-6-phosphate dehydrogenase deficiency (OMIM #300908) and thalassemia minor, B-thalassemia, or sickle cell trait.[9,10] The true frequency of multiple molecular diagnoses has only begun to emerge with the availability of genome-wide investigations such as

chromosome microarray analysis (CMA)[11] and WES; larger-scale analyses of diagnostic WES in a clinical setting have revealed that 3.2 to 7.2% of unselected diagnostic WES cases have two or more molecular diagnoses contributing to their observed clinical phenotype. [8,12–17] This rate can be even higher, 12% (5/41), in selected cohorts.[18] From a clinical standpoint, recognition of multiple molecular diagnoses can have important implications for both treatment and estimates of familial recurrence risk.[8]

Genome-wide analyses have been instrumental for identifying molecular diagnoses in the setting of both atypical presentations of well-characterized syndromes,[7] as well as cases for which the set of observed phenotypic features extended beyond those previously reported in association with a particular locus, termed 'phenotypic expansion'.[19] Examples of phenotypic expansion attributed to previously-known disease loci abound, with a striking 18.2% (174/956) of discoveries reported by the NIH-funded Centers for Mendelian Genomics (CMG) representing apparent phenotypic expansion.[19] Given the relative frequency of both phenotypic expansion and multiple molecular diagnoses, we explored the hypothesis that cases of apparent phenotypic expansion may sometimes represent a blended phenotype resulting from variation at more than one locus. We tested this hypothesis using extensive phenotyping and personal genome data (WES) from 106 unrelated subjects from a well-characterized neurodevelopmental cohort, including 19 families with apparent phenotypic expansion that had previously been attributed to variation at a known disease gene.[20] A high prevalence of identity-by-descent enabled a focus on phenotypic and genotypic variability among siblings with shared genetic backgrounds.[20]

MATERIALS AND METHODS

Patients

Patients provided informed consent, including consent to publish photographs, under the Baylor Hopkins Center for Mendelian Genomics (BHCMG) protocol H-29697 approved by the Baylor College of Medicine (BCM) Institutional Review Board.

RESULTS

We investigated the hypothesis that an observed clinical phenotype, initially interpreted as phenotypic expansion associated with pathogenic variation at one known disease gene locus, may actually represent a blended phenotype due to multilocus variation. We compared the frequency of multilocus variation between cases with and without apparent phenotypic expansion, using a pre-existing, well-phenotyped cohort for which WES had previously demonstrated a high rate of molecular diagnoses (84.4%, 108/128 families).[20] Of 108 families for whom a molecular diagnosis was identified, 19 families were classified as having phenotypic expansion due to the presence of observed clinical features that extended beyond those previously reported in association with variation at the identified locus (Table 1).[20] In the present pilot study, we performed a reanalysis of exome variant data from 106 of these 108 families - two families were removed from analysis as the exome variant data were previously generated at another institution using a different platform and capture.

Using an objective, very stringent set of Variant Call File filtering criteria to parse the personal genome sequencing data, followed by manual review, we identified a set of rare variants predicted to be damaging to protein function in 51 loci in 37 families for which previously reported disease-gene associations indicated a phenotypic overlap with the proband's observed clinical features. From this set of variants, candidate etiologic variants were selected if they followed expected segregation patterns in affected and unaffected available relatives and were confirmed through independent orthogonal dideoxy Sanger sequencing. This final analysis step identified variants in 10 loci in 8 of 106 (7.5%) families (Table 1, Table S1). Notably, the originally reported variants/loci were parsed and not identified in 60.2% (65/108) of families using these strict filtering criteria which increase the specificity of personal genome analysis but decrease the sensitivity, highlighting the limitations of utilizing a purely computational approach to vcf data filtration prior to manual variant review of individual genomes within a large dataset. Variants at 8 loci were identified in 6 families (6/19, 31.6%) classified as having phenotypic expansion, and variants at 2 loci were identified in 2 families (2/87, 2.3%) without phenotypic expansion, demonstrating that apparent phenotypic expansion is more frequently associated with multilocus variation (two-tailed Fisher's exact p=0.0004). Notably, this analysis demonstrated a second variant in HOU1960 (BAB4892, *C2CD3*) that was interpreted to explain the full phenotype of the proband; because the phenotypic contribution of the originally identified *IQSEC2* variant was less clear, this family was not considered to have multilocus variation. Candidate novel disease genes were identified in 2 additional families previously classified as having phenotypic expansion, raising the potential rate of multilocus variation in cases of apparent phenotypic expansion to 42.1% (8/19); these candidate genes are discussed in further detail below. Taken together with the 3 cases of multilocus variation described in the original study (listed in Table 1)[20], the overall rate of multilocus variation for this neurodevelopmental cohort is 12.0% (13/108).

In the phenotypic expansion cohort, variants at a second or third locus involving known disease genes were observed in 6 families (31.6%, 6/19), and included *COL3A1* (HOU1341), *AMPD2* (HOU1838), *GPR126* (HOU1857), *TUBA1A* (HOU1965), *ALG8* and *CLN5* (HOU2293), and *MED17* and *SPG7* (HOU2337, Table 1). In each case, variation at the second and sometimes third locus provided a molecular etiology for the expanded objectively determined phenotype (Table 1). In the non-phenotypic expansion cohort, variants at a second locus related to the observed clinical phenotype were identified in 2 families: *GNAO1* (HOU2092) and *ATRX* (HOU2442). Unlike the phenotypic expansion cohort, the initially reported (Table 1) variants identified in the non-phenotypic expansion group were interpreted as causative of all phenotypic features in each case. Objective re-analysis of variant data demonstrated that each of these cases had an additional rare variant in a second gene associated with a subset of the observed clinical features. Notably, one case (HOU2268, BAB5990) with macrocephaly with extended subarachnoid spaces, a thin corpus callosum, and hypotonia was previously reported as having biallelic variation in *ZNF423*.[20] In this re-analysis, a *de novo AHDC1* frameshifting variant was identified and because of the comprehensive phenotypic match with *AHDC1* pathogenic variation,[21,22] the proband's reported clinical diagnosis of Joubert syndrome (OMIM 614844) was considered less likely. This case was not classified as having multilocus variation in the present study.

Intrafamilial variability allows a dissection of genotype-phenotype correlation

Two families initially classified as having phenotypic expansion in particular demonstrated phenotypic variability between affected siblings. In pedigree HOU2293, the proband (BAB3640) was reported to have intellectual disability, neurodevelopmental delay, microcephaly, and schizencephaly; this individual's sister (BAB3636), however, had intellectual disability, neurodevelopmental delay, and microcephaly without schizencephaly (Figure 1). Initial WES analysis demonstrated a molecular diagnosis of primary microcephaly (OMIM 251200) due to biallelic pathogenic, frameshifting variants in *MCPH1* in both siblings.[20] Reanalysis of WES data in the proband and sibling demonstrated additional molecular diagnoses of congenital disorder of glycosylation type Ih (OMIM 608104) and neuronal ceroid lipofuscinosis 5 (OMIM 256731) due to rare biallelic variants in *ALG8* and *CLN5* (Figure 1) present only in the more affected proband; notably, none of the three unaffected siblings were homozygous for these variants. Pathogenic variation at both of these loci has been associated with structural brain anomalies including brain atrophy,[23,24] and we propose that the additional mutational burden conferred by rare variation at *ALG8* and *CLN5* led to the schizencephaly phenotype identified in the more severely affected proband.

In pedigree HOU2337, the proband (BAB6228) was reported to have cortical atrophy, agenesis of the corpus callosum, microphthalmia, and severe neurodevelopmental delay characterized by the inability to walk, even with support. This individual's brother (BAB6229) had a milder phenotype with cortical atrophy, partial agenesis of the corpus callosum, and microphthalmia; despite neurodevelopmental delay, he was eventually able to walk independently. A maternally inherited hemizygous variant in *BCOR* consistent with a molecular diagnosis of syndromic microphthalmia 2 (OMIM 300166) was identified in both siblings (Figure 2).[20] Re-analysis of WES data further demonstrated biallelic variants in *MED17* and *SPG7,* consistent with molecular diagnoses of postnatal progressive microcephaly with seizures and brain atrophy (OMIM 613668) and autosomal recessive spastic paraplegia (OMIM 607259), in the more severely affected proband only (Figure 2). Although the proband does not demonstrate the full set of features for each of these diagnoses, both loci were hypothesized to provide genotypic correlation in the form of additional mutational burden to the more severe neurodevelopmental phenotype observed in the proband. Siblings pairs demonstrating phenotypic variability allow a true functional analysis of shared and private variation at genomic loci associated with Mendelian disease and support the contention that all affected individuals in a family or clan should be thoroughly investigated from a clinical phenotyping perspective.

Multilocus variation involving candidate novel disease genes

Variants at a second locus suggesting a potentially novel disease gene were observed in 2 families (10.5%, 2/19, Table 1). The proband in HOU2091 has frontotemporal dementia. Initial WES analysis revealed a heterozygous variant in *EPB41L1* associated with Mental Retardation AD 11 (OMIM 614257). On reanalysis, a rare heterozygous variant at a second locus, *SRR* encoding serine racemase, was identified in the affected proband (BAB5368), as well as the affected mother (BAB5369) and affected maternal uncle (BAB5371). *SRR* is expressed in mammalian brain, where it converts L-serine to D-serine, a relevant compound

for glutamatergic neurotransmission. Studies of *Srr* knockout mice have demonstrated both a reduction of D-serine in the brain and a reduction of NMDA-induced neurotoxicity, the latter of which has been proposed to be a potential mechanism for the development of the dementia observed in Alzheimer disease.[25] Based on the known function of serine racemase and its expression pattern, we propose that the identified *SRR* variant may function in a gain of function manner, resulting in a predisposition to frontotemporal dementia.

The proband of family HOU2128 has microcephaly, agenesis of the corpus callosum, cerebellar hypoplasia, seizures, horseshoe kidney, scoliosis, hemivertebrae, asymmetric extremities, and hypopigmented skin macules. This individual was initially found to have a rare and potentially deleterious variant in *ITGA7*, conferring a molecular diagnosis of congenital muscular dystrophy (OMIM 613204, Table 1). Reanalysis of the WES data demonstrated variation in *MAGI3*, encoding membrane-associated guanylate kinase with inverted structure 3, and not previously associated with disease. The membrane associated guanylate kinase proteins, MAGI1, MAGI2, and MAGI3, are expressed in the mammalian brain.[26] MAGI2 in particular has been described as causative of mental retardation, autosomal dominant 10 (OMIM 614256) in humans.[27] Disruption of *MAGI2* in humans can lead to a seizure phenotype, as atypical 7q11.23 deletions of the Williams-Beuren syndrome (OMIM 194050) region that extend more distally to include *MAGI2*, are associated with seizures and infantile spasms.[28] MAGI3 enhances the ability of PTEN to regulate AKT kinase activity, particularly under conditions of low PTEN expression, [29] and also plays a role in efficient trafficking of TGF-alpha to the cell surface in polarized epithelial cells.[26] Based on the expression pattern and function of MAGI3, and the known disease association of *MAGI2*, we propose that *MAGI3* may be a novel candidate gene associated with structural and functional neurologic abnormalities and also potentially with asymmetric extremities.

Absence of Heterozygosity (AOH)-Mediated Mutation Burden

One of the advantages of the cohort chosen for this analysis is the high rate of reported consanguinity, enabling pathogenic variants to be reduced to homozygosity within a single generation,[30] and facilitating study of recessive disease, particularly in sibling pairs sharing a similar genetic background. Considering the 8 cases with a second and sometimes third molecular diagnosis identified in a known disease gene, there were a total of 10 variants in four families found within regions of absence of heterozygosity, determined based on the calculated B-allele frequency from exome variant data using BafCalculator (https://github.com/BCM-Lupskilab/BafCalculator) (Table S2).[31] The calculated total AOH for probands of each of the 19 cases with phenotypic expansion ranged from 62.0 Mb to 542.2 Mb, and this was not significantly different from that of the studied brain malformation cohort (128 original cases), which ranged from 28.3 Mb to 866.7 Mb. Although the relatively small number of cases with phenotypic expansion and/or multilocus variation was not enriched for AOH, the studied brain malformation cohort did demonstrate an enrichment for AOH compared to all other Turkish individuals in the CMG, and the Turkish cases within the CMG (~1,000 exomes) were similarly enriched for AOH compared to non-Turkish individuals in our local CMG exome variant database (~6,000 exomes) (Figures S1, S2).

The utility of AOH for the study of multilocus variation is perhaps best illustrated by family HOU1838 for which there are biallelic variants at 2 distinct loci, each with evidence that they contribute to the phenotype, and each within the same 44.8 Mb region of AOH on chromosome 1 (Figure 3). BAB4474, the proband of HOU1838, was initially described as having a novel pathogenic variant in the gene *AP4B1*, providing a molecular diagnosis of spastic paraplegia 47, autosomal recessive (OMIM 614066). This diagnosis was consistent with the observed phenotypes of DD/ID, microcephaly, and agenesis of the corpus callosum, as well as the segregation pattern observed in the pedigree. Extensive re-analysis of the WES data demonstrated additional pathogenic variation at *AMPD2*, providing etiologic molecular diagnoses potentially explaining the pontocerebellar hypoplasia, thin retina, and protruding tongue (Pontocerebellar hypoplasia 9, OMIM 615809 associated with pathogenic variants in *AMPD2,* Figure 3). Notably, a homozygous splice site variant in *NOTCH2* (c.4512–3T>G) could be consistent with the features of cholelithiasis, thin retina, atrial septal defect and hirsutism (Alagille syndrome, OMIM 610205 due to variants in *NOTCH2*), however the zygosity of the identified variant is not consistent with the autosomal dominant inheritance observed with Alagille syndrome, and we propose that the potential for this locus to be involved in both autosomal recessive and dominant modes of disease trait inheritance requires further exploration.[32,33]

In family HOU1857, siblings BAB4635 and BAB4636 were found to have potentially etiologic variants at two unlinked loci within regions of AOH (Figure 4). Both siblings presented with agenesis of the corpus callosum, optic atrophy, sensorimotor neuropathy, distal contractures, and a cardiac valve anomaly. After the initial identification of variants in *C12ORF65* (Combined oxidative phosphorylation deficiency 7, OMIM 613559), homozygous variants in *GPR126* (Lethal congenital contracture syndrome, OMIM 616503) were identified as causative of the expanded phenotype (Figure 4A, B). Both the *C12ORF65* and *GPR126* variants are located in distinct but overlapping regions of AOH on chromosome 12 and 6: *C12ORF65* within a region of 11.3 Mb (BAB4635) and 11.0 Mb (BAB4636); *GPR126* within a region of 9.5 Mb (BAB4635) and 12.1 Mb (BAB4636, Figure 4C).

DISCUSSION

The availability of WES has enabled a genomic interrogation of phenotypes in a hypothesis-free manner, leading to increasing discoveries of both phenotypic expansion associated with a single genomic locus, and blended phenotypes resulting from multilocus variation. To investigate whether cases of apparent phenotypic expansion may have an increased frequency of multilocus variation, we focused on a cohort of well-characterized neurodevelopmental phenotypes , a subset of which were previously categorized as having phenotypic expansion, occurring in the setting of a high rate of reported consanguinity.[20] This strategy enabled a focus on predominantly recessive disease and greater likelihood of shared phenotypes amongst siblings with similar genetic backgrounds resulting from identity by descent.[30] We found that 31.6% (6/19) of cases characterized as demonstrating 'phenotypic expansion' actually represent blended phenotypes[8] with more than one molecular diagnosis, compared to 2.3% (2/87) of cases not characterized as demonstrating phenotypic expansion (two-tailed Fisher's exact p=0.0004).

Such clinical cases are difficult to recognize as potential dual molecular diagnoses, because blended phenotypes may present as phenotypic expansion, more severe disease along a spectrum, or an apparently novel disease.[8] Our findings suggest that cases of apparent phenotypic expansion are more likely to present with a blending of distinct disease phenotypes due to multiple molecular diagnoses. We further demonstrate that cases that do not have apparent phenotypic expression can also result from multilocus variation, and may present as a manifestation of more severe disease. Considering this neurodevelopmental cohort in its entirety, which includes 3 cases with multilocus variation reported in the initial study[20], a total of 108 cases have one or more molecular diagnoses, of which 13 cases (12.0%, 13/108) have two or more molecular diagnoses resulting from known or potential novel disease genes, underscoring the frequency with which multiple molecular diagnoses can occur. Notably, the overall rate of multilocus variation in this selected cohort (12.0%) is greater than the 4.9% reported in a diagnostic cohort of mixed phenotypes from an outbred population[8]; this observation may reflect the genomics of the Turkish population, the high rate of consanguinity, and/or the studied phenotype. As the field moves toward study of oligogenic disease and multi-locus pathogenic variation, we will rely heavily on clinical expertise to untangle the relationships between phenotypic expression observed in the clinic and molecular diagnoses identified by genome-wide assays.

Intrafamilial variability enables a dissection of genotype-phenotype correlation

The interplay among genomic variants at more than one locus and the resulting impact on phenotypic expression can be difficult to predict, and a genotype-forward approach to analysis of phenotypic expression is limited by the rarity of each condition, and combination of conditions, studied. Here, we report 2 families with intrafamilial clinical variability that was ultimately explained by genotypic variability. In family HOU2337, we identified extremely rare and potentially deleterious homozygous variants in both *MED17* and *SPG7* genes in patient BAB6228, in addition to a hemizygous mutation in the *BCOR* gene that he shares with his less severely affected brother (Figure 2). Similarly, in family HOU2293, we identified variants in *ALG8* and *CLN5* in BAB3640, who additionally shared a homozygous *MCPH1* variant with her less severely affected sibling (Figure 1). In both families, the expanded phenotype observed in the probands was in fact explained by variation at additional loci not present in their less severely affected siblings. These cases demonstrate the importance of a thorough clinical history and examination of all affected family members. Genomic analysis in families with more than one affected individual allows a dissection of phenotypes and phenotypic severity attributed to variation at each identified locus within the context of a shared genetic background.

AOH-mediated recessive mutation burden as a tool

In both families described above, AOH contributed to a recessive mutation burden, with total AOH, as determined from the calculated B-allele frequency from exome variant data,[31] measuring 509.0 Mb in BAB6228 and 493.1 Mb in BAB3640 (Table S3). In 14 known or potential autosomal recessive loci identified in this cohort, 11 loci had variants that were homozygous and located within regions of AOH. This finding is particularly striking for patient BAB4474, for whom analysis revealed rare, homozygous variants in three genes within a single region of 44.8 Mb AOH: *AP4B1* (OMIM 614066), *AMPD2*, and a potential

contribution from *NOTCH2* (OMIM 610205, **Figure 3, Table 1).** Although Alagille syndrome has previously been described as an autosomal dominant condition, BAB4474 is homozygous for the identified variant, which was inherited from apparently unaffected parents. Several examples of genes for which both mono- and biallelic variation lead to disease have been previously described,[32–35] and we propose that this *NOTCH2* variant may represent a mild or hypomorphic allele for which biallelic variation is required for phenotypic expression.

We also describe a family (HOU1857) for which AOH-mediated recessive mutation burden involves two causative loci located within unlinked regions of AOH (Figure 4). In this case, the *C12ORF65*: c.248delT: p.V83fs (Spastic Paraplegia 55, AR OMIM 615035) variant explained the majority of the clinical features including intellectual disability, optic atrophy, agenesis of the corpus callosum, sensorimotor neuropathy and contractures (Figure 4, Table 1). A second locus, *GPR126*, has been shown to be associated with Schwann cell and peripheral nerve development and myelination. We propose contribution of the homozygous *GPR126* variant (chr12:123738468_GT>G; c.1880A>G; p.K627R) at this locus to the neuropathy and contractures observed in this family (Figure 4C).[36,37] In addition, the role of Gpr126 in cardiovascular development has been identified in several animal model studies and may be consistent with the cardiac finding of valvular anomaly in this family.[38] While total AOH was not significantly greater in cases with multilocus variation and/or phenotypic expansion compared to the remainder of the neurodevelopmental cohort (Figures S1, S2), the cohort demonstrated enrichment for AOH compared to other Turkish individuals and non-Turkish individuals within our local exome database. The burden of homozygous variants identified in regions of AOH in this cohort demonstrate that AOH-driven recessive disease is one genetic mechanism leading to multiple molecular diagnoses and mutational burden in a patient.

Candidate novel genes

We propose two novel candidate disease genes, each with the potential to impact our understanding of the pathophysiology of neurologic disease and perhaps delineate a precision medicine-guided pathway to potential therapeutic interventional strategies. In HOU2019, a rare variant in *SRR* was identified as contributing to the frontotemporal dementia phenotype in the proband (BAB5368), affected mother (BAB5369), and maternal uncle (BAB5371). *SRR* is a serine racemase that impacts levels of L- and D-serine in the brain, and is thought to be important for glutamatergic neurotransmission, which has been proposed to be involved in the development of complex neurodegenerative disease, in particular Alzheimer disease.[25] We propose that *SRR* missense variation results in a gain-of-function allele affecting glutamatergic neurotransmission and resulting in excitatory neurotoxicity and dementia. Additionally, the role of *SRR* in this phenotype suggests that therapy aimed at modification of glutamatergic neurotransmission may be of potential clinical benefit. In the proband of HOU2128, a rare variant in *MAGI3* was identified and may provide both a molecular explanation for the Aicardi phenotype as well as the observed hemihypertrophy. Hemihypertrophy and overgrowth syndromes can be associated with variation in *PTEN*, and MAGI3 has been demonstrated to impact PTEN regulation of AKT activity.[29] These examples demonstrate that re-analysis of cases characterized as phenotypic

expansion can lead to potential novel candidate disease gene discovery and potentially inform the molecular pathophysiology of disease.

Conclusion

In conclusion, each genome is truly "personal" and phenotypic expression, what the clinician observes, can be impacted by variation at more than one locus, often resulting in more than one molecular diagnosis.[8] In some cases, phenotypic expansion may be a hallmark of a blended phenotype resulting from multilocus variation. This variation can be driven by the genetic substructure of the family, clan or population (i.e. isolated, consanguinity), leading to AOH-mediated recessive disease burden. Implementation of WES has enabled genomic medicine to move from individual study of single gene Mendelian disorders to a more comprehensive understanding of the interplay of genetic variation at more than one locus and resulting phenotypic expression. These findings further emphasize the important role of the clinician and robust physical examination and clinical characterization of each patient in order to recognize blended phenotypes due to multilocus variation. Such blended phenotypes may present clinically as distinct phenotypes manifesting as apparent phenotypic expansion associated with the first identified locus, or overlapping phenotypes that may manifest as an apparent worsening of disease severity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

REFERENCES

1. Fitzsimmons JS, Guilbert PR. Spastic paraplegia associated with brachydactyly and cone shaped epiphyses. J Med Genet. 1987;24(11):702–705. [PubMed: 3430547]

2. Hennekam RC. Spastic paraplegia, dysarthria, brachydactyly, and cone shaped epiphyses: confirmation of the Fitzsimmons syndrome. J Med Genet. 1994;31(3):251–252. [PubMed: 8014978]

3. Armour CM, Humphreys P, Hennekam RC, Boycott KM. Fitzsimmons syndrome: spastic paraplegia, brachydactyly and cognitive impairment. Am J Med Genet A. 2009;149A(10):2254–2257. [PubMed: 19760657]

4. Armour CM, Smith A, Hartley T, et al. Syndrome disintegration: Exome sequencing reveals that Fitzsimmons syndrome is a co-occurrence of multiple events. Am J Med Genet A. 2016;170(7): 1820–1825. [PubMed: 27133561]

5. Shroyer NF, Lewis RA, Yatsenko AN, Lupski JR. Null missense *ABCR* (*ABCA4*) mutations in a family with stargardt disease and retinitis pigmentosa. Invest Ophthalmol Vis Sci. 2001;42(12): 2757–2761. [PubMed: 11687513]

6. Lee W, Xie Y, Zernant J, et al. Complex inheritance of *ABCA4* disease: four mutations in a family with multiple macular phenotypes. Hum Genet. 2016;135(1):9–19. [PubMed: 26527198]

7. Jehee FS, de Oliveira VT, Gurgel-Giannetti J, et al. Dual molecular diagnosis contributes to atypical Prader-Willi phenotype in monozygotic twins. Am J Med Genet A. 2017.

8. Posey JE, Harel T, Liu P, et al. Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. N Engl J Med. 2017;376(1):21–31. [PubMed: 27959697]

9. Cahill KM, Ley AB. Favism and thalassemia minor in a pregnant woman. JAMA. 1962;180:119–121. [PubMed: 13875561]

10. Fraser GR, Stamatoyannopoulos G, Kattamis C, et al. Thalassemias, Abnormal Hemoglobins and Glucose-6-Phosphate Dehydrogenase Deficiency in the Arta Area of Greece: Diagnostic and Genetic Aspects of Complete Village Studies. Ann N Y Acad Sci. 1964;119:415–435. [PubMed: 14219423]

11. Girirajan S, Rosenfeld JA, Coe BP, et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N Engl J Med. 2012;367:1321–1331. [PubMed: 22970919]

12. Farwell KD, Shahmirzadi L, El-Khechen D, et al. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. Genet Med. 2015;17(7):578–586. [PubMed: 25356970]

13. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA. 2014;312(18):1870–1879. [PubMed: 25326635]

14. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013;369(16):1502–1511. [PubMed: 24088041]

15. Retterer K, Juusola J, Cho MT, et al. Clinical application of whole-exome sequencing across clinical indications. Genet Med. 2016;18(7):696–704. [PubMed: 26633542]

16. Posey JE, Rosenfeld JA, James RA, et al. Molecular diagnostic experience of whole-exome sequencing in adult patients. Genet Med. 2016;18(7):678–685. [PubMed: 26633545]

17. Balci TB, Hartley T, Xi Y, et al. Debunking Occam's razor: Diagnosing multiple genetic diseases in families by whole-exome sequencing. Clin Genet. 2017;92(3):281–289. [PubMed: 28170084]

18. Tarailo-Graovac M, Shyr C, Ross CJ, et al. Exome Sequencing and the Management of Neurometabolic Disorders. N Engl J Med. 2016;374(23):2246–2255. [PubMed: 27276562]

19. Chong JX, Buckingham KJ, Jhangiani SN, et al. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. Am J Hum Genet. 2015;97(2):199–215. [PubMed: 26166479]

20. Karaca E, Harel T, Pehlivan D, et al. Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease. Neuron. 2015;88(3):499–513. [PubMed: 26539891]

21. Xia F, Bainbridge MN, Tan TY, et al. De novo truncating mutations in AHDC1 in individuals with syndromic expressive language delay, hypotonia, and sleep apnea. Am J Hum Genet. 2014;94(5): 784–789. [PubMed: 24791903]

22. Jiang Y, Wangler M, McGuire A, et al. The Phenotypic Spectrum of Xia-Gibbs Syndrome. 2018; Am J Med Genet A. 76(6):1315–1326.

23. Hock M, Wegleiter K, Ralser E, et al. ALG8-CDG: novel patients and review of the literature. Orphanet J Rare Dis. 2015;10:73. [PubMed: 26066342]

24. Savukoski M, Klockars T, Holmberg V, Santavuori P, Lander ES, Peltonen L. CLN5, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis. Nat Genet. 1998;19(3):286–288. [PubMed: 9662406]

25. Inoue R, Hashimoto K, Harai T, Mori H. NMDA- and beta-amyloid1–42-induced neurotoxicity is attenuated in serine racemase knock-out mice. J Neurosci. 2008;28(53):14486–14491. [PubMed: 19118183]

26. Franklin JL, Yoshiura K, Dempsey PJ, et al. Identification of MAGI-3 as a transforming growth factor-alpha tail binding protein. Exp Cell Res. 2005;303(2):457–470. [PubMed: 15652357]

27. Hamdan FF, Gauthier J, Araki Y, et al. Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. Am J Hum Genet. 2011;88(3): 306–316. [PubMed: 21376300]

28. Marshall CR, Young EJ, Pani AM, et al. Infantile spasms is associated with deletion of the MAGI2 gene on chromosome 7q11.23-q21.11. Am J Hum Genet. 2008;83(1):106–111. [PubMed: 18565486]

29. Wu Y, Dowbenko D, Spencer S, et al. Interaction of the tumor suppressor PTEN/MMAC with a PDZ domain of MAGI3, a novel membrane-associated guanylate kinase. J Biol Chem. 2000;275(28):21477–21485. [PubMed: 10748157]

30. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. Cell. 2011;147(1):32–43. [PubMed: 21962505]

31. Gambin T, Coban Akdemir Z, Yuan B, et al. Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. Nucleic Acids Res. 2017;45(4):1633–1648. [PubMed: 27980096]

32. Harel T, Yesil G, Bayram Y, et al. Monoallelic and Biallelic Variants in *EMC1* Identified in Individuals with Global Developmental Delay, Hypotonia, Scoliosis, and Cerebellar Atrophy. Am J Hum Genet. 2016;98(3):562–570. [PubMed: 26942288]

33. Harel T, Yoon WH, Garone C, et al. Recurrent *De Novo* and Biallelic Variation of *ATAD3A*, Encoding a Mitochondrial Membrane Protein, Results in Distinct Neurological Syndromes. Am J Hum Genet. 2016;99(4):831–845. [PubMed: 27640307]

34. Monies D, Maddirevula S, Kurdi W, et al. Autozygosity reveals recessive mutations and novel mechanisms in dominant genes: implications in variant interpretation. Genet Med. 2017.

35. Rainger J, Pehlivan D, Johansson S, et al. Monoallelic and biallelic mutations in *MAB21L2* cause a spectrum of major eye malformations. Am J Hum Genet. 2014;94(6):915–923. [PubMed: 24906020]

36. Mogha A, Benesh AE, Patra C, et al. Gpr126 functions in Schwann cells to control differentiation and myelination via G-protein activation. J Neurosci. 2013;33(46):17976–17985. [PubMed: 24227709]

37. Ravenscroft G, Nolent F, Rajagopalan S, et al. Mutations of *GPR126* are responsible for severe arthrogryposis multiplex congenita. Am J Hum Genet. 2015;96(6):955–961. [PubMed: 26004201]

38. Cui H, Wang Y, Huang H, et al. GPR126 protein regulates developmental and pathological angiogenesis through modulation of VEGFR2 receptor signaling. J Biol Chem. 2014;289(50): 34871–34885. [PubMed: 25217645]
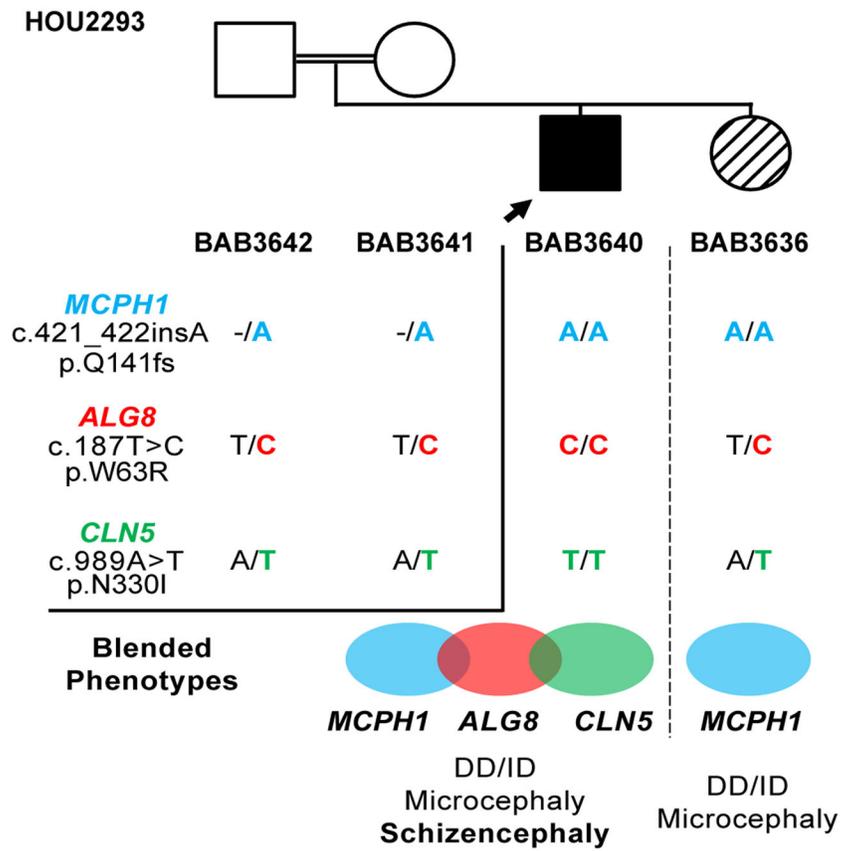
FIGURE 1. Intrafamilial genotypic and phenotypic variability in HOU2293.
Pedigree structure of family HOU2293. Both affected siblings share a homozygous variant in *MCPH1*; however, the proband (BAB3640) has additional homozygous variants in *ALG8* and *CLN5*, resulting in a more severe blended phenotype.
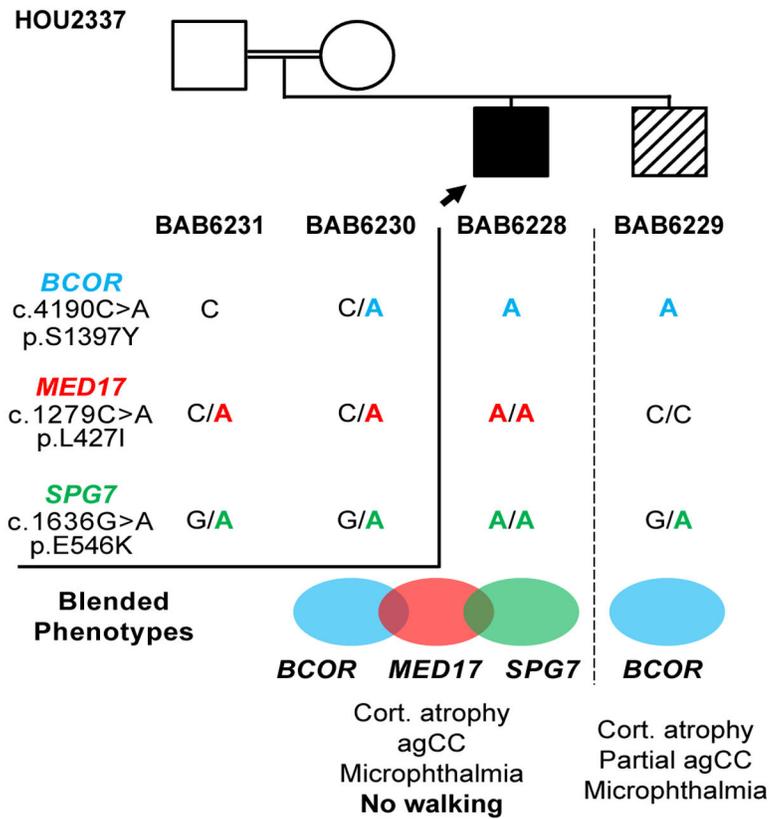
FIGURE 2. Intrafamilial genotypic and phenotypic variability in HOU2337.
Pedigree structure of family HOU2337. Both affected siblings share a hemizygous variant in *BCOR*; the proband (BAB6228) has additional homozygous variants in *MED17* and *SPG7*, resulting in a more severe phenotype due to the additional mutational burden. Cort, cortical; agCC, agenesis of the corpus callosum.
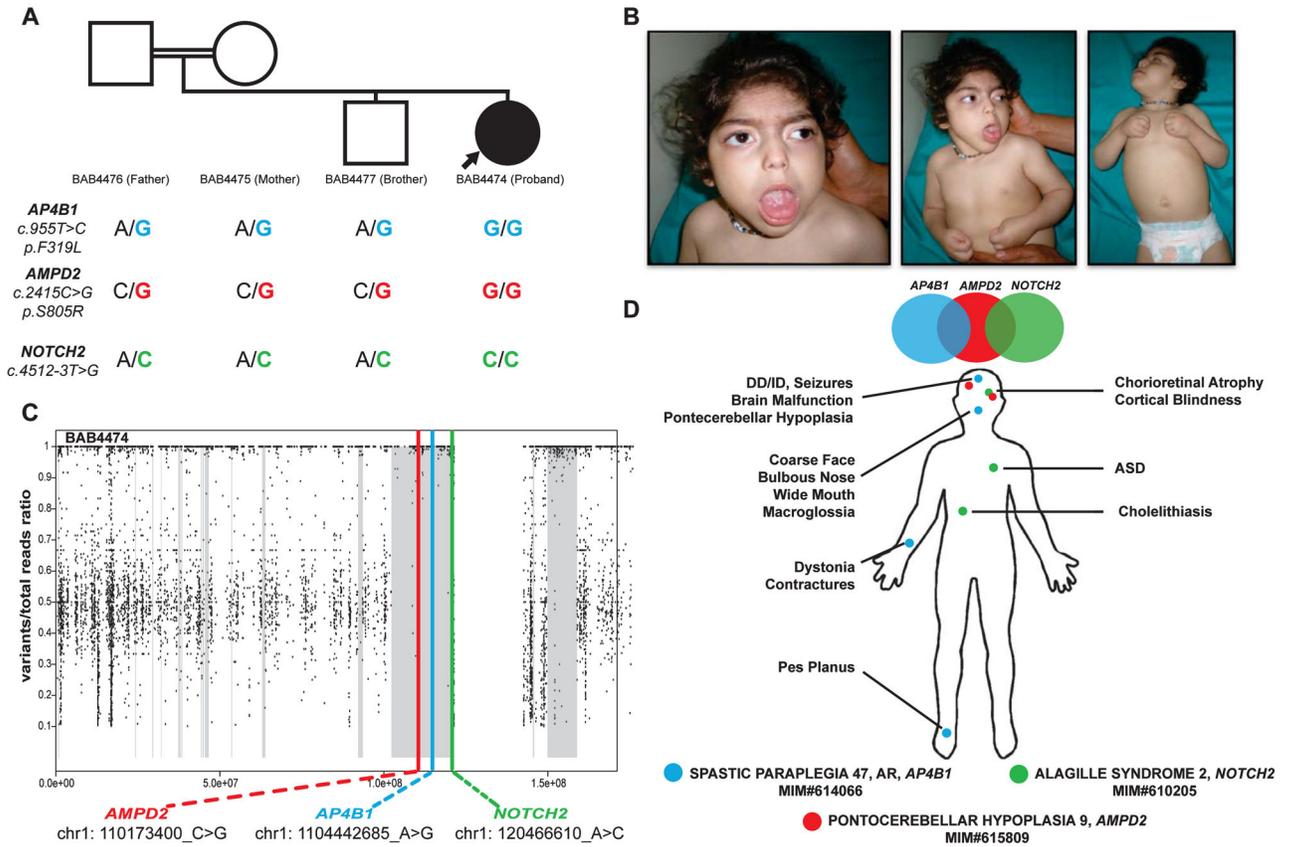
FIGURE 3. Absence of heterozygosity mediates mutational burden at a single locus.
(**A**) Pedigree structure of family HOU1838 and segregation of the variants identified in
AMPD2, AP4B1 and *NOTCH2*. (**B**) Images of proband (BAB4474) demonstrate narrow
forehead, esotropia of the left eye, short nose and protruding tongue. (**C**) B-allele frequency
calculated from exome variant data demonstrates several regions of AOH, marked by grey
zones. Variants in *AMPD2, AP4B1*, and *NOTCH2* are located within the same region of
AOH. (**D**) Blended phenotypic features and associated syndromes are presented. ASD, atrial
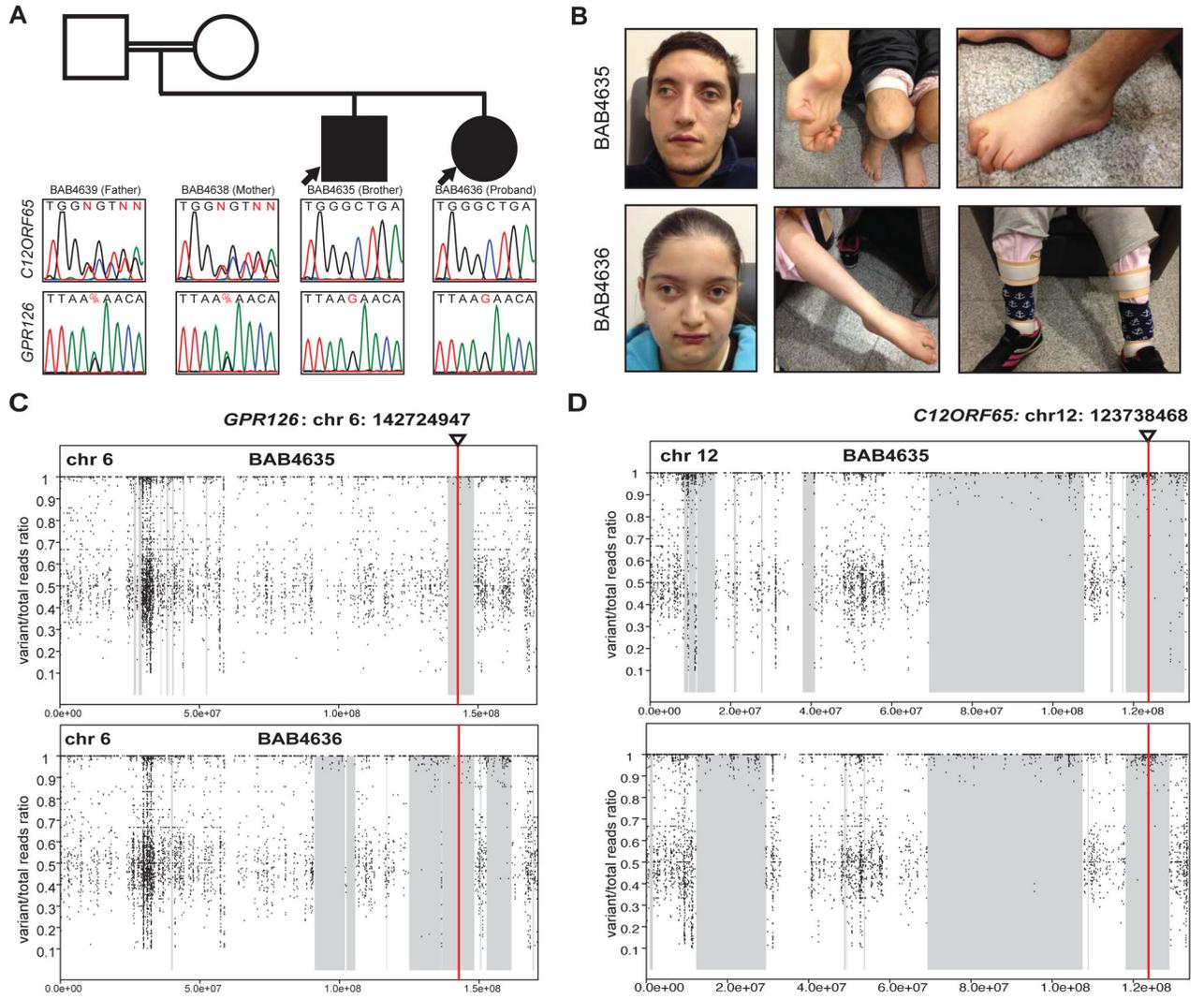septal defect.

FIGURE 4. Absence of heterozygosity mediates mutational burden at more than one locus.
(**A**) Pedigree structure of family HOU1857 and segregation of variants identified in *C12orf65* and *GPR126*. (**B**) Both siblings had esotropia, progressive sensorimotor polyneuropathy with impaired walking and resultant contractures and deformities in their extremities. (**C**) B-allele frequency calculated from exome variant data demonstrates several regions of AOH within chromosome 6, marked by grey zones. Variants in *GPR126* (red line) are located in a region of AOH. (**D**) B-allele frequency calculated from exome variant data demonstrates several regions of AOH within chromosome 12, marked by grey zones. Variants in *C12ORF65* (red line) are located in a region of AOH.

Table 1.

Phenotypic features explained by multiple molecular diagnoses involving 26 known disease genes or loci and 2 candidate disease genes in 13 families. Phenotypes that are italicized indicate expanded phenotypes described in association with a particular molecular diagnosis: bold/red font indicates a phenotype due to a second molecular diagnosis; underscored/blue font indicates a phenotype due to a third molecular diagnosis. ***Severe phenotype proposed to result from burden due to combination of second and third molecular diagnoses. CC, corpus callosum; DD/ID, developmental delay/intellectual disability; PE, phenotypic expansion. * De novo **Phenotype attributed to more than one molecular diagnosis.

Group	Pedigree #	BAB#	1st Dx	MIM#	Patient Phenotypes	2nd Dx	MIM #	3rd Dx	MIM#
	HOU1341	BAB3405	*GRIP1*	617667	**DD/ID, autistic features, agenesis of CC, subependymal heterotopia, hypoplastic vermis, dilated 3rd and 4th ventricles, hydrocephalus, delayed bone age, joint laxity**	*COL3A1*	130050		
	HOU1838	BAB4474	*AP4B1*	614066	**DD/ID, microcephaly, agenesis of CC, Hypoplasia of bulbus and pons, blindness, protruding tongue, atrial septal defect, cholelithiasis, hirsutism**	*AMPD2*	615809		
	HOU1857	BAB4635 BAB4636	*C12ORF65*	613559	ID, agenesis of CC, optic atrophy, sensorimotor neuropathy, *distal contractures, cardiac valve abnormality*	*GPR126*	616503		
PE-KNOWN	HOU1965	BAB4918 BAB4919	*PTCHD1*	611603	**DD/ID**,*frontal lissencephalic cortical dysplasia, seizures, autism spectrum disorder*	*TUBA1A**	611603		
	HOU2293	BAB3640	*MCPH1*	251200	**DD/ID***, microcephaly, *schizencephaly***	*ALG8*	608104	CLN5	256731
		BAB3636			DD/ID, microcephaly				
	HOU2337	BAB6228	*BCOR*	300166	**ID**, _cortical atrophy_ **, agenesis of CC, microphthalmia, nystagmus,** atrial septal defect, _no walking_	*MED17*	613668	*SPG7*	607259
		BAB6229	*BCOR*		ID, cortical atrophy, partial agenesis of CC, microphthalmia, ASD				
PE-NOVEL	HOU2091 BAB5371	BAB5368	*EPB41L1*	614257	*Frontotemporal dementia*	*SRR*			
	HOU2128	BAB5492	*ITGA7*	613204	**DD/ID, microcephaly, agenesis of CC, cerebellar hypoplasia, seizures, hypotonia, chorioretinal lacunae, asymetric extremities, hypopigm macules, horseshoe kidney, scoliosis, hemivertebrae**	*MAGI3*			
NON PE-KNOWN	HOU2092	BAB5373	*INA*	n/a	**ID**, microcephaly, seizures* , cerebral palsy**	*GNAO1**	615473		
	HOU2442	BAB6530	*ATRX*	309580	**ID**, microcephaly** , epilepsy	*CTNNB1*	615075		
	HOU2215	BAB5804 BAB5805	*RARS2 SNX14*	611523 616354	ID, hypoplasia of CC, atrophy of bilateral cerebellum, hypoplastic vermis, seizures				
PREV REPORT	HOU2231	BAB5866 BAB5867	*DHCR7 DMD ex*	270400 310200	Smith-Lemli-Opitz, elevated creaktine kinase				
	HOU2124	BAB5481	15q11.2 del *ASXL3*	615616 615485	DD/ID, seizures, microcephaly, diffuse cortical atrophy, gastroesophageal reflux				

*
De novo

Page 18

**
Phenotype attributed to more than one molecular diagnosis.

***
Severe phenotype proposed to result from burden due to combination of second and third molecular diagnoses.