Human Genetics
https://doi.org/10.1007/s00439-020-02187-7

ORIGINAL INVESTIGATION

Autozygosity-driven genetic diagnosis in consanguineous families from Italy and the Greater Middle East

Flavia Palombo[1,2] · Claudio Graziano[1] · Nadia Al Wardy[3] · Nayereh Nouri[4,5] · Caterina Marconi[6] · Pamela Magini[1] · Giulia Severi[1] · Chiara La Morgia[2,7] · Gaetano Cantalupo[8,9] · Duccio Maria Cordelli[6,10] · Simone Gangarossa[11] · Mohammed Nasser Al Kindi[3] · Mazin Al Khabouri[3,12] · Mansoor Salehi[4] · Elisa Giorgio[13] · Alfredo Brusco[13] · Francesco Pisani[14] · Giovanni Romeo[6] · Valerio Carelli[2,7] · Tommaso Pippucci[1] · Marco Seri[1,6]

Received: 26 November 2019 / Accepted: 25 May 2020

Abstract

Autozygosity-driven exome analysis has been shown effective for identification of genes underlying recessive diseases especially in countries of the so-called Greater Middle East (GME), where high consanguinity unravels the phenotypic effects of recessive alleles and large family sizes facilitate homozygosity mapping. In Italy, as in most European countries, consanguinity is estimated low. Nonetheless, consanguineous Italian families are not uncommon in publications of genetic findings and are often key to new associations of genes with rare diseases. We collected 52 patients from 47 consanguineous families with suspected recessive diseases, 29 originated in GME countries and 18 of Italian descent. We performed autozygosity-driven exome analysis by detecting long runs of homozygosity (ROHs > 1.5 Mb) and by prioritizing *candidate clinical* variants within. We identified a pathogenic synonymous variant that had been previously missed in *NARS2* and we increased an initial high diagnostic rate (47%) to 55% by matchmaking our candidate genes and including in the analysis shorter ROHs that may also happen to be autozygous. GME and Italian families contributed to diagnostic yield comparably. We found no significant difference either in the extension of the autozygous genome, or in the distribution of *candidate clinical* variants between GME and Italian families, while we showed that the average autozygous genome was larger and the mean number of *candidate clinical* variants was significantly higher ($p = 0.003$) in mutation-positive than in mutation-negative individuals, suggesting that these features influence the likelihood that the disease is autozygosity-related. We highlight the utility of autozygosity-driven genomic analysis also in countries and/or communities, where consanguinity is not widespread cultural tradition.

Introduction

In genetics, consanguineous marriages are commonly defined as unions between individuals related as second cousins or closer, resulting in a pedigree-based coefficient of inbreeding (F) in their progeny ≥ 0.0156 (Bittles 2001). Individuals whose parents are so closely related are expected to have an increased proportion of their autosomal genome that is autozygous, where two identical haplotypes descend from a recent common ancestor; the closer the degree of relatedness, the greater the proportion of the genome that is expected to be autozygous (Smith 1974). Autozygosity is considered to be the genomic hallmark of inbreeding, manifesting as long runs of homozygosity (ROHs), i.e., sizeable stretches of homozygous genotypes at consecutive polymorphic DNA marker positions, which are spread throughout an individual genome.

The most well-known medical impact of parental consanguinity is the increased risk of rare autosomal recessive diseases in the progeny. Notably, the excess risk is inversely proportional to the frequency of the disease-related allele in the gene pool (Bittles 2001). As a result, genetic analysis of consanguineous families paved the way to the identification of many genes underlying ultra-rare Mendelian conditions (Alkuraya 2013; Alazami et al. 2015). Since it was first proposed in 1987 (Lander and Botstein 1987) homozygosity mapping, which exploits the occurrence of

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00439-020-02187-7) contains supplementary material, which is available to authorized users.

✉ Tommaso Pippucci
tommaso.pippucci@unibo.it

Extended author information available on the last page of the article

Published online: 02 June 2020

long ROHs surrounding the disease-related variant, proved to be a powerful tool to map recessive genes, and its combination with Whole Exome Sequencing (WES) rapidly boosted novel gene discoveries in first years of the next generation sequencing era (Aldahmesh et al. 2011; Becker et al. 2011; Chiang et al. 2012; Shaheen al. 2011; Sobreira et al., 2010). More recently, diagnostic yield up to 60% has been reported in clinical settings, where homozygosity mapping and WES have been used to establish genetic diagnosis in children to consanguineous parents (Makrythanasis et al. 2014; Yavarna et al. 2015; Charng et al. 2016). It is usually emphasized how marriage between consanguineous spouses is a cultural tradition in countries of the so-called Greater Middle East (GME) (Scott et al. 2016), while is no longer common among most European populations, where, however, it was socially accepted and even favored at least until the nineteenth century (Bittles et al. 2001). In Italy, profound demographic changes occurred during the 2nd half of the twentieth century causing a rapid decrease in rates of consanguineous marriages, especially in the north (Cavalli Sforza et al. 2004). It has been argued that autozygosity-driven WES analysis is most powerful in populations of GME countries, where the identification of genes underlying autosomal recessive diseases is facilitated by high population inbreeding rate and large family size (Monies et al. 2017a, b). Nonetheless, reports of genetic diagnoses and novel disease-gene associations in Italian consanguineous families are not uncommon in the literature (see Spataro et al. 2019; Ramos et al. 2019; Milev et al. 2018 as illustrative recent papers). We aimed to investigate whether autozygosity-driven WES analysis could be as effective in a country like Italy, where population inbreeding rate is lower and family size is on average smaller than in GME countries. Here we report on our findings in 52 patients from 47 consanguineous families with suspected autosomal recessive diseases, 29 that originated in GME countries and 18 of Italian descent.

Materials and methods

Patients. This study includes 47 families referred in years 2012–2017 to the Medical Genetics Unit, Sant'Orsola-Malpighi University Hospital in Bologna, Italy or collaborative centers (Bellaria Hospital in Bologna, Italy; Città della Salute e della Scienza University Hospital, Turin, Italy; Department of Biochemistry, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Oman; Department of ENT, Al Nahdha Hospital, Ministry of Health, Muscat, Oman; Alzahra University Hospital, Isfahan University of Medical Sciences, Isfahan, Iran) for suspected genetic condition in consanguineous families.

Written informed consent was obtained from all subjects and study was approved by the local institution ethical committee and review board (IRB protocol 3206/2016 at Policlinico S. Orsola-Malpighi, Bologna (Italy); Ref No. SQU – EC/121/16, MREC # 1311 (Oman); Ethics Committee of the Città della Salute e della Scienza University Hospital, Torino (Italy) (approval number 0060884); IRM. MUI. REC of Isfahan University of Medical science (Iran)) and performed according to the Declaration of Helsinki protocol.

Whole exome sequencing. Genomic DNA was extracted from peripheral blood samples collected in EDTA anticoagulant with the GenElute Blood Genomic DNA Kit (Sigma Aldrich, Missouri, USA), following the manufacturer's instruction, and with the QIAamp DNA Blood Mini (Qiagen, Venlo, Netherlands), following a modified protocol (800 µl of blood instead of 200 µl and 2 additional washes with the kit's buffer) to obtain 3 µg of DNA.

Targeted capture and enrichment were performed using different commercial kits: BGI exome (BGI Tech Solutions, Shenzhen, China), Nextera Rapid CaptureExome (Illumina Inc., San Diego, CA), TruSeqExomeLibrary Prep Kit (Illumina), SeqCap EZ Exome Enrichment v2 and v3 (Nimblegen Inc., Madison, WI). Libraries were sequenced as 91-bp or 100-bp paired-end reads on Illumina HiSeq2000 or HiSeq2500 platforms (Illumina).

Generated reads were treated following a general pipeline elsewhere described (Magini et al. 2014) including alignment with BWA (Li and Durbin 2009) to the reference genome hg19, realignment and base quality score recalibration with GATK (DePristo et al. 2011) and duplicate removal with PicardTools (https://picartools.sourceforge.net). Alignment and coverage statistics were collected with SAMtools (Li et al. 2009) and GATK. Variants were called and filtered by quality with GATK HaplotypeCaller and Variant Quality Score Recalibration, and then annotated with RefSeq using ANNOVAR (Wang et al. 2010). H3M2 (Magi et al. 2014) was used for the identification of ROHs from WES alignments.

Autozygosity-based variant prioritization workflow. To prioritize variants according to their probability of being causal for autozygosity-related pathology, we classified ROHs identified by H3M2 into 2 size classes reflecting the presumed ROH origin (Pemberton et al. 2012):

1. Long ROH (larger than 1.5 Mb), typical of consanguineous families as a consequence of close parental relatedness and thus likely to be autozygous.
2. Short-medium ROH (smaller than 1.5 Mb), common in inbred communities but also present in outbred populations as a consequence of background parental relatedness;

Candidate clinical variants were defined as variants with potential to alter the protein product (missense, nonsense, small insertion/deletions and splicing-affecting variants) with allele frequency lower than 0.01 (rare variants from here on) and not seen in homozygous state in gnomAD database (https://gnomad.broadinstitute.org/). Effect on splicing was assumed if variants impacted conventional splice-sites (±2 bp from intron–exon junction) and/or were predicted as unconventional splicing-affecting variants (uSAVs) by MutPredSplice v1.3.2 (https://mutdb.org/mutpredsplice).

We subsequently stratified homozygous *candidate clinical* variants following their probability of being autozygous: autosomal homozygous variants in long ROHs (increased chance to be autozygous) and out of long ROHs (less likely to be autozygous).

Finally, we prioritized *candidate clinical* variants in 3 layers. Based on the American College for Medical Genetics (ACMG) (Richards et al. 2015) guidelines the first two layers were:

(a) *Pathogenic*:

- variant already reported as pathogenic or likely pathogenic in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) and/or in HGMD® (https://www.hgmd.cf.ac.uk/ac/index.php) or for which pathogenicity is convincingly supported by data in the literature;
- variant predicted to be Loss of Function (LOF) (nonsense changes, frame-shift indels and splicing-affecting variants) within a gene reported in Online Mendelian Inheritance in Man (OMIM) to be associated with a Mendelian phenotype;

(b) *Likely pathogenic*:

- non-LOF variant within a gene reported in OMIM to be associated with a Mendelian phenotype, consistent with the clinical diagnosis;

  The third layer included *potentially pathogenic* variants predicted to be deleterious by Combined Annotation Dependent Depletion (CADD v.1.3) score greater than 20 and in genes not reported in OMIM to be associated with a Mendelian phenotype at the time when bioinformatic analysis was completed (31/12/2017).

Sanger sequencing. Variants were validated and their segregation within parents and siblings was checked through Sanger sequencing. Specific primers (Eurofins Genomics) were manually designed and PCRs were carried out using KAPA ReadyMix (KAPA Biosystems) under standard condition. Sanger sequencing was performed using the Big Dye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems), following the manufacturer's instructions. Sequencing analysis was performed by the automated sequencer with 48 capillaries (3730 DNA analyzer, Applied Biosystems) and output files were analyzed through the software Sequencher 5.0 (Gene Code Corporation).

NARS2 cDNA sequencing and real time PCR. Total RNA was extracted from cultured skeletal muscle cells and fibroblasts of a healthy individual and the proband of family ITA_9, using the tissue protocol of QIAmp RNA blood mini kit (Qiagen). 500 ng of extracted RNA and of human colon and whole brain RNA pools (Clontech) were retrotranscribed to cDNA through the high-capacity cDNA reverse transcription kit (Thermo Fisher Scientific). To evaluate possible splicing alterations around exon 9 in the patient, *NARS2* cDNA was specifically amplified through standard PCR with primers complementary to exon 7 (forward) and exon 10 (reverse) sequences. Bands from gel electrophoresis of *NARS2* cDNA amplicons were separated, purified through the QIAquick gel extraction kit (Qiagen), sequenced using the BigDye terminator v1.1 cycle sequencing kit (Thermo Fisher Scientific) and run on the 3730 DNA Analyzer (Thermo Fisher Scientific). Expression of *NARS2* wt transcript, including exon 9, was evaluated in patient's and controls' fibroblasts and skeletal muscle cells by quantitative PCR through the Universal Probe Library (UPL) system (Roche). *ACTB* mRNA was used as endogenous normalizer and amplified separately. Each reaction was performed in triplicate. The ΔCt method was applied to real-time data to obtain a relative quantification of *NARS2* wt transcript expression.

Statistical analysis. Differences between mean values of coverage, proportion of autozygous genome and number of variants were assessed by Welch's t-test, or unequal variances *t*-test, with R version 3.5.1. (https://www.r-project.org/). A p value ≤ 0.05 was considered significant.

Results

Patients' population and autozygosity-driven WES variant prioritization. In this study, 52 individuals affected with different disorders of suspected genetic origin, belonging to 47 consanguineous families, had WES performed. In detail, parents' probands were related as first cousins (40, 85.1%), first cousins once-removed (3, 6.4%), second cousins (3, 6.4%) or second cousins once-removed (1, 2.1%). Family history was negative for 43 probands (91.5%), while the remaining had 1 (3 families, 6.5%) or 2 (1 family, 2%) affected siblings but no record of disease was reported for parents or relatives in upper generations, suggesting autosomal

recessive inheritance. The majority of families (29, 62%) were from GME countries, while the remaining (18, 38%) were from Italy. There was no major difference in the distribution of the degree of parental relatedness between Italian and GME families, with first cousins being preponderant in both (13/18 [72%] and 27/29 [93%], respectively). Clinical diagnoses reflected a composition of suspected Mendelian disorders, including either groups of families from single GME countries with homogeneous phenotype (12 Omani families with non-syndromic congenital deafness) or with syndromic conditions unified by a common sign (8 Iranian families with syndromic cleft palate), or families with heterogeneous multi-systemic (neuro)developmental disorders and of prevalent Italian ancestry (18/27, 64%) (Table 1). Genetic analyses performed prior to WES in 32 probands (68%) were inconclusive and laboratory tests prescribed in specific instances after clinical examination were negative, as summarized in Supplemental Table 1.

Notwithstanding differences in library preparation and capture, all WES experiments achieved comparable coverage which was adequate for the identification of homozygous variants with, on average, a mean coverage of 82.5X (±27.2X) and % of bases covered above 20X of 86.3% (±5.3%) (Supplemental Table 2). Mean number of exonic and canonical splice-site variants (±2 bp from exon/intron boundary) was 22,920 (±1,271) per individual, of which 8,853 (±461) were homozygous. Autozygosity-driven prioritization restricted the list to a mean of 931 (±477) variants within large ROHs including 38 uSAVs in 23 families (mean 1.6±0.7) as detailed in Supplemental Table 3, of which 19 (±12) per individual were *candidate clinical*.

Diagnostic yield of the autozygosity-driven WES analysis. In 22 out of 47 families, where we identified a variant classified as pathogenic (Table 2) or likely pathogenic (Table 3) according to our criteria, we claimed that we established diagnosis (47%).

Nine variants were LOF, while 11 were missense. Clinical significance of 4 missense variants was convincingly supported by ClinVar (Table 2), while one, *ACP5* p.Q120R, was described in another Italian individual (Patient 21 in Briggs et al. 2016). MutpredSplice predicted 11 uSAVs within genes associated with autosomal recessive Mendelian diseases, 5 of which consistent with the patient's clinical presentation. Four were nonsynonymous changes already classified as pathogenic or likely pathogenic, three in established deafness genes (*CDH23*, *COL9A2*, *PCDH15*) and one in *TRAPPC2L* (Milev et al. 2018), while one was the novel synonymous *NARS2* p.N90N variant. This unconventional splice-site variant was predicted to produce an Exonic Splicing Enhancer (ESE) loss and Exonic Splicing Silencer (ESS) gain and it was, therefore, added to the list of pathogenic variants. We subsequently confirmed that the functional consequence of this variant was *NARS2* exon 9 skipping resulting in an out-of-frame truncated protein (p.N90HfsX4) by analysis of cDNA from patient's fibroblasts. With specific primers, we obtained a single *NARS2* amplicon (318 bp) in all tested control tissues. In the patient, bands with different size indicated the activation of alternative splice sites, including those eliminating exon 9 (280 bp). Interestingly, patient's fibroblasts maintained the wild-type band, while skeletal muscle cells, which are one of the disease targets, lost it completely (Supplemental Fig. 1A, B). To assess the presence of the *NARS2* wt transcript, a Real-Time PCR was carried out on RNA from both fibroblasts and muscle biopsy: in the proband the wt transcript resulted expressed at very low levels with respect to controls in both tissues (Supplemental Fig. 1C).

In two instances, the *candidate clinical* variants were found in association with phenotypes only partially overlapping the classical presentation ascribable to the genes. *DDC* p.Arg375Cys was described to expand the phenotypic spectrum of aromatic aminoacid decarboxylase deficiency related to *DDC* mutations (Graziano et al. 2015). *MC2R* p.L283R only partially explains the composite clinical picture in patient from family GME_14, since while it was assessed as causative of the glucocorticoid deficiency phenotype, it could not clarify patient's craniostenosis, which is a feature that has never been observed so far in *MC2R*-related disorders. In two further families, homozygous variants in two genes normally underlying autosomal dominant diseases, *RAD21* (p.A622T) and *NOTCH3* (p.C966X), were demonstrated to cause recessive phenotypes that partially overlapped the dominant ones (Bonora et al. 2015; Pippucci et al. 2015).

We then searched for variants out of long ROHs, to uncover disease-related variants within short-medium ROH. *MYO15A* p.Y393CfsX41, identified in three siblings affected with congenital deafness from family GME_17, was within a ROH ranging about 900 Kb and we previously demonstrated that it is a founder mutation in Oman (Palombo et al. 2017), the country from which this family originates.

We also evaluated *potentially pathogenic* variants identified in mutation-negative families after previous steps, resulting in 20 candidate genes. For 3 of these genes (*TRAPPC2L*, *SMPD4* and *CCDC32*) we were connected through Gene Matcher to colleagues that had identified defects in one of the same genes in patients presenting with convincingly overlapping phenotypes. We were thus able to collect a series of genetic and functional evidence to support novel disease-gene association which are described in separate publications (Milev et al. 2018; Magini et al. 2019; Harel et al. 2020). Thanks to these collaborative efforts we were eventually able to ascertain the causative role of these three variants and to upgrade their pathogenicity classification from potentially pathogenic to pathogenic (Table 2). Overall, with the *MYO15A* variants and variants that were initially

Table 1 Overview of patients and pedigrees in this study

Family	Country of origin	Closest relationship between parents of probands	Phenotype	WES (affected)
ITA_1	Italy	First cousins	Short stature, facial hirsutism, alopecia, dyslipidemia and extreme insulin resistance	1 † (1)
ITA_2	Italy	Second cousins once removed	ID	1† (1)
ITA_3	Italy	First cousins	ID	1† (1)
ITA_4	Italy	First cousins	Mitochondrial DNA depletion syndrome	1# (1)
ITA_5	Italy	First cousins	Epileptic Encephalopathy	1# (1)
ITA_6	Italy	Second cousins	Metabolic encephalomyopathy mitochondrial suspected	1‡ (1)
ITA_7	Italy	Second cousins	Syndromic ID	2# (2)
ITA_8	Italy	First cousins	Bilateral optic neuritis	2# (1)
ITA_9	Italy	First cousins	Reversible COX deficiency	2# (1)
ITA_10	Italy	First cousins	HSP and AIHA	1¶ (1)
ITA_11	Italy	First cousins	Leukoencephalopathy	1§ (1)
ITA_12	Italy	First cousins	Agammaglobulinemia	1§ (1)
ITA_13	Italy	First cousins	Corpus callosum agenesia	1† (1)
ITA_14	Italy	First cousins once removed	Microphtalmia and hypoacusia	1‡ (1)
ITA_15	Italy	First cousins	HSP, diplopia and nistagmo	1¶ (1)
ITA_16	Italy	Second cousins	ID (CNS malformation)	1† (1)
ITA_17	Italy	First cousins	Diaphragmatic hernia	1‡ (1)
ITA_18	Italy	First cousins	Neuroaxonal dystrophy	1† (1)
GME_1	Oman	First cousins	Deafness	1‡ (1)
GME_2	Oman	First cousins	Deafness	1‡ (1)
GME_3	Oman	First cousins	Deafness	1‡ (1)
GME_4	Oman	First cousins	Deafness	1‡ (1)
GME_5	Oman	First cousins	Deafness	1‡ (1)
GME_6	Oman	First cousins	Deafness	1‡ (1)
GME_7	Oman	First cousins	Deafness	1‡ (1)
GME_8	Oman	First cousins	Deafness	1‡ (1)
GME_9	Oman	First cousins	Deafness	1‡ (1)
GME_10	Oman	First cousins	Deafness	1‡ (1)
GME_11	Oman	First cousins	Deafness	1‡ (1)
GME_12	Turkey	First cousins	CIPO	2§ (2)
GME_13	Morocco	First cousins	HSP	2‡ (2)
GME_14	Pakistan	First cousins	Craniostenosis and glucocorticoid deficiency	1‡ (1)
GME_15	Bangladesh	First cousins	Severe psychomotor delay, aggression, early puberty	1‡ (1)
GME_16	Pakistan	First cousins	Microcephaly, mild developmental delay, atrial septal defect and multicystic left kidney	1‡ (1)
GME_17	Oman	First cousins	Deafness	3# (3)
GME_18	Morocco	First cousins once removed	Lissencephaly	2‡ (1)
GME_19	Iran	First cousins	Syndromic CP	1‡ (1)
GME_20	Iran	First cousins once removed	Syndromic CP	1‡ (1)
GME_21	Iran	First cousins	Syndromic CP	1‡ (1)
GME_22	Iran	First cousins	Syndromic CP	1‡ (1)
GME_23	Iran	First cousins	Syndromic CP	1‡ (1)
GME_24	Iran	First cousins	Syndromic CP	1‡ (1)
GME_25	Iran	First cousins	Syndromic CP	1‡ (1)
GME_26	Iran	First cousins	Syndromic CP	1‡ (1)
GME_27	Pakistan	First cousins	Leukodystrophy	1¶ (1)
GME_28	Saudi Arabia	First cousins	ARMs	1‡ (1)
GME_29	Iran	First cousins	THC	1† (1)

AIHA autoimmune haemolytic anaemia, *ARMs* ano-rectal malformations, *CIPO* chronic intestinal pseudo-obstruction, *CNS* Central Nervous System, *CP* Cleft Palate, *HSP* Hereditary Spastic Paraplegia, *ID* Intellectual Disability, *THC* thrombocytopenia

In the last column the target enrichment kits are indicated: †BGI_target; ‡Nextera; §SeqCapV2; ¶SeqCapV3; #Truseq

Table 2 Pathogenic variants in patients

Family	Disease	Gene	Variant	gnomAD†	gnomAD† allele	GME‡ frequency	GME‡ genotype count	HGMD/Clinvar	CADD	ROH size
ITA_1	Short stature, facial hirsutism, alopecia, dyslipidemia and extreme insulin resistance	POC1A	NM_015426.5:c.1007dupC:p.G337Rfs*25	4.295e-6	1/232,832	NA	NA	No	NA	26 Mb
ITA_4	Mitochondrial DNA depletion syndrome	POLG	NM_001126131.2:c.1943C>G:pP648R	8.259e-6	2/242,174	NA	NA	Yes	24.7	3.5 Mb
ITA_5	Epileptic encephalopathy	CACNA2D2	NM_001005505.3:c.1295delT:p.N432Tfs*35	NA	NA	NA	NA	No	NA	18.1 Mb
ITA_9	Reversible COX deficiency	NARS2	NM_001243251.1:c.270C>T:p.N90N	NA	NA	NA	NA	No	NA	6.5 Mb
ITA_10	HSP and AIHA	ACP5	NM_001611.5:c.359A>G:p.Q120R	NA	NA	NA	NA	No	16.56	4 Mb
ITA_11	Leukoencephalopathy	NOTCH3	NM_000435.3:c.2898C>A:pC966*	NA	NA	NA	NA	No	39	3.7 Mb
ITA_16	ID (CNS malformations)	TRAPPC9	NM_001160372.4:c.224_225delGC:p.R75Qfs*22	NA	NA	NA	NA	No	NA	13 Mb
ITA_18	Neuroaxonal dystrophy	TFG	NM_001007565.2:c.316C>T:p.R106C	3.918e-5	11/280,752	0.0005035	992:1:0	Yes	24.7	22.6 Mb
GME_2	Deafness	CDH23	NM_022124.6:c.6049G>A:p.G2017S	2.873e-5	7/243,642	NA	NA	Yes	34	4 Mb
GME_5	Deafness	SLC26A4	NM_000441.2:c.1246A>C:p.T416P	1.949e-4	55/282,196	NA	NA	Yes	19.47	19.5 Mb
GME_6	Deafness	COL4A3	NM_000091.5:c.828+1G>C	NA	NA	NA	NA	No	16	10 Mb
GME _8	Deafness	LOXHD1	NM_001145473.2:c.771C>A:p.Y257*	6.313e-6	1/158,404	NA	NA	No	41	27 Mb
GME _9	Deafness	COL9A2	NM_001852.4:c.1807A>T:p.K603*	NA	NA	NA	NA	No	39	10 Mb
GME _11	Deafness	PCDH15	NM_001142769.3:c.4705G>A:p.Q1569*	NA	NA	NA	NA	No	35	16 Mb
GME _13	HSP	SPG20	NM_001142294.1:c.892dupA:p.T298Nfs*16	NA	NA	NA	NA	No	NA	3.7 Mb
GME _17	Deafness	MYO15A	NM_016239.4:c.1171_1177dupGCCATCT:p.Y393Cfs*41	NA	NA	NA	NA	No	NA	800 kb
GME _18	Lissencephaly	SMPD4	NM_001171083.2:c.1264+1G>A	NA	NA	NA	NA	No	17	22 Mb
GME _23	Syndromic CP	CCDC32	NM_001080791.2:c.190_191insGG:p.E64Gfs*12	NA	NA	NA	NA	No	NA	5.23 Mb

†gnomAD = genome aggregation database (https://gnomad.broadinstitute.org/); ‡GME = Greater Middle East Variome Project (https://igm.ucsd.edu/gme/index.php). ROH = Run Of Homozygosity

Table 3 Likely pathogenic variants in patients

Family	Disease	Gene	Variant	gnomAD†	gnomAD† allele	GME‡ frequency	GME‡ genotype count	CADD	ROH size
ITA_6	Metabolic encephalomyopathy mitochondrial suspected	TRAPPC2L	NM_016209.5:c.109G > T:p.D37Y	1.592e-5	4/251,318	NA	NA	27.1	12.5 Mb
ITA_7	Syndromic ID	DDC	NM_001242888.2:c.1123C > T:p.R375C	1.591e-5	4/251,370	NA	NA	16.21	3.3 Mb
GME_10	Deafness	CDH23	NM_001171933.1:c.716A > C:p.D239A	NA	NA	NA	NA	28	40 Mb
GME _12	CIPO	RAD21	NM_006265.3:c.1864G > A:p.A622T	7.997e-6	2/250,086	NA	NA	15.82	7.1 Mb
GME_14	Craniostenosis and glucocorticoid deficiency	MC2R	NM_000529.2:c.848 T > G:p.L283R	NA	NA	NA	NA	21.7	2.3 Mb
GME _16	Microcephaly, mild developmental delay, atrial septal defect and multicystic left kidney	CENPF	NM_016343.4:c.5447 T > C:p.V1816A	1.041e-4	26/249,662	NA	NA	0.28	26 Mb
GME _19	Syndromic CP	ECEL1	NM_004826.4:c.1339C > T:p.R447C	7.967e-6	2/251,048	NA	NA	20.9	17 Mb
GME _22	Syndromic CP	COLEC11	NM_001255986.1:c.364G > A:p.E122K	NA	NA	NA	NA	17.58	2.4 Mb

†gnomAD = genome aggregation database (https://gnomad.broadinstitute.org/); ‡GME = Greater Middle East Variome Project (https://igm.ucsd.edu/gme/index.php). ROH = Run Of Homozygosity

classified as potentially pathogenic but were later assessed to be causative, we established diagnosis in 26/47 families (11/18 Italian families, 60%; 15/29 GME families, 51%) and we were thus able to increase the diagnostic yield to 55% (26/47). Diagnostic yields varied substantially between the principal different disease groups, as we achieved genetic diagnosis in 8/12 of the probands with non-syndromic congenital deafness (66%), in 3/8 of those with syndromic cleft palate (37%), and in 10/16 of those with neurodevelopmental disorders (62.5%).

Among the remaining 17 novel candidate genes (Table 4), we deem that at least *NDUFAF7* deserves a comment although we were not able to provide likewise evidence that it is causative of the patient's clinical manifestation. *NDUFAF7* p.A278T was identified in a patient with a severe form of leukodystrophy with suspected mitochondrial origin. This gene, which has been functionally characterized only recently (Rhein et al. 2013), encodes a S-adenosylmethionine-dependent methyltransferase located in the mitochondrial matrix that symmetrically dimethylates the residue Arg-85 in the *NDUFS2* subunit during the early stage of complex I assembly. Moreover, *NDUFAF7* seems to be essential for normal vertebrate development as its knockout in mice is embryonically lethal (Zurita Rendón et al. 2014).

Proportion of autozygous genome and variants and evaluation of their impact on diagnostic yield. The cumulative length of individual long ROHs, assumed to be the autozygous genome, ranged from 62.6 Mb to 432 Mb (mean: 253 Mb ± 97 Mb). Such a wide span in the autozygous genome is consistent with literature data reporting a 86 Mb-345 Mb range in the offspring to first cousin unions (Leutenegger et al. 2003). Although GME families had on average a larger autozygous genome than Italian families (266 Mb ± 98 Mb vs. 231 Mb ± 96 Mb), there was

Table 4 Potentially pathogenic variants in patients

Family	Autozygous genome	Disease	Gene	Variant	gnomAD†	gnomAD† allele	GME‡ frequency	GME‡ genotype count	CADD	ROH size
ITA _12	373 Mb	Agammaglobulinemia	BNIP1	NM_001205.3:c.164G > A:p.R55H	1.45e-4	41/282,830	NA	NA	22	9.8 Mb
			TMC7	NM_001160364.2:c.1479delC:p.L494fs*5	3.98e-6	1/251,492	NA	NA	NA	11 Mb
ITA_13	217 Mb	CC Agenesia	AP1G1	NM_001128.6:c.728C > A:p.P243H	NA	NA	NA	NA	31	9.5 Mb
ITA _14	216 Mb	Microphtalmia and hypoacusia	EVI2A	NM_014210.4:c.343A > T:p.K115*	1.77e-5	5/282,384	NA	NA	26	4.5 Mb
ITA _15	265 Mb	HSP, diplopia and nistagmo	NRXN3	NM_004796.6:c.928G > A:p.A310T	5.983e-5	15/250,692	0.00151	990:3:0	26	33 Mb
ITA _17	311 Mb	Diaphragmatic hernia	RFC4	NM_002916.4:c.481G > A:p.A161T	NA	NA	NA	NA	21	7.8 Mb
			TTC28	NM_001145418.1:c.4255C > T:p.R1419W	8.87e-4	164/184,926	0.00232	428:2:0	24	4.2 Mb
			INPP5J	NM_001284287.1:c.413A > G:p.D138G	7.86e-5	22/279,882	0.00151	988:3:0	27	13.2 Mb
GME_3	304 Mb	Deafness	LATS1	NM_004690.4:c.2416C > T:p.R806*	NA	NA	NA	NA	41	13 Mb
GME_4	178 Mb	Deafness	ECE2	NM_032331.3:c.643delG:p.G216Sfs*82	3.58e-5	9/251,276	0.00151	990:3:0	NA	5.2 Mb
GME _15	223 Mb	Severe psychomotor delay, aggression, early puberty	C11orf49	NM_001278222.1:c.71C > G:p.S24*	2.03e-5	5/246,342	NA	NA	NA	2.5 Mb
GME _25	352 Mb	Syndromic CP	BRF2	NM_018310.4:c.481G > T:p.G161*	NA	NA	NA	NA	32	5.6 Mb
			LRFN2	NM_020737.3:c.2363C > T:p.T788M	NA	NA	NA	NA	20	10.4 Mb
GME _27	251 Mb	Leukodystrophy	NDUFAF7	NM_144736.5:c.832G > A:p.A278T	3.582e-5	9/251,256	NA	NA	27	3.2 Mb
			TRMT11	NM_001031712.3:c.1195G > A:p.E399K	9.94e-5	25/251,420	NA	NA	36	12.4 Mb
GME _28	359 Mb	ARMs	IQGAP3	NM_178229.5:c.3281 + 2 T > G	NA	NA	NA	NA	14	7.6 Mb

†gnomAD = genome aggregation database (https://gnomad.broadinstitute.org/); ‡GME = Greater Middle East

substantial overlap in range (62.6–432 Mb vs. 96–406 Mb) resulting in no significant difference (p value = 0.3096).

The autozygous genome positively correlated with the total number of variants in long ROH (Adjusted R-squared: 0.8103, p value: < 2.2e-16), as well as, although showing progressively weaker correlation, with the number of *rare* variants in long ROHs (Adjusted R-squared: 0.5711, p value: 5.713e−11) or of only *candidate clinical* variants in long ROHs (Adjusted R-squared: 0.2364, p value: 0.000153) (Supplemental Fig. 2).

GME individuals had a higher mean number of variants in long ROHs compared to individuals of Italian ancestry for all total, *rare* and *candidate clinical* variants, although this difference was not significant (total: 32,390 vs. 16,029, mean 981.5 std ± 483.6 vs. 843.6 ± 465.1, p value = 0.3166; *rare*: 1156 vs. 473, 35 ± 22.1 vs. 24.9 ± 14.3, p value = 0.05105; *candidate clinical*: 126 vs. 68, 3.8 ± 2.7 vs. 3.5 ± 2.9, p value = 0.7766).

We eventually examined variables that could mark a difference between mutation-positive and mutation-negative families. First, since the mean depth of sequence coverage on target is commonly defined as a parameter of good experimental quality, we compared mean on-target coverage in mutation-positive (83.8×) and mutation-negative (80.5×) families finding no significant difference (p value = 0.6719). Differences in sequence coverage are not thus likely to determine failure or success to detect the diagnostic variant in this study.

We then wondered whether the proportion of autozygous genome and/or the proportion of *candidate clinical* variants in long ROHs was on average higher in families, where a causative variant was identified. Mutation-positive families showed trend towards a higher proportion of autozygous genome compared to mutation-negative families (Fig. 1) (282 Mb vs. 234 Mb; p value = 0.08548) and a statistically significant higher number of *candidate clinical* variants (Fig. 2) (5 vs. 2.5; p value = 0.003356), this trend surviving even when diagnostic variants in mutation-positive families were excluded from the analysis (4 vs.. 2.5; p value = 0.0698).

Discussion

In this study, a collection of 47 consanguineous families with occurrence of a variety of suspected genetic conditions underwent WES analysis under suspicion of an autozygosity-related autosomal recessive disease. Therefore, an autozygosity-driven bioinformatics workflow similar to previous studies (Makrythanasis et al. 2014; Harripaul et al. 2018), with the combined analysis of uSAVs by MutPred-Splice, achieved an initial diagnostic yield of 47% (22/47 families). This number includes a family displaying a non-canonical phenotype ascribable to *DDC* mutations (Graziano et al. 2015), as well as a proband who received a partial diagnosis. This proband had glucocorticoid deficiency explained by a *MC2R* variant and sagittal synostosis, the most common form of craniostenosis known for a low contribution from monogenic forms (Wilkie et al. 2017), suggesting that the two defects are independent and that craniostenosis in this patient is not likely caused by a single variant. Moreover, in two families (Bonora et al. 2015; Pippucci et al. 2015) recessive variants were found in genes that are normally described in dominant clinically-overlapping phenotypes, a phenomenon observed also in other reports (Harel et al. 2016; Monies et al. 2017a, b). In addition, identification of an *ACP5* variant allowed clinical re-assessment of a patient

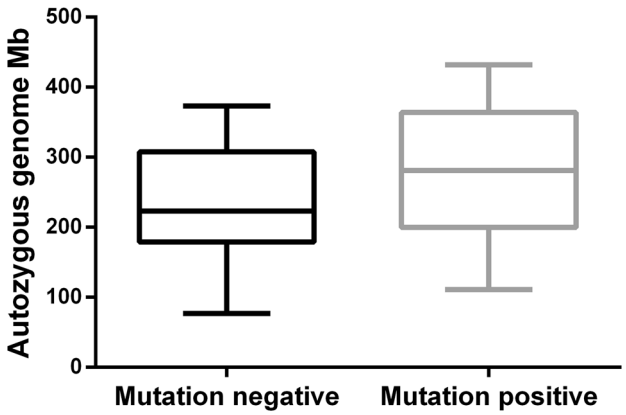

Fig. 1 Boxplot showing proportion of autozygous genome in mutation -positive and -negative families. Mutation-positive families showed trend towards a higher proportion of autozygous genome compared to mutation-negative families (282 Mb vs. 234 Mb; p value = 0.08548)

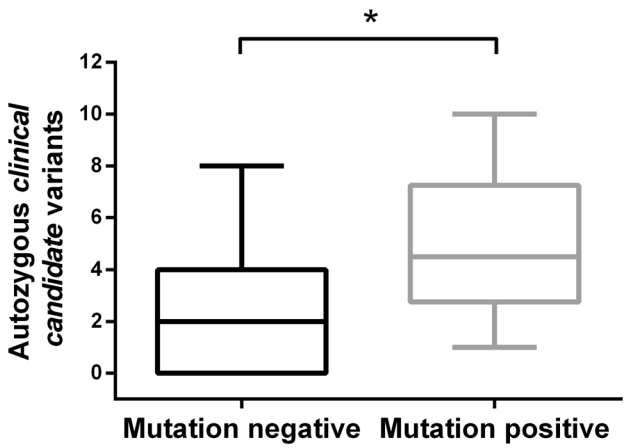

Fig. 2 Boxplot showing proportion of autozygous *clinical candidate* variants in mutation -positive and -negative families. Mutation-positive families showed a statistically significant higher number of autozygous pathogenic/likely pathogenic/potentially pathogenic variants (5 vs. 2.5; p value = 0.003356)

initially diagnosed with pure hereditary spastic paraplegia with associated autoimmune haemolitic anemia. Involvement of *ACP5*, a gene associated with spondyloenchondrodysplasia with immune dysregulation (MIM 607944) which has spasticity among its clinical signs, motivated to perform X-rays in this patient. Radiologic findings revealed the presence of vertebra's platispondilia and epiphysial radiolucencies of the radius fully consistent with the *ACP5*-related phenotypic spectrum.

Interestingly, a homozygous uSAV in *NARS2*, otherwise discarded by canonical variant filtering workflows as it caused a synonymous codon change, was eventually demonstrated to be an actual LOF variant that by impacting the normal splicing of *NARS2* exon 9 led to generation of an out-of-frame transcript, and as such pathogenic (Supplemental Fig. 1). As their functional effect is more difficult to predict than that of missense changes, synonymous variants and uSAVs in general are usually overlooked in WES studies. The autozygosity-driven strategy allowed to narrow the search space to long ROHs, thus shortening the list of MutPredSplice variants among which it was then easy to point to the *NARS2* one. Since uSAVs can behave as LOF variants, it is becoming more and more urgent to investigate their role in autosomal recessive diseases and an autozygosity-driven analysis can be of great help in this task.

We were subsequently able to increase the diagnostic yield to 55% in two ways. First, we looked for homozygous variants outside long ROHs. This revealed a *MYO15A* homozygous frame-shift small duplication that could explain the deafness phenotype, surrounded by a 849 Kb ROH shared by three Omani siblings. We demonstrated (Palombo et al. 2017) that this short-medium ROH reflected a founder haplotype introduced in Oman within the past 2–3 centuries and with an estimated carrier frequency of about 1% in Northern Oman, suggesting that focusing on large ROHs (Prasad et al. 2018; Wakeling et al. 2019) may cause to miss clinical variants that are not linked to long haplotypes arising as a result of recent parental relatedness. Second, gene matchmaking (Sobreira et al. 2015) allowed to establish candidate genes as disease-associated before they were published (Milev et al. 2018; Magini et al. 2019; Harel et al. 2020), with important implications for patient management and family counseling.

We achieved the highest diagnostic yield in probands with non-syndromic congenital deafness (66%), exceeding that obtained by WES and microarray approaches combined (56%) in another recent study on Australian non-consanguineous families (Downie et al., 2019), likely explained by the consanguinity of our families. Two studies that embraced consanguineous cohorts with multiple congenital anomalies and neurocognitive/neurodevelopmental disorders reported diagnostic yields of 38.8% (Al-Dewick et al. 2019) and 46–54% (Al-Dewick et al. 2019; Charng et al. 2016) for the two classes, respectively. These findings were comparable to our similar disease groups of syndromic CP (37%) and neurodevelopmental disorders (62.5%).

A 55% overall diagnostic yield is a striking result, but it is not unexpected in a WES study of consanguineous families. Makrythanasis et al. (2014) established diagnosis in 18/50 families of predominantly Arab ancestry (36%) (Makrythanasis et al. 2014). Yavarna et al. (2015) detected pathogenic or likely pathogenic variants in 89/149 Qatari probands (60%) (Yavarna et al. 2015). Charng et al. (2016) reported a potential molecular diagnosis in 17/31 families in Saudi Arabia (54.8%) (Charng et al. 2016). Finally, WES was used as first-line diagnostic tool in Palestinian and Israeli Arab consanguineous families, reaching a diagnostic rate of 51% (42/83) (Hengel et al. 2020). Similar high diagnostic rates were recently reported in large cohorts of consanguineous families: in a multicenter clinical exome study, Alfares et al. (2017) identified likely disease-causing variants in 222/454 probands (49%) (Alfares et al. 2017), while Al-Dewik et al. (2019) reported the Clinical Exome Sequencing experience in Qatar with 246/509 probands receiving a molecular diagnosis (48.3%) (Al-Dewik et al. 2019). In these studies, almost the entire cohorts have GME ancestry, which raises the question whether in our study GME patients, coming from countries with high inbreeding rates, received more autozygosity-related diagnoses than Italian patients. Indeed, the two groups contributed comparably to the achieved diagnostic yield. On average, the autozygous genome was larger (and the number of variants in long ROHs was higher) in GME than in Italian patients, maybe a consequence of the closer parental relatedness of GME parents. However, these differences were not significant, indicating that patients belonging to a GME population known to have high inbreeding levels and/or consanguinity rates does not necessarily have significant autozygosity excess compared to patients from consanguineous families of European ancestry. Moreover, going down from total to rare and then to *candidate clinical* variants, the excess seen in GME patients is progressively smaller, suggesting that the augmented autozygosity does not necessarily reflects in significant over-burden of deleterious variants as also indicated by the progressively weaker correlation between extent of the autozygous genome and the total/*rare/candidate clinical* variants in the whole sample.

We thus wanted to understand what could be significantly different between mutation-positive and mutation-negative patients, to get an insight into reasons that caused failure to identify the causative variant. We advanced three hypotheses:

- Differences in exome target coverage could explain why we failed to establish diagnosis in the remaining 21 families;

- Although most family loops had 1st cousins parents, extent of the autozygous genome varied substantially from family to family: families with smaller autozygous genome are, therefore, less likely to carry an autozygosity-related variant than families with larger autozygous genome;
- Even with comparable extent of the autozygous genome, in some families there may be higher occurrence of deleterious variants in shared parental haplotypes resulting in a higher chance to inherit an autozygosity-related causative variant.

While there was no apparent discrepancy in exome target coverage that could justify detection failures, we noticed that mutation-positive probands showed a trend towards augmented autozygosity and, most importantly, had significantly more autozygous deleterious variants. Notably, the extent of the "autozygome" was recently observed to correlate with the detection rate of recessive disease-causing variants in patients from a high-consanguinity community (Hengel et al. 2020). Moreover, probands of Pakistani ancestry in the Deciphering Developmental Disorders (DDD) study who were characterized by higher levels of genomic inbreeding also showed an increased rate in the diagnosis of autosomal recessive diseases (Martin et al. 2018). These observations have important implications, as they suggest that probands with lower degrees of genomic autozygosity, irrespective of the high pedigree-based inbreeding coefficient, are less likely to have their disease explained through an autozygosity-driven approach. The observation is consistent with the recent finding in 1000 genomes project data that as the genome is progressively covered by autozygous regions the rate in gain of homozygous damaging alleles outpaces that for less damaging alleles (Pemberton and Szpiech 2018). This is well illustrated by two of our families, where all patients had an extent of the autozygous genome below the median value of our cohort (265 Mb). First, the three siblings from family GME_17 with the *MYO15A* variant, a founder mutation that is not linked to recent common parental haplotypes, had relatively low autozygosity levels (ranging 62–204 Mb) although these individuals belong to a population with one of the highest inbreeding level worldwide (Islam 2012).

Second, in one of the Iranian probands the review of clinical images suggested the hypothesis of Kabuki syndrome: *ad-hoc* inspection of *KMT2D* revealed a pathogenic LOF mutation that turned out to be de novo after parental testing. Notably, also this patient had low autozygosity levels (214 Mb) and was the only one with a pathogenic/likely pathogenic variant in a known dominant gene. Assessment of this case was facilitated by a straightforward diagnosis after review, but it suggests that it is worth to consider that probands with low autozygosity levels may rather be affected by non autozygosity-related diseases. That probands from consanguineous families may have, e.g., a dominant disorder is an obvious consideration; however, the hypothesis that it is more likely to occur while autozygosity levels decrease may have consequences on study design (*trio* to have higher chance of detecting de novo variants in place of proband-only sequencing that can instead be adequate when searching for homozygous causative variants).

In conclusion, we highlighted the utility of autozygosity-driven WES analysis in an European population, where consanguinity is not widespread cultural tradition. Analysis of uSAVs, inspection of short-medium ROHs and genetic match-making were key in our study to improve diagnostic yield. Irrespective of patient's population or ethnicity, augmented genome autozygosity reflected in higher diagnostic chance under the assumption of an autozygosity-related disease. These observations may influence study design and clinical prioritization of genetic variants in autozygosity-driven WES studies.

Acknowledgements NH and MS are supported by the Deputy of Research, Isfahan University of Medical Sciences, Isfahan, Iran (Grant no. 189087). We are thankful to Dr.ssa Alessandra Maresca, IRCCS Istituto delle Scienze Neurologiche di Bologna, UOC Clinica Neurologica, Bologna, Italy, for her assistance with the Real-Time data analysis.

Author contributions FP performed research, analyzed and interpreted data. FP, TP, and CG wrote the manuscript. CG, NAW, NN, GS, CLM, GC, DMC, SG, MNAK, MAK, MS, EG, AB, and VC collected patients' clinical data and contributed the DNA samples. CM and PM performed the experiments. TP, GR, and MS designed and conceived the general overview of the study. All the authors critically revised the manuscript and accepted the final version.

Compliance with ethical standards

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

References

Alazami AM, Patel N, Shamseldin HE et al (2015) Accelerating novel candidate gene discovery in neurogenetic disorders via whole-exome sequencing of prescreened multiplex consanguineous families. Cell Rep 10(2):148–161. https://doi.org/10.1016/j.celrep.2014.12.015

Aldahmesh MA, Mohamed JY, Alkuraya HS, Verma IC, Puri RD, Alaiya AA, Rizzo WB, Alkuraya FS (2011) Recessive mutations in ELOVL4 cause ichthyosis, intellectual disability, and spastic quadriplegia. Am J Hum Genet 89(6):745–750. https://doi.org/10.1016/j.ajhg.2011.10.011

Al-Dewik N, Mohd H, Al-Mureikhi M et al (2019) Clinical exome sequencing in 509 Middle Eastern families with suspected Mendelian diseases: the Qatari experience. Am J Med Genet Part A 179A:927–935. https://doi.org/10.1002/ajmg.a.61126

Alfares A, Alfadhel M, Wani T et al (2017) A multicenter clinical exome study in unselected cohorts from a consanguineous

population of Saudi Arabia demonstrated a high diagnostic yield. Mol Genet Metab 121(2):91–95. https://doi.org/10.1016/j.ymgme.2017.04.002

Becker J, Semler O, Gilissen C et al (2011) Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta. Am J Hum Genet 88(3):362–371. https://doi.org/10.1016/j.ajhg.2011.01.015

Bittles AH (2001) Consanguinity and its relevance to clinical genetics. Clin Genet 60:89–98. https://doi.org/10.1034/j.1399-0004.2001.600201.x

Bonora E, Bianco F, Cordeddu L et al (2015) Mutations in RAD21 disrupt regulation of APOB in patients with chronic intestinal pseudo-obstruction. Gastroenterology 148(4):771–782.e11. https://doi.org/10.1053/j.gastro.2014.12.034

Briggs TA, Rice GI, Adib N et al (2016) Spondyloenchondrodysplasia due to mutations in ACP5: a comprehensive survey. J Clin Immunol 36(3):220–234. https://doi.org/10.1007/s10875-016-0252-y

Cavalli-Sforza LL, Moroni A, Zei G (2004) In: Levin SA, Horn HS (eds) Consanguinity, inbreeding, and genetic drift in Italy. Inbreeding in Italy. Monographs in Population Biology. Princeton University Press, Oxford, p 320

Charng WL, Karaca E, Coban Akdemir Z et al (2016) Exome sequencing in mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate. BMC Med Genom 9(1):42. https://doi.org/10.1186/s12920-016-0208-3

Chiang PW, Wang J, Chen Y et al (2012) Exome sequencing identifies NMNAT1 mutations as a cause of Leber congenital amaurosis. Nat Genet 44(9):972–974. https://doi.org/10.1038/ng.2370

Chong JX, Buckingham KJ, Jhangiani SN et al (2015) The genetic basis of mendelian phenotypes: discoveries, challenges, and opportunities. Am J Hum Genet 97(2):199–215. https://doi.org/10.1016/j.ajhg.2015.06.009

DePristo MA, Banks E, Poplin R et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43(5):491–501. https://doi.org/10.1038/ng.806

Downie L, Halliday J, Burt R et al (2020) Exome sequencing in infants with congenital hearing impairment: a population-based cohort study. Eur J Hum Genet 28(5):587–596. https://doi.org/10.1038/s41431-019-0553-8

Graziano C, Wischmeijer A, Pippucci T et al (2015) Syndromic intellectual disability: a new phenotype caused by an aromatic amino acid decarboxylase gene (DDC) variant. Gene 559(2):144–148. https://doi.org/10.1016/j.gene.2015.01.026

Harel T, Yoon WH, Garone C et al (2016) Recurrent de novo and biallelic variation of ATAD3A, encoding a mitochondrial membrane protein, results in distinct neurological syndromes. Am J Hum Genet 99(4):831–845. https://doi.org/10.1016/j.ajhg.2016.08.007

Harel T, Griffin JN, Arbogast T et al (2020) Loss of function mutations in CCDC32 cause a congenital syndrome characterized by craniofacial, cardiac and neurodevelopmental anomalies. Hum Mol Genet. https://doi.org/10.1093/hmg/ddaa073

Harripaul R, Vasli N, Mikhailov A et al (2018) Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. Mol Psychiatry 23(4):973–984. https://doi.org/10.1038/mp.2017.60

Hengel H, Buchert R, Sturm M et al (2020) First-line exome sequencing in Palestinian and Israeli Arabs with neurological disorders is efficient and facilitates disease gene discovery. Eur J Hum Genet. https://doi.org/10.1038/s41431-020-0609-9

Islam MM (2012) The practice of consanguineous marriage in Oman: prevalence, trends and determinants. J Biosoc Sci 44:571–594. https://doi.org/10.1017/S0021932012000016

Lander ES, Botstein D (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236:1567–1570. https://doi.org/10.1126/science.2884728

Leutenegger AL, Prum B, Genin E, Verny C, Lemainque A, Clerget-Darpoux F, Thompson EA (2003) Estimation of the inbreeding coefficient through use of genomic data. Am J Hum Genet 73(3):516–523. https://doi.org/10.1086/378207

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Magi A, Tattini L, Palombo F et al (2014) H3M2: Detection of run oh Homozygosity from whole-exome sequencing data. Bioinformatics 30(20):2852–2859. https://doi.org/10.1093/bioinformatics/btu401

Magini P, Pippucci T, Tsai IC et al (2014) A mutation in PAK3 with a dual molecular effect deregulates the RAS/MAPK pathway and drives an X-linked syndromic phenotype. Hum Mol Genet 23(13):3607–3617. https://doi.org/10.1093/hmg/ddu070

Magini P, Smits DJ, Vandervore L et al (2019) Loss of SMPD4 causes a developmental disorder characterized by microcephaly and congenital arthrogryposis. Am J Hum Genet. https://doi.org/10.1016/j.ajhg.2019.08.006

Makrythanasis P, Nelis M, Santoni FA et al (2014) Diagnostic exome sequencing to elucidate the genetic basis of likely recessive disorders in consanguineous families. Hum Mutat 35(10):1203–1210. https://doi.org/10.1002/humu.22617

Martin HC, Jones WD, McIntyre R et al (2018) Quantifying the contribution of recessive coding variation to developmental disorders. Science 362(6419):1161–1164

McQuillan R, Leutenegger AL, Abdel-Rahman R et al (2008) Runs of homozygosity in European populations. Am J Hum Genet 83(3):359–372

Milev MP, Graziano C, Karall D et al (2018) Bi-allelic mutations in TRAPPC2L result in a neurodevelopmental disorder and have an impact on RAB11 in fibroblasts. J Med Genet 55(11):753–764. https://doi.org/10.1136/jmedgenet-2018-105441

Monies D, Abouelhoda M, AlSayed M et al (2017a) The landscape of genetic diseases in Saudi Arabia based on the first 1000 diagnostic panels and exomes. Hum Genet 136(8):921–939. https://doi.org/10.1007/s00439-017-1821-8

Monies D, Maddirevula S, Kurdi W et al (2017b) Autozygosity reveals recessive mutations and novel mechanisms in dominant genes: implications in variant interpretation. Genet Med 19(10):1144–1150. https://doi.org/10.1038/gim.2017.22

Ng SB, Buckingham KJ, Lee C et al (2010) Exome sequencing identifies the cause of a Mendelian disorder. Nat Genet 42(1):30–35. https://doi.org/10.1038/ng.499

Palombo F, Al-Wardy N, Ruscone GA et al (2017) A novel founder MYO15A frameshift duplication is the major cause of genetic hearing loss in Oman. J Hum Genet 62(2):259–264. https://doi.org/10.1038/jhg.2016.120

Pemberton TJ, Szpiech ZA (2018) Relationship between deleterious variation, genomic autozygosity, and disease risk: insights from the 1000 genomes project. Am J Hum Genet 102(4):658–675. https://doi.org/10.1016/j.ajhg.2018.02.013

Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ (2012) Genomic patterns of homozygosity in worldwide human populations. Am J Hum Genet 91(2):275–292. https://doi.org/10.1016/j.ajhg.2012.06.014

Pippucci T, Parmeggiani A, Palombo F et al (2013) A novel null homozygous mutation confirms CACNA2D2 as a gene mutated in epileptic encephalopathy. PLoS ONE 8(12):e82154. https://doi.org/10.1371/journal.pone.0082154

Pippucci T, Maresca A, Magini P et al (2015) Homozygous NOTCH3 null mutation and impaired NOTCH3 signaling in recessive early-onset arteriopathy and cavitating leukoencephalopathy. EMBO Mol Med 7(6):848–858. https://doi.org/10.15252/emmm.201404399

Prasad A, Sdano MA, Vanzo RJ et al (2018) Clinical utility of exome sequencing in individuals with large homozygous regions detected by chromosomal microarray analysis. BMC Med Genet 19(1):46. https://doi.org/10.1186/s12881-018-0555-3

Ramos EM, Roca A, Chumchim N et al (2019) Primary familial brain calcification caused by a novel homozygous MYORG mutation in a consanguineous Italian family. Neurogenetics 20(2):99–102. https://doi.org/10.1007/s10048-019-00571-8

Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M (2018) CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 47(D1):D886–D894. https://doi.org/10.1093/nar/gky1016

Rhein VF, Carroll J, Ding S, Fearnley IM, Walker JE (2013) NDUFAF7 methylates arginine 85 in the NDUFS2 subunit of human complex I. J Biol Chem 288(46):33016–33026. https://doi.org/10.1074/jbc.M113.518803

Richards S, Aziz N, Bale S et al (2015) 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 17(5):405–424. https://doi.org/10.1038/gim.2015.30

Shaheen R, Faqeih E, Sunker A et al (2011) Recessive mutations in DOCK6, encoding the guanidine nucleotide exchange factor DOCK6, lead to abnormal actin cytoskeleton organization and Adams-Oliver syndrome. Am J Hum Genet 89(2):328–333. https://doi.org/10.1016/j.ajhg.2011.07.009

Smith CAB (1974) Measures of homozygosity and inbreeding in populations. Ann Hum Genet 37(4):377–391. https://doi.org/10.1111/j.1469-1809.1974.tb01844.x

Sobreira NL, Cirulli ET, Avramopoulos D et al (2010) Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. PLoS Genet 6(6):e1000991. https://doi.org/10.1371/journal.pgen.1000991

Sobreira N, Schiettecatte F, Valle D, Hamosh A (2015) GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat 36(10):928–930. https://doi.org/10.1002/humu.22844

Spataro R, Kousi M, Farhan SMK et al (2019) Mutations in ATP13A2 (PARK9) are associated with an amyotrophic lateral sclerosis-like phenotype, implicating this locus in further phenotypic expansion. Hum Genom 13(1):19. https://doi.org/10.1186/s40246-019-0203-9

Wakeling MN, Laver TW, Wright CF et al (2019) Homozygosity mapping provides supporting evidence of pathogenicity in recessive Mendelian disease. Genet Med 21(4):982–986. https://doi.org/10.1038/s41436-018-0281-4

Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from next-generation sequencing data. Nucleic Acids Res 38(16):e164. https://doi.org/10.1093/nar/gkq603

Wilkie AOM, Johnson D, Wall SA (2017) Clinical genetics of craniosynostosis. Curr Opin Pediatr 29(6):622–628. https://doi.org/10.1097/MOP.0000000000000542

Yavarna T, Al-Dewik N, Al-Mureikhi M et al (2015) High diagnostic yield of clinical exome sequencing in Middle Eastern patients with Mendelian disorders. Hum Genet 134(9):967–980. https://doi.org/10.1007/s00439-015-1575-0

Zurita Rendón O, Silva Neiva L, Sasarman F, Shoubridge EA (2014) The arginine methyltransferase NDUFAF7 is essential for complex I assembly and early vertebrate embryogenesis. Hum Mol Genet 23(19):5159–5170. https://doi.org/10.1093/hmg/ddu239

Affiliations

Flavia Palombo[1,2] · Claudio Graziano[1] · Nadia Al Wardy[3] · Nayereh Nouri[4,5] · Caterina Marconi[6] · Pamela Magini[1] · Giulia Severi[1] · Chiara La Morgia[2,7] · Gaetano Cantalupo[8,9] · Duccio Maria Cordelli[6,10] · Simone Gangarossa[11] · Mohammed Nasser Al Kindi[3] · Mazin Al Khabouri[3,12] · Mansoor Salehi[4] · Elisa Giorgio[13] · Alfredo Brusco[13] · Francesco Pisani[14] · Giovanni Romeo[6] · Valerio Carelli[2,7] · Tommaso Pippucci[1] · Marco Seri[1,6]

1 Medical Genetics Sant'Orsola, Malpighi University Hospital of Bologna, Via Massarenti 9, 40138 Bologna, Italy

2 IRCCS Istituto Delle Scienze Neurologiche Di Bologna, UOC Clinica Neurologica, Bologna, Italy

3 Department of Biochemistry, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Oman

4 Department of Genetics and Molecular Biology, Isfahan University of Medical Sciences, Isfahan, Iran

5 Craniofacial and Cleft Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

6 Department of Medical and Surgical Sciences (DIMEC), University of Bologna, Bologna, Italy

7 Department of Biomedical and Neuromotor Sciences (DIBINEM), University of Bologna, Bologna, Italy

8 Child Neuropsychiatry, Department of Surgical Sciences, Dentistry, Gynecology and Pediatrics, University of Verona, Verona, Italy

9 UOC Neuropsichiatria Infantile, DAI Materno-Infantile, AOUI Verona, Verona, Italy

10 Neuropsychiatry Sant'Orsola-Malpighi University Hospital of Bologna, Bologna, Italy

11 ASP7 Ragusa, Ragusa, Italy

12 Department of ENT, Al Nahdha Hospital, Ministry of Health, Muscat, Oman

13 Department of Medical Sciences, University of Torino, Turin, Italy

14 Child Neuropsychiatry Unit, Department of Medicine & Surgery, University of Parma, Parma, Italy