Incidental Medical Information in Whole-Exome Sequencing

AUTHORS: Benjamin D. Solomon, MD,[a] Donald W. Hadley, MS, CGC,[b] Daniel E. Pineda-Alvarez, MD,[a] NISC Comparative Sequencing Program,[c] Aparna Kamat, MSc,[d] Jamie K. Teer, PhD,[e] Praveen F. Cherukuri, PhD,[d] Nancy F. Hansen, PhD,[d] Pedro Cruz, MSc,[d] Alice C. Young, MSc,[d] Benjamin E. Berkman, JD, MPH,[f] Settara C. Chandrasekharappa, PhD,[g] and James C. Mullikin, PhD[d]

[a]Medical Genetics Branch, [b]Social and Behavioral Research Branch, [c]National Institutes of Health Intramural Sequencing Center, [d]Genome Technology Branch, [e]Genetic Disease Research Branch, [f]Department of Bioethics, Clinical Center, and Office of the Clinical Director, and [g]Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland

abstract

Genomic technologies, such as whole-exome sequencing, are a powerful tool in genetic research. Such testing yields a great deal of incidental medical information, or medical information not related to the primary research target. We describe the management of incidental medical information derived from whole-exome sequencing in the research context. We performed whole-exome sequencing on a monozygotic twin pair in which only 1 child was affected with congenital anomalies and applied an institutional review board–approved algorithm to determine what genetic information would be returned. Whole-exome sequencing identified 79 525 genetic variants in the twins. Here, we focus on novel variants. After filtering artifacts and excluding known single nucleotide polymorphisms and variants not predicted to be pathogenic, the twins had 32 novel variants in 32 genes that were felt to be likely to be associated with human disease. Eighteen of these novel variants were associated with recessive disease and 18 were associated with dominantly manifesting conditions (variants in some genes were potentially associated with both recessive and dominant conditions), but only 1 variant ultimately met our institutional review board–approved criteria for return of information to the research participants. *Pediatrics* 2012;129:e1605–e1611

The advent of "Next Generation" sequencing technologies has allowed large-scale genomic sequencing to become widely used in genetic research. One type of genomic analysis, whole-exome sequencing, refers to the sequencing of the exome, or all the known coding regions (~1%) of the genome.[1] This technology is an efficient, affordable, and powerful research tool.[1–6] Genomic sequencing necessarily reveals incidental medical information, or medical information that has potential health or reproductive importance, but is not related to the primary research question.[1,7–9] Managing this incidental medical information presents many logistical and ethical challenges, especially as relates to novel variants.[10,11] Guidelines have been proposed to help guide the return of incidental genetic information, but it can be difficult to conceptualize the ramifications of a given algorithm in the abstract.[12–14] Here, we illustrate the incidental information revealed by whole-exome sequencing and provide an example of the decision process by which novel variants were determined to meet criteria for return to participants.

PATIENT PRESENTATION

Through our National Human Genome Research Institute institutional review board (IRB)-approved protocol on VACTERL (vertebral defects–anal atresia–cardiovascular anomalies–tracheoesophageal fistula with esophageal atresia–radial and renal dysplasia–limb defects) association, which uses whole-exome sequencing (among other research modalities), we studied a monozygotic twin-pair in which only 1 twin was affected, with appropriate consent obtained from all participants, and with a separate consent required to perform genomic sequencing. Potential incidental medical information is discussed in detail during the consent process, and research participants choose whether to learn the results of genomic sequencing, both regarding the primary

research target, as well as related to incidental medical information.

We extracted DNA from peripheral blood samples. Confirmation of mutations that met criteria for return was undertaken through Clinical Laboratory Improvement Amendment, 1988 (CLIA)-approved laboratories, either through DNA extracted from lymphoblastoid cell lines, or from a new blood sample. For specific genes for which CLIA testing is not available, we use commercial laboratories that are able to perform CLIA-based confirmation of any genetic variant identified through research laboratories (we perform confirmation through GeneDx, Gaithersburg, MD). Of note, some research laboratories also have their own, noncommercial CLIA laboratories that are able to perform CLIA verification on any genetic variant.

See Supplemental Information for detailed whole-exome sequencing and analysis methods. After variant analysis, the data were initially filtered to eliminate likely false-positives; genotypes were called at all positions with high-quality sequence bases (Phred-like Q20 or greater) by using a Bayesian algorithm (Most Probable Genotype [MPG]).[15] Genotypes with MPG score ≥ 10 demonstrate $>99.89\%$ concordance with high-density single-nucleotide polymorphism (SNP) array data, and were considered high quality if they also had an MPG score/coverage ratio of ≥ 0.5. Variants predicted not to result in potential pathogenicity were excluded (based on Conserved Domain-based Prediction scores that predict pathogenicity; see http://research.nhgri.nih.gov/software/VarSifter for details). High-quality predicted pathogenic variants were analyzed by using standard available databases, including the Human Genome Mutation Database (professional version) (http://www.hgmd.cf.ac.uk/ac/index.php) and Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/omim), and through literature review. Further comparison of

variants of interest was performed versus results of whole-exome sequencing of 479 whole-exome samples (sequenced at the same sequencing facility) ascertained from the ClinSeq cohort, which ascertains patients with a phenotypic continuum from unaffected, to those who have had myocardial infarctions.[16] A working committee consisting of board-certified clinical geneticists, board-certified molecular geneticists, board-certified genetic counselors, bioethicists, and National Human Genome Research Institute IRB members, as well as other genetic researchers, convened to discuss variants that met the above criteria. Finally, in several cases, experts in the study of individual genes and conditions were contacted when results remained equivocal. IRB-approved guidelines regarding what incidental medical information would be returned to study participants are summarized as follows:

1. The genetic change must be known or predicted to be of urgent clinical significance.

2. Knowledge of the finding must have a clear direct benefit that would be lost if diagnosis was made later; that is, knowledge of this risk factor would substantially alter medical or reproductive decision-making.

3. The potential benefit of knowing a genetic disorder exists clearly outweighs the potential risks of anxiety and subsequent medical testing that could result from this knowledge.

4. Unless they add substantial risk, risk factors for multifactorial disorders are not reported.

5. Recessive mutations will be reported only if (1) the carrier frequency for mutations in that specific gene is $>1\%$ (such that the disease incidence is more than 1/40 000); (2) the syndrome results in significant morbidity; or (3) early diagnosis and intervention would have significant benefit.

Monozygosity was initially confirmed by high-density SNP microarray. Whole-exome

sequencing did not reveal an obvious genetic cause of the congenital anomalies in the affected twin, although studies are ongoing regarding several variants of interest. High-quality genotypes were 98.9% concordant in the twins; see Table 1 for details of overall sequencing results. After eliminating known SNPs not meeting criteria outlined above and likely nonpathogenic variants, as well as artifacts (both by examining MPG and coverage data from next-generation sequencing, as well as by comparison with 479 other whole-exome samples), a total of 412 variants were identified. Of these, 32 novel variants (25 missense and 7 insertion-deletions) in 30 genes were felt to be likely associated with human disease (Table 2). Eighteen variants were associated with recessive diseases (manifesting in the homozygous or compound heterozygous state), and 17 were associated with conditions for which the presence of a single mutation (heterozygosity) was associated with disease in a dominant model (variants in some genes were associated with both recessive and dominant conditions). Some of these dominant-model genes were associated with increased susceptibility to a complex condition, such as schizophrenia or inflammatory bowel disease, whereas others were related to more traditional Mendelian disorders, although there was no clinical evidence for these disorders in the

participants or their families, indicating that the variant was nonpathogenic, nonpenetrant, or that the reported gene-disease association was spurious.

Three heterozygous variants initially met criteria for return of information: *CACNA1S*, associated with hypokalemic periodic paralysis and malignant hyperthermia[17,18]; *CPSI*, associated with Carbamoyl Phosphate Synthetase I deficiency, as well as pulmonary artery hypertension,[19–21] the latter of which this participant had; and *CYP21A2*, associated with 21-hydroxylase deficiency leading to congenital adrenal hyperplasia.[22] Experts in each of the diseases/genes were contacted, as pathogenicity was not clear in all cases. Of these, after extensive discussion, it was unclear (but felt to be unlikely) whether the specific variant identified in *CACNA1S* was pathogenic, and because the research participants and their relatives showed no signs of these *CACNA1S*-related conditions, it was determined that this result did not meet criteria for return. In discussion with individual researchers, strong evidence emerged that the variant in *CYP21A2* was nonpathogenic through (private, unpublished) work of researchers specializing in this disease (Maria New, MD; Tony Yuen, PhD). There was evidence that the *CPSI* mutation was pathogenic as relates to postoperative pulmonary artery hypertension, which the affected twin suffered, and, as there is

a potential intervention (special anesthesiology considerations in future surgeries), this was deemed to meet requirements for return to the participants.

DISCUSSION

New technologies allow researchers to examine large portions of the genome with relative ease. These are valuable tools, and reveal a large amount of potentially medically significant information that, with in-depth analysis, can be used to facilitate health care[8,9]; however, the presence of incidental medical information may also impede the use of genomic sequencing. Challenges directly related to incidental genomic information in clinical practice involve complex and resource-consuming interpretation and validation of data, the possibility of subjecting patients to risky and unnecessary follow-up testing, and questions about the overall risk-benefit ratio of conducting such testing.[11] In the research setting, similar issues also apply, and hinge on specific IRB-approved guidelines.

One central issue in the research context involves defining and determining what genetic information should be returned to research participants who undergo this type of sequencing. In our experience, and in discussions with a number of other researchers using these sequencing methods, there is a wide range of opinions. These opinions range from returning no incidental medical information (because of logistical concerns as well as the argument that that the overall risk would outweigh the benefits), to returning large amounts of information involving personal and familial genetic risk factors for disease, to returning all genetic data in an uncurated fashion so that research participants and their physicians can access this information prospectively. Likewise, although there is no single accepted algorithm in the medical literature, various criteria have been proposed

TABLE 1 Overview of Variants Detected in the Monozygotic Twins Through Whole-Exome Sequencing

Total variants[a]	79 525
Total variants with high-quality genotypes[b]	Twin A: 62 504
	Twin B: 64 926
Total variants where there is a high-quality genotype available for both twins	52 143 (98.9% concordance: 51 566 concordant variants; 577 discordant variants)
Variants in dbSNP (build 130)	43 300
Synonymous variants	8428
Nonsynonymous variants	7660
In-frame insertions/deletions	77
Frameshift mutations	68
Nonsense variants	49
Splice-site mutations	37

[a] Differences from the human reference sequence, build National Center for Biotechnology Information 36.
[b] Based on MPG score $\geq$10 and score/coverage ratio $\geq$0.5.

TABLE 2 Identified Genes With Novel Variants Potentially Associated With Disease, Including Carrier States for Recessive Conditions

Number	Gene	Variant Type	Disease(s)[a]	Type of Association	Notes/Reason Information Not Returned	Reference(s)
1	ABCC8	M	Hyperinsulinemic hypoglycemia, permanent neonatal diabetes with neurologic features, transient and permanent neonatal diabetes mellitus	Disease-associated gene (dominant)[a] and/or carrier for recessive disorder	No signs in participants or family, not clear if this specific variant is associated with this condition in this family	23–26
2	AK2	M	Reticular dysgenesis	Carrier for recessive disorder	Rare recessive disorder	27
3	CACNA1S	M	Malignant hyperthermia, hypokalemic periodic paralysis	Susceptibility for conditions and/or disease-associated gene (dominant)	The specific variant in this family was ultimately not felt to be pathogenic per discussion with multiple experts in the conditions studied	17,18
4	CD320	DIV	Methylmalonic aciduria (owing to transcobalamin receptor defect)	Carrier for recessive disorder	Rare recessive disorder	28
5	CISD2[b]	M	Wolfram syndrome	Carrier for recessive disorder	Rare recessive disorder	29
6	COL5A2	M	Ehlers-Danlos syndrome type I	Disease-associated gene (dominant)	The individuals and their family do not meet criteria for diagnosis, but preliminary evidence suggests that this variant may be related to the congenital anomaly that is the primary focus of this research	30
7	COL6A3	DIV	Bethlem myopathy, Ullrich congenital muscular dystrophy (autosomal recessive and dominant forms)	Disease-associated gene (dominant) and/or carrier for recessive disorder	No signs in participants or family, not clear if this specific variant is associated with this condition in this family	31–33
8	CPS1	M	Carbamoyl phosphate synthetase deficiency, pulmonary artery hypertension	Carrier for recessive disorder and possibly related to susceptibility to conditions	Rare recessive disorder	19–21
9	CRYGD	M	Progressive juvenile-onset punctate cataracts	Disease-associated gene (dominant)	No signs in participants or family, not clear if this specific variant is associated with this condition in this family	34
10	CYP21A2	DIV	Congenital adrenal hyperplasia	Carrier for recessive disorder	Normal variant in private databases of researchers studying this gene/condition	22
11	DBH	M	Dopamine β-hydroxylase deficiency, norepinephrine deficiency	Carrier for recessive disorders	Rare recessive disorder	35–37
12	DDX11	M	Warsaw breakage syndrome	Carrier for recessive disorder	Rare recessive disorder	38
13	DISC1	M	Possible association with schizophrenia	Possibly related to susceptibility for a condition	Susceptibility factor, not clear if this specific variant is associated with this condition in this family	39–42
14	DNAH5	M	Primary ciliary dyskinesia type 3	Carrier for recessive disorder	Primary ciliary dyskinesia is more common than 1/40 000, but only about 25% of cases are caused by mutations in DNAH5	43
15	FOXD4[b]	M	Cardiomyopathy, obsessive-compulsive disorder, suicidality	Disease-associated gene (dominant)	Complex purported gene-associated phenotype reported in 1 family, susceptibility factor	44
16	FUT7	M	Possible autosomal recessive association with multiple autoimmune conditions	Carrier for recessive disorder	Rare recessive disorder	45
17	GIGYF2	DIV	Familial Parkinson's disease (incompletely penetrant)	Possibly related to susceptibility for a condition	No signs in family, not clear if this specific variant is associated with this condition in this family	46,47
18	HMOX1	M	Heme-oxygenase 1 deficiency	Carrier for recessive disorder and possibly related to susceptibility for a condition	Rare recessive disorder; possible association with COPD is a susceptibility factor	48–50
19	HSPG2	M	Schwartz-Jampel syndrome type 1; dyssegmental dysplasia, Silverman-Handmaker type	Carrier for recessive disorders	Rare recessive disorder	51
20	KRT18	M	Possible association with susceptibility to cirrhosis	Possibly related to susceptibility for a condition	Susceptibility factor	52–55
21	MST1	M	Possible association with inflammatory bowel disease	Possibly related to susceptibility for a condition	Susceptibility factor	56,57

TABLE 2 Continued

Number	Gene	Variant Type	Disease(s)[a]	Type of Association	Notes/Reason Information Not Returned	Reference(s)
22	NLRPI2	M	Familial cold autoinflammatory syndrome	Disease-associated gene (dominant)	No signs in participants or family, not clear if this specific variant is associated with this condition in this family	58
23	ORAI1	DIV	Immune dysfunction	Carrier for recessive disorder	Rare recessive disorder, although heterozygotes may have subtle subclinical manifestations	59
24	PRDM9	M	Possible association with azoospermia	Possibly related to susceptibility for a condition	Susceptibility factor	60, 61
25	PRKCSH	DIV	Autosomal dominant polycystic liver disease	Disease-associated gene (dominant)	No signs in participants or family, not clear if this specific variant is associated with this condition in this family	62
26	PRKRA	M	Dystonia 16	Carrier for recessive disorder	Generally viewed as a rare recessive disorder, although affected heterozygote reported (possibly related to compound heterozygosity or other explanation)	63,64
27	PSPH[b]	M	Phosphoserine phosphatase deficiency	Carrier for recessive disorder	Rare recessive disorder	65
28	SCARB2	M	Action myoclonus-renal failure	Carrier for recessive disorder	Rare recessive disorder	66
29	SLC1A3	DIV	Episodic ataxia, type 6	Disease-associated gene (dominant)	No signs in participants or family, not clear if this specific variant is associated with this condition in this family	67
30	TH	M	Segawa syndrome	Carrier for recessive disorder	Rare recessive disorder	68
31	TRPC7	M	Possible association with bipolar affective disorder, ALS, and Parkinsonism	Possibly related to susceptibility for conditions	Susceptibility factors, no family history of these conditions and not clear if this specific variant is associated with this condition in this family	69,70
32	USP26	M	Possible association with impaired spermatogenesis	Possibly related to susceptibility for a condition	Susceptibility factors, not clear if this specific variant is associated with this condition in this family	71–74

As known SNPs do meet criteria for return of information when associated with increased susceptibility for a genetic condition, these are not included here. ALS, amyotrophic lateral sclerosis; COPD, chronic obstructive pulmonary disease; DIV, in-frame deletion or insertion variant; M, missense (nonsynonymous) variant.

a "'Disease-associated genes (dominant)'" indicates that studies show that only 1 mutant gene is required for the presence of disease, as opposed to recessive conditions, in which carriers of a single mutation are classically considered to be unaffected.

b Variants called in only 1 of the twins; however, in the twin in whom the variant was not called, the variant is likely present, as the coverage was low for that base in the twin without the variant.

regarding which information should be returned. Unifying themes of these recommendations involve the following: allowing research participants to choose whether they want to learn about incidental medical information; the need for filtering and confirmation of variants owing to the inevitable false-positive variant calls; the requirement to show that the findings are valid in terms of pathogenicity, which can be far more demanding; and requirements that the returned genetic information involves a relatively high risk for disease, that the information is actionable, and that the benefits of knowing the genetic information outweigh the risks.[7,11,13] Other recommendations focus on specific context, taking into account the nature of the variant, the parameters of the research study, and the characteristics of the research participants.[12,14]

Despite optimal guidance and careful algorithms, there will inevitably be variants that fall into "gray zones" in the decision-making process. For such variants, we use a multidisciplinary approach drawing on experts from multiple fields, but this can be a laborious process that draws resources away from direct research goals. This report demonstrates that a large amount of effort was required to return only 1 incidental genomic variant found through whole-exome sequencing. One proposal to help manage the workload in a streamlined manner is to establish collective groups experienced in interpretation of genomic sequencing to automate the process of both analyzing research results, as well as aiding in interpretation, validation, and return of results.

REFERENCES

1. Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. *Am J Hum Genet*. 2009;85(2):142–154

2. Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet*. 2010;19(R2):R145–R151

3. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*. 2010;42(1):30–35

4. Roach JC, Glusman G, Smit AF, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science*. 2010;328(5978):636–639

5. Ng SB, Bigham AW, Buckingham KJ, et al. Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome. *Nat Genet*. 2010;42(9):790–793

6. Ng SB, Nickerson DA, Bamshad MJ, Shendure J. Massively parallel sequencing and rare disease. *Hum Mol Genet*. 2010;19(R2):R119–R124

7. Wolf SM, Lawrenz FP, Nelson CA, et al. Managing incidental findings in human subjects research: analysis and recommendations. *J Law Med Ethics*. 2008;36(2):219–248, 211

8. Lupski JR, Reid JG, Gonzaga-Jauregui C, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med*. 2010;362(13):1181–1191

9. Ashley EA, Butte AJ, Wheeler MT, et al. Clinical assessment incorporating a personal genome. *Lancet*. 2010;375(9725):1525–1535

10. McGuire AL, Lupski JR. Personal genome research: what should the participant be told? *Trends Genet*. 2010;26(5):199–201

11. Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. *JAMA*. 2006;296(2):212–215

12. Ravitsky V, Wilfond BS. Disclosing individual genetic results to research participants. *Am J Bioeth*. 2006;6(6):8–17

13. Fabsitz RR, McGuire A, Sharp RR, et al; National Heart, Lung, and Blood Institute working group. Ethical and practical guidelines for reporting genetic research results to study participants: updated guidelines from a National Heart, Lung, and Blood Institute working group. *Circ Cardiovasc Genet*. 2010;3(6):574–580

14. Beskow LM, Burke W. Offering individual genetic research results: context matters. *Sci Transl Med*. 2010;2(38):38cm20

15. Teer JK, Bonnycastle LL, Chines PS, et al; NISC Comparative Sequencing Program. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. *Genome Res*. 2010;20(10):1420–1431

16. Biesecker LG, Mullikin JC, Facio FM, et al; NISC Comparative Sequencing Program. The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine. *Genome Res*. 2009;19(9):1665–1674

17. Jurkat-Rott K, Lehmann-Horn F, Elbaz A, et al. A calcium channel mutation causing hypokalemic periodic paralysis. *Hum Mol Genet*. 1994;3(8):1415–1419

18. Monnier N, Procaccio V, Stieglitz P, Lunardi J. Malignant-hyperthermia susceptibility is associated with a mutation of the alpha 1-subunit of the human dihydropyridine-sensitive L-type voltage-dependent calcium-channel receptor in skeletal muscle. *Am J Hum Genet*. 1997;60(6):1316–1325

19. Hoshide R, Matsuura T, Haraguchi Y, Endo F, Yoshinaga M, Matsuda I. Carbamyl phosphate synthetase I deficiency. One base substitution in an exon of the *CPS I* gene causes a 9-basepair deletion due to aberrant splicing. *J Clin Invest*. 1993;91(5):1884–1887

20. Pearson DL, Dawling S, Walsh WF, et al. Neonatal pulmonary hypertension—urea-cycle intermediates, nitric oxide production, and carbamoyl-phosphate synthetase function. *N Engl J Med*. 2001;344(24):1832–1838

21. Summar ML, Gainer JV, Pretorius M, et al. Relationship between carbamoyl-phosphate synthetase genotype and systemic vascular function. *Hypertension*. 2004;43(2):186–191

22. White PC, Chaplin DD, Weis JH, Dupont B, New MI, Seidman JG. Two steroid 21-hydroxylase genes are located in the murine S region. *Nature*. 1984;5;312(5993):465–467

23. Thomas PM, Cote GJ, Wohllk N, et al. Mutations in the sulfonylurea receptor gene in familial persistent hyperinsulinemic hypoglycemia of infancy. *Science*. 1995;268(5209):426–429

24. Babenko AP, Polak M, Cavé H, et al. Activating mutations in the *ABCC8* gene in neonatal diabetes mellitus. *N Engl J Med*. 2006;355(5):456–466

25. Proks P, Arnold AL, Bruining J, et al. A heterozygous activating mutation in the sulphonylurea receptor SUR1 (ABCC8) causes neonatal diabetes. *Hum Mol Genet*. 2006;15(11):1793–1800

26. Ellard S, Flanagan SE, Girard CA, et al. Permanent neonatal diabetes caused by dominant, recessive, or compound heterozygous SUR1 mutations with opposite functional effects. *Am J Hum Genet*. 2007;81(2):375–382

27. Pannicke U, Hönig M, Hess I, et al. Reticular dysgenesis (aleukocytosis) is caused by mutations in the gene encoding mitochondrial adenylate kinase 2. *Nat Genet*. 2009;41(1):101–105

28. Quadros EV, Lai SC, Nakayama Y, et al. Positive newborn screen for methylmalonic aciduria identifies the first mutation in *TCblR/CD320*, the gene for cellular uptake of transcobalamin-bound vitamin B(12). *Hum Mutat*. 2010;31(8):924–929

29. Amr S, Heisey C, Zhang M, et al. A homozygous mutation in a novel zinc-finger protein, *ERIS*, is responsible for Wolfram syndrome 2. *Am J Hum Genet*. 2007;81(4):673–683

30. Michalickova K, Susic M, Willing MC, Wenstrup RJ, Cole WG. Mutations of the alpha2(V) chain of type V collagen impair matrix assembly and produce Ehlers-Danlos syndrome type I. *Hum Mol Genet*. 1998;7(2):249–255

31. Pan TC, Zhang RZ, Pericak-Vance MA, et al. Missense mutation in a von Willebrand factor type A domain of the alpha 3(VI) collagen gene (*COL6A3*) in a family with Bethlem myopathy. *Hum Mol Genet*. 1998;7(5):807–812

32. Demir E, Sabatelli P, Allamand V, et al. Mutations in *COL6A3* cause severe and mild phenotypes of Ullrich congenital muscular dystrophy. *Am J Hum Genet*. 2002;70(6):1446–1458

33. Baker NL, Mörgelin M, Peat R, et al. Dominant collagen VI mutations are a common cause of Ullrich congenital muscular dystrophy. *Hum Mol Genet*. 2005;14(2):279–293

34. Stephan DA, Gillanders E, Vanderveen D, et al. Progressive juvenile-onset punctate cataracts caused by mutation of the gammaD-crystallin gene. *Proc Natl Acad Sci U S A*. 1999;96(3):1008–1012

35. McKinney EF, Walton RT, Yudkin P, et al. Association between polymorphisms in dopamine metabolic enzymes and tobacco consumption in smokers. *Pharmacogenetics*. 2000;10(6):483–491

36. Lea RA, Dohy A, Jordan K, Quinlan S, Brimage PJ, Griffiths LR. Evidence for allelic association of the dopamine beta-hydroxylase gene (DBH) with susceptibility to typical migraine. *Neurogenetics*. 2000;3(1):35–40

37. Kim CH, Zabetian CP, Cubells JF, et al. Mutations in the dopamine beta-hydroxylase gene are associated with human norepinephrine deficiency. *Am J Med Genet*. 2002;108(2):140–147

38. van der Lelij P, Chrzanowska KH, Godthelp BC, et al. Warsaw breakage syndrome, a cohesinopathy associated with mutations in the XPD helicase family member *DDX11/ChlR1*. *Am J Hum Genet*. 2010;86(2):262–266

39. Millar JK, Wilson-Annan JC, Anderson S, et al. Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum Mol Genet*. 2000;9(9):1415–1423

40. Ekelund J, Hovatta I, Parker A, et al. Chromosome 1 loci in Finnish schizophrenia families. *Hum Mol Genet.* 2001;10(15):1611–1617

41. Song W, Li W, Feng J, Heston LL, Scaringe WA, Sommer SS. Identification of high risk DISC1 structural variants with a 2% attributable risk for schizophrenia. *Biochem Biophys Res Commun.* 2008;367(3):700–706

42. Schumacher J, Laje G, Abou Jamra R, et al. The DISC locus and schizophrenia: evidence from an association study in a central European sample and from a meta-analysis across different European populations. *Hum Mol Genet.* 2009;18(14):2719–2727

43. Olbrich H, Häffner K, Kispert A, et al. Mutations in *DNAH5* cause primary ciliary dyskinesia and randomization of left-right asymmetry. *Nat Genet.* 2002;30(2):143–144

44. Minoretti P, Arra M, Emanuele E, et al. A W148R mutation in the human FOXD4 gene segregating with dilated cardiomyopathy, obsessive-compulsive disorder, and suicidality. *Int J Mol Med.* 2007;19(3):369–372

45. Bengtson P, Larson C, Lundblad A, Larson G, Påhlsson P. Identification of a missense mutation (G329A;Arg(110)—> GLN) in the human FUT7 gene. *J Biol Chem.* 2001;276 (34):31575–31582

46. Lautier C, Goldwurm S, Dürr A, et al. Mutations in the *GIGYF2* (*TNRC15*) gene at the PARK11 locus in familial Parkinson disease. *Am J Hum Genet.* 2008;82(4):822–833

47. Tan EK, Lin CH, Tai CH, et al. Non-synonymous *GIGYF2* variants in Parkinson's disease from two Asian populations. *Hum Genet.* 2009;126 (3):425–430

48. Yachie A, Niida Y, Wada T, et al. Oxidative stress causes enhanced endothelial cell injury in human heme oxygenase-1 deficiency. *J Clin Invest.* 1999;103(1):129–135

49. Yamada N, Yamaya M, Okinaga S, et al. Microsatellite polymorphism in the *heme oxygenase-1* gene promoter is associated with susceptibility to emphysema. *Am J Hum Genet.* 2000;66(1):187–195

50. Guénégou A, Leynaert B, Bénessiano J, et al. Association of lung function decline with the *heme oxygenase-1* gene promoter microsatellite polymorphism in a general population sample. Results from the European Community Respiratory Health Survey (ECRHS), France. *J Med Genet.* 2006;43 (8):e43

51. Nicole S, Davoine CS, Topaloglu H, et al. Perlecan, the major proteoglycan of basement membranes, is altered in patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia). *Nat Genet.* 2000;26 (4):480–483

52. Arikawa-Hirasawa E, Wilcox WR, Le AH, et al. Dyssegmental dysplasia, Silverman-Handmaker type, is caused by functional null mutations of the *perlecan* gene. *Nat Genet.* 2001;27(4): 431–434

53. Ku NO, Wright TL, Terrault NA, Gish R, Omary MB. Mutation of human keratin 18 in association with cryptogenic cirrhosis. *J Clin Invest.* 1997;99(1):19–23

54. Ku NO, Darling JM, Krams SM, et al. Keratin 8 and 18 mutations are risk factors for developing liver disease of multiple etiologies. *Proc Natl Acad Sci U S A.* 2003;100(10):6063–6068

55. Strnad P, Zhou Q, Hanada S, et al; Acute Liver Failure Study Group. Keratin variants predispose to acute liver failure and adverse outcome: race and ethnic associations. *Gastroenterology.* 2010;139(3):828–835, 835.e1-3

56. McGovern DP, Gardet A, Törkvist L, et al; NIDDK IBD Genetics Consortium. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat Genet.* 2010;42(4):332–337

57. Tong P, Prendergast JG, Lohan AJ, et al. Sequencing and analysis of an Irish human genome. *Genome Biol.* 2010;11(9):R91

58. Jéru I, Duquesnoy P, Fernandes-Alnemri T, et al. Mutations in *NALP12* cause hereditary periodic fever syndromes. *Proc Natl Acad Sci U S A.* 2008;105(5):1614–1619

59. Feske S, Gwack Y, Prakriya M, et al. A mutation in *Orai1* causes immune deficiency by abrogating CRAC channel function. *Nature.* 2006;441(7090):179–185

60. Irie S, Tsujimura A, Miyagawa Y, et al. Single-nucleotide polymorphisms of the *PRDM9* (*MEISETZ*) gene in patients with nonobstructive azoospermia. *J Androl.* 2009;30 (4):426–431

61. Miyamoto T, Koh E, Sakugawa N, et al. Two single nucleotide polymorphisms in PRDM9 (MEISETZ) gene may be a genetic risk factor for Japanese patients with azoospermia by meiotic arrest. *J Assist Reprod Genet.* 2008;25(11-12):553–557

62. Drenth JP, te Morsche RH, Smink R, Bonifacino JS, Jansen JB. Germline mutations in *PRKCSH* are associated with autosomal dominant polycystic liver disease. *Nat Genet.* 2003;33(3):345–347

63. Camargos S, Scholz S, Simón-Sánchez J, et al. DYT16, a novel young-onset dystonia-parkinsonism disorder: identification of a segregating mutation in the stress-response protein PRKRA. *Lancet Neurol.* 2008;7(3):207–215

64. Seibler P, Djarmati A, Langpap B, et al. A heterozygous frameshift mutation in *PRKRA* (DYT16) associated with generalised dystonia in a German patient. *Lancet Neurol.* 2008;7(5):380–381

65. Veiga-da-Cunha M, Collet JF, Prieur B, et al. Mutations responsible for 3-phosphoserine phosphatase deficiency. *Eur J Hum Genet.* 2004;12(2):163–166

66. Berkovic SF, Dibbens LM, Oshlack A, et al. Array-based gene discovery with three unrelated subjects shows SCARB2/LIMP-2 deficiency causes myoclonus epilepsy and glomerulosclerosis. *Am J Hum Genet.* 2008; 82(3):673–684

67. Jen JC, Wan J, Palos TP, Howard BD, Baloh RW. Mutation in the glutamate transporter *EAAT1* causes episodic ataxia, hemiplegia, and seizures. *Neurology.* 2005;65(4):529–534

68. Lüdecke B, Dworniczak B, Bartholomé K. A point mutation in the tyrosine hydroxylase gene associated with Segawa's syndrome. *Hum Genet.* 1995;95(1):123–125

69. McQuillin A, Bass NJ, Kalsi G, et al. Fine mapping of a susceptibility locus for bipolar and genetically related unipolar affective disorders, to a region containing the *C21ORF29* and *TRPM2* genes on chromosome 21q22.3. *Mol Psychiatry.* 2006;11(2):134–142

70. Hermosura MC, Cui AM, Go RC, et al. Altered functional properties of a TRPM2 variant in Guamanian ALS and PD. *Proc Natl Acad Sci U S A.* 2008;105(46):18029–18034

71. Stouffs K, Lissens W, Tournaye H, Van Steirteghem A, Liebaers I. Possible role of USP26 in patients with severely impaired spermatogenesis. *Eur J Hum Genet.* 2005;13 (3):336–340

72. Paduch DA, Mielnik A, Schlegel PN. Novel mutations in testis-specific ubiquitin protease 26 gene may cause male infertility and hypogonadism. *Reprod Biomed Online.* 2005;10(6):747–754

73. Stouffs K, Lissens W, Tournaye H, Van Steirteghem A, Liebaers I. Alterations of the *USP26* gene in Caucasian men. *Int J Androl.* 2006;29(6):614–617

74. Zhang J, Qiu SD, Li SB, et al. Novel mutations in ubiquitin-specific protease 26 gene might cause spermatogenesis impairment and male infertility. *Asian J Androl.* 2007;9 (6):809–814