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, a Aparna Kamat, MSc, d Jamie K. Teer, PhD, d Praveen F. Cherukuri, PhD, d Nancy F. Hansen, PhD, d Pedro Cruz, MSc, d Alice C. Young, MSc, d Benjamin E. Berkman, JD, MPH, d Settara C. Chandrasekharappa, PhD, d 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

KEY WORDS
whole-exome sequencing, incidental medical information

ABBREVIATIONS
CLIA—Clinical Laboratory Improvement Amendment, 1988; IRB—institutional review board; MPG—Most Probable Genotype; SNP—single-nucleotide polymorphism

All authors have made substantive intellectual contributions to this study. All authors have made substantial contributions in each of the following categories: (1) conception and design, acquisition of data, or analysis and interpretation of data; (2) drafting the article or revising it critically for important intellectual content; and (3) final approval of the version to be published.

www.pediatrics.org/cgi/doi/10.1542/peds.2011-0080
doi:10.1542/peds.2011-0080

Accepted for publication Jan 11, 2012

Address correspondence to Benjamin D. Solomon, MD, NIH, MSC 3717, Building 35, Room 1B-207, Bethesda, MD 20892. E-mail: solomonn@mail.nih.gov

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275); published in the public domain by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: Supported by the Division of Intramural Research, National Human Genome Research Institute, National Institutes of Health, Department of Health and Human Services. Funded by the National Institutes of Health (NIH).
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. This technology is an efficient, affordable, and powerful research tool. 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. Managing this incidental medical information presents many logistical and ethical challenges, especially as relates to novel variants. 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.

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 monzygotic 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]). 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. 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: \(\text{CACNA1S} \), associated with hypokalemic periodic paralysis and malignant hyperthermia\(^1\); \(\text{CPSI} \), associated with Carbamoyl Phosphate Synthetase I deficiency, as well as pulmonary artery hypertension\(^1\); \(\text{CYP21A2} \), associated with 21-hydroxylase deficiency leading to congenital adrenal hyperplasia\(^2\). 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 \(\text{CACNA1S} \) was pathogenic, and because the research participants and their relatives showed no signs of these \(\text{CACNA1S} \)-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 \(\text{CYP21A2} \) was nonpathogenic through (private, unpublished) work of researchers specializing in this disease (Maria New, MD; Tony Yuen, PhD). There was evidence that the \(\text{CPSI} \) 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\(^3\); 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\(^4\). 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

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

* Differences from the human reference sequence. build National Center for Biotechnology Information 36.

* Based on MPG score ≥10 and score/coverage ratio ≥0.5.
<table>
<thead>
<tr>
<th>Number</th>
<th>Gene</th>
<th>Variant Type</th>
<th>Disease(s) *</th>
<th>Type of Association</th>
<th>Notes/Reason Information Not Returned</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABCC8</td>
<td>M</td>
<td>Hyperinsulinemic hypoglycemia, permanent neonatal diabetes with neurologic features, transient and permanent neonatal diabetes mellitus</td>
<td>Disease-associated gene (dominant) and/or carrier for recessive disorder</td>
<td>No signs in participants or family, not clear if this specific variant is associated with this condition in this family</td>
<td>23–26</td>
</tr>
<tr>
<td>2</td>
<td>AK2</td>
<td>M</td>
<td>Reticular dysgenesis</td>
<td>Carrier for recessive disorder</td>
<td>Rare recessive disorder</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>CACNA1S</td>
<td>M</td>
<td>Malignant hyperthermia, hypokalemic periodic paralysis</td>
<td>Susceptibility for conditions and/or disease-associated gene (dominant)</td>
<td>The specific variant in this family was ultimately not felt to be pathogenic per discussion with multiple experts in the conditions studied</td>
<td>17,18</td>
</tr>
<tr>
<td>4</td>
<td>CD520</td>
<td>DIV</td>
<td>Methylmalonic aciduria (owing to transcobalamin receptor defect)</td>
<td>Carrier for recessive disorder</td>
<td>Rare recessive disorder</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>CISD2</td>
<td>M</td>
<td>Wolfram syndrome</td>
<td>Carrier for recessive disorder</td>
<td>Rare recessive disorder</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>COL5A2</td>
<td>M</td>
<td>Ehlers-Danlos syndrome type I</td>
<td>Disease-associated gene (dominant)</td>
<td>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</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>COL6A3</td>
<td>DIV</td>
<td>Bethlem myopathy, Ullrich congenital muscular dystrophy (autosomal recessive and dominant forms)</td>
<td>Disease-associated gene (dominant) and/or carrier for recessive disorder</td>
<td>No signs in participants or family, not clear if this specific variant is associated with this condition in this family</td>
<td>31–33</td>
</tr>
<tr>
<td>8</td>
<td>CPSI</td>
<td>M</td>
<td>Carbamoyl phosphate synthetase deficiency, pulmonary artery hypertension</td>
<td>Carrier for recessive disorder and possibly related to susceptibility to conditions</td>
<td>Rare recessive disorder</td>
<td>19–21</td>
</tr>
<tr>
<td>9</td>
<td>CRYGD</td>
<td>M</td>
<td>Progressive juvenile-onset punctate cataracts</td>
<td>Disease-associated gene (dominant)</td>
<td>No signs in participants or family, not clear if this specific variant is associated with this condition in this family</td>
<td>34</td>
</tr>
<tr>
<td>10</td>
<td>CYP21A2</td>
<td>DIV</td>
<td>Congenital adrenal hyperplasia</td>
<td>Carrier for recessive disorder</td>
<td>Normal variant in private databases of researchers studying this gene/condition</td>
<td>22</td>
</tr>
<tr>
<td>11</td>
<td>DBH</td>
<td>M</td>
<td>Dopamine β-hydroxylase deficiency, norepinephrine deficiency</td>
<td>Carrier for recessive disorders</td>
<td>Rare recessive disorder</td>
<td>35–37</td>
</tr>
<tr>
<td>12</td>
<td>DDX11</td>
<td>M</td>
<td>Warsaw breakage syndrome</td>
<td>Carrier for recessive disorder</td>
<td>Rare recessive disorder</td>
<td>38</td>
</tr>
<tr>
<td>13</td>
<td>DISC1</td>
<td>M</td>
<td>Possible association with schizophrenia</td>
<td>Possibly related to susceptibility for a condition</td>
<td>Susceptibility factor; not clear if this specific variant is associated with this condition in this family</td>
<td>39–42</td>
</tr>
<tr>
<td>14</td>
<td>DNAH5</td>
<td>M</td>
<td>Primary ciliary dyskinesia type 3</td>
<td>Carrier for recessive disorder</td>
<td>Primary ciliary dyskinesia is more common than 1/40,000, but only about 29% of cases are caused by mutations in DNAH5</td>
<td>43</td>
</tr>
<tr>
<td>15</td>
<td>FOXD4</td>
<td>M</td>
<td>Cardiomyopathy, obsessive-compulsive disorder; suicidality</td>
<td>Disease-associated gene (dominant)</td>
<td>Complex purported gene-associated phenotype reported in 1 family, susceptibility factor</td>
<td>44</td>
</tr>
<tr>
<td>16</td>
<td>FUT7</td>
<td>M</td>
<td>Possible autosomal recessive association with multiple autoimmune conditions</td>
<td>Carrier for recessive disorder</td>
<td>Rare recessive disorder</td>
<td>45</td>
</tr>
<tr>
<td>17</td>
<td>GIGYF2</td>
<td>DIV</td>
<td>Familial Parkinson's disease (incompletely penetrant)</td>
<td>Possibly related to susceptibility for a condition</td>
<td>No signs in family, not clear if this specific variant is associated with this condition in this family</td>
<td>46,47</td>
</tr>
<tr>
<td>18</td>
<td>HM0X1</td>
<td>M</td>
<td>Heme-oxygenase 1 deficiency</td>
<td>Carrier for recessive disorder and possibly related to susceptibility for a condition</td>
<td>Rare recessive disorder; possible association with COPD is a susceptibility factor</td>
<td>48–50</td>
</tr>
<tr>
<td>19</td>
<td>HSPG2</td>
<td>M</td>
<td>Schwartz-Jampel syndrome type 1, dyssegmental dysplasia, Silverman-Handmaker type</td>
<td>Carrier for recessive disorders</td>
<td>Rare recessive disorder</td>
<td>51</td>
</tr>
<tr>
<td>20</td>
<td>KRT18</td>
<td>M</td>
<td>Possible association with susceptibility to cirrhosis</td>
<td>Possibly related to susceptibility for a condition</td>
<td>Susceptibility factor</td>
<td>52–55</td>
</tr>
<tr>
<td>21</td>
<td>MST1</td>
<td>M</td>
<td>Possible association with inflammatory bowel disease</td>
<td>Possibly related to susceptibility for a condition</td>
<td>Susceptibility factor</td>
<td>56,57</td>
</tr>
</tbody>
</table>
### TABLE 2 Continued

<table>
<thead>
<tr>
<th>Number</th>
<th>Gene</th>
<th>Variant Type</th>
<th>Disease(s) a</th>
<th>Type of Association</th>
<th>Notes/Reason Information Not Returned</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>NLRP12</td>
<td>M</td>
<td>Familial cold autoinflammatory syndrome</td>
<td>Disease-associated gene (dominant)</td>
<td>No signs in participants or family, not clear if this specific variant is associated with this condition in this family</td>
<td>58</td>
</tr>
<tr>
<td>23</td>
<td>ORAI1</td>
<td>DIV</td>
<td>Immune dysfunction</td>
<td>Carrier for recessive disorder</td>
<td>Rare recessive disorder; although heterozygotes may have subtle subclinical manifestations</td>
<td>59</td>
</tr>
<tr>
<td>24</td>
<td>PROM1</td>
<td>M</td>
<td>Possible association with azoospermia</td>
<td>Possibly related to susceptibility for a condition</td>
<td>Susceptibility factor</td>
<td>60, 61</td>
</tr>
<tr>
<td>25</td>
<td>PRKCSH</td>
<td>DIV</td>
<td>Autosomal dominant polycystic liver disease</td>
<td>Disease-associated gene (dominant)</td>
<td>No signs in participants or family, not clear if this specific variant is associated with this condition in this family</td>
<td>62, 63, 64</td>
</tr>
<tr>
<td>26</td>
<td>PRKRA</td>
<td>M</td>
<td>Dystonia 16</td>
<td>Carrier for recessive disorder</td>
<td>Generally viewed as a rare recessive disorder; although affected heterozygote reported (possibly related to compound heterozygosity or other explanation)</td>
<td>65</td>
</tr>
<tr>
<td>27</td>
<td>PSPH b</td>
<td>M</td>
<td>Phosphoserine phosphatase deficiency</td>
<td>Carrier for recessive disorder</td>
<td>Rare recessive disorder</td>
<td>66</td>
</tr>
<tr>
<td>28</td>
<td>SCABR2</td>
<td>M</td>
<td>Action myelogenous renal failure</td>
<td>Carrier for recessive disorder</td>
<td>Rare recessive disorder</td>
<td>67</td>
</tr>
<tr>
<td>29</td>
<td>SLCTA3</td>
<td>DIV</td>
<td>Episodic ataxia, type 6</td>
<td>Disease-associated gene (dominant)</td>
<td>No signs in participants or family, not clear if this specific variant is associated with this condition in this family</td>
<td>68</td>
</tr>
<tr>
<td>30</td>
<td>TH</td>
<td>M</td>
<td>Segawa syndrome</td>
<td>Carrier for recessive disorder</td>
<td>Rare recessive disorder</td>
<td>69</td>
</tr>
<tr>
<td>31</td>
<td>TRPC7</td>
<td>M</td>
<td>Possible association with bipolar affective disorder, ALS, and Parkinsonism</td>
<td>Possibly related to susceptibility for conditions</td>
<td>Susceptibility factors, no family history of these conditions and not clear if this specific variant is associated with this condition in this family</td>
<td>70</td>
</tr>
<tr>
<td>32</td>
<td>USP26</td>
<td>M</td>
<td>Possible association with impaired spermatogenesis</td>
<td>Possibly related to susceptibility for a condition</td>
<td>Susceptibility factors, not clear if this specific variant is associated with this condition in this family</td>
<td>71–74</td>
</tr>
</tbody>
</table>

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.

---

### ACKNOWLEDGMENT

The authors are extremely grateful to Dr Leslie G. Biesecker (Chief and Senior Investigator, Genetic Disease Research Branch, National Human Genome Research Institute) for access to large-scale comparison samples.

---

**CASE REPORT**

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, experienced in interpretation of genomics, to construct the proposal to help maintain the integrity of the research study. The parameters of the research study and the characteristics of the research participants 12, 13.

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.
REFERENCES


Incidental Medical Information in Whole-Exome Sequencing

Benjamin D. Solomon, Donald W. Hadley, Daniel E. Pineda-Alvarez, NISC Comparative Sequencing Program, Aparna Kamat, Jamie K. Teer, Praveen F. Cherukuri, Nancy F. Hansen, Pedro Cruz, Alice C. Young, Benjamin E. Berkman, Settara C. Chandrasekharappa and James C. Mullikin

Pediatrics 2012;129:e1605; originally published online May 14, 2012;
DOI: 10.1542/peds.2011-0080

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/129/6/e1605.full.html