

Incidental copy-number variants identified by routine genome testing in a clinical population

Philip M. Boone, BS¹, Zachry T. Soens, BS¹, Ian M. Campbell, BS¹, Pawel Stankiewicz, MD, PhD^{1,2}, Sau Wai Cheung, PhD^{1,2}, Ankita Patel, PhD^{1,2}, Arthur L. Beaudet, MD^{1,2-5}, Sharon E. Plon, MD, PhD^{1,4-6}, Chad A. Shaw, PhD^{1,2}, Amy L. McGuire, JD, PhD⁷ and James R. Lupski, MD, PhD^{1,2,4-6}

Purpose: Mutational load of susceptibility variants has not been studied on a genomic scale in a clinical population, nor has the potential to identify these mutations as incidental findings during clinical testing been systematically ascertained.

Methods: Array comparative genomic hybridization, a method for genome-wide detection of DNA copy-number variants, was performed clinically on DNA from 9,005 individuals. Copy-number variants encompassing or disrupting single genes were identified and analyzed for their potential to confer predisposition to dominant, adult-onset disease. Multigene copy-number variants affecting dominant, adult-onset cancer syndrome genes were also assessed.

Results: In our cohort, 83 single-gene copy-number variants affected 40 unique genes associated with dominant, adult-onset disorders and unrelated to the patients' referring diagnoses (i.e., incidental) were found. Fourteen of these copy-number variants

are likely disease-predisposing, 25 are likely benign, and 44 are of unknown clinical consequence. When incidental copy-number variants spanning up to 20 genes were considered, 27 copy-number variants affected 17 unique genes associated with dominant, adult-onset cancer predisposition.

Conclusion: Copy-number variants potentially conferring susceptibility to adult-onset disease can be identified as incidental findings during routine genome-wide testing. Some of these mutations may be medically actionable, enabling disease surveillance or prevention; however, most incidentally observed single-gene copy-number variants are currently of unclear significance to the patient.

Genet Med 2013;15(1):45–54

Key Words: copy-number variation; genetic load; genomic mutational load; incidentalome; structural variation; variants of unknown significance

INTRODUCTION

Human genomes exhibit substantial variation; the average diploid human genome differs from the reference genome by ~3 million to 3.5 million single-nucleotide variants and about a thousand copy-number variants (CNVs; e.g., DNA deletions and duplications) >500 base pairs in size.¹ Genomic variation can result in positive traits but can also make an individual susceptible to disease. Even apparently healthy individuals possess genetic load, i.e., suboptimal alleles that diminish fitness, and may also harbor carrier mutations and mutations predisposing to illness later in life.

Disease predisposition mutations constitute genetic vulnerabilities that could impact a patient's life and medical care. Yet, most healthy individuals do not undergo genetic testing unless they are tested for mutations found in relatives manifesting a genetic disease, are tested with a personal genomic panel, or are part of a research study. Previous literature has addressed the medical, ethical, and psychological considerations relating to such testing²⁻⁵ and concerns regarding the identification of incidental variants identified by genomic assays (the "incidentalome").⁶⁻¹⁰

The availability of genome-wide data for a number of control subjects has provided a genome-scale glimpse of potential disease-causing and susceptibility alleles in healthy individuals.^{1,11,12} However, the variant alleles analyzed have generally consisted of point mutations and other single-nucleotide variants, leaving the contribution of structural genomic variation, including CNVs, to be fully determined. Two preliminary studies examined clinical cohorts analyzed by array comparative genomic hybridization (aCGH; also known as chromosomal microarray analysis)^{13,14} for CNVs spanning cancer syndrome genes to identify CNVs potentially predisposing to cancer later in life.^{15,16} A subset of these CNVs were interpreted as likely contributing to the patients' presenting symptoms, whereas others were not (i.e., incidental). Only a subset of potential disease predisposition genes were examined in these studies: the few dozen known cancer syndrome genes. Therefore, despite discussion about the possibility of discovering incidental findings during genomic testing, the frequency and characteristics of such incidental variants at a genome- or "phenome"-wide scale in a clinical population remains unknown.

¹Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; ²Medical Genetics Laboratories, Baylor College of Medicine, Houston, Texas, USA; ³Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas, USA; ⁴Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA; ⁵Texas Children's Hospital, Houston, Texas, USA; ⁶Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA; ⁷Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, Texas, USA. Correspondence: James R. Lupski (jlupski@bcm.edu)

Submitted 22 March 2012; accepted 26 June 2012; advance online publication 9 August 2012. doi:10.1038/gim.2012.95

We sought to determine to what extent CNV mutations potentially conferring predisposition to adult-onset disease are detected incidentally during routine clinical aCGH. We used exon-targeted aCGH, a technique enabling CNV mutations as small as one exon of a target gene to be identified,¹⁷ to examine DNA from 9,005 individuals in a clinical cohort. We show that, in this large clinical population, CNVs are identified that encompass and disrupt disease genes for late-onset disorders unrelated to the current diagnoses of the patients. Some of these incidentally identified CNVs may predict disease susceptibility and constitute medically actionable alleles.

MATERIALS AND METHODS

Subjects

Subjects were patients and fetuses referred to our diagnostic laboratory between June 2009 and July 2011 because of clinical suspicion of a genetic or genomic disorder, and their parents ($n = 9,005$ subjects). For most patients, a short clinical description was available. Clinical information was not available for parents. Informed consent, approved by the Institutional Review Board for Human Subjects Research of Baylor College of Medicine, was obtained from subject 82, for whom detailed molecular and phenotypic data are presented. All other subject data were anonymized for our analyses.

aCGH

aCGH was performed with gender-matched controls using postnatal DNA from whole blood, or fetal DNA from cells obtained by amniocentesis or chorionic villus sampling as described,¹⁸ with minor modifications. The CGH microarray (V8 OLIGO),¹⁷ procedures,¹⁹ and computational analysis^{13,19,20} have been described. Briefly, V8 OLIGO contains both genome-wide backbone probe coverage and enhanced probe resolution within the exons and introns of ~1,700 manually curated known and putative disease genes. Minor additions to the list of exon-targeted genes were made during the study; details are available at <https://www.bcm.edu/geneticlabs/>. Array CGH data for all 9,005 individuals were independently reviewed by clinical cytogeneticists.

Assigning inheritance patterns to genes

Phenotypes in the Online Mendelian Inheritance in Man database (<http://www.omim.org/>) with at least one known, associated disease gene were identified ($n = 3,430$ as of July 2011). The inheritance pattern of each was determined computationally or by review, then each dominant and recessive phenotype was matched to its causal gene(s), producing a list of 667 “dominant” and 1,080 “recessive” disease genes (Supplementary Table S1 online).

CNV selection

Phase I (single-gene CNVs). CNVs for which the minimum span of altered copy number encompassed or disrupted a single gene (“single-gene CNVs”) were selected, as (i) the potential phenotypic consequences of single-gene CNVs may be more

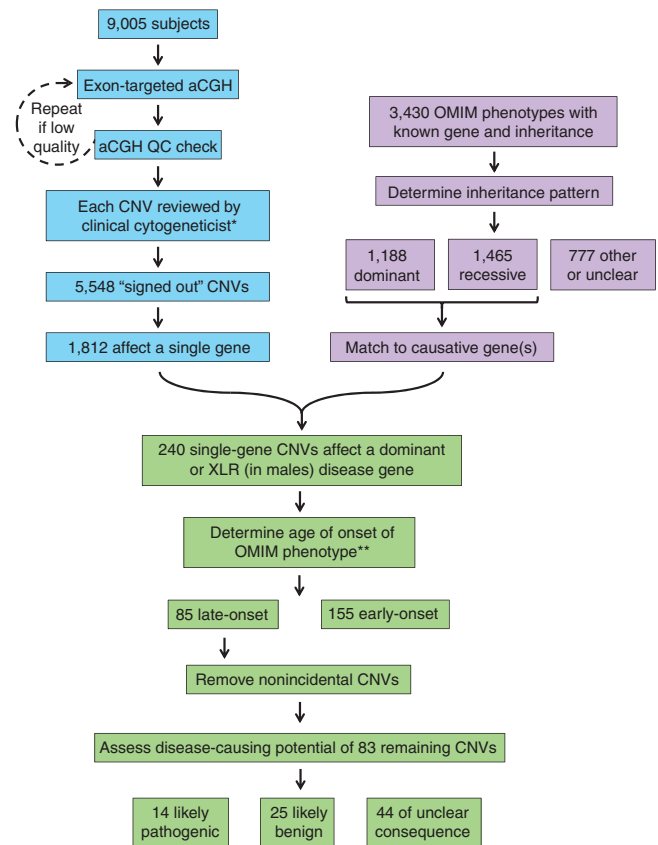


Figure 1 Strategy for identifying single-gene copy-number variants (CNVs) potentially conferring predisposition to dominant, late-onset disease. (Blue) Single-gene CNVs passing quality measures were detected in a clinical population. (Purple) The genes associated with dominant and recessive Online Mendelian Inheritance in Man database (OMIM) phenotypes were identified computationally. (Green) CNVs affecting dominant and X-linked recessive (in males) late-onset disease genes and unrelated to the patients’ current clinical symptoms (i.e., incidental) were identified and analyzed for disease-causing potential. *Review included visual inspection of each CNV log₂ plot and exclusion of common polymorphic CNVs. **Late-onset was defined as a $\geq 5\%$ probability of disease presentation in adulthood. QC, quality control; XLR, X-linked recessive.

predictable than those of larger CNVs, enabling them to more readily be assigned as incidental (i.e., noncontributory to the subjects’ current symptoms) and (ii) smaller CNVs are expected to be more representative of the mutational load found in the general, “healthy” population, given the allele frequency spectrum of CNV size.^{1,21,22} Next, incidental single-gene CNVs affecting genes for adult-onset disorders were identified and analyzed for potential pathogenicity (Figure 1). As the majority of CNVs detected by aCGH were expected to be heterozygous or hemizygous, we examined only genes associated with dominant and X-linked recessive conditions, the latter only in males, because hemizygous mutations are sufficient to cause these diseases. For ease of discussion, these conditions are referred to as “dominant” in some sections of this report. Late-onset disorders were defined as those that usually or exclusively present in adulthood, as well as typically childhood- or adolescent-onset conditions that may present in adulthood in $\geq 5\%$ of

cases (e.g., OMIM #118220, Charcot-Marie-Tooth disease, type 1A). This definition was constructed to identify conditions that could remain asymptomatic in pediatric patients (the majority of patients referred to our lab for clinical aCGH are children; data not shown) and thus potentially be predicted by “presymptomatic” incidental genetic findings. OMIM phenotypes that are not disease states or that do not display clear Mendelian inheritance were excluded.

Phase II (multi-gene CNVs affecting cancer-predisposition genes). This analysis replicated the first (single-gene) CNV discovery phase, but parsed for CNVs spanning more than one gene and affecting at least one gene associated with dominant cancer predisposition.

CNV validation

A subset of CNVs was re-assessed by one or more independent molecular methods, including fluorescence *in situ* hybridization, long-range PCR, multiplex ligation-dependent probe amplification, and di-deoxynucleotide sequencing, as described¹⁷ (**Supplementary Materials and Methods** online).

All genomic coordinates are based on the March 2006 assembly of the reference human genome (NCBI36/hg18) unless otherwise specified.

RESULTS

Incidental single-gene CNVs affecting dominant, adult-onset disease genes

Of 5,548 CNVs identified by aCGH among our cohort and passing quality measures, 1,812 deleted or duplicated part or all of a single gene (**Figure 1**). Eighty-five of these CNVs, present in 84 individuals, affected 41 unique genes associated with adult-onset disorders displaying either dominant or X-linked recessive (if identified in a male) inheritance (**Table 1**; **Supplementary Table S2** online). Affected genes were present on 17 autosomes and the X chromosome. All autosomal CNVs were heterozygous (**Supplementary Table S2** online). Sixty-nine of the 85 CNVs listed in **Table 1** and **Supplementary Table S2** online were present in patients, including one fetus (subject 14). Both parents, one parent, or a grandparent were assessed by aCGH or fluorescence *in situ* hybridization for 25 of these patients. In 21 cases, the CNV was inherited, in three it was found by aCGH not to be maternally inherited, and one CNV was found by aCGH and PCR to be *de novo* (**Supplementary Table S2** online).

A variant uncovered by a genome-wide assay is only incidental if it does not explain the patient’s current clinical symptoms. Referring diagnoses were available for all but four of the nonparental and nonfetal subjects (**Supplementary Table S2** online). In two cases, this clinical description was highly suggestive of early-onset disease corresponding to the affected gene (e.g., subject 32, a 3-year-old boy diagnosed clinically with rhabdomyolysis who has an in-frame deletion in *DMD*). These patients’ CNVs are thus not incidental and are not considered further. In nine cases, the presenting diagnosis raised the possibility of early-onset disease corresponding to the affected

gene (e.g., subject 60, a 4-year-old girl diagnosed clinically with autism spectrum disorder and attention deficit hyperactivity disorder who has a deletion in *MYO6*, a deafness gene), although none of these phenotypes were pathognomonic for any disorder. Therefore, 83 of 85 of the single-gene CNVs we describe likely represent incidental findings. Clinical information was not available for parents.

To determine the anticipated consequence of each incidental CNV (disease-causing/predisposing, benign, or unknown), we considered the ploidy (deletion vs. duplication) and minimum and maximum predicted boundaries (or exact coordinates, if determined by DNA sequencing) in light of mutations reported in the Human Gene Mutation Database (HGMD; <http://www.hgmd.org/>), OMIM, and/or the primary literature. This algorithm, similar to the variant “binning” strategy proposed by Berg et al.,¹⁰ involved manual annotation and judgment by both the authors and the clinical cytogeneticists who originally reviewed the array data. This annotation is detailed for each CNV in **Supplementary Table S2** online. **Figure 2** demonstrates the extent of each CNV relative to the exon–intron structure of the gene of interest. Among the possible rearrangement types are CNVs affecting whole genes, overlapping one end of a gene, affecting several exons, or disrupting or encompassing a single exon. Fourteen of 83 incidental mutations are predicted to be likely disease-predisposing or causing, 25 are likely benign, and 44 are of unknown consequence (**Table 1**; **Supplementary Table S2** online). No CNV was present in the Database of Genomic Variants, a catalog of structural variation in control populations, in the UCSC “structural variation” track, which aggregates data from nine population-based CNV studies (<http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=237402995&c=chrX&g=cnp>), or in the 1000 Genomes Project CNV data.²¹

Figure 3 profiles four CNVs (subjects 17, 41, 78, and 82) and confirmatory experiments for one (subject 82). Subject 82 (**Figure 3a–f**) is a 16-year-old boy referred for aCGH because of autism and epilepsy. A CNV deleting exon 13 of *SPAST*, the disease gene for autosomal dominant hereditary spastic paraplegia type 4 (OMIM #182601), was identified, although he did not show signs of spastic paraplegia upon physical examination. This mutation, predicted to be likely disease-causing (**Figure 3f**),²³ thus constitutes an incidental finding. Other CNVs depicted include (i) a deletion in *CHEK2* (subject 17; **Figure 3g**), a gene associated with a moderately increased risk of breast cancer and risk of other cancers²⁴ in a subject with severe global developmental delay; (ii) a deletion in *SDHB* (subject 78; **Figure 3h**), the disease gene for paragangliomas, type 4 (OMIM #115310), a tumor syndrome characterized by carotid body tumors and extra-adrenal pheochromocytomas, in a subject with an unknown clinical phenotype. This CNV was also present in his mother (subject 79); and (iii) a deletion in *GRHL2* (subject 41; **Figure 3i**), the disease gene for autosomal dominant deafness type 28 (OMIM #608641), characterized by progressive sensorineural hearing loss, in a patient with dysmorphic features and failure to thrive. Breakpoints of

Table 1 Single-gene CNVs detected by aCGH affecting genes associated with adult-onset conditions displaying dominant or X-linked recessive (in males) inheritance

Number and ploidy of events	Subject no(s).	Gene affected by CNV(s)	Adult-onset OMIM phenotype(s) associated with affected gene (OMIM ID)	Is/are CNV(s) predicted to be disease-predisposing?	Number of different families affected	CNV diversity, if ≥2 affected families ^c
1 del	1	<i>ABCA4</i>	153800	No	1	
13 dels	2–14	<i>ABCC6</i>	177850	No	11	11 Identical dels
1 del	15	<i>APP</i>	104760; 605714	No	1	
1 del	16	<i>BDNF</i>	164230	No	1	
2 dels	17–18	<i>CHEK2</i>	609265; 259500; 114480	?	2	2 Identical dels
2 dups	19–20	<i>CHM</i>	303100	No	2	2 Different dups
1 del	21	<i>COL4A3</i>	104200	No	1	
11 dels	22–32	<i>DMD</i>	300376; 302045	6 Yes, 4?, 1 N/A ^a	11	11 Different dels
2 dels	33–34	<i>DSP</i>	607450	Yes	1	
1 del	35	<i>FBN1</i>	129600	No	1	
2 dups	36–37	<i>FECH</i>	177000	No	1	
1 del	38	<i>FLNC</i>	609524	?	1	
2 dups	39–40	<i>FTL</i>	606159	No	1	
1 del	41	<i>GRHL2</i>	608641	?	1	
1 del	42 ^b	<i>HPRT1</i>	300323	N/A ^a	1	
1 dup	43	<i>INSR</i>	609968	?	1	
4 dels	44–47	<i>ITPR1</i>	606658	?	2	2 Different dels
2 dels	48 ^b –49	<i>KIF1B</i>	171300	?	1	
1 del	50	<i>LIPC</i>	151670	?	1	
1 del	51	<i>LMNB1</i>	169500	No	1	
1 dup	52	<i>LRRK2</i>	607060	?	1	
1 del	53	<i>MFN2</i>	609260	Yes	1	
1 del	54	<i>MYBPC3</i>	115200	?	1	
2 dels	55–56	<i>MYH14</i>	600652	?	1	
4 dups	42 ^b , 57–59	<i>MYH9</i>	115100; 603622; 605249	?	3	3 Identical dups
1 del	60	<i>MYO6</i>	606346	?	1	
1 del	61	<i>MYOC</i>	137750	Yes	1	
2 dups	62–63	<i>PMP22</i>	118220; 162500	?	1	
3 dels	64–66	<i>PMS2</i>	614337	2 Yes, 1?	3	3 Different dels
2 dels	67–68	<i>POLG2</i>	610131	?	1	
1 del	69	<i>PON1</i>	168820	?	1	
2 dup	70–71	<i>PRKCG</i>	605361	?	1	
3 del	72–74	<i>PRODH</i>	600850	?	2	2 Identical dels
1 dup	75	<i>RP1</i>	145750; 180100	?	1	
2 dup	76–77	<i>RP2</i>	312600	?	1	
2 del	78–79	<i>SDHB</i>	115310	?	1	
1 dup	80	<i>SGCE</i>	159900	?	1	
1 dup	81	<i>SNCA</i>	127750; 168601; 605543	?	1	
1 del	82	<i>SPAST</i>	182601	Yes	1	
1 dup	83	<i>T</i>	215400	Yes	1	
1 dup	84	<i>WNK1</i>	145260	?	1	

aCGH, array comparative genomic hybridization; CNV, copy-number variant; del, deletion (loss); dup, duplication (gain); N/A, not applicable; OMIM, Online Mendelian Inheritance in Man database; ?, unknown.

^aThe current phenotype of the subject is highly suggestive of an early-onset disease associated with the disrupted disease gene (**Supplementary Table S2** online).

^bReported previously.¹⁷ ^cAs assessed by aCGH; CNV breakpoints were not determined with nucleotide resolution in the majority of cases.

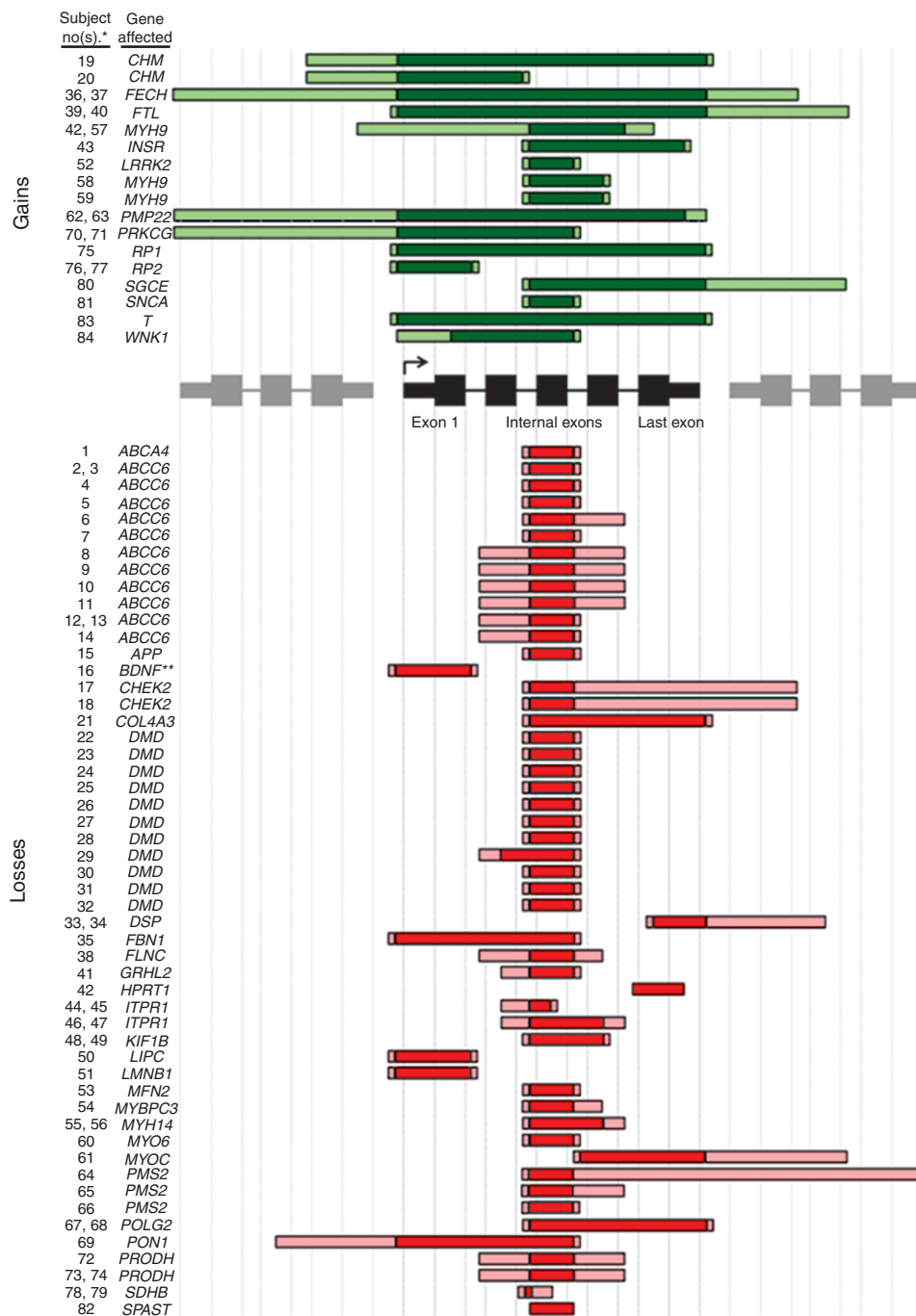


Figure 2 Copy-number variants (CNVs) rearrange adult-onset disease genes in various ways. Each CNV (colored bars) listed in [Table 1](#) and [Supplementary Table S2](#) online is displayed relative to a schematized structure of the adult-onset disease gene it rearranges (black) and the most proximal upstream (left, gray) and downstream (right, gray) genes. The central part of each colored bar (dark green or red) is the minimum region of altered copy number; the flanking parts of each bar (light green or pink) are the maximum affected interval. If the exact size of the CNV was determined, no flanking bars are shown. For some CNVs, disparity between the minimum and maximum affected intervals renders the exact exons affected unclear. *Parent-child pairs are combined into a single line. **Several (but not all) possible alternative first exons are affected. All but one are noncoding.

the *CHEK2*, *SDHB*, and *GRHL2* deletions were not determined with nucleotide resolution; thus, the exact exonic extent and therefore the functional and potential phenotypic consequence of each is unknown ([Supplementary Table S2](#) online). Subject 17 may have a previously described and somewhat common cancer risk allele deleting two internal exons of *CHEK2*,^{25,26}

potentially providing a clue as to the pathogenicity of this mutation.

Eight genes were interrupted by CNVs in multiple unrelated subjects ([Table 1](#); [Supplementary Table S2](#) online). Losses in *ABCC6*, *CHEK2*, and *PRODH*, and a gain in *MYH9* appeared to be recurrent (the same mutation in multiple unrelated

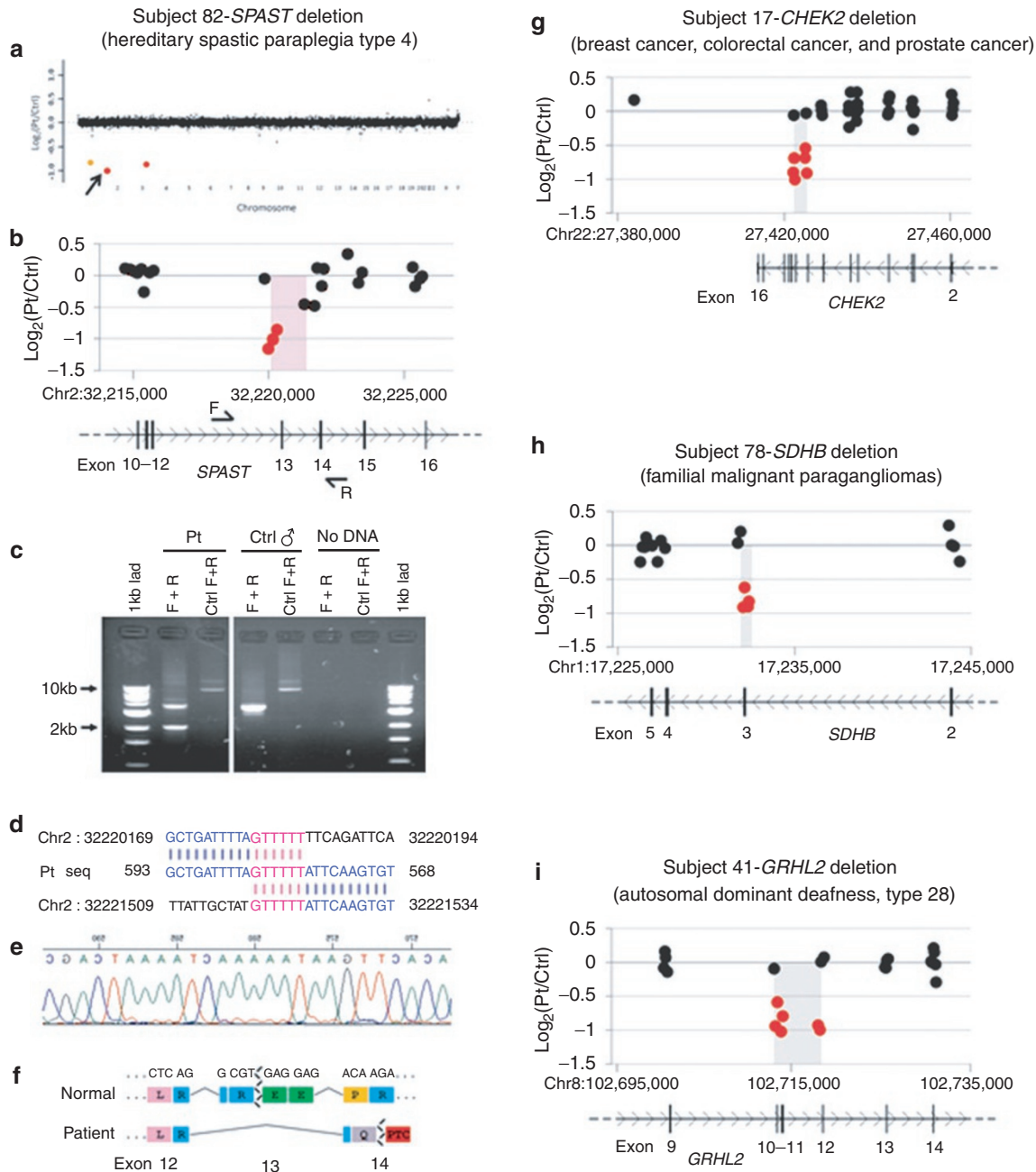


Figure 3 Examples of incidental single-gene copy-number variants (CNVs) affecting dominant, adult-onset disease genes. (a–f) Complete genomic DNA analysis for a single subject (subject 82; BAB 3166), a child referred for array comparative genomic hybridization (aCGH) because of autism and epilepsy. (a) A genome-wide aCGH log₂ plot displays (i) a known, benign CNV (orange dot), (ii) a deletion CNV potentially contributory to the patient's phenotype affecting *SLC9A9* (red dot, no arrow), and (iii) a deletion disrupting *SPAST*, the disease gene for type 4 autosomal dominant hereditary spastic paraplegia (SPG4) (red dot, arrow). (b) Local view of the *SPAST* CNV (pink shading), which deletes exon 13. F and R primers were used to confirm the presence of the CNV. (c) PCR confirms a heterozygous deletion in the patient. An ~2 kb band amplified from patient DNA was sequenced to determine the exact size, location, and breakpoint characteristics of the deletion (Supplementary Materials and Methods online). Pt, patient DNA; Ctrl ♂, control male DNA; Ctrl F+R, control primers. (d) DNA sequencing confirms a 1,340 base pair (bp) deletion. Genomic sequence telomeric (top) and centromeric (bottom) to the deletion breakpoint, which displays six bp of microhomology (magenta), are displayed, as is the mutant patient sequence ("Pt seq"; middle). (e) DNA sequencing trace of the breakpoint region. A flipped reverse complement sequence is displayed, as sequencing was performed in reverse genomic orientation. (f) The predicted effect of this incidentally discovered removal of a 43 bp exon is a shift of translational frame, introducing a premature termination codon (PTC) in exon 14, likely leading to nonsense-mediated decay and thus acting as a null allele. Deletions of exon 13 have been described in individuals with SPG4.²³ (g–i) aCGH log₂ plots displaying heterozygous CNVs disrupting genes associated with dominant, adult-onset conditions in three additional patients. Gray shading indicates the minimum span of altered copy number. (g) Subject 17. (h) Subject 78. (i) Subject 41.

individuals/families), although CNV boundaries were not determined with nucleotide resolution. Losses in *DMD*, *ITPR1*, *PMS2*, and gains in *CHM* were nonrecurrent (different mutations in each unrelated individual/family).

CNV validation

Of 23 CNVs assessed by PCR, fluorescence *in situ* hybridization, or multiplex ligation-dependent probe amplification, 20 were confirmed, and one (in subject 38) was not (**Supplementary Materials and Methods** and **Supplementary Table S2** online); PCRs for subjects 33 and 78 were equivocal. In 12 additional instances, CNVs were found in two family members by aCGH, adding to the likelihood that these CNVs represent true losses or gains.

Incidental CNVs spanning more than one gene

We hypothesized that CNVs encompassing or disrupting more than one gene may serve as an important source of incidentally identified mutations. As a proof-of-principal experiment, we investigated this possibility for the same limited group of phenotypes studied by Adams et al.¹⁵ and Pichert et al.¹⁶; the inherited cancer syndromes. We searched among our cohort for CNVs spanning up to 20 genes and affecting at least 1 dominant, adult-onset cancer-predisposition syndrome gene (43 genes; **Supplementary Table S3** online).²⁴ Twenty-seven CNVs encompassed or disrupted 17 unique cancer genes (**Figure 4**, **Supplementary Table S4** online). In no case were the patient's current clinical symptoms highly suggestive of a phenotype related to the affected gene (**Supplementary Table S4** online). The CNV of only one patient encompassed a gene associated with a noncancer, dominant, single-gene disorder (*MPZ*, associated with several inherited peripheral neuropathies: OMIM #118200 and others), which did not explain the patient's referring diagnosis ("seizure disorder"). Therefore, all 27 cases of rearrangements including a cancer susceptibility gene likely constitute incidental findings.

DISCUSSION

Genome-wide diagnostic tests have entered the clinic.²⁷ These assays examine multiple parts of the genome; thus, in addition to identifying mutations causative of a patient's presenting symptoms, they also have the potential to incidentally uncover other mutations, including medically actionable variants not related to the patient's current clinical findings. The contribution of CNVs to the incidentalome is under-explored. Two studies have provided a preliminary analysis of variants such as these in clinical cohorts: Adams et al.¹⁵ and Pichert et al.¹⁶ identified CNVs encompassing cancer-predisposition genes in cohorts of patients referred for aCGH for various indications. Some of these CNVs were incidental, whereas others were not. Nonetheless, these studies demonstrate the potential clinical utility of knowing that a cancer gene is in a CNV, even if the CNV is not itself unexpected.

We reported previously that exon-focused aCGH enables the detection of clinically relevant CNVs affecting single genes or

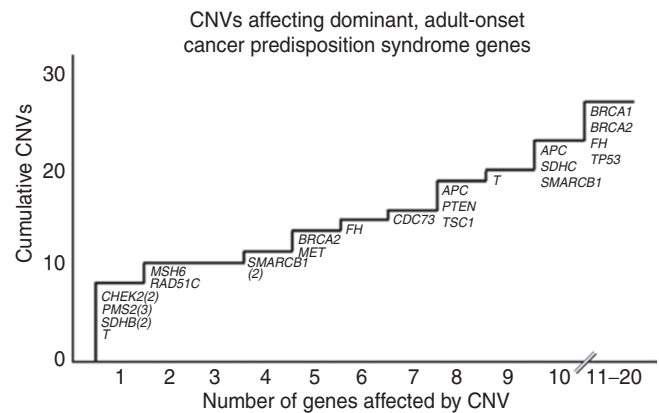


Figure 4 An analysis of multigene copy-number variants (CNVs) reveals additional incidental variants. CNVs affecting disease genes for dominant Mendelian cancer-predisposition syndromes and spanning up to 20 genes per CNV were identified. As CNVs extending beyond a single gene are examined, additional incidental variants are revealed. Genes listed under each "step" of the graph are those rearranged by CNVs affecting the number of genes indicated on the x-axis. The number of subjects with such a CNV, if greater than one, is shown in parentheses. Nineteen additional CNVs, affecting 13 unique genes, identified as CNVs spanning from 2 to 20 genes are considered.

even single exons.¹⁷ We now show that, in addition to detecting CNVs associated with patients' current medical conditions, exon-targeted aCGH uncovers, incidentally, CNVs affecting late-onset disease genes, potentially predicting future disease susceptibility and constituting medically actionable variants.

Incidental single-gene CNVs affecting dominant, adult-onset disease genes

Eighty-four of 9,005 subjects in our cohort had a CNV affecting a single gene associated with dominant, adult-onset disease. Most were novel, being absent from databases of benign (e.g., the Database of Genomic Variants, 1000 Genomes Project) and pathogenic (e.g., the Human Gene Mutation Database) CNVs. A few were recurrent, potentially indicative of founder mutations (e.g., deletion of *CHEK2* exons 10 and 11 in subjects 17 and 18, a mutation conferring increased risk of cancer and identified in ~1 in 250 Polish individuals).^{25,26}

We restricted our initial analysis to single-gene CNVs to enhance the interpretability of variants; nonetheless, most variants remain of unknown consequence. Uncertainty may arise for many reasons, including (i) the mutation is novel; (ii) follow-up DNA sequencing or another method has not been performed to determine the exact boundaries, location, orientation, and potential complexity of the CNV; (iii) penetrance of the mutation, i.e., the way the variant interacts with the rest of the individual's genome, environmental factors, and stochastic processes to produce a clinical phenotype, is unknown; (iv) predicted loss-of-function mutations near the 3' end of a gene may instead result in escape from nonsense-mediated decay and gain of function; and (v) the possibility of mosaicism exists. Further complicating variant interpretation is the fact that many genes are associated with two or more disease phenotypes sometimes

displaying different inheritance patterns. For example, deletions in the *ABCC6* gene (subjects 2–14), which stand out among mutations we report for their recurrence in multiple unrelated individuals, probably do not represent causative mutations for the dominant condition pseudoxanthoma elasticum, *forme fruste* (OMIM #177850), but rather most likely represent a previously-described, common carrier mutation for recessive pseudoxanthoma elasticum (OMIM #264800).^{28,29}

Incidental CNVs spanning more than one gene

In addition to single-gene variants, we investigated CNVs encompassing or disrupting more than one gene and demonstrate that these mutations are a source of incidentally identified mutations affecting Mendelian cancer predisposition genes. How many of these CNVs are expected to be pathogenic was not assessed. Rather, our aim was to demonstrate that as CNVs of larger size are examined, an increasing number of variants are identified that may constitute incidental findings. The frequency of CNVs affecting cancer genes in our cohort (27 CNVs among 9,005 subjects (0.30%) affecting 17 of 43 dominant cancer syndrome genes) was similar to that of Adams et al.¹⁵ (34 CNVs among 18,437 subjects (0.18%) affecting 10 of 22 genes for childhood onset syndromes including a cancer phenotype) and Pichert et al.¹⁶ (29 CNVs among 4,805 subjects (0.60%) with a noncancer clinical indication affecting 14 of 47 dominant cancer syndrome genes); however, this comparison is potentially misleading due to the differing array platforms, number and genomic position of the interrogating probes, CNV size cutoffs, and gene lists queried in each investigation. CNVs in cancer-predisposition genes are potentially medically actionable (for example, in cases of deleterious mutations in *SDHB*, surveillance for paragangliomas is recommended); prevention or tailored therapy may be possible in the future.

Even by extending our analysis to a subset of multigene CNVs, our study is by no means a complete survey of incidental variants. The following limitations of our study suggest avenues for future research: (i) aCGH screens only for CNVs, and the array we used had exon-by-exon probe coverage for ~1,700 genes; therefore, >90% of genes in the genome were not interrogated for exonic CNV; (ii) we defined “adult-onset” conditions as those likely to present at age ≥ 18 years, as most subjects were below this age. This strategy likely led to mutations associated with conditions that present later in childhood, but were not yet present at the time of testing, to be excluded (e.g., an 11-month-old patient with a predicted deleterious mutation in *EXT2*, the disease gene for multiple exostoses type 2 (OMIM #133701), but no reported exostoses)¹⁷; (iii) parents of patients were tested in fewer than half of cases; (iv) other classes of mutation not systematically searched for in our study include parental mosaic mutations in dominant, early-onset disease genes; homozygous mutations in recessive, adult-onset disease genes; and mutations affecting X-linked recessive disease genes in females. An example of the last class of mutation involves subject 80, a 16-year-old male referred for attention deficit hyperactivity disorder who possesses a duplication of exons 2–12 of the 12-exon

gene *SGCE*, the disease gene for myoclonic dystonia, type 11 (OMIM #159900) (**Supplementary Figure S1a,b** online). This mutation is of unknown disease-causing potential, and was not found in the subject’s mother (**Supplementary Figure S1c** online). Instead, the mother was found to have an in-frame deletion of exons 49–51 (of 79) of *DMD* (**Supplementary Figure S1c,d** online), a mutation reported previously in individuals with dilated cardiomyopathy and Duchenne muscular dystrophy.^{30,31} Females can occasionally present with symptoms of X-linked recessive conditions. Therefore, this mutation type could be of personal, clinical importance as well as being of relevance to potential current and/or future pregnancies; and (v) we did not formally consider the potential phenotypic consequences of multiple CNVs being present in a single individual (e.g., the *SPAST* and *SLC9A9* CNVs present in subject 82; **Figure 3a**). The possibility that two or more CNVs might interact additively, synergistically, or might modify each other to affect the overall phenotype of an individual is intriguing, and may offer etiologic insights into complex phenotypes not explained by a single genetic lesion and may explain how CNVs found separately in healthy individuals may cause disease when present together.^{32–35}

Practical and ethical considerations

The potential of genomic diagnosis to identify the basis for illness in a patient and to possibly guide patient-tailored therapy is immense. Yet, with sufficient resolution, clinical genome analysis will identify, incidentally, additional variants in each patient tested. Which of these variants should be reported to the patient, the physician, or even the genomic diagnostician is the subject of ongoing debate.^{36,37} Categories of genes can be constructed based on their clinical utility, as can categories of variants based on their known or predicted pathogenicity. Berg et al.¹⁰ have proposed a formal “binning” strategy that combines these two steps to assess the “reportability” of a given variant. As demonstrated by our data, the significance of many variants is currently unclear, limiting the precision with which decisions about reportability may be made.

Clinical genome analysis has the potential to identify incidental CNVs and other types of variants that could influence medical management in presymptomatic individuals. Uncovering incidental mutations conferring susceptibility to untreatable conditions may also be useful, as an individual with such a finding would have an *a priori* knowledge of his/her mutation, potentially minimizing the time, cost, and confusion that often accompanies the accurate diagnosis of a rare illness.³⁸ Furthermore, in all cases the possibility exists of using reproductive strategies to prevent the mutation from being transmitted.

There are currently no standards for the communication of incidental genetic findings to patients in a clinical setting; this decision is at the discretion of the diagnostic laboratory. It has been proposed that reporting of incidental variants by laboratories may benefit from a standardized approach. Necessary for such an approach is the assignment of known genes and

variants into reportability categories based on clinical validity and actionability.³⁹ Our findings suggest that these efforts might benefit from explicit consideration of the reportability of incidental CNVs.

Clinical genomic tests are ordered by a physician often without written informed consent from the patient, and incidental findings may or may not be discussed during pretest counseling. Furthermore, patient choice in return of results is likely the exception rather than the norm. The potential to reveal incidental findings may influence patient decision-making about testing and therefore it may be beneficial to routinely discuss this possibility during pretest counseling.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

ACKNOWLEDGMENTS

P.M.B. is a fellow of the Baylor College of Medicine Medical Scientist Training Program (T32GM007330) and is supported by the National Eye Institute (T32EY007102), the Wintermann Foundation, and the Baylor Research Advocates for Student Scientists. This work was supported in part by the National Cancer Institute (R01CA138836 to S.E.P.), the National Institute of Neurological Disorders and Stroke (R01NS058529 to J.R.L.), and the National Human Genome Research Institute (U54HG006542). The authors thank Bryce Daines and Jason Salvo for technical advice, Patricia Hixson and Audrey Ester for technical support, and Frank Probst for clinical insights.

DISCLOSURE

J.R.L. is a paid consultant for Athena Diagnostics, holds stock ownership in 23andMe and Ion Torrent Systems, and is a co-inventor on US and European patents related to molecular diagnostics. The Department of Molecular and Human Genetics at Baylor College of Medicine derives revenue from genetic testing offered in the Medical Genetics Laboratories. A.L.M. declares no conflict of interest.

REFERENCES

- Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. *Annu Rev Med* 2012;63:35–61.
- Surbone A. Social and ethical implications of BRCA testing. *Ann Oncol* 2011;22(suppl 1):i60–i66.
- Hogarth S, Javitt G, Melzer D. The current landscape for direct-to-consumer genetic testing: legal, ethical, and policy issues. *Annu Rev Genomics Hum Genet* 2008;9:161–182.
- Evans JP, Green RC. Direct to consumer genetic testing: Avoiding a culture war. *Genet Med* 2009;11:568–569.
- Wolf SM, Lawrenz FP, Nelson CA, et al. Managing incidental findings in human subjects research: analysis and recommendations. *J Law Med Ethics* 2008;36:219–248, 211.
- Kohane IS, Hsing M, Kong SW. Taxonomizing, sizing, and overcoming the incidentalome. *Genet Med* 2012;14:399–404.
- Evans JP, Berg JS. Next-generation DNA sequencing, regulation, and the limits of paternalism: the next challenge. *JAMA* 2011;306:2376–2377.
- Johnson AD, Bhimavarapu A, Benjamin EJ, et al. CLIA-tested genetic variants on commercial SNP arrays: potential for incidental findings in genome-wide association studies. *Genet Med* 2010;12:355–363.
- Cho MK. Understanding incidental findings in the context of genetics and genomics. *J Law Med Ethics* 2008;36:280–285, 212.
- Berg JS, Khoury MJ, Evans JP. Deploying whole genome sequencing in clinical practice and public health: meeting the challenge one bin at a time. *Genet Med* 2011;13:499–504.
- Ashley EA, Butte AJ, Wheeler MT, et al. Clinical assessment incorporating a personal genome. *Lancet* 2010;375:1525–1535.
- MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012;335:823–828.
- Cheung SW, Shaw CA, Yu W, et al. Development and validation of a CGH microarray for clinical cytogenetic diagnosis. *Genet Med* 2005;7:422–432.
- Shaffer LG, Kashork CD, Saleki R, et al. Targeted genomic microarray analysis for identification of chromosome abnormalities in 1500 consecutive clinical cases. *J Pediatr* 2006;149:98–102.
- Adams SA, Coppinger J, Saitta SC, et al. Impact of genotype-first diagnosis: the detection of microdeletion and microduplication syndromes with cancer predisposition by aCGH. *Genet Med* 2009;11:314–322.
- Pichert G, Mohammed SN, Ahn JW, Ogilvie CM, Izatt L. Unexpected findings in cancer predisposition genes detected by array comparative genomic hybridisation: what are the issues? *J Med Genet* 2011;48:535–539.
- Boone PM, Bacino CA, Shaw CA, et al. Detection of clinically relevant exonic copy-number changes by array CGH. *Hum Mutat* 2010;31:1326–1342.
- Bi W, Breman AM, Venable SF, et al. Rapid prenatal diagnosis using uncultured amniocytes and oligonucleotide array CGH. *Prenat Diagn* 2008;28:943–949.
- Ou Z, Kang SH, Shaw CA, et al. Bacterial artificial chromosome-emulation oligonucleotide arrays for targeted clinical array-comparative genomic hybridization analyses. *Genet Med* 2008;10:278–289.
- Shaw CJ, Shaw CA, Yu W, et al. Comparative genomic hybridisation using a proximal 17p BAC/PAC array detects rearrangements responsible for four genomic disorders. *J Med Genet* 2004;41:113–119.
- Mills RE, Walter K, Stewart C, et al. Mapping copy number variation by population-scale genome sequencing. *Nature* 2011;470:59–65.
- Girirajan S, Brkanac Z, Coe BP, et al. Relative burden of large CNVs on a range of neurodevelopmental phenotypes. *PLoS Genet* 2011;7:e1002334.
- Depienne C, Fedirko E, Forlani S, et al. Exon deletions of SPG4 are a frequent cause of hereditary spastic paraplegia. *J Med Genet* 2007;44:281–284.
- Lindor NM, McMaster ML, Lindor CJ, Greene MH. Concise handbook of familial cancer susceptibility syndromes - second edition. *J Natl Cancer Inst Monographs* 2008;38:1–93.
- Walsh T, Casadei S, Coats KH, et al. Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *JAMA* 2006;295:1379–1388.
- Cybulski C, Wokolorczyk D, Huzarski T, et al. A large germline deletion in the Chek2 kinase gene is associated with an increased risk of prostate cancer. *J Med Genet* 2006;43:863–866.
- Heger M. Sequencing breaks into the clinic in 2011, but ethical, regulatory, and technical hurdles remain. *Clinical Sequencing News*, 2012. <http://www.genomeweb.com/sequencing/sequencing-breaks-clinic-2011-ethical-regulatory-and-technical-hurdles-remain>
- Ringpfeil F, Nakano A, Uitto J, Pulkkinen L. Compound heterozygosity for a recurrent 16.5-kb Alu-mediated deletion mutation and single-base-pair substitutions in the ABCC6 gene results in pseudoxanthoma elasticum. *Am J Hum Genet* 2001;68:642–652.
- Plomp AS, Hu X, de Jong PT, Bergen AA. Does autosomal dominant pseudoxanthoma elasticum exist? *Am J Med Genet A* 2004;126A:403–412.
- Muntoni F, Di Lenarda A, Porcu M, et al. Dystrophin gene abnormalities in two patients with idiopathic dilated cardiomyopathy. *Heart* 1997;78:608–612.
- Mital A, Kumari D, Gupta M, Goyle S. Molecular characterisation of Duchenne muscular dystrophy and phenotypic correlation. *J Neurol Sci* 1998;157:179–186.
- Potocki L, Chen KS, Koeuth T, et al. DNA rearrangements on both homologues of chromosome 17 in a mildly delayed individual with a family history of autosomal dominant carpal tunnel syndrome. *Am J Hum Genet* 1999;64:471–478.
- Lupski JR. Structural variation in the human genome. *N Engl J Med* 2007;356:1169–1171.
- El-Hattab AW, Zhang F, Maxim R, et al. Deletion and duplication of 15q24: molecular mechanisms and potential modification by additional copy number variants. *Genet Med* 2010;12:573–586.

35. Girirajan S, Rosenfeld JA, Cooper GM, et al. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* 2010;42:203–209.
36. Cassa CA, Savage SK, Taylor PL, Green RC, McGuire AL, Mandl KD. Disclosing pathogenic genetic variants to research participants: quantifying an emerging ethical responsibility. *Genome Res* 2012;22:421–428.
37. 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;3:574–580.
38. Field MJ, Boat TF (eds). *Rare Diseases and Orphan Products: Accelerating Research and Development*. National Academy of Sciences, 2011.
39. Hayden EC. Geneticists debate what to tell patients about clinical genome sequences. *Nature News Blog*, 2012. <http://blogs.nature.com/news/2012/03/geneticists-debate-what-to-tell-patients-about-clinical-genome-sequences.html>