Contemplating effects of genomic structural variation

Janet A. Buchanan, PhD[1], and Stephen W. Scherer, PhD[1,2]

Two developments have sparked new directions in the genetics-to-genomics transition for research and medical applications: the advance of whole-genome assays by array or DNA sequencing technologies, and the discovery among human genomes of extensive submicroscopic genomic structural variation, including copy number variation. For health care to benefit from interpretation of genomic data, we need to know how these variants contribute to the phenotype of the individual. Research is revealing the spectrum, both in size and complexity, of structural genotypic variation, and its association with a broad range of human phenotypes. Genomic disorders associated with relatively large, recurrent contiguous variants have been recognized for some time, as have certain Mendelian traits associated with functional disruption of single genes by structural variation. More recent examples from phenotype- and genotype-driven studies demonstrate a greater level of complexity, with evidence of incremental dosage effects, gene interaction networks, buffering and modifiers, and position effects. Mechanisms underlying such variation are emerging to provide a handle on the bulk of human variation, which is associated with complex traits and adaptive potential. Interpreting genotypes for personalized health care and communicating knowledge to the individual will be significant challenges for genomics professionals. *Genet Med* 2008:10(9):639–647.

Key Words: *Structural variation, copy number variants, CNV, genome scanning, autism, complex traits*

In a medical context, we investigate human genomes to explain, anticipate, or mitigate their effects on the affiliated phenotypes. Researchers focus on collections of data about groups, trends, and mechanisms, but health care workers need to take the knowledge gleaned back to the individual. Whether for research or for clinical intervention, the questions may be driven primarily by phenotype or by genotype. Phenotype-driven research begins with a cohort of individuals who share characteristics, and commonality is sought among their genetic variants. Genotype-driven research ascertains individuals according to particular genetic variants and then documents the associated phenotypes.[1] In a clinical context, phenotype-driven investigation is for diagnosis. A trait or condition brings to medical attention an individual whose genome may then be assayed for evidence of a particular genomic variant to confirm a suspected diagnosis, or scanned for evidence of anything unusual and assessed for the likelihood of a causal relationship. In contrast, a genotype-driven clinical investigation, such as a family study or population screening, characterizes a genotype to anticipate the possible phenotypic outcome (and perhaps to intervene).

In the first 50 or so years of clinically applied genetics, traditional karyotype analysis has always been a global genomic assay with limited (albeit improving) resolution; whereas, other laboratory investigations have been relatively targeted in nature, typically interrogating one genetic locus at a time. With recent technologic developments, the cytogenetic and molecular approaches are merging into one that is global in scope, but with high sensitivity and resolution. Not only it is becoming feasible—indeed practical—to scan the entire genome simultaneously in search of particular genetic flags, but the unity of the data will eventually allow a comprehensive interpretation of the genomic findings.

Two recent technologies are rapidly changing our entire approach to studying the human genome. Microarrays in various forms—some relatively targeted and others with genome-wide capacity—have been developed for comparative genomic hybridization and detection of chromosome imbalance, or for single nucleotide genotype analysis. During the same time, rapidly evolving DNA sequencing methods produced the first two genome sequences, each from a single individual, published in 2007[2] and 2008.[3] These were each accomplished at a fraction of the expense of the Human Genome Project reference sequence, and ongoing cost improvements are ushering in the era of the personal genome and the means to directly assay genomic variation.

Perhaps the most striking finding to emerge from these new technologies has been the extent of interindividual variation accounted for, not by single base-pair differences such as single nucleotide polymorphism (SNPs) or rare mutations, but by

From [1]The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children; and [2]Department of Molecular Genetics, University of Toronto, Toronto, Canada.

Stephen W. Scherer, The Centre for Applied Genomics, The Hospital for Sick Children, 14th Floor, Toronto Medical Discovery Tower/MaRS Discovery District, 101 College Street, Toronto, Ontario, M5G 1L7 Canada. E-mail: steve@genet.sickkids.on.ca.

structural variants involving larger segments of DNA.[4–7] These include both balanced rearrangements (inversions and translocations) and copy number variants (CNVs).[7–9] The genome is neither as binary nor as static as we might have surmised; rather, it can be dynamic, with plenty of iteration and absence. Some variants in this class are structurally simple, but others are complex, and the CNVs can reflect either loss or gain of genetic material relative to a designated reference genome.

Of particular interest is that these structural variants are associated with a full spectrum of phenotypic outcomes, from unrecognizable or inconsequential through to those that may be incompatible with life (Fig. 1). Researchers are documenting these variant genomic sites at an exponential pace, which we anticipate will approach an asymptote within the next 5 years, at least with respect to the polymorphic variants. The concomitant activity is to catalogue the nature and extent of human variation associated with each of these variant loci—an activity that is likely to be ongoing indefinitely. For this genotypic information to be useful, particularly in a clinical context, it needs to be related to phenotypic outcomes, and to that end, we have barely scratched the surface. This area of investigation, in the realm that is intermediate between microscopic chromosome analysis and gene mutation assays, is already revealing both genotypes and phenotypes that can be far more complex than those associated with classical cytogenetic or Mendelian traits. From that complexity, however, is likely to emerge the explanations not only for overtly maladaptive syndromes, disorders and diseases but also for adaptive traits, variable responses and susceptibilities, common and complex traits, subtle individual distinguishing features or idiosyncrasies, and the opportunity to accommodate changing environments.

THE GENOTYPIC SPECTRUM

Array technologies and whole genome sequencing are finally drawing our focus to the kind of variation that is intermediate in size, completing the spectrum between single base variants (mutations or SNPs) and microscopically-visible aneuploidies or heteromorphisms. Most of the present discussion will pertain to CNVs, which seem to be the more prevalent form of structural variation,[2,10–12] though currently accessible methods detect the quantitative variants more readily than balanced translocations or inversions.[13]

CNVs have been defined operationally as involving segments of DNA that are 1 kb or greater in size,[7] though this limit is somewhat arbitrary from a functional perspective. Smaller variants, such as minute insertions and deletions or variable number of tandem repeats are excluded from the working definition and discussion, but are recognized as part of the full genotypic spectrum.[13]

Many genomic structural variants characterized to date have been associated with structures called "segmental duplications" or "low-copy repeats": segments that predispose to genomic rearrangement during meiosis by nonallelic homologous recombination (NAHR).[14] Because these sequences are vulnerable, the resultant rearrangements tend to recur, creating clusters of variants with common endpoints. Other rearrangements that are not in association with such duplicated elements are more randomly distributed and nonrecurrent. Two mechanisms have been proposed to explain the latter: nonhomologous end joining[15] and replication fork stalling and template switching.[16] Recent higher-resolution data are contradictory as to the predominant mechanism mediating the majority of genomic imbalances. For example, two studies es-
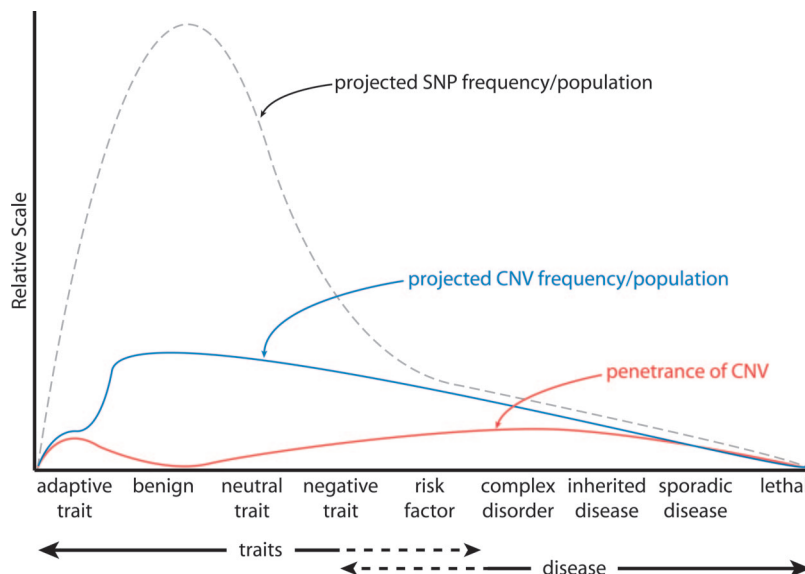

Fig. 1. CNV characteristics and frequency. Conceptual curves show projected frequencies in the population for SNPs and point mutations (dashed gray) and CNVs (blue) with different characteristic associations. Some examples of traits that correspond to the different groupings are described in Table 1 and Figure 2. Although real data would not currently follow these curves, we anticipate that with time, the thorough identification of associations will show that CNVs follow the exhibited trend. These curves will also be strongly affected by environmental conditions and relationships with other variants in each genome. A penetrance curve (red) is shown for the CNVs. This is a relative curve, where 100% penetrance is indicated by a height equal to the CNV frequency curve. Two recent studies caught our interest, exemplifying how CNVs previously designated to be 'benign' might move to the 'adaptive trait' or 'neutral trait' groupings, involving α- amylase[104] and testosterone metabolism.[105] Two other new studies provide examples of rare recurrent susceptibility CNVs found in schizophrenia[106] and novel mechanisms for germ-line CNV effects in cancer predisposition.[107]

timated that only 9%[17] or 14%[18] of structural breakpoints fall within repetitive sequences, suggesting that nonrecurrent mechanisms predominate, whereas another study[10] demonstrated that 47% of breakpoints follow NAHR rules. Some of these differences can be attributed to ascertainment biases in the technologies and the size of variants being assayed, but more data will be required before we fully understand the genesis of structural variation. There are only a few primary reports assessing new mutation rates for CNVs.[6,19] Emerging observations suggest a locus-specific rate of $1.6 \times 10^{-6} - 1.2 \times 10^{-4}$, which is 3–4 times greater than that observed for SNPs.[20,21] Moreover, it seems that most CNV gains are local duplication events, but new studies of both human and *Drosophila* also demonstrate transposition events.[10,22]

The larger CNVs (>50 kb) described in recent population surveys seem skewed toward rare variants.[6] Their distribution also seems to be nonrandom, with more in the subtelomeric and centromeric regions of chromosomes.[23,24] Overall, however, the spectrum of structural variation is extensive. In terms of phenotypic impact, the location of these variants in relation to genes is particularly germane. They may occur anywhere, but are more common in regions devoid of genes, known as "gene deserts." Some comprise multigene segments that are deleted, duplicated or moved; others involve segments contained within functional genes; yet others are in nongene segments that nonetheless have a regulatory role on gene function. When genes are involved, impact of the variant will be contingent upon the function(s) of these genes. "Essential" genes are less likely to be tolerant of any disruption, and de novo variants that affect them may face strong selection[25] (Fig. 1). The functions of "disease-associated" genes are sufficiently important that their disruption or copy number change may lead to a clinically-recognizable phenotype. Other genes can have more subtle effects on phenotype and fitness, being more robust or more discretionary, and it is the genetic elements at this end of the spectrum that seem to have the greatest relationship with CNVs and other structural variants.

We can classify the genomic structural variants according to form. "Balanced" rearrangements involve no loss or gain of genetic material, and include intrachromosomal and interchromosomal translocations and inversions. These are not detectable by current array-based methods, but are revealed by direct comparison of genome sequences or cytogenomic approaches. Some ostensibly balanced rearrangements in individuals with a clinical phenotype have, on closer scrutiny with the higher resolution methods, been found to comprise subtle deletions or duplications at their breakpoints, or to be associated with additional changes elsewhere in the genome[26–29] (Fig. 2). Truly balanced rearrangements, even when they have no functional effect in a carrier, can, nonetheless, create genomic instability for future generations.[30–33]

The remaining structural variants are "unbalanced" with respect to DNA content and are called CNVs. These can involve a relative loss or gain (deletion or replication) of genetic material. Methods to detect CNVs include comparative and directly-quantitative array screening strategies, sequence analysis, and site-focused assays such as quantitative polymerase chain reaction (qPCR) and fluorescence in situ hybridization (FISH).
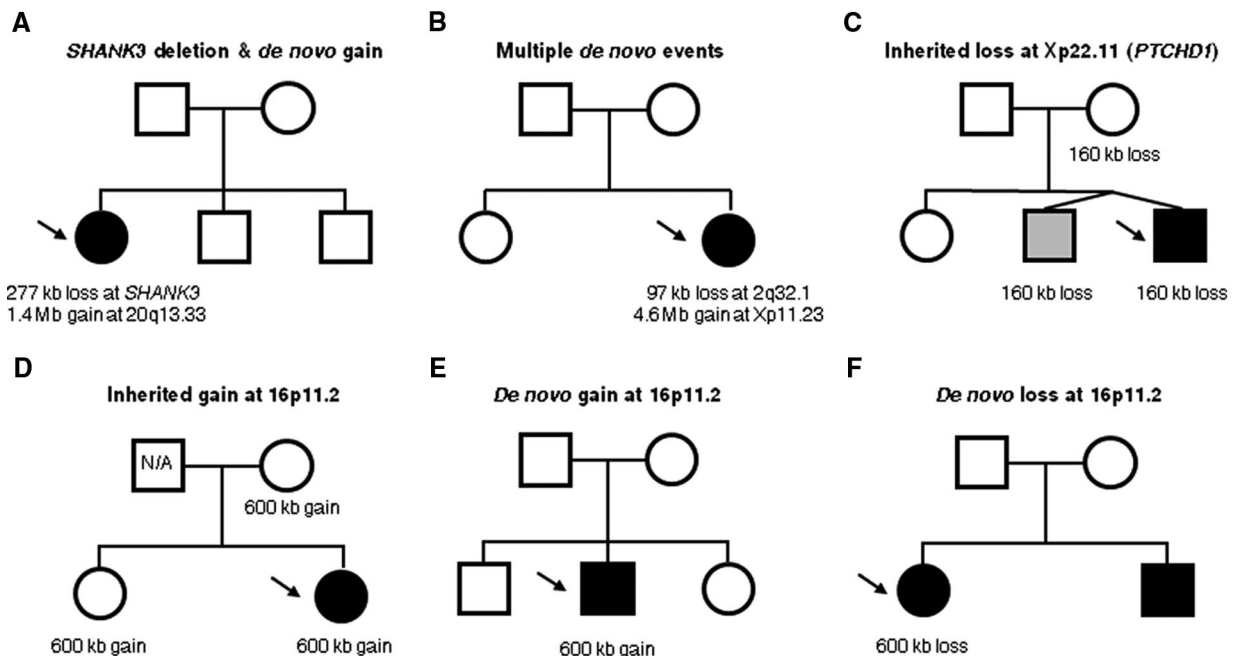

Fig. 2. CNV and phenotypic complexity in autism spectrum disorder (ASD) pedigrees. The size of each de novo or inherited CNV is shown below each family member. Arrows identify probands. ASD cases have filled symbols (gray denotes developmental delay but not a definitive ASD diagnosis); open symbols denote individuals who do not have ASD. (A) ASD probands may carry multiple de novo events including those overlapping genes known to be associated with ASD (*SHANK3*)[73] or (B) large de novo rearrangements accompanied by smaller de novo events at different loci. (C) Male probands may inherit chromosome X deletions (*PTCHD1*) from female carriers, unmasking an identical CNV in fraternal siblings with variable expression of ASD. (D). Recurrent CNV gains and losses (representing the reciprocal events) in unrelated probands (at 16p11.2) may be inherited from non-ASD parents or (E, F) present as de novo events.

The form for an individual structural variant can be as simple as a segmental deletion, or a highly complex genomic rearrangement involving multiple elements.[6,10,34,35] (For detailed description of variant classes.[6,34]) Collectively, there is also diverse complexity for the loci at which these events occur, since a given variant region may demonstrate overlapping but nonidentical rearrangements when genomes of different individuals are compared, and CNVs can be multiallelic. A challenge going forward, particularly for complex traits and diseases, will be to determine how structural variants and single nucleotide variants might interact, looking at mutation rates and linkage disequilibrium, and to develop new models to extract these data.[36,37]

Table 1 classifies some structural variants according to their genotypic form and features, from deletion CNVs through balanced rearrangements to CNVs with large relative gains of material. Clinically-relevant illustrative examples are listed for each. Other good reviews on this topic have been published.[7–9,38–40]

PHENOTYPIC SPECTRUM

The observable qualities of an organism comprise its phenotype. As with individual variation directed by single nucleotide variants, the phenotypic impact related to structural variants in the genome can be as severe as to cause embryonic lethality, or at the other end of the spectrum, to have little or no discernable outcome. In between, they can be associated with degrees of dysfunction, which, beyond a certain threshold, are called "disease,"[41,42] though that threshold can sometimes be moved

Table 1
Spectra of CNV Genotypes and phenotypes[a]

CNV genotype features	Illustrative Phenotypes (*gene symbol* or locus)
Single gene/functional disruption or loss	
Recessive	Duchenne/Becker Muscular Dystrophy (*DMD*)[57–59]
Dominant	Neurofibromatosis 1 (*NF1*)[60–62]
	Tuberous sclerosis (*TSC1* or *TSC2*)[63]
	Sotos syndrome (*NSD1*)[33,64,65]
	Coloboma, heart anomaly, choanal atresia, retardation, genital and ear anomalies (CHARGE) syndrome (*CHD7*)[54]
Single gene/dosage effects	
Recessive	Spinal muscular atrophy (*SMN1*, *SMN2*)[66,67]
	Pelizaeus-Merzbacher disease (*PLP1*)[16,68]
Dominant	Early-onset Alzheimer disease (*APP*)[69,71]
	22q autism (*SHANK 3*)[27,72,73]
Position effects	Aniridia (*PAX6*)[74,75]
	Triphalangeal thumb-polysyndactyly (TPT-PS) syndrome[76]
Multiallelic effects	Crohn disease predisposition (*HBD-2*)[77]
	Systemic autoimmunity predisposition (*FCGR3*)[78,79]
	Parkinson disease (*SNCA*)[80–82]
	HIV/AIDS susceptibility – Kawasaki disease susceptibility rheumatoid arthritis predisposition (*CCL3L1*)[83–85]
Multigene CNV/contiguous	
Recurrent	Williams-Beuren syndrome[30,86,87]—7q11.23 duplication syndrome[86,88]
	17q21.3 microdeletion syndrome[31,32,89,90]
	1q41q42 microdeletion syndrome[1]
	16p11.2-p12.2 microdeletion syndrome(s)[91]
Nonrecurrent	Potocki-Lupski syndrome (dup(17)(p11.2p11.2))[92]
	Wilms tumor, aniridia, genitourinary anomalies, mental retardation (WAGR) syndrome (11p13)[75]
Multigene/noncontiguous, heterogeneous	Autism[27,48–50,93–95]
	Bipolar disorder[96]
	Schizophrenia[97,98]
	Age-related macular degeneration[99]
Buffering or modifier effects	Thrombocytopenia/absent radius (TAR) syndrome (1q21.1 deletion/modifier)[100]
	Spinal muscular atrophy (*SMN1*/*SMN2*)[66]
Epigenetic effects	Silver-Russell syndrome (11p15 duplication)[101]
	Developmental verbal dyspraxia del(7)(q31) (*FOXP2*)[102]
Somatic mosaicism	Rubenstein-Taybi syndrome (*CREBBP*)[103]
	Tuberous sclerosis (*TSC1*, *TSC2*)[63]
	Aniridia (*PAX6*)[75]

[a]Selected examples, focusing on more recent discoveries (2006–2008) that are shaping the CNV field.

by clinical interventions. Traits may also be relatively adaptive or maladaptive in different environmental contexts. From a clinical perspective, structural variants can be the basis for severely disabling syndromes or diseases, for single-gene disorders and those involving large chromosomal segments. Their impact is being recognized much more, however, on the more quantitative traits where they can have somewhat incremental effects on phenotype and fitness. They are anticipated to be even more important for predisposition to common threats to health, such as heart disease, diabetes, cancer, or dementia, particularly in those with apparently complex etiology. Much of structural variation is not gene- or disease-associated and has become widely dispersed in the absence of selective pressure (Fig. 1). It is becoming clear that these variants are important contributors to traits that not only create a state of disease or health, but influence quality of life and simple human differences.

The earlier phenotype-driven research has detected more genomic deletions than duplications—a bias probably due to the typically milder phenotype associated with gain of genetic material.[24] The corollary, however, is less selection pressure, and the relative abundance of CNV gains is becoming apparent with genotype-driven approaches.

TECHNOLOGY AND DATABASES

Both array-based and sequencing technologies are evolving quickly to adapt to the recognition of CNVs and other structural variants as important genomic elements to be ascertained, documented, and interpreted.[13] Initially, there has been a detection bias in favor of medium-to-large and non-complex variants. The genome-wide arrays are designed for breadth of detection and have limited ability to resolve endpoints of variant sequences with precision, or to determine whether variants are exactly the same or overlapping. Targeted arrays, and more labor-intensive approaches such as qPCR or FISH, can add information to allow more detailed interpretation. Repetitive elements are inherently challenging for DNA sequencing, and more variant regions are being discovered as gaps in the reference sequences are gradually conquered. The higher-density arrays and higher-throughput sequencing will also increase detection of variant regions that are smaller, more complex, and more difficult to interpret. A particular challenge for relating specified genotypes to phenotypes is that the high-throughput array technologies reveal relative copy-number differences but have limited ability to resolve absolute copy number—a matter particularly relevant to multiallelic loci.[9,13] Further, they ascertain a diploid genotype and do not directly discern the component haploid variant alleles. Finally, the yet limited ability to resolve CNV breakpoints will, for some time to come, compromise their interpretation, particularly for predictive purposes.

The Database of Genomic Variants (DGV) (http://projects.tcag.ca/variation/) was established to catalogue genomic variation from human control samples, as a support for research correlating genomic variation with phenotypes.[4,43] It is important to keep in mind that it derives from individuals deemed to be "healthy controls," but the amount of phenotypic documentation is limited. A control subject for a cancer study, for example, may not have been assessed for health status with respect to blood pressure. Health is not static, and the status of a research participant could change. The DGV comprises structural variants not known to cause overt disease, but does not necessarily exclude alterations associated with complex, variable, mild, or late-onset phenotypes. The database is an essential research tool, but caution is needed in its use for prediction of health outcomes.

Databases such as Database of Chromosomal Imbalance and Phenotype in Humans using Ensemble Resources (DECIPHER) (https://decipher.sanger.ac.uk/) and others (reviewed in[44]) are intended to marry clinical phenotypic descriptions with data about structural variation. Currently, such databases house, primarily, information on highly penetrant variants that cause overt phenotypes such as dysmorphic syndromes and cognitive impairment. As the field moves from examining the role of structural variants in rare, highly penetrant disorders to that in common and complex traits and disease, the overlap of content in "control" and "disease" databases such as DGV and DECIPHER, respectively, will increase. Moreover, as depicted in Figure 1 and Figure 2D, some structural variants previously annotated as benign or neutral in their effect will be reclassified as predisposing, risk factors, or partially penetrant alleles.[9,34,45]

MODELING RELATIONSHIPS BETWEEN STRUCTURAL VARIANT GENOTYPE AND PHENOTYPE

Some structural variants influence single genes and behave as simple Mendelian traits, and others merge with the realm of traditional cytogenetics. They are coming to the fore, however, as contributors to the "everything else" category from textbook genetics—that of complex traits. CNVs underlie common variation that may be selectively advantageous, neutral, or detrimental in different contexts. To understand the relationships to phenotype, our thinking and analysis will need to evolve from models with simple, linear, binary, and discontinuous concepts to those that are complex, networked, multifocal, and continuous.[25] CNVs will be responsible for complex additive and/or epistatic effects and for buffering. More elements will have individually small incremental effects, and be associated, not only with threshold traits, but also with those that are continuously variable or quantitative.

The structural variants can impact gene function[8,35] (or not[46]) in several ways. They can create functional loss through deletion or disruption of one or more genes, behaving as dominant or recessive alleles according to the cellular function of the impacted gene product(s). They may cause disruption of a regulatory element with any number of possible positive and negative sequelae, including imprinting and differential allelic gene expression.[47] Replication of genes may increase the protein product, or buffer the impact of other genetic variation.[41] Rearrangements can have position effects on gene expression[37] by separating genes from their regulatory elements or putting

them into a different genomic context with new epigenetic factors. They may also generate novel fusion products.

A number of features of the variant genotype will be relevant to the concomitant phenotype:

1. The location of the structural variant with respect to genes or regulatory regions.
2. Dosage characteristics of the variant—whether there is a loss or gain of genetic material.
3. When functional genes are impacted, the dosage-sensitivity of the related gene. Proteins involved in complexes are more likely to require dosage balance for optimal function.
4. Extent of the variant—involving dosage effects on one or on multiple genes.
5. Cellular role of the impacted gene product.

Phenotypes associated with particular variant regions can be relatively consistent, or highly variable. Consistency may reflect involvement of a single gene, but could also be due to a multigene segmental variant passed on from a common ancestor. Alternatively, concordance is often the result of recurrent rearrangements driven by predisposing genomic sequences, such as nearby segmental duplications, or a balanced variant such as an inversion. Phenotypic variability can have many causes. A syndrome or disorder, for example, might be defined by its core gene(s), but the extent of the CNV and its encompassing of nearby genes may influence the phenotype. Most importantly, the overall genomic context in which a given structural variant functions, and the environmental variables, will be different for each individual in which the variant is found.

In Figure 2, we present results from our group's studies of CNVs in individuals with autism spectrum disorder (ASD).[27] Albeit still simplistic, the data begin to reveal some of the complexities to be considered when attempting to make proper genotype-to-phenotype associations.[48] For example, with higher-resolution arrays, multiple de novo CNV events may be identified in individual samples (Fig. 2, A and B). CNVs can be unmasked, depending on their position and context in the genome, which might influence expressivity and penetrance (Fig. 2, C and D). Gains and losses at the same locus can lead to overlapping phenotypes, with variable penetrance and potential contributions from other risk alleles (Fig. 2, D–F). Among individuals with de novo structural variants in our ASD cohort, more than 10% had two or more variants, cautioning against assigning independent causation to all de novo structural variants observed.[27] As expected for a common, complex disorder with potentially numerous contributing loci, in the family illustrated in Figure 2, F, the CNV deletion is detected in only one of the two ASD sibs. Although the 16p11.2 CNV may be more prevalent in autism families than among control subjects,[27,49,50] the genomic characteristics demonstrated in ASD families in Figure 2, D–F suggest that this variant is neither necessary nor sufficient to cause ASD. We need to consider additional independent potential risk factors, including those that are genetic, epigenetic, gender-related, environmental, or

stochastic in origin. A recent comparison of CNVs between monozygotic twins[51] also draws attention to the prevalence of somatic events that can create mosaicism for structural variants, with possible contribution to a variety of phenotypes.

IMPLICATIONS IN THE APPLICATION TO HEALTH CARE

Genotyping arrays are already well established as front-line research tools, and are rapidly being integrated into mainstream medical practice, and, more recently, into consumer genomics. Whole genome sequencing is likely but a few years behind, and all of these approaches already generate far more genomic data than we can translate or interpret. Untargeted investigations, in particular, will yield huge amounts of data about variants—concerning both groups and individuals— that may not be interpretable for some time, but will still be on the table.[52] In a research context, the question of how to manage incidental findings is but one looming dilemma being anticipated and contemplated by lawyers and ethicists.[53] It is exciting to watch the resurgence of discovery as newly-recognized CNVs open investigative paths to issues that had been stalled. Even when CNVs are infrequent contributors to a particular phenotype, the rare cases are beginning to draw attention to relevant genes for further investigation of nucleotide variation (see e.g., Ref. 54), and on to functional studies. As research findings are turned into applications for medical practice, we must keep in mind that statistical inferences about groups and populations are not the same as implications for individual risk. Our ability to use these genotypic data to explain a phenotype (i.e., diagnosis) will be far ahead of our ability to predict outcomes, and this will be particularly true as the structural variants allow access to the massive realm of common and complex traits and disease.

At present, it is very difficult to know whether a given de novo variant is pathologic, and some of the reason for this is our still rudimentary knowledge of mutation rates in different classes of structural variants.[21] It is also difficult to know whether an inherited variant is necessarily benign in a particular genomic environment. Efforts to document and catalogue genotypic data in relation to phenotypic information from thousands of phenotype-classified individuals and controls, should eventually make it possible to do so. As our focus evolves from primarily gene-specific investigations to what will eventually be routine whole-genome analysis, we will be driven to scrutinize individual phenotypes more closely. Some who are classified in research protocols as being unaffected by a particular trait or disease may, upon retrospective evaluation, be found to carry subtle signs, and such observations will help in understanding the spectrum of variation associated with particular structural variants.

As we come to recognize the extent of structural variation, it is making us aware of the degree to which the genome is fluid and unstable, both through germlines and in somatic events. Knowledge of this area of human variation is both filling in gaps and reminding us of the extent of complexity in biological systems. This should keep us cognizant of the opportunities to

make diagnostic or predictive errors through erroneous assumptions, and the further we become dependent upon interactive and computer-based interpretations, the more such risks will emerge.

If the process of documenting and cataloguing the complex genomic variants and rearrangements is challenging, it is an almost daunting task to do the same systematically for phenotypic traits. Nomenclature needs to be standardized,[55] and means found to accommodate subjectivity and the fact that health status is not static, among other complexities.[44] Relational databases can then go forward, connecting observed variation in genomes to phenotypic outcomes, to allow a knowledge base for application to health care.

As personalized medicine becomes more common, interpreting the amount of information potentially available for an individual could quickly overwhelm. We are still inclined to look at one locus or variant at a time, and manually interpret the observations in isolation. This approach will continue to be appropriate for many of the genetic variants described to date, that have individually significant impact on phenotypes. As the bulk of information from variants is brought forward, however, particularly from CNVs, more and more will we recognize those with small incremental effects, and complex interactions with other elements. The holistic opportunities offered by genome-wide assays will gradually be realized, facilitated by tools with which to interpret complex networks of genomic interactions and to account for epigenetic and nongenetic factors, such as time, place, environment, and experience.

There will be an expanding role for professionals trained in such interpretation, as research delivers into the arena of applications for individual health care. Today's clinical molecular geneticists and cytogeneticists will merge skills and acquire completely new ones to fill this niche. There will be an enhanced role for counselors to communicate the interpreted information to the individuals, families or communities who will be impacted.[52,56] They will be particularly challenged with issues of complexity, subtlety, and uncertainty at the same time as a sheer volume increase in demand for their services. The delivery of information about human variation will be as much an art as a science for a long time to come.

We are reminded of Charles Scriver's wise insight that, "genetic variation itself is normal; it is dis-ease only when we experience it as illness. The professional will understand the process underlying the disease; the healer will alter the perception of illness."[42] In contemplation, we add that these emerging studies of genomic structural variation will bring the professional to the phenotype and the healer to the genome, in pursuit of common answers to their respective questions.

References

1. Shaffer LG, Theisen A, Bejjani BA, et al. The discovery of microdeletion syndromes in the post-genomic era: review of the methodology and characterization of a new 1q41q42 microdeletion syndrome. *Genet Med* 2007;9:607–616.
2. Levy S, Sutton G, Ng PC, et al. The diploid genome sequence of an individual human. *PLoS Biol* 2007;5:e254.
3. Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008;452:872–876.
4. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. *Nat Genet* 2004;36:949–951.
5. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. *Science* 2004;305:525–528.
6. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature* 2006;444:444–454.
7. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet* 2006;7:85–97.
8. Feuk L, Marshall CR, Wintle RF, Scherer SW. Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum Mol Genet.* 2006;15 Spec No 1:R57–66.
9. Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* 2007; 8:639–646.
10. Kidd JM, Cooper GM, Donahue WF, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature* 2008;453:56–64.
11. Khaja R, Zhang J, MacDonald JR, et al. Genome assembly comparison identifies structural variants in the human genome. *Nat Genet* 2006;38:1413–1418.
12. Tuzun E, Sharp AJ, Bailey JA, et al. Fine-scale structural variation of the human genome. *Nat Genet* 2005;37:727–732.
13. Scherer SW, Lee C, Birney E, et al. Challenges and standards in integrating surveys of structural variation. *Nat Genet* 2007;39(suppl 7):S7–S15.
14. Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. *Trends Genet* 2002;18:74–82.
15. Moore JK, Haber JE. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae. Mol Cell Biol* 1996;16:2164–2173.
16. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 2007; 131:1235–1247.
17. Korbel JO, Urban AE, Affourtit JP, et al. Paired-end mapping reveals extensive structural variation in the human genome. *Science* 2007;318:420–426.
18. Perry GH, Ben-Dor A, Tsalenko A, et al. The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet* 2008;82:685–695.
19. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. *Science* 2007;316:445–449.
20. van Ommen GJ. Frequency of new copy number variation in humans. *Nat Genet* 2005;37:333–334.
21. Lupski JR. Genomic rearrangements and sporadic disease. *Nat Genet* 2007; 39(suppl 7):S43–S47.
22. Emerson JJ, Cardoso-Moreira M, Borevitz JO, Long M. Natural selection shapes genome-wide patterns of copy-number polymorphism in Drosophila melanogaster. *Science* 2008;320:1629–1631.
23. Freeman JL, Perry GH, Feuk L, et al. Copy number variation: new insights in genome diversity. *Genome Res* 2006;16:949–961.
24. Turner DJ, Miretti M, Rajan D, et al. Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nat Genet* 2008;40:90–95.
25. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. *Proc Natl Acad Sci U S A* 2007;104:8685–8690.
26. Gribble SM, Prigmore E, Burford DC, et al. The complex nature of constitutional de novo apparently balanced translocations in patients presenting with abnormal phenotypes. *J Med Genet* 2005;42:8–16.
27. Marshall CR, Noor A, Vincent JB, et al. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 2008;82:477–488.
28. Baptista J, Mercer C, Prigmore E, et al. Breakpoint mapping and array cgh in

translocations: comparison of a phenotypically normal and an abnormal cohort. *Am J Hum Genet* 2008;82:927–936.

29. Higgins AW, Alkuraya FS, Bosco AF, et al. Characterization of apparently balanced chromosomal rearrangements from the developmental genome anatomy project. *Am J Hum Genet* 2008;82:712–722.

30. Scherer S, Osborne L. Williams-Beuren Syndrome. In: Lupski J, Stankiewicz P, editors. Genomic disorders: the genomic basis of disease. Totowa, NJ: Humana Press, 2006:221–236.

31. Shaw-Smith C, Pittman AM, Willatt L, et al. Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat Genet* 2006;38:1032–1037.

32. Koolen DA, Vissers LE, Pfundt R, et al. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat Genet* 2006;38:999–1001.

33. Visser R, Shimokawa O, Harada N, et al. Identification of a 3.0-kb major recombination hotspot in patients with Sotos syndrome who carry a common 1.9-Mb microdeletion. *Am J Hum Genet* 2005;76:52–67.

34. Conrad DF, Hurles ME. The population genetics of structural variation. *Nat Genet* 2007;39(suppl 7):S30–S36.

35. Hurles ME, Dermitzakis ET, Tyler-Smith C. The functional impact of structural variation in humans. *Trends Genet* 2008;24:238–245.

36. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. *Nat Genet* 2007;39(suppl 7):S37–S42.

37. Stranger BE, Forrest MS, Dunning M, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007;315:848–853.

38. Lee JA, Lupski JR. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron* 2006;52:103–121.

39. Estivill X, Armengol L. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet* 2007;3:1787–1799.

40. Lee C, Iafrate AJ, Brothman AR. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nat Genet* 2007;39(suppl 7):S48–S54.

41. Hartman JLt, Garvik B, Hartwell L. Principles for the buffering of genetic variation. *Science* 2001;291:1001–1004.

42. Scriver CR. The human genome project will not replace the physician. *CMAJ* 2004; 171:1461–1464.

43. Zhang J, Feuk L, Duggan GE, Khaja R, Scherer SW. Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. *Cytogenet Genome Res* 2006;115:205–214.

44. Van Vooren S, Coessens B, De Moor B, Moreau Y, Vermeesch JR. Array comparative genomic hybridization and computational genome annotation in constitutional cytogenetics: suggesting candidate genes for novel submicroscopic chromosomal imbalance syndromes. *Genet Med* 2007;9:642–649.

45. Cooper GM, Nickerson DA, Eichler EE. Mutational and selective effects on copy-number variants in the human genome. *Nat Genet* 2007;39(suppl 7):S22–S29.

46. Barber JC, Maloney VK, Kirchhoff M, Thomas NS, Boyle TA, Castle B. Transmitted duplication of 12q21.32–12q22 includes 48 genes and has no apparent phenotypic consequences. *Am J Med Genet A* 2007;143:615–618.

47. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A. Widespread monoallelic expression on human autosomes. *Science* 2007;318:1136–1140.

48. Tabor HK, Cho MK. Ethical implications of array comparative genomic hybridization in complex phenotypes: points to consider in research. *Genet Med* 2007;9: 626–631.

49. Kumar RA, KaraMohamed S, Sudi J, et al. Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet* 2008;17:628–638.

50. Weiss LA, Shen Y, Korn JM, et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 2008;358:667–675.

51. Bruder CE, Piotrowski A, Gijsbers AA, et al. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am J Hum Genet* 2008;82:763–771.

52. Caulfield T, McGuire AL, Cho M, et al. Research ethics recommendations for whole-genome research: consensus statement. *PLoS Biol* 2008;6:e73.

53. 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.

54. Vissers LE, van Ravenswaaij CM, Admiraal R, et al. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet* 2004;36:955–957.

55. Giardine B, Riemer C, Hefferon T, et al. PhenCode: connecting ENCODE data with mutations and phenotype. *Hum Mutat* 2007;28:554–562.

56. Darilek S, Ward P, Pursley A, et al. Pre- and postnatal genetic testing by array-comparative genomic hybridization: genetic counseling perspectives. *Genet Med* 2008;10:13–18.

57. Stockley TL, Akber S, Bulgin N, Ray PN. Strategy for comprehensive molecular testing

for Duchenne and Becker muscular dystrophies. *Genet Test* 2006;10:229–243.

58. White SJ, Aartsma-Rus A, Flanigan KM, et al. Duplications in the DMD gene. *Hum Mutat* 2006;27:938–945.

59. White SJ, den Dunnen JT. Copy number variation in the genome; the human DMD gene as an example. *Cytogenet Genome Res* 2006;115:240–246.

60. De Luca A, Bottillo I, Dasdia MC, et al. Deletions of NF1 gene and exons detected by multiplex ligation-dependent probe amplification. *J Med Genet* 2007;44:800–808.

61. Raedt TD, Stephens M, Heyns I, et al. Conservation of hotspots for recombination in low-copy repeats associated with the NF1 microdeletion. *Nat Genet* 2006;38:1419–1423.

62. Wimmer K, Yao S, Claes K, et al. Spectrum of single- and multiexon NF1 copy number changes in a cohort of 1,100 unselected NF1 patients. *Genes Chromosomes Cancer* 2006;45:265–276.

63. Kozlowski P, Roberts P, Dabora S, et al. Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations. *Hum Genet* 2007;121:389–400.

64. Saugier-Veber P, Bonnet C, Afenjar A, et al. Heterogeneity of NSD1 alterations in 116 patients with Sotos syndrome. *Hum Mutat* 2007;28:1098–1107.

65. Visser R, Matsumoto N. Genetics of Sotos syndrome. *Curr Opin Pediatr* 2003;15: 598–606.

66. Prior TW. Spinal muscular atrophy diagnostics. *J Child Neurol* 2007;22:952–956.

67. Lefebvre S, Burglen L, Reboullet S, et al. Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 1995;80:155–165.

68. Inoue K. PLP1-related inherited dysmyelinating disorders: Pelizaeus-Merzbacher disease and spastic paraplegia type 2. *Neurogenetics* 2005;6:1–16.

69. Sleegers K, Brouwers N, Gijselinck I, et al. APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. *Brain* 2006; 129(Pt 11):2977–2983.

70. Brouwers N, Sleegers K, Engelborghs S, et al. Genetic risk and transcriptional variability of amyloid precursor protein in Alzheimer's disease. *Brain* 2006;129(Pt 11):2984–2991.

71. Rovelet-Lecrux A, Hannequin D, Raux G, et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet* 2006;38:24–26.

72. Durand CM, Betancur C, Boeckers TM, et al. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet* 2007;39:25–27.

73. Moessner R, Marshall CR, Sutcliffe JS, et al. Contribution of SHANK3 mutations to autism spectrum disorder. *Am J Hum Genet* 2007;81:1289–1297.

74. Fantes J, Redeker B, Breen M, et al. Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. *Hum Mol Genet* 1995;4:415–422.

75. Robinson DO, Howarth RJ, Williamson KA, van Heyningen V, Beal SJ, Crolla JA. Genetic analysis of chromosome 11p13 and the PAX6 gene in a series of 125 cases referred with aniridia. *Am J Med Genet A* 2008;146:558–569.

76. Klopocki E, Ott CE, Benatar N, Ullmann R, Mundlos S, Lehmann K. A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome. *J Med Genet* 2008;45:370–375.

77. Fellermann K, Stange DE, Schaeffeler E, et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet* 2006;79:439–448.

78. Aitman TJ, Dong R, Vyse TJ, et al. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* 2006;439:851–855.

79. Fanciulli M, Norsworthy PJ, Petretto E, et al. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet* 2007;39:721–723.

80. Chartier-Harlin MC, Kachergus J, Roumier C, et al. Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* 2004;364:1167–1169.

81. Singleton AB, Farrer M, Johnson J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science* 2003;302:841.

82. Ibanez P, Bonnet AM, Debarges B, et al. Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet* 2004;364:1169–1171.

83. Burns JC, Shimizu C, Gonzalez E, et al. Genetic variations in the receptor-ligand pair CCR5 and CCL3L1 are important determinants of susceptibility to Kawasaki disease. *J Infect Dis* 2005;192:344–349.

84. Gonzalez E, Kulkarni H, Bolivar H, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 2005;307:1434–1440.

85. McKinney C, Merriman ME, Chapman PT, et al. Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. *Ann Rheum Dis* 2008;67:409–413.

86. Osborne LR, Mervis CB. Rearrangements of the Williams-Beuren syndrome locus: molecular basis and implications for speech and language development. *Expert Rev Mol Med* 2007;9:1–16.

87. Cusco I, Corominas R, Bayes M, et al. Copy number variation at the 7q11.23 segmental duplications is a susceptibility factor for the Williams-Beuren syndrome deletion. *Genome Res* 2008;18:683–694.

88. Somerville MJ, Mervis CB, Young EJ, et al. Severe expressive-language delay related to duplication of the Williams-Beuren locus. *N Engl J Med* 2005;353:1694–1701.

89. Lupski JR. Genome structural variation and sporadic disease traits. *Nat Genet* 2006;38:974–976.

90. Sharp AJ, Hansen S, Selzer RR, et al. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet* 2006;38:1038–1042.

91. Ballif BC, Hornor SA, Jenkins E, et al. Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. *Nat Genet* 2007;39:1071–1073.

92. Potocki L, Bi W, Treadwell-Deering D, et al. Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. *Am J Hum Genet* 2007;80:633–649.

93. Szatmari P, Paterson AD, Zwaigenbaum L, et al. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 2007;39:319–328.

94. Ullmann R, Turner G, Kirchhoff M, et al. Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum Mutat* 2007;28:674–682.

95. Schaefer GB, Mendelsohn NJ. Genetics evaluation for the etiologic diagnosis of autism spectrum disorders. *Genet Med* 2008;10:4–12.

96. Lachman HM, Pedrosa E, Petruolo OA, et al. Increase in GSK3beta gene copy number variation in bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet* 2007;144:259–265.

97. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008;320:539–543.

98. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 2008;40:880–885.

99. Hughes AE, Orr N, Esfandiary H, Diaz-Torres M, Goodship T, Chakravarthy U. A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration. *Nat Genet* 2006;38:1173–1177.

100. Klopocki E, Schulze H, Strauss G, et al. Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. *Am J Hum Genet* 2007;80:232–240.

101. Schonherr N, Meyer E, Roos A, Schmidt A, Wollmann HA, Eggermann T. The centromeric 11p15 imprinting centre is also involved in Silver-Russell syndrome. *J Med Genet* 2007;44:59–63.

102. Feuk L, Kalervo A, Lipsanen-Nyman M, et al. Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. *Am J Hum Genet* 2006;79:965–972.

103. Gervasini C, Castronovo P, Bentivegna A, et al. High frequency of mosaic CREBBP deletions in Rubinstein-Taybi syndrome patients and mapping of somatic and germ-line breakpoints. *Genomics* 2007;90:567–573.

104. Perry GH, Dominy NJ, Claw KG, et al. Diet and the evolution of human amylase gene copy number variation. *Nat Genet* 2007;39:1256–1260.

105. Schulze JJ, Lundmark J, Garle M, Skilving I, Ekstrom L, Rane A. Doping test results dependent on genotype of ugt2b17, the major enzyme for testosterone glucuronidation. *J Clin Endocrinol Metab* 2008;93:2500–2506.

106. Stefansson H, Rejescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature* Jul 30 2008 [Epub ahead of print].

107. Shlien A, Tabori U, Marshall CR, et al. Excessive genomic DNA copy number variation in the Li-Fraumeni cancer predisposition syndrome. *Proc Natl Acad Sci* 2008;105:11264–11269.