Genetics inMedicine | BRIEF REPORT

Open

Estimates of penetrance for recurrent pathogenic copy-number variations

Jill A. Rosenfeld, MS[1], Bradley P. Coe, PhD[2], Evan E. Eichler, PhD[2,3], Howard Cuckle, DPhil[4] and Lisa G. Shaffer, PhD[1,5]

Purpose: Although an increasing number of copy-number variations are being identified as susceptibility loci for a variety of pediatric diseases, the penetrance of these copy-number variations remains mostly unknown. This poses challenges for counseling, both for recurrence risks and prenatal diagnosis. We sought to provide empiric estimates for penetrance for some of these recurrent, disease-susceptibility loci.

Methods: We conducted a Bayesian analysis, based on the copy-number variation frequencies in control populations ($n = 22,246$) and in our database of >48,000 postnatal microarray-based comparative genomic hybridization samples. The background risk for congenital anomalies/developmental delay/intellectual disability was assumed to be ~5%. Copy-number variations studied were 1q21.1 proximal duplications, 1q21.1 distal deletions and duplications, 15q11.2 deletions, 16p13.11 deletions, 16p12.1 deletions, 16p11.2 proximal and distal deletions and duplications, 17q12 deletions and duplications, and 22q11.21 duplications.

Results: Estimates for the risk of an abnormal phenotype ranged from 10.4% for 15q11.2 deletions to 62.4% for distal 16p11.2 deletions.

Conclusion: This model can be used to provide more precise estimates for the chance of an abnormal phenotype for many copy-number variations encountered in the prenatal setting. By providing the penetrance, additional, critical information can be given to prospective parents in the genetic counseling session.

Genet Med 2013:15(6):478–481

Key Words: copy-number variation; genomic disorder; microarray; penetrance; prenatal diagnosis

INTRODUCTION

Over the past several years, our understanding of copy-number variation (CNV) within the human genome and its relation to disease has rapidly evolved. Molecular cytogenetic techniques such as microarray-based comparative genomic hybridization have identified disease-causing CNVs in a variety of disorders, ranging from pediatric disease (congenital anomalies, intellectual disability, epilepsy, and autism spectrum disorders) to adult-onset conditions such as schizophrenia. Of note, many CNVs were identified in multiple, variable disease cohorts, indicating that identical genetic changes could result in different phenotypes.[1,2] Furthermore, some of these CNVs were inherited from phenotypically normal parents.[1,2] Although the genetics community was already familiar with variable expressivity in the classic example of 22q11.21 deletions, traditional cytogenetics had taught us to use inheritance of a genetic change as a definitive factor for pathogenicity. Specifically, *de novo* aberrations are thought to be more deleterious, whereas inherited rearrangements (such as a marker chromosome) are considered more benign. However, for newly described CNVs like the distal 1q21.1 microdeletions/microduplications, despite variable phenotypes and inheritance from normal parents, enrichment of the CNVs among affected individuals in comparison with healthy controls implicated them as pathogenic.[3] As increasing numbers of cases and controls are studied for CNVs, we are discovering many additional examples of these "predisposing," or "susceptibility," loci.[1,2,4,5]

Microarray analysis is now recommended as a first-tier test for many pediatric neurodevelopmental disorders.[6,7] Postnatal identification of one of these susceptibility CNVs explains at least one part of the genetic etiology of the disorder in the individual, although additional factors, either genetic or environmental, are likely to ultimately influence the phenotypic expression of these loci.[1,8] Additional genetic factors, such as other CNVs, may be identified via microarray testing, but in many cases, the other influences on the phenotype remain unknown. This poses challenges to recurrence-risk counseling because subsequent children inheriting the CNV could have more or less severe, or no, phenotypic consequences, and specific testing is not available to inform such predictions. In addition, as the use of microarrays in prenatal settings increases, fetuses without a known family history of these CNVs will be identified as carriers. This can lead to counseling dilemmas and parental anxiety, especially in low-risk pregnancies, because the associated neurodevelopmental phenotypes cannot be ascertained prenatally and it is difficult to quantify the risk to the fetus. To

[1]Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, Washington, USA; [2]Department of Genome Sciences, University of Washington, Seattle, Washington, USA; [3]Howard Hughes Medical Institute, University of Washington, Seattle, Washington, USA; [4]Department of Obstetrics and Gynecology, Columbia University Medical Center, New York, New York, USA; [5]Current affiliation: Genetic Veterinary Sciences, Inc., Spokane, Washington, USA. Correspondence: Lisa G. Shaffer (lshaffer@pawprintgenetics.com)

Submitted 4 September 2012; accepted 7 November 2012; advance online publication 20 December 2012. doi:10.1038/gim.2012.164

aid in counseling for these CNVs, we calculated empiric estimates for penetrance on the basis of the CNV frequencies in our population of postnatal microarray-based comparative genomic hybridization samples and in control populations.

MATERIALS AND METHODS

We examined postnatal specimens received by our laboratory, mostly from the United States, for clinical microarray-based comparative genomic hybridization between March 2004 and April 2012. The analysis of indications for study among samples received in the first quarter of 2008 and of 2011 showed that 51–54% of individuals have developmental delay/intellectual disability and 10–11% have epilepsy, whereas cases with autism spectrum disorders have increased from 10% to 14%, those with congenital anomalies have increased from 16% to 23%, but those with dysmorphic features have decreased from 25% to 16%. Cases with unspecified indications for study have decreased from 7% to 5%. These are likely underestimates of actual phenotypes because not all phenotypic features are recorded on the test requisition form. The array platform used depended on the date of specimen receipt because array designs changed over time. Samples were tested on targeted, bacterial artificial chromosome–based arrays (SignatureChip versions 1–4; Signature Genomic Laboratories, Spokane, WA; n = 15,411), whole-genome, bacterial artificial chromosome–based arrays (SignatureChipWG versions 1–2; Signature Genomic Laboratories; n = 8,113), or whole-genome, oligonucleotide-based arrays (SignatureChipOS version 1; manufactured by Agilent Technologies, Santa Clara, CA; SignatureChipOS versions 2–3; manufactured by Roche NimbleGen, Madison, WI; all custom designed by Signature Genomic Laboratories; n = 25,113) according to previously described methods.[9–12] For the CNVs analyzed here, the targeted, bacterial artificial chromosome–based arrays only had coverage of 22q11.21 and proximal 1q21.1, whereas the whole-genome arrays had coverage of all studied CNVs. Frequencies for 15q11.2 deletions were calculated only for cases studied on oligonucleotide-based arrays because the CNV was initially interpreted as likely benign and, therefore, not captured in our database for the cases studied with bacterial artificial chromosome–based arrays. For determination of CNV frequencies, only those CNVs that are of the recurrent size, as determined within the limits of resolution of

Table 1 Penetrance estimates with case and control frequencies for recurrent CNVs

Region (gene within region)	Copy number	Coordinates (hg18)	Frequency, postnatal aCGH cases	Frequency, controls	*P* value (Fisher exact one-tailed test)	Frequency of *de novo* occurrence in cases	Penetrance estimate, % (95% CI)
Proximal 1q21.1 (*RBM8A*)	Duplication	chr1: 144.0–144.5 Mb	85/48,637 (0.17%)	10/22,246 (0.04%)	<<0.0001	0/13 (0%)	17.3 (10.8–27.4)
Distal 1q21.1 (*GJA5*)	Deletion	chr1: 145.0–146.35 Mb	97/33,226 (0.29%)	6/22,246 (0.03%)	<<0.0001	7/39 (17.9%)	36.9 (23.0–55.0)
Distal 1q21.1 (*GJA5*)	Duplication	chr1: 145.0–146.35 Mb	68/33,226 (0.20%)	6/22,246 (0.03%)	<<0.0001	5/30 (16.7%)	29.1 (16.9–46.8)
15q11.2 (*NIPA1*)	Deletion	chr15: 20.3–20.8 Mb	203/25,113 (0.81%)	84/22,246 (0.38%)	<<0.0001	0/27 (0%)	10.4 (8.45–12.7)
16p13.11 (*MYH11*)	Deletion	chr16: 14.9–16.4 Mb	50/33,226 (0.15%)	12/22,246 (0.05%)	<0.0005	5/23 (21.7%)	13.1 (7.91–21.3)
16p12.1 (*CDR2*)	Deletion	chr16: 21.85–22.4 Mb	62/33,226 (0.19%)	16/22,246 (0.07%)	<0.0002	1/28 (3.6%)	12.3 (7.91–18.8)
Distal 16p11.2 (*SH2B1*)	Deletion	chr16: 28.65–29.0 Mb	46/33,226 (0.14%)	1/22,246 (0.005%)	<<0.0001	7/21 (33.3%)	62.4 (26.8–94.4)
Distal 16p11.2 (*SH2B1*)	Duplication	chr16: 28.65–29.0 Mb	35/33,226 (0.11%)	10/22,246 (0.04%)	<0.01	1/8 (12.5%)	11.2 (6.26–19.8)
Proximal 16p11.2 (*TBX6*)	Deletion	chr16: 29.5–30.15 Mb	146/33,226 (0.44%)	6/22,246 (0.03%)	<<0.0001	33/47 (70.2%)[a]	46.8 (31.5–64.2)
Proximal 16p11.2 (*TBX6*)	Duplication	chr16: 29.5–30.15 Mb	93/33,226 (0.28%)	9/22,246 (0.04%)	<<0.0001	7/30 (23.3%)	27.2 (17.4–40.7)
17q12 (*HNF1B*)	Deletion	chr17: 31.8–33.3 Mb	29/33,226 (0.09%)	2/22,246 (0.01%)	<0.0001	5/9 (55.6%)	34.4 (13.7–70.0)
17q12 (*HNF1B*)	Duplication	chr17: 31.8–33.3 Mb	37/33,226 (0.11%)	5/22,246 (0.02%)	<0.0001	2/9 (22.2%)	21.1 (10.6–39.5)
22q11.21 (*TBX1*)	Duplication	chr22: 17.2–19.9 Mb	136/48,637 (0.28%)	12/22,246 (0.05%)	<<0.0001	12/47 (25.5%)	21.9 (14.7–31.8)

aCGH, microarray-based comparative genomic hybridization; CI, confidence interval; CNV, copy-number variation; <<, much less than.

[a]Deletions of the proximal 16p11.2 region showed a maternal transmission bias (14/68 mothers identified to be carriers vs. 0/38 fathers; two-tailed P = 0.0018, Fisher exact test); no parental transmission bias was detected for any other CNV.

the array used, are counted; any CNVs that do not include the entire region or extend into surrounding regions are excluded. However, individuals who harbor CNVs at other loci are included in the CNV frequencies.

Control specimens included samples from 8,329 previously described adult controls profiled on Illumina single-nucleotide polymorphism arrays.[4] Additional control specimens were collected from the Atherosclerosis Risk in Communities study (dbGaP accession phs000090.v1.p1) and the Wellcome Trust Case Control Consortium (WTCCC2 1958 British birth cohort). Both the Atherosclerosis Risk in Communities and WTCCC2 data were derived from Affymetrix SNP6.0 (Affymetrix, Santa Clara, CA) array profiles and processed using Affymetrix Genotyping Console 4.1 with hg18 chromosome annotations. Samples were filtered using the default contrast quality control parameters, and segmentation was also performed using default settings. Additional filtering was applied to remove cases with excessive CNV counts, and a threshold of >72 CNVs per case was established using an outlier detection method for skewed data.[13] After quality control filtering, the final control set consisted of 11,305 controls from the Atherosclerosis Risk in Communities study, 2,612 controls from the WTCCC2 58C cohort, and 8,329 previously published controls.

CNVs chosen for study were recurrent, identified in controls, and significantly enriched in cases (**Table 1**). A Bayesian analysis was performed, based on the method used by Vassos et al.[14] for the calculation of the penetrance of CNVs associated with schizophrenia, although we differed from their methods by using the observed population CNV frequencies directly in the following calculation. In brief, penetrance was calculated as:

$$P(D \mid G) = \frac{P(G \mid D)\,P(D)}{P(G \mid D)\,P(D) + P(G \mid \overline{D})\,P(\overline{D})},$$

where D = disease, G = genotype (i.e., the presence of the CNV), and $\overline{D}$ = absence of disease. Because we intended to calculate the probability of any abnormal pediatric phenotype when the CNV was identified on prenatal microarray testing, we defined the frequency of disease (P(D)) to be 5.12%, which is derived from the work of Baird et al.,[15] who estimated the population frequency of diseases with an important genetic component among individuals younger than 25 to be 53 in 1,000. We subtracted from this 1.8 per 1,000, the frequency of chromosomal disorders, because these will have been ruled out in most cases through karyotyping. The 95% confidence intervals (CIs) for our penetrance estimates were calculated using the binomial CI for case and control counts calculated by the Clooper–Pearson exact tail area method. Using penetrance samples from both case and control distributions, we first calculated maximal and minimal likely counts in which the probability of generating a more extreme count for either cases or controls is 15.8% ($\sqrt{0.025}$); thus the probability of sampling a more extreme combination of case and control counts is 0.158 × 0.158 = 0.025 per tail. In the case of observed proportions of 0 and 1, the upper and lower binomial confidence bounds are fixed at 0 and 1, respectively. The lower penetrance bound is thus defined by substituting the maximal likely control count and minimal likely case count, whereas the upper bound is defined by substituting the minimal likely control count and maximal likely case count. Because this methodology is based on two one-tailed analyses, the actual CI will approach 97.5% as case and control counts approach their respective minima and maxima.

RESULTS

Penetrance estimates for these CNVs range from 10.4% (95% CI, 8.45–12.7%) for 15q11.2 deletions, which only represents about a twofold increase in risk over the background population risk, to 62.4% (95% CI, 26.8–94.4%) for distal 16p11.2 deletions (**Table 1**). The lower penetrance figures are seen with CNVs that show less marked differences in frequencies between cases and controls, including distal 16p11.2 duplications, 16p12.1 deletions, and 16p13.11 deletions. The CNVs with a larger difference between cases and controls, including 16p11.2 proximal deletions and 1q21.1 distal deletions, have higher penetrance rates. In addition, higher penetrance is seen with CNVs that have higher *de novo* frequencies (**Table 1**; P = 0.0029, Spearman correlation). For some of these CNVs that are still rare in controls, such as the distal 16p11.2 deletions, screening a larger control group would help to ensure a more precise estimate of penetrance. For still other CNVs that were not found in controls and therefore not part of this study (as penetrance would be estimated at 100%), such as the BP4-BP5 15q13.2q13.3 microdeletion, the CNVs may be inherited from apparently healthy parents in some cases,[16] so penetrance is not complete, yet our data could not be used to estimate a value. Although similar penetrance estimates based on a subset of these data were recently part of a corrigendum to the article by Cooper et al.,[4] our increased population sizes and inclusion of only postnatal cases give increased power to the estimates in this current study.

DISCUSSION

By using a patient population with a variety of phenotypes, we are able to provide penetrance estimates for our group of disease susceptibility CNVs for a range of abnormal pediatric phenotypes. This is both a strength and a weakness, because our estimates apply simply to the presence or absence of any abnormal pediatric phenotype without providing information about expressivity. It is well established that these CNVs lead to a spectrum of phenotypes, and predictions about severity (expressivity) are not possible on the basis of the data presented here. Some CNVs may have an association with a specific phenotype, and different calculations could provide separate estimates for a phenotype of concern. For example, Bayesian analysis for proximal 16p11.2 deletions and autism spectrum disorders, with P(G|D) being 0.5%[17] and P(D) being 1/110,[18] yields a penetrance estimate of 14.5% for an autism spectrum disorder phenotype in the presence of a proximal 16p11.2 deletion. Notably, this is lower than our penetrance

for any abnormal phenotype, which supports the use of our estimates to include specific phenotypes among a number of other possible manifestations. In addition, our estimates do not include risks for adult-onset or other conditions, such as obesity, that alone would likely not lead to an individual to be referred for clinical microarray-based comparative genomic hybridization testing. Although subclinical phenotypes may not be of concern, adult-onset conditions might be, and penetrance has been estimated for some of these CNVs and, for example, schizophrenia.[14] Finally, it should be noted that we could have underestimated penetrance because the controls studied did not have in-depth phenotyping and may include mildly affected individuals. Also, these estimates are based on populations that are assumed to be mostly Caucasian, so it is also unclear whether estimates would vary in other ethnic groups.

The calculation model and estimates provided here will hopefully be a useful tool in prenatal genetic counseling, providing one more piece of information to inform prospective parents on the risks associated with carrying a specific CNV. Although counseling should still include information about the range of possible phenotypic outcomes, penetrance estimates can help to put the degree of risk into perspective; for example, counseling about a 15q11.2 deletion could be relatively reassuring with a ~90% likelihood of a normal phenotype, as compared with an ~50% chance of a normal outcome with a 16p11.2 proximal deletion. The ultimate phenotype of the child is probably affected by his/her genetic background and other environmental factors, the vast majority of which are unknown and therefore cannot be tested. Even when microarray testing identifies an additional CNV, it is not possible to predict how the CNVs may interact. Although it is still possible that prenatal microarray testing will identify a novel CNV of unclear clinical significance, in which case data do not exist to apply this model, large population studies have estimated that ~1/200 low-risk pregnancies carry a clinically significant CNV, many of which are at these recurrent loci;[19,20] and so these penetrance estimates are likely applicable for many abnormal prenatal microarray results.

ACKNOWLEDGMENTS

We thank Greg Cooper for his assistance with statistical calculations. E.E.E. is an investigator of the Howard Hughes Medical Institute.

DISCLOSURE

J.A.R. is an employee of Signature Genomic Laboratories, a subsidiary of PerkinElmer, Inc. E.E.E. is on the scientific advisory boards for Pacific Biosciences, Inc., SynapDx Corp., and DNAnexus, Inc. The other authors declare no conflict of interest.

REFERENCES

1. Girirajan S, Eichler EE. Phenotypic variability and genetic susceptibility to genomic disorders. *Hum Mol Genet* 2010;19(R2):R176–R187.
2. Girirajan S, Campbell CD, Eichler EE. Human copy number variation and complex genetic disease. *Annu Rev Genet* 2011;45:203–226.
3. Mefford HC, Sharp AJ, Baker C, et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 2008;359:1685–1699.
4. Cooper GM, Coe BP, Girirajan S, et al. A copy number variation morbidity map of developmental delay [forthcoming corrigendum in Nat Genet]. *Nat Genet* 2011;43:838–846.
5. Kaminsky EB, Kaul V, Paschall J, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet Med* 2011;13:777–784.
6. Miller DT, Adam MP, Aradhya S, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet* 2010;86:749–764.
7. Shen Y, Dies KA, Holm IA, et al. Clinical genetic testing for patients with autism spectrum disorders. *Pediatrics* 2010;125:e727–e735.
8. Girirajan S, Rosenfeld JA, Coe BP, et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N Engl J Med* 2012;367:1321–1331.
9. Ballif BC, Theisen A, Coppinger J, et al. Expanding the clinical phenotype of the 3q29 microdeletion syndrome and characterization of the reciprocal microduplication. *Mol Cytogenet* 2008;1:8.
10. Ballif BC, Theisen A, McDonald-McGinn DM, et al. Identification of a previously unrecognized microdeletion syndrome of 16q11.2q12.2. Clin Genet 2008;74:469–475.
11. Duker AL, Ballif BC, Bawle EV, et al. Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *Eur J Hum Genet* 2010;18:1196–1201.
12. Bejjani BA, Saleki R, Ballif BC, et al. Use of targeted array-based CGH for the clinical diagnosis of chromosomal imbalance: is less more? *Am J Med Genet* A 2005;134:259–267.
13. Hubert M, Van der Veeken S. Outlier detection for skewed data. *J Chemometr* 2008;22:235–246.
14. Vassos E, Collier DA, Holden S, et al. Penetrance for copy number variants associated with schizophrenia. *Hum Mol Genet* 2010;19:3477–3481.
15. Baird PA, Anderson TW, Newcombe HB, Lowry RB. Genetic disorders in children and young adults: a population study. Am J Hum Genet 1988;42:677–693.
16. van Bon BW, Mefford HC, Menten B, et al. Further delineation of the 15q13 microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. *J Med Genet* 2009;46:511–523.
17. Walsh KM, Bracken MB. Copy number variation in the dosage-sensitive 16p11.2 interval accounts for only a small proportion of autism incidence: a systematic review and meta-analysis. *Genet Med* 2011;13:377–384.
18. Centers for Disease Control and Prevention. Prevalence of autism spectrum disorders—Autism and Developmental Disabilities Monitoring Network, United States, 2006. Surveillance Summaries. *Morb Mortal Wkly Rep* 2009;58:1–20.
19. Lee CN, Lin SY, Lin CH, Shih JC, Lin TH, Su YN. Clinical utility of array comparative genomic hybridisation for prenatal diagnosis: a cohort study of 3171 pregnancies. *BJOG* 2012;119:614–625.
20. Shaffer LG, Dabell MP, Fisher AJ, et al. Experience with microarray-based comparative genomic hybridization for prenatal diagnosis in over 5000 pregnancies. *Prenat Diagn* 2012;32:976–985.