Molecular Syndromology

Mol Syndromol 2013;4:27–31
DOI: 10.1159/000345406

Published online: November 20, 2012

Exome Sequencing and High-Density Microarray Testing in Monozygotic Twin Pairs Discordant for Features of VACTERL Association

B.D. Solomon[a] D.E. Pineda-Alvarez[a] D.W. Hadley[a] N.F. Hansen[b, c] A. Kamat[d] F.X. Donovan[d] S.C. Chandrasekharappa[d] S.-K. Hong[a] E. Roessler[a] J.C. Mullikin[b, c]
NISC Comparative Sequencing Program

[a]Medical Genetics Branch, [b]NIH Intramural Sequencing Center, [c]Genome Technology Branch, and [d]Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Md., USA

Key Words
Exome sequencing · Twin studies · VACTERL association

Abstract
Exome sequencing offers an efficient and affordable method to interrogate genetic factors involved in human disease. Performing exome sequencing of monozygotic twins discordant for VACTERL (Vertebral anomalies, Anal atresia, Cardiac malformations, Tracheo-Esophageal fistula, Renal anomalies, and Limb abnormalities) association-type congenital malformations was hypothesized to potentially reveal discordant variants that could demonstrate disease cause(s). After demonstrating monozygosity, we applied high-density microarrays and exome sequencing to 2 twin pairs in which 1 twin had features of VACTERL association while the other was phenotypically normal (demonstrated through comprehensive clinical and radiological evaluation). No obvious discordant genotypic results were found that would explain phenotypic discordance. We conclude that VACTERL association is a complex disease, and while performing microarray analysis and exome sequencing on phenotypically discordant monozygotic twins may hypothetically reveal genetic causes of disorders, challenges remain in applying these methods in this circumstance.

Exome sequencing (sometimes, perhaps redundantly referred to as 'whole exome sequencing') is a powerful tool for investigating the genetic underpinnings of human disease [Ng et al., 2010]. Exome sequencing refers to the use of 'next-generation' technology to sequence the exome or all the coding regions (exons) of the genome, which makes up approximately 1–2% of the human genome (depending on the exact regions included), as well as some surrounding regions. As Mendelian pathogenic mutations are frequently exonic, exome sequencing allows an efficient method to examine many coding regions simultaneously, and has rapidly proven to be an important tool in genetic research [Teer et al., 2010; Biesecker et al., 2011].

Congenital malformations are more common in twins. A satisfying explanation for this phenomenon is lacking, but likely involves genetic, epigenetic, and environmental factors [Zwijnenburg, 2010]. The presence of molecularly determined point mutations accounting for disease in 1 of monozygotic twins have been described in several conditions including Darier's disease, van der Woude syndrome, and otopalatodigital syndrome [Sakuntabhai et al., 1999; Kondo et al., 2002; Robertson et al., 2006]. Relevant to the current study, however, it is important to note that the concordance rate in monozygotic twins with VACTERL association is very similar to that reported in dizygotic twins [Bartels et al., 2012].

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2012 S. Karger AG, Basel
1661–8769/13/0042–0027$38.00/0

Accessible online at:
www.karger.com/msy

Benjamin D. Solomon
NIH, MSC 3717
Building 35, Room 1B-207
Bethesda, MD 20892 (USA)
E-Mail solomonb@mail.nih.gov

VACTERL association is typically defined as involving at least 3 of the following congenital malformations without evidence for an alternative diagnosis: Vertebral anomalies, Anal atresia, Cardiac malformations, Tracheo-Esophageal fistula, Renal anomalies, and Limb abnormalities [Quan and Smith, 1973; reviewed in Solomon, 2011]. Though the condition is not exceedingly rare, with an incidence estimated at approximately 1 in 10,000–40,000 live-born infants, relatively little is known about causes [reviewed in Solomon, 2011]. This dearth of causative knowledge is likely related to issues including typical sporadic occurrence, clinical and molecular heterogeneity, and a relatively large number of distinct but clinically similar conditions [Solomon et al., 2010; reviewed in Solomon, 2011].

In order to interrogate potential genetic discordance in twins with VACTERL association, we applied high-density microarray analysis and exome sequencing to 2 monozygotic twin pairs in which only 1 twin had features of VACTERL association.

Patients and Methods

Patients and Initial Analysis

Patients participated in our National Human Genome Research Institute institutional review board-approved protocol on VACTERL association, with appropriate consent obtained from all participants, including for genomic sequencing, such as exome sequencing. For the purpose of this specific study, inclusion criteria involved at least 3 component malformations of VACTERL association without clinical or laboratory-based evidence of an alternative diagnosis.

For both pairs of monozygotic twins, both the affected and unaffected child underwent a detailed physical examination by a clinical geneticist very familiar with VACTERL association, as well as X-rays of the entire spine and limbs, abdominal ultrasound and echocardiogram.

As part of study participation, patients gave peripheral blood samples from which DNA was extracted via phenol:chloroform extraction. Peripheral blood samples were used to create lymphoblastoid cell lines, but all analyses described here were performed from directly extracted DNA from whole blood rather than from cell lines. Parental microarray/exome studies were not performed (the inheritance of individual variants of interest were studied on a variant-by-variant basis).

Microarray Performance and Analysis

Prior to exome sequencing, microarray analysis was performed to confirm monozygosity and to detect clinically significant pathogenic copy number variants using the Illumina Omni1-Quad SNP array (Illumina Inc., San Diego, Calif., USA), which contains over 1 million SNP loci, with 300 ng of DNA (4 μl of 75 ng/μl DNA) according to the Illumina 'infinium assay' protocol [Gunderson et al., 2005]. We collected data using the BeadArray scanner and we visualized data with the GenomeStudio (v2009.2, www.Illumina.com) genotyping module, using human genome build 36.1 (NCBI36/hg18). The call rates for all the DNA samples were >99%. CNVs were detected using PennCNV software and were initially filtered to annotate regions with at least 3 contiguous SNPs reflecting the same abnormality with a confidence interval of at least 50 [Wang et al., 2007]. Detected regions of genomic imbalances were compared with known CNVs in control populations via the Database of Genomic Variants [Zhang et al., 2006].

Exome Sequencing

We performed solution hybridization exome capture with the SureSelect Human All Exon 38-Mb (twin pair 1) and 50-Mb (twin pair 2) systems (Agilent Technologies, Santa Clara, Calif., USA), using biotinylated RNA baits to hybridize to sequences corresponding to exons [Gnirke et al., 2009]. Manufacturer's protocol version 1.0 compatible with Illumina paired-end sequencing were used, except that we measured DNA fragment size and quality using a 2% agarose gel with Sybr Gold stain rather than an Agilent Bioanalyzer. Manufacturer's specifications for the 38-Mb kit report that the capture regions total approximately 38 Mb, accounting for 1.22% of the human genome, corresponding to the Consensus Conserved Domain Sequences database and over 1,000 non-coding RNAs. The 50-Mb kit additionally includes Gencode Project defined exons (http://www.sanger.ac.uk/resources/databases/encode/). Targeted regions included the exons of 18,113 genes of the Consensus Conserved Domain Sequences database, totaling 37,640,396 bases (All Exon 38-Mb). The All Exon 50-Mb kit includes all regions in the All Exon 38-Mb kit, and adds exons of additional genes, miRNAs, and non-coding RNA genes, totaling 30,241 genomic features and a total of 51,646,629 targeted bases. Flowcell preparation and sequencing were performed according to the protocol for the GAIIx sequencer (Illumina) [Bentley et al., 2008] using 75- or 100-bp paired-end reads to generate sufficient data such that at least 85% of the targeted bases were accurately genotyped (see next paragraph). Image analysis and base calling were performed on all data lanes using Illumina Genome Analyzer Pipeline software (GAPipeline versions 1.4.0 or greater) with default parameters.

Variant Analysis

Small sequence variants were determined using custom variant-calling software. In brief, we aligned reads to human genome build 36.1 (NCBI36/hg18) for analysis using 'efficient large-scale alignment of nucleotide databases' (ELAND, Illumina). Reads were grouped aligning uniquely into genomic sequence intervals of approximately 100 kb. Non-aligning reads were binned with their paired-end mates. Reads in each bin underwent a Smith-Waterman-based local alignment algorithm, cross_match using the parameters minscore 21 and masklevel 0 to their respective 100-kb genomic sequence (http://www.phrap.org). A total of 6 Gb per individual of high-confidence mappable sequence data were generated in autosomal targeted regions. Genotypes were called at all positions using high-quality sequence bases (Phred-like Q20 or greater) with a Bayesian algorithm (most probable genotype; goal read-depth is an average of at least 85% in targeted regions) [Teer et al., 2010]. Genotypes were required to have a most probable genotype score ≥10 and a score/coverage ratio ≥0.5. These are criteria which have been shown to yield >99.89% concordance with SNP microarray data. Targeted regions included the exons of 17,134 genes, with a total of 37,640,396 bases in the human genome (All Exon 38-Mb: twin pair 1) or the exons of 30,241 genes and

Table 1. Summary of the clinical data

Patient	V	A	C	TE	R	L	Other features
1	no anomalies (distal kinking on spinal MRI)	none	VSD	TEF with distal EA	abnormally echogenic kidneys bilaterally, with atrophic right kidney	none	none
2	multiple dysplastic supernumerary lumbar vertebral bodies, left cervical rib, 13 R ribs	anal atresia	none (history of PDA and PFO)[a]	TEF with distal EA	no frank anomalies; grade I VUR and chordee	none, though mild UE asymmetry	only consistent with prematurity

Twin pair 1 has additionally been described in Solomon et al. [2011]. EA = Esophageal atresia; PDA = patent ductus arteriosus; PFO = patent foramen ovale; TEF = tracheo-esophageal fistula; UE = upper extremity; VSD = ventricular septal defect; VUR = vesicoureteral reflux.

[a] History of PDA and PFO is supposed to be most likely related to prematurity rather than reflective of a true congenital cardiovascular malformation.

total 51,646,629 bases (All Exon 50-Mb: twin pair 2). The annotation of coding SNVs (single nucleotide variants) was based on the 'known genes' dataset of the University of California, Santa Cruz. We classified SNVs and short deletion-insertion variants with a custom suite of annotation scripts (PIANNO) as those in intronic, UTR, or within coding regions. The software categorized variants as belonging to 1 of the following subsets: UTR3, UTR5, downstream variants, frameshift (deletion, insertion, or substitution), intergenic, intronic, ncRNA (UTR3, UTR5, exonic, intronic, or splicing), non-frameshift (deletion, insertion, or substitution), non-synonymous SNV, splicing, stop-gain SNV, stop-loss SNV, synonymous SNV, or upstream. Resulting variants were then analyzed and filtered using VarSifter software (available at: http://research.nhgri.nih.gov/software/VarSifter/) [Teer et al., 2012].

Results

Clinical Features

See table 1 for clinical characteristics of the studied individuals. Both families were seen in person at the NIH Clinical Center. There was no clinical, radiological, or laboratory-based evidence for an alternative diagnosis. Neither unaffected twin demonstrated (via detailed physical and radiological examination) any evidence of anomalies observed in the affected twin, or anomalies seen in VACTERL association more generally, or any other congenital malformation.

Genetic/Genomic Analysis

Monozygosity was confirmed in the twin pairs, first by high-density microarray analysis, then as part of exome sequencing. In brief, neither microarray analysis nor exome sequencing revealed an obvious discordant genetic anomaly (CNV or exonic mutation) that would readily explain the presence of congenital anomalies in 1 child in the twin pair (see tables 2 and 3 for details). Additionally, no copy number variants or exonic mutations were identified in or affecting known disease-associated loci that would explain the congenital anomalies according to a model taking into account the possibility of incomplete penetrance. However, further studies are ongoing related to several genetic variants of high interest (not located in genes previously shown to be associated with human disease) that were found in both twins and which may act as susceptibility factors in concert with putative interacting genetic and/or environmental factors.

Discussion

In this study, neither microarray analysis nor exome sequencing of 2 pairs of monozygotic but phenotypically discordant twins revealed a rapid explanation (based on genotypic discordance) explaining why only 1 member of each twin pair was affected with features of VACTERL association. There are multiple possible explanations. First, if there is in fact a genetic explanation for the congenital malformations in these patients, it may occur in a region not interrogated by current methods of CNV analysis or exome sequencing, such as a regulatory factor that affects gene expression. Second, on a related note, even if a coding-region CNV or mutation were present in only the affected individuals, these methods do not result in complete coverage of all coding regions, even of regions tested. Eventually, similar applications of genome testing may be

Table 2. Summary of microarray-detected CNVs in the 2 twin pairs, filtered for the presence of at least 3 contiguous SNPs with the same abnormality [Wang et al., 2007]

Twin pair	Total CNVs[a]	CNVs not including known genes	CNVs containing only intronic elements	CNVs containing coding regions of genes[b] (overlapping with known CNVs of the same type[c])	CNVs containing coding regions of genes, and which are not known CNVs	Discordant CNVs
1	290	169	83	36	2	0
2	377	207	98	55	17	0

Detected CNVs were compared with control populations described in the Database of Genomic Variants [Zhang et al., 2006]. [a] When filtered for ≥10 contiguous SNPs with confidence interval of ≥50, twin pair 1 had 15 such CNVs, while twin pair 2 had 16 such CNVs. [b] These CNVs may additionally contain intronic elements. [c] The 'same type' refers to deletions versus duplications.

Table 3. Summary of exome-based sequence variants in the 2 twin pairs

Twin pair	Detected variants[a]	Variants not in dbSNP	Discordant variants	Discordant coding region variants predicted to potentially result in pathogenicity[b]	Predicted pathogenic discordant variants in affected twin only, with data above threshold[c]	Discordant variants confirmed by Sanger sequencing
1	79,525	19,976	1,070	99	1	0
2	125,630	34,423	1,787	124	0	0

The difference in the number of variants in the 2 different twin pairs largely reflects different methods of exome sequencing (SureSelect Human All Exon 38-Mb vs. 50-Mb kits). [a] Differences in the number of detected variants reflect greater coverage in the second twin pair analyzed. [b] Includes deletion-insertion variants, non-synonymous variants, splice-site variants, and nonsense variants, but excludes variants in 3′UTR, 5′UTR, intron variants, noncoding variants, as well as synonymous variants and variants in repeat regions. [c] Our threshold score was set such that we required a most probable genotype score ≥10 and a score/coverage ratio ≥0.5 (for details see Patients and Methods under the Variant Analysis section). The false-positive discordant variant found in twin pair 1, which was shown to be an artifact by Sanger sequencing, was later found as an artifact in multiple samples sequenced at the same facility [Biesecker et al., 2009].

used to test these hypothetical explanations that there is a discordant mutation not detected by the applied CNV or exome methodology. Third, we conducted microarray studies and exome sequencing on DNA extracted from peripheral blood, and similar testing based on other tissue types may yield more success, as has recently been demonstrated in Proteus syndrome [Lindhurst et al., 2011]. Finally, on a level more specific to VACTERL association, it is likely that multiple interacting genetic and environmental factors are necessary to produce the appropriate phenotype. In fact, specific evidence pointing to this distinct possibility has recently been described [Bartels et al., 2012]. The occurrence of a twin pregnancy may act on susceptible alleles to result in congenital malformations. Parental studies were not performed in these twin pairs. However, future molecular studies including more family members, such as the parents, may help filter likely susceptibility alleles (e.g. if the variants are not present in the parents). Additional, other testing modalities, such as methylation analysis, in conjunction with CNV or exome-based results, may shed more light on disease pathogenesis. Finally, it is possible that the causes of these congenital malformations may not be directly gene-related, and may involve a primary, currently unidentified environmental factor.

Acknowledgements

This research was supported by the Division of Intramural Research, National Human Genome Research Institute, National Institutes of Health and Human Services, United States of America. B.D.S. would like to thank Maximilian Muenke for his support and mentorship. The authors are extremely grateful to the patients and families who participated in our study, as well as for the superb collaborative efforts by the extremely helpful team at the NIH Intramural Sequencing Center.

References

Bartels E, Schulz AC, Mora NW, Pineda-Alvarez DE, Wijers CH, et al: VATER/VACTERL association: identification of seven new twin pairs, a systematic review of the literature, and a classical twin analysis. Clin Dysmorphol 21:191–195 (2012).

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al: Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–59 (2008).

Biesecker LG, Mullikin JC, Facio FM, Turner C, Cherukuri PF, et al: The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine. Genome Res 19: 1665–1674 (2009).

Biesecker LG, Shianna KV, Mullikin JC: Exome sequencing: the expert view. Genome Biol 12:128 (2011).

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, et al: Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 27:182–189 (2009).

Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS: A genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet 37:549–554 (2005).

Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, et al: Mutations in *IRF6* cause Van der Woude and popliteal pterygium syndromes. Nat Genet 32:285–289 (2002).

Lindhurst MJ, Sapp JC, Teer JK, Johnston JJ, Finn EM, et al: A mosaic activating mutation in *AKT1* associated with the Proteus syndrome. N Engl J Med 365:611–619 (2011).

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, et al: Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42:30–35 (2010).

Quan L, Smith DW: The VATER association. Vertebral defects, anal atresia, T-E fistula with esophageal atresia, radial and renal dysplasia: a spectrum of associated defects. J Pediatr 82:104–107 (1973).

Robertson SP, Thompson S, Morgan T, Holder-Espinasse M, Martinot-Duquenoy V, et al: Postzygotic mutation and germline mosaicism in the otopalatodigital syndrome spectrum disorders. Eur J Hum Genet 14:549–554 (2006).

Sakuntabhai A, Ruiz-Perez V, Carter S, Jacobsen N, Burge S, et al: Mutations in *ATP2A2*, encoding a Ca2+ pump, cause Darier disease. Nat Genet 21:271–277 (1999).

Solomon BD: VACTERL/VATER Association. Orphanet J Rare Dis 6:56 (2011).

Solomon BD, Pineda-Alvarez DE, Raam MS, Bous SM, Keaton AA, et al: Analysis of component findings in 79 patients diagnosed with VACTERL association. Am J Med Genet A 152A:2236–2244 (2010).

Solomon BD, Pineda-Alvarez DE, Hadley DW, NISC Comparative Sequencing Program, Teer JK, et al: Personalized genomic medicine: lessons from the exome. Mol Genet Metab 104:189–191 (2011).

Teer JK, Bonnycastle LL, Chines PS, Hansen NF, Aoyama N, et al: Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. Genome Res 20:1420–1431 (2010).

Teer JK, Green ED, Mullikin JC, Biesecker LG: VarSifter: visualizing and analyzing exome-scale sequence variation data on a desktop computer. Bioinformatics 28:599–600 (2012).

Wang K, Li M, Hadley D, Liu R, Glessner J, et al: PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 17: 1665–1674 (2007).

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 115:205–214 (2006).

Zwijnenburg PJ, Meijers-Heijboer H, Boomsma DI: Identical but not the same: the value of discordant monozygotic twins in genetic research. Am J Med Genet B Neuropsychiatr Genet 153B:1134–1149 (2010).