HUMAN MUTATION 28(12), 1225–1235, 2007

METHODS

Visualization of Uniparental Inheritance, Mendelian Inconsistencies, Deletions, and Parent of Origin Effects in Single Nucleotide Polymorphism Trio Data With SNPtrio

Jason C. Ting,[1] Elisha D.O. Roberson,[1,2] Nathaniel D. Miller,[1,3] Alana Lysholm-Bernacchi,[4] Dietrich A. Stephan,[4] George T. Capone,[1] Ingo Ruczinski,[5] George H. Thomas,[6] and Jonathan Pevsner[1–3]*

[1]Department of Neurology, Kennedy Krieger Institute, Baltimore, Maryland; [2]Program in Human Genetics, Johns Hopkins School of Medicine, Baltimore, Maryland; [3]Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, Maryland; [4]Neurogenomics Division, Translational Genomics Research Institute, Phoenix, Arizona; [5]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; [6]Department of Genetics, Kennedy Krieger Institute, Baltimore, Maryland

Communicated by Anthony J. Brookes

A variety of alterations occur in chromosomal DNA, many of which can be detected using high density single nucleotide polymorphism (SNP) microarrays. These include deletions and duplications (assessed by observing changes in copy number) and regions of homozygosity. The analysis of SNP data from trios can provide an additional category of information about the nature and origin of inheritance patterns, including uniparental disomy (UPD), loss of transmitted allele (LTA), and nonparental relationship. The main purpose of SNPtrio is to locate regions of uniparental inheritance (UPI) and Mendelian inconsistency (MI), identify the type (paternal vs. maternal, iso- vs. hetero-), and assess the associated statistical probability of occurrence by chance. SNPtrio's schema permits the identification of hemizygous or homozygous deletions as well as UPD. We validated the performance of SNPtrio on three platforms (Affymetrix 10 K and 100 K arrays and Illumina 550 K arrays) using SNP data obtained from DNA samples of patients known to have UPD and diagnosed with Prader-Willi syndrome, Angelman syndrome, Beckwith-Wiedemann syndrome, pseudohypoparathyroidism, and a complex chromosome 2 abnormality. We further validated SNPtrio using DNA from patients previously shown to have microdeletions that were verified by fluorescence in situ hybridization (FISH). SNPtrio successfully identified previously known UPD and deletion regions, and generated associated probability values. SNPtrio analysis of trisomy 21 (Down syndrome) cases and their parents permitted identification of the parent of origin of the extra chromosomal copy. SNPtrio is freely accessible at http://pevsnerlab.kennedykrieger.org/SNPtrio.htm (Last accessed: 20 June 2007). Hum Mutat 28(12), 1225–1235, 2007. **© 2007 Wiley-Liss, Inc.**

KEY WORDS: uniparental disomy; isodisomy; heterodisomy; trio; Mendelian inconsistency; genotype calls; web tool; visualization; parent of origin; autozygosity; trisomy

INTRODUCTION

Chromosomal DNA is characterized in part by variations between individuals at the level of entire chromosomes (e.g., aneuploidy), segmental changes (including deletions, duplications, inversions, and rearrangements such as translocations), and changes to small genomic regions (e.g., deletions, insertions, and single nucleotide polymorphisms [SNPs]). SNP microarray technology permits the estimation of genotype and copy number for up to hundreds of thousands of SNPs in genomic DNA. Since 2004 a variety of studies using SNP arrays, array comparative genome hybridization (aCGH), and other related technologies have been used to detect genomic variations including, in particular, deletions and duplications [Conrad et al., 2006; Hinds et al., 2006; Iafrate et al., 2004; McCarroll et al., 2006; Newman et al., 2006; Redon et al., 2006; Sebat et al., 2004; Sharp et al., 2006; Tuzun et al., 2005; Wong et al., 2007]. These studies [reviewed in Eichler, 2006; Freeman et al., 2006] have established the presence of hundreds of copy number variants (CNVs) in apparently normal individuals occurring in loci that span as much as 12% of the human genome.

The Supplementary Material referred to in this article can be accessed at www.interscience.wiley.com/jpages/1059–7794/suppmat.

Received 2 February 2007; accepted revised manuscript 18 May 2007.

*Correspondence to: Dr. Jonathan Pevsner, Dept. of Neurology, Kennedy Krieger Institute, 707 North Broadway, Baltimore, MD 21205. E-mail: pevsner@kennedykrieger.org

Grant sponsor: Taishoff Foundation; Grant sponsor: National Institutes of Health; Grant number: R01 HD046598; Grant sponsor: Mental Retardation and Developmental Disabilities Research Center (MRDDRC); Grant number: HD24061.

DOI 10.1002/humu.20583

Published online 27 July 2007 in Wiley InterScience (www.interscience.wiley.com).

SNP array data provide information on both chromosomal copy number (based on intensity measurements) and genotypes (i.e., AA and BB homozygous biallelic calls, and AB heterozygous biallelic calls). Because current software does not distinguish between homozygous and hemizygous states, regions containing homozygous calls and a reduced copy number most likely reflect the occurrence of a hemizygous deletion. The availability of SNP data obtained from trios is further useful to define whether a particular inheritance pattern, such as a deletion, has occurred in a child de novo or by descent. In this work we describe SNPtrio, a software tool that is useful to identify assorted variations in chromosomal inheritance in SNP data from trios.

While chromosomal DNA is typically inherited from both parents, under some circumstances an individual inherits two copies of a chromosome from one parent. This is known as uniparental disomy (UPD) [Engel, 2006]. UPD does not involve an aneuploidy (change in chromosomal copy number), and the region of UPD may extend over an entire chromosome or segmentally across a subregion of a chromosome. If the two copies inherited from one parent are identical, the condition is termed uniparental isodisomy (iUPD); this may occur when sister chromatids fail to separate during the second meiotic division. If the two inherited copies are different homologs, the result is uniparental heterodisomy (hUPD). This occurs when bivalent chromatids fail to separate during meiosis I. A mixture of iUPD and hUPD is also possible, due to meiotic recombination [Flori et al., 2005].

Recent advances in microarray-based assays for SNPs allow the genomewide assessment of regions of homozygosity that may be caused by UPD or autozygosity. SNP experiments allow the determination of both copy number (based on the intensity of a hybridization signal) and SNP genotype calls. This consists of allele calls that are heterozygous (AB), homozygous (AA or BB), or null (no call). SNP patterns can be used to infer the presence of regions of homozygosity in the absence of copy number gain or loss, consistent with an interpretation of UPD or autozygosity. By examining SNPs from a father/mother/child trio, one can identify potential UPD in cases such as: AA/BB/AA (both A alleles in the child are derived from the father), AA/BB/BB (both B alleles in the child are derived from the mother), and AA/AB/BB (both B alleles in the child are identical and are derived from the mother). In each of these examples, the genotype of the child is derived from one parent, with possible interpretations of iUPD and hUPD. A series of consecutive SNPs of the same type aligned along chromosomal positions provides evidence of the inheritance of a child's chromosome or chromosomal segment.

We created SNPtrio as a web-based tool to visualize tracts of uniparental inheritance (UPI), biparental inheritance (BPI), hemizygous and homozygous deletions, and Mendelian inconsistencies (MI) in biallelic SNP data from trios. Our primary focus was on the analysis of SNPs that are homozygous in the child and that are informative for identifying regions of UPI. MI in SNP trio data may be caused by a variety of technical or biological sources. Current SNP technologies classify SNPs as biallelic, such that true monoallelic genotype calls of A/- or B/- are reported as AA or BB, respectively. Nonetheless SNPtrio can detect deletions in such regions in trio data, and the use of a copy number track allows deletions to be distinguished from UPD or autozygosity. An approach to analyzing UPD from SNP genotype data using trios was reported by Bruce et al. [2005]. This approach differs from SNPtrio, for example in its basic assumption that no copy number variations occur, i.e., that calls are biallelic. Therefore deletions cannot be distinguished from UPD.

MATERIALS AND METHODS

SNPtrio consists of three parts: a central schema consisting of various informative genotype calls in a trio, a web server where the schema is implemented, and SNP data from several platforms can be uploaded and analyzed, and an output consisting of assorted tables and a plot that visualizes regions of UPI of the child.

SNPtrio Schema

Fig. 1 outlines a schema showing the 27 possible combinations of SNP genotype calls when each of the three individuals in a trio has a biallelic call. The columns (1–5) in the schema indicate five SNP combinations suitable for detecting particular types of BPI, UPI, or MI as a function of chromosomal position. We employ the term UPI to describe UPI that may be attributed to copy neutral changes (e.g., UPD) or to changes in copy number (e.g., heterozygous or homozygous deletions). The five classes are: 1) paternal UPI (UPI-P); 2) BPI; 3) maternal UPI (UPI-M); 4) single Mendelian inconsistency (MI-S); and 5) double Mendelian inconsistency (MI-D). Among these five informative classes, the two UPI classes are further color-coded for identification of iso- vs. hetero-: "iUPI-P" denotes uniparental iso-inheritance that is paternal; "hUPI-M" denotes uniparental hetero-inheritance that is maternal. BPI refers to situations in which the parents are homozygous for a given SNP (one parent having an AA genotype, the other BB) and the child is heterozygous (AB), specifying that the child's genotype must have been derived from both parents. The BPI calls in SNPtrio provide a baseline set of calls in which BPI positions are observed. We define "event regions" as UPI (heterozygosity due to deletions or homozygosity interpreted as UPD or autozygosity) or MI phenomena occurring in consecutive SNPs on a chromosomal arm. Supplementary Figure S1 (available online at http://www.interscience.wiley.com/jpages/1059–7794/suppmat) depicts the associated pedigrees of the schema in Fig. 1.

SNPtrio is designed to accept biallelic genotype calls as input, with or without copy number information (in terms of $\log_2$ ratio). The explanation of the SNPtrio schema (Fig. 1) is based on two alleles for every individual in a trio. Therefore, chromosome X of a son, or other regions of hemizygous deletion in a child, are flagged by SNPtrio as UPI regions. When copy number information is available, regions of UPI identified by SNPtrio can further be interpreted as a hemizygous deletion in the child or a copy-number neutral UPD in the child (Supplementary Figure S2).

The "X" marks in Fig. 1 correspond to possible classes consistent with a particular combination of SNP genotype calls. Those 18 genotype call combinations which can identify a SNP as being exclusively consistent with one of the five SNPtrio classes are highlighted with a blue background. It is formally possible that erroneous calls occur for any SNP genotype. Depending on the platform, the false-positive and false-negative error rates for SNP calls are typically low ($<1\%$), and such errors are unlikely to occur in consecutive informative SNP positions that are identified by SNPtrio.

SNPtrio Input

SNPtrio accepts input of SNP data in three formats: exported from GTYPE software of Affymetrix (Santa Clara, CA), exported from BeadStudio software of Illumina Inc. (San Diego, CA), or downloaded from the HapMap website (www.hapmap.org). An input text file consists of tab-separated columns of: SNP_ID (dbSNP, Affymetrix SNP, etc.), chromosome, position, plus one genotype call column per individual, in terms of AA, AB, or BB. Since not all SNP data have associated copy number information,

			1		2	3		4	5
Father	Mother	Child	iUPI-P	hUPI-P	BPI	hUPI-M	iUPI-M	MI-S	MI-D
AA	AA	AA	X	X	X	X	X		
		AB						X	
		BB							X
	AB	AA	X	X	X		X		
		AB			X	X			
		BB					X		
	BB	AA	X	X					
		AB			X				
		BB				X	X		
AB	AA	AA	X		X	X	X		
		AB		X	X				
		BB	X						
	AB	AA	X		X		X		
		AB		X	X	X			
		BB	X		X		X		
	BB	AA	X						
		AB		X	X				
		BB	X		X	X	X		
BB	AA	AA				X	X		
		AB			X				
		BB	X	X					
	AB	AA					X		
		AB			X	X			
		BB	X	X	X		X		
	BB	AA							X
		AB						X	
		BB	X	X	X	X	X		

FIGURE 1. **SNPtrio schema. Informative SNPs are highlighted (light blue background). The color of the X mark for each informative SNP is selected to show the class (iUPI-P, BPI, etc.) to which it belongs. The same corresponding colors are also used in SNPtrio plot. Notably, even for hemizygous deletions, current technologies report homozygous biallelic calls. Thus an AA call may represent either AA (biallelic) or A/- (monoallelic); similarly, BB may correspond to BB (biallelic) or B/- (monoallelic). See Supplementary Figure S2 for detailed examples.**

the per individual $\log_2$ ratio (copy number) columns are optional. The null SNPs (e.g., "NoCall," "null," or "NN") are noninformative in the SNPtrio schema and are ignored. Extra columns are also ignored. Sample input files are provided on the SNPtrio website (http://pevsnerlab.kennedykrieger.org/SNPtrio.htm). SNP data from other platforms can be accepted if the column headers are set to match one of the three supported formats.

Optional Settings

The user can specify the source of SNP data (Affymetrix GTYPE, or Illumina BeadStudio, or HapMap), the fonts and style of the output plot, a variable window size for the statistical assessment of significant event regions, a combination of chromosomes to be displayed, the human genome build information (affecting chromosomal position), and whether the corresponding probability, or copy number panel, is included for each event region when plotting a single chromosome.

SNPtrio Implementation

SNPtrio is implemented via an Apache web server, running on an HP server installed with RedHat Enterprise Linux AS operating system. The SNPtrio software includes a combination of HTML, Perl, C, and R codes.

SNPtrio Output

SNPtrio generates an output plot with the following formats available: PNG, TIFF, PostScript, and PDF. It also provides four text outputs: a ".sns" file, which includes detailed SNP calls from the trios for all of the informative SNPs (as defined in Fig. 1), a ".blk" file for all of the event regions (blocks) found, a ".csv" file (Table 1), which lists the probability calculated for each block, and a ".cnt" file, which consists of per chromosome, per trio counts for each of the informative classes (Supplementary Table S1). The copy number information, in the $\log_2$ ratio form, is included in the .sns and .blk files, if it is present in the user's input file.

TABLE 1. **Example of a SNPtrio ".csv" Output File***

	Child ID	Chr	Start	End	Type	Size	p value
1	Family 1	15	23321583	99638601	iUPI-M	33	0
2	Family 4	2	2907122	88317231	iUPI-P	63	0
3	Family 4	2	99306329	237018374	iUPI-M	97	0
4	Family 5	11	5503448	45185781	iUPI-P	28	0
5	Family 5	X	4544856	46704028	iUPI-M	15	NA
6	Family 5	X	67088612	148840899	iUPI-M	30	NA
7	Family 6	20	38912414	62622258	iUPI-P	12	0
8	Family 6	X	3667078	55853855	iUPI-M	21	NA
9	Family 6	X	67271956	149514662	iUPI-M	42	NA

*Each row contains information on each uniparental inheritance block identified by SNPtrio, including the associated p value. In the p value column, the entry NA (not available) is returned for X chromosome queries. For small p values (approximately $<10^{-16}$) SNPtrio reports a p value of 0. Size refers to number of consecutive informative SNPs of a given type.
chr, chromosome.

SNPtrio plots those informative SNPs, one track per class, for users to visualize event regions (UPI and MI) across the genome. If any one SNP of the trio data for a SNP site is a "NoCall" (null), it renders that SNP uninformative. Uninformative SNP sites are not plotted by SNPtrio.

The BPI track plots the "known biparental inheritance" SNPs, in which the parents have AA and BB calls and the child has an AB call. It is included to demonstrate the presence of a region in which there is not an event such as UPI. The BPIs are also necessary for the calculations of statistical significance of event regions. Examples of plots for which BPI calls are lacking in a UPI region are shown in Figs. 2 and 3 (discussed below), and plots for

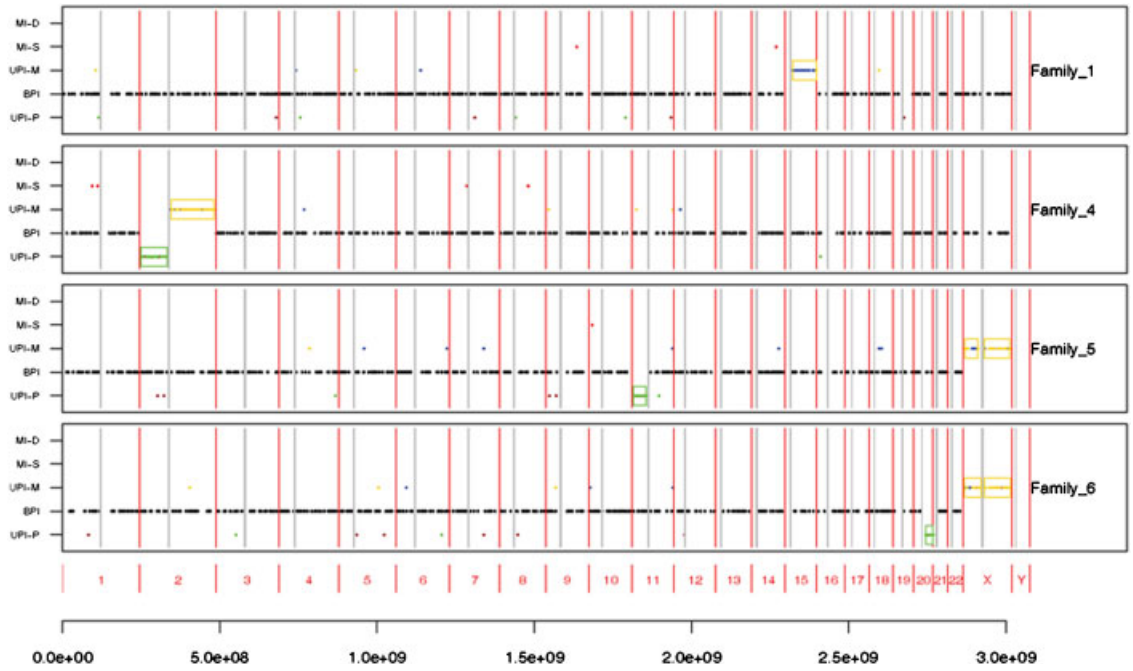

FIGURE 2. **SNPtrio plot for four trios having a proband known to have UPD, based on Affymetrix 10 K SNP array data from Altug-Teber et al. [2005]. SNPtrio identified iUPI-M on 15q in Family_1; iUPI-P on 2p and iUPI-M on 2q in Family_4; iUPI-P on 11p in Family_5; and iUPI-P on 20q in Family_6. SNPtrio also identified iUPD-M on chrX in Family_5 and Family_6, since the proband in both cases is a son. Each dot is plotted at a position and with a color based on the Fig. 1 schema. The x-axis on all SNPtrio plots corresponds to physical map position. Vertical red bars separate the chromosomes, and vertical gray bars indicate the centromeres.**

which BPI calls are lacking in regions of hemizygous deletion are shown in Supplementary Figures S3 and S4.

The classes of UPI-P and UPI-M trio-call combinations are further subdivided into two subclasses. One is "isodisomy," in which, for example, one parent has an AA call, the other has AB, and the child has a BB call. Such a combination precludes the possibility of heterodisomy. The other subclass is "disomy," in which for example one parent has an AA call, the other has BB, and the child has a BB call. In this case, the child must have inherited both of the B copies from only one parent, but the inheritance could reflect either heterodisomy or isodisomy. Alternatively, the child has only one B allele (because of a hemizygous deletion), but it is called as biallelic and homozygous by SNP software. The two subclasses of UPI-P and UPI-M are plotted on the same track, using different colors. A uniparental iso-inheritance region occurs when the plotted BPI track remains blank for a region while both color of dots are presented in a mixture of iUPI-P and hUPI-P or iUPI-M and hUPI-M. If the BPI track remains blank and only dots with the color of "disomy" are plotted, this potentially indicates a uniparental hetero-inheritance region.

MI is plotted on two separate tracks. MI-S indicates that the child has a single allele that is inconsistent with a Mendelian inheritance pattern. For example, MI-S can occur when both of the parents have an AA call and the child has an AB call. MI-D occurs when both of the child's genotype calls are inconsistent with the Mendelian rule, as in the case of both parents having an AA call while the child has a BB call.

UPI Event Region

A UPI event region is defined as: 1) a user-defined set of consecutive SNPs (set to a default minimum length of three); 2) all have the same type (MI-S, MI-D, UPI-P, or UPI-M; with iso- and hetero- combined); and 3) occurrence along a chromosomal arm. According to this definition, no BPI SNPs are located within a UPI region. A box is drawn for each UI region, using color corresponding to the dots of that region. When both iso- and hetero- types are presented, the color of the iso- type is used for the box.

Probability Calculation

When SNPtrio is used to analyze a single chromosome, an ideogram is displayed, as well as rectangles enclosing stretches of three or more consecutive SNPs on the MI-D, MI-S, UPI-M, and UPI-P tracks. These rectangles include an associated value corresponding to the likelihood of occurrence of these particular patterns of SNPs by chance, under the assumption of normal segregation. If there were no abnormal regions in the chromosomes of the trios (i.e., if all copy numbers were equal to two and no Mendelian errors had occurred), the only non-BPI points on the tracks would be due to genotyping errors (which is our null hypothesis). Assuming the independence of genotyping errors, we calculate the probability of observing stretches (or clusters) of consecutive non-BPI data points as large as or larger than the ones observed under the null. If we see large stretches of non-BPI points with low p values, we infer that this cannot be explained reasonably well by chance events (i.e., genotyping errors) alone. We then reject the null and declare that the data support that an event has occurred. This approach provides a useful guide to highlight regions identified by the algorithm as potentially biologically significant. The SNPtrio output includes a table listing the chromosome for each event, the start and end positions, the type (e.g., iUPI-M), the size (in consecutive SNPs), and the probability value (see Table 1).

SNPtrio provides the probability of occurrence of a particular event region based only upon each individual SNP array result uploaded to the program. No reference library is used in such calculations (see Discussion). We calculate the following probability: given the number of SNPs for a particular event region type (*M*), among all uploaded informative autosomal SNPs (*X*) of that trio, what is the probability to have the observed number (*N*) or more SNPs of that type clustered together solely by chance? In other words, if the M SNPs of a particular event region type were randomly dispersed among the *X* informative autosomal SNPs, how probable is it to observe an event of the same or larger magnitude (defined by the number of consecutive SNPs of that event region)? Note that since this is a global assessment across all autosomes, this approach does not require a Bonferroni or other multiple comparison correction. These calculations are based solely on the assumption of independence of genotyping errors, since in trios of apparently normal individuals we expect only random deviations from BPI caused by genotyping errors.

To derive the solution, we consider this question as a general binomial problem. Let *P* be the probability of a positive outcome in a Bernoulli trial (e.g., a one for zero/one outcomes), and assume that we have Z trials. Let *LS* be the length of the largest block of consecutive ones in those Z trials, and let $P_Z(LS<N)$ be the probability that *LS* is smaller than some number *N*. Clearly, if $Z<N$, then $P_Z(LS<N)=1$, as we can never have more than Z ones in a row. If $Z=N$, then $P_Z(LS<N)=1–P^N$: only if all trials had outcome one the number *LS* would be equal to *N*. The derivation of $P_Z(LS<N)$ in the case when Z is larger than *N* (as is always the case in the SNP data) uses a recursive argument. Any sequence of zeros and ones with $LS<N$ can not start with *N* or more ones, otherwise *LS* would be bigger or equal to *N*. Hence any of those sequences starts with *k* ones (where k is between 0 and *N*–1) before the first zero comes. The remainder of the sequence, which has length Z–k–1, also must not contain any stretch of ones equal to or longer than *N*. Hence

$$P_Z(LS<N)=\sum_{k=0}^{N-1}P^k\cdot(1-P)\cdot P_{Z-k-1}(LS<N)$$

Therefore, we first tabulate the probabilities $P_1(LS<N),\ldots,P_N(LS<N)$ as outlined above, and then iteratively, as given by the above equation, calculate $P_{N+1}(LS<N)$, $P_{N+2}(LS<N)$, etc., until $P_X(LS<N)$. Using the values from observed data and $P=M/X$, the probability of having *N* or more of SNPs of the particular MI type clustered together solely by chance therefore is equal to $1–P_X(LS<N)$.

SNP Data From Trios

All studies were performed with informed consent and approval of the Institutional Review Board at Johns Hopkins. Transformed lymphoblastoid cells were obtained from: 1) a female patient with a chromosome 3p deletion and her apparently normal parents, 2) a patient with a chromosome 3q deletion (with symptoms including self-injurious behavior [SIB]) and her apparently normal parents, and 3) two trios in which the child was diagnosed with Down syndrome. The deletions were detected by karyotyping and confirmed by fluorescence in situ hybridization (FISH). The Down syndrome cases were confirmed to have trisomy 21 by karyotyping. For the two trios involving deletion cases, DNA was purified from cell lines from the probands and parents and submitted to the Center for Inherited Disease Research (CIDR, Baltimore, MD) for analysis on Illumina 550 K arrays. Data analysis was performed using BeadStudio software (version 2.2; Illumina) then a text file was exported for analysis by SNPtrio. Data from unrelated parents were obtained using the identical platform. For the two trios involving Down syndrome, DNA was purified and submitted to the Translational Genomics Institute (TGen, Phoenix, AZ) for analysis on the Affymetrix 100 K platform. Data analysis was performed using GTYPE software (Affymetrix) and CRLMM [Carvalho et al., 2006].

RESULTS

SNPtrio Performance in Known UPD Cases

We analyzed SNP data from four trios described by Altug-Teber et al. [2005] using Affymetrix 10 K SNP arrays, shown in Fig. 2. These include a case of maternal UPD on chromosome 15 (Prader-Willi syndrome; Family_1), paternal UPD on chromosome 2p in combination with maternal UPD on chromosome 2q (Family_4), paternal UPD on chromosome 11p (Beckwith-Wiedemann syndrome; Family_5), and paternal UPD on chromosome 20q (pseudohypoparathyroidism; Family_6). In the Prader-Willi case, SNPtrio plotted all informative SNPs on chromosome 15q on the UPI-M track, while no informative SNPs for other tracks (anomaly types) were plotted for that chromosome arm. In this way, the UPD phenomenon was clearly highlighted. The lack of BPI on 15q indicates that there were no cases of normal inheritance from both parents in that region. In the chromosome 2 case (Fig. 2, second panel) an unusual case of paternal UPD (2p) and maternal UPD (2q) is highlighted. This individual was a healthy carrier of two isochromosomes, i(2p) and i(2q) [Albrecht et al., 2001]. Informative profiles corresponding to paternal UPD events are also displayed in Family_6 (bottom panel), showing homozygous regions consistent with those described by Altug-Teber et al. [2005]. For chromosome X, a female child of a trio presented a SNP profile typical of the autosomes (Family_1 and Family_4) while a male child having only one copy of X (inherited from the mother) presented a pattern typical of maternal inheritance (Family_5 and Family_6). The yellow colored dots (and boxes; discussed below) correspond to SNPs that were consistent with a trio in which iUPD has occurred (e.g., AA/AB/BB in the father/mother/child), while the blue dots correspond to SNPs that are consistent with patterns of either isodisomy or heterodisomy (e.g., AA/BB/BB in father/mother/child).

We further analyzed data from two trios based on the Affymetrix 100 K platform, containing 116,000 SNPs. The trios each included a child with Prader-Willi syndrome (described in [Slater et al., 2005]). SNPtrio revealed an expected pattern of maternal UPD on chromosome 15q (Supplementary Figure S5). SNPtrio plotted copy number estimates for each trio, indicating that there were no apparent copy number changes on chromosome 15.

SNPtrio Performance in Deletion Cases

In the absence of copy number data, SNPtrio analyzes SNP genotype data from trios assuming that all genotypes are biallelic (disomic), as indicated in the Fig. 1 schema. In the case of hemizygous deletions, the genotypes are monoallelic. We developed a schema describing possible outcomes given the observed genotype calls of paternal AA, maternal BB, and child AA, AB, or BB (Supplementary Fig. S2). In this schema, the true copy number is shown (involving a deletion in father, mother, or child, or UPD with no deletion). The SNPtrio interpretation of UPI (Fig. 1) thus may be consistent with a true deletion or UPD event. By including copy number information, SNPtrio allows copy number neutral cases (e.g., UPD) to be distinguished from hemizygous or homozygous deletions. The male X chromosome, discussed

above, presents an example of a monosomic chromosome resembling a hemizygous deletion, and SNPtrio effectively displays maternal inheritance of that chromosome. We further assessed the performance of the tool on known deletion cases.

Case L92_1135 is a child who displayed SIB and was diagnosed with a de novo hemizygous deletion on chromosome 3q. This deletion was confirmed by FISH. We obtained SNP data from the Illumina 550 K platform for this proband and her parents, and plotted the results using SNPtrio (Fig. 3, panels 3 and 4). The Illumina $\log_2$(Rsub/Rref) ratio track (reflecting copy number) revealed the deletion region on 3q, overlapped by an event region of UPI-M. This indicates that the father's copy of this region was deleted in the child. For comparison, we plotted data from a trio that was apparently normal on chromosome 3 (Fig. 3, panels 1 and 2). For the 3q deletion case SNPtrio displayed the reduced copy number and the parent of origin of the deletion. In order to further infer whether this deletion was inherited or de novo, it would be necessary to assess copy number change in the father (e.g., using BeadStudio, GTYPE, or SNPscan software), or to examine the genotype calls in the affected region of the father. Heterozygous calls would be expected to be absent in the case of an inherited deletion, but present in the case of a de novo deletion.

A case of 3p hemizygous deletion (Case L99-2297), confirmed by FISH, was reported by Cargile et al. [2002]. We previously confirmed the deletion using array comparative genomic hybridization and SNP array data from this individual as visualized by SNPscan [Ting et al., 2006]. We obtained lymphoblastoid cell lines from a trio consisting of the 3p deletion case and his parents, isolated genomic DNA, and obtained data on 550,000 SNPs using the Illumina 550 K platform. The SNPtrio results are shown for chromosome 3 (Fig. 3, panels 5 and 6). For both the 3p and the 3q deletion cases, the profile in Fig. 3 showed a lack of BPI in the region of the deletion. This result is expected, since the proband can have A or B calls that are interpreted by the manufacturer's software as AA or BB. However, AB calls in the proband are not expected or observed in a hemizygous deletion region, and thus BPI is not observed (Fig. 3, panels 3 and 5). SNPtrio also provided tabular results for each of seven categories (MI-D, MI-S, UPI-M, iUPI-M, BPI, iUPI-P, and UPI-P). Analysis of this Illumina dataset showed large numbers of UPI-M and iUPI-M on chromosome 3 (as expected). Supplementary Table S1 presents a tabular output from SNPtrio for trios including probands L92_1135 and L99_2297. These results quantify the chromosome 3 and other changes. We further validated SNPtrio by analyzing additional cases of patients with mental retardation having single copy deletions [Friedman et al., 2006] (Supplementary Fig. S3).

Identification of Maternal Homozygosity in Trios With a Male Child

It may be of interest to identify regions of homozygosity in females. A son's hemizygous genotype on the X chromosome (A or B) will be called as AA or BB, respectively, given of the

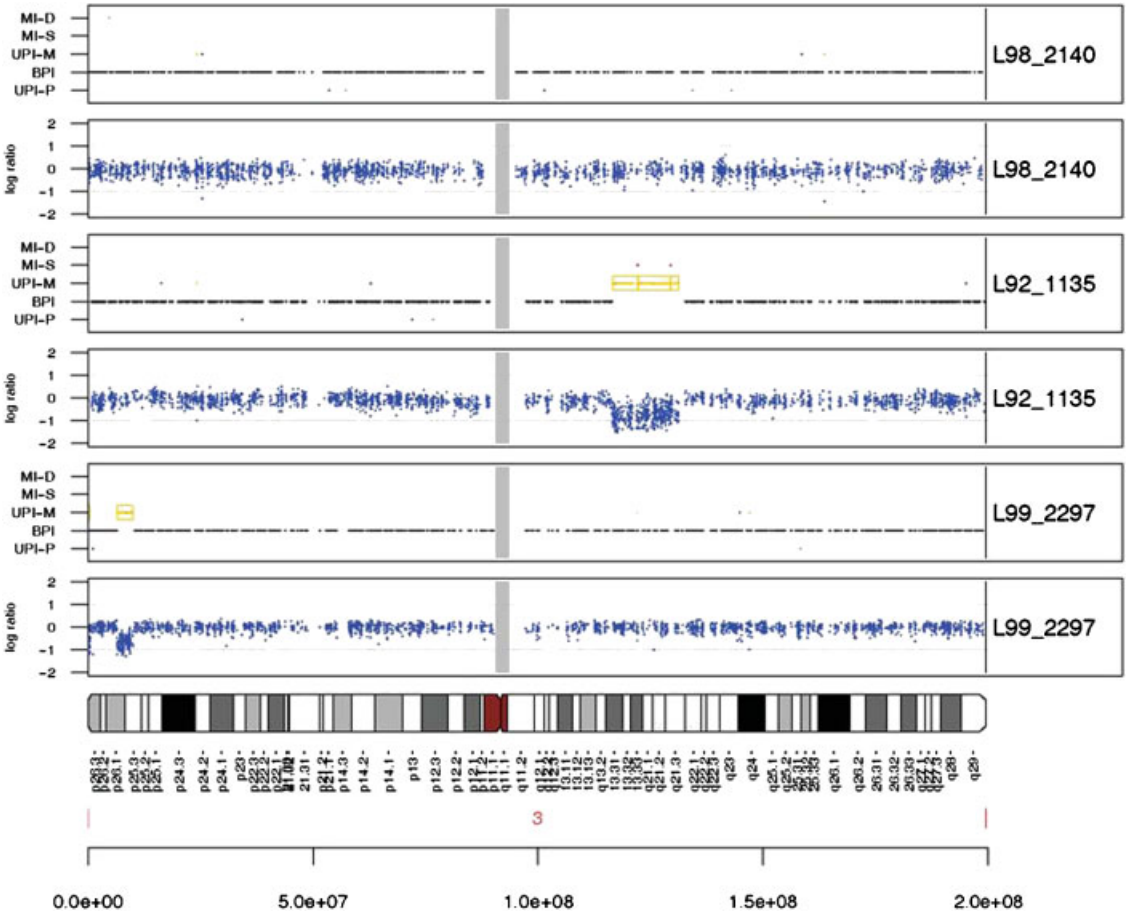

FIGURE 3. **SNPtrio performance in cases of hemizygous deletion. Data are plotted for three cases along chromosome 3 (x-axis) with the main SNPtrio output tracks above an optional copy number track (based on 550 K Illumina BeadStudio $\log_2$Rsub/Rref data). The cases are Case L98-2140 (apparently normal trio), Case L92-1135 (deletion on 3q), and Case L99-2297 (deletion on 3p). Note that for each of the two lower trios, the UPI-M signal occurs in a deletion region, as confirmed by the copy number measurements.**

limitations of current software to making biallelic calls. If a son's call matches his mother's call (i.e., maternal AA or BB) she must be homozygous. If on the other hand, a son's call (AA or BB) does not match his mother's, she must be heterozygous (AB). Thus, regions of blue dots along the X chromosome of trio having a male son reflect homozygosity in the mother. The father's X chromosome genotype is not expected to be of importance in this analysis as it is not inherited.

We analyzed HapMap trios and found the X chromosome had regions of blue dots (but not yellow dots) near the centromere for some trios with a male child. One such example (NA19120) is shown in Supplementary Figure S4. This result indicated the occurrence of maternal homozygosity on the X chromosome corresponding to this region. An expectation of this result is that the mother (NA19116) has extended regions of homozygosity on the X chromosome corresponding to this region. Analysis of HapMap data for the mother (NA19116) confirmed the presence of pericentric homozygous calls (Supplementary Fig. S4). This could be due to a hemizygous deletion in the mother, or to copy number neutral homozygosity caused by phenomena such as UPD or autozygosity.

SNPtrio Performance in Trios With Unrelated Parents

As proof of principle of the functionality of SNPtrio and to enhance the understanding of Fig. 1 and the patterns that arise, we analyzed two trios in which SNP data for the parents were substituted with those of unrelated individuals (Supplementary Fig. S6). First, we plotted both chromosome 3 and an overview of all chromosomes for proband L92-1135 (having a 3q deletion as shown in Fig. 3) with her biological parents. We substituted SNP data for the mother with those of an unrelated female, then we substituted SNP data for the father with those of an unrelated male, and finally we substituted the data for both parents. These changes resulted in a dramatic series of altered profiles. For example, in the deletion region we observed primarily UPI-M calls in the related trios (as shown in Fig. 3), UPI-M, UPI-P, and MI-D calls using data from an unrelated mother, UPI-M calls using an unrelated male (thus, as expected, this had no effect on the outcome), and a combination of MI-D, UPI-M,and UPI-P calls for two unrelated parents.

Loss of Transmitted Allele

Redon et al. [2006] recently described loss of transmitted allele (LTA) as a phenomenon in which a parent transmits an allele to a normal child, and subsequently the transmitted allele is deleted in that parent. As a consequence, the genotype of a heterozygous SNP in that parent changes from AB to A/- (interpreted as AA by current SNP software) or B/- (interpreted as BB). This could cause an apparent MI event region. For example, consider the case of a father's genotype call changing from AB to AA after transmitting the B allele to the child. (This change could potentially occur as a cell culture artifact.) There are three scenarios based on the mother's possible genotype: 1) If the mother is AA, then the child must have an AB genotype. This is classified as MI-S by SNPtrio. 2) If the mother is AB, then the child is either AB (a noninformative SNP) or BB (classified as iUPI-M). 3) For mother BB and child BB, the SNPtrio interpretation is UPI-M. Therefore, for the affected region, SNPtrio will display SNPs in classes of MI-S, UPI-M, and normal BPI (from the father's original homozygous SNPs).

We used SNPtrio to identify a possible case of LTA in trio data, from Release 21 of the HapMap Project, including a father (NA12872), mother (NA12873) and child (NA12864). Chromosome 22q11.22 included a region consisting of UPI-P, BPI, and MI-S (Fig. 4A). We used the SNPscan program [Ting et al., 2006] to visualize copy number data generated from Affymetrix GeneChip® Human Mapping 500 K Array Set for these three individuals and observed a hemizygous deletion in the mother but not the father or child (Fig. 4B). This is consistent with an interpretation of LTA with the loss of a maternal allele.

SNPtrio Analysis of Duplication in Trisomy 21

We assessed the performance of SNPtrio on duplicated regions by obtaining SNP data (Affymetrix 100 K platform) on individuals diagnosed with Down syndrome and their apparently normal parents. Copy number estimates from the SNP data indicated the presence of an extra copy of chromosome 21 in the affected children but not their parents (data not shown). In these trisomic cases, most or all of the child's genotypes for chromosome 21 are triallelic. We employed two data analysis approaches that only produce biallelic calls: GTYPE (the Affymetrix software package), and CRLMM [Carvalho et al., 2006]. The results for two trios are shown in Fig. 5. CRLMM showed extended regions of UPI-P consistent with a paternal origin of inheritance of the third copy of chromosome 21 for Case 1505 (Fig. 5, top panel). Of 73 informative SNP sites, there were 70 calls consistent with paternal inheritance, and only three calls consistent with maternal inheritance. Notably, the majority of calls was consistent with isodisomy of the paternal alleles. Using GTYPE (Fig. 5, third panel) all nine informative calls were consistent with paternal inheritance, and zero calls consistent with maternal inheritance. The two data analysis algorithms yielded consistent results. A greater number of total informative calls occurred using CRLMM because of improved genotyping performance and because CRLMM does not permit no calls, while GTYPE does so.

For another trio with a Down syndrome child (Case 1852), SNPtrio indicated maternal UPI which we interpret as maternal inheritance of the third copy of chromosome 21 (Fig. 5, panels 2 and 4). Using the CRLMM algorithm, we observed 59 instances of maternal inheritance (22 maternal iso-UPI and 37 iso-/hetero-UPI), and only five instances of paternal inheritance. We analyzed five microsatellite markers from chromosome 21 for the parents and child, and three of these markers were informative and consistent with maternal inheritance of the third copy of chromosome 21 (data not shown). This supports the hypothesis that SNPtrio can identify the parent of origin of an extra chromosome copy, even using only biallelic genotype calls.

Recently, Wang et al. [2006b] reported an algorithm (CN-RLMM) capable of making triallelic genotype calls. We employed this algorithm for the genotype calls of the two Down syndrome cases, and CRLMM for the parents. For the trio including Case 1505 there were 107 informative calls consistent with paternal inheritance of the third copy of chromosome 21 and only 24 calls consistent with maternal inheritance. For the trio including Case 1852, there were 102 calls consistent with maternal inheritance of the extra chromosomal copy, and only 35 calls consistent with paternal inheritance. These results support the assignments made by SNPtrio using biallelic calls.

DISCUSSION

SNPtrio allows the identification, visualization and analysis of UPI resulting from either UPD or deletions with SNP data collected from trios. We anticipate that it will be useful for

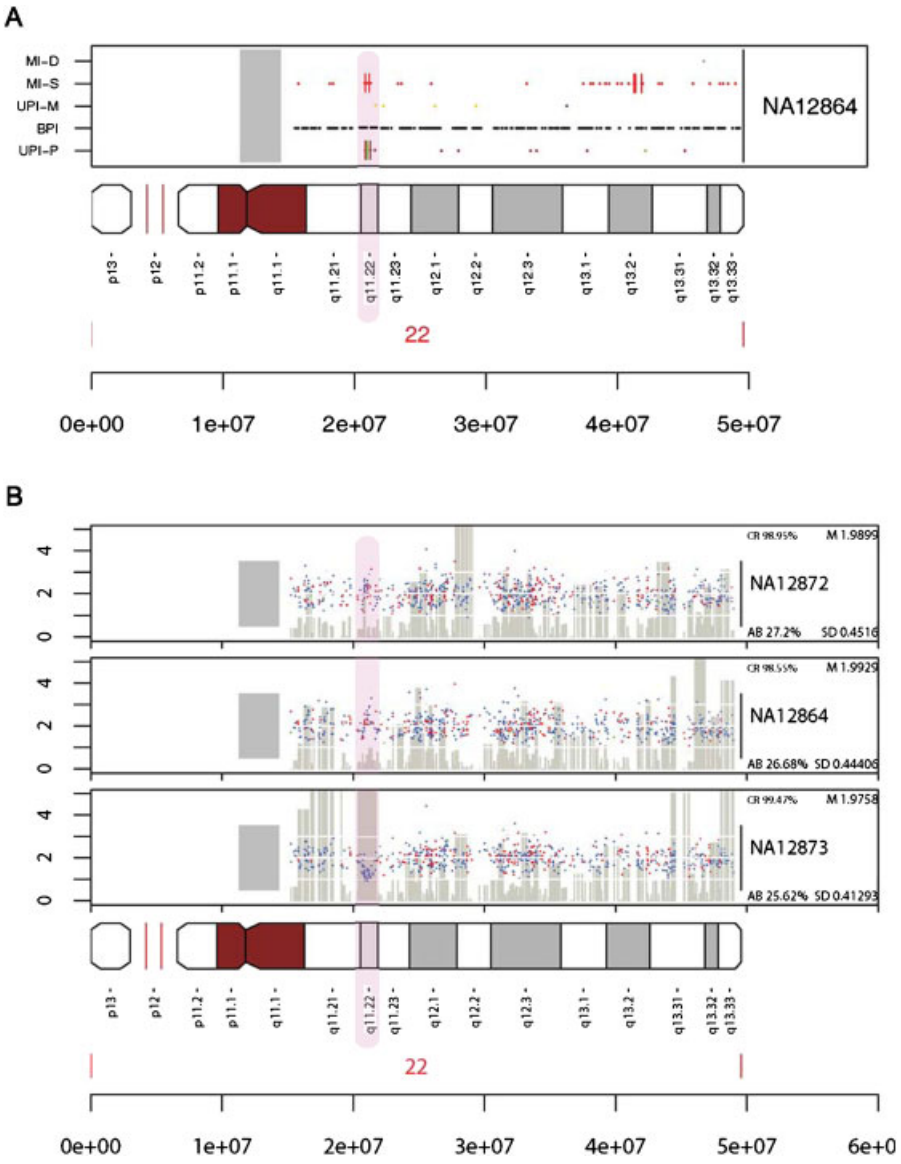

FIGURE 4. **Identification of loss of transmitted allele. A: A HapMap Release 21 trio was plotted using SNPtrio, including a region on chromosome 22q (purple oval) having UPI-P, BPI, and MI-S patterns. B: Analysis of the copy numbers of this trio, using SNPscan with SNP data generated from Affymetrix's Human Mapping 500 K Array Set, revealed a copy number loss in the affected region (purple oval) in the mother. The three panels correspond to Affymetrix SNP data plotted for the father, child, and mother, respectively. The x-axis is physical map position across chromosome 22. The y-axis displays information from GTYPE software as visualized by SNPscan. This includes the copy number (y-axis) of heterozygous SNPs (colored red), homozygous SNPs (colored blue), no calls (colored green), -log$_{10}$ p value for copy number change (gray bars; note the gray bars in the affected region of the mother indicating a copy number loss with $p < 10^{-6}$), and p values for regions of homozygosity (yellow bars). The pattern of UPI-P, BPI, and MI-S in (A) is consistent with the interpretation of loss of transmitted allele based on the copy number loss indicated in (B).**

researchers who perform any SNP experiments involving trios, because the chromosomal alterations it reports may not be detectable using other available software. The present study described examples of large-scale anomalies, but smaller events may be detected as well. We developed SNPtrio based on a schema of 27 possible combinations of biallelic calls in a trio (Fig. 1). This allows the program to identify regions of UPI and MI, including an assessment of the statistical significance of the event. The phenomena that were detected include maternal or paternal UPD (affecting the child in a trio), hemizygous deletion (affecting a child), homozygous deletion (affecting any member of trio), and LTA (occurring in a parent). An approach related to the one presented here was described by Bruce et al. [2000]. SNPtrio is unique because it includes copy number information, it generates a p value for each event, it is web-accessible and (in contrast to other software described below) it reveals a variety of combinations of anomalous events in SNP trio data through the combined analysis of UPI, BPI, and MI.

SNPtrio includes an exact probability calculation which does not require external references or adjustments for multiple comparisons. In particular our approach avoids the use of a reference library of genotype calls in order to address the question of whether a particular event region observed using SNPtrio is likely to have occurred in a large sample of genotype data obtained from apparently normal individuals. In a typical analysis of data with SNPtrio it is likely that SNP data from a

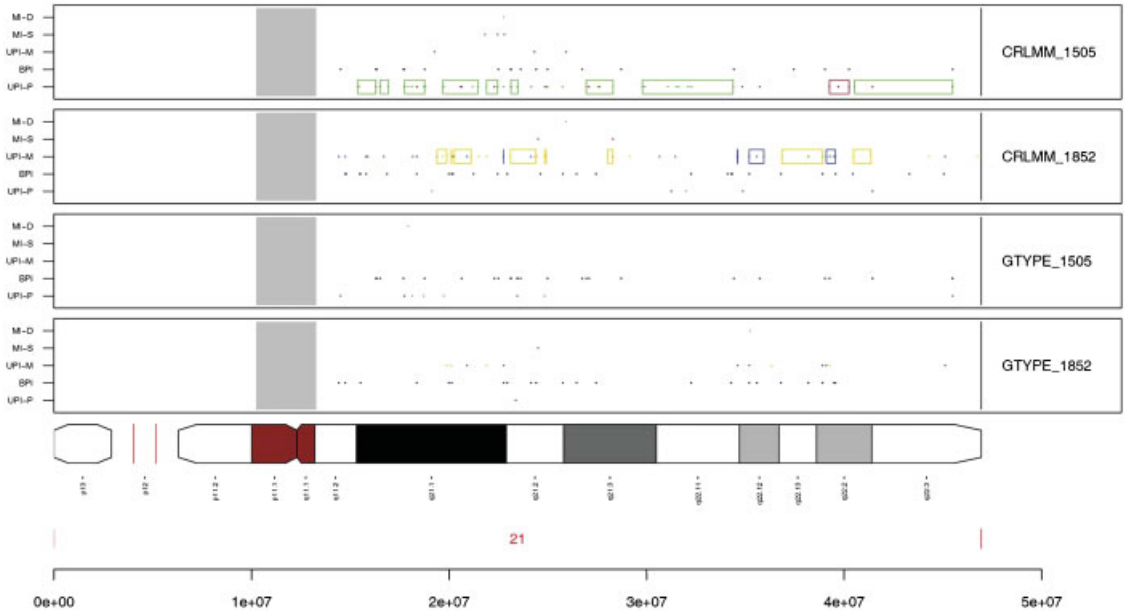

FIGURE 5. **Performance of SNPtrio with two trios in which the child had Down syndrome. SNP data were collected using the Affymetrix 100 K platform. Top two panels include the results of genotype analysis using CRLMM for trios 1505 and 1850. The bottom two panels include results from the same SNP chips, but using Affymetrix GTYPE software for the genotype calls.**

variety of platforms will be used, and each of these may have varying accuracy (i.e., different rates of genotyping errors). Many factors affect the accuracy of SNP data. For a given sample, there is typically variability between labs in the overall call rate, and in the AB call rate for the male X chromosome (a measure of false-positive genotype calls). Other sources of variability may include the quality of DNA preparation, hybridization protocol, and data analysis. There is additional variability within the same facility (e.g., between samples processed at different times). Thus, the meaning of a p value calculated from a user's samples relative to a control data set may be unclear or invalid.

The SNPtrio schema is based on the presence of two alleles, because current commercial SNP microarray software packages typically produce only biallelic calls. However, in cases of hemizygous deletion the observed disomic calls (e.g., AA) correspond to true deletions (e.g., A/-). SNPtrio provides a copy number track that allows the detection of deletions. Supplementary Figure S2 presents a schema describing true and observed genotype calls as well as the corresponding SNPtrio output.

We used SNPtrio to visualize data from trios in which a child has UPD. Since its description in 1980, an increasing number of UPD cases has been reported. In principle, there are 47 possible ways to derive UPI of an entire chromosome pair (22 autosomes and the X chromosome in each parent, plus the XY paternal pair) [Engel, 2006]. Of these scenarios, 32 have been identified to date. Robinson [2000] estimated that the frequency of UPD for any chromosome is one in ~3,500 births and the frequency for an average chromosome is one in ~80,000 births, based on the frequency of UPD15.

In some cases, UPD is thought to be benign. In many other situations, it is associated with disease, due to several distinct mechanisms [Robinson, 2000; Siegel and Slavotinek, 2005]. UPD can disrupt genomic imprinting, such that imprinted genes (expressed preferentially from the paternal or maternal alleles) fail to be expressed. Such disorders include Prader-Willi syndrome (involving chromosome 15q), Angelman syndrome (15q), and Beckwith-Wiedemann syndrome (11p). Also, UPD can cause homozygosity for an autosomal recessive trait. This has been demonstrated for at least 28 autosomal recessive diseases [Zlotogora, 2004] including retinitis pigmentosa and cystic fibrosis.

The advent of high throughput SNP technology has recently permitted the identification of UPD in DNA samples from clinically affected individuals [Altug-Teber et al., 2005; Andersen et al., 2007; Middleton et al., 2006; Pei et al., 2006; Walker et al., 2006]. To assess the clinical significance of UPD, it will be necessary to document the frequency and nature of UPD in the general population. Recently, several groups have analyzed high-density SNP data from 270 apparently normal individuals studied in the HapMap project [HapMap Consortium, 2005]. The HapMap project includes trio data from individuals of northern European (CEU) and Yoruban (YRI) descent, as well as unrelated individuals of Japanese and Chinese descent. McCarroll et al. [2006] identified 541 deletion variants, based on apparent deviations from Mendelian inheritance, apparent deviations from Hardy-Weinberg equilibrium, and null genotypes. They focused on describing these events in the context of deletion rather than copy number neutral changes such as UPD. Also, their schema includes the use of null genotype calls (no calls), which we do not employ in SNPtrio because they are noninformative with respect to UPD. A similar emphasis on polymorphic deletion events has been applied by Conrad et al. [2006], Hinds et al. [2006], and Redon et al. [2006]. Thus, there has been great interest in describing CNVs that are defined as copy number changes involving DNA fragments that are ~1 kb or larger [Feuk et al., 2006; Freeman et al., 2006]. We suggest that the complementary study of UPD in HapMap data will be important to assess its prevalence and significance.

Other investigators have noted large tracts of homozygosity in apparently outbred populations. Gibson et al. [2006] reported 1,393 tracts extending at least 1 Mb in length among 209 unrelated HapMap individuals, and interpreted this as most likely due to autozygosity. Simon-Sanchez et al. [2007] also described

extended regions of homozygosity (contiguous tracts greater than 5 Mb) in 26 of 272 elderly, neurologically normal subjects. They suggested that autozygosity is a more likely explanation than segmental UPD because the presence of multiple tracts of homozygosity within some individuals supports the likelihood of parental consanguinity as the cause. However, trio data were not generated with which to test this hypothesis. Autozygosity implies homozygosity in which two alleles are identical by descent (i.e., they are copies of an ancestral DNA segment). While autozygosity is manifested by stretches of homozygosity in a child, such a phenomenon is not visualized by the SNPtrio tool because those homozygous patterns are classified as noninformative. For all instances of UPI calls in the SNPtrio schema, the child is homozygous for a given SNP while the parents have nonidentical genotypes (with either one or both parents having homozygosity for a given SNP). We predict that SNPtrio could detect autozygosity by the absence of BPI calls or any UPI calls in the presence of normal copy number. If each parent shares an ancestral allele, the situation in which one parent is AA and the other BB (with child AB) is never expected to occur.

SNPtrio also detected the parent of origin of an extra copy of chromosome 21 in two trisomic trios. The genotypes of the Down syndrome individuals are triallelic, consisting of the calls AAA, AAB, ABB, or BBB. We used biallelic calls (AA, AB, or BB) obtained from the available GTYPE and BeadStudio software packages, which do not generate triallelic calls. The rate of heterozygote biallelic calls increased dramatically in the trisomic region (data not shown), and the influence of triallelic genotypes on biallelic calls has not been assessed. Nonetheless SNPtrio revealed an extremely strong signal in the SNP data that reflected the parent of origin of the extra chromosome. The paternal calls in SNPtrio were largely consistent with isodisomy of the two paternal copies. We speculate that SNPtrio succeeded at identifying the parent of origin because the child has a preponderance of genotypes derived from that parent, thus influencing which biallelic calls were generated by GTYPE and BeadStudio software. We also employed the CN-RLMM algorithm to obtain triallelic calls, and these indicated a parent of origin of the extra copy of chromosome 21 that was consistent with that shown using biallelic calls. As a future development of the SNPtrio package we are developing a schema for trisomic genotypes, analogous to the biallelic schema of Fig. 1. Thus we may be able to model regions of chromosomal duplication.

A variety of prior tools have been developed to analyze and visualize SNP data. The HapMap website includes a genome browser as well as integration with the HaploView tool [Barrett et al., 2005]. This is one of many software packages useful for visualizing SNP data relevant to linkage disequilibrium, SNP tagging [Stram, 2004], phasing of trios [Marchini et al., 2006], detection of MI (e.g., PedCheck) [O'Connell and Weeks, 1998], detection of recombination hotspots [Fearnhead, 2006], or SNP browsers [Shah et al., 2005; Tebbutt et al., 2005; Wang et al., 2006a; Yang et al., 2006]. AutoSNPa is a tool for the visualization of autozygosity regions using Affymetrix SNP array data [Carr et al., 2006]. That tool relies on pedigrees with affected or unaffected members, and shows patterns of shared homozygosity without statistical analyses. We previously developed SNPscan [Ting et al., 2006], a web-based tool used to display SNP data with copy number and genotype information from single array experiments. We also created SNPchip, a related R package available at BioConductor (www.bioconductor.org) [Scharpf et al., 2007]. To our knowledge SNPtrio is unique in its approach to visualizing and analyzing SNP from trios.

ACKNOWLEDGMENTS

We thank the laboratories of Drs. Altug-Teber, Olaf Riess, Harold Slater, and Marco Marra for providing raw data files from published data sets, and Benilton Carvalho, Wenyi Wang, and Dr. Rafael Irizarry for discussions and assistance with SNP data analysis. We thank Drs. David Cutler, Ada Hamosh, Julie E. Hoover-Fong, and N. Varg for helpful discussions. J.P. was supported by grant R01 HD046598 and MRDDRC grant HD24061 from the National Institutes of Health. G.C. and J.P. received support from the Taishoff Foundation.

REFERENCES

Albrecht B, Mergenthaler S, Eggermann K, Zerres K, Passarge E, Eggermann T. 2001. Uniparental isodisomy for paternal 2p and maternal 2q in a phenotypically normal female with two isochromosomes, i(2p) and i(2q). J Med Genet 38:214.

Altug-Teber O, Dufke A, Poths S, Mau-Holzmann UA, Bastepe M, Colleaux L, Cormier-Daire V, Eggermann T, Gillessen-Kaesbach G, Bonin M, Riess O. 2005. A rapid microarray based whole genome analysis for detection of uniparental disomy. Hum Mutat 26:153–159.

Andersen CL, Wiuf C, Kruhoffer M, Korsgaard M, Laurberg S, Orntoft TF. 2007. Frequent occurrence of uniparental disomy in colorectal cancer. Carcinogenesis 28:38–48.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265.

Bruce S, Leinonen R, Lindgren CM, Kivinen K, Dahlman-Wright K, Lipsanen-Nyman M, Hannula-Jouppi K, Kere J. 2005a. Global analysis of uniparental disomy using high density genotyping arrays. J Med Genet 42:847–851.

Cargile CB, Goh DL, Goodman BK, Chen XN, Korenberg JR, Semenza GL, Thomas GH. 2002. Molecular cytogenetic characterization of a subtle interstitial del(3)(p25.3p26.2) in a patient with deletion 3p syndrome. Am J Med Genet 109:133–138.

Carr IM, Flintoff KJ, Taylor GR, Markham AF, Bonthron DT. 2006. Interactive visual analysis of SNP data for rapid autozygosity mapping in consanguineous families. Hum Mutat 27:1041–1046.

Carvalho B, Bengtsson H, Speed TP, Irizarry RA. 2006. Exploration, normalization, and genotype calls of high density oligonucleotide SNP array data. Biostatistics 8:485–489.

Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. 2006. A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 38:75–81.

Consortium IH. 2005. A haplotype map of the human genome. Nature 437:1299–1320.

Eichler EE. 2006. Widening the spectrum of human genetic variation. Nat Genet 38:9–11.

Engel E. 2006. A fascination with chromosome rescue in uniparental disomy: Mendelian recessive outlaws and imprinting copyrights infringements. Eur J Hum Genet 14:1158–1169.

Fearnhead P. 2006. SequenceLDhot: detecting recombination hotspots. Bioinformatics 22:3061–3066.

Feuk L, Carson AR, Scherer SW. 2006. Structural variation in the human genome. Nat Rev Genet 7:85–97.

Flori E, Girodon E, Samama B, Becmeur F, Viville B, Girard-Lemaire F, Doray B, Schluth C, Marcellin L, Boehm N, Goossens M, Pingault V. 2005. Trisomy 7 mosaicism, maternal uniparental heterodisomy 7 and Hirschsprung's disease in a child with Silver-Russell syndrome. Eur J Hum Genet 13:1013–1018.

Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW, Lee C. 2006. Copy number variation: new insights in genome diversity. Genome Res 16:949–961.

Friedman JM, Baross A, Delaney AD, Ally A, Arbour L, Asano J, Bailey DK, Barber S, Birch P, Brown-John M, Cao M, Chan S, Charest DL, Farnoud N, Fernandes N, Flibotte S, Go A, Gibson WT, Holt RA, Jones SJ, Kennedy GC, Krzywinski M, Langlois S, Li HI, McGillivray BC,

Nayar T, Pugh TJ, Rajcan-Separovic E, Schein JE, Schnerch A, Siddiqui A, Van Allen MI, Wilson G, Yong SL, Zahir F, Eydoux P, Marra MA. 2006. Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. Am J Hum Genet 79:500–513.

Gibson J, Morton NE, Collins A. 2006. Extended tracts of homozygosity in outbred human populations. Hum Mol Genet 15:789–795.

Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA. 2006. Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat Genet 38:82–85.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. 2004. Detection of large-scale variation in the human genome. Nat Genet 36:949–951.

Marchini J, Cutler D, Patterson N, Stephens M, Eskin E, Halperin E, Lin S, Qin ZS, Munro HM, Abecasis GR, Donnelly P, International HapMap Consortium. 2006. A comparison of phasing algorithms for trios and unrelated individuals. Am J Hum Genet 78:437–450.

McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM, International HapMap Consortium. 2006. Common deletion polymorphisms in the human genome. Nat Genet 38:86–92.

Middleton FA, Trauzzi MG, Shrimpton AE, Gentile KL, Morley CP, Medeiros H, Pato MT, Pato CN. 2006. Complete maternal uniparental isodisomy of chromosome 4 in a subject with major depressive disorder detected by high density SNP genotyping arrays. Am J Med Genet B Neuropsychiatr Genet 141:28–32.

Newman TL, Rieder MJ, Morrison VA, Sharp AJ, Smith JD, Sprague LJ, Kaul R, Carlson CS, Olson MV, Nickerson DA, Eichler EE. 2006. High-throughput genotyping of intermediate-size structural variation. Hum Mol Genet 15:1159–1167.

O'Connell JR, Weeks DE. 1998. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63: 259–266.

Pei J, Kruger WD, Testa JR. 2006. High-resolution analysis of 9p loss in human cancer cells using single nucleotide polymorphism-based mapping arrays. Cancer Genet Cytogenet 170:65–68.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. 2006. Global variation in copy number in the human genome. Nature 444:444–454.

Robinson WP. 2000. Mechanisms leading to uniparental disomy and their clinical consequences. Bioessays 22:452–459.

Scharpf RB, Ting JC, Pevsner J, Ruczinski I. 2007. SNPchip: R classes and methods for SNP array data. Bioinformatics 23:627–628.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Yc K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. 2004. Large-scale copy number polymorphism in the human genome. Science 305:525–528.

Shah N, Teplitsky MV, Minovitsky S, Pennacchio LA, Hugenholtz P, Hamann B, Dubchak IL. 2005. SNP-VISTA: an interactive SNP visualization tool. BMC Bioinformatics 6:292.

Sharp AJ, Cheng Z, Eichler EE. 2006. Structural variation of the human genome. Annu Rev Genomics Hum Genet 7:407–442.

Siegel DH, Slavotinek A. 2005. Uniparental disomy. Pediatr Dermatol 22:482–487.

Simon-Sanchez J, Scholz S, Fung HC, Matarin M, Hernandez D, Gibbs JR, Britton A, de Vrieze FW, Peckham E, Gwinn-Hardy K, Crawley A, Keen JC, Nash J, Borgaonkar D, Hardy J, Singleton A. 2007. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. Hum Mol Genet 16:1–14.

Slater HR, Bailey DK, Ren H, Cao M, Bell K, Nasioulas S, Henke R, Choo KH, Kennedy GC. 2005. High-resolution identification of chromosomal abnormalities using oligonucleotide arrays containing 116,204 SNPs. Am J Hum Genet 77:709–726.

Stram DO. 2004. Tag SNP selection for association studies. Genet Epidemiol 27:365–374.

Tebbutt SJ, Opushnyev IV, Tripp BW, Kassamali AM, Alexander WL, Andersen MI. 2005. SNP Chart: an integrated platform for visualization and interpretation of microarray genotyping data. Bioinformatics 21: 124–127.

Ting JC, Ye Y, Thomas GH, Ruczinski I, Pevsner J. 2006. Analysis and visualization of chromosomal abnormalities in SNP data with SNPscan. BMC Bioinformatics 7:25.

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE. 2005. Fine-scale structural variation of the human genome. Nat Genet 37: 727–732.

Walker BA, Leone PE, Jenner MW, Li C, Gonzalez D, Johnson DC, Ross FM, Davies FE, Morgan GJ. 2006. Integration of global SNP-based mapping and expression arrays reveals key regions, mechanisms, and genes important in the pathogenesis of multiple myeloma. Blood 108:1733–1743.

Wang P, Dai M, Xuan W, McEachin RC, Jackson AU, Scott LJ, Athey B, Watson SJ, Meng F. 2006a. SNP Function Portal: a web database for exploring the function implication of SNP alleles. Bioinformatics 22: e523–e529.

Wang W, Caravalho B, Miller N, Pevsner J, Chakravarti A, Irizarry RA. 2006. Estimating genome-wide copy number using allele specific mixture models. Johns Hopkins University, Department of Biostatistics Working Papers. Working Paper 122:1–18.

Wong KK, Deleeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, Macaulay C, Ng RT, Brown CJ, Eichler EE, Lam WL. 2007. A comprehensive analysis of common copy-number variations in the human genome. Am J Hum Genet 80:91–104.

Yang TP, Chang TY, Lin CH, Hsu MT, Wang HW. 2006. ArrayFusion: a web application for multi-dimensional analysis of CGH, SNP and microarray data. Bioinformatics 22:2697–2698.

Zlotogora J. 2004. Parents of children with autosomal recessive diseases are not always carriers of the respective mutant alleles. Hum Genet 114: 521–526.