Uniparental disomy analysis in trios using genome-wide SNP array and whole-genome sequencing data imply segmental uniparental isodisomy in general populations

Kensaku Sasaki,[a] Hiroyuki Mishima,[a] Kiyonori Miura,[b] and Koh-ichiro Yoshiura,[a,*]

[a]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

[b]Department of Obstetrics and Gynecology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

*Corresponding author: Dr K-i Yoshiura, Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan. Tel: +81-95-819-7118, fax: +81-95-819-7121

Email: kyoshi@nagasaki-u.ac.jp

Author email addresses:

Kensaku Sasaki: nell2gene@yahoo.co.jp

Hiroyuki Mishima: hmishima@nagasaki-u.ac.jp

Kiyonori Miura: kiyonori@nagasaki-u.ac.jp

Conflict of interest

None of the authors of this paper declares a conflict of interest.

Abstract

Whole chromosomal and segmental uniparental disomy (UPD) is one of the causes of imprinting disorder and other recessive disorders. Most investigations of UPD were performed only using cases with relevant phenotypic features and included few markers. However, the diagnosis of cases with segmental UPD requires a large number of molecular investigations. Currently, the accurate frequency of whole chromosomal and segmental UPD in a normal developing embryo is not well understood. Here, we present whole chromosome and segmental UPD analysis using single nucleotide polymorphism (SNP) microarray data of 173 mother-father-child trios (519 individuals) from six populations (including 170 HapMap trios). For two of these trios, we also investigated the possibility of shorter segmental UPD as a consequence of homologous recombination repair (HR) for DNA double strand breaks (DSBs) during the early developing stage using high-coverage whole-genome sequencing (WGS) data from 1000 Genomes Project. This could be overlooked by SNP microarray. We identified one obvious segmental paternal uniparental isodisomy (iUPD) (8.2 mega bases) in one HapMap sample from 173 trios using Genome-Wide Human SNP Array 6.0 (SNP6.0 array) data. However, we could not identify shorter segmental iUPD in two trios using WGS data. Finally, we estimated the rate of segmental UPD to be one per 173 births (0.578%) based on the UPD screening for 173 trios in general populations. Based on the autosomal chromosome pairs investigated, we estimate the rate of segmental UPD to be one per 3806 chromosome pairs (0.026%). These data imply the possibility of hidden segmental UPD in normal individuals.

Abbreviations: DSBs, double strand breaks; HR, homologous recombination; NHEJ, non-homologous end joining; UPD, uniparental disomy; hUPD, uniparental heterodisomy; iUPD, uniparental isodisomy; NGS, next-generation sequencing; WGS, whole-genome sequencing; LCLs, lymphoblastoid cell lines; SNPs, single-nucleotide polymorphisms; SNP6.0 array, Genome-Wide Human SNP Array 6.0; PartekGS, Partek Genomics Suite; INDELs, short insertions and deletions; SVs, structural variants; GATK, Genome Analysis Toolkit; CNVs, copy number variants; LTA, loss of transmitted allele; LOH, loss of heterozygosity; ROH, runs of homozygosity; QPCR, quantitative polymerase chain reaction; ESCs, embryonic stem cells;

Keywords: Human genome, Genomic integrity, DNA repair, Gene conversion, International HapMap Project, 1000 Genomes Project

1. Introduction

Uniparental disomy (UPD) is defined as the inheritance of a chromosome pair derived only from one parent (Engel, 1980). Chromosomal UPD can occur because of gamete complementation, trisomic rescue, monosomic rescue and postfertilization error (Robinson, 2000). Uniparental heterodisomy (hUPD) is defined as the inheritance of both homologous chromosomes from one parent and occurs when bivalent chromatids fail to separate during meiosis I. Uniparental isodisomy (iUPD) is defined as the inheritance of two copies of one chromosome from one parent and may occur when sister chromatids fail to separate during meiosis II. The region of UPD may extend over an entire or segmental (interstitial or telomeric) chromosome. Segmental UPD is defined as UPD of one part of a chromosome (Kotzot, 2008), and occurs due to postzygotic somatic recombination between maternal and paternal homologues (Kotzot, 2008). Problems associated with UPD include aberrant genomic imprinting and homozygosity of autosomal recessively inherited mutations.

To maintain genome integrity, cells repair DNA damage including DNA double strand breaks (DSBs), by one of two major pathways, non-homologous end-joining (NHEJ) and homologous recombination (HR) (Wyman and Kanaar, 2006). NHEJ repair performs error-prone repair by joining DNA ends directly, independent of extensive DNA sequence homology, while HR repair performs error-free repair by utilizing the undamaged homologous sequence as the template for repair (Hartlerode and Scully, 2009). DNA damage during DNA replication can be repaired by HR using the intact sister chromatid (Sonoda et al., 2006) and inter-sister chromatid HR during S phase will

not result in segmental iUPD. However, several imprinting disorders such as Beckwith-Wiedemann syndrome (BWS; OMIM #130650), Prader Willi syndrome (PWS; OMIM #176270), Angelman syndrome (AS; OMIM #105830) can be caused by UPD. In BWS almost all patients with UPD have segmental UPD, in contrast, in PWS/AS patients mostly have UPD of the whole chromosome. In addition to those imprinting disorders, recessive hereditary disease can be caused by segmental iUPD (Kotzot, 2001; Pérez et al., 2011). Because segmental iUPD can be found in some disorders, it is possible that segmental UPD can occur in normal development without any disease phenotype. Segmental iUPD could be considered the signature of HR between maternal and paternal homologues during the early stages of embryogenesis. UPD can be detected using microsatellite analysis (Hannula et al., 2000) and methylation testing (Baumer et al., 2001), based on a limited number of markers in the chromosomal region of interest. The advent of high throughput single nucleotide polymorphism (SNP) microarray technology has recently permitted the identification of UPD in DNA samples from clinically affected individuals (Altug-Teber et al., 2005; Pérez et al., 2011), and the number of UPD case reports are increasing (Pérez et al., 2011). To assess the clinical significance of UPD, it is necessary to document the frequency and nature of UPD in the general population. Recently, several studies reported mosaic genomic variations (copy-neutral loss of heterozygosity (LOH) or acquired UPDs, trisomies and CNVs) in blood and buccal genomic DNA samples from cancer cases and controls (Jacobs et al., 2012; Laurie et al., 2012; Rodríguez-Santiago et al., 2010). However, assessing the segmental UPD in general populations using trios

and genome-wide SNP array has not been performed to date.

Two thousand cases of UPD have been reported thus far (http://www.fish.uniklinikum-jena.de/UPD.html). UPD is one of the causes of "imprinting disorders" and is found at a high rate (7% for AS and 25% for PWS: Amor and Halliday, 2008). BWS has segmental UPD11p in 20% of cases (Amor and Halliday, 2008). Until 2010, 122 cases were reported as segmental UPD, and ~65% of those cases were due to BWS and segmental paternal UPD 11p (Liehr, 2010). However, segmental UPD of other chromosomes not associated with a cytogenetically abnormal karyotype is extremely rare (Kotzot, 2001).,and UPD has no effect on phenotype at many chromosomal region. Although UPD cases without clinical abnormalities have been reported in the literature, they were found by chance or were due to repeated abortions in a family with chromosomal rearrangement (Liehr, 2010). Thus, despite the increasing importance of UPD as a disease causing mechanism, the precise UPD rate, including segmental UPD, in the general population is unknown.

Little information is available regarding DNA repair in the early development of zygotes. But it is clear that segmental iUPD detected systemically in adult can be the result from inter-allelic HR during the postzygotic period to the early embryonic stage. Therefore, we attempted to identify segmental iUPD in individuals without an abnormal phenotype. To this aim, we analyzed parent-offspring trios from SNP microarray data and also whole-genome sequencing (WGS) data of genomic DNA from two trios derived from lymphoblastoid cell lines (LCLs) during the pilot 2 data of the 1000 Genomes Project (http://www.1000genomes.org/) (Altshuler et al., 2010). WGS data

129 was used to identify shorter iUPD, because it is difficult to identify shorter segmental
130 iUPD by SNP microarray due to limited SNP information of the whole genome.
131 In this paper, we evaluated the frequency of UPD in healthy normal development.
132

133 **2. Materials and methods**

134 **2.1. HapMap 3 samples**

135 We downloaded and studied a set of 170 trios (510 samples) data from SNP6.0 arrays
136 from 5 populations in HapMap 3
137 (ftp://ftp.ncbi.nlm.nih.gov/hapmap/raw_data/hapmap3_affy6.0/); 159 individuals from
138 the Centre d'Etude du Polymorphisme Humain collected in Utah, USA, with ancestry
139 from northern and western Europe (CEU); 33 Africans with ancestry in the
140 southwestern USA (ASW); 81 Maasai in Kinyawa, Kenya (MKK); 174 Yoruba in
141 Ibadan, Nigeria (YRI); and 63 Mexicans with ancestry in Los Angeles, California
142 (MXL) (Supplementary Table 1).
143

144 **2.2. Genomic DNA**

145 We attempted to identify UPD in 173 trios. Three trios (original trio 1, trio 2 and trio 3)
146 in this study were Japanese (JPT) and were healthy volunteers (not included in HapMap
147 samples). The three trio's genomic DNA was extracted from peripheral blood following
148 standard protocols. Genomic DNA for two HapMap trios (CEU family ID (FID) 1463
149 and YRI Y117 trio) was obtained from the Coriell Institute
150 (http://ccr.coriell.org/sections/collections/NHGRI/?SsId=11).

151

152 **2.3. Microarray analysis**

153 We performed high-resolution genome-wide SNP genotyping and DNA copy number
154 detection using Genome-Wide Human SNP Array 6.0 (SNP6.0 array) following the
155 manufacturer's instructions (Affymetrix, Inc., Santa Clara, California, USA).
156 Genotyping were performed using the default parameters in the Birdseed v2 algorithm
157 of Genotyping Console (GTC) 4.1 software (Affymetrix). As a quality control for the
158 genotyping, Contrast QC values were calculated as implemented in the GTC 4.1, and
159 samples used passed the recommended values for contrast QC > 0.4. Genomic positions
160 of the SNPs corresponded to the March 2006 human genome (hg18). Copy number and
161 allele ratio analysis was performed by Partek Genomics Suite (PartekGS) version 6.5
162 (Partek Inc., St. Louis, Missouri, USA). For 3 trios of healthy volunteers and 170 trios
163 from HapMap, the copy number reference generated from the intensities of 20 normal
164 sample profiles in our laboratory and 100 HapMap sample profiles (no overlapping 170
165 trios) were used, respectively. The Hidden Markov Model (HMM) method was used to
166 detect amplified or deleted regions using PartekGS with default parameters, and
167 required at least 5 genomic markers to obtain CNVs call. We considered the 27 possible
168 combinations of genotypes when each of the mother/father/child in a trio had a biallelic
169 genotype (Supplementary Table 2). UPD genotypes were identified using in-house
170 Ruby script from trio genotyping information exported from GTC. A UPD region was
171 defined as a set of consecutive SNPs, where all plots had the same type (paternal and
172 maternal UPD segment) and occurred along a chromosome. We used the criteria of a

minimum of 6 consecutive UPD SNPs, with segments extending over 200 kilo bases (kb). In this study, we focused on the autosome, and chromosome X only when the offspring in the trio was a daughter. We visualized tracts of paternal uniparental inheritance (UPI-P), maternal uniparental inheritance (UPI-M), biparental inheritance (BPI), MI-S, single Mendelian inconsistencies (MI-S), double Mendelian inconsistencies (MI-D) and not informative (NI) in biallelic SNP data from trios using PartekGS'SNPtrio. Current software does not distinguish between homozygous and hemizygous states. In addition, it is known that UPD type genotypes can result from the loss of transmitted allele (LTA) (Ting et al., 2007). LTA was defined as a phenomenon in which the transmitted allele is lost (due to deletion or UPD) in the parent after the transmission to a normal child (Redon et al., 2006). Therefore, the putative UPD genotype overlapping with CNVs in trios confirmed using BEDtools (version 2.12.0) (Quinlan and Hall, 2010). The distinction between the segmental UPD as opposed to homozygosity due to small deletion, is difficult to determine just by inspection of the SNP array data alone. To exclude false segmental UPD due to undetectable small CNVs, we adopted a cutoff value of a length of 200 kb or smaller. Finally, we confirmed whether known imprinting genes were present in the identified segmental UPD region. Imprinting genes are based on Geneimprint (http://www.geneimprint.com/site/genes-by-species.Homo+sapiens.imprinted-All)

2.4. Next-generation sequencing (NGS) data

HapMap CEU 1463 and YRI Y117 trios were sequenced using multiple platforms, as

195 described elsewhere (Altshuler et al., 2010). We downloaded BAM files (aligned to the
196 NCBI36 reference genome using Maq v0.7) of two trios (CEU and YRI) sequenced
197 using Illumina Genome Analyzer I, II and IIx in the 1000 Genomes Project pilot 2
198 (ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/pilot_data/data/) with high coverage. We
199 focused on the autosomal and X chromosomes. Each included one offspring (daughter),
200 father and mother: CEU daughter NA12878, father NA12891 and mother NA12892;
201 and YRI daughter NA19240, father NA19239 and mother NA19238.
202
203 **2.5. NGS Bioinformatics**
204 After downloading the BAM files, duplicate reads from samples were identified and
205 removed using Picard (version 1.38) (http://picard.sourceforge.net/). Base quality scores
206 were recalibrated and reads were locally realigned with the Genome Analysis Toolkit
207 (GATK) (version 1.0.5974) (DePristo et al., 2011; McKenna et al., 2010). Coverage
208 statistics were calculated as default using GATK'DepthOfCoverageWalker. The diploid
209 consensus sequences and variants for autosomal and X chromosomes were obtained by
210 the 'EMIT_ALL_CONFIDENT_SITES (using -stand_call_conf 50.0 and
211 -stand_emit_conf 10.0)' command of the GATK'UnifiedGenotyper. SNPs and short
212 insertions and deletions (INDELs) were detected with the GATK's UnifiedGenotyper
213 according to the Best Practice Variant Detection with the GATK v2
214 (http://www.broadinstitute.org/gsa/wiki/index.php/Best_Practice_Variant_Detection_wi
215 th_the_GATK_v2). SNPs and INDELs were then filtered for the removal of low quality
216 variants with GATK's VariantFiltrationWalker tools. We filtered out any SNPs

matching the following criteria: (1) greater than 10% of aligned reads included at the site have a mapping quality of 0 (MAPQ0), or (2) overlaps INDELs, or (3) DP > 100 || MQ0 > 40 || SB > −0.10. We filtered out any INDELs matching the following criteria: (1) greater than 10% of aligned reads included at the site have a mapping quality of 0 (MAPQ0), or (2) SB >= -1.0, (3) QUAL < 10. Identified SNPs were annotated based on the dbSNP132 with ANNOVAR (Wang et al., 2010). Once the trio genotypes were determined, we extracted any iUPD genotypes that did not comply with the rules of Mendelian inheritance.

Filters were applied to exclude genomic regions in which false positive iUPD calls might be picked up. Since some genome regions are problematic for mapping and assembly, including regions of CNV in the each daughter, a putative iUPD call was not attempted in these regions (Altshuler et al., 2010; Conrad et al., 2011). We used the following filters: Simple Repeats, Segmental Duplications, CNV regions (Conrad et al., 2010; Kidd et al., 2008; McCarroll et al., 2008; Mills et al., 2011), and read depth (sites where at least one trio member has no mapped Illumina reads). BEDTools was used to confirm the intersections between putative iUPD genotypes and above-mentioned regions (Quinlan and Hall, 2010). Other annotations are based on The National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/) and The University of California Santa Cruz (UCSC; http://genome.ucsc.edu/) databases. Finally, we required each genotype in a trio to have qualities GQ40 or greater for more efficient identification of the true iUPD genotypes.

239 **2.6. Capillary sequencing**

240 Validation experiments were performed on the DNA extracted from LCLs in each trio
241 by a standard capillary sequencing approach. For CEU 1463 and YRI Y117 trios,
242 primers were designed for 140 and 178 sites, respectively. We designed PCR primers
243 using PrimerZ (http://genepipe.ngc.sinica.edu.tw/primerz/beginDesign.do) (Tsai et al.,
244 2007) or Primer3Plus
245 (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) (Untergasser et al.,
246 2007). Primers for each data set are provided in Supplementary Table 3.
247

248 **2.7. Quantitative polymerase chain reaction (qPCR)**

249 qPCR analysis was performed to measure the genomic copy number using a
250 LightCycler 480 (Roche Diagnostics, Basel, Switzerland) and the Thunderbird SYBR
251 qPCR Mix (Toyobo Co., Ltd.) according to the manufacturer's experimental protocol.
252 Two sets of primers, zinc finger protein 80 (*ZNF80*) and G protein-coupled receptor 15
253 (*GPR15*) (D'haene et al., 2010), were used as references for quantification. Data
254 analysis was performed with the second derivative maximum method of LightCycler
255 480 software (version 1.5.0.39) (Roche Diagnostics). qPCR amplification was carried
256 out in triplicate. Primers for target regions were designed to surround the putative iUPD
257 genotype by PrimerZ. Primers for each data set are provided in Supplementary Table 3.
258

3. Results

3.1. A whole chromosomal and segmental UPD analysis in 173 trios using SNP6.0 array

To investigate whole chromosome and segmental UPD in general populations using SNP6.0 array data, we examined the genotypes of the 173 trios that included 3 JPT trios in Nagasaki and 170 HapMap trios. Screening of UPD segments identified 46 putative segments (Table 1). A whole chromosomal UPD was not found in any chromosome except the Y chromosome in all samples tested. To rule out false segmental UPD due to CNVs and LTAs, we performed CNV analysis (Supplementary Table 4) and then cross-referenced with regions of putative segmental UPD in each trio (Table 1). As a result, we identified 24 CNVs, 21 LTAs (18 results from CNV and 3 possible copy number neutral LOH in the investigated parent's genome) (Supplementary Fig. 1, 2 and 3) and 1 obvious segmental iUPD (Table 1). This one segmental iUPD indicated a paternal iUPD range from p-terminal to physical position 8,202,065 on chromosome 17p13.3-13.1 (about 8.2 mega bases (Mb)) in NA19918 (HapMap ASW FID 2431) (Fig. 1).

3.2. Base calling and detection of iUPD genotype

We investigated the possibility of the shorter segmental iUPD being undetectable by SNP6.0 array in the human genome using sequence data with a high coverage by Illumina platform during the pilot phase 2 of the 1000 Genomes Project. For each of the trios, we called the genotype of the three genomes independently using the GATK

281 framework. In the CEU trio (NA12878, NA12891 and NA12892), mapped sequence
282 coverage of 31.9×, 30.3× and 25.6×, respectively, and 2.32, 2.33 and 2.31 gigabases
283 (Gb) of accessible genome included 2.85, 2.85, 2.79 million SNPs. In the YRI trio
284 (NA19240, NA19239 and NA19238), mapped sequence coverage of 33.4×, 24.5× and
285 20.6×, respectively, and 2.36, 2.30 and 2.23 Gb of accessible genome included 3.60,
286 3.40 and 3.10 million SNPs. The accessible genome per CEU and YRI trio set were
287 2.24 Gb and 2.14 Gb, respectively. Statistics for each data set are provided in Table 2.
288 Of these accessible genomes in each trio set, in the CEU 1463 and YRI Y117 trios,
289 1,094 and 1,474 putative iUPD genotypes were selected, respectively (Fig. 2). To
290 exclude false iUPD genotypes, we filtered out the putative iUPD genotypes overlapping
291 with regions of the simple repeats and segmental duplications and previously reported
292 CNVs in the trio's daughter (Supplementary Table 5 and 6). As a result, we identified
293 502 and 965 putative iUPD genotypes in the CEU 1463 and YRI Y117 trio, respectively
294 (Fig. 2).
295

296 **3.3. GQ threshold and filtering for iUPD genotypes**

297 Our approach was simple and would allow false iUPD candidates in the initial screening.
298 Therefore, of 502 putative iUPD genotypes in the CEU 1463 trio, 100 candidate sites
299 (300 genotypes in the trio) were selected at random, and validated by capillary
300 sequencing on the LCLs DNA. We used this data to estimate the accuracy of the
301 genotype and to determine the threshold quality more efficiently for identification of the
302 true iUPD genotype. Of the 300 validated genotypes, the correct and incorrect

303 genotypes were 189 (63%) and 111 (37%), respectively, and true iUPD genotype was
304 not confirmed (Supplementary Table 7). For more efficient screening, we focused on
305 genotype quality (GQ), encoded as a Phred quality and read depth (DP) at genotype
306 position. The 300 genotypes validated had a mean GQ of 71.13 (from a minimum of
307 1.61 to a maximum of 99.00) and a mean DP of 31.79 (from a minimum of 8.00 to a
308 maximum of 75.00), respectively. Studying the relationship between GQ and accuracy
309 of the genotypes with GQ10 or more, the correct genotype rate was 64.9% (189/291),
310 72.5% (182/251) with GQ40 or more, 91.0% (162/178) with GQ60 or more and 99.3%
311 (150/151) with GQ80 or more. Thus, a higher GQ showed a higher reliability
312 (Supplementary Fig. 4A). In contrast, increasing DP simply did not have much power to
313 remove incorrect genotypes (Supplementary Fig. 4B). Furthermore, the majority of false
314 positives for putative iUPD genotypes arose from an inaccuracy of genotyping in any
315 one of the trio (81.1%, 90/111). Therefore, we required all genotypes in the trio with
316 GQ40 or greater for identification of the true iUPD genotype. After filtering with a
317 threshold GQ40, we identified 100 and 178 putative iUPD genotypes in the CEU 1463
318 and YRI Y117 trio, respectively (Fig. 2, Supplementary Table 8 and 9).
319
320 **3.4. Validation of the putative iUPD genotypes by capillary sequencing and qPCR**
321 We attempted to validate these candidates by capillary sequencing. Of these, only 1
322 putative iUPD genotype (Validation ID C1383 and Y3887, respectively), in the CEU
323 1463 and YRI Y117 trio was confirmed as a true iUPD genotype (Fig. 2, Supplementary
324 Table 8 and 9, Supplementary Fig. 5A and B). Although iUPD candidates were not

present in the known CNVs regions in the daughter, qPCR analysis with DNA from each trio was performed with primers C1383 and Y3887 to confirm the copy number on the putative iUPD loci. The results revealed a deletion on the C1383 locus in the daughter (NA12878) and mother (NA12892). Similarly, the results revealed a deletion on the Y3887 locus in the daughter (NA19240) and father (NA19239) (Supplementary Fig. 5C). In our investigation, we could not identify shorter segmental iUPD by SNP6.0 array in the daughters from the two trios (Fig. 2).

3.5. Genes in identified segmental UPD regions in normal individuals

Finally, we identified one segmental paternal iUPD on 17p13.3-13.1 from 173 individuals. This segmental UPD region was included in the 233 RefSeq genes (Supplementary Table 10), but which are not “imprinted genes”. According to the conventional concept, UPD has no practical impact on phenotypes with the exception of the disruption of imprinting and homozygosity for recessive mutations.

4. Discussion

At any stage of the life cycle, from gamete formation to fetal post-natal life, exposure to genotoxic stress may affect the genomic integrity and fate of the organism (Jaroudi and SenGupta, 2006; Vinson and Hales, 2002). In undifferentiated cells, such as the embryo and progenitor cells, mutations are propagated to multiple differentiated cell types within the organism. Therefore, undifferentiated cells would require error-less repair mechanisms. HR would be a suitable repair mechanism for such cells, because intact homologous chromosomes are used as repair templates. Indeed, embryonic stem cells (ESCs) repair DSBs more frequently using the error-free HR pathway rather than the error-prone NHEJ (Tichy, 2011; Tichy and Stambrook, 2008). HR (also called gene conversion) can occur between sister chromatids, homologous chromosomes or homologous sequences on either the same chromatid or different chromosomes (Chen et al., 2007). Although the extent of genetic loss is minimal if HR results in a non-crossover gene conversion, crossover gene conversion leads to iUPD of the large region of the chromosome in daughter cells (Moynahan and Jasin, 1997 and 2010; Stark and Jasin, 2003). The occurrence of inter-allelic HR causing human inherited disease is rare (Chen et al., 2007). To our knowledge, homozygous nonsense mutations due to inter-allelic HR have been reported in a patient with campomelic dysplasia (Y440X) in SRY-box 9 (SOX9) (Pop et al., 2005). This case indicates that inter-allelic HR in early stage embryogenesis can occur.

To assess the possibility that inter-allelic HR occurs in the human genome during the period between postzygotic cells and the early embryonic stage to maintain the higher

362 fidelity of genomic integrity, we investigated the traits of iUPD genotypes using NGS
363 data during the pilot phase 2 of the 1000 Genomes Project. However, we could not find
364 direct evidence of segmental iUPD after the accurate reconfirmation process including
365 capillary sequencing and qPCR. Some parts of the reference sequence are inaccessible
366 because of high-copy repeats or segmental duplications. This is a limitation of the
367 current NGS technology producing short sequence reads. Indeed, 20% of the reference
368 genome was inaccessible in the trio project (Altshuler et al., 2010). From our data, the
369 accessible genome per CEU and YRI trio set were 2.24 Gb and 2.14 Gb, respectively
370 (Table 1). Because the total length of the human reference genome, including the gap
371 was composed of about 3.08 Gb, 27.2% and 30.5% of data in CEU and YRI trio,
372 respectively, were not analyzed in this study. Furthermore, the use of only two trios
373 might be too small a scale and low-level mosaicism is often difficult to detect accurately.
374 However, the data presented here provides evidence that segmental UPD during normal
375 development could not be a constitutive event in order to maintain genomic integrity.
376 Constitutive UPD is very rare. Robinson et al. determined that UPD for an average
377 chromosome occurs in 1/80,000 births (0.00125%) and UPD for any chromosome can
378 be expected in roughly 1/3,500 births (0.02857%), based on the frequency of UPD15
379 (Robinson, 2000). Liehr suggested that the rate of UPD in human population might be
380 even lower than 1 in 5,000 or less (Liehr, 2010). We studied 173 trios using
381 genome-wide SNP array and WGS data using NGS, and identified one case with
382 segmental iUPD. Segmental UPD for any chromosome can be expected in 1/173 births
383 which equals a rate of 0.57803%. Based on the investigated autosomal chromosome

pairs, we estimate the rate of segmental UPD to be one per 3806 chromosome pairs that equals a rate of 0.02627%. We found a higher frequency of UPD events than the previously reported frequency by Robinson et al and Liehr. These data imply the possibility of hidden segmental UPD in normal individuals. However, we found just only one UPD in 3806 chromosome pairs, we need analyze more trio samples and that would give the accurate rate of whole chromosomal and/or segmental UPD.

iUPD resulting from a somatic recombination can cause LOH. Somatic recombination leading to mosaic segmental UPD could occur in any individual and it is likely to be mosaic or in a heterogeneous cell population with increased cell division. In fact, the studies by Laurie et al. and Jacobs et al. found that detectable mosaic genomic variations including segmental UPD were rare (1%) in adults younger than 50 but that its prevalence increased to 2-3% in individuals older than 70 (Jacobs et al., 2012; Laurie et al., 2012). We detected 21 LTAs over 200 kb in the process of UPD screening using SNP microarray (Table 1 and supplementary fig. 2 and 3). These genomic alterations may reflect that CNVs or segmental UPD result from somatic recombination in restricted soma (for example, in hematopoietic cells) or during cell culture, as with aging. Although most sample data analyzed here was derived from DNA of LCLs (170 trios from HapMap), we suggest that segmental UPD occurring in early developmental stages in individuals in the general population can be detected. However, we cannot totally negate the possibility that one segmental UPD identified in this study arose during passage in the artificial culture.

Studies of UPD have only been performed in cases with relevant phenotypic features

406 and included only a few markers. These facts suggest that researchers may overlook
407 UPD in normal development and miss shorter segmental UPD, because UPD of many
408 chromosomal regions results in no obvious abnormalities (Kotzot and Utermann, 2005;
409 Robinson, 2000). In addition, lethal genotypes due to UPD during early embryonic
410 development would be undetectable. We suggest that trio genome analysis with
411 enhanced sequence accuracy could provide new findings for the risk of recessive
412 disorders, because one mutant allele from one parent can be transmitted to a child and
413 result in a homozygous state due to iUPD. To the best of our knowledge, this is the first
414 systematic study over whole chromosomal and segmental UPD in the human genome
415 without abnormal phenotype using familial trios.
416

417 **5. Conclusions**

418 The current study assessed the presence of whole chromosome and segmental UPD in
419 general populations using genome-wide SNP microarray and WGS data. We provided
420 evidence that segmental UPD in normal development is not a constitutive event in order
421 to maintain genomic integrity. Although we identified one obvious segmental paternal
422 iUPD in one HapMap sample, we could not find direct evidence of shorter segmental
423 iUPD. This suggested three possibilities, 1) human cells repress the usage of inter-allelic
424 homologous sequences as a template for HR, even at the early embryonic stage, 2)
425 shorter iUPD segments are unidentifiable because of absent informative markers within
426 the limited short segment, 3) UPD could be present in inaccessible genome regions
427 when using current NGS with short reads. Investigation of segmental UPD in general

428 populations will help to expand our general understanding of normal development in
429 humans.

Appendices (Supplementary Information)

Supplementary data associated with this article can be found in the online version.

Supplementary Fig. 1. SNP6.0 array plots of hemizygous deletion on chromosome 1 in a child in the HapMap YRI trio (FID Y003). Using the set for the YRI trio Y003, a hemizygous deletion and UPI-M was observed in the child (NA18497) but not in both parents. A red box indicates a false UPD locus due to CNV. M, mother; F, father; C, child.

Supplementary Fig. 2. SNP6.0 array plots of LTA due to hemizygous deletion of chromosome 6 in HapMap CEU trio (FID 1423). Set for CEU trio 1423 observed a hemizygous deletion in the mother (NA11920) but not in the child or father. The pattern of MI-S, UPI-P and BPI is consistent with an interpretation of LTA with the loss of an allele in the mother. A red box indicates a false UPD locus due to LTA. M, mother; F, father; C, child.

Supplementary Fig. 3. SNP6.0 array plots of LTA due to putative LOH on chromosome 11 in HapMap CEU trio (FID 1463). We also detected two UPD genotype segments in NA12865 (CEU FID 1459) and one UPD genotype segment in NA12877 (CEU FID 1463) that were not CNVs in any of the individuals from the trio. However, these segments showed UPI-M, MI-S and BPI, and large contiguous long runs of homozygosity (ROH) in the fathers genome. ROH in the father indicated that the UPD

452 genotype might result from LTA due to copy number neutral LOH, but not CNV. A red
453 box indicates the cluster of UPI-M, MI-S and BPI genotypes. Arrowhead indicates
454 ROH in the father. We did not consider these regions as UPD segments in this study.
455 The data on chromosome 11 from CEU trio 1463 is shown as a representative example.
456 M, mother; F, father; C, child.
457
458 **Supplementary Fig. 4.** GQ and DP in correct and incorrect genotypes confirmed by
459 capillary sequencing in CEU 100 putative iUPD genotypes (300 genotypes). (A) The
460 number of correct and incorrect genotypes falling within each genotype quality (GQ)
461 score threshold are shown on the bar. (B) The number of correct and incorrect
462 genotypes falling within each read depth (DP) score threshold are shown on the bar. GT,
463 genotype.
464
465 **Supplementary Fig. 5.** Results of putative iUPD validated by capillary sequencing and
466 qPCR. (A) Result of capillary sequencing and genotypes registered in Personal Genome
467 Variants on UCSC in NA12878, NA12891 and NA12892 at candidate locus C1383
468 (NA19892 is not registered in Personal Genome Variants on UCSC). The genotype of
469 putative iUPD C1383 site in NA12878 on UCSC was incorrect. Electropherograms of
470 DNA sequences in the CEU trio show a paternal iUPD genotype in NA12878
471 (daughter). (B) Result of capillary sequencing and genotypes registered in Personal
472 Genome Variants on UCSC in NA19240, NA19239 and NA19238 at candidate locus
473 Y3887 (NA19239 and NA19238 are not registered in Personal Genome Variants on

474 UCSC). Electropherograms of DNA sequences in candidate locus Y3887 from the YRI
475 trio show the maternal iUPD genotype in NA19240 (daughter). (C) The daughter had a
476 microdeletion. qPCR was performed with primers C1383 and Y3887 for putative iUPD,
477 and with primers *ZNF80* and *GPR15*, respectively, as standards. C1471 and Y3350
478 primers demonstrated the known deletion control (Kidd et al., 2008; Mills et al., 2011).
479 Normalized mean values for triplicates are shown for each interest target versus *ZNF80*
480 (blue) and *GPR15* (red), respectively.
481
482 **Supplementary Table 1.** Samples from the UPD study. FID, family ID; IID, individual
483 ID.
484
485 **Supplementary Table 2.** The 27 possible combinations of genotypes in a trio.
486
487 **Supplementary Table 3.** Primers for capillary sequencing and qPCR. This table lists
488 the primer sets used for this analysis. CHR, chromosome; POS, position; REF, reference
489 base.
490
491 **Supplementary Table 4.** Identified CNVs in each trio by SNP6.0 array analysis.
492
493 **Supplementary Table 5.** Collating and filtering of putative iUPD genotypes and
494 problematic regions for mapping and assembly of regions or previous reported CNVs in
495 the daughter from the CEU trio 1463. CHR, chromosome; POS, position; REF,

reference base; ALT, alternative base; GT, genotype; GQ, genotype quality; DP, read depth; CNV, copy number variant; ND, not detectable; SR, simple repeat; SD, segmental duplication.

Supplementary Table 6. Collating and filtering of putative iUPD genotypes and problematic regions for mapping and assembly or regions of previous reported CNVs in the daughter from the YRI trio Y117. CHR, chromosome; POS, position; REF, reference base; ALT, alternative base; GT, genotype; GQ, genotype quality; DP, read depth; CNV, copy number variant; ND, not detectable; SR, simple repeat; SD, segmental duplication.

Supplementary Table 7. Summary of validation of 100 putative iUPD genotypes (300 genotypes) in the CEU 1463 trio by capillary sequencing of LCLs DNA. CHR, chromosome; POS, position; REF, reference base; ALT, alternative base; GT, genotype; GQ, genotype quality; DP, read depth.

Supplementary Table 8. Putative iUPD genotypes with GQ40 or greater of each genotype in the CEU 1463 trio. CHR, chromosome; POS, position; REF, reference base; ALT, alternative base; GT, genotype; GQ, genotype quality; DP, read depth; FP, false positive; ND, not detectable.

Supplementary Table 9. Putative iUPD genotypes with GQ40 or greater of each

genotype in the YRI Y117 trio. CHR, chromosome; POS, position; REF, reference base; ALT, alternative base; GT, genotype; GQ, genotype quality; DP, read depth; FP, false positive; ND, not detectable.

Supplementary Table 10. RefSeq Genes in identified segmental iUPD region 17p13.3-13.1.

Acknowledgements

We express our gratitude to the families for their participation in this research and the anonymous HapMap families for contributing samples for research. We also thank Ms Chisa Hayashida for technical assistance. K.Y. was supported by a Grant-in-Aid for Challenging Exploratory Research (No. 22659071) from the Japan Society for the Promotion of Science.

532 **References**

533

534 Altshuler, D., et al., 2010. A map of human genome variation from population-scale
535 sequencing. Nature 467, 1061-73.
536 Altug-Teber, O., et al., 2005. A rapid microarray based whole genome analysis for
537 detection of uniparental disomy. Hum Mutat. 26, 153-9.
538 Amor, DJ., Halliday, J., 2008. A review of known imprinting syndromes and their
539 association with assisted reproduction technologies. Hum Reprod. 23, 2826-34.
540 Baumer, A., et al., 2001. A novel MSP/DHPLC method for the investigation of the
541 methylation status of imprinted genes enables the molecular detection of low cell
542 mosaicisms. Hum Mutat. 17, 423-30.
543 Chen, JM., et al., 2007. Gene conversion: mechanisms, evolution and human disease.
544 Nat Rev Genet. 8, 762-75.
545 Conrad, DF., et al., 2010. Origins and functional impact of copy number variation in the
546 human genome. Nature 464, 704-12.
547 Conrad, DF., et al., 2011. Variation in genome-wide mutation rates within and between
548 human families. Nat Genet. 43, 712-4.
549 D'haene, B., et al., 2010. Accurate and objective copy number profiling using real-time
550 quantitative PCR. Methods 50, 262-70.
551 DePristo, MA., et al., 2011. A framework for variation discovery and genotyping using
552 next-generation DNA sequencing data. Nat Genet. 43, 491-8.
553 Engel, E., 1980. A new genetic concept: uniparental disomy and its potential effect,

554 isodisomy. Am J Med Genet. 6, 137-43.
555 Hannula, K., et al., 2000. A narrow segment of maternal uniparental disomy of
556 chromosome 7q31-qter in Silver-Russell syndrome delimits a candidate gene region.
557 Am J Hum Genet. 68, 247-53.
558 Hartlerode, AJ., Scully, R., 2009. Mechanisms of double-strand break repair in somatic
559 mammalian cells. Biochem J. 423, 157-68.
560 Jacobs, KB., et al., 2012. Detectable clonal mosaicism and its relationship to aging and
561 cancer. Nat Genet. 44, 651-8.
562 Jaroudi, S., SenGupta, S., 2006. DNA repair in mammalian embryos. Mutat Res. 635,
563 53-77.
564 Kidd, JM., et al., 2008. Mapping and sequencing of structural variation from eight
565 human genomes. Nature 453, 56-64.
566 Kotzot, D., 2001. Complex and segmental uniparental disomy (UPD): review and
567 lessons from rare chromosomal complements. J Med Genet. 38, 497-507.
568 Kotzot, D., 2008. Complex and segmental uniparental disomy updated. J Med Genet. 45,
569 545-56.
570 Kotzot, D., Utermann, G., 2005. Uniparental disomy (UPD) other than 15: phenotypes
571 and bibliography updated. Am J Med Genet A. 136, 287-305.
572 Laurie, CC., et al., 2012. Detectable clonal mosaicism from birth to old age and its
573 relationship to cancer. Nat Genet. 44, 642-50.
574 Liehr, T., 2010. Cytogenetic contribution to uniparental disomy (UPD). Mol Cytogenet.
575 3, 8.

McCarroll, SA., et al., 2008. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet. 40, 1166-74.

McKenna, A., et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297-303.

Mills, RE., et al., 2011. Mapping copy number variation by population-scale genome sequencing. Nature 470, 59-65.

Moynahan, ME., Jasin, M., 1997. Loss of heterozygosity induced by a chromosomal double-strand break. Proc Natl Acad Sci U S A. 94, 8988-93.

Moynahan, ME., Jasin, M., 2010. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. Nat Rev Mol Cell Biol. 11, 196-207.

Pérez, B., et al., 2011. Segmental uniparental disomy leading to homozygosity for a pathogenic mutation in three recessive metabolic diseases. Mol Genet Metab. 105, 270-1.

Pop, R., et al., 2005. A homozygous nonsense mutation in SOX9 in the dominant disorder campomelic dysplasia: a case of mitotic gene conversion. Hum Genet. 117, 43-53.

Quinlan, AR., Hall, IM., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-2.

Redon, R., et al., 2006. Global variation in copy number in the human genome. Nature 444, 444-54.

Robinson, WP., 2000. Mechanisms leading to uniparental disomy and their clinical

598 consequences. Bioessays 22, 452-9.
599 Rodríguez-Santiago, B., et al., 2010. Mosaic uniparental disomies and aneuploidies as
600 large structural variants of the human genome. Am J Hum Genet. 87, 129-38.
601 Sonoda, E., et al., 2006. Differential usage of non-homologous end-joining and
602 homologous recombination in double strand break repair. DNA Repair (Amst) 5,
603 1021-9.
604 Stark, JM., Jasin, M., 2003. Extensive loss of heterozygosity is suppressed during
605 homologous repair of chromosomal breaks. Mol Cell Biol. 23, 733-43.
606 Tichy, ED., 2011. Mechanisms maintaining genomic integrity in embryonic stem cells
607 and induced pluripotent stem cells. Exp Biol Med (Maywood). 236, 987-96.
608 Tichy, ED., Stambrook, PJ., 2008. DNA repair in murine embryonic stem cells and
609 differentiated cells. Exp Cell Res. 314, 1929-36.
610 Ting, JC., et al., 2007. Visualization of uniparental inheritance, Mendelian
611 inconsistencies, deletions, and parent of origin effects in single nucleotide
612 polymorphism trio data with SNPtrio. Hum Mutat. 28, 1225-35.
613 Tsai, MF., et al., 2007. PrimerZ: streamlined primer design for promoters, exons and
614 human SNPs. Nucleic Acids Res. 35(Web Server issue), W63-5.
615 Untergasser, A., et al., 2007. Primer3Plus, an enhanced web interface to Primer3.
616 Nucleic Acids Res. 35(Web Server issue), W71-4.
617 Vinson, RK., Hales, BF., 2002. DNA repair during organogenesis. Mutat Res. 509,
618 79-91.
619 Wang, K., et al., 2010. ANNOVAR: functional annotation of genetic variants from

high-throughput sequencing data. Nucleic Acids Res. 38, e164.

Wyman, C., Kanaar, R., 2006. DNA double-strand break repair: all's well that ends well. Annu Rev Genet. 40, 363-83.

Web references

1000 Genomes Project pilot 2 data. ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/pilot_data/data/.

1000 Genomes Project. http://www.1000genomes.org/.

Best Practice Variant Detection with the GATK v2. http://www.broadinstitute.org/gsa/wiki/index.php/Best_Practice_Variant_Detection_with_the_GATK_v2.

Coriell Institute. http://ccr.coriell.org/sections/collections/NHGRI/?SsId=11.

Geneimprint. http://www.geneimprint.com/site/genes-by-species.Homo+sapiens.imprinted-All.

HapMap 3 raw data. ftp://ftp.ncbi.nlm.nih.gov/hapmap/raw_data/hapmap3_affy6.0/.

HapMap project. http://www.sanger.ac.uk/resources/downloads/human/hapmap3.html.

Liehr T, 2012. Cases with uniparental disomy. http://www.fish.uniklinikum-jena.de/UPD.html.

Picard (version 1.38). http://picard.sourceforge.net/.

Primer3Plus. http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi.

642 PrimerZ. http://genepipe.ngc.sinica.edu.tw/primerz/beginDesign.do.

643 The National Center for Biotechnology Information (NCBI).

644 http://www.ncbi.nlm.nih.gov/.

645 The University of California Santa Cruz (UCSC). http://genome.ucsc.edu/.

646

647

Figure Legends

Fig. 1.

Segmental paternal iUPD in HapMap ASW sample (NA19918). SNP6.0 data analyzed with PartekGS software shows the plots for the allele ratio, copy number state, and inheritance pattern by SNPtrio on chromosome 17 in HapMap ASW trio (FID 2431) (M, mother; F, father; C, child). (A) The allele ratio graph represents the genotypes for each individual single nucleotide polymorphism (SNP). Dots with a value of 1, −1, and 0 represent SNPs with AA, BB, and AB genotypes, respectively. (B) Plots represent chromosome copy number state (0.0 ~ 4.0). (C) SNPtrio displayed five classes of inheritance pattern. The five classes are 1) double Mendelian inconsistency (MI-D); 2) single Mendelian inconsistency (MI-S); 3) maternal uniparental inheritance (UPI-M); 4) paternal uniparental inheritance (UPI-P); 5) biparental inheritance (BPI). NI indicates not informative. The BPI plots represent the biparental inheritance SNPs, in which the parents have AA and BB calls and the child has an AB call. A red box indicates the segmental paternal iUPD locus.

Fig. 2.

Study design and summary of iUPD segment analysis using whole-genome sequencing (WGS) data of HapMap FID CEU 1463 and YRI Y117 trios, respectively. GQ, genotype quality; qPCR, quantitative polymerase chain reaction. *Previously reported CNV regions (Conrad et al., 2010; Kidd et al., 2008; McCarroll et al., 2008; Mills et al.,

670 2011).

671

672 **Table Legends**

673

674 **Table 1**

675 Summary of putative segmental UPD segments in 173 trios detected by SNP6.0 array

676 data analysis. Chr, chromosome; CNV, copy number variant; ND, not detectable; LTA,

677 loss of transmitted allele; iUPD, uniparental isodisomy; LOH, loss of herozygosity.

678

679 **Table 2**

680 Summary of alignment and base calling in two trios. AC+X: Autosomal chromosome

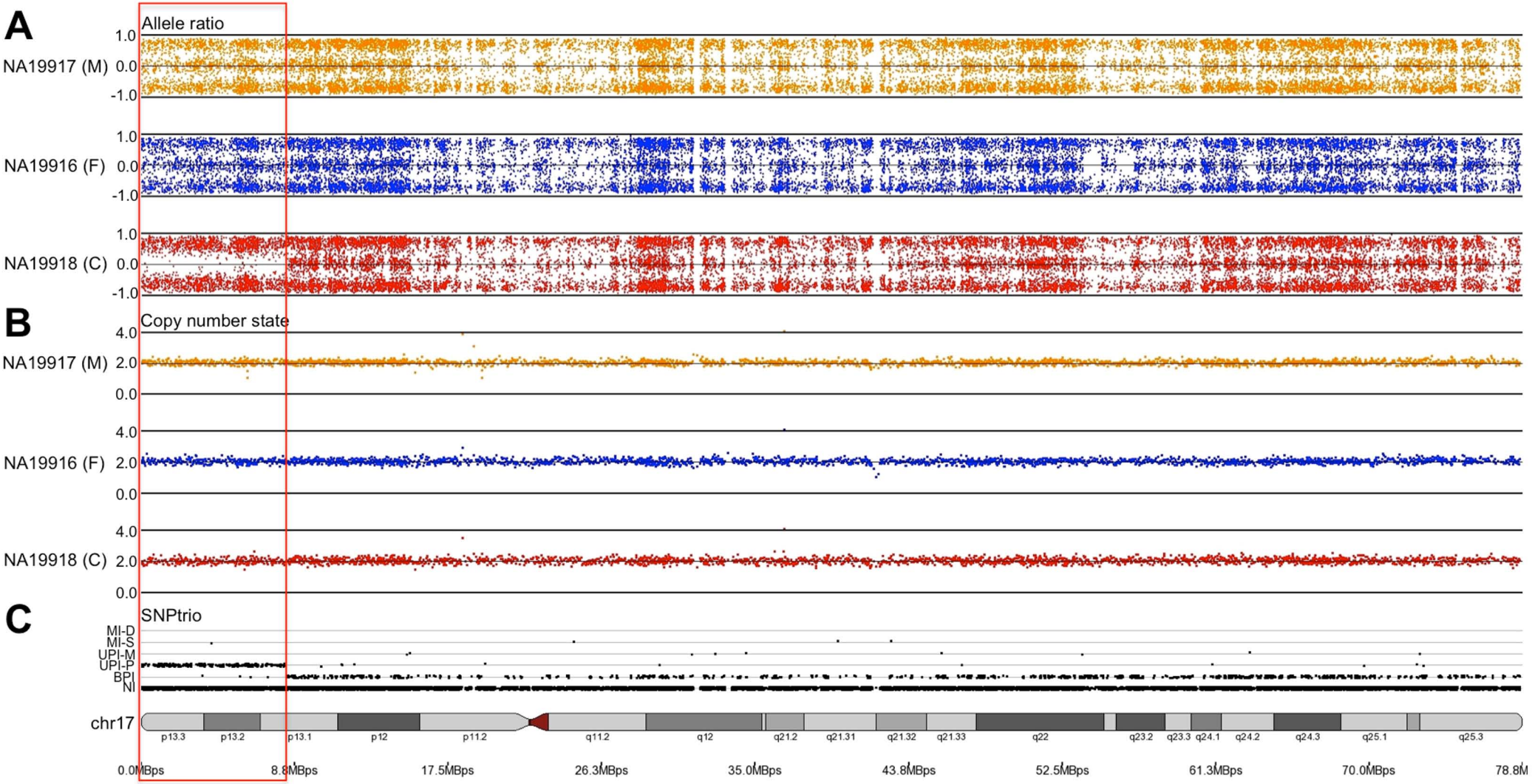
681 (1-22) and X chromosome (exclude gap) = total length 2,706,959,439 bases (about 2.71

682 Gb).

683

684

Fig. 1



Segmental paternal iUPD in 17p13.3-13.1

Fig. 2

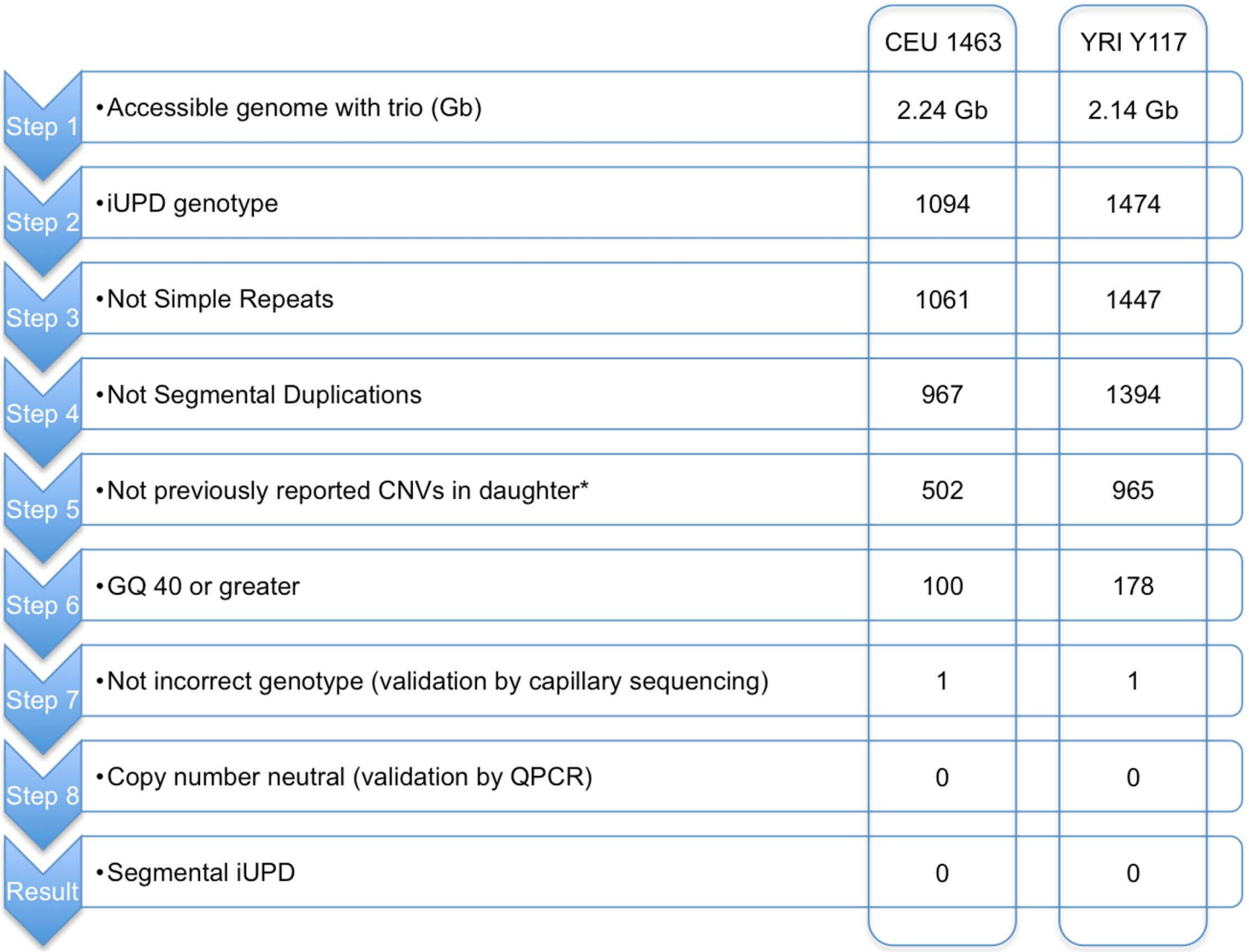

CEU 1463
YRI Y117
Step 1
•Accessible genome with trio (Gb)
2.24 Gb
2.14 Gb
Step 2
•iUPD genotype
1094
1474
Step 3
•Not Simple Repeats
1061
1447
Step 4
•Not Segmental Duplications
967
1394
Step 5
•Not previously reported CNVs in daughter*
502
965
Step 6
•GQ 40 or greater
100
178
Step 7
•Not incorrect genotype (validation by capillary sequencing)
1
1
Step 8
•Copy number neutral (validation by QPCR)
0
0
Result
•Segmental iUPD
0
0

Table 1

Chr	Start position	End position	Population	HapMap FID	Sex	UPD type	Mother CNV	Father CNV	Child CNV	Result	UPD probe number	Length (bp)
11	51,078,178	51,359,581	ASW	2368	XY	Paternal	CNV	ND	ND	LTA	7	281,404
17	6,689	8,202,065	ASW	2431	XY	Paternal	ND	ND	ND	Paternal iUPD	301	8,195,377
7	119,133,278	119,393,868	ASW	2427	XY	Maternal	ND	CNV	CNV	CNV	8	260,591
6	137,300,451	143,369,018	CEU	1423	XX	Paternal	CNV	ND	ND	LTA	122	6,068,568
5	107,513,060	107,716,753	CEU	1350	XY	Paternal	ND	ND	CNV	CNV	15	203,694
8	14,677,944	15,701,490	CEU	1375	XX	Paternal	CNV	ND	CNV	CNV	64	1,023,547
7	88,496,196	88,887,792	CEU	1330	XY	Paternal	CNV	ND	CNV	CNV	15	391,597
2	85,751,279	88,861,509	CEU	1330	XX	Paternal	CNV	ND	ND	LTA	34	3,110,231
12	129,502	131,942,726	CEU	1444	XY	Paternal	ND	ND	CNV	CNV	825	131,813,225
11	81,131,219	81,387,538	CEU	1447	XX	Paternal	CNV	ND	CNV	CNV	11	256,320
22	20,825,481	21,201,922	CEU	1459	XY	Paternal	CNV	ND	ND	LTA	7	376,442
X	28,498,460	31,437,190	CEU	1463	XX	Paternal	CNV	CNV	ND	LTA	34	2,938,731
4	118,785,685	119,509,766	CEU	1340	XX	Maternal	ND	CNV	ND	LTA	16	724,082
12	33,468,716	34,188,071	CEU	1345	XX	Maternal	ND	CNV	ND	LTA	9	719,356
22	20,784,680	21,191,527	CEU	1420	XX	Maternal	ND	CNV	ND	CNV	11	406,848
X	276,282	154,127,693	CEU	1349	XX	Maternal	ND	ND	CNV	CNV	2,170	153,851,412
15	21,205,648	45,731,444	CEU	1377	XY	Maternal	ND	CNV	ND	LTA	85	24,525,797
18	65,224,346	76,085,336	CEU	1328	XX	Maternal	ND	CNV	ND	LTA	369	10,860,991
22	20,927,130	21,243,931	CEU	1330	XY	Maternal	ND	CNV	ND	LTA	8	316,802
1	206,304,300	246,785,226	CEU	1330	XX	Maternal	ND	CNV	ND	LTA	144	40,480,927
17	69,586,313	70,111,013	CEU	13281	XY	Maternal	CNV	ND	CNV	CNV	12	524,701
X	140,182,100	140,575,068	CEU	1354	XX	Maternal	ND	CNV	CNV	CNV	18	392,969
22	20,718,086	21,107,920	CEU	1358	XY	Maternal	ND	ND	CNV	CNV	15	389,835
1	144,979,429	145,700,719	CEU	1459	XX	Maternal	ND	ND	ND	LTA (putative LOH in father)	51	721,291
1	236,789,304	246,590,204	CEU	1459	XX	Maternal	ND	ND	ND	LTA (putative LOH in father)	396	9,800,901
11	114,231,222	134,235,117	CEU	1463	XY	Maternal	ND	ND	ND	LTA (putative LOH in father)	543	20,003,896
13	82,217,462	83,042,185	MXL	M019	XX	Maternal	ND	ND	CNV	CNV	9	824,724
6	140,718,454	141,182,824	MXL	M027	XX	Maternal	ND	CNV	CNV	CNV	9	464,371
X	4,726,561	145,198,977	MKK	2596	XX	Paternal	CNV	CNV	ND	LTA	611	140,472,417

22	20,909,341	21,181,447	MKK	2699	XY	Paternal	ND	ND	CNV	CNV	7	272,107
11	55,060,441	55,440,561	MKK	2588	XX	Maternal	ND	CNV	CNV	CNV	8	380,121
5	110,517,276	110,787,436	MKK	2634	XY	Maternal	ND	CNV	CNV	CNV	17	270,161
9	11,947,750	12,155,758	MKK	2634	XY	Maternal	ND	CNV	CNV	CNV	11	208,009
22	24,042,173	24,292,988	MKK	2634	XY	Maternal	ND	CNV	ND	LTA	9	250,816
1	245,373,155	247,137,334	YRI	Y014	XX	Paternal	ND	ND	CNV	CNV	66	1,764,180
3	136,397,878	137,151,871	YRI	Y014	XX	Paternal	CNV	ND	ND	LTA	10	753,994
11	329,969	26,983,000	YRI	Y014	XX	Paternal	ND	ND	CNV	CNV	263	26,653,032
X	2,386,344	25,622,488	YRI	Y014	XX	Paternal	CNV	ND	ND	LTA	81	23,236,145
19	22,821,274	23,413,380	YRI	Y074	XY	Paternal	ND	ND	CNV	CNV	7	592,107
7	119,175,698	119,393,868	YRI	Y038	XX	Paternal	CNV	ND	CNV	CNV	8	218,171
12	73,201,200	91,388,277	YRI	Y112	XY	Paternal	CNV	ND	ND	LTA	70	18,187,078
1	22,392,010	28,325,476	YRI	Y003	XY	Maternal	ND	ND	CNV	CNV	97	5,933,467
15	20,318,185	20,773,725	YRI	Y009	XY	Maternal	ND	CNV	CNV	CNV	11	455,541
22	24,012,780	24,238,616	YRI	Y071	XY	Maternal	ND	CNV	CNV	CNV	10	225,837
13	18,759,817	19,002,511	YRI	Y039	XY	Maternal	ND	CNV	ND	LTA	9	242,695
2	151,462,078	153,347,758	YRI	Y048	XY	Maternal	ND	CNV	ND	LTA	36	1,885,681

Table 2

Family	CEU 1463			YRI Y117		
Sample	NA12878	NA12891	NA12892	NA19240	NA19239	NA19238
Relation	Daughter	Father	Mother	Daughter	Father	Mother
Total bases (Gb)	102.24 Gb	100.89 Gb	85.68 Gb	108.25 Gb	84.03 Gb	71.56 Gb
Mapped bases (Gb)	99.63 Gb	97.25 Gb	80.2 Gb	104.25 Gb	79.94 Gb	65.24 Gb
Total reads	2,507,012,490	2,264,396,064	2,051,935,811	2,738,304,812	2,296,647,842	1,971,737,379
Mapped reads	2,443,207,477	2,189,660,230	1,952,966,402	2,632,175,898	2,184,252,515	1,796,606,841
Mean mapped depth	31.9	30.3	25.6	33.4	24.5	20.6
Accessible genome (Gb)	2.32 Gb	2.33 Gb	2.31 Gb	2.36 Gb	2.30 Gb	2.23 Gb
Accessible genome (% of AC+X)	85.61 (%)	85.98 (%)	85.24 (%)	87.08 (%)	84.87 (%)	82.29 (%)
Accessible genome with trio (Gb)	2.24 Gb			2.14 Gb		
Accessible genome with trio (Gb) (% of AC+X)	82.66 (%)			78.97 (%)		
SNPs (N)	2,854,439	2,846,437	2,785,908	3,602,569	3,395,713	3,090,355
SNPs in dbSNP132 (N)	2,838,282	2,831,464	2,773,304	3,576,164	3,371,626	3,070,416
SNPs in dbSNP132 (%)	99.43 (%)	99.47 (%)	99.55 (%)	99.27 (%)	99.29 (%)	99.35 (%)