<table>
<thead>
<tr>
<th>Title</th>
<th>Uniparental disomy analysis in trios using genome-wide SNP array and whole-genome sequencing data imply segmental uniparental isodisomy in general populations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Sasaki, Kensaku; Mishima, Hiroyuki; Miura, Kiyonori; Yoshiura, Kohichiro</td>
</tr>
<tr>
<td>Citation</td>
<td>Gene, 512(2), pp.267-274; 2013</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2013-01-10</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10069/32451">http://hdl.handle.net/10069/32451</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2012 Elsevier B.V.; NOTICE: this is the author’s version of a work that was accepted for publication in Gene. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Gene, 512, 2(2013)</td>
</tr>
</tbody>
</table>

NAOSITE: Nagasaki University’s Academic Output SITE
http://naosite.lb.nagasaki-u.ac.jp
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; Rodriguez-Santiago et al., 2010). However, assessing the segmental UPD in general populations using trios
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
was used to identify shorter iUPD, because it is difficult to identify shorter segmental iUPD by SNP microarray due to limited SNP information of the whole genome.

In this paper, we evaluated the frequency of UPD in healthy normal development.

2. Materials and methods

2.1. HapMap 3 samples

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

2.2. Genomic DNA

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

We performed high-resolution genome-wide SNP genotyping and DNA copy number detection using Genome-Wide Human SNP Array 6.0 (SNP6.0 array) following the manufacturer’s instructions (Affymetrix, Inc., Santa Clara, California, USA). Genotyping were performed using the default parameters in the Birdseed v2 algorithm of Genotyping Console (GTC) 4.1 software (Affymetrix). As a quality control for the genotyping, Contrast QC values were calculated as implemented in the GTC 4.1, and samples used passed the recommended values for contrast QC > 0.4. Genomic positions of the SNPs corresponded to the March 2006 human genome (hg18). Copy number and allele ratio analysis was performed by Partek Genomics Suite (PartekGS) version 6.5 (Partek Inc., St. Louis, Missouri, USA). For 3 trios of healthy volunteers and 170 trios from HapMap, the copy number reference generated from the intensities of 20 normal sample profiles in our laboratory and 100 HapMap sample profiles (no overlapping 170 trios) were used, respectively. The Hidden Markov Model (HMM) method was used to detect amplified or deleted regions using PartekGS with default parameters, and required at least 5 genomic markers to obtain CNVs call. We considered the 27 possible combinations of genotypes when each of the mother/father/child in a trio had a biallelic genotype (Supplementary Table 2). UPD genotypes were identified using in-house Ruby script from trio genotyping information exported from GTC. A UPD region was defined as a set of consecutive SNPs, where all plots had the same type (paternal and 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’SNPtri. 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
described elsewhere (Altshuler et al., 2010). We downloaded BAM files (aligned to the
NCBI36 reference genome using Maq v0.7) of two trios (CEU and YRI) sequenced
using Illumina Genome Analyzer I, II and IIx in the 1000 Genomes Project pilot 2
(ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/pilot_data/data/) with high coverage. We
focused on the autosomal and X chromosomes. Each included one offspring (daughter),
father and mother: CEU daughter NA12878, father NA12891 and mother NA12892;
and YRI daughter NA19240, father NA19239 and mother NA19238.

2.5. NGS Bioinformatics

After downloading the BAM files, duplicate reads from samples were identified and
removed using Picard (version 1.38) (http://picard.sourceforge.net/). Base quality scores
were recalibrated and reads were locally realigned with the Genome Analysis Toolkit
(GATK) (version 1.0.5974) (DePristo et al., 2011; McKenna et al., 2010). Coverage
statistics were calculated as default using GATK’DepthOfCoverageWalker. The diploid
consensus sequences and variants for autosomal and X chromosomes were obtained by
the ‘EMIT_ALL_CONFIDENT_SITES (using -stand_call_conf 50.0 and
-stand_emit_conf 10.0)’ command of the GATK’UnifiedGenotyper. SNPs and short
insertions and deletions (INDELs) were detected with the GATK’s UnifiedGenotyper
according to the Best Practice Variant Detection with the GATK v2
(http://www.broadinstitute.org/gsa/wiki/index.php/Best_Practice_Variant_Detection wi
th_the_GATK_v2). SNPs and INDELs were then filtered for the removal of low quality
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.
2.6. Capillary sequencing
Validation experiments were performed on the DNA extracted from LCLs in each trio by a standard capillary sequencing approach. For CEU 1463 and YRI Y117 trios, primers were designed for 140 and 178 sites, respectively. We designed PCR primers using PrimerZ (http://genepipe.ngc.sinica.edu.tw/primerz/beginDesign.do) (Tsai et al., 2007) or Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) (Untergasser et al., 2007). Primers for each data set are provided in Supplementary Table 3.

2.7. Quantitative polymerase chain reaction (qPCR)
qPCR analysis was performed to measure the genomic copy number using a LightCycler 480 (Roche Diagnostics, Basel, Switzerland) and the Thunderbird SYBR qPCR Mix (Toyobo Co., Ltd.) according to the manufacturer’s experimental protocol. Two sets of primers, zinc finger protein 80 (ZNF80) and G protein-coupled receptor 15 (GPR15) (D'haene et al., 2010), were used as references for quantification. Data analysis was performed with the second derivative maximum method of LightCycler 480 software (version 1.5.0.39) (Roche Diagnostics). qPCR amplification was carried out in triplicate. Primers for target regions were designed to surround the putative iUPD genotype by PrimerZ. Primers for each data set are provided in Supplementary Table 3.
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
framework. In the CEU trio (NA12878, NA12891 and NA12892), mapped sequence coverage of 31.9×, 30.3× and 25.6×, respectively, and 2.32, 2.33 and 2.31 gigabases (Gb) of accessible genome included 2.85, 2.85, 2.79 million SNPs. In the YRI trio (NA19240, NA19239 and NA19238), mapped sequence coverage of 33.4×, 24.5× and 20.6×, respectively, and 2.36, 2.30 and 2.23 Gb of accessible genome included 3.60, 3.40 and 3.10 million SNPs. The accessible genome per CEU and YRI trio set were 2.24 Gb and 2.14 Gb, respectively. Statistics for each data set are provided in Table 2.

Of these accessible genomes in each trio set, in the CEU 1463 and YRI Y117 trios, 1,094 and 1,474 putative iUPD genotypes were selected, respectively (Fig. 2). To exclude false iUPD genotypes, we filtered out the putative iUPD genotypes overlapping with regions of the simple repeats and segmental duplications and previously reported CNVs in the trio’s daughter (Supplementary Table 5 and 6). As a result, we identified 502 and 965 putative iUPD genotypes in the CEU 1463 and YRI Y117 trio, respectively (Fig. 2).

### 3.3. GQ threshold and filtering for iUPD genotypes

Our approach was simple and would allow false iUPD candidates in the initial screening. Therefore, of 502 putative iUPD genotypes in the CEU 1463 trio, 100 candidate sites (300 genotypes in the trio) were selected at random, and validated by capillary sequencing on the LCLs DNA. We used this data to estimate the accuracy of the genotype and to determine the threshold quality more efficiently for identification of the true iUPD genotype. Of the 300 validated genotypes, the correct and incorrect
genotypes were 189 (63%) and 111 (37%), respectively, and true iUPD genotype was not confirmed (Supplementary Table 7). For more efficient screening, we focused on genotype quality (GQ), encoded as a Phred quality and read depth (DP) at genotype position. The 300 genotypes validated had a mean GQ of 71.13 (from a minimum of 1.61 to a maximum of 99.00) and a mean DP of 31.79 (from a minimum of 8.00 to a maximum of 75.00), respectively. Studying the relationship between GQ and accuracy of the genotypes with GQ10 or more, the correct genotype rate was 64.9% (189/291), 72.5% (182/251) with GQ40 or more, 91.0% (162/178) with GQ60 or more and 99.3% (150/151) with GQ80 or more. Thus, a higher GQ showed a higher reliability (Supplementary Fig. 4A). In contrast, increasing DP simply did not have much power to remove incorrect genotypes (Supplementary Fig. 4B). Furthermore, the majority of false positives for putative iUPD genotypes arose from an inaccuracy of genotyping in any one of the trio (81.1%, 90/111). Therefore, we required all genotypes in the trio with GQ40 or greater for identification of the true iUPD genotype. After filtering with a threshold GQ40, we identified 100 and 178 putative iUPD genotypes in the CEU 1463 and YRI Y117 trio, respectively (Fig. 2, Supplementary Table 8 and 9).

3.4. Validation of the putative iUPD genotypes by capillary sequencing and qPCR

We attempted to validate these candidates by capillary sequencing. Of these, only 1 putative iUPD genotype (Validation ID C1383 and Y3887, respectively), in the CEU 1463 and YRI Y117 trio was confirmed as a true iUPD genotype (Fig. 2, Supplementary 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
fidelity of genomic integrity, we investigated the traits of iUPD genotypes using NGS data during the pilot phase 2 of the 1000 Genomes Project. However, we could not find direct evidence of segmental iUPD after the accurate reconfirmation process including capillary sequencing and qPCR. Some parts of the reference sequence are inaccessible because of high-copy repeats or segmental duplications. This is a limitation of the current NGS technology producing short sequence reads. Indeed, 20% of the reference genome was inaccessible in the trio project (Altshuler et al., 2010). From our data, the accessible genome per CEU and YRI trio set were 2.24 Gb and 2.14 Gb, respectively (Table 1). Because the total length of the human reference genome, including the gap was composed of about 3.08 Gb, 27.2% and 30.5% of data in CEU and YRI trio, respectively, were not analyzed in this study. Furthermore, the use of only two trios might be too small a scale and low-level mosaicism is often difficult to detect accurately. However, the data presented here provides evidence that segmental UPD during normal development could not be a constitutive event in order to maintain genomic integrity. Constitutive UPD is very rare. Robinson et al. determined that UPD for an average chromosome occurs in 1/80,000 births (0.00125%) and UPD for any chromosome can be expected in roughly 1/3,500 births (0.02857%), based on the frequency of UPD15 (Robinson, 2000). Liehr suggested that the rate of UPD in human population might be even lower than 1 in 5,000 or less (Liehr, 2010). We studied 173 trios using genome-wide SNP array and WGS data using NGS, and identified one case with segmental iUPD. Segmental UPD for any chromosome can be expected in 1/173 births 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
and included only a few markers. These facts suggest that researchers may overlook
UPD in normal development and miss shorter segmental UPD, because UPD of many
chromosomal regions results in no obvious abnormalities (Kotzot and Utermann, 2005;
Robinson, 2000). In addition, lethal genotypes due to UPD during early embryonic
development would be undetectable. We suggest that trio genome analysis with
enhanced sequence accuracy could provide new findings for the risk of recessive
disorders, because one mutant allele from one parent can be transmitted to a child and
result in a homozygous state due to iUPD. To the best of our knowledge, this is the first
systematic study over whole chromosomal and segmental UPD in the human genome
without abnormal phenotype using familial trios.

5. Conclusions

The current study assessed the presence of whole chromosome and segmental UPD in
general populations using genome-wide SNP microarray and WGS data. We provided
evidence that segmental UPD in normal development is not a constitutive event in order
to maintain genomic integrity. Although we identified one obvious segmental paternal
iUPD in one HapMap sample, we could not find direct evidence of shorter segmental
iUPD. This suggested three possibilities, 1) human cells repress the usage of inter-allelic
homologous sequences as a template for HR, even at the early embryonic stage, 2)
shorter iUPD segments are unidentifiable because of absent informative markers within
the limited short segment, 3) UPD could be present in inaccessible genome regions
when using current NGS with short reads. Investigation of segmental UPD in general
populations will help to expand our general understanding of normal development in 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
genotype might result from LTA due to copy number neutral LOH, but not CNV. A red box indicates the cluster of UPI-M, MI-S and BPI genotypes. Arrowhead indicates ROH in the father. We did not consider these regions as UPD segments in this study. The data on chromosome 11 from CEU trio 1463 is shown as a representative example. M, mother; F, father; C, child.

Supplementary Fig. 4. GQ and DP in correct and incorrect genotypes confirmed by capillary sequencing in CEU 100 putative iUPD genotypes (300 genotypes). (A) The number of correct and incorrect genotypes falling within each genotype quality (GQ) score threshold are shown on the bar. (B) The number of correct and incorrect genotypes falling within each read depth (DP) score threshold are shown on the bar. GT, genotype.

Supplementary Fig. 5. Results of putative iUPD validated by capillary sequencing and qPCR. (A) Result of capillary sequencing and genotypes registered in Personal Genome Variants on UCSC in NA12878, NA12891 and NA12892 at candidate locus C1383 (NA19892 is not registered in Personal Genome Variants on UCSC). The genotype of putative iUPD C1383 site in NA12878 on UCSC was incorrect. Electropherograms of DNA sequences in the CEU trio show a paternal iUPD genotype in NA12878 (daughter). (B) Result of capillary sequencing and genotypes registered in Personal Genome Variants on UCSC in NA19240, NA19239 and NA19238 at candidate locus Y3887 (NA19239 and NA19238 are not registered in Personal Genome Variants on
25

UCSC. Electropherograms of DNA sequences in candidate locus Y3887 from the YRI
trio show the maternal iUPD genotype in NA19240 (daughter). (C) The daughter had a
microdeletion. qPCR was performed with primers C1383 and Y3887 for putative iUPD,
and with primers ZNF80 and GPR15, respectively, as standards. C1471 and Y3350
primers demonstrated the known deletion control (Kidd et al., 2008; Mills et al., 2011).
Normalized mean values for triplicates are shown for each interest target versus ZNF80
(blue) and GPR15 (red), respectively.

Supplementary Table 1. Samples from the UPD study. FID, family ID; IID, individual ID.

Supplementary Table 2. The 27 possible combinations of genotypes in a trio.

Supplementary Table 3. Primers for capillary sequencing and qPCR. This table lists
the primer sets used for this analysis. CHR, chromosome; POS, position; REF, reference base.

Supplementary Table 4. Identified CNVs in each trio by SNP6.0 array analysis.

Supplementary Table 5. Collating and filtering of putative iUPD genotypes and
problematic regions for mapping and assembly of regions or previous reported CNVs in
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
Supplementary Table 10. RefSeq Genes in identified segmental iUPD region
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.


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


Wang, K., et al., 2010. ANNOVAR: functional annotation of genetic variants from


Web references

1000 Genomes Project pilot 2 data.


Best Practice Variant Detection with the GATK v2.


Geneimprint.


http://www.fish.uniklinikum-jena.de/UPD.html.


The National Center for Biotechnology Information (NCBI).


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


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.,...
Table Legends

Table 1
Summary of putative segmental UPD segments in 173 trios detected by SNP6.0 array data analysis. Chr, chromosome; CNV, copy number variant; ND, not detectable; LTA, loss of transmitted allele; iUPD, uniparental isodisomy; LOH, loss of heterozygosity.

Table 2
Summary of alignment and base calling in two trios. AC+X: Autosomal chromosome (1-22) and X chromosome (exclude gap) = total length 2,706,959,439 bases (about 2.71 Gb).
Fig. 1

A

Allele ratio

NA19917 (M) 0.0
NA19916 (F) 0.0
NA19918 (C) 0.0

B

Copy number state

NA19917 (M) 2.0
NA19916 (F) 2.0
NA19918 (C) 2.0

C

SNP trio

chr17

Segmental paternal IUPD in 17p13.3-13.1
Fig. 2

<table>
<thead>
<tr>
<th>Step</th>
<th>CEU 1463</th>
<th>YRI Y117</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>2.24 Gb</td>
<td>2.14 Gb</td>
</tr>
<tr>
<td>Step 2</td>
<td>1094</td>
<td>1474</td>
</tr>
<tr>
<td>Step 3</td>
<td>1061</td>
<td>1447</td>
</tr>
<tr>
<td>Step 4</td>
<td>967</td>
<td>1394</td>
</tr>
<tr>
<td>Step 5</td>
<td>502</td>
<td>965</td>
</tr>
<tr>
<td>Step 6</td>
<td>100</td>
<td>178</td>
</tr>
<tr>
<td>Step 7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Step 8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Result</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr</td>
<td>Start position</td>
<td>End position</td>
</tr>
<tr>
<td>-----</td>
<td>----------------</td>
<td>--------------</td>
</tr>
<tr>
<td>11</td>
<td>51,078,178</td>
<td>51,359,581</td>
</tr>
<tr>
<td>17</td>
<td>6,689</td>
<td>8,202,065</td>
</tr>
<tr>
<td>7</td>
<td>119,133,278</td>
<td>119,393,868</td>
</tr>
<tr>
<td>6</td>
<td>137,300,451</td>
<td>143,369,018</td>
</tr>
<tr>
<td>5</td>
<td>107,513,060</td>
<td>107,716,753</td>
</tr>
<tr>
<td>8</td>
<td>14,677,944</td>
<td>15,701,490</td>
</tr>
<tr>
<td>7</td>
<td>88,496,196</td>
<td>88,887,792</td>
</tr>
<tr>
<td>2</td>
<td>85,751,279</td>
<td>88,861,509</td>
</tr>
<tr>
<td>12</td>
<td>129,502</td>
<td>131,942,726</td>
</tr>
<tr>
<td>11</td>
<td>81,131,219</td>
<td>81,387,538</td>
</tr>
<tr>
<td>22</td>
<td>20,825,481</td>
<td>21,201,922</td>
</tr>
<tr>
<td>15</td>
<td>21,205,648</td>
<td>45,731,444</td>
</tr>
<tr>
<td>18</td>
<td>65,224,346</td>
<td>76,085,336</td>
</tr>
<tr>
<td>22</td>
<td>20,927,130</td>
<td>21,243,931</td>
</tr>
<tr>
<td>1</td>
<td>206,304,300</td>
<td>246,785,226</td>
</tr>
<tr>
<td>17</td>
<td>69,586,313</td>
<td>70,111,013</td>
</tr>
<tr>
<td>X</td>
<td>140,182,100</td>
<td>140,575,068</td>
</tr>
<tr>
<td>22</td>
<td>20,718,086</td>
<td>21,107,920</td>
</tr>
<tr>
<td>1</td>
<td>144,979,429</td>
<td>145,700,719</td>
</tr>
<tr>
<td>1</td>
<td>236,789,304</td>
<td>246,590,204</td>
</tr>
<tr>
<td>11</td>
<td>114,231,222</td>
<td>134,235,117</td>
</tr>
<tr>
<td>13</td>
<td>82,217,462</td>
<td>83,042,185</td>
</tr>
<tr>
<td>6</td>
<td>140,718,454</td>
<td>141,182,824</td>
</tr>
<tr>
<td>X</td>
<td>4,726,561</td>
<td>145,198,977</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td>Sample</td>
<td>CEU 1463</td>
</tr>
<tr>
<td>----------</td>
<td>--------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>NA12878</td>
<td>NA12891</td>
</tr>
<tr>
<td>Relation</td>
<td>Daughter</td>
<td>Father</td>
</tr>
<tr>
<td>Total bases (Gb)</td>
<td>102.24 Gb</td>
<td>100.89 Gb</td>
</tr>
<tr>
<td>Mapped bases (Gb)</td>
<td>99.63 Gb</td>
<td>97.25 Gb</td>
</tr>
<tr>
<td>Total reads</td>
<td>2,507,012,490</td>
<td>2,264,396,064</td>
</tr>
<tr>
<td>Mapped reads</td>
<td>2,443,207,477</td>
<td>2,189,660,230</td>
</tr>
<tr>
<td>Mean mapped depth</td>
<td>31.9</td>
<td>30.3</td>
</tr>
<tr>
<td>Accessible genome (Gb)</td>
<td>2.32 Gb</td>
<td>2.33 Gb</td>
</tr>
<tr>
<td>Accessible genome (% of AC+X)</td>
<td>85.61 (%)</td>
<td>85.98 (%)</td>
</tr>
<tr>
<td>Accessible genome with trio (Gb)</td>
<td>2.24 Gb</td>
<td></td>
</tr>
<tr>
<td>Accessible genome with trio (Gb) (% of AC+X)</td>
<td>82.66 (%)</td>
<td></td>
</tr>
<tr>
<td>SNPs (N)</td>
<td>2,854,439</td>
<td>2,846,437</td>
</tr>
<tr>
<td>SNPs in dbSNP132 (N)</td>
<td>2,838,282</td>
<td>2,831,464</td>
</tr>
<tr>
<td>SNPs in dbSNP132 (%)</td>
<td>99.43 (%)</td>
<td>99.47 (%)</td>
</tr>
</tbody>
</table>