ARTICLE
Journal of Cellular Biochemistry 113:3520–3527 (2012)

Journal of Cellular Biochemistry

Chromosome Dynamic Changes in Two Cultured Chinese Human Embryonic Stem Cell Lines: Single Nucleotide Polymorphism, Copy Number Variation and Loss of Heterozygosity

Xue-Mei Chen,[1,2] Quan-Cheng Kan,[3] Fang Wang,[1] Hui-Juan Kong,[1] Yong-Yong Zhang,[1] Wen-Zhu Yu,[1] and Ying-Pu Sun[1*]

[1] *Reproductive Medical Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China*
[2] *Department of Human Anatomy, College of Basic Medical Sciences, Zhengzhou University, Zhengzhou 450001, China*
[3] *The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China*

ABSTRACT

The quality and safety of human embryonic stem cells (hESCs) in clinical application depend on gene stability. Two Chinese hESC lines, Zh1 and Zh21, were incubated over a long period. We observed and compared the gene stability in the passage numbers 20, 17 for Zh1 cell line and passage numbers 27, 60, 68 for Zh21 cell line. Single nucleotide polymorphisis analysis indicated that hESCs in early passages had relative gene stability; and with the increase in passage number, gene instability became strong. We also found that there were copy number variations (CNVs) in both Zh21 and Zh1. We analyzed the CNVs of Chinese Han Beijing man (CHB; normal Chinese people) and found that the all CNV forms were the loss in Zh21, Zh1, and CHB. We also analyzed and compared the related pathways of the mutant genes. We propose three steps to ensure hESC safety. Firstly, besides the conventional methods such as pluripotent genes, chromosome G-banding and teratoma, high-resolution DNA chip analysis should also be adopted; secondly, chromosomal properties are monitored every 10 passages in less than passage 50 and every 5 passages in more than passage 50; thirdly, the related pathways of mutant genes should be observed because only the mutant genes with variations of their related pathways may affected cell functions. J. Cell. Biochem. 113: 3520–3527, 2012.

KEY WORDS: HUMAN EMBRYONIC STEM CELLS; COPY NUMBER VARIATION; SINGLE NUCLEOTIDE POLYMORPHISM; LOSS OF HETEROZYGOSITY; KARYOTYPE

Human embryonic stem cells (hESCs) are a cell source of biotherapy because they have unique properties including self-differentiation, self-renewal, and multi-directional differentiation. Gene stability of hESCs is very important for both the clinical application and scientific research. It is necessary to perform strict analysis of genome stability before large-scale clinical applications of hESCs [Laurent et al., 2011]. Traditional methods for analysis of hESC gene stability refer to cytogenetic analysis techniques such as G-banding karyotype analysis and fluorescence in situ hybridization (FISH) [Mitalipova et al., 2005]. Karyotypic variations during hESC cultivation have been found by G-banding karyotype analysis [Draper et al., 2004]. Karyotypic abnormalities have been reported and the gains of chromosomes 12, 17, and X were the most common [Baker et al., 2007]. The related genes in these regions include antiapoptotic gene *BIRC5*, pluripotent gene *NANOG*, cell-cycle genes *DPPA3* and *GDF3*, proto-oncogenes *KRAS*, *TOP3A*, *COPS3*, *MAPK7* and *SOX5*, X inactive gene *XIST*, and cell signaling molecules *ELK* and *ARAF* [Roelofs et al., 2000; Azuhata et al., 2001;

Additional supporting information may be found in the online version of this article.

No competing financial interests exist.
Grant sponsor: National Natural Science Foundation of China; Grant number: 30771105.

*Correspondence to: Dr. Ying-Pu Sun, Number one, Constructive East Road, Zhengzhou 450052, China.
E-mail: syp2008@vip.sina.com

Manuscript Received: 1 February 2012; Manuscript Accepted: 6 June 2012
Accepted manuscript online in Wiley Online Library (wileyonlinelibrary.com): 18 June 2012
DOI 10.1002/jcb.24229 • © 2012 Wiley Periodicals, Inc.

Sperger et al., 2003; Yang et al., 2003; Zafarana et al., 2003; van Dartel and Hulsebos, 2004; Imreh et al., 2006].

It is necessary to evaluate karyotypic abnormalities during hESC passage [Moralli et al., 2011]. However, the resolution of conventional karyotype analysis (3–20 Mb) and metaphase chromosome FISH (4–10 Mb) was low [Trask et al., 1991; Närvä et al., 2010; Slovak et al., 2010]. At present, high-resolution single nucleotide polymorphisms (SNPs) genotyping is used to detect micro-abnormalities such as copy number variation (CNV) and loss of heterozygosity (LOH) in hESCs and induced pluripotent stem cells [Mayshar et al., 2010; Närvä et al., 2010; Laurent et al., 2011]. Compared with other methods, it can detect the genic regions that are more susceptible to mutation in hESCs, such as repeated gene duplication on chromosome 12. Therefore, Stephenson et al. [2010] proposes that SNP, CNV, and LOH analysis should be served as mandatory diagnostic tests for each hESC line before it enters the national stem cell bank.

Our study aimed to evaluate the dynamic changes in chromosome of the two cultured Chinese hESCs (Zh1 and Zh21). Based on the paper [Närvä et al., 2010], we chose passage number 20, 27 for Zh1 cell line and passage number 27 for Zh21 cell line as low passage numbers, and passage number 60, 68 for Zh21 cell line as high passage numbers. We performed high-resolution SNP genotyping (1140,419 SNPs) in the two hESC lines. The array is suitable for detecting karyotype, CNV, LOH, and SNP. To observe the changes in hESC passage, the results were compared between different passage numbers in the same line, and CNVs were also compared between the two hESCs and Chinese Han Beijing man (CHB; normal Chinese people chosen by HapMap project).

MATERIALS AND METHODS

All study methods were approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University. All the subjects enrolled into the study gave written formal consent to participate.

CELL CULTURE AND hESC KARYOTYPE ANALYSIS

The two hESC lines Zh1 and Zh21 (ZZU-hESC-1 and ZZU-hESC-21, ZZU: Zhengzhou University) were investigated in this study. hESC lines were established by isolating the inner cell mass of in vitro fertilized 5AA and 3AA blastocysts provided by the Embryonic Stem Cell Laboratory of the Reproductive Medical Center of the First Affiliated Hospital of Zhengzhou University. The isolated hESCs were incubated in an inactivated mixed feeder layer. Passages 1–6 of primary mouse embryonic fibroblasts (MEF) and passages 10–20 of primary human foreskin fibroblasts (HFF) were inactivated with mitomycin C, and then mixed according to about 1:1 of MEF: HFF to prepare the mixed feeder layer. One passage was performed with the mechanical method every 4–5 days. hESC culture media mainly consisted of 80% KO-DMEM, 20% knockout serum replacement (KOSR), 1% nonessential amino acids (NEAA), 2 mM L-glutamine, 0.1 mM β-mercaptoethanol, and 8 ng/ml of basic fibroblast growth factor (all from GIBCO, USA) [Li et al., 2011]. After passage, hESCs were analyzed with the techniques of conventional karyotype and FISH [Caisander et al., 2006; Li et al., 2011]. For FISH, monoclonals of Zh1-P20, Zh1-P27, Zh21-P27, Zh21-P60, and Zh21-P68 were respectively collected with mechanical method, and then degested into single cell with 0.25% of pancreatin. The single cell was fixed on the glass slide followed by the procedures of X/Y probe kit (CEP X (DXZ1)/Y (DYZ3) Probe, Vysis, USA) containing probes CEPY (orange) and CEPX (green).

DNA EXTRACTION AND PURIFICATION

Whole-genome DNA was extracted and purified with the DNeasy Blood & Tissue Kit according to the manufacturer's protocol (QIAGEN, Cat. no.69504, German). The concentration and quality of the samples were determined by a spectrophotometer (Nanodrop, Thermo Scientific, USA) and 1% agarose gel electrophoresis with Reference DNA as control. Two hundred nanograms of DNA was used for amplification. High-resolution SNP genotyping (the Illumina Human omni1-quad beadchip, 1,140,419 SNPs, USA) was performed in five samples including Zh1-P20, Zh1-P27, Zh21-P27, Zh21-P60, Zh21-P68 followed by scanning on a BeadArray Reader (Illumina, Inc., USA). There was gene SNP probe, tag SNPs, CNV and other main genome regions in this chip. Reference DNA used for the SNP analysis is from Genotyping of more than 100 healthy individuals including Caucasion (CEU), Asian (CHB + JPT) and Yoruban (YRI; from the first paragraph in page 3 of supplementary file). The gene call threshold was set at 0.15, and the call rates were between 0.9882988 and 0.9961972.

SNP IDENTIFICATION IN SAMPLES AND THE STATISTICS OF BASIC DATA

The genotype of each sample, logR ratio and B Allele frequency of each locus were calculated with Genome Studio software (Illumina, Inc.). The numbers of homo (BB) and heter (AB) of every sample were analyzed referring to dbSNP database (from UCSC hg18 version). The SNP categories for coding regions were annotated and different mutation types were analyzed. There were four mutation types including synonymous, nonsynonymous, tolerated and damaging

TABLE I. SNP Mutation Forms in the Coding Region of Different Passages of Zh21 and Zh1

		Zh21 %			Zh1 %	
Cases	Subtype	P27	P60	P68	P20	P27
Homo	Tolerated	63	58	58	58	58
	Damaging	37	42	42	42	42
	Nonsynonymous	26	35	35	35	35
	Synonymous	25	16	16	16	16
Heter	Tolerated	70	71	71	74	74
	Damaging	30	29	29	26	26
	Nonsynonymous	28	27	27	27	27
	Synonymous	24	24	24	25	25

Citation: For example, homo "#Tolerated+#Damaging" is standard for coding variants predicted. Tolerated% = #Tolerated/(#Tolerated+# Damaging) × 100%; Damaging% = # Damaging/(#Tolerated+# Damaging) × 100%; Nonsynonymous% = (#Nonsynonymous/#coding variants) × 100%; Synonymous% = (#synonymous/#coding variants) × 100%.

variations. Forecast of damaging mutations was calculated based on SIFT algorithm [Kumar et al., 2009].

CNV ANALYSIS WITH HIDDEN MARKOV MODEL (HMM)

According to SNP chip data, CNV regions were identified with HMM algorithm. This algorithm can more accurately capture CNV in specific population, because HMM can efficiently reflect LRR and BAF intensity from SNP (LRR: Log R ratio and BAF: B allelefrequency), and the information of SNP allele frequency distribution. Firstly, the presumed original CNV regions underwent

TABLE II. CHB-CNV and CNV Changes in Genome During hESC Passages (%)

Samples	Loss	Heter	Gain
Zh1-P27/20	94.95	93.56	5.05
Zh21-P60/27	92.00	97.83	8.00
Zh21-P68/60	95.33	98.77	4.67
HapMap-CHB	91.15	69.52	8.85
Zh1-P20	96.77	91.67	3.23
Zh21-P27	99.38	95.49	0.62

Heter % = heter deletion/(heter deletion + homo deletion) × 100%; loss % = loss/(loss + gain) × 100%, gain % = gain/(loss + gain) × 100%.

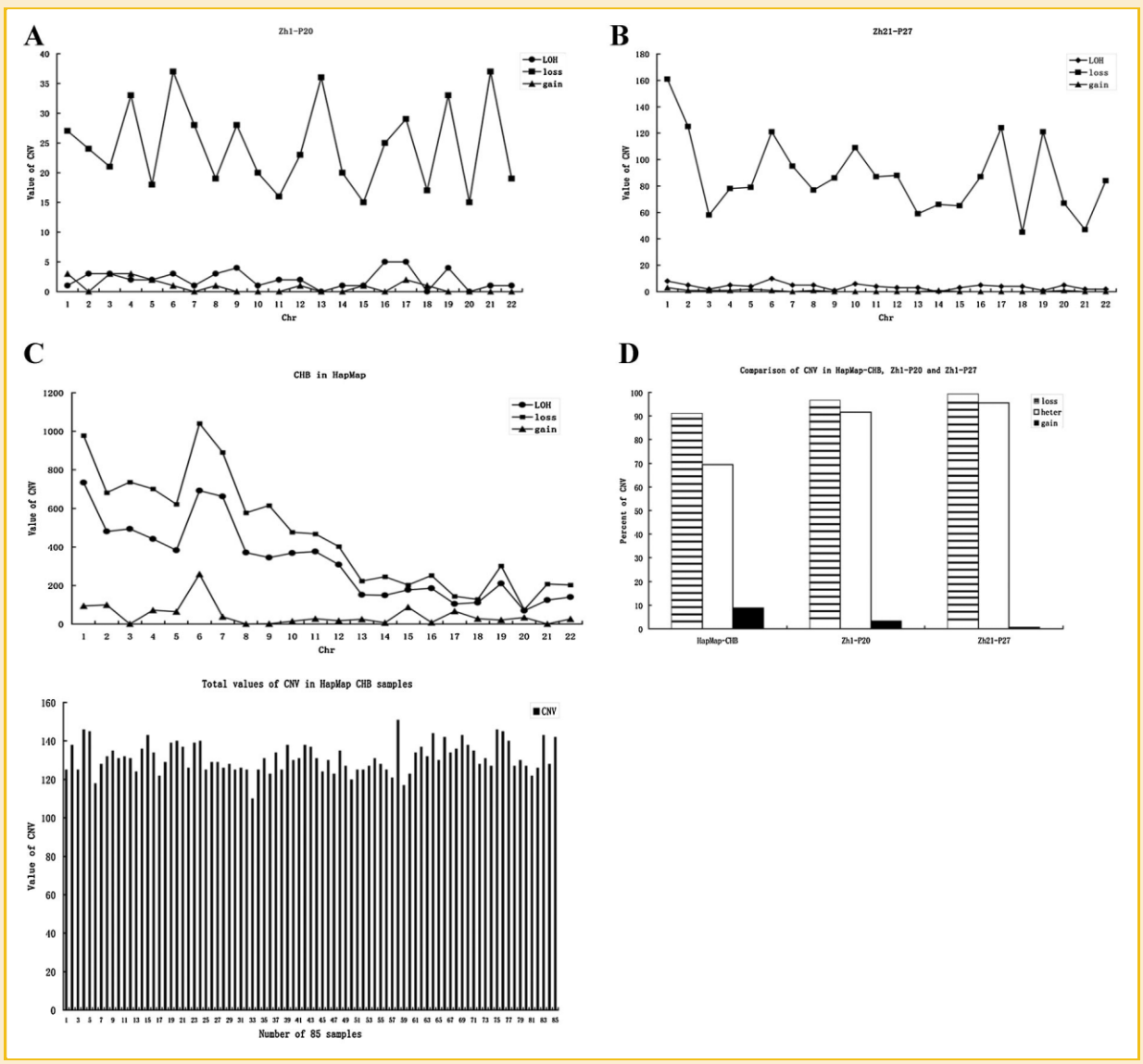

Fig. 1. Majority of CNV is loss in human embryonic stem cells (hESCs). A: CNV in Zh1-P20: Loss accounts for 96.77% in Zh1-P20. B: CNV in Zh21-P27: Loss accounts for 99.38% in Zh21-P27. C: CNV in CHB HapMap population: Loss accounts for 91.15% in CHB HapMap population. D: Comparison of loss between Zh1-P20, Zh21-P27 and CHB: X axis indicates CHB, Zh1-P20 and Zh21-P27. From CHB-91.15%, Zh1-P20-96.77%, and Zh21-P27-99.38%, we can see that Chinese hESC lines (Zh1 and Zh21) have a parallel CNV (both loss and gain) distribution to CHB. E: The total values of CNV in HapMap CHB samples: There are average 131 CNVs in each sample of the 85 Chinese Han Beijing man.

TABLE III. Gene Pathway of CNV Changes Detected in hESCs Samples During hESC Passages

Chromosome	SNP no.	Length	Gene	Pathway
Zh21-P60/27				
4p13-p12	143	896,128	GABRG1	ko04080 Neuroactive ligand-receptor interaction
Zh21-P68/60				
1p31.1	337	1,725,882	NEGR1	ko04514 Cell adhesion molecules (CAMs)
11p11.2-p11.12	175	1,454,522	FOLZH1	ko04977 Vitamin digestion and absorption
Zh1-P27/20				
13q14.2	45	137,039	LPAR6,RB1	LPAR6: map04080 Neuroactive ligand-receptor interaction RB1:K06618 Retinoblastoma-associated protein

three-round screening (including immune regions, centromeric regions and SNP limit, for example, SNP20_Len 100 kbp) to obtain CNV regions, and then CNV-related genes were annotated with UCSC Genome Browser database [Wang et al., 2007, 2008; Diskin et al., 2008; Staaf et al., 2008].

CNV ANALYSIS WITH KARYOSTUDIO SOFTWARE

KaryoStudio software is suitable to the analysis of the data generated by Illumina Infinium HD Bead Chips. The software can display gene information and found regions including the regions associated with phenotypes.

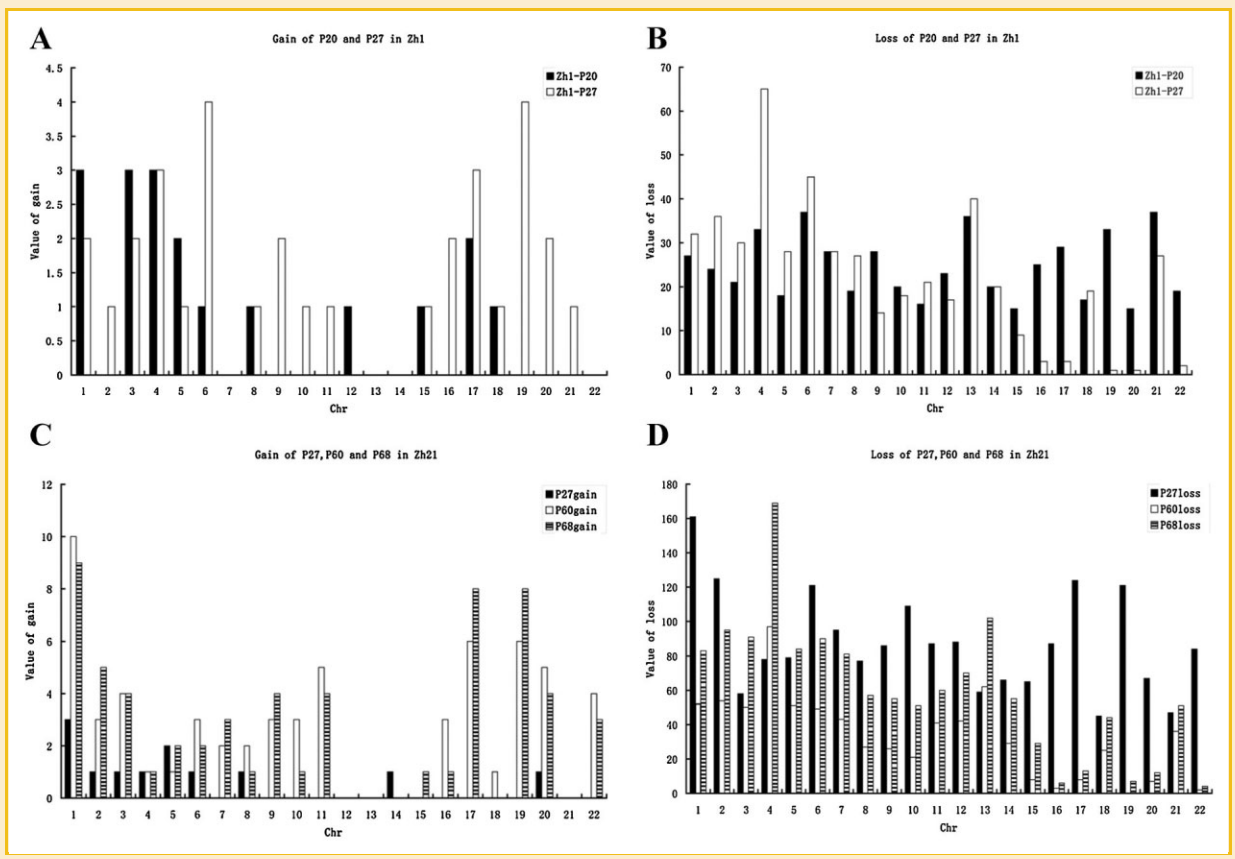

Fig. 2. Gains and losses in Zh1 and Zh21. *X*-axis displays chromosomes and *Y*-axis displays the number of CNV (gain or loss). A, B, C, and D all show the dynamic changes in CNV during hESC passages. A: Gains of P20 and P27 in Zh1: There are no CNVs in chromosomes 7, 13, 14, and 22 of Zh1-P20 and P27, and no CNVs in chromosomes 19 of Zh1-P20, but 4 gains occur in chromosomes 19 of Zh1-P27. B: Losses of P20 and P27 in Zh1: losses occur in Zh1-P20 and Zh1-P27. Losses are increased from 33 to 65 in chromosomes 4 of Zh1-P20 to Zh1-P27. Losses are decreased from 33 to 1 in chromosomes 19 of Zh1-P20 to Zh1-P27. C: Gains of P27, P60, and P68 in Zh21: There are no gains in chromosomes 12, 13, and 21 of Zh21-P27/60/68, and the most gain locuses (10) are in chromosome 1 of Zh 21-P60. The total gains are increased from 12 to 62 of Zh21-P27 to Zh21-P60, and the total gains (61) of Zh21-P68 are similar to that of Zh21-P60. D: Losses of P27, P60 and P68 in Zh21: There are no losses in chromosome 19 of Zh21-P60, and the most loss locuses (161) are in chromosome 1 of Zh21-P27. The total losses are decreased from 1929 to 733 of Zh21-P27 to Zh21-P60, but increase from 733 to 1309 of Zh21-P60 to Zh21-P68.

LOH ANALYSIS

Genomic LOH distribution was also calculated with HMM algorithm.

STATISTICAL ANALYSIS

All data were put in EpiData3.1 software, then were analyzed with SPSS17.0 software.

RESULTS

SAMPLES

The hESC lines, Zh1 and Zh21, displayed the typical markers and properties of undifferentiated hESCs [Li et al., 2011], (Supplementary Table 1). To investigate genomic changes in hESC passage, undifferentiated hESCs with different passage numbers were harvested. All hESC lines (Zh1-P20, Zh1-P27, Zh21-P27, Zh21-P60, Zh21-P68) have normal karyotype indicated by G-bind (46, XY) and FISH (Diploid, XY; Supplementary Figs. 1–5).

To observe the complex changes in genes during the long-term cultivation, the three genetic variation parameters including SNP, CNV, and LOH were observed and analyzed in the same celllinewith different passage numbers (Zh1-20/27, Zh21-27/60/68).

SNP ANALYSIS

The variations of Genomic SNP include heter genotype (AB) and homo genotype (BB). The four kinds of SNP variations are tolerated, damaging, nonsynonymous, and synonymous. The homo damaging and nonsynonymous variations were respectively increased by 5% and 9% from P27 to P60 of Zh21 in the coding region. The variation in the coding region was unchanged from P60 to P68 of Zh21 (Table I). There were no variations from P20 to P27 of Zh1 (lower passage numbers) in the coding region. These results suggest that there is gene stability in the hESCs with early passage numbers. With the increase in passage numbers, the genomic instability is increased, especially the genomic variations in the coding region may affect protein expression.

CNV CHANGES IN hESC PASSAGE

The five samples (Zh1-P20, Zh1-P27, Zh21-P27, Zh21-P60, Zh21-P68) have normal karyotype. However, the CNV changes were discovered in P20/27 of Zh1, and P27/60 and P60/68 of Zh21. To compare the CNVs between the two hESCs and normal Chinese human genome, we analyzed the CNVs of Chinese Han Beijing man (CHB; normal Chinese people chosen by HapMap project) and found that there were average 131 CNVs in 85 HapMap samples from CHB. The loss proportion of CNV in the whole genome was 91.15% in CHB, 96.77% in Zh1-P20 and 99.38% in Zh21-P27, respectively. We could see from this result that Chinese hESC lines, Zh1 and Zh21, have a parallel CNV (loss and gain) distribution to the HapMap CHB samples (Table II, Fig. 1).

During passage, the loss ratios of Zh1-P20/27 (94.95%) and Zh21-P60/27 (92.00%) were closed to that of HapMap samples (Table II). According to the more credible standard SNP20_length 100 kb, we identified 10 CNVs with more than 0.5 Mb in Zh1-P27/20 (1p21.1, 1p21.3, 4q34.2, 6q12, 7q21.3, 7q21.2, 8p22, 12q12–q13.11, 13q14.2, and 21q21.1), 5 CNVs with more than 0.5 kb in Zh21-P60/27 (2p12, 3p12.1, 4p13–p12, 6q12, and 11p14.3), and 9 CNVs with more than 1 Mb in Zh21-P68/60 (1p31.1, 2q32.1, 3p12.1, 4q28.3, 5q21.2-q21.3, 6q12, 7q31.31, 11p11.2-p11.12, 12q21.31; Supplementary Table 2). In these CNV-related genes, we found the pathways of five genes including *LPAR6 (Neuroactive ligand-receptor interaction)*, *RBI (Retinoblastoma-associated protein)*, *GABRG1 (Neuro active ligand-receptor interaction)*, *NEGR1 (Cell adhesion molecules, CAMs)* and *FOLZH1 (Vitamin digestion and absorption*; Table III).

In early passage of Zh1-P27/20, besides the 11 CNVs with morethan 100 kb, there were no large CNV loss (>200 kb; Supplementary Table 2). It was very interesting that there was nogain in chromosomes 12 and 13 of Zh21-P27/60/68 and Zh1-P20,but loss occurred in all chromosome of the two cell lines (Zh1-P20/27 and Zh21-P27/60/68) except chromosome 19 of Zh21-P60. The most loss locuses (161) were in chromosome 1 of Zh21-P27, and the most gain locuses (10) were in chromosome 1 of Zh21-P60 (Fig. 2).

These results indicate that the fragment length of CNVs is smaller in early passage; while in late passage, the fragment length of CNVs is larger. This may be another cause of increased gene instability in hESCs with late passage numbers.

Moreover, we analyzed P27/60 and P60/68 of Zh21 with KaryoStudio software and found the CNV-related genes, *Azospermia factor a (AZFa)* and *Azospermia factor b (AZFb)* in P60; and *AZFa*, *AZFb* and *Azospermia factor c (AZFc)* in P68. *Azospermia factor (AZF)* includes *AZFa*, *AZFb*, and *AZFc*. Microdeletions in these factors will lead to azoospermia and oligospermia

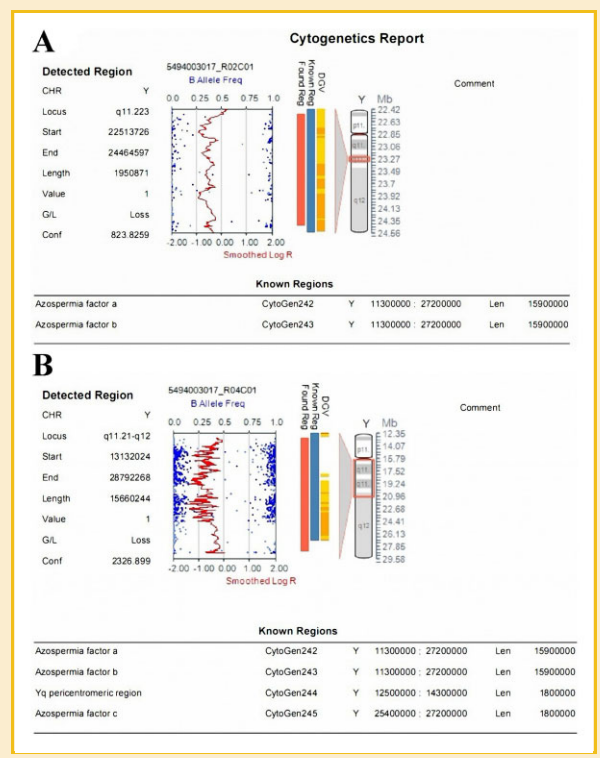

Fig. 3. Losses in the chromosome Y of Zh21 cell line. A: P60: The losses at Yq11.223 with related genes, *AZFa* and *AZFb*. B: P68: The loss at Yq11.21-q12 with related genes, *AZFa*, *AZFb*, and *AZFc*.

with male infertility (Fig. 3). The results suggest that Zh21 can not be used in induction and differentiation of germ cells-spermatozoa.

LOH ANALYSIS

Other chromosomes had LOH except chromosomes 13, 14, 21 and 22 of Zh21, and chromosome 22 of Zh1. There were 25 LOHs (>1 Mb) in P27/60 of Zh21, and 13 LOHs in P60/68 of Zh21 (Fig. 4). In Zh1-P20/27, there was only one large LOH at 3q26.1.

GENES AFFECTED BY LOH

In Zh21, the LOH-related differential genes including *DAB2*, *RFX5*, *PSMB4*, and *SMPD3* were identified in chromosomes 1, 5, 9, 11, 12, 16, 17, 18 (Table IV, Supplementary Table 3). These genes belonged to these pathways (Table V). In Zh1-P20/27, the LOH-related gene was *SLITRK3*.

DISCUSSION

The application of hESCs such as substitution therapy depends on whether hESCs can keep normal karyotype during cultivation and passage. In most studies, hESC lines seem to can maintain their pluripotency and normal karyotypes during long-term cultivation in vitro [Darnfors et al., 2005]. However, the abnormalities occurring in sensitive areas of genomes (e.g., the sub-telomeric regions, gains of chromosomes 17q and 12, loss of chromosomes 10, isodicentric X chromosome in HS237-P61) are likely to result in marked changes in functional and biological effects, leading to loss of hESC applications value [Draper et al., 2004; Inzunza et al., 2004; Närvä et al., 2010]. These findings suggest that each hESC possesses its unique genomic properties and alteration during cultivation. Stem cells are faced with different outcomes including self-renewal, differentiation and death; and genomic variation in late passage exhibits an adaptability to long-term cultivation [Enver et al., 2005]. The gains of chromosomes 12, 17, and 20 are conducive to self-renewal of hESCs [International Stem Cell Initiative et al., 2011].

Recently, some reports indicated that more genetic variations could be detected with high resolution DNA analysis methods such as SNP genotyping (locuses 1, 114, and 419) and Affimetrix SNP6.0 array [Inzunza et al., 2004; Darnfors et al., 2005; Maitra et al., 2005; Närvä et al., 2010; Laurent et al., 2011] (Supplementary Table 4). There were CNVs in chromosome 10, 15, 3, 5, 12, 17, and 20. It was reported that 72% of CNVs manifested amplifiable form in hESCs; while in the HapMap Caucasians samples, gains and losses were basically equal [Närvä et al., 2010].

Our data suggest that Zh1 and Zh21 possess their own genetic features during cultivation. SNP analysis indicates that the genome is stable in hESCs with early passage number; and with the increase in passage number, genomic instability is gradually increases. CNVs occurred in chromosomes 1, 2, 3, 4, 5, 6, 7, 11, and 12 of Zh21-P27/

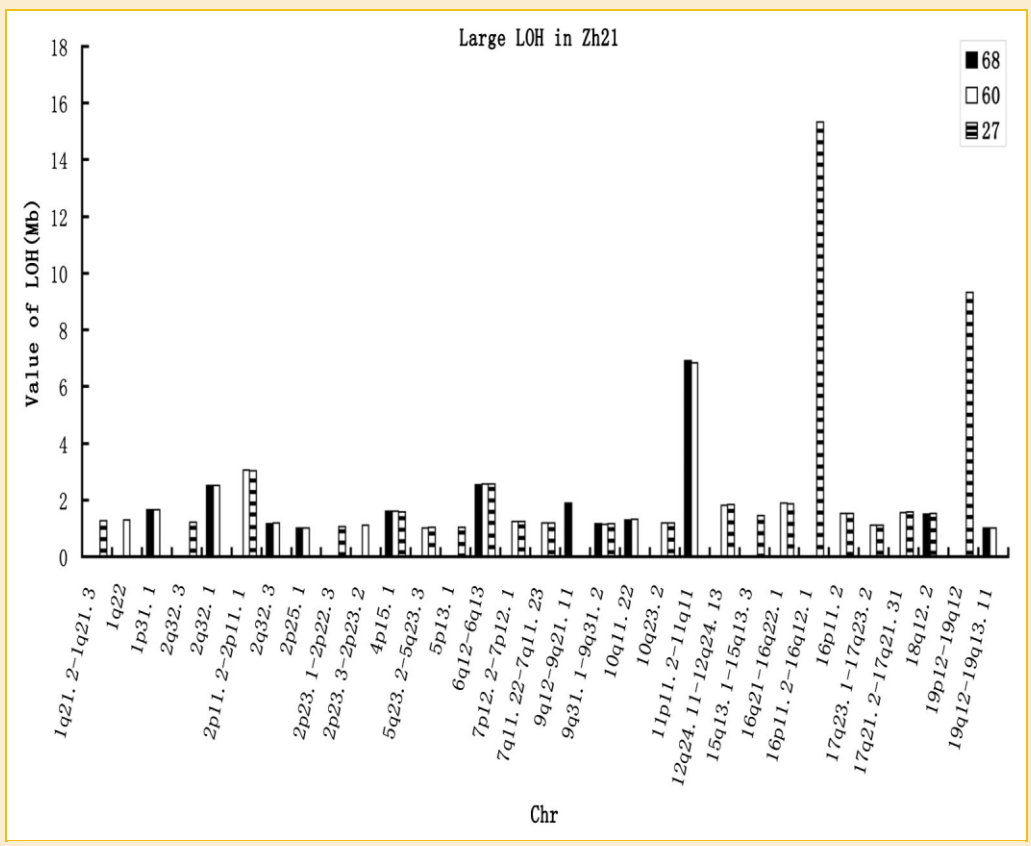

Fig. 4. Large loss of heterozygosity (LOH; >1 Mb) in Zh21. *X*-axis displays chromosomal localization of LOH and *Y*-axis displays the length (Mb) of LOH. There were 25 large LOH (>1 Mb) from P27 to P60 of Zh21, and 13 large LOH from P60 to P68 of Zh21.

TABLE IV. The Diversity Genes With Large LOH Changes (>1 Mb in size) in Different Samples Pairs (P27/P60, P60/P68) of Zh21

Chr	Passage	LOH sites	LOH length (>1 Mb)	Difference genes
1	27	1q21.2-1q21.3	1.278496	Loss: RFX5, PSMB4, SELENBP1, POGZ
1	60	1q21.2	1.103154	
5	27	5p13.1	1.026553	Loss: DAB2
9	68	9q12-9q21.11	1.883613	Loss: FOXD4L5, 4, 2, CBWD3,CBWD5, CBWD3, FOXD4L3, LOC572558, PGM5, C9orf71, PIP5K1, FAM122A
11	60	11p11.2-11q11	6.828672	
11	68	11p11.2-11q11	6.907441	Gain: OR4A15
12	27	12q24.11	1.206816	
12	60	12q24.11	1.349395	Gain: TCTN1, HVCN1
16	27	16q21-16q22.1	1.856016	
16	60	16q21-16q22.1	1.894063	Gain: SMPD3
17	27	17q21.2-17q21.31	1.568086	
17	60	17q21.2-17q21.31	1.543447	Loss: MIR2117
18	27	18q12.2	1.167521	
18	60	18q12.2	1.517969	Gain: MIR187,C18orf21, RPRD1A

TABLE V. Diversity Gene Pathway of LOH

Gene	Pathway	Class
DAB2	ko04144 endocytosis	Cellular processes; transport and catabolism; endocytosis
RFX5	ko04612 antigen processing and presentation	Organismal systems; immune system; antigen processing and presentation
	ko05152 tuberculosis	
	ko05340 primary immunodeficiency	Human diseases; infectious diseases; tuberculosis; human diseases; immune system diseases; primary immunodeficiency
PSMB4	ko03050 proteasome	Genetic information processing; folding, sorting, and degradation; proteasome
SMPD3	map00600 sphingolipid metabolism	Metabolism; lipid metabolism
	map01100 metabolic pathways	

The data in Table V are from KEGG database. Its web address is as follows: http://www.genome.jp/kegg/pathway.html

60/68. CNVs were 91.20% in HapMap-CHB, 92.00% in Zh21-P60/27 and 95.33% in Zh21-P68/60, and mostly were loss. These data suggest that the two Chinese hESC lines have a parallel CNV distribution to CHB in both loss and gain. However, Närvä et al. [2010] reported that 72% of genomic CNVs were gain in hESCs; while in the 90 Caucasian samples from HapMap database, gains and losses were basically equal. The differences between Narva' data and ours may be caused by racial difference. The International Stem Cell Initiative analyzed 125 hESC lines and 11 induced pluripotent stem cell lines from 38 laboratories worldwide, and their data reflected far greater ethnic diversity [International Stem Cell Initiative et al., 2011].

hESCs should have a normal genetic composition for clinical application. However, there is a spontaneous mutation rate of 10^{-9} in each nucleotide of hESCs [Maitra et al., 2005]. SNP rate is the highest and conservation is the lowest in coding site 3, while the lowest SNP rate occurs in coding site 2, reflecting codon degeneracy of amino acid encoding [Castle, 2011]. Our data showed that high-resolution analysis platform could detect kilobase–length aberrations, hESC genetic changes continuously increased during cultivation, the homozygous damaging and nonsynonymous changes of coding SNP (cSNP) were respectively increased by 5% and 9% in P60/27 of Zh21, and heterozygous damaging and nonsynonymous variations were all decreased by 1% in P60/27 of Zh21. These findings suggest that cSNP is likely to result in amino acid substitutions, which may be related to phenotype or diseases.

Obviously, shortening cultivation time is very important for clinical applications. However, it is very difficult to define ''normal'' because even a smallest genetic change may substantially affect functions such as cellular oncogenic potential and other functions. In *AZF* family, CNV-related genes, *AZFa* and *AZFb* occurred in P60 of Zh21; and *AZFa*, *AZFb*, and *AZFc* occurred in P68 of Zh21 during long-term cultivation. These results suggest that Zh21 can not be used in induction and differentiation of germ cells-spermatozoa.

Although our data and previous reports can not allow us to define a safe cut-off passage, we still propose three steps to ensure the safety of hESCs. Firstly, besides the conventional methods such as pluripotent genes, G-banding karyotype and teratoma, high-resolution DNA chip analysis should also be used to display the genetic features; secondly, chromosomal properties are monitored every 10 passages in early passage number and every 5 passages in late passage number; thirdly, the related pathways of mutant genes should be observed because only the mutant genes with variations of their related pathways may affected cell functions.

ACKNOWLEDGMENTS

We are grateful to everyone who has contributed to this work. We especially acknowledge professor Kai-Juan Wang, Zhao-shu Zeng, and Li Wang of Zhengzhou University for their valuable advice about data analysis, integration, and statistical analysis.

REFERENCES

Azuhata T, Scott D, Takamizawa S, Wen J, Davidoff A, Fukuzawa M, Sandler A. 2001. The inhibitor of apoptosis protein survivin is associated with high-risk behavior of neuroblastoma. J Pediatr Surg 36:1785–1791.

Baker DE, Harrison NJ, Maltby E, Smith K, Moore HD, Shaw PJ, Heath PR, Holden H, Andrews PW. 2007. Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. Nat Biotechnol 25:207–215.

Caisander G, Park H, Frej K, Lindqvist J, Bergh C, Lundin K, Hanson C. 2006. Chromosomal integrity maintained in five human embryonic stem cell lines after prolonged in vitro culture. Chromosome Res 14:131–137.

Castle JC. 2011. SNPs occur in regions with less genomic sequence conservation. PLoS ONE 6:e20660.

Darnfors C, Flodin A, Andersson K, Caisander G, Lindqvist J, Hyllner J, Wahlström J, Sartipy P. 2005. High-resolution analysis of the subtelomeric regions of human embryonic stem cells. Stem Cells 23:483–488.

Diskin SJ, Li M, Hou C, Yang S, Glessner J, Hakonarson H, Bucan M, Maris JM, Wang K. 2008. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res 36:e126.

Draper JS, Smith K, Gokhale P, Moore HD, Maltby E, Johnson J, Meisner L, Zwaka TP, Thomson JA, Andrews PW. 2004. Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. Nat Biotechnol 22: 53–54.

Enver T, Soneji S, Joshi C, Brown J, Iborra F, Orntoft T, Thykjaer T, Maltby E, Smith K, Abu Dawud R, Jones M, Matin M, Gokhale P, Draper J, Andrews PW. 2005. Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells. Hum Mol Genet 14:3129–3140.

Imreh MP, Gertow K, Cedervall J, Unger C, Holmberg K, Szöke K, Csöregh L, Fried G, Dilber S, Blennow E, Ahrlund-Richter L. 2006. In vitro culture conditions favoring selection of chromosomal abnormalities in human ES cells. J Cell Biochem 99:508–516.

International Stem Cell Initiative, Amps K, Andrews PW, Anyfantis G, Armstrong L, Avery S, Baharvand H, Baker J, Baker D, Munoz MB, Beil S, Benvenisty N, Ben-Yosef D, Biancotti JC, Bosman A, Brena RM, Brison D, Caisander G, Camarasa MV, Chen J, Chiao E, Choi YM, Choo AB, Collins D, Colman A, Crook JM, Daley GQ, Dalton A, De Sousa PA, Denning C, Downie J, Dvorak P, Montgomery KD, Feki A, Ford A, Fox V, Fraga AM, Frumkin T, Ge L, Gokhale PJ, Golan-Lev T, Gourabi H, Gropp M, Lu G, Hampl A, Harron K, Healy L, Herath W, Holm F, Hovatta O, Hyllner J, Inamdar MS, Irwanto AK, Ishii T, Jaconi M, Jin Y, Kimber S, Kiselev S, Knowles BB, Kopper O, Kukharenko V, Kuliev A, Lagarkova MA, Laird PW, Lako M, Laslett AL, Lavon N, Lee DR, Lee JE, Li C, Lim LS, Ludwig TE, Ma Y, Maltby E, Mateizel I, Mayshar Y, Mileikovsky M, Minger SL, Miyazaki T, Moon SY, Moore H, Mummery C, Nagy A, Nakatsuji N, Narwani K, Oh SK, Oh SK, Olson C, Otonkoski T, Pan F, Park IH, Pells S, Pera MF, Pereira LV, Qi O, Raj GS, Reubinoff B, Robins A, Robson P, Rossant J, Salekdeh GH, Schulz TC, Sermon K, Sheik Mohamed J, Shen H, Sherrer E, Sidhu K, Sivarajah S, Skottman H, Spits C, Stacey GN, Strehl R, Strelchenko N, Suemori H, Sun B, Suuronen R, Takahashi K, Tuuri T, Venu P, Verlinsky Y, Ward-van Oostwaard D, Weisenberger DJ, Wu Y, Yamanaka S, Young L, Zhou Q. 2011. Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. Nat Biotechnol 29:1132–1144.

Inzunza J, Sahlén S, Holmberg K, Strömberg AM, Teerijoki H, Blennow E, Hovatta O, Malmgren H. 2004. Comparative genomic hybridization and karyotyping of human embryonic stem cells reveals the occurrence of an isodicentric X chromosome after long-term cultivation. Mol Hum Reprod 10:461–466.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081.

Laurent LC, Ulitsky I, Slavin I, Tran H, Schork A, Morey R, Lynch C, Harness JV, Lee S, Barrero MJ, Ku S, Martynova M, Semechkin R, Galat V, Gottesfeld J, Izpisua Belmonte JC Murry C, Keirstead HS, Park HS, Schmidt U, Laslett AL, Muller FJ, Nievergelt CM, Shamir R, Loring JF. 2011. Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. Cell Stem Cell 8:106–118.

Li PF, Wang F, Kong HJ, Zhao F, Bai AH, Chen XM, Sun YP. 2012. Establishment of polycystic ovary syndrome-derived human embryonic stem cell lines. Gynecol Endocrinol 28:25–28.

Maitra A, Arking DE, Shivapurkar N, Ikeda M, Stastny V, Kassauei K, Sui G, Cutler DJ, Liu Y, Brimble SN, Noaksson K, Hyllner J, Schulz TC, Zeng X, Freed WJ, Crook J, Abraham S, Colman A, Sartipy P, Matsui S, Carpenter M, Gazdar AF, Rao M, Chakravarti A. 2005. Genomic alterations in cultured human embryonic stem cells. Nat Genet 37:1099–1103.

Mayshar Y, Ben-David U, Lavon N, Biancotti JC, Yakir B, Clark AT, Plath K, Lowry WE, Benvenisty N. 2010. Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell 7:521–531.

Mitalipova MM, Rao RR, Hoyer DM, Johnson JA, Meisner LF, Jones KL, Dalton S, Stice SL. 2005. Preserving the genetic integrity of human embryonic stem cells. Nat Biotechnol 23:19–20.

Moralli D, Yusuf M, Mandegar MA, Khoja S, Monaco ZL, Volpi EV. 2011. An improved technique for chromosomal analysis of human ES and iPS cells. Stem Cell Rev 7:471–477.

Närvä E, Autio R, Rahkonen N, Kong L, Harrison N, Kitsberg D, Borghese L, Itskovitz-Eldor J, Rasool O, Dvorak P, Hovatta O, Otonkoski T, Tuuri T, Cui W, Brüstle O, Baker D, Maltby E, Moore HD, Benvenisty N, Andrews PW, Yli-Harja O, Lahesmaa R. 2010. High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity. Nat Biotechnol 28:371–377.

Roelofs H, Mostert MC, Pompe K, Zafarana G, van Oorschot M, van Gurp RJ, Gillis AJ, Stoop H, Beverloo B, Oosterhuis JW, Bokemeyer C, Looijenga LH. 2000. Restricted 12p amplification and RAS mutation in human germ cell tumors of the adult testis. Am J Pathol 157:1155–1166.

Slovak ML, Smith DD, Bedell V, Hsu YH, O'Donnell M, Forman SJ, Gaal K, McDaniel L, Schultz R, Ballif BC, Shaffer LG. 2010. Assessing karyotype precision by microarray-based comparative genomic hybridization in the myelodysplastic/myeloproliferative syndromes. Mol Cytogenet 3:23.

Sperger JM, Chen X, Draper JS, Antosiewicz JE, Chon CH, Jones SB, Brooks JD, Andrews PW, Brown PO, Thomson JA. 2003. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. Proc Natl Acad Sci USA 100:13350–13355.

Staaf J, Lindgren D, Vallon-Christersson J, Isaksson A, Göransson H, Juliusson G, Rosenquist R, Höglund M, Borg A, Ringnér M. 2008. Segmentation-based detection of allelic imbalance and loss-of-heterozygosity in cancer cells using whole genome SNP arrays. Genome Biol 9:R136.

Stephenson E, Ogilvie CM, Patel H, Cornwell G, Jacquet L, Kadeva N, Braude P, Ilic D. 2010. Safety paradigm: Genetic evaluation of therapeutic grade human embryonic stem cells. J R Soc Interface 7(Suppl 6):S677–S688.

Trask BJ, Massa H, Kenwrick S, Gitschier J. 1991 Mapping of human chromosome Xq28 by two-color fluorescence in situ hybridization of DNA sequences to interphase cell nuclei. Am J Hum Genet 48:1–15.

van Dartel M, Hulsebos TJM. 2004. Amplification and overexpression of genes in 17p11.2∼p12 in osteosarcoma. Cancer Genet Cytogenet 153: 77–80.

Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. 2007. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 17:1665–1674.

Wang K, Chen Z, Tadesse MG, Glessner J, Grant SF, Hakonarson H, Bucan M, Li M. 2008. Modeling genetic inheritance of copy number variations. Nucleic Acids Res 36:e138.

Yang SH, Jaffray E, Hay RT, Sharrocks AD. 2003. Dynamic interplay of the SUMO and ERK pathways in regulating Elk-1 transcriptional activity. Mol Cell 12:63–74.

Zafarana G, Grygalewicz B, Gillis AJ, Vissers LE, van de Vliet W, van Gurp RJ, Stoop H, Debiec-Rychter M, Oosterhuis JW, van Kessel AG, Schoenmakers EF, Looijenga LH, Veltman JA. 2003. 12p-amplicon structure analysis in testicular germ cell tumors of adolescents and adults by array CGH. Oncogene 22:7695–7701.