SHORT REPORT

Breakpoint mapping by whole genome sequencing identifies *PTH2R* gene disruption in a patient with midline craniosynostosis and a de novo balanced chromosomal rearrangement

Juwon Kim,[1] Hong-Hee Won,[2,3,4] Yoonjung Kim,[1] Jong Rak Choi,[5] Nae Yu,[5] Kyung-A Lee[5]

ABSTRACT

Background Craniosynostosis (CRS) is a premature closure of calvarial sutures caused by gene mutation or environmental factors or interaction between the two. Only a small proportion of non-syndromic CRS (NSC) patients have a known genetic cause, and thus, it would be meaningful to search for a causative gene disruption for the development NSC. We applied a whole genome sequencing approach on a 15-month-old boy with sagittal and metopic synostosis to identify a gene responsible for the development of the disease.

Methods and results Conventional chromosome study revealed a complex paracentric inversion involving 2q14.3 and 2q34. Array comparative genomic hybridisation did not show any copy number variation. Multicolour banding analysis was carried out and the breakpoints were refined to 2q14 and 2q34. An intronic break of the *PTH2R* gene was detected by whole genome sequencing and fluorescence in situ hybridisation analysis confirmed disruption of *PTH2R*.

Conclusions We report *PTH2R* as a gene that is disrupted in NSC. The disruption of the *PTH2R* gene may cause uncontrolled proliferation and differentiation of chondrocytes, which in turn results in premature closure of sutures. This addition of *PTH2R* to the list of genes associated with NSC expands our understanding of the development of NSC.

INTRODUCTION

Craniosynostosis (CRS) is defined by the premature osseous obliteration of one or more of the cranial vault sutures. The disruption of a fine balance between proliferation and differentiation during embryogenesis or early childhood can alter the patency of these sutures. Non-syndromic CRS (NSC) cases, which occur as an isolated anomaly, account for >80% of all CRS cases and most commonly affect sagittal sutures.[1] Although multiple reports have identified mutations in several genes such as *FGFR1-3*, *TWIST1/2*, *MSX2*, *FGFRL1*, *SNAIL*, *SLUG*, *NELL1* and *RUNX2* in NSC, only a small proportion of patients have a known genetic cause, which indicates that in addition to strong genetic components the premature closure of the sutures is a complex trait.[1] Recently, in an attempt to find a genetic cause of NSC, a locus near *BMP2* and within *BBS9* was implicated in the association of non-syndromic sagittal suture.[2] Each sutural synostosis shows distinct characteristics, even among cases with known genetic components. Of note, chromosomal aberrations and submicroscopic chromosomal rearrangements have been associated more frequently with midline synostoses.[3] Recent advances in whole genome sequencing (WGS) technologies have made it possible to identify genomic rearrangements and breakpoints involved in these rearrangements, facilitating rapid identification of disease genes in chromosomal breakpoint regions.[4] Here, our objective is to identify a gene that is disrupted to cause the disease in a patient with midline NSC using cytogenetic analysis and WGS.

MATERIALS AND METHODS

Patient information

The patient was a 15-month-old boy, the first child of healthy non-consanguineous parents without any family history of CRS. He was born at 40 weeks gestation by caesarean section with normal birth parameters (weight of 3660 g). He presented with hearing impairment and delayed development. At the age of 15 months, clinical examination revealed hypertelorism, crumpled ear, bulging anterior fontanelle and premature closure of the coronal suture. Brain MRI suggested sequelae of previous periventricular white matter injury (periventricular leukomalacia). A three-dimensional facial bone CT scan showed synostosis of the sagittal and metopic sutures (figure 1A). His weight, height and head circumference at the age of 18 months were 9000 g, 78 cm and 41.5 cm (less than third centile), respectively. A craniofacial distraction osteogenesis was performed and followed up with postoperative helmet moulding therapy. Chromosomal study of peripheral blood using the G-banding with trypsin-Giemsa (GTG) technique at 550 band resolution revealed a complex paracentric inversion involving 2q14.3 and 2q34 (figure 1B). His parents and sister had normal karyotypes.

Array comparative genomic hybridisation (CGH)

We considered local copy number (amplification and deletion) enriched regions using the NimbleGen CGX-12 array-chip (Roche NimbleGen, Wisconsin, USA) and the Agilent SurePrint G3 Human 1M Custom array CGH (Agilent Technologies Inc., California, USA), which was scanned using the Agilent Microarray Scanner, all according to the

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2015-103001).

[1]Department of Laboratory Medicine, Yonsei University Wonju College of Medicine, Wonju, Korea
[2]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA
[3]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA
[4]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
[5]Department of Laboratory Medicine, Yonsei University College of Medicine, Seoul, Korea

Correspondence to
Dr Kyung-A Lee, Department of Laboratory Medicine, Yonsei University College of Medicine, 211 Eonju-ro, Gangnam-gu, Seoul 135-720, Korea; kal1119@yuhs.ac

JK and H-HW contributed equally.

Received 13 January 2015
Revised 3 May 2015
Accepted 11 May 2015
Published Online First 4 June 2015

To cite: Kim J, Won H-H, Kim Y, *et al*. *J Med Genet* 2015;**52**:706–709.

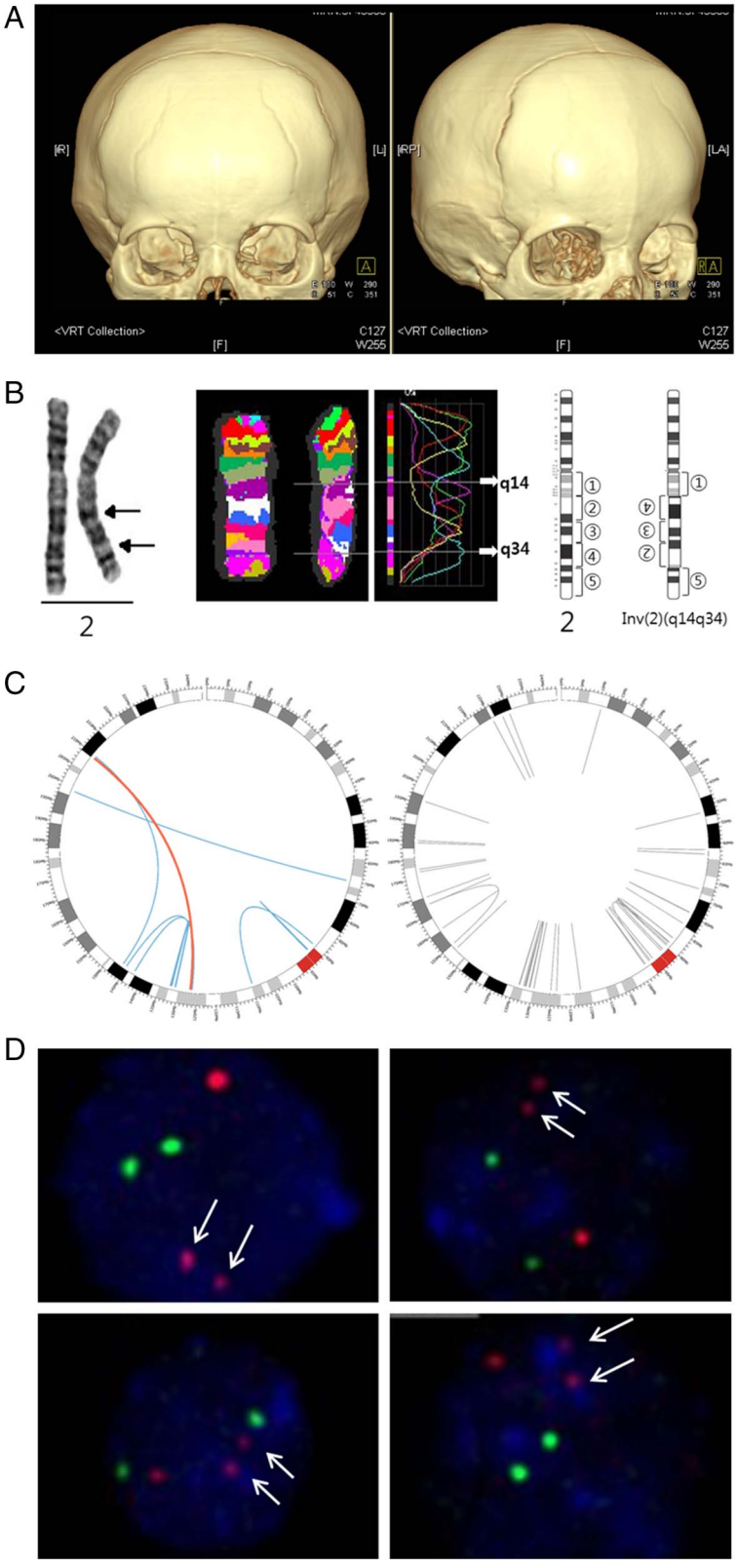


Figure 1 *PTH2R* is a gene that is disrupted in craniosynostosis. (A) Three-dimensional CT scan showing a synostosis of the metopic and sagittal sutures. (B) Conventional chromosome analysis, which revealed a paracentric inversion involving 2q14.3 and 2q34 and multicolour banding (mBAND) analysis refined breakpoints to 2q14 and 2q34. G-banding (left end), multicolour banding (middle) and partial G-banded ideogram (right end) of chromosome 2 are shown. (C) Circos plot of eight inversions on chromosome 2 detected by analysis of whole genome sequence data. Left and right figures show inversions observed in the patient and in 22 normal controls, respectively. Curves indicate both boundaries of each inversion detected by analysis of whole genome sequence data; blue and red curves represent eight candidate inversions that passed multiple filters and grey curves represent other inversions (see online supplementary table S1 for details). The red curve represents the inversion that led to the intronic break of the *PTH2R* gene, which resides on 2q34, and that was not observed in normal controls. This intronic break was consistent with the breakpoints detected by mBAND analysis and was confirmed by fluorescence in situ hybridisation (FISH) analysis. (D) FISH analysis showed one brighter red signal and two small red ones (arrow), indicating a breakpoint within the *PTH2R* gene. Cohybridisation of green 5-fluorescein 2′-deoxyuridine 5′-triphosphate-labelled control probes, which hybridised to the centromeric region of chromosome 2, revealed two normal green signals. L, left; LA, left anterior; R, right; RP, right posterior; VRT, volume rendering technique.

manufacturer's instructions. Copy number variations were detected by the ADM-2 algorithm in the Agilent Genomic Workbench Lite Edition 6.5.0.18.

Multicolour banding analysis

To better localise the breakpoints, multicolour banding (mBAND) analysis was carried out according to standard techniques and hybridised with a commercially available mBAND probe (MetaSystems, Altlussheim, Germany) for chromosome 2 (XCyte 2). Hybridisation, post-hybridisation washes and signal detection conditions were carried out following the method described by Tawn *et al*.[5]

Whole genome sequencing

We sequenced the entire genome of the patient using HiSeq 2000 (Illumina, California, USA) with paired reads of 101 base

pairs (bp) and inserts of 500 bp. We checked the quality of the sequence reads using the FastQC 0.1 software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and trimmed the last 41 bp sequences from reads using the NGSQCToolkit 2.2.3 software.[6] Using the BWA V.0.5.9 software, we aligned the paired-end reads to the human reference genome (February 2009 assembly of the human genome (hg19, GRCh37)).[7] In addition, we also used the first Korean individual genome sequence (SJK) data as a reference genome.[8]

Detection and filtering of inversions

We extracted anomalously mapped paired-end reads using the SVDetect 0.8 software,[9] including reads mapped on two different chromosomes with an incorrect strand orientation or pair order, or with an insert size longer than threefold of the SD from the mean. Based on calculation for one million reads, we set the mean insert size (μ) at 542.93 bp and the SD (σ) at 38.34 bp. We detected structural variations from the anomalously mapped reads using the BreakDancer max 1.1 software.[10] The fixed window size used by BreakDancer is defined as $w=\mu+3\sigma-2l$ bp, where μ and σ are the mean and the SD estimated from mapped read pairs, respectively, and l is the average read length. The inversions detected by BreakDancer were again filtered if they were not detected by the Manta software (Illumina).

In addition, we filtered clinically tolerable inversions observed in the case subject by comparing with inversions observed in normal subjects of Korean ancestry. Inversions of the case subject whose boundaries resided within the window interval *w* from those of normal subjects were excluded in further analysis.

Fluorescence in situ hybridisation

In order to confirm involvement of the *PTH2R* gene in chromosomal rearrangement, fluorescence in situ hybridisation (FISH) analysis was performed according to the manufacturer's instructions with commercially available *PTH2R* FISH probes and the control probe for chromosome 2q11 (Empire Genomics, New York, USA).

Variant calling

After mapping sequence reads to the human reference genome, we marked duplicate sequence reads using the Picard tool and recalibrated base quality scores using the Genome Analysis Toolkit (GATK) V.1.2 TableRecalibration. We realigned sequence reads using the GATK IndelRealigner. We performed variant calling and variant filtering using the GATK UnifiedGenotyper and VariantRecalibrator, respectively. We annotated the variants called with genotype quality >30 using the SnpEff v2.1b software with the GRCh37.66 reference genome and analysed 'high impact' variants annotated as nonsense, splice site and frameshift. We excluded variants with minor allele frequency >0.5% in all populations and East populations from the 1000 Genomes and ExAC data (Exome Aggregation Consortium, http://exac.broadinstitute.org (April 2015 accessed)).

RESULTS

The mBAND analysis was carried out and the breakpoints were refined to 2q14 and 2q34 (figure 1B). Array CGH analysis did not reveal any copy number variation in the vicinity of the breakpoints.

To precisely detect the breakpoints for the balanced inversions observed in karyotype analysis, we sequenced the patient's whole genome and generated paired-end sequences of 585 million reads with 118 billion bp. The sequence reads were mapped to the human reference genome and anomalously mapped reads were used to detect structural variants. Among a total of 800 inversions detected from the whole genome, 175 inversions were considered highly confident because these were supported by at least five sequence reads and had a BreakDancer confidence score >80, which was previously shown to indicate ~90% of the validation rate (see online supplementary table S1 and figure S1) and were not observed in 22 normal Korean individuals.[10] Furthermore, 73 inversions were also detected by Manta, and 39 inversions led to chromosomal breaks within 15 kilobases (kb) of nearby genes and, of those, 31 were expected to affect gene function because most regulatory elements fall close to transcribed region (figure 1C and see online supplementary table S2).[11] Assuming that the resolution of cytogenetic analysis is about 5 megabases (Mb), we discarded the inversions >5 Mb, which were not detected cytogenetically and were likely false-positives (see online supplementary tables S1 and S2). Analysing all chromosomes left only two inversions involving *PTH2R* or *LRP1B* (see online supplementary table S2; lines shaded in light blue) on chromosome 2 to be considered as a gene that might be disrupted to cause the disease. Since both involve 2p14.3 on the left side, only one of the two should be a plausible candidate. However, although the inversion involving *LRP1B* has a fragment size over 5 Mb, it was not visible in the mBAND analysis which allows delineation of chromosomal regions with a resolution of a few megabases and regarded as a false-positive. The quality per base of sequence reads decreased while the length of the reads increased and the 10th percentile of average quality score at the 60th bp position was lower than 20 (see online supplementary figure S2). To check sensitivity of our analysis, we performed additional analyses by removing the last 41 bp sequences from each read and mapping the trimmed reads to the GRCh37 and Korean (SJK) human reference genomes. All of the eight candidate inversions were detected by analysis of trimmed reads mapped to the two different genomes (see online supplementary figure S3).

Consequently, the disruption of the *PTH2R* gene was confirmed by FISH analysis (figure 1D). The directly labelled probes for *PTH2R* (red) and the control locus at 2q11 (green) were used and generated two green and three red (one normal red signal and two split red signals) signals corresponding to the 2q11 locus and *PTH2R*, respectively. The two small red signals indicated the break of the *PTH2R* gene, and their close proximity showed that the gene was separated due to an inversion within the chromosome.

To examine if homozygotes of rare variants affect the patient in a recessive form, we analysed variants that were predicted to have high impact on the corresponding protein (nonsense, splice site and frameshift) and were rare (minor allele frequency <0.5% in both all populations and East Asian populations). The patient harboured seven homozygotes for the variants, but none of the corresponding genes was biologically relevant to the disease phenotype (see online supplementary table S3).

DISCUSSION

We used whole genome paired-end sequencing to characterise breakpoints of a balanced chromosomal rearrangement in a patient with non-syndromic midline CRS. The whole genome analysis refined the breakpoint of the inversion to genomic location at 126 101 173 (2q14) and 209 310 827 (2q34) causing separation of the *PTH2R* gene at intron 7. The 2q14 breakpoint is about 430 kb and 1 Mb away from the nearest transcriptional units (see online supplementary figure S4). Consequently, the *PTH2R* gene is thought to be fused with the intergenic sequence

between *CNTNAP5* and *GYPC*. In genetic diseases, among the possible molecular consequences of chromosome aberrations in genetic diseases,[12] it is most likely that the phenotype was the result of deregulation of gene expression through direct disruption of the *PTH2R* gene in this case.

The *PTH2R* gene is a member of the family B (type II) of G-protein coupled receptors and is activated by parathyroid hormone and tuberoinfundibular peptide of 39 residues (TIP39) but not by parathyroid hormone-related protein (PTHrP).[13] In situ hybridisation revealed that PTH2R is distributed within the cells residing primarily in chondrocytes in the growth plate subarticular cartilage, and the expression was particularly strong in developing bones.[14] It was found that both TIP39, a ligand of PTH2R, and PTH2R are expressed in the growth plate of cartilage, with PTH2R being produced by chondrocytes in the resting/early proliferating zone while TIP39 is synthesised by chondrocytes in the prehypertrophic and hypertrophic zones.[15] While the main function of PTHrP/PTH1R signalling is to maintain chondrocytes in a proliferative state and prevent hypertrophy, TIP39/PTH2R signalling inhibits the proliferation of chondrocytes.[15] In a subsequent study, it was found that targeted expression of PTH2R in growth plate chondrocytes impaired endochondral ossification by decreasing proliferation while impairing differentiation of chondrocytes, which suggests that TIP39/PTH2R has a crucial role in the regulation of chondrocytic proliferation and differentiation.[16] Thus, it is likely that disruption of the *PTH2R* gene in our patient may have caused reduced expression of PTH2R or production of aberrant PTH2R, which led to decreased TIP29/PTH2R signalling and uncontrolled proliferation and differentiation of chondrocytes that in turn resulted in premature closure of sutures.

The IHH protein, which is expressed in prehypertrophic chondrocytes, is known to play a major role in endochondral bone formation by regulating proliferation and differentiation via the PTHrP controlled feedback loop.[16] The duplication of locus 2q35, which includes the *IHH* gene, causes increased IHH expression, which is likely to cause osteoblast proliferation and consecutive overgrowth that results in CRS.[17] IHH has also been shown to stimulate chondrocyte proliferation independently of PTHrP signalling, indicating a possible interaction between PTH2R and IHH in regulating chondrocyte proliferation.[18] In view of our findings, it will be interesting to determine the function of the *PTH2R/PTH2* complex using an animal model of CRS.

Acknowledgements Whole genome sequence data of Korean control subjects was provided by the Genome Research Foundation, Personal Genomics Institute, and Korean Personal Genome Project.

Contributors JK and H-HW performed experiments, analysed whole genome sequence data and wrote the manuscript. JRC was in charge of the patients, performed mBAND analysis and analysed data. YK and NY performed chromosome analysis, FISH and array CGH. K-AL coordinated and designed the study and reviewed the manuscript.

Funding This study was supported by a grant of the Korean Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (A120030).

Competing interests None declared.

Patient consent Obtained.

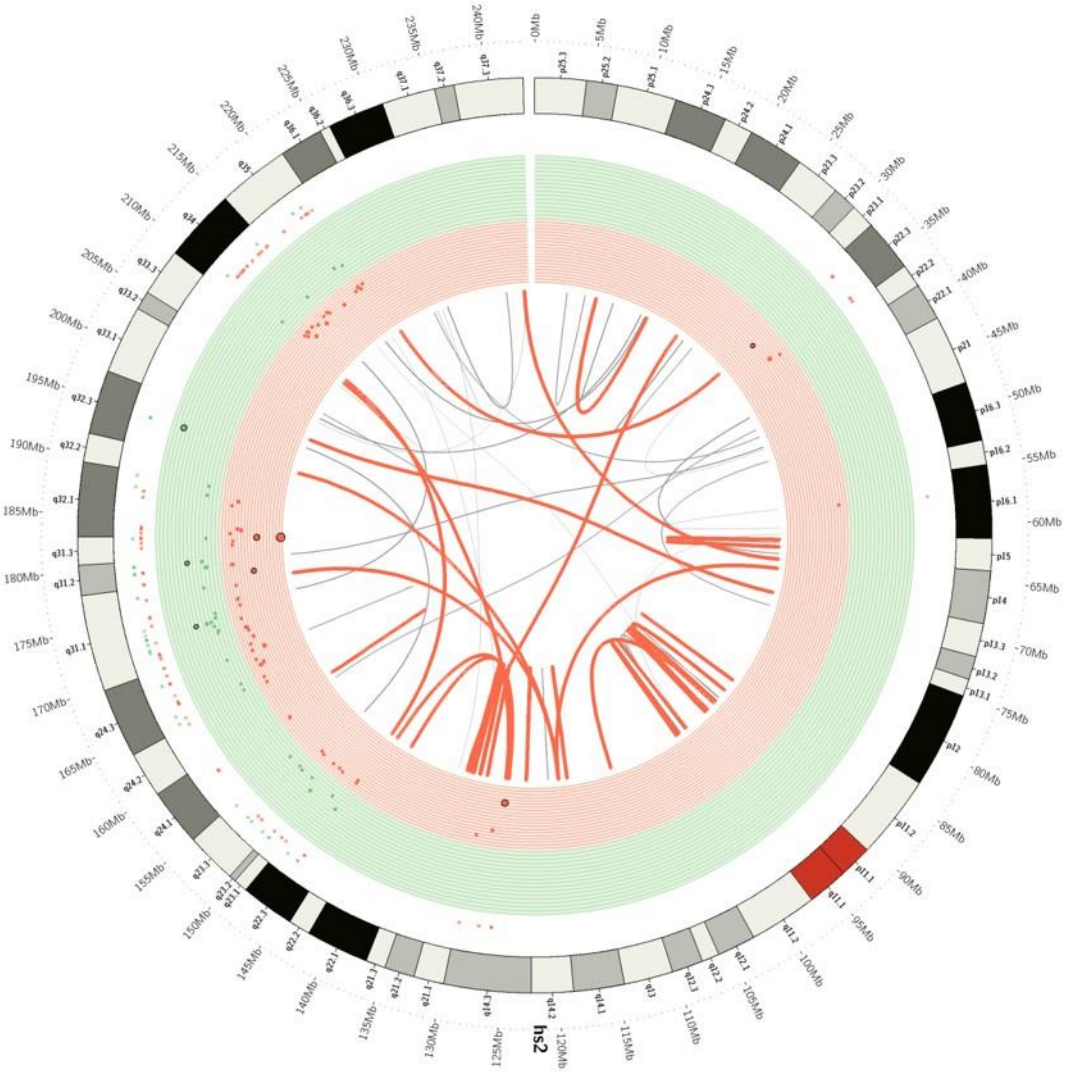
Ethics approval Institutional Review Board of the Gangnam Severance Hospital (IRB#3-2014-0075).

Provenance and peer review Not commissioned; externally peer reviewed.

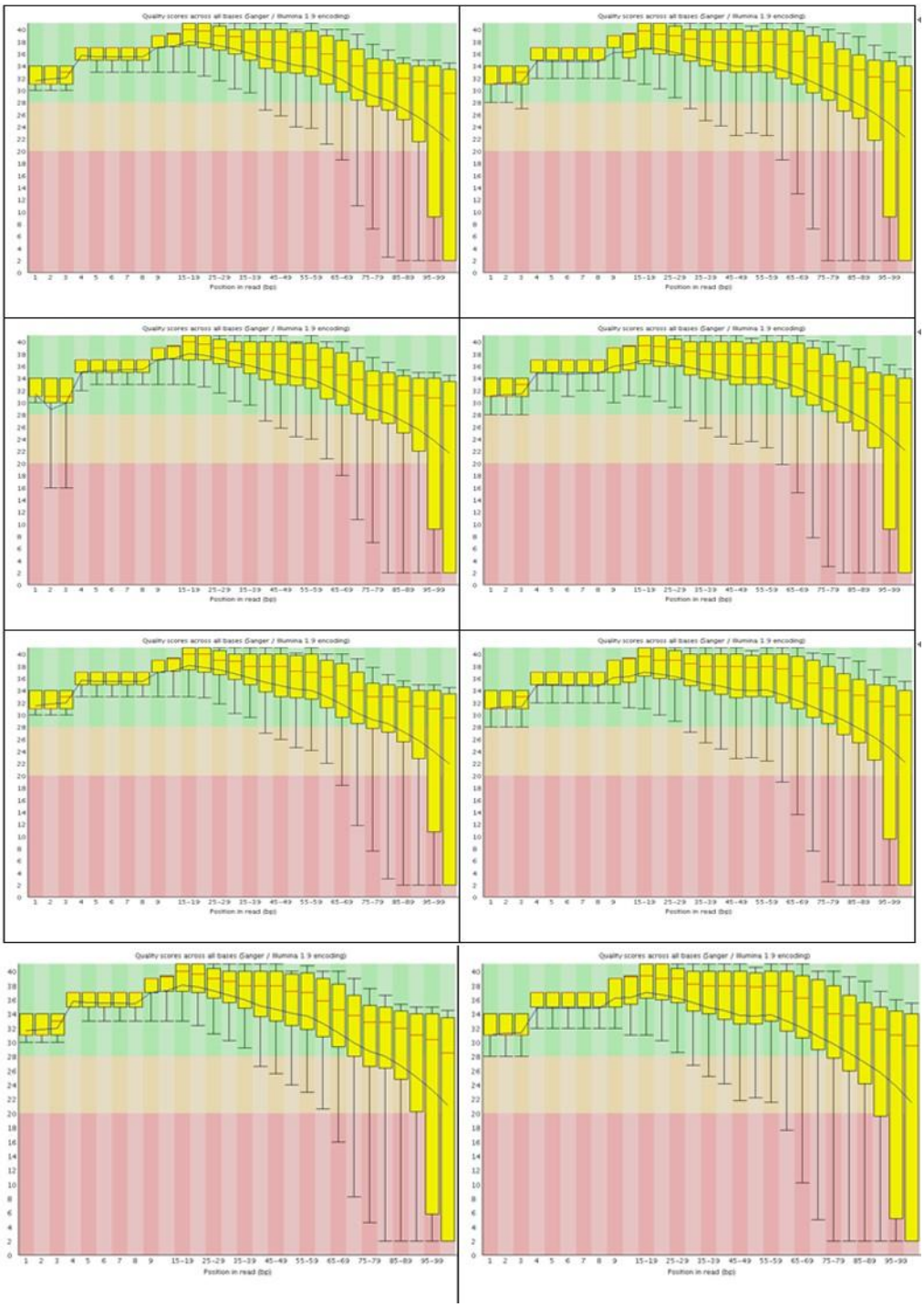
REFERENCES

1 Cohen MM, MacLean RE. *Craniosynostosis: diagnosis, evaluation, and management*. New York: Oxford University Press, 2000.
2 Justice CM, Yagnik G, Kim Y, Peter I, Jabs EW, Erazo M, Ye X, Ainehsazan E, Shi L, Cunningham ML, Kimonis V, Roscioli T, Wall SA, Wilkie AO, Stoler J, Richtsmeier JT, Heuze Y, Sanchez-Lara PA, Buckley MF, Druschel CM, Mills JL, Caggana M, Romitti PA, Kay DM, Senders C, Taub PJ, Klein OD, Boggan J, Zwienenberg-Lee M, Naydenov C, Kim J, Wilson AF, Boyadjiev SA. A genome-wide association study identifies susceptibility loci for nonsyndromic sagittal craniosynostosis near BMP2 and within BBS9. *Nat Genet* 2012;44:1360–4.
3 Lattanzi W, Bukvic N, Barba M, Tamburrini G, Bernardini C, Michetti F, Di Rocco C. Genetic basis of single-suture synostoses: genes, chromosomes and clinical implications. *Childs Nerv Syst* 2012;28:1301–10.
4 Schluth-Bolard C, Labalme A, Cordier MP, Till M, Nadeau G, Tevissen H, Lesca G, Boutry-Kryza N, Rossignol S, Rocas D, Dubruc E, Edery P, Sanlaville D. Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations. *J Med Genet* 2013;50:144–50.
5 Tawn EJ, Whitehouse CA, Holdsworth D, De Ruyck K, Vandenbulcke K, Thierens H. mBAND analysis of chromosome aberrations in lymphocytes exposed in vitro to alpha-particles and gamma-rays. *Int J Radiat Biol* 2008;84:447–53.
6 Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* 2012;7:e30619.
7 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
8 Ahn SM, Kim TH, Lee S, Kim D, Ghang H, Kim DS, Kim BC, Kim SY, Kim WY, Kim C, Park D, Lee YS, Kim S, Reja R, Jho S, Kim CG, Cha JY, Kim KH, Lee B, Bhak J, Kim SJ. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res* 2009;19:1622–9.
9 Zeitouni B, Boeva V, Janoueix-Lerosey I, Loeillet S, Legoix-ne P, Nicolas A, Delattre O, Barillot E. SVDetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data. *Bioinformatics* 2010;26:1895–6.
10 Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, Shi X, Fulton RS, Ley TJ, Wilson RK, Ding L, Mardis ER. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* 2009;6:677–81.
11 Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 2010;464:768–72.
12 Kloosterman WP, Hochstenbach R. Deciphering the pathogenic consequences of chromosomal aberrations in human genetic disease. *Mol Cytogenet* 2014;7:100.
13 Usdin TB, Bonner TI, Hoare SR. The parathyroid hormone 2 (PTH2) receptor. *Receptors Channels* 2002;8:211–18.
14 Usdin TB, Hilton J, Vertesi T, Harta G, Segre G, Mezey E. Distribution of the parathyroid hormone 2 receptor in rat: immunolocalization reveals expression by several endocrine cells. *Endocrinology* 1999;140:3363–71.
15 Panda D, Goltzman D, Juppner H, Karaplis AC. TIP39/parathyroid hormone type 2 receptor signaling is a potent inhibitor of chondrocyte proliferation and differentiation. *Am J Physiol Endocrinol Metab* 2009;297:E1125–36.
16 Panda DK, Goltzman D, Karaplis AC. Defective postnatal endochondral bone development by chondrocyte-specific targeted expression of parathyroid hormone type 2 receptor. *Am J Physiol Endocrinol Metab* 2012;303:E1489–501.
17 Klopocki E, Lohan S, Brancati F, Koll R, Brehm A, Seemann P, Dathe K, Stricker S, Hecht J, Bosse K, Betz RC, Garaci FG, Dallapiccola B, Jain M, Muenke M, Ng VC, Chan W, Chan D, Mundlos S. Copy-number variations involving the IHH locus are associated with syndactyly and craniosynostosis. *Am J Hum Genet* 2011;88:70–5.
18 Chung UI, Schipani E, McMahon AP, Kronenberg HM. Indian hedgehog couples chondrogenesis to osteogenesis in endochondral bone development. *J Clin Invest* 2001;107:295–304.

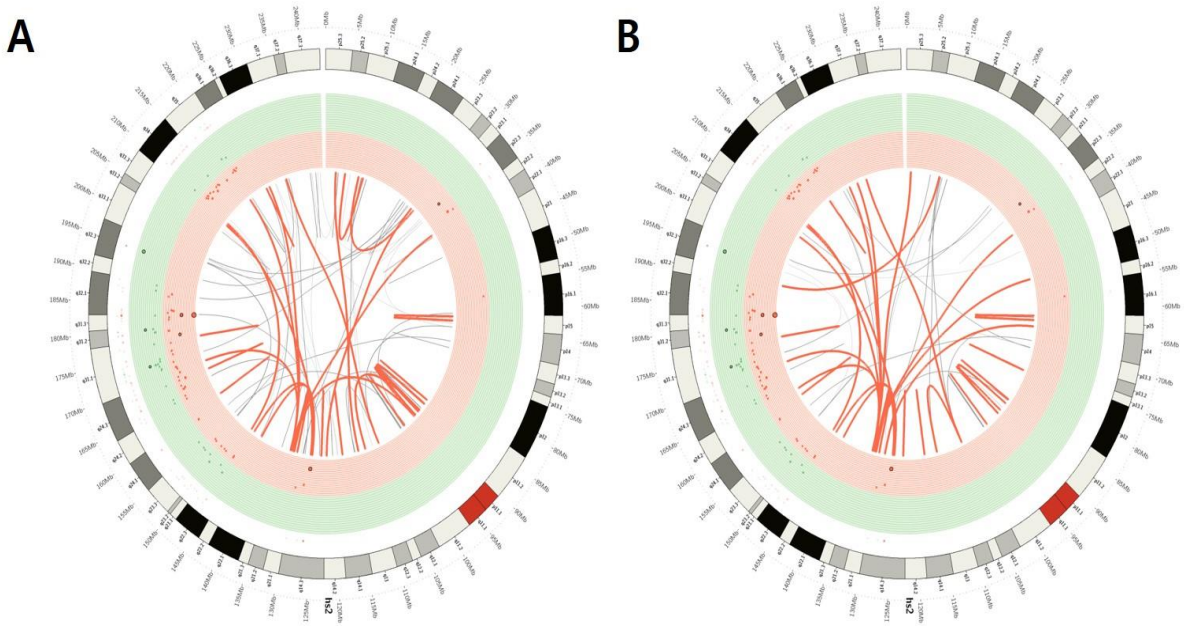
Supplementary figure 1S. Circos plot of inversions on chromosome 2 detected by whole genome sequence data. Outer green dots and inner red dots represent copy number amplification and deletion in the array comparative genomic hybridization **(**CGH), respectively. Red curves are plotted to point both boundaries of each inversion with a high confidence score (>80) and grey curves are plotted to point to those with lower confidence scores.



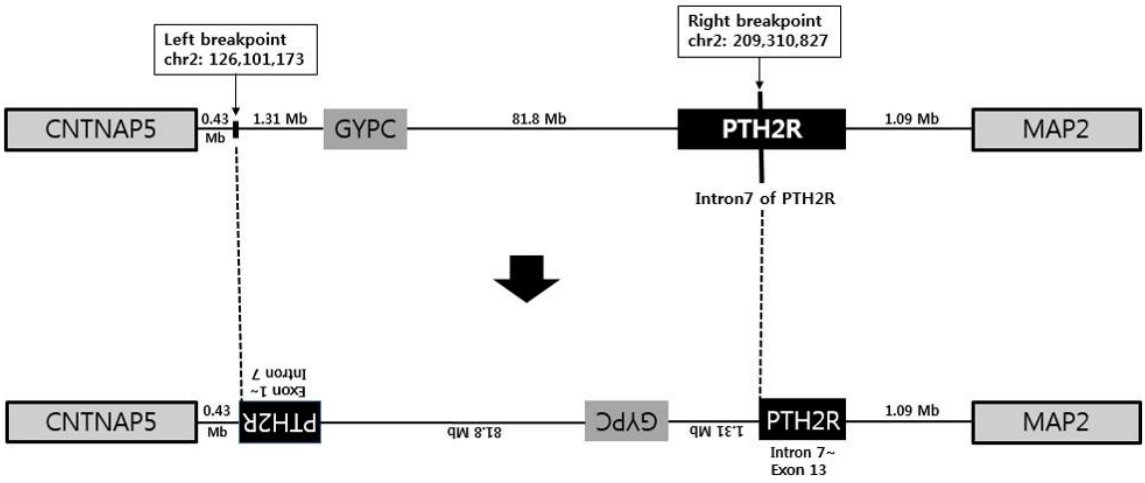
Supplementary figure 2S. Quality per base of sequence reads. Quality per base of sequence reads decreased while the length of the reads increased and the 10th percentile of the average quality score at the 60th base pair position was lower than 20.



Supplementary figure 3S. Circos plot of inversions on chromosome 2 detected by trimmed reads. Trimmed reads were mapped to (A) GRCh37 and (B) Korean (KSJ) human reference genomes. Outer green dots and inner red dots represent copy number amplification and deletion in the array CGH, respectively. Red curves are plotted to point to both boundaries of each inversion with high confidence scores (>80) and grey curves are plotted to point those with lower confidence scores.



Supplementary figure 4S. Schematic overview of genomic position of the analyzed inversion and neighboring genes. The genomic positions of the 5' and 3' breakpoints are given before (upper) and after (Lower) inversion, and the sizes (Mb) of the fragments involved are indicated. The whole genome analysis validated the breakpoint of inversion to genomic location at 126,101,173 (2q14) and 209,310,827 (2q34) causing separation of the *PTH2R* gene at intron 7. The 2q14 breakpoint is about 430kb and 1Mb away from the nearest transcriptional units, and those nearby transcriptional units (*CNTNAP5* and *GYPC*) are not be disrupted by the inversion. Consequently, the *PTH2R* gene is thought to be fused with the intergenic sequence between *CNTNAP5* and *GYPC*.



SUPPLEMENTARY TABLES

Supplementary table 1S. Identification of candidate inversions on all chromosomes and chromosome 2

Filters	**Number of inversions passing filters**	
	All chromosomes	**Chromosome 2**
Inversions detected by BreakDancer	800	75
Inversions with confidence score >80	352	38
Inversions supported by at least five reads	303	35
Inversions not found in 22 control samples	175	21
Inversions detected by Manta	73	8
Genes near inversion boundaries (<15 kilobases)	39	6
Intronic or exonic breaks near inversion boundaries	31	4

Supplementary table 2S. 31 inversions selected after all the filtration process of whole genome sequence data

CHR#	Left boundary	Annotation	Genes	Left band	CHR#	Right boundary	Annotation	Genes	Right band	Size (Mb)
1	17185929	upstream	MIR3675	p36.13	1	145382685	intronic	NBPF10, NBPF20	q21.1	128.2
1	54271697	intronic	NDC1	p32.3	1	55706676	intergenic	MIR4422(dist=15280), PPAP2B(dist=1253743)	p32.3	1.4
1	147731513	intronic	NBPF8	q21.1	1	148559669	ncRNA_intronic	NBPF25P	q21.2	0.8
1	145109494	intronic	NBPF20, NBPF9, SEC22B	q21.1	1	148184080	intronic	NBPF8	q21.2	3.1
2	130944046	intronic	MZT2B	q21.1	2	132246786	ncRNA_intronic	RNU6-81P	q21.2	1.3
2	70184039	intergenic	MXD1(dist=13963), ASPRV1(dist=3185)	p14	2	198761941	intronic	PLCL1	q33.1	128.6
2	126701173	intergenic	CNTNAP5(dist=1028219), GYPC(dist=712338)	q14.3	2	209310827	intronic	PTH2R	q34	82.6
2	126700323	intergenic	CNTNAP5(dist=1027369), GYPC(dist=713188)	q14.3	2	142666659	intronic	LRP1B	q22.2	16.0
3	187860533	intergenic	BCL6(dist=397020), LPP-AS2(dist=8461)	q27.3	3	193081071	exonic	ATP13A5	q28	5.2
3	187860533	intergenic	BCL6(dist=397020), LPP-AS2(dist=8461)	q27.3	3	193081071	exonic	ATP13A5	q28	5.2

3	187860533	intergenic	BCL6(dist=397020), LPP-AS2(dist=8461)	q27.3	3	193080245	exonic	ATP13A5	q28	5.2
3	187860533	intergenic	BCL6(dist=397020), LPP-AS2(dist=8461)	q27.3	3	193080245	exonic	ATP13A5	q28	5.2
4	71878179	intronic	DCK	q13.3	4	71956181	intergenic	DCK(dist=59552), SLC4A4(dist=96822)	q13.3	0.1
7	149737329	intergenic	ATP6V0E2(dist=159528), ACTR3C(dist=206972)	q36.1	7	153760892	intronic	DPP6	q36.2	4.0
8	48844731	intronic	PRKDC	q11.21	8	48847240	intronic	PRKDC	q11.21	0.0
11	24509946	intergenic	MIR8054(dist=1069210), LUZP2(dist=8570)	p14.3	11	34973254	intronic	PDHX	p13	10.5
11	24509946	intergenic	MIR8054(dist=1069210), LUZP2(dist=8570)	p14.3	11	34973254	intronic	PDHX	p13	10.5
12	133141845	intronic	FBRSL1	q24.33	12	133344199	intergenic	ANKLE2(dist=5725), GOLGA3(dist=1296)	q24.33	0.2
12	133141845	intronic	FBRSL1	q24.33	12	133344199	intergenic	ANKLE2(dist=5725), GOLGA3(dist=1296)	q24.33	0.2
14	58259536	intronic	SLC35F4	q23.1	14	58412621	intergenic	SLC35F4(dist=80029), C14orf37(dist=58187)	q23.1	0.2
15	79057976	intronic	ADAMTS7	q25.1	15	82620316	ncRNA_exonic	ADAMTS7P1	q25.2	3.6
16	24121355	intronic	PRKCB	p12.1	16	80277677	ncRNA_intronic	LOC102724084	q23.2	56.2

17	746316	intronic	NXN	p13.3	17	11598770	intronic	DNAH9	p12	10.9
19	53193320	intronic	ZNF83	q13.32	19	53869176	ncRNA_intronic	ZNF525	q13.33	0.7
19	53193320	intronic	ZNF83	q13.32	19	53869176	ncRNA_intronic	ZNF525	q13.33	0.7
19	53193320	intronic	ZNF83	q13.32	19	53870170	ncRNA_intronic	ZNF525	q13.33	0.7
19	53193320	intronic	ZNF83	q13.32	19	53870170	ncRNA_intronic	ZNF525	q13.33	0.7
19	10447827	intronic	ICAM3	p13.2	19	31795439	intronic	TSHZ3	q12	21.3
19	10869740	intronic	DNM2	p13.2	19	54685855	intronic	MBOAT7	q13.33	43.8
X	149570895	intronic	MAMLD1	q28	X	149585449	intronic	MAMLD1	q28	0.0
X	49013681	intergenic	GPKOW(dist=33530), MAGIX(dist=5500)	p11.23	X	49019827	intronic	MAGIX	p11.23	0.0

Supplementary table 3S. List of genes highly affected by homozygous variants

CHR#	Position	Variant ID	Ref	Alt	Effect	Gene
3	194080395	rs113243841	G	A	Nonsense	*LRRC15*
4	155244401	rs112727159	TTTTG	T	Frameshift	*DCHS2*
4	3590823	rs34083130	GACAC	G	Splice site acceptor	*RP3-368B9.1.1*
6	96034869	NA	G	GTATA	Frameshift	*MANEA*
13	100517195	rs3831038	CTG	C	Frameshift	*CLYBL*
17	71188971	rs57350092	G	GC	Splice site donor	*COG1*
22	20779973	rs5844418	C	CG	Frameshift	*SCARF2*

* The 1000 Genomes, and ExAC databases were consulted upon identification of variants.