Genomics 101 (2013) 134–138

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Gene-based copy number variation study reveals a microdeletion at 12q24 that influences height in the Korean population

Yun Kyoung Kim [1], Sanghoon Moon [1], Mi Yeong Hwang, Dong-Joon Kim, Ji Hee Oh, Young Jin Kim, Bok-Ghee Han, Jong-Young Lee, Bong-Jo Kim *

Division of Structural and Functional Genomics, Center for Genome Science, National Institute of Health, Chungcheongbuk-do, 363-951, Republic of Korea

A R T I C L E I N F O

Article history:
Received 2 July 2012
Accepted 1 November 2012
Available online 10 November 2012

Keywords:
Copy number variation
Height
Genome-wide association study
Structural variation
Deletion

A B S T R A C T

Height is a classic polygenic trait with high heritability ($h^2=0.8$). Recent genome-wide association studies have revealed many independent loci associated with human height. In addition, although many studies have reported an association between copy number variation (CNV) and complex diseases, few have explored the relationship between CNV and height. Recent studies reported that single nucleotide polymorphisms (SNPs) are highly correlated with common CNVs, suggesting that it is warranted to survey CNVs to identify additional genetic factors affecting heritable traits such as height.

This study tested the hypothesis that there would be CNV regions (CNVRs) associated with height nearby genes from the GWASs known to affect height. We identified regions containing > 1% copy number deletion frequency from 3667 population-based cohort samples using the Illumina HumanOmni1-Quad BeadChip. Among the identified CNVRs, we selected 15 candidate regions that were located within 1 Mb of 283 previously reported genes. To assess the effect of these CNVRs on height, statistical analyses were conducted with samples from a case group of 370 taller (upper 10%) individuals and a control group of 1828 individuals (lower 50%).

We found that a newly identified 17.7 kb deletion at chromosomal position 12q24.33, approximately 171.6 kb downstream of *GPR133*, significantly correlated with height; this finding was validated using quantitative PCR. These results suggest that CNVs are potentially important in determining height and may contribute to height variation in human populations.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Height is a typical polygenic complex trait with high heritability ($h^2=0.8$) [1]. Genome-wide association studies (GWASs) have identified hundreds of variants associated with human height in 180 genetic loci [2]. Most of these association studies have investigated common single nucleotide polymorphisms (SNPs). These common variants, however, explain only a small percentage of heritable variation in height [3].

CNVs are multi-kilobase genomic regions that include varying degrees of deletion and duplication in different individuals [4]. Recent studies have suggested an important role for CNVs in complex traits [5–7]. To date, two studies have reported an association between CNVs and height in children and in Chinese populations [8,9]. Although these studies have confirmed that CNVs may contribute to variation in height, the specific role of CNVs in height determination is still largely unknown. Recent CNV studies, however, have shown that common CNVs were highly correlated with SNPs and that most common CNVs contained relatively high densities of SNPs [5]. Therefore investigation of CNVs nearby genes associated with SNPs will contribute to unraveling the genetic basis of complex diseases and phenotypes.

This study tested the hypothesis that CNVRs affecting height would be located near genes associated with SNP previously reported to be associated with height. To investigate the contribution of these CNVRs to height, we carried out an association study between selected CNVRs and height in 3667 Korean individuals using the Illumina HumanOmni1-Quad BeadChip. This study revealed a deletion, which had not previously been reported, that was associated with height. This finding provides a new perspective on our understanding of the genetic basis of height variation.

2. Results

2.1. Characteristics of discovered CNVs

We identified 657,466 autosomal CNVs, including 460,395 deletions and 197,071 duplications, within the genome of 3667 individuals. The average CNV size was approximately 11.7 kb (Supplementary Table 1) and the median and the mean number of CNVs were 15 and 17.31,

* Corresponding author at: Division of Structural and Functional Genomics, Center for Genome Science, Korea National Institute of Health, Chungcheongbuk-do, 363-951, Republic of Korea. Fax: +82 43 719 8908.
E-mail address: kbj6181@cdc.go.kr (B.-J. Kim).
[1] These authors contributed equally to this work.

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2012.11.002

respectively. Regions of more than 1 kb with at least 10 consecutive probes were designated CNVRs. Among these, there were 805 common CNVRs with >1% CNV frequency (Supplementary Fig. 1). Distributions of CNVs in the case group and control group are shown in Supplementary Fig. 2.

2.2. Selecting candidate CNVRs nearby known genes

Using the GWAS Integrator, we identified 276 loci within the HuGE Navigator database that correlated with height. Excluding redundant genes, we obtained a list of 283 non-redundant genes (see Supplementary Table 2). We selected a set of 11 CNVRs that were determined to be within 500 kb of one of 283 genes previously correlated with height and that fell into the “multiple-class” genotype cluster (Fig. 1; Supplementary Fig. 3; Supplementary Table 3). Regarding the quality score of the CNVtools, all of the multiple-class CNVRs were greater than threshold 4 which was used in the WTCCC CNV association study [5].

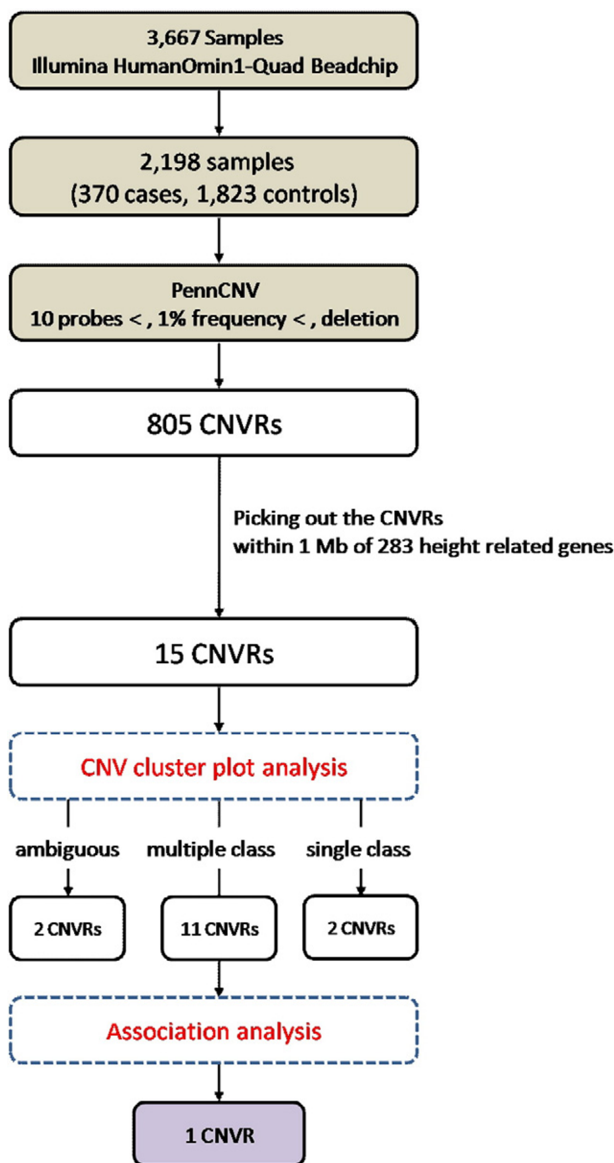

Fig. 1. Overall scheme of this study. We conducted two stage analyses to estimate more exact CNV genotypes. In the first stage, we used the PennCNV software to examine CNVRs. And then, we selected 15 CNVRs that were within the 1 Mb ranges of genes previously correlated with height. In the second stage, we estimated CNV genotypes of 15 CNVRs using the CNVtools software package regarding signal intensity values. After regression analysis, only 1 CNVR significantly correlated with height.

2.3. Statistical analysis

The tallest 10% of our study subjects (370 individuals) were classified as our case group and the shortest 50% (1828 individuals) were classified as controls. Logistic regression analyses revealed only 1 well-clustered multiple-class CNVR, near the gene *GPR133* at chromosomal position 12q24.33, that showed significant differences between the case group and control group ($P=0.0043$; $P<0.05$ after Bonferroni correction) (see Figs. 2(A) and (B); Supplementary Fig. 4). The result of the regression analysis is shown in Table 2.

We also conducted a linear regression analysis between identified CNVRs and height. The CNVR located at 12q24.33 showed significant association with height ($P=0.0096$, adjusted for age and gender; Supplementary Table 4).

2.4. Assessment of accuracy of CNVRs by experimental validation

To confirm the reliability of CNVR designation, we performed experimental validation with the TaqMan Copy Number Assay. We defined positive predictive value (PPV) as the standard of accuracy and verified that the deletions of several CNVs were correctly predicted with high positive predictive value (PPV = 0.97; Table 3 and Fig. 2(C)).

2.5. Genetic difference of CNVR near the gene GPR133 at 12q24.33 among populations

To identify the genetic difference of this CNVR, we compared the minor allele frequency (MAF) of SNP, rs1569019, that was a previously reported height associated SNP located on 12q24.33 in different populations as a proxy for this CNVR (Supplementary Table 5) [10].

For comparison of genetic diversity, we also analyzed the CNV genotype information of HapMap samples. HapMap Europeans showed only diploid copy numbers while one sample with a copy deletion was observed in both HapMap Asians and Yoruba (see supplementary Table 6) [11].

3. Discussion

This study examined the association of CNVs with human height. The hypothesis underlying this study was that common CNVRs would neighbor genes involved in height. This study highlighted technical issues with the CNV genotyping process and the importance of selecting well-defined CNV genotypes for association study. Recently, Wineinger et al. assessed statistical approaches to estimate CNVs and their effect on downstream association analyses [12]. They reported that estimations of copy number deletions are less prone to error than copy number duplications and that estimates were greatly improved by setting a threshold of 10 consecutive probes in genome array studies [12]. Based on this knowledge, we identified CNV deletion regions with greater than 1% frequency, that were larger than 1 kb in length and that were detected by more than 10 consecutive probes on the Illumina BeadChip. We examined the association of these CNVRs with multiple near target genes which have been reported to influence height from GWAS.

A deletion at chromosomal position 12q24.33, neighboring *GPR133* showed the most notable difference in frequency between taller and shorter individuals. Tall individuals had about a 2 fold higher deletion frequency at this position than the shorter controls (Table 2). Previous study has reported that several SNPs near *GPR133*, that fall in 12q24.33 region, are associated with height [10]. The association between this CNV locus and height, however, has not previously been reported. Our logistic and linear regression analyses showed that the variation in the size of this CNVR has a greater effect on height (height range ~ 0.7–1.2 cm) than the formerly reported height-associated SNP near *GPR133*, rs1569019 (height range 0.949 cm) [10].

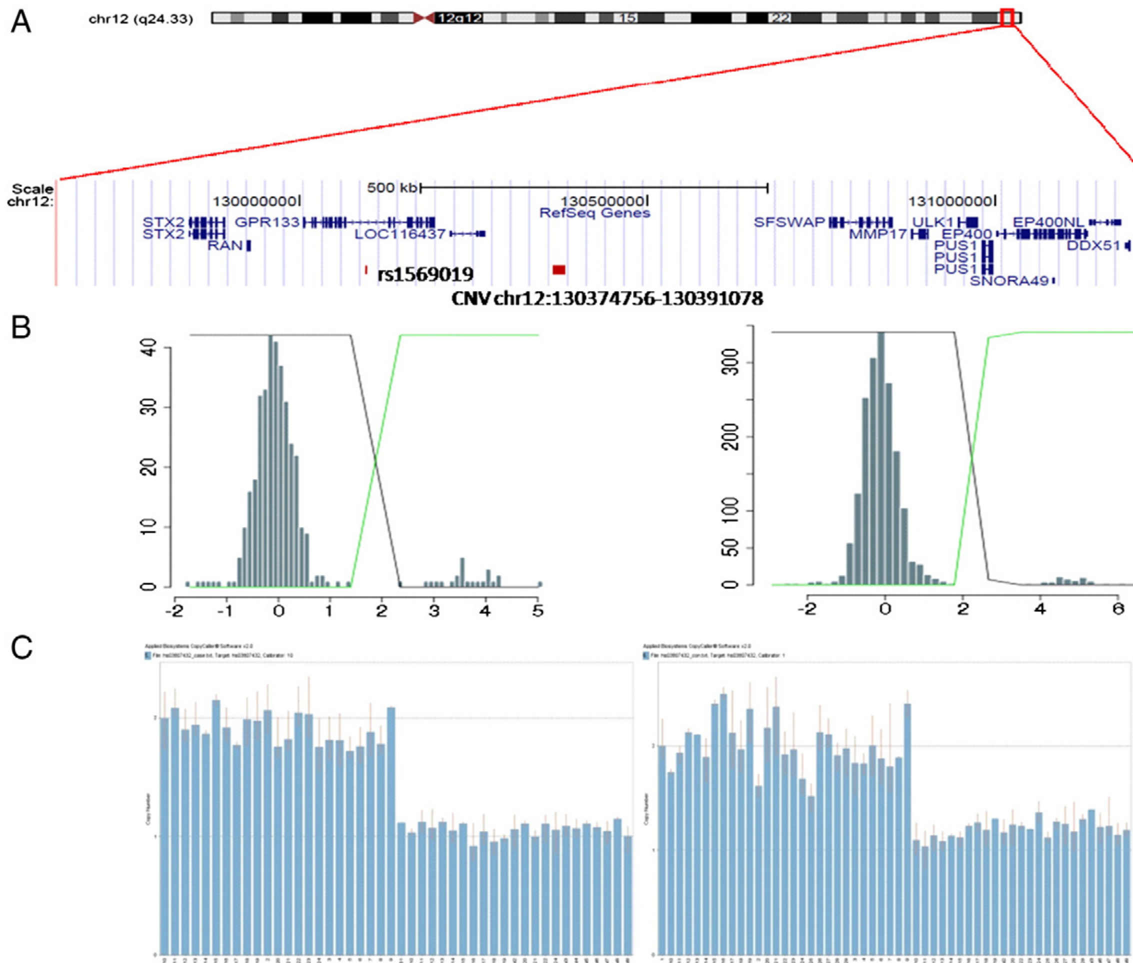

Fig. 2. A deletion at chromosomal position 12q24.33 is significantly associated with height. (A) Position of the CNV deletion within 12q24.33. The identified CNVR in our study is marked by red filled square and red vertical line correspond to the position of rs1569019 reported by Tonjes et al. (B) Histograms depict LDF plot of a deletion in 12q24.33 using the CNVtools. Copy number genotypes of this region were clearly separated into two groups (normal and hemizygous deletion). Left side and right side of histogram represent case and control, respectively. Higher peak of each histogram represents an estimated normal copy (2 copy), whereas lower peak represents an estimated hemizygous deletion (1 copy). (C) Quantitative PCR results of the CNV region on 12q24.33 near *GPR133*. We conducted validation experiment for 4 states: normal copy and single copy for cases and normal copy and single copy for controls. Samples in each state were randomly selected. The copy number state of cases (left) and control samples (right). Higher bars of each figure represent a validated normal copy (2 copy) whereas lower bars represent a validated hemizygous deletion (1 copy).

Tonjes et al. [10] and our study strongly support the association of the locus around *GPR133* with height. Given reported genetic impact of CNV, the deletion in this study would have more strong genetic effect than previously reported SNP at 12q24.33. In this context, one might think that previously height associated SNP rs1569019 in Europeans may be a proxy for the CNV locus at 12q24.33 in this study. However, rs1569109 was not effective tagger for the CNV in Koreans (see Supplementary Fig. 5). Moreover, rs1569109 was not significantly associated with height in Koreans. This would imply possible genetic diversity at 12q24.33 among populations. Interestingly, multiple studies reported that distinct genetic pattern was observed at 12q24 across populations [13–15]. In this study, we observed difference in MAF of rs1569109 between Asians and Europeans (see supplementary Table 5). In detail, MAF of CEU is 0.106 whereas those of CHB and JPT are 0.012 and 0.017, respectively (MAF of Korean population = 0.019). We further examined a linkage disequilibrium (LD) block patterns near rs1569019 for each ethnics using Haploview software (Supplementary Fig. 6) [16]. As discussed above, distinct pattern of LD blocks were shown among populations. Given the population genetic diversity in genomic region of the locus near *GPR133*, the designated region of Asians may have different genetic effect on height compared to Europeans. This difference in genetic architecture was also shown in CNVs. We analyzed the CNV genotype information of HapMap samples. HapMap Europeans showed only diploid copy numbers while one sample with a copy deletion was observed in both HapMap Asians and Yoruba (see Supplementary Table 6) [11]. Together, these results imply that *GPR133* at 12q24.33 is associated with height yet different genetic effect exists among populations. Therefore, CNVR that we identified in this study may have a stronger effect on height than rs1569019 in Korean population.

GPR133 encodes the adhesion G-protein-coupled receptor (GPCR or GPR), a membrane-bound protein with a long N termini containing seven transmembrane domains [17]. GPCRs transduce extracellular signals through heterotrimeric G proteins and are known to play a role in regulating most physiological functions, especially cell growth and development [17,18]. Activating mutations of the Gα_s subunit of G protein have been found in growth hormone-secreting pituitary tumors. These properties present a rationale for the association of *GPR133* variations with height [19]. Moreover, variants in *GPR126*, another GPCRs, have also been shown to be associated with height, including trunk length [20–22].

Interestingly, various transcription factor-binding sites have been identified within this CNV region. Among them, binding sites for three transcription factors, BATF, BCL3 and NF-kappa B have high cluster scores, 530, 567 and 1000, respectively (see Supplementary Fig. 7). BATF controls the differentiation of helper T cells [23]. BCL3 and NF-kappa B also control the differentiation of helper T cells, deeply engaged in the immune response, and interact strongly with each other [24]. The role of these transcription factors in cell differentiation

Table 1
Basic characteristics of the study subjects.

Trait	Total (N = 3667)	Male (N = 1375)	Female (N = 2292)
Age (years)	59.83 ± 10.05	61.39 ± 9.89	58.90 ± 10.02
Height (cm)	157.48 ± 8.55	165.17 ± 6.38	152.88 ± 5.99
BMI (kg/m^2)	23.79 ± 3.07	23.49 ± 2.98	23.97 ± 3.10

BMI, body mass index; data are shown as mean ± standard deviation.

suggests a possible relationship with cell growth in general and, by extension, with height.

It should be noted that this study has some limitations. First, we did not confirm the present finding in other populations. Second, the platform used in this study (SNP microarray) has lower resolution for CNV detection than array comparative genomic hybridization (aCGH) or next-generation sequencing (NGS) [25]. Therefore, further studies are necessary to replicate our findings in other large-scale populations. Higher resolution technologies, such as NGS, will also help to uncover causal relationships between CNVs and height.

In conclusion, we have examined the association of selected CNVRs and height in a select population. We found a CNVR located near *GPR133* which may play a key role in regulating human height. This study suggests an approach that may be useful in uncovering other genomic variations associated with complex traits such as height.

4. Materials and methods

4.1. Study subjects

All of study subjects were selected from the NongChon1 cohort, one of the Korean Genome Epidemiology Study (KoGES) population-based cohort, that enrolled 5678 healthy individuals aged 38 to 90 living in Yangpyeong, Goryeong and Namwon. Participants responded to a written/oral epidemiological questionnaire and were subjected to a series of laboratory tests to establish biochemical properties. A total of 3667 individuals (1375 male and 2292 female) were deemed appropriate for this study based on the SNP quality control criteria (filtering out samples with call rate ≤ 99%, excessive heterozygosity, inconsistent gender, duplication, and cancer patients). The report file containing input signal intensity of the 3667 samples was converted using the Illumina BeadStudio software package.

4.2. Phenotyping

Height was measured as the maximum distance from the floor to the highest point on the head after removal of shoes using a stadiometer and recorded to the nearest 0.1 cm. The average height of males was 165.2 ± 6.4 cm and body mass index (BMI) was 23.5 ± 3.0 kg/m^2. The average height and BMI of females was 152.9 ± 6.0 cm and 24.0 ± 3.1 kg/m^2, respectively. Females made up 62.5% of the test group (2292 of 3667). Basic characteristics of our subjects are summarized in Table 1. With consideration of age and gender, the upper 10% in height (370 individuals) were defined as the "taller" case group and the lower 50% in height (1828 individuals) were defined as the "control" group.

4.3. Characteristics of CNV genotyping platform

Genomic DNA was extracted from peripheral blood with informed consent. Genotyping was performed using the Illumina HumanOmni1-Quad BeadChip, with ~200 ng of genomic DNA quantified by using a NanoDrop spectrophotometer. This BeadChip contains markers for 1,140,419 SNPs and 123,996 CNVs with median CNV marker spacing of 1.2 kb.

4.4. CNV discovery

It has been noted that detection of copy number duplication is more prone to error than those of deletions [12]. For this reason, we only focused on detection of common autosomal CNV deletions (> 1% frequency) that contained at least 10 consecutive probes greater than 1 kb in length as determined by the PennCNV software (Fig. 1) [26]. The chromosomal coordinates of the identified CNVs were determined using the UCSC version hg18/NCBI Build 36 database.

4.5. CNV cluster plot analysis

Traditionally, genotypes of CNVs depended on threshold such as segmental mean value (log R ratio) which is chosen by users. This approach is prone to error because thresholds are set strictly to reduce spurious calls. The recent CNV association study of the Wellcome Trust Case Control Consortium (WTCCC) reported that only about 40% of CNVs that they identified were well separated enough to be genotyped [5]. Moreover, due to the poor signal-to-noise ratio of CNVs relative to SNPs, the WTCCC study used statistical method to estimate CNV genotype. Wineinger et al. recently recommended signal intensity as an alternative approach to CNV genotype estimation [12]. Moreover, some studies have used a mixture-based model to separate samples into more exact CNV genotype groups [5,27,28], Similarly, we used two different software packages, PennCNV [26] and CNVtools [28], for CNV detection and genotyping, respectively. Using this approach, we assigned our selected CNVRs to three genotype groups; "single-class", "multiple-class", and "ambiguous" [5,27]. Only CNVRs in the "multiple-class" group were used for further analysis (Fig. 1).

4.6. Candidate CNVR selection

We targeted CNV deletion regions within 500 kb (upstream and downstream) of non-redundant genes reported to be involved in height determination from GWASs using the GWAS Integrator of the HuGE Navigator knowledge base (see Supplementary Table 2) [29].

4.7. Statistical analysis

Associations of selected CNVRs and height were performed by logistic regression analysis using the SAS program (version 9.1; SAS Institute Inc.). We only considered CNVRs that showed nominal significance between cases and controls ($P < 0.05$). For the multiple comparisons, we performed the Bonferroni correction. Adjusted *P*-value of less than 0.05, were considered significant (Table 2). We also tested for correlation of selected CNVRs and height with linear regression analysis after adjusting for age and gender using SAS program.

Table 2
Data of the selected CNV for association analysis with height in Korean population.

CHR	Locus	Start(bp)	End(bp)	Nearby gene	Type	Frequency		*P*-value	DGV
						Case	Control		
12	12q24.33	130363561	130381275	*GPR133*	Deletion	0.062	0.031	0.0043	Overlapped

CHR = chromosome; DGV = database of genomic variants.
P-value that is < 0.05 after Bonferroni correction is shown in boldface.

Table 3
Result of TaqMan copy number assays of the CNV at 12q24.33.

CHR	Start	End	Length	Assay ID	State	TP	FP	PPV
12	130363561	130381275	17.7 kb	Hs03807432_cn	Deletion	101	3	0.97

CHR = chromosome; TP = true positive; FP = false positive; PPV = positive predictive value; PPV = number of TP/(number of TP + number of FP).

4.8. Validation of CNVs by TaqMan copy number assays

To verify the identified CNVRs, we carried out quantitative PCR using the TaqMan Copy Number Assay (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's protocols. The samples used for validation were randomly selected (case 48/control 56). Copy number for each sample was calculated by Copy Caller v2.0 using the manufacturer's protocol. Table 3 shows the validation result of identified CNVRs. All experiments were replicated four times to enhance the validation accuracy.

Acknowledgments

This work was supported by an intramural grant from Korea National Institute of Health (2012-N73004-00) and grants from the Korea Centers for Disease Control and Prevention (4845-301, 4851-302, 4851-307).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2012.11.002.

References

[1] M. Perola, et al., Combined genome scans for body stature in 6,602 European twins: evidence for common Caucasian loci, PLoS Genet. 3 (2007) e97.
[2] H. Lango Allen, et al., Hundreds of variants clustered in genomic loci and biological pathways affect human height, Nature 467 (2010) 832–838.
[3] G. Lettre, Recent progress in the study of the genetics of height, Hum. Genet. 129 (2011) 465–472.
[4] S.A. McCarroll, Extending genome-wide association studies to copy-number variation, Hum. Mol. Genet. 17 (2008) R135–R142.
[5] N. Craddock, et al., Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls, Nature 464 (2010) 713–720.
[6] I. Jarick, et al., Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis, Hum. Mol. Genet. 20 (2011) 840–852.
[7] E.G. Bochukova, et al., Large, rare chromosomal deletions associated with severe early-onset obesity, Nature 463 (2010) 666–670.
[8] A. Dauber, et al., Genome-wide association of copy-number variation reveals an association between short stature and the presence of low-frequency genomic deletions, Am. J. Hum. Genet. 89 (2011) 751–759.
[9] X. Li, et al., A genome wide association study between copy number variation (CNV) and human height in Chinese population, J. Genet. Genomics 37 (2010) 779–785.
[10] A. Tonjes, et al., Genetic variation in GPR133 is associated with height: genome wide association study in the self-contained population of Sorbs, Hum. Mol. Genet. 18 (2009) 4662–4668.
[11] D.M. Altshuler, et al., Integrating common and rare genetic variation in diverse human populations, Nature 467 (2010) 52–58.
[12] N.E. Wineinger, H.K. Tiwari, The impact of errors in copy number variation detection algorithms on association results, PLoS One 7 (2012) e32396.
[13] N. Soranzo, et al., A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium, Nat. Genet. 41 (2009) 1182–1190.
[14] N. Kato, et al., Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians, Nat. Genet. 43 (2011) 531–538.
[15] Y.J. Kim, et al., Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits, Nat. Genet. 43 (2011) 990–995.
[16] J.C. Barrett, B. Fry, J. Maller, M.J. Daly, Haploview: analysis and visualization of LD and haplotype maps, Bioinformatics 21 (2005) 263–265.
[17] T.K. Bjarnadottir, et al., The human and mouse repertoire of the adhesion family of G-protein-coupled receptors, Genomics 84 (2004) 23–33.
[18] J. Bohnekamp, T. Schoneberg, Cell adhesion receptor GPR133 couples to Gs protein, J. Biol. Chem. 286 (2011) 41912–41916.
[19] R.T. Dorsam, J.S. Gutkind, G-protein-coupled receptors and cancer, Nat. Rev. Cancer 7 (2007) 79–94.
[20] D.F. Gudbjartsson, et al., Many sequence variants affecting diversity of adult human height, Nat. Genet. 40 (2008) 609–615.
[21] G. Lettre, et al., Identification of ten loci associated with height highlights new biological pathways in human growth, Nat. Genet. 40 (2008) 584–591.
[22] N. Soranzo, et al., Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size, PLoS Genet. 5 (2009) e1000445.
[23] W. Ise, et al., The transcription factor BATF controls the global regulators of class-switch recombination in both B cells and T cells, Nat. Immunol. 12 (2011) 536–543.
[24] B. Ge, O. Li, P. Wilder, A. Rizzino, T.W. McKeithan, NF-kappa B regulates BCL3 transcription in T lymphocytes through an intronic enhancer, J. Immunol. 171 (2003) 4210–4218.
[25] C. Alkan, B.P. Coe, E.E. Eichler, Genome structural variation discovery and genotyping, Nat. Rev. Genet. 12 (2011) 363–376.
[26] K. Wang, et al., PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data, Genome Res. 17 (2007) 1665–1674.
[27] J.C. Marioni, et al., Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization, Genome Biol. 8 (2007) R228.
[28] C. Barnes, et al., A robust statistical method for case–control association testing with copy number variation, Nat. Genet. 40 (2008) 1245–1252.
[29] W. Yu, et al., GWAS Integrator: a bioinformatics tool to explore human genetic associations reported in published genome-wide association studies, Eur. J. Hum. Genet. 19 (2011) 1095–1099.