Published in final edited form as:

Hum Genet. 2014 August ; 133(8): 967–974. doi:10.1007/s00439-014-1437-1.

Genome-wide association tests of inversions with application to psoriasis

Jianzhong Ma^{1,2}, Momiao Xiong³, Ming You⁴, Guillermina Lozano⁵, and Christopher I. Amos^{6,*}

¹Biostatistics/Epidemiology/Research Design Core, Center for Clinical and Translational Sciences, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

²Division of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

³Division of Biostatistics, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

⁴Department of Pharmacology and Toxicology, Medical College of Wisconsin, Milwaukee, WI 53226, USA

⁵Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁶Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon, NH 03766, USA

Abstract

Although inversions have occasionally been found to be associated with disease susceptibility through interrupting a gene or its regulatory region, or by increasing the risk for deleterious secondary rearrangements, no association study has been specifically conducted for risks associated with inversions, mainly because existing approaches to detecting and genotyping inversions do not readily scale to a large number of samples. Based on our recently proposed approach to identifying and genotyping inversions using principal components analysis (PCA), we herein develop a method of detecting association between inversions and disease in a genome-wide fashion. Our method uses genotype data for single nucleotide polymorphisms (SNPs), and is thus cost-efficient and computationally fast. For an inversion polymorphism, local PCA around the inversion region is performed to infer the inversion genotypes of all samples. For many inversions, we found that some of the SNPs inside an inversion region are fixed in the two lineages of different orientations and thus can serve as surrogate markers. Our method can be applied to case-control and quantitative trait association studies to identify inversions that may interrupt a gene or the connection between a gene and its regulatory agents. Our method also offers a new venue to identify inversions that are responsible for disease-causing secondary rearrangements. We illustrated our proposed approach to case-control data for psoriasis and identified novel associations with a few inversion polymorphisms.

*Corresponding to: Department of Community and Family Medicine Geisel School of Medicine, Dartmouth College 46 Centerra Parkway, Suite 330, Lebanon, NH 03766 Tel: 603-653-6696 Christopher.I.Amos@Dartmouth.edu.

Keywords

Chromosomal inversion; Principal components analysis; Genome-wide association scan; Single-Nucleotide Polymorphism; Psoriasis

Introduction

Inversion polymorphisms represent a significant class of structural variant (Feuk 2010; Alves et al., 2012). However, study of inversions in genetic epidemiology and statistical genetics is still in its infancy, especially compared to other structural variants, such as copy number variants (CNV) (Lakich et al. 1993). This is largely because of the lack of high-throughput methods for detecting inversions: without gain or loss of genetic materials, inversions cannot be readily detected with copy number arrays. Also, again because of their balanced nature, inversions have been believed to be mostly phenotypically neutral. So far, to our best knowledge, no genetic association study has been conducted either for targeted inversions or in a genome-wide fashion.

There has been growing evidence that inversions may be associated with a disease or a quantitative trait in different ways (Antonarakis et al. 1995; Bondeson et al. 1995; Small et al. 1997; Osborne et al. 2001; Knoll et al. 1998; Gimelli et al. 2003; Sharp et al. 2008). First, an inversion may disrupt a gene and thus mutate it. Second, the breakpoints of an inversion may fall between a gene and its regulatory region and alter its normal expression. Finally and more often, an inversion in the genome of a parent may induce secondary rearrangements in offspring that in turn may cause disease. Specifically, it has been suggested that the presence of an inverted segment in heterozygotes originates an unpaired region at pachtene making the region prone to misalignment (Feuk 2010; Osborn et al. 2001; Gimelli et al. 2003; Milina et al. 2012). This type of misalignment may cause non-allelic homologous recombination (NAHR), which generates *de novo* rearrangements, and segmental aneuploidies, such as insertions and deletions. Somatic, among individuals heterozygous for an inversion, recombination repair across the inversion could also lead to segmental aneuploidy.

Recent advances in mapping inversions with the advent of sequencing-based algorithms (Feuk 2010), such as the approach of paired-end mapping (Tuzun 2005; Kidd et al. 2008), have shown that inversions may be much more abundant than expected previously (Tuzun et al. 2005; Kidd et al. 2008; Ahn et al. 2009; Korbel et al. 2007; Chen et al. 2009), implying that there may be more inversions than previously expected that are associated with disease. In view that genetic variants identified through current genome-wide association studies (GWAS) using SNPs and CNV can only explain a small proportion of heritability (Manolio et al. 2009; Eichler et al. 2010), it is desirable to develop an approach to detecting inversion-disease association. In one of our recent studies, we demonstrated that targeted principal components analysis (PCA) can be used to identify and genotype inversions polymorphisms (Ma & Amos 2012a). Our approach utilizes merely unphased SNP genotype data and is cost-efficient and computationally fast. More importantly, our approach can be readily applied to very large sample sets and is thus more suitable for powerful association tests,

compared to other statistical and/or computational approaches (Bansal et al. 2007; Sindi et al. 2010; Cáceres et al. 2012; Salm et al. 2012; Bosch et al. 2009; Deng et al. 2008; Tuner et al. 2006) that are based on the linkage disequilibrium induced by an inversion. In the present paper, we propose a PCA-based approach of genome-wide association test of inversions and illustrate its application using two case-control data sets for psoriasis. Specifically, we describe our approach for determining surrogate SNPs of inversions based on the results of PCA. The identified surrogate SNPs can be used in lieu of the corresponding inversions for data sets in which there are no sufficient number of SNPs in an inverted region for meaningful PCA, making our approach robust to various existing GWAS data sets for a quick and efficient scan of inversions that may contribute to complex diseases or quantitative traits.

Methods and Materials

Detecting and genotyping inversions using PCA

We first briefly review the PCA-based approach to detecting and genotyping inversions. In a previous publication (Ma & Amos 2010), we developed a theoretical formulation of PCA, in which individuals are treated as features and markers (SNPs) as “realizations” of a random vector. In the space spanned by the first few eigenvectors, individuals sampled from different populations are distributed into different clusters. This formulation of PCA was later extended to the case of admixture (Ma & Amos 2012b). We demonstrated that admixed individuals are distributed along the line segment connecting the centroids of the clusters formed by individuals from the two parental populations according to the admixture proportions. Recently, we found that application of this approach identifies admixture in population structure that results from chromosomal inversions (Ma & Amos 2012a). Because recombination is usually suppressed between the inverted and non-inverted segments, these two segments evolve independently and create an effect similar to a simple population substructure: two distinct populations of inversion homozygotes of different orientations and their 1:1 admixture, namely the inversion heterozygotes. Locally performing PCA around this inverted region results in a pattern consisting of three equidistant strips, with the middle stripe representing the inversion heterozygous individuals. This pattern can be used to genotype an inversion polymorphism. Identification of this equidistant three-stripe pattern in the space spanned by the first two eigenvectors is a major criterion for discovering a novel inversion.

Identification of surrogate SNP for inversions

We have developed a PCA-based scanner for inversions and conducted a genome-wide scan using data for the two Caucasian populations, CEU and TSI, from Phase III of the International HapMap project (Gibbs et al. 2003), which resulted in 2,040 inversions, including 159 of those previously discovered using cytogenetic approaches and/or sequencing-based approaches. For existing GWAS data sets, however, many of those predicted inversions may not be genotyped by performing local PCA because the subjects are usually not genotyped as dense as for the HapMap subjects. Fortunately, we found that for some of those predicted or known inversions, many SNPs inside or around the inversion regions are fixed in the two lineages corresponding to the two orientations and thus can

serve as surrogate markers to the inversion polymorphism. For example, in the well-known inversion at 17q21.31, allele A of SNP rs241031 is homozygous for all HapMap samples in the inversion homozygote group, but only appears in 0.006% of those in the non-inverted homozygote group, and has an allele frequency 50% in the inversion heterozygote group (Figure 1). For an inversion polymorphism, we identified its surrogate SNPs using the HapMap data by analyzing the allele frequencies of SNPs as follows. First, local PCA was performed around the inverted region and individuals were genotyped for this inversion. Then, allele frequencies of all SNPs inside the inversion region were calculated for each of the three inversion-genotype groups. A SNP was defined to be fixed to the inversion if its allele frequency in the inversion heterozygous group, p_2 , was close to 0.5, $|p_2 - 0.5| < 0.01$, its allele frequency in the two inversion homozygous groups, p_1 and p_3 , are close to 0 and 1, $p_1 < 0.01$ and $p_3 > 0.99$, respectively. These cutoffs were chosen as a trade-off between the number of resultant surrogate SNPs and the degree of their correlation with the inversion, given the sample size of the HapMap populations. In total, we identified surrogate SNPs for 340 inversions including 37 known inversions (Online Resource 1: Table S1).

Construction of a preliminary inversion map

A map of inversion (Online Resource 1: Table S1) was constructed by integrating known and/or validated inversions from the literature, including the Database of Genomic Variants (DGV) (<http://projects.tcag.ca/variation/>), and those predicted using our PCA-based scanner. For the known and/or validated inversions, we found that not all of them can be genotyped using our PCA-based approach. One of the reasons could be that some of the inversion regions are too short to include a sufficient number of SNPs for meaningful PCA. Also, it is possible that some of the known inversions are not polymorphic or have too low frequencies for the rare orientation of inversion in the HapMap populations. We therefore included in the inversion map only those known inversions that can be genotyped using the PCA approach (i.e. showing an equidistant, three-stripe pattern). For each of the inversions in the map, information on the corresponding surrogate SNPs was also annotated. Currently, the inversion map includes 2,050 inversions including 179 that are known and/or validated.

Inversion-disease association test

With the proposed approach to genotyping inversions and the inversion map, performing an inversion-disease association test is straightforward for genetic data of independent individuals, such as the psoriasis data sets. The phenotype can be either dichotomous (case-control design) or quantitative (linear regression). The purpose of this type of test is to detect inversions that may directly disrupt a gene or alter a gene and its regulatory region. For each of the inversions in our inversion map, if there are a sufficient number of SNPs inside the inversion region genotyped, local PCA will be performed to infer the inversion genotypes of all individuals. The number of SNPs is considered sufficient if a clear three-stripe pattern, as defined in Ma and Amos (2012a) based on the K-means algorithm, can be observed for the data in an inversion region. Only when there are too few SNPs genotyped inside an inversion, should we use the surrogate SNPs for the inversion-disease association test. Because a surrogate SNP usually is not fully associated with an inversion, whenever possible, we should use PCA to infer the inversion genotypes of the individuals in order to improve statistical power. If there are two or more surrogate SNPs for an inversion, the one

with the strongest correlation with the inversion should be used for the inversion-disease association test.

For a genome-wide inversion association study, an important issue for inversion association tests is how to control for false positive rate. We will leave the issue of correcting for multiple testing for future studies, because the current preliminary inversion map is far from complete, including only 2,050 inversions, and the correlation among inversions is to be explored. In fact, the number of inversions that can be tested using existing GWAS data may be much smaller than the total number of inversions in the inversion map due to the lack of sufficient number of SNPs. For example, in the application to the psoriasis data, only 235 and 231 inversions can be tested. Therefore, we will simply report the original p-values and indicate what the Bonferroni correction will result in.

Results

Psoriasis is a chronic, inflammatory skin disease affecting 2-3% of the world population. To illustrate the proposed approach of association test for inversions, we analyzed data from the Psoriasis GAIN study downloaded from the dbGaP. The data we analyzed included (1) SNP data from an original GWAS with 955 patients and 693 controls; and (2) SNP data from a replication study with 466 patients and 732 controls. After the quality control process, we had 915 cases and 675 controls with genotype data for 443,018 SNPs for the original GWAS, and 431 cases and 702 controls with genotype data for 439,201 SNPs for the replication study.

Logistic regression analysis was performed for each of the inversions in the inversion map using the following five genetic models: full, additive, dominant, recessive, and over Dominant, with sex and age as covariates. We considered the over-dominant model because the presence of an inverted and a non-inverted segment may cause a secondary, deleterious rearrangement (Feuk 2010; Osborne et al. 2001; Gimelli et al. 2003; Molina et al. 2012). Compared to the full model, an over-dominant model should have a greater power because the inversion heterozygous are compared to a pool of the two homozygous. Due to lack of surrogate SNPs, some of the inversions could not be tested. Among the inversions on the map, 235 and 231 were tested either directly using the PCA genotyping method or through surrogate SNPs for the two data sets, respectively. In the original data set, association ($P < 0.05$) for at least one of the five genetic models was observed at 37 inversions, as shown in Table S2 (Online Resource 2). Results for the replicate data set are shown in Table S3 (Online Resource 3), where significant association was observed for 33 inversions. There were eight inversions were associated in the test data shown in Table S2 and were replicated in Table S3. However, only three (3) of these are likely to be true positive findings because the rest of the findings were associated with risk in opposite direction of association. The most significantly associated inversion is the region of chr6: 31363479-31385967 with $P < 1.1 \times 10^{-9}$ and $P < 1.1 \times 10^{-13}$ for the original and the replicate data set (Table 2). This predicated inversion overlaps with the HLA region and has almost perfect tag SNPs listed in Table 1. Among these tag SNPs, rs2243868 was found to be tag SNP of the HLA-Cw*0602 allele, which has been described as the primary risk allele for psoriasis. Therefore, it is possible that the observed association here may be due to effects from the HLA-C gene, instead of

the predicted inversion. This region of chromosome 6p is extremely heterogeneous and previous sequencing studies have identified inversions in the region (Nair et al. 2006), but there are also highly conserved haplotypes associated with psoriasis. Our results indicate that a dominant model best explains the association with the predicted inversion with disease risk: the odds ratio between patients with and without the inverted segments was estimated to be 1.99 (1.59~2.48) and 2.41 (1.81~3.24), respectively, for the two data sets. In Table S4, the genotype frequencies of this predicted inversion were given for different HapMap populations.

The next most significant association ($P < 0.00012$), shown in Table 2, was seen for a predicted inversion, Chr5:131893690-132042870 and was confirmed in the replicated data, but with less significance ($P = 0.02$). If the less prevalent orientation is taken as the inverted segment, as in our previous publication, the inversion frequency in the CEU population is estimated to be 0.21, close to 0.18 in both the original and replicate psoriasis data sets (Online Resource 4: Table S5). Then, our analysis showed that the non-inverted allele is deleterious with an odds ratio 1.51 (1.22~1.86) and 1.35 (1.04~1.75) in the two data sets for a recessive model. Similarly, a much less significant association was observed for another predicted inversion at Chr5:161842217-161963696. In addition, marginally significant association ($P \sim 0.05$) for both data sets was seen for a predicted inversion at Chr11:23451453-23460539. Other predicted inversions showed significant associations with the risk of psoriasis for both data sets, but with opposite directions, such as Chr17:3854057-4232178.

Six (6) known inversions listed in DGV were found to be significant in either the original data set (Table S2) or the replicate data set (Table S3). A 1.2-Mpb inversion at Chr15:72151413-73356183 was detected by Kidd et al (2008) from clone-based end-paired sequence data and validated using fluorescent in situ hybridization assays. This inversion was later confirmed by Ahn et al. (2009) using the paired-end sequence and mapping approach in a Korean genome. Using our PCA-based approach, we detected this inversion and estimated the inversion frequencies for the non-African HapMap populations (Online Resource 4: Table S6). If we define the orientation that is rare in the two Caucasian populations (9% for CEU and 12% for TSI), the inversion is highly prevalent in the Asia population (>44%). This inversion was found to be associated with the risk of psoriasis with the non-inverted orientation being deleterious in a dominant model with an odds ratio 1.47 (1.18~1.82) and $P = 0.0006$ or an additive model with an odds ratio 1.32 (1.09~1.60) and $P = 0.004$, respectively (Table 2). This association was, however, not validated in the replicate data set ($P > 0.1$). It should be noted that this inversion corresponds to the site of the 15q24 recurrent microdeletion.

A second known inversion that was found to be significantly associated with the risk of psoriasis is located at Chr14:75534288:75547088 (Table 2), which was identified by Kidd et al (2008). Our PCA-based approach detected this inversion and found that it is more prevalent (~33%) in the Caucasian populations including CEU, TSI and GIH than in the East Asia populations (<24%), as shown in (Online Resource 4: Table S7). Significant association (0.001) was seen in the original psoriasis data set with the inverted segment being deleterious with odds ratio 1.27 (1.10~1.48) in an additive model.

In addition, significant association was also observed for a known inversion at Chr2:234100014-234189000 ($P=0.006$), shown in Table 2. This inversion was identified and validated by several studies including Kobel et al. (2007), Kidd et al. (2008), Levy et al. (2007) and Pang et al. (2010). We detected this inversion and estimated the frequencies for the non-African HapMap populations (Online Resource 4: Table S8). This inversion has relatively higher frequencies in the Caucasian populations (23% and 27% for CEU and TSI, respectively) than in other populations. The inverted orientation was found to be deleterious in an additive model with an odds ratio 1.27(1.07~1.52) in the original data set. This association was not replicated in the second data set.

Finally, for the original discovery data, the Bonferroni corrected alpha-level is $0.05/235=0.000213$. Therefore, if Bonferroni correction is applied, only the first two inversions listed in Table 2 are still significant. It should be noted that they are the only two that were replicated using the replicated data set.

Discussion

We have proposed for the first time an approach to conducting inversion-disease association test using widely available SNP genotype data. Our approach is based on PCA and utilizes only unphased genotype and is thus extremely cost-efficient and computationally fast. More importantly, our approach can readily scale to large number of samples, making it possible to perform powerful association test for this unique type of structural variant. Indeed, most of the inversions given in our inversion map are just predicted from our genome-wide scan of inversion using our PCA-based approach and demand computational or experimental validation for their existences. Nevertheless, we wish to emphasize here the following two points. First, although there may be significant false positive rate in identifying novel inversions using our PCA-based approach for a genome wide scan of inversions, the PCA approach works very well when applied to genotype known inversion polymorphisms, as demonstrated in Ma and Amos (2012a) using simulated data and for the two well-known inversions at 8p23.1 and 17q21.31, for which experimental results are available for many of the HapMap individuals. Therefore, the approach to conducting inversion association test proposed here should be trusted when applied to known inversions discovered using experimental or other computational approaches using sequence data. Second, even if a predicted inversion would turn out to be some sort of variant other than an inversion, the association test proposed here may still shed a light on the genetic architecture of the disease of interest.

The approach we proposed here works not only to very dense SNP data, but also to genotype data after quality control as in a typical GWAS and also to single nucleotide variants called from sequencing data. The surrogate SNPs identified to many of the inversions enable us to assess the inversion characteristics and its association with a complex trait instantly, once the genotype of a surrogate SNP is known from either SNP array data, sequence data, or experimental results. We will continue to identify surrogate SNPs for more inversions and also for different populations.

For the real data analyses, we have identified several candidate known and predicted inversions that may be associated with the susceptibility of psoriasis. An interesting question is how one should interpret results of our inversion association test comparing with those of the SNP-based association test. First, if the number of SNPs inside an inversion is not large enough for performing local PCA, the SNP with the strongest correlation with the inversion should be used for the inversion association test, and thus the resultant test statistic and the p-value will be the same as those obtained in the usual SNP-based association test. However, compared to the SNPs, the number of inversions to be tested to date is small and thus requires less conservative correction for multiple testing. Therefore, the power of detecting association for an inversion should be greater than using a SNP-based GWAS. Second, if there is sufficient number of SNPs inside an inversion, the inferred inversion status should be used for association test of the inversion. In this case, the resultant test statistic and p-value should be more reliable than those obtained from the SNP-based association test, unless a SNP is perfectly correlated with the inversion.

In this work, we only applied the proposed approach to the association test of case-control data. This type of association study applies to situations in which inversions directly disrupt a gene or the connection between a gene and its regulatory region. It is, however, more likely that an inversion of a parent's DNA increases the risk of a secondary rearrangement that affects the offspring. Our approach can also be used to perform association test between the inversion status in parents and the occurrence of disease or genomic disorder in offspring. For example, using SNP data for case-parent trios, we may conduct statistical analysis to test whether the rate of inversions in the patients is significantly higher than that in the general population for an inversion. Of specific interest is the rate of inversion heterozygote in the parents, because the presence of an inverted and a non-inverted segment originates an unpaired region at the pachytene stage, which may result in a misalignment and non-allelic homologous recombination. In addition, we may also perform an analysis to investigate the influence of inversions on gene expression. Information about the identified inversions associated with a gene expression, an e-inversion, could be incorporated with that from the direct inversion association test in order to gain a more comprehensive understanding about how the inversions interrupt a gene and/or its regulatory agents and thus increase the risk of disease.

Given that all other types of structural variants, including single nucleotide variants, copy number variants, translocations, have been investigated in their susceptibility to various disease, we believe that the proposed approach here will prove to be a useful tool for the inversion polymorphisms as the last class of structural variants. Since our approach works well with the usual SNP genotype data, which are abundant thanks to the success of the GWAS, it is expected to provide further insight into the etiology of many diseases and may explain some of the missing heritability.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work has been supported by NIH grant R01CA134682. JM also acknowledge the support provided by the Biostatistics/ Epidemiology/ Research Design (BERD) component of the Center for Clinical and Translational Sciences (CCTS) for this project. CCTS is mainly funded by the NIH Centers for Translational Science Award (NIH CTSA) grant (UL1 RR024148), awarded to University of Texas Health Science Center at Houston in 2006 by the National Center for Research Resources (NCRR) and its renewal (UL1 TR000371) by the National Center for Advancing Translational Sciences (NCATS).

Reference

- 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; 19:1622–1629. [PubMed: 19470904]
- Antonarakis SE, Rossiter JP, Young M, Horst J, de Moerloose P, Sommer SS, Ketterling RP, Kazazian HH Jr, Negrier C, Vinciguerra C, Gitschier J, Goossens M, Girodon E, Ghanem N, Plassa F, Lavergne JM, Vidaud M, Costa JM, Laurian Y, Lin SW, Lin SR, Shen MC, Lillicrap D, Taylor SA, Windsor S, Valleix SV, Nafa K, Sultan Y, Delpech M, Vnencak-Jones CL, et al. Factor VIII gene inversions in severe hemophilia A: results of an international consortium study. *Blood.* 1995; 86:2206–2212. [PubMed: 7662970]
- Alves JM, Lopes AM, Chikhi L, Amorim A. On the structural plasticity of the human genome, chromosomal inversions revisited. *Current Genome.* 2012; 13:623–632.
- Bansal V, Bashir A, Bafna V. Evidence for large inversion polymorphisms in the human genome from HapMap data. *Genome Res.* 2007; 17:219–230. [PubMed: 17185644]
- Bondeson ML, Dahl N, Malmgren H, Kleijer WJ, Tonnesen T, Carlberg BM, Pettersson U. Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the Hunter syndrome. *Hum Mol Genet.* 1995; 4:615–621. [PubMed: 7633410]
- Bosch N, Morell M, Ponsa I, Mercader JM, Armengol L, et al. Nucleotide, cytogenetic and expression impact of the human chromosome 8p23.1 inversion polymorphism. *PLoS ONE.* 2009; 4:e8269. doi: 10.1371/journal.pone.0008269. [PubMed: 20011547]
- Cáres A, Sindi S, Raphael B, Cáres M, Gonzáz J. Identification of polymorphic inversions from genotypes. *BMC Bioinformatics.* 2012; 13:28. [PubMed: 22321652]
- Chen K, Wallis JW, McLellan MD, David E, Larson DE, Kalick DJ, Pohl CS, McGrath SD, Wendt 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 Meth.* 2009; 6:667–781.
- Deng L, Zhang Y, Kang J, Liu T, Zhao H, et al. An unusual haplotype structure on human chromosome 8p23 derived from the inversion polymorphism. *Hum Mutat.* 2008; 29:1209–1216. [PubMed: 18473345]
- Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moor JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet.* 2010; 11:446–450. [PubMed: 20479774]
- Feuk L. Inversion variants in the human genome: role in disease and genome architecture. *Genome Med.* 2010; 2:11–19. [PubMed: 20156332]
- Gibbs R, Belmont J, Hardenbol P, Willis T, Yu F, et al. The international HapMap project. *Nature.* 2003; 426:789–796. [PubMed: 14685227]
- Gimelli G, Pujana MA, Patricelli MG, Russo S, Giardino D, Larizza L, Cheung J, Armengol L, Schinzel A, Estivill X, Zuffardi O. Genomic inversions of human chromosome 15q11–q13 in mothers of Angelman syndrome patients with class II (BP2/3) deletions. *Hum Mol Genet.* 2003; 12:849–858. [PubMed: 12668608]
- Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, Haugen E, Zerr T, Yamada NA, Tsang P, Newman TL, Tuzun E, Cheng Z, Ebling HM, Tusneem N, David R, Gillett W, Phelps KA, Weaver M, Saranga D, Brand A, Tao W,

- Gustafson E, McKernan K, Chen L, Malig M, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature*. 2008; 453:56–64. [PubMed: 18451855]
- Knoll JH, Nicholls RD, Magenis RE, Graham JM Jr, Lalande M, Latt SA. Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion. *Am J Med Genet*. 1998; 32:285–290. [PubMed: 2564739]
- Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M. Paired-end mapping reveals extensive structural variation in the human genome. *Science*. 2007; 318:420–426. [PubMed: 17901297]
- Lakich D, Kazazian HH Jr, Antonarakis SE, Gitschier J. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nat Genet*. 1993; 5:236–241. [PubMed: 8275087]
- Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, Macdonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Kravitz SA, Busam DA, Abril Ferrando, Josep Francesc, The Diploid genome sequence of an individual human. *PLoS Biol*. 2007 doi:10.1371.
- Ma J, Amos CI. Theoretical formulation of principal components analysis to detect and correct for population stratification. *PLoS ONE*. 2010; 5:e12510. [PubMed: 20862251]
- Ma J, Amos CI. Investigation of inversion polymorphisms in the human genome using principal components analysis. *PLoS ONE*. 2012a; 7:e40224. [PubMed: 22808122]
- Ma J, Amos CI. Principal Components Analysis of Population Admixture. *PLoS One*. 2012b; 7:e40115. [PubMed: 22808102]
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindoff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009; 461:747–753. [PubMed: 19812666]
- Molina O, Anton E, Vidal F, Blanco J. High rates of de novo 15q11q13 inversions in human spermatozoa. *Mol Cytogenetics*. 2012 doi: 10.1186/1755-8166-5-11.
- Osborne LR, Li M, Pober B, Chitayat D, Bodurtha J, Mandel A, Costa T, Grebe T, Cox S, Tsui LC, Scherer SW. A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. *Nat Genet*. 2001; 29:321–325. [PubMed: 11685205]
- Pang AW, MacDonald JR, Pinto D, Wei J, Rafiq MA, Conrad DF, Park H, Hurles ME, Lee C, Venter JC, Kirkness EF, Levy S, Feuk L, Scherer SW. Towards a comprehensive structural variation map of an individual human genome. *Genome Biol*. 2010 doi: 10.1186/gb-2010-11-5-r52.
- Salm A, Horswell S, Hutchison C, Speedy H, Yang X, et al. The origin, global distribution, and functional impact of the human 8p23 inversion polymorphism. *Genome Res*. 2012; 22:1144–1153. [PubMed: 22399572]
- Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE, Schroer RJ, Novara F, De Gregori M, Ciccone R, Broome A, Casuga I, Wang Y, Xiao C, Barbacioru C, Gimelli G, Bernardina BD, Torniero C, Giorda R, Regan R, Murday V, Mansour S, Fichera M, Castiglia L, Failla P, Ventura M, Jiang Z, Cooper GM, Knight SJ, Romano C, et al. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet*. 2008; 40:322–328. [PubMed: 18278044]
- Sindi S, Raphael B. Identification and frequency estimation of inversion polymorphisms from haplotype data. *J Comp Biol*. 2010; 17:517–531.
- Small K, Iber J, Warren ST. Emerin deletion reveals a common X chromosome inversion mediated by inverted repeats. *Nat Genet*. 1997; 16:96–99. [PubMed: 9140403]
- Turner DJ, Shendure J, Porreca G, Church G, Green P, Tyler-Smith C, Hurles ME. Assaying chromosomal inversions by single-molecule haplotyping. *Nat Meth*. 2006; 3:439–445.
- Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE. Fine-scale structural variation of the human genome. *Nat Genet*. 2005; 37:727–732. [PubMed: 15895083]

Fig.1.

Identification of surrogate SNPs for the inversion at 17q21.31 from the SNP allele frequencies in each inversion genotype group

Table 1

Surrogate SNPs of a predicted inversion at chr6: 31363479-31385967 and their allele frequencies in the three inversion-genotype groups. The inversion genotypes of all individuals were inferred using principal components analysis.

SNP	Location	Inv-hom	Inv-heter	NonInv-hom
rs2844603	31358833	0.992806	0.498113	0
rs2853935	31361857	0.994604	0.498113	0
rs2853933	31362067	0.005396	0.501887	1
rs2524040	31365604	0.005396	0.501887	1
rs2524163	31367558	0.994604	0.498113	0
rs2524156	31368376	0.994604	0.498113	0
rs2243868	31369255	0.992806	0.498113	0
rs2524095	31374096	0.994604	0.5	0
rs2853922	31374169	0.998201	0.5	0
rs2524089	31374501	0.998201	0.5	0
rs2524066	31377133	0.001799	0.503774	1

Table 2

Top inversions associated with psoriasis and the corresponding ODDS ratios (95% CI) and p-values
 Replicated: Significant association was confirmed using the replicated data set with the same effect direction as in the original data set. Validated: “PCA predicted” refers to inversions to be validated; “DGV” refers to known inversion listed in the Database of Genomic Variants

Inversion	ODDS Ratio (95% CI)	P	Model	Replicated	Validated
chr6: 31363479- 31385967	1.99 (1.59~2.48)	1.10E- 09	Recessive	yes	PCA predicted
Chr5:131893690- 132042870	1.51 (1.22~1.86)	0.00012	Recessive	yes	PCA- predicted
Chr15: 72151413- 73356183	1.47 (1.18~1.82)	0.0006	Dominant	no	DGV
Chr14:75534288- 75547088	1.27 (1.10~1.48)	0.001	Additive	no	DGV
Chr2:234100014- 234189000	1.27(1.07~1.52)	0.006	Additive	no	DGV