

PROCEEDINGS

Open Access



Identity-by-descent mapping for diastolic blood pressure in unrelated Mexican Americans

Xiao-Qing Liu^{1,2,3*}, Jillian Fazio¹, Pingzhao Hu^{2,4} and Andrew D. Paterson^{5,6}

From Genetic Analysis Workshop 19
Vienna, Austria. 24-26 August 2014

Abstract

Population-based identity by descent (IBD) mapping is a statistical method for detection of genetic loci that share an ancestral segment among “unrelated” pairs of individuals for a disease. As a complementary method to genome-wide association studies, IBD mapping is robust to allelic heterogeneity and may identify rare inherited variants when combined with sequence data.

Our objective is to identify the causal genes for diastolic blood pressure (DBP). We applied a population-based IBD mapping method to 105 unrelated individuals selected from the family data provided for the Genetic Analysis Workshop 19. Using the genome-wide association study data (ie, the microarray data), chromosome 3 was scanned for IBD sharing segments among all pairs of these individuals. At the chromosomal region with the most significant relationship between IBD sharing and DBP, the whole genome sequence data were examined to identify the risk variants for DBP.

The most significant chromosomal region that was identified to have a relationship between the IBD sharing and DBP was at 3q12.3 ($p = 0.0016$), although it did not achieve the chromosome-wide significance level ($p = 0.00012$). This chromosomal region contains 1 gene, *ZPLD1*, which has been reported to be associated with cerebral cavernous malformations, a disease with enlarged small blood vessels (capillaries) in the brain. Although 24 deleterious variants were identified at this region, no significant association was found between these variants and DBP ($p = 0.40$).

We presented a mapping strategy which combined a population-based IBD mapping method with sequence data analyses. One gene was located at a chromosomal region identified by this method for DBP. However, further study with a large sample size is needed to assess this result.

Background

As more sequence data become available, one of the challenges is how to identify the disease-causing variants among hundreds of deleterious variants an individual carries [1]. Various research strategies, such as the combined homozygosity mapping and sequence data analysis approach, have been successfully applied [2].

However, other methods are needed for complex diseases and traits that are not autosomal recessive.

Traditional identity-by-descent (IBD) mapping methods have been successfully applied in studies of Mendelian and complex diseases using related cases [3]. With the availability of high-density genetic markers, such as those from genome-wide association and next-generation sequencing studies, it is possible to estimate IBD sharing accurately between 2 randomly chosen individuals in an outbred population for relatively short chromosomal regions, for example, 0.03 to 1 cM instead of 10 to 20 cM from family-based linkage studies [4–7]. In addition, compared to

* Correspondence: liu@umanitoba.ca

¹Department of Obstetrics, Gynecology, and Reproductive Sciences, University of Manitoba, Winnipeg, MB R3E 3P4, Canada

²Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB R3E 3P4, Canada

Full list of author information is available at the end of the article



single-marker association studies, IBD mapping methods are more robust to allelic heterogeneity, which has been observed in complex diseases and traits [8].

To date, a few groups have applied IBD mapping methods as a complementary method to genome-wide association studies for complex diseases in large, unrelated samples [5, 9–11]. For each of these studies, the whole genome was scanned in order to compare IBD sharing of each segment in apparently unrelated case-case pairs to that in case-control and control-control pairs. If there were rare causative variants at a locus, the case-case pairs would be expected to share significantly more at this locus than both the case-control and control-control pairs. Several of the studies found genome-wide significant excess IBD sharing for their disease of interest. In contrast, they did not find any chromosomal regions that reached the level of genome-wide significance with genome-wide association methods using the same data.

The population-based IBD mapping methods may also be applied to quantitative traits [12]. In this study, we test the hypothesis that diastolic blood pressure (DBP) levels are related to rare inherited genetic variants that can be identified using the combined IBD mapping and sequence data analysis approach.

Methods

Data description

The genome-wide association study (GWAS) data from the families, which had 959 individuals and 472,049 genetic markers, were used for quality control procedures, including relationship and population structure analyses. Unrelated individuals and markers from chromosome 3 of the GWAS data were used for IBD mapping. Chromosome 3 was chosen according to the suggestions from the Genetic Analysis Workshop 19 (GAW19) data contributors.

The DBP from the first time point was used as the outcome. For the individuals with medical treatment for hypertension, 5 mmHg was added to the original DBP values [13]. Multiple linear regression was performed with gender, age, smoking status, and ethnicity (the significant principal components retrieved using the Eigensoft program; see “Quality Control” below) as covariates and the DBP residuals were derived and used to test the relationship between IBD sharing and DBP.

Quality control

For the GWAS data, individuals were excluded if they had a genotype missing rate of 5 % or greater or a Mendelian error rate of 1 % or greater, and markers were excluded if they were insertion/deletion, had no unique physical location, a missing rate of 5 % or greater, a Mendelian error rate of 1 % or greater, a minor allele frequency (MAF) of 1 % or less, a MAF of 5 % or less with a missing rate of 1 % or greater, or a Hardy-Weinberg equilibrium

(HWE) p value of less than 0.0001. The markers were further pruned (with $r^2 < 0.33$) for relationship and population structure analyses. These quality control procedures were performed using PLINK v1.07 [4].

The PLINK “-genome” results were used to select unrelated individuals. Because Mexican Americans have ancestors with different ethnic backgrounds and because the genetic variants related to DBP may have different allele frequencies in different ethnic groups, the computer program Eigensoft (v4.2) [14] was applied to estimate population structure for the GWAS individuals. The Tracy-Widom statistic was used to evaluate the statistical significance of each principal component. The eigenvectors corresponding to the statistically significant eigenvalues were retrieved and used as covariates in the multiple regression model mentioned under “Data Description” above.

Identity-by-descent mapping

Beagle Refined IBD (v4.0.r1128) was used to estimate IBD sharing at each location on chromosome 3 [6]. The parameter, *ibdlength* (the minimum IBD length to be detected), was set to 1 cM. Other parameters, *window* (the sliding window that determines the memory usage), *overlap* (overlap between sliding windows), *ibdtrim* (end of a candidate IBD segment trimmed before likelihood calculation), and *ibdwindow* (sliding window for IBD sharing detection), were set to be the number of markers in 12, 1.5, 0.15, and 0.2 cM, respectively. Default settings were used for the remaining parameters. IBD sharing estimation was run 3 times using different seeds. Results from these 3 runs were combined using the *ibdmerge.jar* utility program (see “Website resources” below).

The physical map was from build GRCh37 of the Genome Reference Consortium 2009. The genetic map was generated by Dr. Brian Browning based on the International Haplotype Map Project (HapMap) Phase II genetic map (see “Website resources” below). If a marker was not in this map, its genetic location was interpolated based on its physical location.

Test for the relationship between identity-by-descent sharing and diastolic blood pressure

A Perl script was written to create a matrix with the number of unrelated individual pairs \times the total number of markers on chromosome 3. The observations in the matrix only had 2 values: 0 if no IBD sharing was observed for pair i at marker j , and 1 if IBD sharing was observed for pair i at marker j .

Following the IBD mapping methods proposed for quantitative traits using sibpairs [15–17], we calculated both the squared trait difference (D) and squared trait sum (S) for the DBP residuals for each unrelated pair. Then, for each marker in the matrix, 2 simple linear regressions were performed:

Regression 1: Y^D on the IBD sharing status (π) with estimated slope $\hat{\beta}_D$ and variance σ_D^2

Regression 2: Y^S on π with estimated slope $\hat{\beta}_S$ and variance σ_S^2

The weighted overall slope estimate is:

$$\hat{\beta} = \left(\frac{\sigma_D^2}{\sigma_S^2 + \sigma_D^2} \right) \hat{\beta}_S + \left(\frac{\sigma_S^2}{\sigma_S^2 + \sigma_D^2} \right) \hat{\beta}_D$$

The estimate of the standard error of $\hat{\beta}$ is:

$$SE(\hat{\beta}) = \sqrt{\frac{1}{\left[\frac{1}{\sigma_D^2} + \frac{1}{\sigma_S^2} \right] \left(\sum_{i=1}^n \hat{\pi}_i \right)}}$$

where $\hat{\pi}_i$ is the observed IBD sharing (0 for no sharing or 1 for sharing) for pair i .

Linkage was tested using a one-sided t test of the slope estimate. Under the null hypothesis of no linkage, the slope was zero whereas under the alternative hypothesis, the locus was linked to the trait and the slope was negative. The t statistic was calculated as the weighted overall slope estimate divided by its standard error:

$$t = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

Because the pairs of individuals were not independent from each other, the significance threshold for the real data was obtained using a permutation procedure.

Whole genome sequence data

For the chromosomal region with the most significant relationship between IBD sharing and DBP, the sequence data were examined to identify the risk variants for DBP. First, variants with no variation (MAF = 0), more than 2 alleles, or a missing rate of more than 15 % were excluded. The remaining genetic variants were annotated with the Combined Annotation-Dependent Depletion (CADD) scaled scores [18]. As recommended by the authors, a genetic variant with a scaled score higher than 20 was regarded as deleterious. The relationships between the deleterious variants and DBP were analyzed using the optimal sequence kernel association test (SKAT-O [19]) with gender, age, smoking status, and the first 3 principal components as covariates.

Results

After the quality control procedures, 914 individuals and 374,179 markers were selected for relationship and population structure analyses; and 105 unrelated Mexican Americans and 52,216 markers from chromosome 3 of the GWAS data were selected for IBD mapping (Table 1). Three principal components were determined to be

Table 1 Characteristics of the 105 unrelated individuals

Variable	Category	Frequency (%) or Mean (SD, range)
Gender	Male	43 (41 %)
Treatment	Yes	22 (21 %)
Smoking status	Yes	25 (24 %)
Age (years)		56.6 (15.5, 20.3–91.3)
Original DBP (mm Hg)		71.8 (9.3, 51–101)
Adjusted DBP (mm Hg)		72.9 (10.3, 51–101)

significant for population structure and were used as covariates in the multiple linear regression model for DBP. None of the covariates was significantly associated with DBP (ie, with $p < 0.05$); however, the first principal component had a p value of 0.1 (ie, the DBP increased 18.2 mmHg for each unit increase in the first principal component).

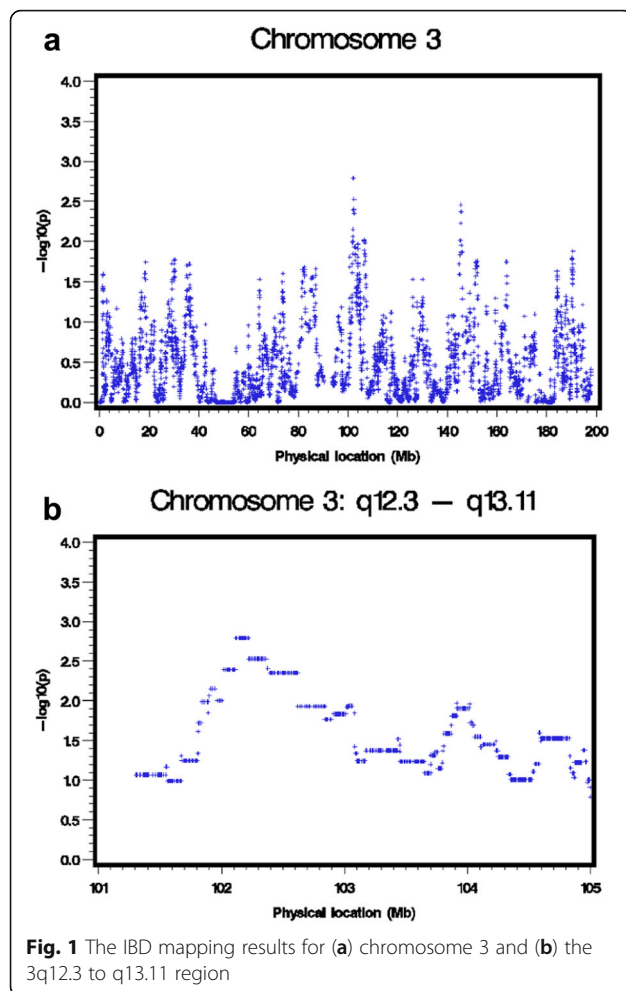
For the 105 unrelated individuals, there were 5460 pairs; 1573 (28.8 %) of them had detectable chromosomal segments shared by IBD. The average length of the IBD sharing segments was 1.47 cM with a range of 1.00 to 6.28 cM. The average chromosome-wide IBD sharing rate was 0.0029 with a standard deviation of 0.0025 and a range of 0 to 0.019. For each region, the rate was calculated with the number of pairs with IBD sharing at this region divided by the total number of pairs (ie, 5460).

For the IBD mapping of DBP, the most significant chromosomal region was at 3q12.3 with the peak single-nucleotide polymorphisms (SNPs) located at 101,901,465 to 102,620,049 bp, a region bounded by the GWAS markers that had IBD mapping p values of less than 0.01 (Fig. 1). The most significant p value for these markers was 0.0016, which did not reach the estimated chromosome-wide significance level 0.00012. For the most significant region, the IBD sharing rate was 0.0057, which was among the top 8 % of the rates for all the regions on chromosome 3.

Of the 105 unrelated individuals, 71 had the whole-genome sequence data. The most significant region overlapped with 1 gene, the zona pellucida-like domain containing 1 gene (*ZPLDI*), which is located at 101,818,088 to 102,196,462 bp. There were 3811 variants at the region after the quality control procedures, 24 of which were identified as deleterious (2 within the gene *ZPLDI*). Based on the result from SKAT-O, these variants were not associated with DBP ($p = 0.40$).

Discussion

We applied a population-based IBD mapping method to a quantitative trait, DBP. Because of the relatively small sample size ($n = 105$), statistical power to detect the causal chromosomal region was low, and the most



significant IBD mapping result did not reach the estimated significance level.

Interestingly, however, the gene at the most significant IBD mapping region, *ZPLD1*, has been reported to be related to cerebral cavernous malformations, a disease with enlarged small blood vessels (capillaries) in the brain, in a patient with a balanced translocation [20]. From Genetic Analysis Workshop 18, Bonner et al. [21] also reported an association between a sparse principal component (which included 28 SNPs at the intergenic region between *ZPLD1* and *MIR548A3*) and systolic blood pressure (SBP) using unrelated individuals from the family data ($n = 122$) and the GWAS genotypes. The Spearman correlation coefficient between the DBP and SBP residuals was 0.59 ($p < 0.0001$) in our data. Unfortunately, we were not able to identify causal variants in this gene for DBP. Further study with a large sample size is needed to assess this result.

Using the same IBD sharing estimate method, Browning et al. [6] reported that the average genome-wide IBD sharing rates were 0.015 for a Northern Finland sample and 0.0041 for a United Kingdom (UK) sample [6]. The

average chromosome-wide rate was 0.0029 from our Mexican American sample, which was lower than the UK sample, as expected. A higher IBD sharing rate (ie, on average, a larger combined length of detected IBD sharing segments per pair of individuals) can be obtained from an isolated founder population, such as the Northern Finland sample. Current IBD mapping methods are well powered to detect long IBD sharing segments (>1 cM). However, as marker density increases (eg, with whole genome sequencing data), IBD mapping methods will be able to identify IBD sharing status for small segments from outbred populations.

Conclusions

In summary, we demonstrated a gene mapping strategy for quantitative traits which combines a population-based IBD mapping method with sequence data analyses. However, as a result of the small sample size in this study, more information about this approach and its ability to prioritize causal variants for sequence data analyses should be further explored.

Acknowledgements

The authors gratefully acknowledge the study participants, data providers, and GAW19 organizers. We would also like to thank Mr. Alex Zhao for the R code that converts the Refined IBD output into the Fast IBD output format. Support for the authors was provided by the University of Manitoba (start-up funds for XQ Liu) and Canada Research Chair in Genetics of Complex Diseases (AD Paterson).

Declarations

This article has been published as part of *BMC Proceedings* Volume 10 Supplement 7, 2016: Genetic Analysis Workshop 19: Sequence, Blood Pressure and Expression Data. Summary articles. The full contents of the supplement are available online at <http://bmcpoc.biomedcentral.com/articles/supplements/volume-10-supplement-7>. Publication of the proceedings of Genetic Analysis Workshop 19 was supported by National Institutes of Health grant R01 GM031575.

Authors' contributions

X-QL designed the study, conducted statistical analyses, and drafted the manuscript; JF assisted with statistical analyses and interpretation; and ADP and PH assisted in study design and critical revision of the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Website resources

Genetic map: http://bochet.gcc.biostat.washington.edu/beagle/genetic_maps/
For the *ibdmerge.jar* utility program: http://faculty.washington.edu/browning/beagle_utilities/utilities.html.

Author details

¹Department of Obstetrics, Gynecology, and Reproductive Sciences, University of Manitoba, Winnipeg, MB R3E 3P4, Canada. ²Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB R3E 3P4, Canada. ³The Children's Hospital Research Institute of Manitoba, Winnipeg, MB R3E 3P4, Canada. ⁴George and Fay Yee Centre for Healthcare Innovation, University of Manitoba, Winnipeg, MB R3A 1R9, Canada. ⁵Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada. ⁶Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5G 0A4, Canada.

References

- MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science*. 2012; 335(6070):823–8.
- Zelinski T, Coghlan G, Liu XQ, Reid ME. ABCG2 null alleles define the Jr(a-) blood group phenotype. *Nat Genet*. 2012;44(2):131–2.
- Dawn Teare M, Barrett JH. Genetic linkage studies. *Lancet*. 2005;366(9490): 1036–44.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81(3):559–75.
- Gusev A, Kenny EE, Lowe JK, Salit J, Saxena R, Kathiresan S, Altshuler DM, Friedman JM, Breslow JL, Pe'er I. DASH: a method for identical-by-descent haplotype mapping uncovers association with recent variation. *Am J Hum Genet*. 2011;88(6):706–17.
- Browning BL, Browning SR. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*. 2013;194(2): 459–71.
- Hochreiter S. HapFABIA: identification of very short segments of identity by descent characterized by rare variants in large sequencing data. *Nucleic Acids Res*. 2013;41(22):e202.
- Goldstein DB, Allen A, Keebler J, Margulies EH, Petrou S, Petrovski S, Sunyaev S. Sequencing studies in human genetics: design and interpretation. *Nat Rev Genet*. 2013;14(7):460–70.
- Francks C, Tozzi F, Farmer A, Vincent JB, Rujescu D, St Clair D, Muglia P. Population-based linkage analysis of schizophrenia and bipolar case-control cohorts identifies a potential susceptibility locus on 19q13. *Mol Psychiatry*. 2010;15(3):319–25.
- Browning SR, Thompson EA. Detecting rare variant associations by identity-by-descent mapping in case-control studies. *Genetics*. 2012;190(4):1521–31.
- Lin R, Charlesworth J, Stankovich J, Perreau VM, Brown MA, Taylor BV, ANZgene Consortium. Identity-by-descent mapping to detect rare variants conferring susceptibility to multiple sclerosis. *PLoS One*. 2013;8(3):e56379.
- Sham PC, Cherny SS, Purcell S. Application of genome-wide SNP data for uncovering pairwise relationships and quantitative trait loci. *Genetica*. 2009; 136(2):237–43.
- Cui JS, Hopper JL, Harrap SB. Antihypertensive treatments obscure familial contributions to blood pressure variation. *Hypertension*. 2003;41(2):207–10.
- Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):e190.
- Feingold E. Regression-based quantitative-trait-locus mapping in the 21st century. *Am J Hum Genet*. 2002;71(2):217–22.
- Forrest WF. Weighting improves the “new Haseman-Elston” method. *Hum Hered*. 2001;52(1):47–54.
- Xu X, Weiss S, Xu X, Wei LJ. A unified Haseman-Elston method for testing linkage with quantitative traits. *Am J Hum Genet*. 2000;67(4):1025–8.
- Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310–5.
- Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Christiani DC, Wurfel MM, Lin X, NHLBI GO Exome Sequencing Project—ESP Lung Project Team. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet*. 2012;91(2):224–37.
- Gianfrancesco F, Esposito T, Penco S, Maglione V, Liquori CL, Patrosso MC, Zuffardi O, Ciccodicola A, Marchuk DA, Squitieri F. ZPLD1 gene is disrupted in a patient with balanced translocation that exhibits cerebral cavernous malformations. *Neuroscience*. 2008;155(2):345–9.
- Bonner A, Neupane B, Beyene J. Testing for associations between systolic blood pressure and single-nucleotide polymorphism profiles obtained from sparse principal component analysis. *BMC Proc*. 2014;8(Suppl 1):S95.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

