Published in final edited form as:
Ann Hum Genet. 2013 July ; 77(4): 277–287. doi:10.1111/ahg.12021.

Genome-wide association study of preeclampsia detects novel maternal single nucleotide polymorphisms and copy-number variants in subsets of the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study cohort

Linlu Zhao[1,2], **Michael B. Bracken**[1,2,3], and **Andrew T. DeWan**[1,2,*]

[1]Center for Perinatal, Pediatric and Environmental Epidemiology, Yale Schools of Public Health and Medicine, New Haven, Connecticut, USA

[2]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut, USA

[3]Department of Obstetrics, Gynecology, and Reproductive Sciences, Yale School of Medicine, New Haven, Connecticut, USA

Summary

A genome-wide association study was undertaken to identify maternal single nucleotide polymorphisms (SNPs) and copy-number variants (CNVs) associated with preeclampsia. Case-control analysis was performed on 1070 Afro-Caribbean (n=21 cases and 1049 controls) and 723 Hispanic (n=62 cases and 661 controls) mothers and 1257 mothers of European ancestry (n=50 cases and 1207 controls) from the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study. European ancestry subjects were genotyped on Illumina Human610-Quad and Afro-Caribbean and Hispanic subjects were genotyped on Illumina Human1M-Duo BeadChip microarrays. Genome-wide SNP data were analyzed using PLINK. CNVs were called using three detection algorithms (GNOSIS, PennCNV, and QuantiSNP), merged using CNVision, and then screened using stringent criteria. SNP and CNV findings were compared to those of the Study of Pregnancy Hypertension in Iowa (SOPHIA), an independent preeclampsia case-control dataset of Caucasian mothers (n=177 cases and 116 controls). A list of top SNPs were identified for each of the HAPO ethnic groups, but none reached Bonferroni-corrected significance. Novel candidate CNVs showing enrichment among preeclampsia cases were also identified in each of the three ethnic groups. Several variants were suggestively replicated in SOPHIA. The discovered SNPs and copy-number variable regions present interesting candidate genetic variants for preeclampsia that warrant further replication and investigation.

[*]Corresponding author: Andrew T. DeWan, Yale School of Public Health, Department of Chronic Disease Epidemiology, 60 College Street, Room 403, New Haven, CT 06520, USA; phone: 203-785-3528; fax: 203-785-6279; andrew.dewan@yale.edu.

Conflict of interest

The authors declare that they have no conflicts of interest.

Keywords

copy-number variant; genome-wide association study; microarray analysis; preeclampsia; single nucleotide polymorphism

INTRODUCTION

Preeclampsia (PE) is a multi-system, hypertensive disorder of pregnancy that complicates 2–7% of pregnancies (Sibai, et al. 2005). Characterized by new onset hypertension and proteinuria, PE is a major contributor to maternal morbidity and mortality worldwide. Maternal ethnic differences in PE rates exist: in one study, compared to Caucasians, Afro-American women had higher risk of PE (odds ratio=1.41, 95% confidence interval: 1.25–1.62), while Asian (OR=0.79, 95% CI: 0.72–0.88) and Hispanic (OR=0.90, 95% CI: 0.84–0.97) women had lower risk (Caughey et al. 2005).

Although evidence from family-based studies suggests PE has a heritable component, its exact etiology and specific genetic contributions remain unknown (Arngrimsson et al. 1990, Lie et al. 1998, Mogren et al. 1999, Sutherland et al. 1981). Susceptibility to PE is thought to be influenced by complex interactions between maternal and fetal genetic and environmental factors (Williamson, 2005). Cnattingius et al. (2004) found that 35% of the variance in PE was attributable to maternal genetic effects, 20% to fetal genetic effects, 13% to the “couple effect,” and 32% to unmeasured environmental factors.

Many candidate gene studies have been conducted in attempt to elucidate genetic contributions to PE, but no single gene has been identified with consistent replication (Williamson, 2005). The lack of success with the candidate gene approach may be due to inadequate statistical power and poor study design (Colhoun et al. 2003). In addition, studies examining the genetic factors contributing to PE have focused on single nucleotide polymorphisms (SNPs), which explain only a fraction of overall heritability. An estimated 13% of the genome has been reported to be copy-number variable in one or more individuals, suggesting that copy-number variants (CNVs) may be substantial contributors to the “missing heritability” (Stankiewicz & Lupski. 2010). Prior genome-wide CNV analysis from the Study of Pregnancy Hypertension in Iowa (SOPHIA) conducted by the authors and colleagues discovered three rare deletions that may confer risk for PE, including a potentially functionally important copy-number deletion in *PSG11* (Zhao et al. 2012).

A genome-wide association (GWA) study using a case-control design was conducted to identify PE-associated SNPs and CNVs in Afro-Caribbean, European ancestry, and Hispanic ethnic groups using publicly available genotype data from the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study (HAPO Study Cooperative Research Group, 2002). The present study was also compared to the SOPHIA study to attempt corroboration of findings.

MATERIALS AND METHODS

Study population

The HAPO study is an international, multi-center prospective cohort study of maternal glycemia and adverse pregnancy outcomes. Detailed recruitment methods have been described elsewhere (HAPO Study Cooperative Research Group, 2002). Approximately 4750 Afro-Caribbean (Barbados), European ancestry (Brisbane and Newcastle, Australia; Toronto, Canada; and Belfast, United Kingdom), Hispanic (Bellflower, CA), and Thai (Bangkok, Thailand) mothers were genotyped.

A case-control design was applied to each ethnic group in the present study with subjects drawn from the HAPO study with both phenotype and genotype information available (Afro-Caribbean, n=23 cases and 1075 controls; European ancestry, n=52 cases and 1250 controls; and Hispanic, n=62 cases and 661 controls). Genotype and phenotype information were acquired from the National Center for Biotechnology Information's Genotypes and Phenotypes Database (NCBI dbGaP study accession: phs000096.v4.p1) (HAPO Study Cooperative Research Group, 2002). Thai mothers were not further considered for analysis due to a very limited number of PE cases (n=9). Mothers presenting with PE were eligible as cases and normotensive mothers were eligible as controls. Subjects with chronic hypertension, gestational hypertension, or PE superimposed on chronic hypertension as well as those with unavailable details on hypertension status were excluded.

Phenotype definition

PE was defined according to the International Society for the Study of Hypertension in Pregnancy guidelines (Brown. et al. 2001) as having *de novo* hypertension (systolic blood pressure ≥140 mmHg and/or diastolic blood pressure ≥90 mmHg on two or more occasions at a minimum of six hours apart) and proteinuria (dipstick protein test value ≥1+ or a 24-hour urine collection with protein ≥300 mg) after 20 weeks of gestation.

Whole-genome genotyping

Genomic DNA was isolated from peripheral blood samples using Autopure LS (Gentra Systems, Minneapolis, MN) (Urbanek. et al. 2012). Genotyping was performed at the Broad Institute of MIT and Harvard using Illumina Human610-Quad (European ancestry) and Illumina Human1M-Duo (Afro-Caribbean and Hispanic) BeadChip microarray platforms (Illumina, San Diego, CA), interrogating over 601,000 and 1.1 million SNP and copy-number (non-polymorphic) probes, respectively.

Sample quality was assessed by signal intensity using R and overall call rate using PLINK version 1.07 (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al. 2007). Subjects were excluded if low signal intensity was exhibited across all chromosomes. When examining the mean chromosomal probe intensities, these excluded samples formed an outlying low-intensity "tail" from the main cluster, a pattern consistent throughout the autosomes and X chromosome. No additional samples were excluded due to having overall call rate lower than 90% (default value in PLINK), with Afro-Caribbean, European ancestry, and Hispanic subjects having mean call rates of 99.9%, 95.4%, and 99.6% respectively. A

total of 1070 Afro-Caribbean (n=21 cases and 1049 controls), 1257 European ancestry (n=50 cases and 1207 controls), and 723 Hispanic (n=62 cases and 661 controls) mothers met eligibility criteria for further analysis.

SNP analysis

Genome-wide SNP analysis was performed using PLINK at default settings. A total of 1,145,510 SNPs were genotyped for both Afro-Caribbean and Hispanic study subjects on Illumina Human1M microarrays and 598,821 SNPs for European ancestry subjects on Illumina Human610 microarrays. SNPs with minor allele frequency (MAF) <0.01 (Afro-Caribbean, n=160,642; European ancestry, n=55,303; and Hispanic, n=141,199), call rate <95% (Afro-Caribbean, n=84,828; European ancestry, n=28,141; and Hispanic, n=54,248), or significant deviation from Hardy-Weinberg equilibrium at *p*-value ≤0.001 (Afro-Caribbean, n=4616; European ancestry, n=1529; and Hispanic, n=4390) were excluded, leaving 979,693 (Afro-Caribbean), 541,023 (European ancestry), and 964,533 (Hispanic) SNPs.

Genetic quality control included assessments for both population stratification and cryptic relatedness. Population stratification was assessed using EIGENSTRAT version 3.0 (http://genepath.med.harvard.edu/~reich/Software.htm) by performing a principal components analysis (Price et al. 2006). Cryptic relatedness between subjects in each ethnic group was examined by pairwise identity-by-descent (IBD) estimation in PLINK. Pairs showing pi-hat (estimated proportion of genomic variation shared IBD) >0.2 were inspected and 39, 5 and 3 controls were excluded from downstream SNP and CNV analyses in Afro-Caribbean, European ancestry and Hispanic groups, respectively. Logistic regression analysis was performed to test the association between individual SNPs and PE with adjustment for significant principal components identified from the EIGENSTRAT analysis. Unadjusted genotype association testing was also performed using the Cochrane-Armitage trend test to facilitate comparisons of *p*-values of genotyped and ungenotyped SNPs among the HAPO study groups and the SOPHIA study. Ungenotyped SNPs in Afro-Caribbean subjects, subjects of European descent from HAPO and Caucasian subjects from SOPHIA, and Hispanic subjects were imputed and tested in PLINK using HapMap 3 female founders of African ancestry in Southwest USA (ASW), Northern and Western European ancestry from Utah (CEU), and Mexican ancestry in Los Angeles, CA (MEX), respectively (http://www.broadinstitute.org/~debakker/p3.html). Bonferroni-corrected genome-wide significance thresholds were applied to each ethnic group to maintain an overall study α of 0.05. As Bonferroni correction is conservative, only tested markers were included in the calculation of the corrected significance thresholds; therefore, the significance thresholds used were 5.10×10^{-8} (0.05/979,693) for Afro-Caribbean subjects, 9.24×10^{-8} (0.05/541,023) for subjects of European ancestry, and 5.18×10^{-8} (0.05/964,533) for Hispanic subjects. Mitochondrial SNPs were excluded from analysis.

CNV analysis

CNVision (Sanders et al. 2011), a pipeline specifically designed for use with Illumina microarray data, was used to predict and analyze CNVs using three algorithms with two independent approaches: GNOSIS (Sanders et al. 2011), PennCNV June 16, 2011 version

(Wang et al. 2007), and QuantiSNP version 1.1 (Colella et al. 2007). GNOSIS is a distribution function algorithm that uses sliding windows to detect CNVs. In contrast, PennCNV and QuantiSNP both implement a hidden Markov model. However, all three detection algorithms utilize log R ratio (LRR; a measure of total signal intensity of probes) and B allele frequency (BAF; a measure of relative intensity ratio of allelic probes) information to detect CNVs. Raw data were batch-processed by ethnic group and detection algorithms were run under microarray-specific default CNVision pipeline settings for autosomal and X chromosomes.

Stringent sample and call quality control (QC) measures were undertaken in attempt to increase the accuracy of CNV prediction. Samples with log-transformed number of CNV calls greater than three standard deviations from the mean of the log-transformed calls per sample for at least one of the three algorithms were considered of low quality and removed from further analysis (Afro-Caribbean, n=50 controls; European ancestry, n=50 controls; Hispanic, n=2 cases and 13 controls). Related subjects identified by pairwise IBD were also removed (Afro-Caribbean, n=37 controls; European ancestry, n=5 controls; Hispanic, n=3 controls). Raw CNV calls spanning less than ten microarray probes were excluded. CNVision was then used to merge, analyze, and annotate the QC filtered outputs of GNOSIS, PennCNV, and QuantiSNP. A merged CNV call meeting one or more of the following screening criteria was excluded: 1) <50% overlap between two algorithms and <25% overlap among three algorithms; 2) <10 consecutive microarray probes; or 3) contains conflicting calls (both deletion and amplification) across algorithms within the same genomic region in the given sample. Finally, the following selection criteria were applied to minimal regions of overlap of merged CNV calls across samples to generate a shortlist of candidate regions: 1) CNVs called in ≥3 cases and absence of calls in controls or, if present in controls, having positive ORs comparing PE cases versus normotensive controls and corresponding 95% CIs that excluded the null; 2) does not overlap centromeric or telomeric regions according to PennCNV definitions; and 3) contains ≥10 consecutive SNP or copy-number probes. Figure 1 illustrates the boundaries of a minimal region of overlap of merged CNV calls across samples for a deletion enriched in HAPO subjects of European ancestry.

Reliability of CNVs detected *in silico* was assessed by visual inspection of LRR and BAF plots for a clear change in probe hybridization intensity and zygosity, respectively. Fisher's exact tests were performed to quantify the significance of the difference in CNV frequency. Top SNPs and CNVs identified in the HAPO study were compared to the SOPHIA study, an independent dataset of primiparous Caucasian women, for replication of findings. The SOPHIA study genotyped 177 PE cases and 116 normotensive controls using the Affymetrix Genome-Wide Human SNP Array 6.0 (Zhao et al. 2012). Bonferroni-corrected significance threshold in the SOPHIA study for the SNP analysis was defined as 7.1×10^{-8}. SNPs and CNVs identified in HAPO reaching nominal significance in SOPHIA (*p*-value<0.05) were considered to be suggestively replicated.

Genomic positions are presented according to the NCBI36/hg18 human genome assembly.

RESULTS

SNP associations

SNP analysis was performed on 1031 Afro-Caribbean (n=21 cases and 1010 controls), 1252 European ancestry (n=50 cases and 1202 controls), and 720 Hispanic (n=62 cases and 658 controls) subjects. EIGENSTRAT analysis found evidence of population stratification in Afro-Caribbean and European ancestry study subjects; the ninth eigenvector was significant for Afro-Caribbean subjects (principal component 9 or PC9, p=0.029; see Figure S1 for scatter plots of PC8 vs. PC9 and PC9 vs. PC10) and the first eigenvector was significant for subjects of European descent (PC1, p=0.004; see Figure S2 for a scatter plot of PC1 vs. PC2). PC1 and PC9 were adjusted in the logistic regression models for European ancestry and Afro-Caribbean subjects, respectively. The top 10 principal components for Hispanic subjects were not significant.

After Bonferroni correction for multiple testing, no SNPs were found to be significantly associated with PE in any ethnic group. The most significant SNP identified in Afro-Caribbean subjects adjusted for PC9 was rs11617740 in *FGF14* (OR=16.96, 95% CI: 5.53–51.97, $p=7.32\times10^{-7}$) and in European ancestry subjects adjusted for PC1 was rs7322722 in *MYCBP2* (OR=2.93, 95% CI: 1.90–4.52, $p=1.23\times10^{-6}$). For Hispanic subjects, the most significant SNP was rs17412740 located in an intergenic region near *LZTS1* with a p-value of 2.14×10^{-6} and an OR of 6.08 (95% CI: 2.88–12.81). Top SNP candidates (logistic regression p-value $<10^{-5}$) for the three HAPO ethnic groups are presented in Table S1. LocusZoom (Pruim= et al. 2010) plots visualizing association results for the most significant SNP in each of the three HAPO ethnic groups and its ±500 kb flanking region SNPs are shown in Figures S3-S5. The chromosome 13 region of European ancestry subjects contains a set of SNPs with small p-values ($<10^{-5}$) and relatively high degrees of association ($r^2>0.8$) to rs7322722 within and proximal to the *MYCBP2* gene (Fig. S4).

CNV associations

Afro-Caribbean subjects—Among Afro-Caribbean mothers, 23,962 putative CNVs (15,774 deletions and 8188 amplifications) spanning ten or more SNP or copy-number probes were detected among the 21 PE cases and 962 normotensive controls that passed sample QC. CNVs ranging in size from ~0.8 kb to 21.5 Mb (median=49.0 kb) were merged into 7739 regions of deletion and 4630 regions of amplification. Of these, nine minimal regions of overlap of recurrent CNVs met the selection criteria (see Table 1 for an annotated list of candidate CNVs among Afro-Caribbean mothers). Five of the candidate CNVs achieved nominal significance (p-value<0.05).

Subjects of European ancestry—Among mothers of European descent, 13,733 CNVs containing ten or more consecutive probes (8242 deletions and 5491 amplifications) were detected among the 50 PE cases and 1152 normotensive controls meeting sample QC. CNV calls of ~3.1 kb to 25.4 Mb (median=98.5 kb) were merged into 4493 regions of deletion and 2593 regions of amplification with a total of three genomic regions of minimal call overlap meeting the selection criteria (see Table 2 for an annotated list of candidate CNVs

among mothers of European descent). Nominal significance was achieved for all three CNVs.

Hispanic subjects—A total of 24,046 filtered CNV calls spanning at least 10 probes (20,597 deletions and 3449 amplifications) were made among the eligible 60 PE case and 645 normotensive control Hispanic subjects. Genomic variants ranged in size from ~0.8 kb to 8.6 Mb (median=81.5 kb). CNV calls were merged and screened into 15,263 regions of deletion and 2183 regions of amplification. Fourteen of 15 minimal regions of call overlap enriched among cases that met the pre-specified selection criteria reached nominal significance (see Table 3 for an annotated list of candidate CNVs among Hispanic mothers).

SOPHIA study replication

Among the top SNPs identified in each of the ethnic groups of the HAPO study, none reached Bonferroni-corrected significance and only several were nominally significant ($p<0.05$) in the comparison groups (Table S1). Despite a lack of replication due to the number of examined SNPs, it is interesting to note that two SNPs in HAPO subjects of European ancestry on chromosome 9 (rs7047693 and rs10988989) in an intronic region of the *INVS* gene yielded trend test *p*-values of <0.05 in Caucasian SOPHIA subjects.

When comparing the SOPHIA study to the HAPO CNV findings, similar patterns of calls were observed for the copy-number amplification in 12p11.21 in HAPO subjects of European ancestry and in two regions of amplification enriched in Hispanic cases at 5p13.3 and 7q11.23. The intergenic amplification in chromosome 12 was detected in 8/50 (16.0%) cases and 72/1152 (6.3%) controls of the HAPO study and in 19/169 (11.2%) cases and 6/114 (5.3%) controls of the SOPHIA study. Both chromosome 5 and 7 amplifications are exonic. The former region was detected in 5/60 (8.3%) cases and 10/645 (1.6%) controls of the HAPO study and in 6/169 (3.6%) cases and 2/114 (1.8%) controls of the SOPHIA study while the latter region was detected in 4/60 (6.7%) cases and 8/645 (1.2%) controls of the HAPO study and in 4/169 (2.4%) cases and 1/114 (0.9%) controls of the SOPHIA study. Other candidate CNVs identified in HAPO were not replicated.

DISCUSSION

Numerous SNP candidates were identified for each of the HAPO ethnic groups, but none were significantly associated with PE after Bonferroni correction (Table S1). Because the number of PE cases for each ethnic group was relatively small, the present study had limited statistical power to detect a significant association for SNPs of small to moderate effect size. Post-hoc power calculations were performed in PS: Power and Sample Size Calculation version 3.0 (http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize) using assumptions of minor allele frequency from dbSNP (Sherry et al. 2001) and effect size based on logistic regression ORs for the top ranked SNP (Table S1) of each HAPO study group. This study had ~81.2%, 9.8% and 59.4% power to detect associations at the effect sizes reported for the top SNPs among Afro-Caribbean, European ancestry, and Hispanic subjects, respectively. Another limitation on statistical power is that the principal components analysis used to assess and correct population stratification does not explicitly model cryptic

relatedness; association statistics that account for cryptic relatedness may achieve higher power due to improved weighting of data (Price et al. 2010).

Genome-wide CNV analysis was successfully performed on 983 Afro-Caribbean (n=21 PE cases and 962 normotensive controls), 1202 European ancestry (n=50 cases and 1152 controls), and 705 Hispanic (n=60 cases and 645 controls) subjects. Multiple candidate regions of copy-number deletion or amplification were detected in each of the three ethnic groups, but none were shared among them (Tables 1–3). Although a majority of these candidate CNVs showed nominal significance with PE (p-value<0.05), no region surpassed the threshold for significance accounting for multiple comparisons.

Array-based CNV detection methods are typically plagued by substantial false-positive and false-negative rates (Tsuang et al. 2010). Until recently, studies commonly used a single algorithm detection approach. Due to the limitations of a single algorithm strategy, molecular techniques are necessary to confirm the presence of CNVs. However, HAPO study DNA samples were unavailable for laboratory validation of identified CNVs. Despite this, confidence can be placed in the accuracy of CNV call prediction as three algorithms (GNOSIS, PennCNV, and QuantiSNP) following two independent statistical approaches were employed with stringent algorithm overlap criteria. Moreover, the reliability of calls can be assessed visually using the LRR and BAF plots. For example, LRR drops to around −0.5 and BAF clusters around 0 or 1, but not near 0.5, for a heterozygous copy-number deletion call, as shown in Figure 2. Inspection of the plots for the candidate regions revealed signal intensity patterns consistent with their respective copy-number calls.

Several novel candidate CNVs were identified in the HAPO ethnic groups under study. However, the present study was unable to replicate the top three CNVs identified from the SOPHIA study (Zhao et al. 2012). The lack of replication may be attributable to microarray probe coverage differences in these genomic regions, an issue that is also relevant to the SNP analysis. The Illumina Human610 microarray used to genotype HAPO subjects of European descent only contained one probe in the 16p13.11 deleted region and two probes in the 19p13.31 deleted region, compared to ten and eight probes respectively in these genomic regions on the Affymetrix SNP 6.0 microarray used in SOPHIA. Contrarily, the three novel copy-number variable genomic regions identified among European ancestry subjects in HAPO (Table 2) have comparable probe coverage to SOPHIA. The lack of replication in the present study of the top SOPHIA CNVs may be further attributable to geographic differences between the two study populations. Mothers of European descent from the HAPO study were recruited from Australia, Canada, and the United Kingdom whereas Caucasian mothers from the SOPHIA study were recruited from Iowa. Comparing samples from different geographic populations may be problematic as some variants are specific to certain populations (Redon et al. 2006). In support of this view, CNV frequencies appear to differ between the control subjects of the two studies (see Table S2 for a comparison of the call frequencies in female controls and differences in probe coverage for the top candidate regions identified from SOPHIA with HAPO subjects of European ancestry). Therefore, comparability of the two studies for CNV findings is diminished due to microarray coverage differences and geographic variation of the study populations. At the time of writing, no other PE datasets were available to the authors as replication cohorts for

array-based CNV detection. These issues, in addition to maintaining adequate statistical power, should be taken into consideration for future PE genetic studies.

In this GWA study, a list of top SNP candidates was identified among Afro-Caribbean, European ancestry, and Hispanic HAPO study subjects, but none reached Bonferroni-corrected significance. However, the top SNP candidates reported in this study may be useful for future work in PE genetics as a resource for replication and meta-analysis. Genome-wide analysis of CNVs across the three ethnic groups identified several novel variants enriched in cases that were markedly less frequent in controls. The validity and functional significance of these candidate CNVs need further investigation. Larger replication studies are necessary for confirmation of the variants identified in this study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank M. Geoffrey Hayes for providing additional methodological details of the HAPO study. Computational work was supported by the Yale University Biomedical High Performance Computing Center and the National Institutes of Health (RR19895 and RR029676-01). Funding support for the GWA mapping: Maternal Metabolism-Birth Weight Interactions study was provided through the NIH Genes, Environment and Health Initiative (GEI) (U01HG004415). The GWA mapping: Maternal Metabolism-Birth Weight Interactions study is one of the GWA studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Funding support for genotyping, which was performed at the Broad Institute of MIT and Harvard, was provided by the NIH GEI (U01 HG04424). The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap through dbGaP accession number phs000096.v4.p1.

References

Arngrimsson R, Björnsson S, Geirsson RT, Björnsson H, Walker JJ, Snaedal G. Genetic and familial predisposition to eclampsia and pre-eclampsia in a defined population. Br J Obstet Gynaecol. 1990; 97:762–769. [PubMed: 2242360]

Brown MA, Lindheimer MD, de Swiet M, Van Assche A, Moutquin JM. The classification and diagnosis of the hypertensive disorders of pregnancy: statement from the International Society for the Study of Hypertension in Pregnancy (ISSHP). Hypertens Pregnancy. 2001; 20:IX–XIV. [PubMed: 12044323]

Caughey AB, Stotland NE, Washington AE, Escobar GJ. Maternal ethnicity, paternal ethnicity, and parental ethnic discordance: predictors of preeclampsia. Obstet Gynecol. 2005; 106:156–161. [PubMed: 15994632]

Cnattingius S, Reilly M, Pawitan Y, Lichtenstein P. Maternal and fetal genetic factors account for most of familial aggregation of preeclampsia: a population-based Swedish cohort study. Am J Med Genet A. 2004; 130A:365–371. [PubMed: 15384082]

Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, Bassett AS, Seller A, Holmes CC, Ragoussis J. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. 2007; 35:2013–2025. [PubMed: 17341461]

Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. Lancet. 2003; 361:865–872. [PubMed: 12642066]

HAPO Study Cooperative Research Group. The Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study. Int J Gynaecol Obstet. 2002; 78:69–77. [PubMed: 12113977]

Lie RT, Rasmussen S, Brunborg H, Gjessing HK, Lie-Nielsen E, Irgens LM. Fetal and maternal contributions to risk of pre-eclampsia: population based study. BMJ. 1998; 316:1343–1347. [PubMed: 9563982]

Mogren I, Högberg U, Winkvist A, Stenlund H. Familial occurrence of preeclampsia. Epidemiology. 1999; 10:518–522. [PubMed: 10468424]

Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nat Rev Genet. 2010; 11:459–463. [PubMed: 20548291]

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. [PubMed: 16862161]

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: Regional visualization of genome-wide association scan results. Bioinformatics. 2010; 26:2336–2337. [PubMed: 20634204]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. Nature. 2006; 444:444–454. [PubMed: 17122850]

Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA, Mason CE, Bilguvar K, Celestino-Soper PB, Choi M, Crawford EL, Davis L, Wright NR, Dhodapkar RM, DiCola M, DiLullo NM, Fernandez TV, Fielding-Singh V, Fishman DO, Frahm S, Garagaloyan R, Goh GS, Kammela S, Klei L, Lowe JK, Lund SC, McGrew AD, Meyer KA, Moffat WJ, Murdoch JD, O'Roak BJ, Ober GT, Pottenger RS, Raubeson MJ, Song Y, Wang Q, Yaspan BL, Yu TW, Yurkiewicz IR, Beaudet AL, Cantor RM, Curland M, Grice DE, Günel M, Lifton RP, Mane SM, Martin DM, Shaw CA, Sheldon M, Tischfield JA, Walsh CA, Morrow EM, Ledbetter DH, Fombonne E, Lord C, Martin CL, Brooks AI, Sutcliffe JS, Cook EH Jr, Geschwind D, Roeder K, Devlin B, State MW. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron. 2011; 70:863–885. [PubMed: 21658581]

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001; 29:308–311. [PubMed: 11125122]

Sibai B, Dekker G, Kupferminc M. Pre-eclampsia. Lancet. 2005; 365:785–799. [PubMed: 15733721]

Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annu Rev Med. 2010; 61:437–455. [PubMed: 20059347]

Sutherland A, Cooper DW, Howie PW, Liston WA, MacGillivray I. The indicence of severe pre-eclampsia amongst mothers and mothers-in-law of pre-eclamptics and controls. Br J Obstet Gynaecol. 1981; 88:785–791. [PubMed: 7259998]

Tsuang DW, Millard SP, Ely B, Chi P, Wang K, Raskind WH, Kim S, Brkanac Z, Yu CE. The effect of algorithms on copy number variant detection. PLoS One. 2010; 5:e14456. [PubMed: 21209939]

Urbanek M, Hayes MG, Lee H, Freathy RM, Lowe LP, Ackerman C, Jafari N, Dyer AR, Cox NJ, Dunger DB, Hattersley AT, Metzger BE, Lowe WL Jr. The Role of Inflammatory Pathway Genetic Variation on Maternal Metabolic Phenotypes during Pregnancy. PLoS One. 2012; 7:e32958. [PubMed: 22479352]

Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007; 17:1665–1674. [PubMed: 17921354]

Williamson C. Molecular biology related to pre-eclampsia. Int Congr Ser. 2005; 1279:282–289.

Zhao L, Triche EW, Walsh KM, Bracken MB, Saftlas AF, Hoh J, Dewan AT. Genome-wide association study identifies a maternal copy-number deletion in PSG11 enriched among preeclampsia patients. BMC Pregnancy Childbirth. 2012; 12:61. [PubMed: 22748001]

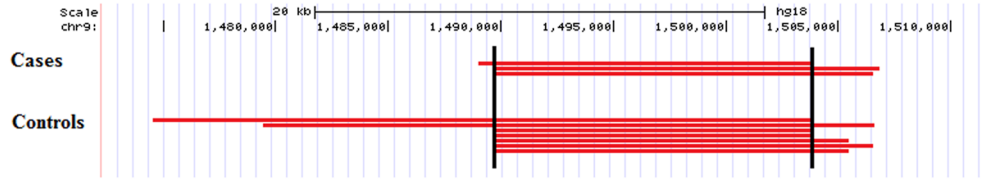

Figure 1. UCSC Genome Browser plot of the copy-number deletion at chr9:1.490–1.503 Mb identified in HAPO subjects of European ancestry

The vertical black lines indicate the minimal region of overlap of merged calls across all subjects harboring the deletion. Each red horizontal bar represents the length and breakpoints of a merged deletion call detected in either PE cases or controls in HAPO subjects of European ancestry. Exact CNV breakpoints are unknown.

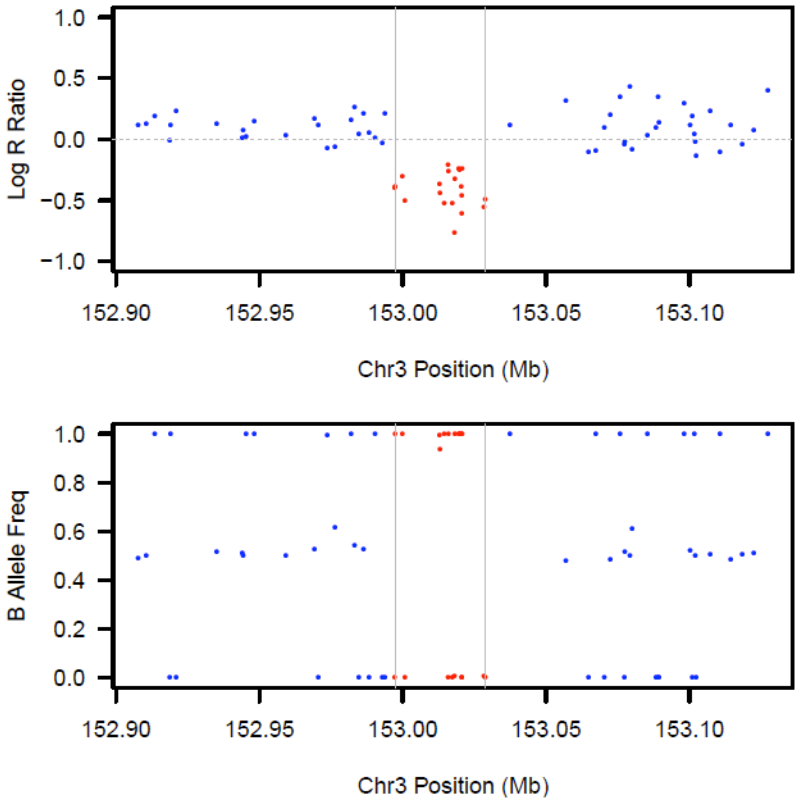

Figure 2. Representative LRR and BAF plots for a genomic region called as a heterozygous deletion

LRR and BAF values for each probe are represented as dots. The algorithm-detected deletion is encompassed between the two vertical grey bars (chr3:152.998–153.028 Mb). LRR values for the SNP and copy-number probes drop to the −0.5 region and BAF values for the SNP probes cluster around 0 or 1 within the deleted region (red dots). LRR values of the bordering copy-normal chromosomal regions bunch around zero and BAF values segregate to three clusters (blue dots).

Table 1

CNVs identified among Afro-Caribbean HAPO study subjects[1]

CNV region (chromosome: start position-stop position)	CNV type	Region size (bp)	# probes within region	# case CNV calls (n=21)	# control CNV calls (n=962)	OR (95% CI)	F-exact *p*-value	# genes (exons) in region	Gene in region	Location	5′ gene	5′ distance (kb)	3′ gene	3′ distance (kb)
chr2:56583659-56599070	Amplification	15412	11	3	15	10.52 (2.80–39.56)	0.006	0 (0)	-	Intergenic	*CCDC85A*	116.8	*VRK2*	1528.2
chr2:133001803-133025106	Amplification	23304	14	4	40	5.42 (1.74–16.86)	0.012	1 (0)	*GPR39*	Intronic	In	0.0	In	0.0
chr2:184807930-184822613	Deletion	14684	11	3	39	3.94 (1.11–13.96)	0.056	0 (0)	-	Intergenic	*NUP35*	1073.3	*ZNF804A*	348.7
chr2:184822613-184842482	Deletion	19870	13	3	38	4.05 (1.14–14.35)	0.053	0 (0)	-	Intergenic	*NUP35*	1088.0	*ZNF804A*	328.9
chr4:69720442-69740862	Amplification	20421	15	3	31	5.01 (1.40–17.89)	0.033	1 (4)	*UGT2B10*	Exonic	In	0.0	In	0.0
chr6:29976032-30001174	Deletion	25143	36	7	143	2.86 (1.14–7.22)	0.030	2 (2)	*HCG2P7, HCG4P6*	Exonic	In	0.0	In	0.0
chr6:31468234-31557306	Deletion	89073	206	3	39	3.94 (1.11–13.96)	0.056	3 (9)	*HCG26, HCP5, MICA*	Exonic	In	0.0	In	0.0
chr6:31557757-31561597	Deletion	3841	12	3	43	3.56 (1.01–12.56)	0.071	0 (0)	-	Intergenic	*HCG26*	9.6	*MICB*	12.2
chr7:78652346-78681225	Deletion	28880	12	4	20	11.08 (3.42–35.92)	0.001	1 (0)	*MAGI2*	Intronic	In	0.0	In	0.0

[1] Shaded areas indicate contiguous merged CNV regions. Genomic positions are indicated according to the NCBI36/hg18 human genome assembly.

Abbreviations: bp, base pair; CI, confidence interval; CNV, copy-number variant; F-exact, Fisher's exact test; OR, odds ratio; PE, preeclampsia.

Table 2

CNVs identified among HAPO study subjects of European ancestry[1]

CNV region (chromosome: start position-stop position)	CNV type	Region size (bp)	# probes within region	# case CNV calls (n=50)	# control CNV calls (n=1152)	OR (95% CI)	F-exact *p*-value	# genes (exons) in region	Gene in region	Location	5′ gene	5′ distance (kb)	3′ gene	3′ distance (kb)
Chr3:152997280-153028731	Deletion	31452	20	4	17	5.81 (1.88–17.94)	0.009	1 (5)	*AADAC*	Exonic	In	0.0	In	0.0
Chr9:1489641-1503917	Deletion	14277	23	3	7	10.44 (2.62–41.65)	0.007	0 (0)	-	Intergenic	*DMRT2*	442.1	*SMARCA2*	501.4
Chr12:31257563-31296067	Amplification	38505	17	8	72	2.86 (1.29–6.31)	0.015	0 (0)	-	Intergenic	*DDX11*	108.6	*FAM60A*	28.7

[1] Black box borders indicate contiguous merged CNV regions. Genomic positions are indicated according to the NCBI36/hg18 human genome assembly.

Abbreviations: bp, base pair; CI, confidence interval; CNV, copy-number variant; F-exact, Fisher's exact test; OR, odds ratio; PE, preeclampsia.

Table 3

CNVs identified among Hispanic HAPO study subjects[1]

CNV region (chromosome: start position-stop position)	CNV type	Region size (bp)	# probes within region	# case CNV calls (n=60)	# control CNV calls (n=645)	OR (95% CI)	F-exact *p*-value	# genes (exons) in region	Gene in region	Location	5′ gene	5′ distance (kb)	3′ gene	3′ distance (kb)
Chr1:12783336-12793914	Deletion	10579	10	4	13	3.47 (1.10–11.01)	0.049	0 (0)	-	Intergenic	*PRAMEF1*	4.5	*HNRNPCL1*	35.9
Chr1:99983403-100017169	Deletion	33767	11	3	3	11.26 (2.22–57.09)	0.010	1 (4)	*FRRS1*	Exonic	In	0.0	In	0.0
Chr2:57832362-57931532	Deletion	99171	23	3	5	6.74 (1.57–28.92)	0.024	0 (0)	-	Intergenic	*CCDC85A*	1365.5	*VRK2*	195.7
Chr2:219032561-219097202	Deletion	64642	24	3	6	5.61 (1.37–23.01)	0.034	1 (13)	*USP37*	Exonic	In	0.0	In	0.0
Chr3:83298713-83407928	Deletion	109216	20	3	8	4.19 (1.08–16.24)	0.059	0 (0)	-	Intergenic	*GBE1*	1405.1	*CADM2*	2450.4
Chr3:129874316-129910429	Deletion	36114	10	3	7	4.80 (1.21–19.06)	0.046	0 (0)	-	Intergenic	*RPN1*	21.9	*RAB7A*	17.2
Chr5:32144879-32175495	Amplification	30617	20	5	12	4.80 (1.63–14.11)	0.011	2 (3)	*GOLPH3, PDZD2*	Exonic	In	0.0	In	0.0
Chr5:32175495-32195274	Amplification	19780	10	5	10	5.77 (1.91–17.49)	0.006	1 (1)	*GOLPH3*	Exonic	In	0.0	In	0.0
Chr6:29982364-30000415	Deletion	18052	23	23	162	1.85 (1.07–3.21)	0.032	1 (1)	*HCG4P6*	Exonic	In	0.0	In	0.0
Chr7:75959305-75977389	Amplification	18085	12	4	5	9.14 (2.39–35.02)	0.004	1 (7)	*DTX2*	Exonic	In	0.0	In	0.0
Chr7:76027543-76115070	Amplification	87528	23	4	8	5.69 (1.66–19.48)	0.014	1 (7)	*POMZP3*	Exonic	In	0.0	In	0.0
Chr7:76115070-76157984	Amplification	42915	10	4	9	5.05 (1.51–16.91)	0.019	0 (0)	-	Intergenic	*POMZP3*	20.5	*PMS2L11*	290.1
Chr7:76157984-76291345	Amplification	133362	56	4	10	4.54 (1.38–14.93)	0.025	0 (0)	-	Intergenic	*POMZP3*	63.4	*PMS2L11*	156.7
Chr7:76291345-76400451	Amplification	109107	35	4	11	4.12 (1.27–13.35)	0.032	0 (0)	-	Intergenic	*POMZP3*	196.8	*PMS2L11*	47.6
Chr22:18274542- 18287303	Deletion	12762	10	3	6	5.61 (1.37–23.01)	0.034	1 (6)	*TXNRD2*	Exonic	In	0.0	In	0.0

[1] Shaded areas indicate contiguous merged CNV regions. Genomic positions are indicated according to the NCBI36/hg18 human genome assembly.

Abbreviations: bp, base pair; CI, confidence interval; CNV, copy-number variant; F-exact, Fisher's exact test; OR, odds ratio; PE, preeclampsia.