

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Johnson, MP;Brennecke, SP;East, CE;Göring, HHH;Kent, JW;Dyer, TD;Said, JM;Roten, LT;Iversen, AC;Abraham, LJ;Heinonen, S;Kajantie, E;Kere, J;Kivinen, K;Pouta, A;Laivuori, H;Austgulen, R;Blangero, J;Moses, EK

Title:

Genome-wide association scan identifies a risk locus for preeclampsia on 2q14, near the inhibin, beta B gene

Date:

2012-03-14

Citation:

Johnson, M. P., Brennecke, S. P., East, C. E., Göring, H. H. H., Kent, J. W., Dyer, T. D., Said, J. M., Roten, L. T., Iversen, A. C., Abraham, L. J., Heinonen, S., Kajantie, E., Kere, J., Kivinen, K., Pouta, A., Laivuori, H., Austgulen, R., Blangero, J. & Moses, E. K. (2012). Genome-wide association scan identifies a risk locus for preeclampsia on 2q14, near the inhibin, beta B gene. *Plos One*, 7 (3), <https://doi.org/10.1371/journal.pone.0033666>.

Persistent Link:

<https://hdl.handle.net/11343/264528>

License:

CC BY

# Genome-Wide Association Scan Identifies a Risk Locus for Preeclampsia on 2q14, Near the Inhibin, Beta B Gene

Matthew P. Johnson<sup>1</sup>, Shaun P. Brennecke<sup>2</sup>, Christine E. East<sup>2</sup>, Harald H. H. Göring<sup>1</sup>, Jack W. Kent Jr.<sup>1</sup>, Thomas D. Dyer<sup>1</sup>, Joanne M. Said<sup>2</sup>, Linda T. Roten<sup>3</sup>, Ann-Charlotte Iversen<sup>3</sup>, Lawrence J. Abraham<sup>4</sup>, Seppo Heinonen<sup>5</sup>, Eero Kajantie<sup>6,7</sup>, Juha Kere<sup>8,9,10</sup>, Katja Kivinen<sup>11</sup>, Anneli Pouta<sup>12</sup>, Hannele Laivuori<sup>10,13</sup> for the FINNPEC Study Group, Rigmor Austgulen<sup>3</sup>, John Blangero<sup>1</sup>, Eric K. Moses<sup>1,14\*</sup>

**1** Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas, United States of America, **2** Department of Perinatal Medicine/Department of Obstetrics and Gynaecology, Royal Women's Hospital and University of Melbourne, Parkville, Victoria, Australia, **3** Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway, **4** The School of Biomedical Biomolecular and Chemical Sciences, The University of Western Australia, Perth, Western Australia, Australia, **5** Department of Obstetrics and Gynecology, Kuopio University Hospital, and University of Eastern Finland, Kuopio, Finland, **6** Department of Chronic Disease Prevention, Diabetes Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland, **7** Children's Hospital, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland, **8** Department of Biosciences and Nutrition, and Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden, **9** Folkhälsan Institute of Genetics, Helsinki, Finland, **10** Haartman Institute, Department of Medical Genetics, University of Helsinki, Helsinki, Finland, **11** Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom, **12** Department of Children, Young People and Families, National Institute for Health and Welfare, Oulu, Finland, **13** Research Programs Unit, Women's Health, University of Helsinki, Helsinki, Finland, **14** The Centre for Genetic Epidemiology and Biostatistics, The University of Western Australia, Perth, Western Australia, Australia

## Abstract

Elucidating the genetic architecture of preeclampsia is a major goal in obstetric medicine. We have performed a genome-wide association study (GWAS) for preeclampsia in unrelated Australian individuals of Caucasian ancestry using the Illumina OmniExpress-12 BeadChip to successfully genotype 648,175 SNPs in 538 preeclampsia cases and 540 normal pregnancy controls. Two SNP associations (rs7579169,  $p = 3.58 \times 10^{-7}$ , OR = 1.57; rs12711941,  $p = 4.26 \times 10^{-7}$ , OR = 1.56) satisfied our genome-wide significance threshold (modified Bonferroni  $p < 5.11 \times 10^{-7}$ ). These SNPs reside in an intergenic region less than 15 kb downstream from the 3' terminus of the Inhibin, beta B (*INHBB*) gene on 2q14.2. They are in linkage disequilibrium (LD) with each other ( $r^2 = 0.92$ ), but not ( $r^2 < 0.80$ ) with any other genotyped SNP  $\pm 250$  kb. DNA re-sequencing in and around the *INHBB* structural gene identified an additional 25 variants. Of the 21 variants that we successfully genotyped back in the case-control cohort the most significant association observed was for a third intergenic SNP (rs7576192,  $p = 1.48 \times 10^{-7}$ , OR = 1.59) in strong LD with the two significant GWAS SNPs ( $r^2 > 0.92$ ). We attempted to provide evidence of a putative regulatory role for these SNPs using bioinformatic analyses and found that they all reside within regions of low sequence conservation and/or low complexity, suggesting functional importance is low. We also explored the mRNA expression in decidua of genes  $\pm 500$  kb of *INHBB* and found a nominally significant correlation between a transcript encoded by the *EPB41L5* gene,  $\sim 250$  kb centromeric to *INHBB*, and preeclampsia ( $p = 0.03$ ). We were unable to replicate the associations shown by the significant GWAS SNPs in case-control cohorts from Norway and Finland, leading us to conclude that it is more likely that these SNPs are in LD with as yet unidentified causal variant(s).

**Citation:** Johnson MP, Brennecke SP, East CE, Göring HHH, Kent JW Jr, et al. (2012) Genome-Wide Association Scan Identifies a Risk Locus for Preeclampsia on 2q14, Near the Inhibin, Beta B Gene. PLoS ONE 7(3): e33666. doi:10.1371/journal.pone.0033666

**Editor:** Struan Frederick Airth Grant, The Children's Hospital of Philadelphia, United States of America

**Received:** November 15, 2011; **Accepted:** February 14, 2012; **Published:** March 14, 2012

**Copyright:** © 2012 Johnson et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** National Institutes of Health grants supported the Australian GWAS (HD049847 to E.K.M., S.P.B. and J.B.) and MEDUSA, the super computer cluster at Texas Biomed (S10RR029392 to J.B.). The AT&T Genomics Computing Center at Texas Biomed is supported by the AT&T Foundation. Transcriptional profiling was supported by the Faye L. and William L. Cowden Charitable Foundation (to M.P.J.). The Norwegian cohort study was supported by the Functional Genomic Programme (FUGE) of the Norwegian Research Council (to R.A.). The FINNPEC study was supported by Jane and Aatos Erkkö Foundation, Päivikki and Sakari Sohlberg Foundation, Academy of Finland, Research Funds of the University of Helsinki, Government Special Subsidiary for Health Sciences (EVO funding) at Helsinki and Uusimaa Hospital District. Novo Nordisk Foundation, Finnish Foundation for Pediatric Research, Emil Aaltonen Foundation, and Sigrid Jusélius Foundation. M.P.J. is supported, in part, by an American Heart Association National Scientist Development grant (09SDG2350008). J.M.S. was supported by a Cornelius Regan Trust Award from the University of Melbourne. This investigation was conducted in facilities constructed with support from Research Facilities Improvement Program grant RR017515 from the National Center for Research Resources, National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: eric.moses@uwa.edu.au

## Introduction

Preeclampsia is a common and serious complication of human pregnancy affecting 3–5% of all primigravid women [1–3]. Delivery of the fetus and placenta is the only intervention for adequate resolution of severe symptoms. It is a major cause of maternal mortality in developing countries, accounting for 50,000

maternal deaths yearly [4]. The maternal and fetal morbidity and mortality associated with preeclampsia and in particular with the adverse consequences of pre-term delivery are a major health burden in the developed world [5–7].

The pathophysiology of preeclampsia is thought to involve two main stages [8,9]. In stage one abnormal fetal-derived cytotrophoblast invasion in the uterine wall in early pregnancy is

associated with failed remodeling of the maternal spiral arteries perfusing the placenta. This is thought to be a 'root' cause. As a result of hypoxia and/or oxidative stress to the placenta there is release of syncytiotrophoblast-derived factors into the maternal circulation that give rise to the second stage of the maternal syndrome. The known placental factor of most relevance to this second stage is the soluble receptor for vascular endothelial growth factor, sVEGFR-1, also called sFlt-1. When present in excess, as in preeclampsia, sFlt-1 binds to, and activates, VEGF, a key survival factor for endothelium [10], and thereby induces systemic endothelial dysfunction.

The principal diagnostic features of preeclampsia are new onset hypertension and proteinuria after 20 weeks gestation [11]. The hypertension is now recognized to be secondary to diffuse endothelial dysfunction [12], and the proteinuria is associated with glomerular endotheliosis [10,13]. Preeclampsia is therefore primarily characterized by endothelial dysfunction, which is also one of the principal pathogenic mechanisms in atherosclerotic vascular diseases such as coronary artery disease and stroke. Consistent with their shared pathogenesis, atherosclerosis and preeclampsia share many common risk factors including hypertension, obesity, insulin resistance, diabetes mellitus, metabolic syndrome, general inflammation, thrombophilia, and family history [14]. A history of preeclampsia increases the risk of future hypertension, ischemic heart disease, stroke and venous thromboembolism. This is true especially for women with a history of early-onset preeclampsia (<34 weeks gestation) than those women who have preeclampsia at term [15]. A popular theory is that pregnancy provides a metabolic stress test to unmask underlying risk of cardiovascular disease [16].

These data have led several investigators to speculate [17,18] that the genetic risk factors for preeclampsia will also be relevant to cardiovascular disease, providing increased impetus and justification for their discovery [19,20]. By far the most effort to date has been focused on candidate genes, primarily those for which a plausible role in the known underlying pathophysiology could be argued, and in particular blood pressure regulation, endothelial dysfunction, lipid metabolism, thrombophilia, placental development and function, and the inflammatory response [21]. There have been many nominal associations reported with a lack of reproducibility a common theme, in many cases most likely due to a lack of uniformity in diagnosis and underpowered study designs. In our attempts to identify risk factors for preeclampsia we have primarily focused on positional cloning strategies, making no *a priori* assumptions about the nature of genes involved. We initially performed genome-wide linkage mapping studies in multiple affected families from Australia and New Zealand, identifying putative susceptibility loci on chromosomes 2q22, 5q and 13q [22,23], with several plausible positional candidate susceptibility genes identified, including the activin receptor gene *ACVR2A* on 2q22 [24,25], the aminopeptidase gene *ERAP2* on 5q [26], and the cytokine encoding *TNFSF13B* gene on 13q [27]. We now report on our continued positional cloning efforts using genome-wide association mapping in a large Caucasian case-control cohort from Australia. We herein report a significant novel SNP association on chromosome 2q14.2, close to the Inhibin, beta B (*INHBB*) gene.

## Results

### Australian GWAS

**Genome-wide association with preeclampsia.** In the set of 1,078 unrelated Australian samples (538 preeclampsia cases, 540 normal pregnancy controls) that passed our quality control criteria, the observed distribution of p-values for 648,175 successfully

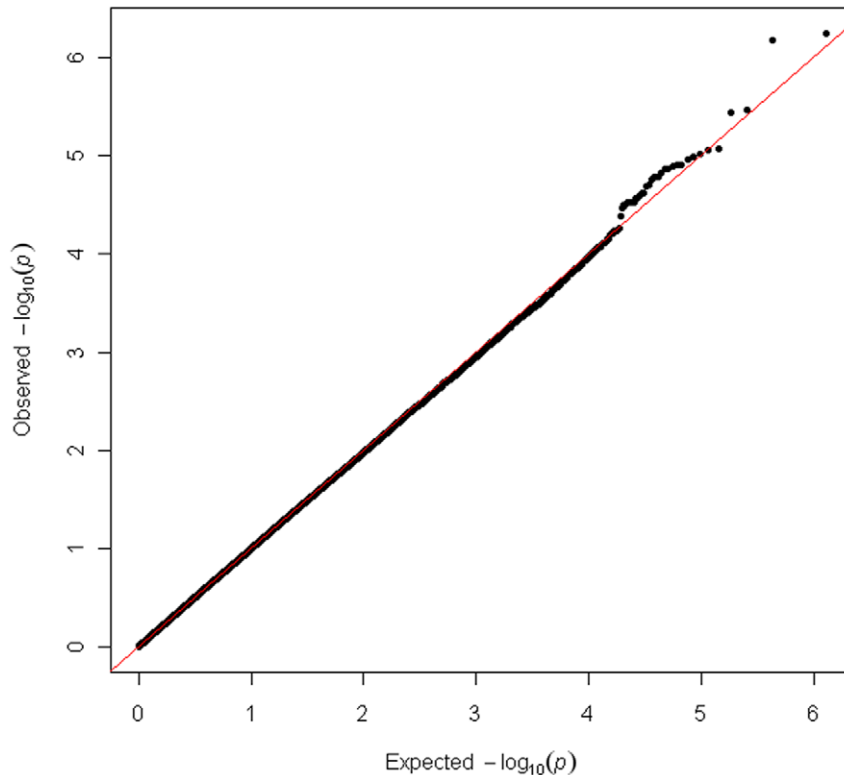
genotyped SNPs exhibited minimal deviation from the expected distribution (Figure 1). As such, this indicates minimal test statistic bias or underlying population structure ( $\lambda = 1.002$ ). The  $-\log_{10}$  transformation of observed p-values across the genome are displayed in Figure 2 and SNPs with a p-value of  $10^{-6}$  or less are presented in Table 1. By accounting for the extent of SNP linkage disequilibrium (LD), per chromosome, the number of independent SNPs ( $\text{SNP}_{\text{INDEP}}$ ) across the genome were estimated (Table S1). The estimated number of independent SNPs, specific to the Australian case-control cohort, was used in a modified Bonferroni procedure to generate an adjusted target alpha level ( $0.05/\text{SNP}_{\text{INDEP}}$ ). The two most significant SNP associations satisfied our genome-wide significance threshold (modified Bonferroni  $p < 5.11 \times 10^{-7}$ ) (Figure 2). The SNP showing the strongest association (rs7579169;  $p = 3.58 \times 10^{-7}$ , OR = 1.57,  $\text{MAF}_{(\text{cases})} = 0.447$ ,  $\text{MAF}_{(\text{controls})} = 0.340$ ) is intergenic and resides  $\sim 8.3$  kb downstream from the 3' terminus of the *INHBB* gene on chromosome 2q14.2. The next strongest association is also for an intergenic SNP (rs12711941;  $p = 4.26 \times 10^{-7}$ , OR = 1.56,  $\text{MAF}_{(\text{cases})} = 0.448$ ,  $\text{MAF}_{(\text{controls})} = 0.342$ ) downstream ( $\sim 13.5$  kb) from the 3' terminus of the *INHBB* gene. Both of these SNPs are strongly correlated with each other ( $r^2 = 0.92$ ), but not ( $r^2 < 0.80$ ) with any other genotyped SNP  $\pm 250$  kb (Figure 3). The SNP chip used in this study accommodated two SNPs within the *INHBB* gene locus itself, a nearGene-5 (rs13419301) and an intronic (rs11902591) SNP. Both of these *INHBB* locus SNPs did not reach nominal significance ( $p = 0.43$  and  $p = 0.22$ , respectively), nor were they correlated with rs7579169 ( $r^2 = 0.03$  and  $r^2 = 0.02$ , respectively). The sample genotype success rates for rs7579169 and rs12711941 were 0.9981 and 1.0, respectively. Furthermore, sample genotype concordance rates for rs7579169 and rs12711941 genotyped on both the Illumina and Sequenom platforms (see methods) were 0.987 and 0.993, respectively.

**HapMap CEU proxy SNPs.** To investigate other potential proxy SNPs to rs7579169, we used the latest (19-Apr-2009) HapMap CEU linkage disequilibrium (LD) data arising from phases I+II+III (rel #27, NCBI B36). Based on current HapMap parameters the search for SNPs flanking rs7579169 was restricted to  $\pm 200$  kb. One additional proxy SNP (rs7576192) was identified to be strongly correlated with rs7579169 in the CEU genotype data ( $r^2 = 1$ ) (Table S2). The rs7576192 SNP also resides downstream from the 3' terminus of the *INHBB* gene and is 93 bp from rs7579169 (Table S2). In the CEU samples, rs7576189 is also strongly correlated with rs12711941 ( $r^2 = 0.96$ ) and not correlated with rs13419301 ( $r^2 = 0.04$ ) or rs11902591 ( $r^2 = 0.03$ ) (Table S2). These data are concordant with our Australian GWAS cohort. In addition, a second *INHBB* nearGene-5 SNP (rs7578624) genotyped in the CEU samples was not correlated with rs7579169 ( $r^2 = 0.02$ ) (Table S2).

**SNP degree of dominance.** We investigated the mode of inheritance of our top two SNP associations by estimating the degree of dominance index (*h*) [28]. We report negligible deviation from additivity and hence, a significant additive effect for both the rs7579169 ( $h = -0.04$ ) and rs12711941 ( $h = -0.13$ ) SNPs. The application of PLINK's [29] `-model` option confirms a stronger additive effect than a dominant or recessive effect for either SNP (data not shown).

### *INHBB* locus sequencing

In an effort to identify other potentially causal variants at the *INHBB* locus we re-sequenced the entire *INHBB* structural gene (NM\_002193.2),  $\sim 2.5$  kb upstream of the *INHBB* translation start site and  $\sim 3$  kb downstream of the *INHBB* STOP codon. We also sequenced a region flanking rs7579169 ( $\sim 2.2$  kb upstream and



**Figure 1. Quantile-quantile (Q-Q) plot of the observed GWAS p-values ( $-\log_{10}(p)$ ).**  
doi:10.1371/journal.pone.0033666.g001

~0.6 kb downstream) that exhibited evolutionary conservation amongst the rhesus monkey (*Macaca mulatta*), dog (*Canis familiaris*) and mouse (*Mus musculus*). Sequencing experiments were conducted in 96 individuals from the Australian GWAS cohort (48 preeclampsia cases, 48 normal pregnancy controls). These individuals were selected on the basis of carrying two copies of the rare allele at both the rs7579169 and rs12711941 SNP loci. A total of 19 SNPs (9 known, 10 novel) plus six novel deletions were identified in our Australian cohort subset (Table 2). Due to the rare 'T' allele for rs7579169 being concordant with our reference sequence template, this SNP locus was not highlighted in our list of identified *INHBB* locus variants.

#### *INHBB* variant genotyping and association analysis

Of the 25 *INHBB* locus variants identified by re-sequencing, 21 were successfully genotyped with a mean (range) genotyping success rate of 0.975 (0.960–0.999) (Table 2). Of the remaining variants, one deletion failed assay design, two deletions failed the assay and one SNP was non-polymorphic due to a discordance between the preeclampsia data set allele and the reference template allele. We observed nominal genetic associations for a novel SNP 2,434 bp upstream of the *INHBB* translation start site (ss469271203;  $p = 0.02$ ,  $MAF_{(cases)} = 0.028$ ,  $MAF_{(controls)} = 0.013$ ) and for a rare novel SNP within *INHBB*'s 3'UTR (ss469271208;  $p = 0.01$ ,  $MAF_{(cases)} = 0.009$ ,  $MAF_{(controls)} = 0.001$ ) (Table 2). A genome-wide significant association was observed for another intergenic SNP residing downstream from the 3' terminus of the *INHBB* gene (rs7576192;  $p = 1.48 \times 10^{-7}$ ,  $OR = 1.59$ ,  $MAF_{(cases)} = 0.449$ ,  $MAF_{(controls)} = 0.339$ ) (Table 2). SNP rs7576192 is in close proximity to, and strongly correlated with the two significant GWAS SNPs ( $r^2 > 0.92$ ) (Table S3). The genotypic correlation data between rs7576192, rs7579169 and rs12711941 in

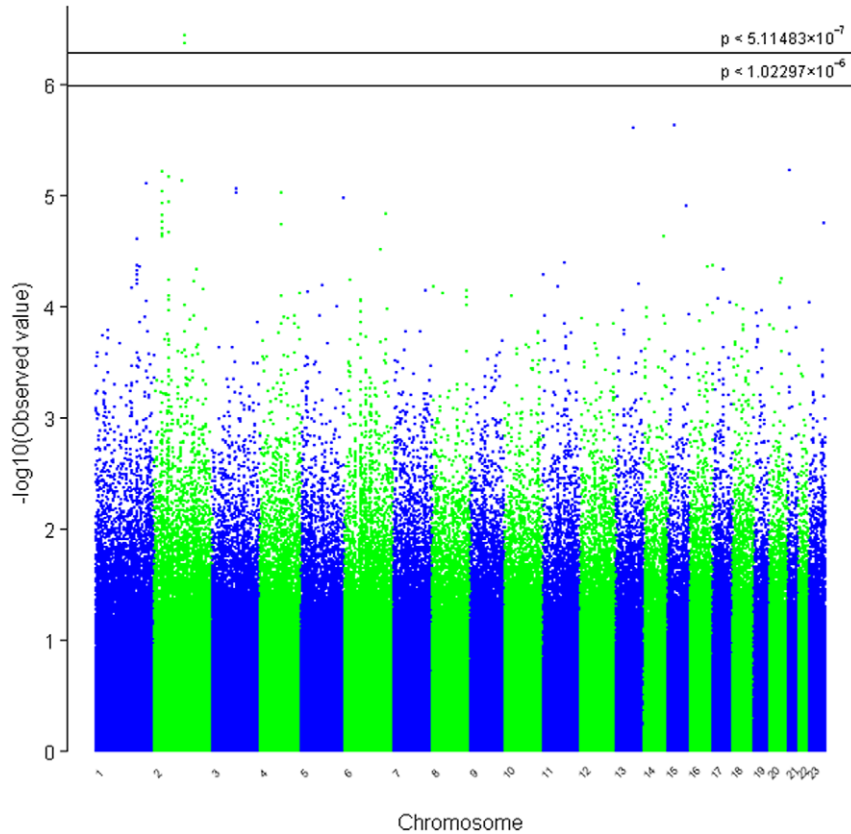
our Australian cohort is concordant with the reported HapMap CEU data.

#### Analysis of gene expression at the *INHBB* locus

To investigate whether expression of the *INHBB* locus was correlated with preeclampsia or the significant GWAS SNPs exhibited regulatory potential, total RNA from decidual basal tissue of 60 individuals from the Australian case-control cohort (25 preeclampsia cases and 35 normal pregnancy controls) were hybridized onto Illumina's HumanHT-12 v4 Expression Bead-Chips. Our analysis of differential mRNA expression at the *INHBB* locus (ILMN\_1685714) was extended  $\pm 500$  kb of *INHBB*, also bringing into consideration several other genes, including *PTPN4* (ILMN\_1793549), *EPB41L5* (ILMN\_1770245, ILMN\_2043306), *TMEM185B* (ILMN\_2231020, ILMN\_2231021), *RALB* (ILMN\_1676358) and *GLI2* (ILMN\_1727577). One transcript (ILMN\_2043306) was not significantly detected (FDR  $p > 0.05$ ) and an additional transcript (ILMN\_1727577) had a mean expression level consistent with background noise (mean average raw signal  $< 50$ ) (Table 3). We observed a nominally significant correlation with the ILMN\_1770245 (*EPB41L5*) transcript and preeclampsia ( $p = 0.03$ ) (Table 3). The independent addition of our GWAS SNPs (rs7579169, rs12711941) yielded neither as significant predictor variables in the 'transcript~preeclampsia' regression model ( $p > 0.05$ ) (Table 3).

#### Bioinformatic analysis of associated SNPs

Using the UCSC genome browser (Human, Feb. 2009 [GRCh37/hg19]) we conducted bioinformatic analyses on the two significantly associated GWAS SNPs (rs7579169, rs12711941) and the rs7576192 SNP identified by re-sequencing to see if they resided within (1) regulatory elements (histone mark H3K4Me1 or



**Figure 2. The genome-wide distribution of asymptotic p-values for each of the quality control filtered SNPs in the Australian cohort (n = 648,175).** Our adjusted genome-wide significant and suggestive thresholds were set at  $p < 5.11483 \times 10^{-7}$  and  $p < 1.02297 \times 10^{-6}$ , respectively. doi:10.1371/journal.pone.0033666.g002

**Table 1. SNP associations with preeclampsia ( $p \leq 10^{-6}$ ).**

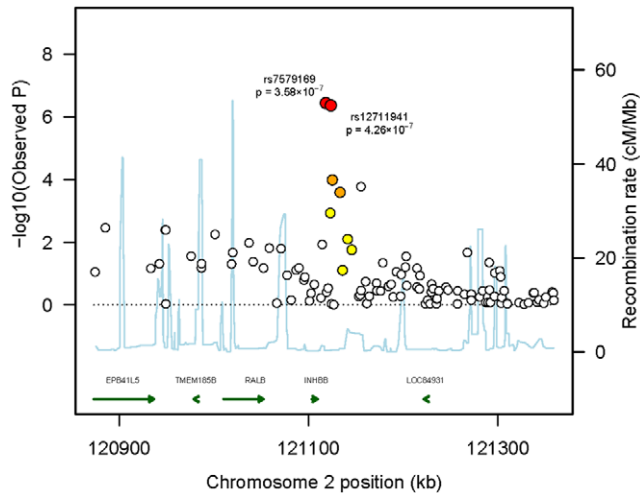
Chr	SNP	bp <sup>1</sup>	Function	Gene	Alleles <sup>2</sup>	MAF <sub>(cases)</sub>	MAF <sub>(controls)</sub>	P-value	OR <sup>3</sup> (95% CI)
2	rs7579169	121118124	intergenic		C/T	0.4471	0.3401	$3.58 \times 10^{-7}$	1.57 (1.32–1.87)
2	rs12711941	121123383	intergenic		G/T	0.4482	0.3420	$4.26 \times 10^{-7}$	1.56 (1.31–1.86)
15	rs2453274	44774202	intronic	<i>CTDSPL2</i>	C/T	0.0462	0.0989	$2.32 \times 10^{-6}$	0.44 (0.31–0.62)
13	rs12431203	77488736	intergenic		G/A	0.2061	0.1303	$2.45 \times 10^{-6}$	1.73 (1.38–2.18)
21	rs2826538	22188735	intergenic		T/C	0.1954	0.2782	$5.93 \times 10^{-6}$	0.63 (0.52–0.77)
2	rs9332419	26322040	intronic	<i>RAB10</i>	G/A	0.3562	0.4518	$6.17 \times 10^{-6}$	0.67 (0.56–0.80)
2	rs4952830	46746963	UTR-5	<i>ATP6V1E2</i>	A/G	0.3343	0.4284	$6.74 \times 10^{-6}$	0.67 (0.56–0.80)
2	rs6542736	108415702	intergenic		G/A	0.4094	0.3167	$7.31 \times 10^{-6}$	1.50 (1.25–1.79)
1	rs6660579	224916592	intronic	<i>CNIH3</i>	C/T	0.1558	0.0924	$7.80 \times 10^{-6}$	1.81 (1.39–2.36)
3	rs2279720	87276699	synonymous	<i>CHMP2B</i>	G/A	0.0592	0.1128	$8.66 \times 10^{-6}$	0.49 (0.36–0.68)
3	rs1044499	87299075	synonymous	<i>CHMP2B</i>	A/C	0.0592	0.1128	$8.66 \times 10^{-6}$	0.49 (0.36–0.68)
2	rs11126375	26344787	intronic	<i>RAB10</i>	A/G	0.1645	0.2412	$9.13 \times 10^{-6}$	0.62 (0.50–0.77)
4	rs7677523	98965696	intronic	<i>C4orf37</i>	T/C	0.0833	0.0379	$9.51 \times 10^{-6}$	2.31 (1.58–3.37)
3	rs17024019	87257423	intergenic		A/G	0.0601	0.1137	$9.57 \times 10^{-6}$	0.50 (0.36–0.68)

<sup>1</sup>Physical coordinate based on NCBI reference assembly build 37.2.

<sup>2</sup>Major allele/minor allele.

<sup>3</sup>Odds ratio for the minor allele.

doi:10.1371/journal.pone.0033666.t001



**Figure 3. Association plot of the chromosome 2 region reaching genome-wide significance with preeclampsia susceptibility (rs7579169±250 kb).** Observed p-values are plotted as  $-\log_{10}$  values as a function of the SNPs physical location (NCBI Build 37.2). Estimated recombination rates were extracted from HapMap data. The local linkage disequilibrium structure is based on the observed allele frequency data in the Australian cohort (red dots,  $r^2 \geq 0.8$ ; orange dots,  $0.5 \leq r^2 < 0.8$ ; yellow dots,  $0.2 \leq r^2 < 0.5$ ; clear dots,  $r^2 < 0.2$ ). Gene annotations were obtained from the UCSC genome browser (Human, Feb. 2009 [GRCh37/hg19]). doi:10.1371/journal.pone.0033666.g003

DNase I hypersensitive sites from ENCODE) or (2) transcription factor (TF) binding sites (ChIP-seq data from ENCODE). Additional TF binding site analysis was performed using P-Match [30,31] and AliBaba 2.1 [30,32]. Histone marks in the regions of rs7579169 and rs7576192 suggest some promoter/enhancer activity to be present, and highest in human umbilical vein endothelial cell (HUVEC) lines. AliBaba indicated a Sp1 (stimulating protein 1) TF binding site in the presence of the minor ‘T’ allele for rs7579169. This would suggest the minor allele for rs7579169 to be affiliated with higher transcriptional activity/expression. Conversely, AliBaba indicated a Sp1 TF binding site in the presence of the major ‘G’ allele for rs12711941. This would suggest the minor allele for rs12711941 to be affiliated with lower transcriptional activity. No TF binding sites were identified in the presence of the major or minor allele for rs7576192. The rs7579169, rs12711941 and rs7576192 SNPs all reside within regions of low sequence conservation and/or low complexity, suggesting functional importance is low. It is therefore more likely that these SNPs are in LD with an as yet unidentified polymorphism of greater functional significance.

### Norwegian and Finnish replication cohorts

The two GWAS SNPs showing significant association with preeclampsia in the Australian cohort were genotyped in two independent case-control cohorts from Norway (1,134 preeclampsia cases, 2,263 normal pregnancy controls) and Finland (760 preeclampsia cases, 759 normal pregnancy controls). The rs7579169 SNP association for the minor ‘T’ allele was not replicated in the Norwegian ( $p = 0.29$ ,  $MAF_{(cases)} = 0.424$ ,  $MAF_{(controls)} = 0.438$ ) or Finnish ( $p = 0.60$ ,  $MAF_{(cases)} = 0.391$ ,  $MAF_{(controls)} = 0.382$ ) cohorts (Table S4). Similarly, the rs12711941 SNP association for the minor ‘T’ allele was not replicated in the Norwegian ( $p = 0.35$ ,  $MAF_{(cases)} = 0.422$ ,  $MAF_{(controls)} = 0.434$ ) or Finnish ( $p = 0.50$ ,  $MAF_{(cases)} = 0.387$ ,  $MAF_{(controls)} = 0.375$ ) cohorts (Table S4).

### Discussion

The determination of the genetic contributions to risk of preeclampsia has proven difficult. In this first GWAS for preeclampsia we have obtained strong evidence for a risk locus on chromosome 2q14.2 defined by significant genetic association with two intergenic SNPs located within 15 kb of the 3’ terminus of the Inhibin, beta B (*INHBB*) gene. Our subsequent resequencing of the *INHBB* locus in a small sample of affected and unaffected individuals from our Australian cohort identified a third intergenic SNP, also residing within 15 kb from the *INHBB* 3’ terminus, to be significantly associated with preeclampsia. While all three intergenic SNPs are in strong LD with each other they are not in LD with any other genotyped SNP within  $\pm 250$  kb.

Our preliminary bioinformatic and transcriptional profiling analyses have not provided compelling data to implicate these SNP variants and/or genes in preeclampsia etiology, and we did not replicate these significant SNP associations in either Norwegian or Finnish case-control cohorts. While successful replication can provide an important and independent verification of a putative genetic association, which helps to prevent the discovery of spurious associations, failure to replicate in a population different from that used in the initial finding does not necessarily invalidate the original observation. The reasons why true associations may not replicate across independent data sets has received considerable attention over the last five years with genetic heterogeneity, environmental interactions, age-dependent effects, epistasis and inadequate statistical power given as possible reasons [33–36]. In this context it is perhaps noteworthy that in our earlier linkage-based positional cloning studies in Australian families where we reported the likely involvement of the activin type 2A receptor (*ACVR2A*) gene [24,37] and the endoplasmic reticulum aminopeptidase 2 (*ERAP2*) gene [26] in risk of preeclampsia, we were also unable to replicate our gene-specific SNP associations in the same Norwegian case-control cohort as that used in this current study. In the case of *ACVR2A* and *ERAP2* we subsequently were able to demonstrate association with preeclampsia in the Norwegian population using other SNPs in these genes, providing evidence of different allele frequencies and LD patterns at these loci [25,26]. These data may be consistent with the existence of as yet unidentified/untyped rare risk variants that exhibit different patterns of linkage disequilibrium in our Australian, Norwegian and Finnish population samples.

While we have not presented compelling functional data to implicate any genes at the 2q14.2 locus marked by our SNP associations, we are encouraged by the striking plausibility of the *INHBB* gene as a positional candidate susceptibility gene for preeclampsia. This is supported by a body of substantive biological data that is consistent with the involvement of the activins, inhibins and other members of the TGF- $\beta$  superfamily in the development of preeclampsia [38–45]. It is worth noting that, during pregnancy activins and inhibins are produced in the human endometrium, decidua and placenta and are thought to inactivate matrix metalloproteases in human endometrial stromal cells during decidualization thereby affecting remodeling of the maternal spiral arteries by the invading cytotrophoblasts [46]. Failed remodeling of these vessels is regarded as an early defining event in the pathophysiology of preeclampsia [47–49]. The fact that *INHBB* is biologically connected to *ACVR2A* leads us to speculate that our positional cloning studies in the Australian Caucasian population, originally using linkage mapping in families and now GWAS in unrelated individuals, have revealed positional candidate genes that define a key pathway involved in susceptibility to preeclampsia. We now propose to focus our efforts on the

**Table 2.** *INHBB* locus variants identified in a sample of the Australian cohort (n = 96).

Variant	bp	Function	Alleles <sup>1</sup>	MAF <sub>(cases)</sub>	MAF <sub>(controls)</sub>	HWEp <sup>2</sup>	P-value
ss469271203	121101331	nearGene-5	G/A	0.0277	0.0133	$1.26 \times 10^{-9}$	0.0201
ss469271213	121101394	nearGene-5	AAG/-	0.0009	0.0019	1.0	1.0*
ss469271214	121101650	nearGene-5	AGCTGG/-	Failed assay			
ss469271215	121101736	nearGene-5	CGCCGCAGCGCC/-	Failed assay			
rs7578624	121102479	nearGene-5	C/G	0.0487	0.0427	1.0	0.5080
rs13419301	121102572	nearGene-5	T/C	0.0580	0.0655	0.28	0.4803
ss469271216	121104688	intronic	TG/-	Failed assay design			
ss469271204	121105292	intronic	C/G	0.0066	0.0038	1.0	0.5478*
rs11902591	121106003	intronic	A/G	0.0673	0.0731	0.81	0.6068
rs4328642	121106850	synonymous	C/T	0.0328	0.0363	1.0	0.6662
ss469271205	121106946	synonymous	G/A	0.0009	0.0	1.0	0.4929*
ss469271217	121107784	UTR-3	AGTC/-	0.0009	0.0	1.0	0.4962*
ss469271206	121107831	UTR-3	G/A	0.0009	0.0	1.0	0.4929*
rs45624437	121108182	UTR-3	T/C	0.0162	0.0083	1.0	0.0987
ss469271207	121108506	UTR-3	G/A	0.0	0.0009	1.0	1.0*
ss469271208	121108585	UTR-3	T/C	0.0087	0.0009	0.02	0.0111*
rs57802235	121109444	nearGene-3	G/A	0.0416	0.0466	0.72	0.5788
rs7568413	121109612	nearGene-3	C/T	Non-polymorphic <sup>3</sup>			
ss469271209	121109737	nearGene-3	C/T	0.0048	0.0019	1.00	0.2813*
rs10183524	121109878	nearGene-3	G/A	0.0491	0.0493	0.74	0.9829
ss469271210	121110151	nearGene-3	G/A	0.0039	0.0	1.0	0.0602*
ss469271218	121116483	intergenic	A/-	0.0028	0.0009	1.0	0.6247*
ss469271211	121116625	intergenic	G/A	0.0009	0.0	1.0	1.0*
ss469271212	121116672	intergenic	C/T	0.0029	0.0009	1.0	0.3709*
rs7576192	121118031	intergenic	G/A	0.4499	0.3392	0.13	$1.48 \times 10^{-7}$

Identified variants were genotyped in Australian individuals passing GWAS quality control criteria (n = 1,078). Novel variants submitted to dbSNP are assigned with their 'ss' submission ID number.

\*Fisher's exact test p-value.

<sup>1</sup>Major allele/Minor allele.

<sup>2</sup>Hardy-Weinberg equilibrium p-value.

<sup>3</sup>Preeclampsia dataset allele discordant to reference template allele.

doi:10.1371/journal.pone.0033666.t002

**Table 3.** Gene expression results of the *INHBB* structural locus  $\pm 500$  kb.

Gene	Transcript	n <sup>1</sup>	p <sup>2</sup>	FDR p <sup>3</sup>	MARS <sup>4</sup>	PE <sup>5</sup>	rs7579169 <sup>6</sup>	rs12711941 <sup>7</sup>
<i>PTPN4</i>	ILMN_1793549	59	$1.7 \times 10^{-77}$	$2.5 \times 10^{-76}$	102.8	0.99	0.37	0.53
<i>EPB41L5</i>	ILMN_1770245	58	$1.9 \times 10^{-74}$	$1.1 \times 10^{-73}$	89.6	0.03	0.90	0.57
	ILMN_2043306	3	0.57	0.91	1.0	NA <sup>8</sup>	NA <sup>8</sup>	NA <sup>8</sup>
<i>TMEM185B</i>	ILMN_2231020	59	$1.7 \times 10^{-77}$	$2.5 \times 10^{-76}$	278.6	0.48	0.25	0.19
	ILMN_2231021	58	$1.9 \times 10^{-74}$	$1.1 \times 10^{-73}$	141.3	0.29	0.22	0.06
<i>RALB</i>	ILMN_1676358	58	$1.9 \times 10^{-74}$	$1.1 \times 10^{-73}$	1,624.2	0.69	0.60	0.84
<i>INHBB</i>	ILMN_1685714	57	$1.1 \times 10^{-71}$	$5.0 \times 10^{-71}$	136.3	0.64	0.17	0.08
<i>GLI2</i>	ILMN_1727577	11	0.00015	0.00031	8.9	NA <sup>8</sup>	NA <sup>8</sup>	NA <sup>8</sup>

<sup>1</sup>Number of samples with a GenomeStudio 'pDetection' p-value  $\leq 0.05$ .

<sup>2</sup>Computed transcript detection p-value.

<sup>3</sup>False discovery rate detection p-value.

<sup>4</sup>Mean average raw signal.

<sup>5</sup>transcript~preeclampsia' regression p-value.

<sup>6</sup>transcript~preeclampsia+rs7579169' regression p-value.

<sup>7</sup>transcript~preeclampsia+rs12711941' regression p-value.

<sup>8</sup>No regression analyses performed.

doi:10.1371/journal.pone.0033666.t003

identification of probable rare and as yet unidentified variants in the inhibins, activins and their receptors as such variation is likely to be critical to the development of preeclampsia in many populations.

## Materials and Methods

### Ethics Statement

**Australian GWAS cohort.** Ethical approval for the recruitment of the Australian women was granted by the RWH Research and Ethics Committees, Melbourne, Australia. Written informed consent was obtained from study participants prior to them being phlebotomized. Permission was also granted from the Australian case-control cohort women to access and examine their medical records in order to confirm/validate Caucasian ancestry and relevant preeclampsia diagnostic criteria. Ethical approval to conduct molecular and statistical analyses of the Australian samples was obtained from the Institutional Review Board (IRB) of the University of Texas Health Science Center at San Antonio (UTHSCSA).

**Norwegian replication cohort.** All HUNT participants provided written informed consent when recruited to the study. Prior approval to use the Norwegian case-control cohort for genetic studies was obtained by the Regional Committee for Medical Research Ethics, Norway and approved by the National Data Inspectorate and The Directorate of Health and Social Welfare. Ethical approval for the molecular and statistical analysis of the Norwegian samples was obtained from the IRB of the UTHSCSA.

**Finnish replication cohort.** All subjects provided a written informed consent. The FINNPEC study protocol was approved by the coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa. The Southern Finnish participant study was approved by the local ethical review committee at the Helsinki University Hospital. Ethical approval for the molecular and statistical analysis of the Finnish samples was in addition obtained from the IRB of the UTHSCSA.

### GWAS case-control sample population

The Australian case-control cohort of 1,092 unrelated women used in this GWAS included 1,018 women of confirmed Caucasian ancestry (471 preeclampsia cases and 547 normal pregnancy controls) retrospectively ascertained from a larger Australian case-control cohort of 1,774 women that were recruited at the Royal Women's Hospital (RWH), Melbourne, Australia over a five period from 2007 to 2011. The Australian population seen at the RWH in Melbourne is ~70% Caucasian and for this study the focus was on the recruitment of Caucasian subjects. The additional 74 women were unrelated preeclampsia cases from our Caucasian Australian and New Zealand family cohort that has been described in detail elsewhere [22–24,26].

### Replication case-control sample populations

The most promising SNPs from the Australian GWAS were assessed in two independent case-control cohorts from Norway and Finland.

**Norway.** All Norwegian samples were retrospectively selected from a large multipurpose health survey conducted over a three period from 1995 to 1997 in Nord-Trøndelag County in Norway [50]. More than 65,000 inhabitants participated. The people living in the Nord-Trøndelag County are considered to be representative of the Norwegian population, and are well suited for genetic studies because of their ethnic homogeneity (<3% non-Caucasians) [50,51]. Information pertaining to all pregnancies

and deliveries has been registered in the Medical Birth Registry of Norway (MBRN) since 1968. The MBRN has established formal classifications of different diseases in pregnancy. The unique 11-digit national identification numbers from HUNT2 women participants were cross referenced with the information registries of the MBRN to identify case-control cohorts. The HUNT study population used to study preeclampsia has been described in detail elsewhere [25,52].

**Finland.** The Finnish patient samples used in this study originate from the Finnish Genetics of Preeclampsia Consortium (FINNPEC) study cohort and the Southern Finland preeclampsia study cohort. FINNPEC is an ongoing multicentre study where DNA samples and data have been collected prospectively at all university hospitals in Finland (i.e. Helsinki, Turku, Tampere, Kuopio and Oulu) from 2008. For each woman with preeclampsia, the next available woman giving birth at the same hospital, with no preeclampsia, is invited as a control. After initial review of hospital records by a research nurse, each diagnosis is confirmed by a study physician based on criteria described below. Information pertaining to the Southern Finnish case-control cohort was obtained from discharge records from the Helsinki University Central Hospital. These records were used to retrospectively identify women with preeclampsia between January 1988 and April 1998 [53]. These women were healthy prior to their first pregnancy with no evidence of renal or autoimmune disease. Blood samples were collected between January 1997 and April 1998 after the index pregnancy [53,54]. During the same period, blood samples from non-preeclamptic (control) patients who had given birth in the same hospital were also collected.

### Preeclampsia diagnosis

**Australian GWAS cohort.** Preeclampsia diagnosis was determined by qualified clinicians using criteria set by the Australasian Society for the Study of Hypertension in Pregnancy [55,56], and the Society of Obstetric Medicine of Australia and New Zealand for the management of hypertensive diseases of pregnancy [57]. Women were considered preeclamptic if they were previously normotensive and if they, on at least two occasions six or more hours apart, had after 20 weeks gestation (i) a rise in systolic blood pressure (SBP) of at least 25 mmHg and/or a rise from baseline diastolic blood pressure (DBP) of at least 15 mmHg, or (ii) SBP $\geq$ 140 mmHg and/or DBP $\geq$ 90 mmHg. Additionally, significant new onset proteinuric levels were either  $\geq$ 0.3 g/l in a 24 hour specimen, at least a '2+' proteinuria dipstick reading from a random urine collection or a spot protein:creatinine ratio  $\geq$ 0.03 g/mmol. Preeclamptic women who also experienced convulsions or unconsciousness in their perinatal period were classified as having eclampsia. Women with pre-existing hypertension or other medical conditions known to predispose for preeclampsia (e.g. renal disease, diabetes, twin pregnancies or fetal chromosomal abnormalities) were excluded. Of the 1,774 unrelated Australian women initially recruited for this study, 1,018 women were of confirmed Caucasian ancestry, meeting our inclusion criteria. Of these, 471 were confirmed, by medical records, as having preeclampsia (cases) and 547 were confirmed as having a normal pregnancy (controls). An additional 74 unrelated preeclamptic (case) women selected for inclusion in our GWAS sample were the probands and/or founders of our previously described 74 preeclampsia families [22–24,26,27,37,58].

**Norwegian replication cohort.** The definition and classification of preeclampsia used for the Norwegian samples was established by the MBRN based on previously reported guidelines [11]. The MBRN definition for preeclampsia was

defined as an increase in SBP to at least 140/90 mmHg (or an increase in SBP  $\geq$  30 mmHg, or in DBP  $\geq$  15 mmHg from the level measured before the 20<sup>th</sup> week of gestation), combined with proteinuria (protein excretion of at least 0.3 g per 24 hours or  $\geq$  1+ on a dip stick). Based on these diagnostic criteria there were 1,179 women registered with preeclampsia (cases) and 2,358 women with a history of a normal, healthy pregnancy (controls). Of these registered women, blood samples were available for 1,134 cases and 2,263 controls at the HUNT Biobank and included for this study.

**Finnish replication cohort.** Finnish women who suffered a preeclamptic pregnancy and had no medical history of chronic hypertension, type 1 diabetes, or renal disease were eligible for the study as cases. Diagnostic criteria used for the FINNPEC study cohort were SBP  $\geq$  140 mmHg and/or DBP  $\geq$  90 mmHg on at least two occasions with new onset proteinuria ( $\geq$  0.3 g/24 hrs, or  $\geq$  0.3 g/L, or in the absence of concurrent quantitative measurement, at least a ‘2+’ or more, or two ‘1+’ proteinuria dipstick readings) after 20 weeks gestation in a previously normotensive woman. Preeclampsia in the Southern Finnish case/control cohort was defined as two SBP/DBP measurements at least 6 hrs apart  $\geq$  140/90 mmHg and proteinuria measurement  $\geq$  0.3 g in a 24 hour urine collection, or at least a ‘1+’ dipstick reading after 20 weeks gestation [53]. A total of 760 preeclamptic (case) women and 664 control women from the FINNPEC study cohort, and 95 control women from the Southern Finland preeclampsia study cohort were included in this study.

## GWAS genotyping

The isolation of genomic DNA (gDNA) from the Australian case-control blood samples was achieved using Qiagen’s Blood & Cell Culture DNA Midi Kit (Qiagen Pty Ltd, Doncaster, VIC, Australia). The individual gDNA samples ( $n = 1,092$ ) were genotyped using Illumina’s Human OmniExpress-12 BeadChip (Illumina Inc., San Diego, CA) containing 731,442 loci derived from phases I, II and III of the International HapMap project [59–61]. A total of 200 ng of gDNA (4  $\mu$ l at 50 ng/ $\mu$ l) for each sample was processed according to Illumina’s Infinium HD Assay Ultra protocol. BeadChips were imaged on Illumina’s iScan System with iScan Control Software (v3.2.45). Normalization of raw image intensity data, genotype clustering and individual sample genotype calls were performed using Illumina’s GenomeStudio software (v2010.2), Genotyping Module (v1.7.4). Illumina’s pre-defined genotype cluster boundaries were used to denote SNP genotype cluster positions (HumanOmniExpress-12v1\_C.egt). Additionally, genotype clusters for all SNPs of interest were visually inspected (Figures S1 & S2). Genotype assay quality control measures were assessed with Illumina’s internal assay performance metrics. Individual SNP loci and individual sample quality control performance measures were assessed using PLINK [29]. Individual SNP loci were excluded, (i) if genotype success rates were  $< 0.95$  ( $n = 4,742$ ); (ii) for deviation from Hardy-Weinberg equilibrium in the control samples with a criterion of  $p < 0.0001$  ( $n = 1,676$ ); (iii) if the observed copies of the minor allele in the population sample (i.e. cases and controls, collectively) was  $< 10$  ( $n = 77,286$ ). This quality control metric equates to a minor allele frequency (MAF) being less than 0.009 (10/1,092); (iv) for any residual non-autosomal or X-linked loci ( $n = 380$  XY-linked loci). Given our female only data set, X-linked loci were retained in our analyses. Individual samples were excluded, (i) if genotype call rates were  $< 0.9$  (1 case, 1 control); (ii) if PLINK’s sex check to estimate X chromosome inbreeding (homozygosity) rates ( $F$ ) was  $\geq 0.2$  (3 cases, 1 control). For this quality control metric a female call is made if  $F < 0.2$  and was conducted to identify probable

random genotype error(s); (iii) using PLINK’s cryptic relatedness metric to examine the possibility of unknown, distant familial relationships amongst the Australian GWAS sample set by estimating the proportion of alleles shared identical by descent ( $\hat{\pi}$ ). Eight pairs of DNA samples putatively exhibited a distant familial relationship ( $\hat{\pi} \geq 0.125$ ), of which 3 cases and 5 controls were excluded from subsequent data analyses. These SNP loci and sample quality control metric thresholds resulted in the passing of 648,175 SNPs to be analyzed in 1,078 unrelated Australian women (538 preeclampsia cases, 540 normal pregnancy controls). The mean (range) genotyping success rate of the quality control filtered data set was 0.9986 (0.9499–1).

## DNA sequencing

Gene-centric and/or conserved intergenic regions flanking prioritized SNPs were sequenced in 96 unrelated Australian samples (48 preeclampsia cases, 48 normal pregnancy controls). These samples were a subset of the final GWAS sample set ( $n = 1,078$ ) that passed our quality control cleaning. Conserved intergenic regions were identified using the ECR Browser [62]. Genomic DNA sequence reference templates were obtained from the UCSC Genome Bioinformatics database (Human, Feb. 2009 [GRCh37/hg19]). All primers were designed using Primer 3 (v0.4.0) and BLASTed to assess their uniqueness to the human genome. Contiguous primer pairs were designed to overlap by  $\sim 100$ –150 bp. Standard PCR was performed with 20 ng of gDNA in a 10  $\mu$ l reaction volume. If standard PCR optimization conditions failed, FailSafe PCR pre-mixes (Epicentre Biotechnologies, Madison, WI) were used in lieu. GeneAmp 9700 thermal cyclers (Life Technologies, Foster City, CA) were used for PCR amplification. PCR amplicons were purified with ExoSAP-IT (USB Corp., Cleveland, OH) according to manufacturers’ instructions. Independent sequencing reactions for both the sense and anti-sense strands were performed on the purified PCR amplicons (1  $\mu$ l) using AB BigDye Terminator v3.1 chemistry (Life Technologies) in a 5  $\mu$ l reaction volume. Sequence reaction amplification was performed on a GeneAmp 9700 thermal cycler using standard cycling conditions. Amplified sequence products were purified with AB BigDye XTerminator purification kits according to manufacturers’ instructions (Life Technologies). Purified sequence reactions were electrophoretically separated on an AB 3730xl DNA Analyzer (Life Technologies). Sequence reaction quality was assessed using Sequencing Analysis software v5.1.1 and sequence variant identification was performed using SeqScape v2.6 (Life Technologies).

## Replication and targeted loci genotyping

Additional genotyping in the Australian cohort and replication genotyping in the Norwegian and Finnish cohorts was performed using Sequenom-based MassArray technology (Sequenom, San Diego, CA). SNP assays were designed using Sequenom’s online design tools in conjunction with Assay Designer v4.0. Variant specific PCR and single-base extension primers were supplied by Integrated DNA Technologies (IDT, Coralville, IA). For each sample, 20 ng of gDNA was used and assayed in accordance with the iPLEX Gold Reaction protocol using the MassARRAY Matrix Liquid Handler. Samples were spotted onto a 384-sample SpectroCHIP II using the MassARRAY Nanodispenser RS1000. SpectroCHIPs were loaded into the MassARRAY Analyzer 4 and the nucleotide mass time-of-flight was recorded using SpectroACQUIRE software (v4.0.2.52). Genotype clustering and individual sample genotype calls were generated using Sequenom’s TyperAnalyzer (v4.0.5). To assess the accuracy of the GWAS

genotypes we re-genotyped our prioritized SNPs back in the Australian GWAS cohort.

### Transcriptional profiling in decidua

Of the 1,078 unrelated Australian women that passed our GWAS quality control decidua basalis tissue was also available from 25 preeclampsia cases and 35 healthy pregnancy controls. These decidua samples were collected at the time of delivery by Caesarean section, from the placental bed by suction curettage, as previously described [63]. Total RNA isolation and quality assessment, and anti-sense RNA (aRNA) synthesis, amplification and purification were performed as previously described [63]. Purified aRNA was hybridized to Illumina's HumanHT-12 v4 Expression BeadChips in accordance with Illumina's Whole-Genome Gene Expression Direct Hybridization assay protocol. All samples were scanned on the Illumina iScan System with iScan Control software (v3.2.45). Illumina's GenomeStudio software (v2010.2), Gene Expression Module (v1.7.0) was used to generate a control summary report to assess assay performance and quality control metrics. One control sample failed the image scan and was subsequently omitted prior to data analysis. The remaining 59 tissue samples yielded high quality expression profile data, without any samples showing a marked reduction in the number of probes detected, in mean average raw signal, or in mean correlation (in raw expression level across probes) with the other samples.

### Data analyses

**Population structure.** To account for potential population structure within the Australian GWAS samples passing quality control ( $n = 1,078$ ), principal components analysis (PCA) was conducted in R (`prcomp`) using a subset of quality control filtered SNPs ( $n = 246,406$ ). The subset of common SNPs ( $MAF \geq 0.05$ ) for PCA was generated using PLINK to compute the genotypic correlation ( $r^2$ ) between SNP pairs within a 50 SNP window. Each SNP window progressed forward by 5 SNPs prior to re-computing pairwise genotypic correlations. One SNP from a pair of SNPs was excluded if  $r^2 > 0.5$ . PCA revealed very minimal population structure in the Australian GWAS samples, so principle components correction was not used in the association analysis. The absence of false positive association due to population structure was confirmed by the calculated genomic inflation factor ( $\lambda$ ) of 1.002.

**Genome-wide data analysis.** Due to minimal population structure, asymptotic p-values for each of the quality control filtered SNPs ( $n = 648,175$ ) were computed to assess minor-allele association with the disease trait (i.e. preeclampsia) using PLINK. The Manhattan plot displaying the  $-\log_{10}$  transformation of observed p-values was generated using the `mhtplot` function of the R package 'gap'. The Q-Q plot depicting  $-\log_{10}$  transformations of observed p-values as a function of expected p-values was generated using R base graphics. The `asplot` function of the R package 'gap' was used to generate a regional association plot for loci of interest ( $\pm 250$  kb) based on, recombination rate (HapMap 2006-10\_rel21\_phaseI+II), PLINK computed pairwise genotypic correlations between all genotyped SNPs in the Australian samples (`-ld-window-r2 0`) and PLINK generated point-wise, asymptotic association test p-values.

**Genome-wide multiple testing correction.** To attain adjusted genome-wide significant and suggestive thresholds we first imputed sporadic missing genotype data using BEAGLE [64]. An effective number of independent SNP tests across our GWAS data set were approximated using the solid spine of linkage disequilibrium (SSLD) measure implemented in HAPLOVIEW [65], as previously described [66]. An approximated effective

number of independent SNP tests were used to calculate modified Bonferroni-adjusted significant and suggestive thresholds. Briefly, juxtaposed chromosome specific SNP windows containing at most 3,000 SNPs were first generated using PLINK. Using HAPLOVIEW, the number of SNP blocks and interblock SNPs were determined with a minimum  $D'$  value of 0.8. Pairwise comparisons of SNPs more than 500 kb apart were ignored. Quality control filtered SNPs that did not satisfy HAPLOVIEW's SSLD default parameters (i.e.  $MAF < 0.01$ ;  $HWE p < 0.001$ ), or were not assigned a chromosomal bp coordinate with the Illumina SNP chip annotation were treated as independent SNPs akin to the interblock SNPs. These additional independent SNPs are herein referred to as 'residual SNPs'. For each chromosome the sum of SNP blocks, interblock SNPs and residual SNPs approximate the effective number of independent SNP tests. The estimated number of independent SNPs ( $SNP_{INDEP}$ ), specific to the Australian case-control cohort, was used to generate an adjusted target alpha level ( $0.05/SNP_{INDEP}$ ). For this study, the adjusted genome-wide significant and suggestive thresholds were set at  $5.11483 \times 10^{-7}$  ( $0.05/97,755$ ) and  $1.02297 \times 10^{-6}$  ( $0.1/97,755$ ), respectively (Table S1).

**Targeted loci data analysis.** Additional association analyses in the Australian cohort and replication association analyses in the Norwegian and Finnish cohorts were performed in PLINK assuming an additive model of gene action. Extremely rare variants ( $MAF < 0.01$ ) were analyzed in PLINK using the conservative Fisher's Exact Test [67].

**Gene expression data analysis.** To further scrutinize sample quality we computed the mean expression signal across all detected probes for each sample independently. We then computed, for each sample, the mean correlation with all other samples across the raw average signals of all detected probes. All 59 samples passing the initial scan were retained. Using the "pDetection" p-values generated by Illumina's GenomeStudio Gene Expression module, and computing the probability that as many or more tissue samples as observed would yield a p-value  $\leq 0.05$  by chance, expression of 24,647 probes (52.2% of all probes) was significantly detected. The raw expression levels of these probes (after background subtraction using GenomeStudio) were shifted upwards to force positive values (minimum expression level of 1.0 across all samples and detected probes),  $\log_2$  transformed, and quantile normalized. To investigate whether expressed probes in the identified candidate 2q14.2 region (*INHBB* structural locus  $\pm 500$  kb) are significantly correlated with preeclampsia, and whether the identified candidate SNPs (rs7579169, rs12711941) are putatively regulatory variants (expression quantitative trait nucleotides), we performed linear regression analysis, using disease status (preeclampsia or no preeclampsia) and/or SNP genotype (coded additively as the number of copies of the minor allele present in a person) as predictors of fully processed expression level.

### Supporting Information

**Figure S1** GenomeStudio genotype cluster plot for rs7579169. (TIFF)

**Figure S2** GenomeStudio genotype cluster plot for rs12711941. (TIFF)

**Table S1** Estimated number of independent GWAS SNPs. (DOC)

**Table S2** Genotypic correlations ( $r^2$ ) of HapMap CEU SNPs flanking the strongest associated preeclampsia SNP (rs7579169  $\pm 200$  kb). bp, distance from rs7579169; SNP Chip, SNPs present

(yes) or absent (no) on the Human OmniExpress-12 BeadChip used in this study; #*INHBB* nearGene-5 SNP; \**INHBB* intronic SNP. (DOC)

**Table S3** Genotypic correlations between rs7576192 and the other 20 re-sequenced *INHBB* locus variants, plus the two GWAS SNPs; rs7579169 and rs12711941. (DOC)

**Table S4** Replication genotyping in Norwegian and Finnish case-control cohorts. Alleles are listed as major/minor.  $r^2$  denotes the genotypic correlation between rs7579169 and rs12711941. (DOC)

## Acknowledgments

For technical assistance we thank Hao-Chang Lan, Yvonne Garcia, Janelle Bentz (Texas Biomed), Anthony Borg (RWH, Melbourne), and Divya Neelam (Sequenom). We sincerely acknowledge the support of the clinical research mid-wives Karen Reidy and Sue Duggan (RWH, Melbourne) who contributed to this study. The HUNT study is a collaboration between HUNT Research Centre, Faculty of Medicine at NTNU, the Norwegian Institute of Public Health and the Nord-Trøndelag County Council. We are indebted to all the Australian, Norwegian and Finnish women whose participation made this work possible.

## References

- Witlin AG, Sibai BM (1997) Hypertension in pregnancy: Current concepts of preeclampsia. *Annu Rev Med* 48: 115–27.
- Roberts JM, Pearson G, Cutler J, Lindheimer M (2003) Summary of the NHLBI working group on research on hypertension during pregnancy. *Hypertension* 41(3): 437–45.
- Roberts JM, Gammill HS (2005) Preeclampsia: Recent insights. *Hypertension* 46(6): 1243–9.
- Duley L (2009) The global impact of pre-eclampsia and eclampsia. *Semin Perinatol* 33(3): 130–137.
- Tang LC, Kwok AC, Wong AY, Lee YY, Sun KO, et al. (1997) Critical care in obstetrical patients: An eight-year review. *Chin Med J (Engl)* 110(12): 936–41.
- Goldenberg RL, Rouse DJ (1998) Prevention of premature birth. *N Engl J Med* 339(5): 313–20.
- Basso O, Rasmussen S, Weinberg CR, Wilcox AJ, Irgens LM, et al. (2006) Trends in fetal and infant survival following preeclampsia. *Jama* 296(11): 1357–62.
- Roberts JM, Hubel CA (2009) The two stage model of preeclampsia: Variations on the theme. *Placenta* 30 Suppl A: S32–7.
- Redman CW, Sargent IL (2009) Placental stress and pre-eclampsia: A revised view. *Placenta* 30 Suppl A: S38–42.
- Maynard SE, Min JY, Merchan J, Lim KH, Li J, et al. (2003) Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension, and proteinuria in preeclampsia. *J Clin Invest* 111(5): 649–658.
- National High Blood Pressure Education Program Working Group on High Blood Pressure in Pregnancy (2000) Report of the national high blood pressure education program working group on high blood pressure in pregnancy. *Am J Obstet Gynecol* 183(1): S1–S22.
- Roberts JM, Taylor RN, Musci TJ, Rodgers GM, Hubel CA, et al. (1989) Preeclampsia: An endothelial cell disorder. *Am J Obstet Gynecol* 161(5): 1200–1204.
- Sugimoto H, Hamano Y, Charytan D, Cosgrove D, Kieran M, et al. (2003) Neutralization of circulating vascular endothelial growth factor (VEGF) by anti-VEGF antibodies and soluble VEGF receptor 1 (sFlt-1) induces proteinuria. *J Biol Chem* 278(15): 12605–12608.
- Rodic VA, Freeman DJ, Sattar N, Greer IA (2004) Pre-eclampsia and cardiovascular disease: Metabolic syndrome of pregnancy? *Atherosclerosis* 175(2): 189–202.
- Irgens HU, Reisaeter L, Irgens LM, Lie RT (2001) Long term mortality of mothers and fathers after pre-eclampsia: Population based cohort study. *Bmj* 323(7323): 1213–7.
- Roberts JM, Hubel CA (2010) Pregnancy: A screening test for later life cardiovascular disease. *Womens Health Issues* 20(5): 304–307.
- Roberts JM, Cooper DW (2001) Pathogenesis and genetics of pre-eclampsia. *Lancet* 357(9249): 53–6.
- Sattar N, Greer IA (2002) Pregnancy complications and maternal cardiovascular risk: Opportunities for intervention and screening? *BMJ* 325(7356): 157–160.
- Johansson A, Curran JE, Johnson MP, Freed KA, Fenstad MH, et al. (2011) Identification of ACOX2 as a shared genetic risk factor for preeclampsia and cardiovascular disease. *Eur J Hum Genet* 19(7): 796–800.
- Roten LT, Fenstad MH, Forsmo S, Johnson MP, Moses EK, et al. (2011) A low COMT activity haplotype is associated with recurrent preeclampsia in a Norwegian population cohort (HUNT2). *Mol Hum Reprod* 17(7): 439–446.
- Williams PJ, Pipkin FB (2011) The genetics of pre-eclampsia and other hypertensive disorders of pregnancy. *Best Pract Res Clin Obstet Gynaecol* 25(4): 405–417.
- Moses EK, Lade JA, Guo G, Wilton AN, Grehan M, et al. (2000) A genome scan in families from Australia and New Zealand confirms the presence of a maternal susceptibility locus for pre-eclampsia, on chromosome 2. *Am J Hum Genet* 67(6): 1581–5.
- Johnson MP, Fitzpatrick E, Dyer TD, Jowett JB, Brennecke SP, et al. (2007) Identification of two novel quantitative trait loci for pre-eclampsia susceptibility on chromosomes 5q and 13q using a variance components-based linkage approach. *Mol Hum Reprod* 13(1): 61–67.
- Fitzpatrick E, Johnson MP, Dyer TD, Forrest S, Elliott K, et al. (2009) Genetic association of the activin A receptor gene (ACVR2A) and pre-eclampsia. *Mol Hum Reprod* 15(3): 195–204.
- Roten LT, Johnson MP, Forsmo S, Fitzpatrick E, Dyer TD, et al. (2009) Association between the candidate susceptibility gene ACVR2A on chromosome 2q22 and pre-eclampsia in a large Norwegian population-based study (the HUNT study). *Eur J Hum Genet* 17(2): 250–257.
- Johnson MP, Roten LT, Dyer TD, East CE, Forsmo S, et al. (2009) The ERAP2 gene is associated with preeclampsia in Australian and Norwegian populations. *Hum Genet* 126(5): 655–666.
- Fenstad MH, Johnson MP, Roten LT, Aas PA, Forsmo S, et al. (2010) Genetic and molecular functional characterization of variants within TNFSF13B, a positional candidate preeclampsia susceptibility gene on 13q. *PLoS One* 5(9): e12993.
- Zintzaras E, Santos M (2011) Estimating the mode of inheritance in genetic association studies of qualitative traits based on the degree of dominance index. *BMC Med Res Methodol* 11(1): 171.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3): 559–575.
- Wingender E, Chen X, Hehl R, Karas H, Liebich I, et al. (2000) TRANSFAC: An integrated system for gene expression regulation. *Nucleic Acids Res* 28(1): 316–319.
- Chekmenov DS, Haid C, Kel AE (2005) P-match: Transcription factor binding site search by combining patterns and weight matrices. *Nucleic Acids Res* 33(Web Server issue): W432–7.
- Grabe N (2002) AliBaba2: Context specific identification of transcription factor binding sites. *In Silico Biol* 2(1): S1–15.
- NCI-NHGRI Working Group on Replication in Association Studies, Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, et al. (2007) Replicating genotype-phenotype associations. *Nature* 447(7145): 655–660.
- Shriner D, Vaughan LK, Padilla MA, Tiwari HK (2007) Problems with genome-wide association studies. *Science* 316(5833): 1840–1842.
- Williams SM, Canter JA, Crawford DC, Moore JH, Ritchie MD, et al. (2007) Problems with genome-wide association studies. *Science* 316(5833): 1840–1842.

## The Finnish Genetics of Pre-eclampsia Consortium (FINNPEC)

We appreciate the collaboration with the following members of the FINNPEC Study Group: Eeva Ekholm (Turku University Central Hospital, Turku, Finland), Kaarin Mäkilä-Anttila, (Oulu University Hospital, Oulu, Finland), Reija Hietala, Susanna Sainio and Terhi Saisto (Helsinki University Central Hospital, Helsinki, Finland), Jukka Uotila, (Tampere University Hospital, Tampere, Finland) Tia Aalto-Viljakainen, Miira Klemetti and Anna Inkeri Lokki (University of Helsinki, Helsinki, Finland), and Leena Georgiadis (Kuopio University Hospital, Kuopio, Finland). The expert technical assistance of Elina Huovari, Eija Kortelainen, Satu Leminen, Aija Lähdesmäki, Susanna Mehtälä, and Christina Salmen is gratefully acknowledged.

## Author Contributions

Conceived and designed the experiments: MPJ SPB JB EKM. Performed the experiments: MPJ EKM. Analyzed the data: MPJ HHHG JWK TDD LJA JB. Contributed reagents/materials/analysis tools: MPJ SPB JMS LTR ACI SH EK JK KK AP HL RA JB EKM. Wrote the paper: MPJ HHHG EKM. Reviewed Australian case-control medical records: SPB CEE. Conceived FINNPEC Study: EK JK HL. FINNPEC Study Core Investigators: SH EK JK KK AP HL.

36. Greene CS, Penrod NM, Williams SM, Moore JH (2009) Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS One* 4(6): e5639.
37. Moses EK, Fitzpatrick E, Freed KA, Dyer TD, Forrest S, et al. (2006) Objective prioritization of positional candidate genes at a quantitative trait locus for pre-eclampsia on 2q22. *Mol Hum Reprod* 12(8): 505–512.
38. Petraglia F, Aguzzoli L, Gallinelli A, Florio P, Zonca M, et al. (1995) Hypertension in pregnancy: Changes in activin A maternal serum concentration. *Placenta* 16(5): 447–54.
39. Muttukrishna S, Knight PG, Groome NP, Redman CW, Ledger WL (1997) Activin A and inhibin A as possible endocrine markers for pre-eclampsia. *Lancet* 349(9061): 1285–8.
40. Fraser RF, 2nd, McAsey ME, Concy P (1998) Inhibin-A and pro-alpha C are elevated in preeclamptic pregnancy and correlate with human chorionic gonadotropin. *Am J Reprod Immunol* 40(1): 37–42.
41. Caniggia I, Winter J, Lye SJ, Post M (2000) Oxygen and placental development during the first trimester: Implications for the pathophysiology of pre-eclampsia. *Placenta* 21 Suppl A: S25–30.
42. Muttukrishna S, North RA, Morris J, Schellenberg JC, Taylor RS, et al. (2000) Serum inhibin A and activin A are elevated prior to the onset of pre-eclampsia. *Hum Reprod* 15(7): 1640–5.
43. Casagrandi D, Bearfield C, Geary J, Redman CW, Muttukrishna S (2003) Inhibin, activin, follistatin, activin receptors and beta-glycan gene expression in the placental tissue of patients with pre-eclampsia. *Mol Hum Reprod* 9(4): 199–203.
44. Venkatesha S, Toporsian M, Lam C, Hanai J, Mammoto T, et al. (2006) Soluble endoglin contributes to the pathogenesis of preeclampsia. *Nat Med* 12(6): 642–649.
45. Wang A, Rana S, Karumanchi SA (2009) Preeclampsia: The role of angiogenic factors in its pathogenesis. *Physiology (Bethesda)* 24: 147–158.
46. Jones RL, Findlay JK, Salamonsen LA (2006) The role of activins during decidualisation of human endometrium. *Aust N Z J Obstet Gynaecol* 46(3): 245–9.
47. Khong TY, De Wolf F, Robertson WB, Brosens I (1986) Inadequate maternal vascular response to placentation in pregnancies complicated by pre-eclampsia and by small-for-gestational age infants. *Br J Obstet Gynaecol* 93(10): 1049–1059.
48. Pijnenborg R, Vercruyse L, Hanssens M (2006) The uterine spiral arteries in human pregnancy: Facts and controversies. *Placenta* 27(9–10): 939–58.
49. Redman CW, Sargent IL (2010) Immunology of pre-eclampsia. *Am J Reprod Immunol* 63(6): 534–543.
50. Holmen J, Midtjell K, Krüger Ø, Langhammer A, Holmen TL, et al. (2003) The Nord-Trøndelag health study 1995–97 (HUNT 2): Objectives, contents, methods and participation. *Norsk Epidemiologi* 13(1): 19–32.
51. Holmen J, Kjelsaas MB, Krüger Ø, Ellekjær H, Bratberg G, et al. (2004) [Attitudes to genetic epidemiology - illustrated by questions for re-consent to 61,426 participants at HUNT]. *Norsk Epidemiologi* 14(1): 27–31.
52. Moses EK, Johnson MP, Tommerdal L, Forsmo S, Curran JE, et al. (2008) Genetic association of preeclampsia to the inflammatory response gene SEPS1. *Am J Obstet Gynecol* 198(3): 336.e1–336.e5.
53. Laivuori H, Kaaja R, Ylikorkala O, Hiltunen T, Kontula K (2000) 677 C→T polymorphism of the methylenetetrahydrofolate reductase gene and preeclampsia. *Obstet Gynecol* 96(2): 277–280.
54. Hiltunen LM, Laivuori H, Rautanen A, Kaaja R, Kere J, et al. (2009) Blood group AB and factor V leiden as risk factors for pre-eclampsia: A population-based nested case-control study. *Thromb Res* 124(2): 167–173.
55. Brown MA, Gallery EDM, Gatt SP, Leslie G, Robinson J (1993) Management of hypertension in pregnancy: Executive summary. *Med J Aust* 158(10): 700–2.
56. Brown MA, Hague WM, Higgins J, Lowe S, McCowan L, et al. (2000) The detection, investigation and management of hypertension in pregnancy: Executive summary. *Aust N Z J Obstet Gynaecol* 40(2): 133–8.
57. Lowe SA, Brown MA, Dekker GA, Gatt S, McLintock CK, et al. (2009) Guidelines for the management of hypertensive disorders of pregnancy 2008. *Aust N Z J Obstet Gynaecol* 49(3): 242–246.
58. Fitzpatrick E, Göring HH, Liu H, Borg A, Forrest S, et al. (2004) Fine mapping and SNP analysis of positional candidates at the preeclampsia susceptibility locus (PREG1) on chromosome 2. *Hum Biol* 76(6): 849–62.
59. International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437(7063): 1299–1320.
60. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449(7164): 851–861.
61. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, et al. (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* 467(7311): 52–58.
62. Ovcharenko I, Nobrega MA, Loots GG, Stubbs L (2004) ECR browser: A tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Res* 32(Web Server issue): W280–6.
63. Løset M, Mundal SB, Johnson MP, Fenstad MH, Freed KA, et al. (2011) A transcriptional profile of the decidua in preeclampsia. *Am J Obstet Gynecol* 204(1): 84.e1–84.27.
64. Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 81(5): 1084–1097.
65. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2): 263–265.
66. Duggal P, Gillanders EM, Holmes TN, Bailey-Wilson JE (2008) Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics* 9: 516.
67. Fisher RA (1922) On the interpretation of  $\chi^2$  from contingency tables, and the calculation of P. *J R Stat Soc* 85(1): 87–94.