Genetic dissection of the pre-eclampsia susceptibility locus on chromosome 2q22 reveals shared novel risk factors for cardiovascular disease

Matthew P. Johnson¹, Shaun P. Brennecke²,³, Christine E. East²,³, Thomas D. Dyer¹, Linda T. Roten⁴, J. Michael Proffitt¹, Phillip E. Melton⁵, Mona H. Fenstad⁶,⁷, Tia Aalto-Viljakainen⁷,⁸,⁹, Kaarin Mäkikallio¹⁰, Seppo Heinonen¹¹, Eero Kajantie¹²,¹³, Juha Kere⁸,¹⁴,¹⁵, Hannele Laivuori⁸,⁹, for the FINNPEC Study Group†, Rigmor Austgulen⁴, John Blangero¹, and Eric K. Moses¹,⁵,¹⁶,*

¹Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX 78227, USA ²Department of Perinatal Medicine, Royal Women’s Hospital, University of Melbourne, Parkville, VIC 3052, Australia ³Department of Obstetrics & Gynaecology, University of Melbourne, Parkville, VIC 3052, Australia ⁴Department of Cancer Research and Molecular Medicine, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim 7006, Norway ⁵Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, Perth, WA 6009, Australia ⁶Department of Immunology and Transfusion Medicine, St. Olav Hospital, Trondheim 7006, Norway ⁷The National Graduate School of Clinical Investigation, University of Helsinki, Helsinki FI-00014, Finland ⁸Department of Medical Genetics, Haartman Institute, University of Helsinki, Helsinki FI-00014, Finland ⁹Research Programs Unit, Women’s Health, University of Helsinki, Helsinki FI-00014, Finland ¹⁰Department of Obstetrics & Gynecology, Oulu University Hospital, Oulu 90220, Finland ¹¹Department of Obstetrics and Gynaecology, Kuopio University Hospital, and University of Eastern Finland, Kuopio 70211, Finland ¹²Department of Chronic Disease Prevention, Diabetes Prevention Unit, National Institute for Health and Welfare, Helsinki FI-00271, Finland ¹³Children’s Hospital, Helsinki University Central Hospital and University of Helsinki, Helsinki FI-00014, Finland ¹⁴Department of Biosciences and Nutrition, and Science for Life Laboratory, Karolinska Institutet, Stockholm SE-171 77, Sweden ¹⁵Folkhalsan Institute of Genetics, Helsinki FI-00014, Finland ¹⁶Present Address: Centre for Genetic Epidemiology & Biostatistics, The University of Western Australia, Perth, WA 6009, Australia

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

Submitted on May 24, 2012; resubmitted on February 4, 2013; accepted on February 7, 2013

ABSTRACT: Pre-eclampsia is an idiopathic pregnancy disorder promoting morbidity and mortality to both mother and child. Delivery of the fetus is the only means to resolve severe symptoms. Women with pre-eclamptic pregnancies demonstrate increased risk for later life cardiovascular disease (CVD) and good evidence suggests these two syndromes share several risk factors and pathophysiological mechanisms. To elucidate the genetic architecture of pre-eclampsia we have dissected our chromosome 2q22 susceptibility locus in an extended Australian and New Zealand familial cohort. Positional candidate genes were prioritized for exon-centric sequencing using bioinformatics, SNPing, transcriptional profiling and QTL-walking. In total, we interrogated 1598 variants from 52 genes. Four independent SNP associations satisfied our gene-centric multiple testing correction criteria: a missense LCT SNP (rs2322659, P = 0.0027), a synonymous LRP1B SNP (rs35821928, P = 0.0001), an UTR-3 RND3 SNP (rs115015150, P = 0.0024) and a missense GCA SNP (rs17783344, P = 0.0020). We replicated the LCT SNP association (P = 0.02) and observed a borderline association for the GCA SNP (P = 0.07) in an independent Australian singleton cohort. Moreover, these four SNP associations could not be replicated in two additional case–control populations from Norway and Finland. These four SNPs, however, exhibit pleiotropic effects with several quantitative CVD-related traits. Our results underscore the genetic complexity of pre-eclampsia and present novel empirical evidence of possible shared genetic mechanisms underlying both pre-eclampsia and other CVD-related risk factors.

Key words: 2q22 / cardiovascular disease risk trait / genetic association / pleiotropy / pre-eclampsia

† The members of the FINNPEC Study Group are given in Appendix.

© The Author 2013. Published by Oxford University Press on behalf of the European Society of Human Reproduction and Embryology. All rights reserved. For Permissions, please email: journals.permissions@oup.com
**Introduction**

Pre-eclampsia is a common and serious complication of human pregnancy affecting 3–5% of all women and it accounts for substantive maternal and fetal morbidity and mortality (Robillard et al., 2011). To date, delivery is the only intervention for adequate resolution of severe symptoms. Clinically, pre-eclampsia is diagnosed by new onset (maternal) hypertension and proteinuria, established after 20 weeks gestation. These maternal symptoms are however said to arise following shallow fetal-derived cytotrophoblast invasion that contributes to the failed remodeling of maternal spiral arteries perfusing the placenta earlier in the pregnancy (Redman and Sargent, 2009; Roberts and Hubel, 2009). The maternal hypertension is now recognized to be secondary to the diffuse endothelial dysfunction occurring in the first half of a pre-eclamptic pregnancy (Roberts et al., 1989) and the proteinuria is associated with glomerular endotheliosis (Maynard et al., 2003; Sugimoto et al., 2003). Pre-eclampsia is therefore primarily characterized by endothelial dysfunction.

Whilst the exact pathophysiological mechanisms of pre-eclampsia are still to be elucidated, it is widely acknowledged that there is a strong genetic component (Cooper et al., 1993; Roberts and Cooper 2001). Identification of pre-eclampsia susceptibility genes may enable early predictive therapies for pre-eclampsia and related conditions such as cardiovascular disease (CVD). The association between pre-eclampsia and the development of later-life maternal CVD has been well established (Bellamy et al., 2007; McDonald et al., 2008; Giguere et al., 2012). Pathological features such as endothelial dysfunction and numerous risk factors (e.g. obesity, general inflammation, insulin resistance, hyperlipidemia) are shared between these conditions (Rodie et al., 2004; Roberts et al., 2011). Studies also show a history of pre-eclampsia exerts an independent risk for later life CVD (Irgens et al., 2001; Lykke et al., 2009; Bushnell and Chireau, 2011). It can therefore be hypothesized that pregnancy provides a metabolic ‘stress test’ to unmask underlying risk of CVD in women (Roberts and Hubel, 2010). These data have led several investigators to speculate that the genetic risk factors for pre-eclampsia will also be relevant to CVD, providing an increased impetus and justification for their discovery (Roberts and Cooper 2001; Sattar and Greer 2002). Evidence supporting shared genetic precursors is now emerging (Johansson et al., 2011; Roten et al., 2011).

In our attempts to identify genetic risk factors for pre-eclampsia, we have primarily focused on positional cloning strategies, making no a priori assumptions about the nature of genes involved. We performed genome-wide linkage mapping studies in multiple affected families from Australia and New Zealand (AUS/NZL), identifying susceptibility loci on chromosomes 2q22, 5q and 13q (Moses et al., 2000; Johnson et al., 2007). More recently, using genome-wide association mapping in a large Caucasian case–control cohort from Australia, we identified a susceptibility locus on chromosome 2q14, near the Inhibin, beta B (INHBB) gene (Johnson et al., 2012).

In this paper we now report on our genetic dissection of the 2q22 susceptibility locus. Our strategy has involved prioritizing genes at this locus for sequencing in founder individuals from an extended set of AUS/NZL families followed by association analyses of these identified variants in the same cohort. Replication association analyses were performed in case–control population cohorts from Australia, Norway and Finland. We herein report novel SNP associations with pre-eclampsia and with a number of CVD-related traits, providing novel empirical evidence of possible shared genetic mechanisms underlying pre-eclampsia and other CVD-related risk factors.

**Materials and Methods**

**Subjects**

**Australian/New Zealand families**

The AUS/NZL families used in this study were of confirmed Caucasian ancestry retrospectively ascertained over a 15-year period from 1984 to 1999. Family members were recruited at the Royal Woman’s Hospital (RWH), Melbourne, Australia, via media advertisements in Sydney, Australia and at the National Women’s Hospital in Auckland, New Zealand. These families consist of the original 34 (26 Australian and 8 New Zealand) that we have used to localize three pre-eclampsia susceptibility QTLs to chromosomes 2q, 5q and 13q (Moses et al., 2000; Johnson et al., 2007), and an additional 40 (Australian) families that we have subsequently ascertained (Fitzpatrick et al., 2009).

**Australian singletons**

The Australian case–control cohort of 1095 women used in this study were of confirmed Caucasian ancestry retrospectively ascertained from a larger Australian case–control cohort of 1774 women that were recruited at the RWH, Melbourne, Australia over a 5-year period from 2007 to 2011. The Australian case–control population seen at the RWH in Melbourne is ~70% Caucasian and for this study the focus was on the recruitment of Caucasian subjects.

**Norwegian singletons**

All Norwegian samples were retrospectively selected from a large multipurpose health survey (HUNT2) conducted over a 3-year period from 1995 to 1997 in Nord-Trøndelag County in Norway (Holmen et al., 2003). The HUNT2 pre-eclampsia case–control cohort used in this study has been described in detail elsewhere (Moses et al., 2008; Roten et al., 2009), and in the online supplementary data.

**Finnish singletons**

The Finnish patient samples used in this study originate from the Finnish Genetics of Pre-eclampsia Consortium (FINNPEC) study cohort and the Southern Finland pre-eclampsia study cohort. FINNPEC is an ongoing multi-centre study where DNA samples and data have been collected prospectively at all university hospitals in Finland since 2008. The Southern Finnish case–control cohort is a retrospective collection of pre-eclamptic and control women from the Helsinki University Central Hospital between January 1997 and April 1998 (Lauvri et al., 2000; Hiltunen et al., 2009). Additional details on these Finnish cohorts can be found in the online supplementary data.

**Mexican American families**

Conceived in 1991, the San Antonio Family Heart Study (SAFHS) was designed to elucidate the genetic aspects of cardiovascular-related health and disease in Mexican Americans (Mitchell et al., 1996). The SAFHS comprises individual families ascertained through a single adult proband (40–60 years of age) with at least six offspring and/or siblings ≥16 years of age from the Mexican American community within San Antonio, TX, USA. Family ascertainment was undertaken without bias towards phenotype or disease status. The enrollment criteria did, however, enrich for large, multigenerational families with first-, second- and third-degree relatives of the proband and/or of the proband’s
Pre-eclampsia diagnosis

Australian cohorts

Pre-eclampsia diagnosis in the AUS/NZL families and Australian singletons was conducted by qualified clinicians using criteria set by the Australasian Society for the Study of Hypertension in Pregnancy (Brown et al., 1993; Brown et al., 2000), and the Society of Obstetric Medicine of Australia and New Zealand for the management of hypertensive diseases of pregnancy (Lowe et al., 2009), respectively. Women were considered pre-eclamptic if they were previously normotensive and if they, on at least two occasions 6 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 ≥ 140 mmHg and/or DBP ≥ 90 mmHg. Additionally, significant new onset proteinuric levels were either ≥ 0.3 g/l in a 24-h specimen, at least a ‘2+’ proteinuria dipstick reading from a random urine collection or a spot protein:creatinine ratio ≥ 0.03 g/mmol. Pre-eclamptic 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 pre-eclampsia (e.g. renal disease, diabetes, twin pregnancies or fetal chromosomal abnormalities) were excluded. For this study, the AUS/NZL familial cohort consisted of 480 individuals originating from 74 families of whom 140 were coded as affected (20 eclamptic, 120 pre-eclamptic) and 146 were coded as unaffected. Of the 1774 case–control women initially recruited for this study, 1095 were of confirmed Caucasian ancestry, making our inclusion criteria for this study. Of these, 499 were confirmed, by medical records, as having pre-eclampsia (cases) and 596 were confirmed as having a normal pregnancy (controls).

Norwegian cohort

The definition and classification of pre-eclampsia used for the Norwegian samples was established by the MBRN based on previously reported guidelines (National High Blood Pressure Education Program Working Group on High Blood Pressure in Pregnancy, 2000). The MBRN definition for pre-eclampsia was defined as an increase in SBP/DBP to at least 140/90 mmHg (or an increase in SBP ≥ 30 mmHg, or in DBP ≥ 15 mmHg from the level measured before the 20th week of gestation), combined with proteinuria (protein excretion of at least 0.3 g per 24 h or ≥ 1+ on a dip stick). Based on these diagnostic criteria there were 1179 women registered with pre-eclampsia (cases) and 2358 women with a history of a normal, healthy pregnancy (controls). Of these registered women, blood samples were available for 1134 cases and 2263 controls at the HUNT Biobank and included for this study.

Finnish cohort

Finnish women who suffered a pre-eclamptic 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 ≥ 140 mmHg and/or DBP ≥ 90 mmHg on at least two occasions with new onset proteinuria (≥ 0.3 g/24 h or ≥ 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. Pre-eclampsia in the Southern Finnish case/control cohort was defined as two SBP/DBP measurements at least 6 hours apart ≥ 140/90 mmHg and proteinuria measurement ≥ 0.3 g in a 24-h urine collection, or at least a ‘1+’ dipstick reading after 20 weeks gestation (Laivuori et al., 2000). A total of 760 pre-eclamptic (case) women and 664 control women from the FINNPEC study cohort and 95 control women from the Southern Finland pre-eclampsia study cohort were included in this study.

SAFHS CVD risk phenotypes

The SAFHS is a rich resource of phenotypic data measurements pertaining, but not limited, to coronary heart disease risk factors [e.g. total cholesterol (TC), HDL, triglycerides], anthropometric measurements of adiposity (e.g. height, weight, hip circumference, skin fold), blood pressure (e.g. systolic and diastolic), diabetes (e.g. glucose and insulin levels), inflammation (e.g. pro-inflammatory cytokines), oxidative stress (e.g. plasma total antioxidant status) and clotting factors (e.g. plasminogen, fibrinogen).

At the time of subject ascertainment pregnant women were ineligible to participate in a medical examination and subsequent acquisition of relevant phenotypic data. A more detailed account of these phenotypic assessments, biochemical methods and CVD-risk related lifestyle factor descriptions are described elsewhere (Mitchell et al., 1996; MacCluer et al., 1999).

Positional candidate gene prioritization

To prioritize validated, protein-coding positional candidate genes for further interrogation we utilized one of the four different strategies: (i) bio-informatics: within the 1-LOD localization support interval (95% confidence) of the 2q22 susceptibility locus we used the computer program GeneSniffer (www.genesniffer.org), as previously described (Moses et al., 2006; Johnson et al., 2007), with user-defined disease-specific keywords, as previously described (Johnson et al., 2009), to objectively select candidate genes based on published database information. (ii) SNPing: known validated coding and putative functional SNPs were selected from the NCBI SNP database (Homo sapiens dbSNP build 125). SNPs associated with pre-eclampsia-related pathophysiological concepts such as CVD, immune- and immunologically related phenotypes were also selected. SNPs genetically associated with these pathophysiology-related phenotypes were identified via dbSNP’s PubMed link, or via the disease-specific categorized literature track of Ingenuity Pathway Analysis’s Gene View (www.ingenuity.com). (iii) Whole-genome transcriptional profiling: differentially expressed genes within the 2q22 QTL were identified from a retrospective collection of Norwegian decidual samples as previously described (Leset et al., 2011). (iv) QTL ‘walking’: positional candidate genes were also triaged for sequencing by systematically ‘walking’ one gene at a time in either direction away from the maximum 2q22 QTL linkage signal (Moses et al., 2000; Moses et al., 2006).

DNA sequencing

The translated and untranslated exons of prioritized candidate genes were sequenced in 48 pre-eclamptic women from the AUS/NZL families. These women were either founder individuals and/or probands chosen from the most informative families. This sequencing sample set provided >99% probability of detecting any variant with a minor allele frequency (MAF) ≥ 5%. This sample set also enhances the potential to discover rarer structural and potentially functional variants that are enriched in these pre-eclamptic women. For higher priority candidate genes identified from our bioinformatic, SNPing or transcriptional profiling strategies, 2 kb of the proximal promoter (i.e. upstream of the translation start site) was also sequenced. Gene sequence reference templates were obtained from the UCSC Genome Bioinformatics database [HUMAN, Mar. 2006 (NCBI36/hg 18)]. All primers were designed using Primer 3 (v 0.4.0) and BLASTed to assess their uniqueness to the human genome. Primers were designed to flank the exons by ~100 bp, and any contiguous primer pairs were designed to overlap by ~100–150 bp. Genomic DNA (gDNA) for sequencing was extracted from peripheral blood as previously described (Moses et al., 2000). Standard PCRs were performed

spouse. More than 1400 individuals from 42 families have been recruited in total.
with 20 ng of gDNA in a 10-µl reaction volume. If standard PCR optimization conditions failed, FailSafe PCR pre-mixes (Epitect 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 amplicons (1 µl) using BigDye Terminator v3.1 chemistry (Life Technologies) in a 5-µl reaction volume. Sequence reaction amplification was performed on a GeneAmp 9700 thermal cycler using standard cycling conditions. Amplified sequence products were purified with BigDye XTerminator purification kits according to manufacturers’ instructions (Life Technologies). Purified sequence reactions were electrophoretically separated on a 3730xl DNA Analyzer, sequence reaction quality was assessed using Sequencing Analysis software v5.1.1 and sequence variant identification was performed using SeqScape v2.6 (Life Technologies).

Genotyping

AUS/NZL families

Stemming from our ‘SNPing’ gene prioritization strategy, a custom Illumina GoldenGate genotyping assay was designed and genotyped as previously described (Fitzpatrick et al., 2009). The remainder of our custom genotyping assays were designed with Illumina’s Assay Design Tool and processed in accordance with Illumina’s GoldenGate assay protocol with VeraCode technology (Illumina, Inc., San Diego, CA). Samples for the custom VeraCode genotyping assays were scanned on Illumina’s BeadXpress Reader System with VeraScan image data acquisition software (v1.1.9.2). All genotype clustering and individual sample genotype calls were performed with Illumina’s GenomeStudio software (v2010.2), Genotyping Module (v1.1.9). Genotype assay quality control measures were assessed with Illumina’s internal contamination and assay performance metrics. Furthermore, genotype call concordance was assessed against the sequence data obtained from the sub-set of AUS/NZL familial individuals.

Genomic variants that either failed Illumina’s custom GoldenGate assay design parameters or the genotyping assay itself were earmarked for genotyping with Sequenom-based MassArray technology (Sequenom, San Diego, CA). Multiplex 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 according to 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). Genotype calls generated by TyperAnalyzer were assessed against the sequence data obtained from our sequencing sub-set of AUS/NZL familial individuals.

Australian singletons

Genotyping was performed using TaqMan SNP Genotyping assays on a 7900HT Fast Real-Time PCR system (Life Technologies). For each TaqMan SNP assay, 50 ng of gDNA was used in a 5-µl reaction volume with 2.5-µl TaqMan Genotyping Master Mix, 0.125-µl TaqMan assay mix (40×) and 1.375 µl water. Four no template (water) controls were incorporated into each 384-well plate and 2% of the samples were blindly genotyped in duplicate. SNP genotype clustering and individual sample genotype calls were generated using Sequence Detection Systems software v2.2.2 (Life Technologies).

Norwegian singletons

SNP genotyping was performed using TaqMan SNP Genotyping assays as previously described for the Australian singletons. Four no template (water) controls were incorporated into each 384-well plate. To assess positive genotype quality control, the HUNT Biobank’s standard-operating procedure is to genotype ~10% of the case–control samples, in duplicate, on an alternative genotyping platform, the LightCycler (Roche Diagnostics Scandinavia AB, Stockholm, Sweden), to assess genotype concordance.

Finnish singletons

SNP genotyping was performed using TaqMan SNP Genotyping assays as previously described for the Australian singletons. Four no template (water) controls were incorporated into each 384-well plate. Two samples, independent to the study participants, were included in each genotyping plate as positive controls.

Mexican American families

Prioritized pre-eclampsia-associated SNPs from the AUS/NZL families were genotyped in the SAFHS participants using Sequenom’s genotyping platform as previously described for the AUS/NZL families.

Statistical analysis

Genotype error checking

In addition to the laboratory-based genotype quality control measures, genotypes pertaining to the AUS/NZL and Mexican American familial cohorts not conforming to Mendelian inheritance laws were identified and assessed using SimWalk2 (Sobel and Lange 1996). Mendelian discrepancies and spurious recombinations were removed by blanking these genotypes identified in SimWalk2 as having a high probability of being in error.

SNP allele frequency estimation

We used SOLAR (Almasy and Blangero, 1998) to estimate SNP allele frequencies and their standard errors by using maximum likelihood methods that account for pedigree structure. Tests for deviations from Hardy–Weinberg equilibrium (HWE) allowing for non-independence amongst family members were also performed in SOLAR.

SNP association analysis

The classical measured genotype approach for association analyses (Hopper and Mathews 1982; Boerwinkle et al., 1986), as implemented in SOLAR (Blangero et al., 2005), was employed to analyze the 2q22 susceptibility locus genomic variants against the pre-eclampsia phenotype in the Australian, Norwegian and Finnish cohorts. The measured genotype procedure was also used to test for significant pleiotropic effects of prioritized pre-eclampsia SNPs upon cardiovascular-related health and disease clinical phenotypes in the Mexican American families. This single degree of freedom test uses a standard threshold model assuming an underlying normal distribution of liability. The threshold model and its assumptions are akin to those used in standard logistic regression but benefits from the ease of interpretation with regard to the genetic effects of the rarer allele. However, the resulting probit-based regression coefficients (beta values) have the opposite sign of those usually obtained in logistic regression because they represent the movement of the disease threshold. Negative signs suggest that the threshold has been moved to the left thereby increasing the expected disease prevalence whilst positive signs are associated with higher thresholds and decreased prevalences. The measured genotype test of association (MGp) can assess the extent of genotypic mean differences (the liability, or risk scale) between case and control individuals assuming a model of additive gene action. This test also takes into account non-independence amongst related individuals.
and thus is a more powerful approach in estimating appropriate Type I error rates.

**Multiple hypothesis testing**

The heterogenic nature of pre-eclampsia and our broad 2q22 susceptibility locus may suggest that more than one pre-eclampsia candidate gene exists in this region. In light of this, a gene-centric approach to conservatively correct for the effect of multiple SNP testing was adopted for the AUS/NZL family data (Moskvin and Schmidt 2008). The effective number of SNPs (SNPEFFNUM), taking into consideration the extent of SNP linkage disequilibrium (LD), is subsequently used in a modified Bonferroni procedure to generate an adjusted target alpha level (0.05/SNPEFFNUM). This adjusted P-value (ADJP) still maintains an overall significance level of 0.05 or less for the SNPs tested in the AUS/NZL families.

**Experiment-wide corrections**

Simulation analyses \( n = 10,000 \) were performed in SOLAR to obtain experiment-wide significance values (EXPp) for SNP associations in the AUS/NZL families.

**Combined cohort association analysis**

To collectively analyze the null hypothesis in the discovery AUS/NZL family cohort and the replication case-control cohorts from Australia, Norway and/or Finland we used METAL (Willer et al., 2010). For each SNP locus, the observed P-value from the discovery cohort and the unidirectional one-sided P-value from one or more replication cohorts were converted to a signed Stouffer’s Z-score and then combined in a weighted sum, with weights proportional to the square-root of each cohort’s sample size.

**Ethics**

Prior approval to use the Australian, Norwegian, Finnish and Mexican American sample cohorts were granted by each country’s relevant ethical review board(s) (see online supplementary data). Ethical approval to conduct molecular and statistical analyses on all sample cohorts was obtained from the Institutional Review Board of the University of Texas Health Science Center at San Antonio.

**Results**

**Validated (protein coding) gene identification**

The 1-LOD localization support interval of the 2q22 susceptibility locus encompasses \( 33.6 \) Mb and extends from \( 132.9 \) to \( 166.5 \) Mb (NCBI Human Genome Builds 36.1 and 36.3). This interval contained 163 potential genes/functional units which comprised 87 validated (protein coding) genes, 35 hypothetical loci, 25 pseudogenes, 15 withdrawn loci, and one mircRNA (micro RNA) species (Supplementary data, Table S1).

**Prioritized gene sequencing**

Of the 87 validated (protein coding) genes, 52 were prioritized for exon-centric re-sequencing as a result of (i) bioinformatic analysis using GeneSniffer, (ii) positional candidate SNP analysis (SNPing), (iii) whole-genome transcriptional profiling in decidual basalis tissue from 37 pre-eclamptic and 58 non-pre-eclamptic women (Leset et al., 2011) and (iv) systematically selecting candidate genes by ‘walking’ away from the maximum linkage signal (QTL ‘walking’) (Supplementary data, Table SII). Higher priority candidate genes identified from our preliminary genetic association results, bioinformatic and/or transcriptional profiling strategies warranted additional sequencing of the proximal promoter region for 13 of these 52 genes, plus introns 49–55 \( (\sim 30 \) kb) of the LRP1B gene and all the introns of the GCA gene \( (\sim 15.3 \) kb) (Supplementary data, Table SII). We identified 203 novel variants (198 SNPs, 3 insertions, 2 deletions) which have subsequently been submitted to NCBI’s dbSNP and assigned an ‘rs’ identifying number (Supplementary data, Table SII). Of particular note was the identification of a novel nonsense SNP (rs141572054) residing in exon 6 of the grancalcin (GCA) gene. This SNP truncates \( \sim 15\% \) of the wild type polypeptide C-terminal (32 amino acids) and is rare (MAF = 0.006). The rs141572054 SNP was observed in six heterozygous individuals who were either pre-eclamptic \( (n = 3) \) or male \( (n = 3) \) from three independent AUS/NZL families.

**Genotyping in the AUS/NZL families**

A total of 1598 variants (1571 SNPs, 16 insertions, 11 deletions) from the 52 prioritized positional candidate genes were designed for custom genotyping assays and genotyped in our AUS/NZL families (74 families; 140 pre-eclamptic females, 146 non-pre-eclamptic females, 34 females with unknown phenotype and 160 males). Of these variants, 54 \( (3.38\%) \) failed the custom assay design or the genotyping assay itself for both the Illumina and Sequenom genotyping platforms, and 258 \( (16.15\%) \) were non-polymorphic. These non-polymorphic variants were either selected from our initial SNPing prioritization strategy \( (n = 196, 75.97\%) \), were not a valid variant based on independent sequence analysis \( (n = 44, 17.05\%) \), or the reference sequence template allele was discordant to the allele present for all of the 48 AUS/NZL pre-eclamptic females sequenced \( (n = 18, 6.98\%) \) (Supplementary data, Table SII). In order to effectively analyze genotype data in the AUS/NZL families, an additional 175 variants \( (10.95\%) \) were excluded due to observing fewer than five copies of the rare allele (Supplementary data, Table SII). A total of 1111 variants \( (1089 \) SNPs, 13 insertions, 9 deletions) were therefore successfully genotyped and analyzed in the AUS/NZL families with an average individual genotype success rate of 98.64\% \( \text{range} 79.9\% – 100\% \) (Fig. 1).

**Association analysis in the AUS/NZL families**

We identified a missense LCT SNP \( \text{rs}2322659, \ MGp = 0.00272, \ MAF = 0.236 \), a synonymous LRPIB SNP \( \text{rs}35821928, \ MGp = 0.00018, \ MAF = 0.077 \), an UTR-3 RND3 SNP \( \text{rs}15051510, \ MGp = 0.00243, \ MAF = 0.021 \) and a missense GCA SNP \( \text{rs}17783344, \ MGp = 0.00210, \ MAF = 0.167 \) to be significantly associated with pre-eclampsia in the AUS/NZL families (Table I). A significant association was denoted as an observed measured genotype P-value (MGp) being less than the corrected or adjusted P-value (ADJP) based on our gene-centric approach to correct for the effective number of actual SNPs (tests). This multiple testing correction takes into consideration the strength of marker genotypic correlations (i.e. LD) for use in a modified Bonferroni procedure to generate an adjusted target alpha level. To ascertain an experiment-wide correction of these four SNPs we used SOLAR to simulate a quantitative trait with a heritability of 0.51 (Johnson et al., 2007) and assigned affection status to a fixed number of affected women \( n = 140 \) with the
highest simulated trait values. The QTLD analysis procedure in SOLAR (Blangero et al., 2005) was repeated 10,000 times to our simulated trait and an experiment-wide P-value (EXPp) for each SNP was determined (x < 0.05). Allowing for the total experiment these four SNPs remain significant (Table I).

**SNP LD**

To evaluate the extent of LD between the significant SNPs in the AUS/NZL families we used the R package snp.plotter (Luna and Nicodemus 2007) to calculate and plot the pairwise genotypic correlations (r²). The LCT (rs2332659), LRP1B (rs35821928), RND3 (rs115015150) and GCA (rs17783344) SNP loci are separated by ≈4.7 Mb, ~10.1 Mb, and ~11.9 Mb, respectively, and all are independent of each other (r² < 0.19) (data not shown). Furthermore, we investigated the extent of LD within a physical window of up to 500 kb either side of each associated SNP with other successfully analyzed variants in the AUS/NZL families. The LCT SNP was strongly correlated (r² > 0.9) with three synonymous LCT SNPs (rs2278544, rs6719488, rs2236783), a nearGene-5 LCT ‘T’ insertion (rs113260867), an intronic MCM6 SNP (rs309180), a nearGene-5 UBXN4 SNP (rs1438307), and an intronic DARS SNP (rs687670) (Fig. 2A). With the exception of the DARS SNP (MGp = 0.063), all other variants correlated with rs2332659 were nominally associated with pre-eclamspia (MGp < 0.05) (Supplementary data, Table SII). The large LRP1B gene (~1.9 Mb) isolated the assessment of LD to within this gene itself, of which no other tested variant was strongly correlated with the synonymous rs35821928 SNP (r² < 0.48) (Fig. 2B). The nearest validated, protein coding genes neighboring RND3 reside ~880 kb (MMADHC) and ~760 kb (RBMA3) away. Hence, the assessment of LD was confined to the three SNPs genotyped within RND3 itself, neither of which were correlated with the UTR-3 rs115015150 SNP (r² < 0.17) (data not shown). The missense GCA rs17783344 SNP was not correlated with any other tested variant in the AUS/NZL families (r² < 0.29) (Fig. 2C).

Of additional relevance to LCT, the genomic region harboring this gene has been subjected to positive selection pressures resulting in lactase persistence in mostly Northern European populations. Lactase persistence in Northern European populations is attributable to the rs4988235 (~13 910*T/C) SNP acting in cis to regulate the promoter activity of LCT (Enattah et al., 2002; Olds and Sibley 2003). Based on our exon-centric sequencing study design we would not have detected rs4988235 (if present) in our AUS/NZL preeclamptic women as this SNP is located within intron 13 of the neighboring MCM6 gene. In light of this and the probable extension of LD (i.e. >500 kb) in the LCT gene region, we extracted Caucasian SNP data from the 1000 Genomes Project (1kGP) database to evaluate the extent of LD with rs2332659 (±1 Mb) (Supplementary data, Table SIII). From a total of 5,732 SNPs in this region (rs1942044 at 135 555 706 bp—rs66519315 at 137 555 352 bp), there were no missense SNPs but 62 synonymous, intronic or intergenic SNPs in strong LD (r² ≥ 0.8) with the missense rs2332659 SNP. These 1kGP LD data are concordant with the AUS/NZL family LD data for five of the seven variants (rs2278544, rs6719488, rs2236783, rs309180, rs687670). The remaining two variants (rs113260867, rs1438307) strongly correlated with rs2332659 in the AUS/NZL families were not present in the 1kGP database. The rs4988235 lactase

![Figure 1](image-url) Successfully genotyped and analyzed 2q22 QTL variants in the AUS/NZL families. The observed measured genotype SNP association, P-values; −log(Observed P) (solid triangles) are superimposed under the linkage signal; LOD score (solid line).

**Table I** SNPs significantly associated with pre-eclampsia in the AUS/NZL familial cohort.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Variant</th>
<th>Function</th>
<th>MjA (%)a</th>
<th>MiA (%)b</th>
<th>HWEp c</th>
<th>MGp d</th>
<th>β* e</th>
<th>ADJp f</th>
<th>EXPpg</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCT</td>
<td>rs2332659</td>
<td>Missense</td>
<td>G (0.764)</td>
<td>A (0.236)</td>
<td>0.07</td>
<td>0.00272</td>
<td>0.35</td>
<td>0.00310</td>
<td>0.0489</td>
</tr>
<tr>
<td>LRP1B</td>
<td>rs35821928</td>
<td>Synonymous</td>
<td>C (0.923)</td>
<td>T (0.077)</td>
<td>0.98</td>
<td>0.00018</td>
<td>0.79</td>
<td>0.00053</td>
<td>0.0184</td>
</tr>
<tr>
<td>RND3</td>
<td>rs115015150</td>
<td>UTR-3</td>
<td>G (0.979)</td>
<td>A (0.021)</td>
<td>0.75</td>
<td>0.00243</td>
<td>-1.41</td>
<td>0.01695</td>
<td>0.0081</td>
</tr>
<tr>
<td>GCA</td>
<td>rs17783344</td>
<td>Missense</td>
<td>A (0.833)</td>
<td>C (0.167)</td>
<td>0.94</td>
<td>0.00210</td>
<td>0.45</td>
<td>0.00319</td>
<td>0.0405</td>
</tr>
</tbody>
</table>

*aMjA (%), observed major allele (frequency).
*bMiA (%), observed minor allele (frequency).
*cHWEp, Hardy–Weinberg equilibrium P-value.
*dMGp, observed measured genotype P-value.
*eRegression coefficient for the rare allele.
*fADJp, gene-centric adjusted P-value.
*gEXPp, experiment-wide P-value.
Figure 2. Measured genotype association results for successfully genotyped variants (solid red triangles) in the AUS/NZL families; $-\log_{10}(P\text{-value})$ (top panel) and corresponding LD heat map; $r^2$ (bottom panel) for (A) the LCT rs2322659 SNP $\pm$ 500 kb, (B) the LRP1B rs35821928 SNP $\pm$ 500 kb and (C) the GCA rs17783444 SNP $\pm$ 500 kb. A physical map of the region is given and depicts other validated protein coding genes in the region on either the sense (→) or anti-sense (←) strand. The down arrow (↓) highlights the significant LCT, LRP1B or GCA SNP, respectively.
persistence/non-persistence regulatory SNP was only moderately correlated with the rs2322659 pre-eclampsia SNP ($r^2 = 0.589$).

**Replication genotyping and association analysis**

Our four experiment-wide significant SNPs in the AUS/NZL families were genotyped in three independent case–control Caucasian populations originating from Australia, Norway and Finland.

**Australian singletons**

We replicated the LCT rs2322659 SNP association (one-sided MGp = 0.02) and observed a borderline genetic association for the GCA rs17783344 SNP (one-sided MGp = 0.07) (Table II). The results attained for the LRPIB (rs35821928) and RND3 (rs115015150) SNPs in the AUS/NZL families were not replicated in the Australian singletons (Table II). The novel GCA nonsense SNP (rs141572054) was also tested in the Australian singletons. We observed two copies of the rare rs141572054-A allele and no genetic association with pre-eclampsia (two-sided MGp = 0.90, MAF = 0.001).

**Norwegian singletons**

The four significant AUS/NZL family SNPs were not replicated in the Norwegian case–control cohort (Table II). We observed 15 copies of the rare GCA rs141572054-A allele in this population and no genetic association with pre-eclampsia (two-sided MGp = 0.59, MAF = 0.002).

**Finnish singletons**

The four significant AUS/NZL family SNPs were not replicated in the Finnish case–control cohort (Table II). We observed 23 copies of the rare GCA rs141572054-A allele in this population and as like the Australian and Norwegian populations, no genetic association with pre-eclampsia (two-sided MGp = 0.51, MAF = 0.008).

**Meta-analysis of pre-eclampsia cohorts**

To determine the total evidence for association we meta-analyzed the observed P-value (MGp) of our four significant SNPs in the discovery AUS/NZL family cohort with the one-sided unidirectional P-values from the replication case–control cohorts from Australia, Norway and/or Finland. The direction of effect (positive beta) for the LCT rs2322659 SNP was the same across all four cohorts and this SNP was significantly associated with higher pre-eclampsia risk (Z-score = 2.76, P-value = 0.0058). The LRPIB rs35821928 SNP was also significantly associated with higher pre-eclampsia risk, but in the AUS/NZL family and Australian case–control cohorts only (Z-score = 2.63, P-value = 0.0085). The RND3 rs115015150 SNP was significantly associated with lower pre-eclampsia risk in the AUS/NZL family and Finnish case–control cohorts only (Z-score = −2.71, P-value = 0.0067). The GCA rs17783344 SNP
was also significantly associated with lower pre-eclampsia risk but in the AUS/NZL family cohort, and the Australian and Norwegian case–control cohorts \( (Z \text{-score} = -2.38, P \text{-value} = 0.0175) \).

**Gene-centric burden tests**

The functional unit within any genome is the gene itself and the collective analysis of gene-centric loci across different populations offers advantages over individual SNP-based or haplotype-based analysis methods. We therefore conducted a gene-centric burden test to jointly analyze all non-synonymous SNPs (nsSNPs) that were identified by exon-centric sequencing and successfully genotyped in our AUS/NZL family data. This individual association test assumes a model of additive gene action for the rare allele with a single degree of freedom. Both GCA \((2 \text{ nsSNPs, } MGp = 0.004, \beta = 0.40)\) and LCT \((3 \text{ nsSNPs, } MGp = 0.014, \beta = 0.16)\) were observed as significant genic functional units. LRP1B was not a significant genic functional unit \((10 \text{ nsSNPs, } MGp = 0.61, \beta = -0.04)\), and no RND3 nsSNPs were identified in the AUS/NZL families.

The strongest genic functional unit \((GCA)\) was not independently associated with pre-eclampsia in either of the Australian \((\text{one-sided } MGp = 0.105, \beta = 0.06)\), Norwegian \((\text{one-sided } MGp = 0.23, \beta = 0.004)\) or Finnish \((\text{one-sided } MGp = 0.50, \beta = -0.06)\) case–control cohorts. However, meta-analysis of the discovery AUS/NZL family, and replication Australian and Norwegian case–control cohorts identified the GCA functional unit \((\text{rs17783344 and rs141572054})\) to be significantly associated with lower pre-eclampsia risk in these three cohorts \((Z \text{-score} = -2.65, P \text{-value} = 0.0081)\). The Finnish case–control cohort was excluded from the GCA burden meta-analysis as this cohort exhibited an opposing direction of effect.

**Bioinformatic analysis of associated SNPs in the AUS/NZL families**

To determine if any of our four associated SNPs in the AUS/NZL families may impact protein folding kinetics, exon splicing regulation, transcriptional regulation and/or post-translational modification(s) we conducted exploratory bioinformatic analyses using F-SNP \((\text{Lee and Shatkay, 2008})\). F-SNP brings together a number of different bioinformatic databases to generate a functional significance (FS) score that predicts queried SNPs to exhibit either neutral (FS score between 0 and 0.5) or deleterious (FS score between 0.51 and 1) characteristics \((\text{Lee and Shatkay, 2009})\). Our replicated LCT \(\text{rs2322659 SNP, in Australian individuals, showed the strongest evidence of demonstrating putative deleterious effects (FS score = 0.774). The non-reference rs2322659-G allele (Serine) is predicted to possess damaging qualities to protein function. This outcome is generated by several databases queried through the F-SNP interface, namely, SIFT protein alignment homologues, SNP effect predicting differences in solvent accessibility and SNPs3D predicting a deleterious impact on protein stability due to the mutant allele (svm profile -1.67). Furthermore, the presence of the rs2322659-G allele creates several putative exon splicing enhancer recognition sequences [SF2/ASF, SF2/ASF (lgM-BRCA1), SRp40], which may result in irregular exon splicing efficacy. As previously noted, rs2322659 is in strong LD with several other nominally associated variants in the AUS/NZL families, of which the prima facie functional classes suggest them to be...}
of lesser FS. Assessment of these variants (rs2278544, rs6719488, rs2236783, rs13260867, rs309180, rs1438307, rs687670) correlated with rs2322659 using F-SNP predicts them to be more neutral than deleterious (FS scores ≤ 0.330). Both the LRP1B rs35821928 and GCA rs17783344 SNPs do present evidence to perturb protein folding kinetics, splicing regulation and/or post-translational modification(s); however, their overall impact was predicted to be more neutral (FS scores ¼ 0.195 and 0.374, respectively). The RND3 rs115015150 SNP, a more recent addition to the public domain, is currently not recognized by F-SNP.

### Pleiotropic effects of candidate pre-eclampsia SNPs upon CVD-related traits

To look for pleiotropic effects upon CVD-related phenotypes the four pre-eclampsia candidate SNPs were genotyped in the SAFHS. The SAFHS is a large cohort of Mexican American families rich in quantitative phenotypic data pertaining, but not limited to, coronary heart disease risk factors, obesity risk factors, blood pressure, diabetes risk factors, inflammatory biomarkers, oxidative stress biomarkers and clotting factors (Mitchell et al., 1996; MacCluer et al., 1999).

The age- and sex-adjusted measured genotype association results (MGp<sup>a</sup>) are presented in Table III. The LCT (rs2322659) SNP was associated with (i) oxidative stress indicators (advanced glycation end products, glutathione reductase), (ii) inflammation (interleukin (IL)-6 protein, soluble vascular cell adhesion molecule 1) and (iii) glucose levels, which are known to be perturbed in diabetes. The LRP1B (rs35821928) SNP was associated with (i) insulin levels, also known to be perturbed in diabetes, (ii) several anthropometric measures of adiposity, known risk factors for obesity, (iii) leptin, another adiposity trait, (iv) total antioxidant status, pertinent to oxidative stress and (v) TNF-alpha, a known pro-inflammatory cytokine. The RND3 (rs115015150) SNP was found only to associate with adiponectin levels, a protein hormone known to modulate glucose metabolism and fatty acid catabolism. The GCA (rs17783344) SNP was found to associate with (i) inflammatory biomarkers (IL-6, IL-8), (ii) insulin levels, (iii) myeloperoxidase, a known toxic by-product of neutrophil degranulation...

### Table III: Pleiotropic effects of candidate pre-eclampsia SNPs upon CVD-related traits in the SAFHS.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Variant</th>
<th>Phenotype group</th>
<th>Trait description</th>
<th>MGp&lt;sup&gt;a&lt;/sup&gt;</th>
<th>β&lt;sub&gt;b&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCT</td>
<td>rs2322659</td>
<td>Oxidative stress</td>
<td>Advanced glycation end products</td>
<td>0.00737</td>
<td>↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inflammation</td>
<td>Interleukin 6 protein</td>
<td>0.00804</td>
<td>↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oxidative stress</td>
<td>Glutathione reductase</td>
<td>0.00892</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diabetes</td>
<td>2 h glucose&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.03728</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inflammation</td>
<td>Soluble VCAM1</td>
<td>0.04204</td>
<td>↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diabetes</td>
<td>Fasting glucose&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.04623</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diabetes</td>
<td>Fasting insulin&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.00036</td>
<td>↓</td>
</tr>
<tr>
<td>LRP1B</td>
<td>rs35821928</td>
<td>Adiposity</td>
<td>Leptin&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.00061</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diabetes</td>
<td>2 h insulin&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.00082</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adiposity</td>
<td>Body mass index</td>
<td>0.00124</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adiposity</td>
<td>Fat pinch test (bicep)</td>
<td>0.00179</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adiposity</td>
<td>Fat free mass (%)</td>
<td>0.00370</td>
<td>↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adiposity</td>
<td>Total body fat (%)</td>
<td>0.00557</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adiposity</td>
<td>Fat pinch test (tricep)</td>
<td>0.02267</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oxidative stress</td>
<td>Total antioxidant status</td>
<td>0.03207</td>
<td>↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adiposity</td>
<td>Waist circumference (cm)</td>
<td>0.04006</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inflammation</td>
<td>TNF-alpha</td>
<td>0.04013</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adiposity</td>
<td>Hip circumference (cm)</td>
<td>0.04211</td>
<td>↓</td>
</tr>
<tr>
<td>RND3</td>
<td>rs115015150</td>
<td>Adiposity</td>
<td>Adiponectin</td>
<td>0.02660</td>
<td>↑</td>
</tr>
<tr>
<td>GCA</td>
<td>rs17783344</td>
<td>Lipoprotein</td>
<td>HDL-C</td>
<td>0.00851</td>
<td>↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Immunity</td>
<td>Myeloperoxidase</td>
<td>0.01132</td>
<td>↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inflammation</td>
<td>Interleukin 6 protein</td>
<td>0.01853</td>
<td>↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inflammation</td>
<td>Interleukin 8 protein</td>
<td>0.02287</td>
<td>↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diabetes</td>
<td>Insulin*</td>
<td>0.02730</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inflammation</td>
<td>C-reactive protein</td>
<td>0.03960</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inflammation</td>
<td>Platelet-activating factor</td>
<td>0.04376</td>
<td>↓</td>
</tr>
</tbody>
</table>

<sup>a</sup>MGp, observed measured genotype F-value.
<sup>b</sup>Direction of regression coefficient for the rare allele; ↑, positive; ↓, negative.
<sup>c</sup>Raw data natural log transformed.
and (iv) HDL-C, a lipoprotein phenotype. However, after accounting for the four SNPs tested across all 57 CVD-risk traits, these pleiotropy association results do not satisfy our Bonferroni-adjusted statistical significance threshold ($P < 0.000219$).

**Discussion**

In this current study we have used a family-based study design to genetically dissect the Australian 2q22 pre-eclampsia susceptibility locus identifying four independent SNPs to be significantly associated with pre-eclampsia in our discovery cohort. These four SNPs reside within four novel pre-eclampsia candidate genes: lactase (LCT, rs2322659), low-density lipoprotein receptor-related protein 1B (LRP1B, rs35821928), rho family GTPase 3 (RRD3, rs115015150) and granulin (GCA, rs17783344). We replicated our LCT SNP association in an independently ascertained Australian case–control cohort only. Failure to replicate the LCT SNP in the Norwegian or Finnish case–control cohorts, or the LRP1B, RRD3 and GCA SNPs in any of the three case–control cohorts is likely due to insufficient power (Supplementary data, Table SIV) or spurious associations in the discovery cohort. Alternatively, these SNP associations may only be specific to our discovery (familial) cohort. Whilst we limit our discussion to the replicated LCT SNP association, the four SNPs pertinent to our discovery cohort also exhibit nominal pleiotropic effects with several quantitative CVD-related traits in an independent population cohort, providing the most substantial empirical evidence to date that CVD and pre-eclampsia share, at least in part, an underlying genetic architecture.

In our earlier positional cloning studies using the same extended set of AUS/NZL families and the same Norwegian case–control cohort only that used in this current study we were able to replicate the ACVR2A (Moses et al., 2006; Fitzpatrick et al., 2009; Roten et al., 2009) and ERAF2 (Johnson et al., 2009) genes, but not specific SNPs themselves, in risk of pre-eclampsia. In this current study we now report a replicated missense SNP association (rs2322659) within the LCT gene in these AUS/NZL families and an independent Australian case–control cohort. We could not replicate this SNP association in the Norwegian case–control cohort, but from our previous observations of allelic heterogeneity we cannot exclude LCT as a pre-eclampsia candidate gene in this population. We have also, for the first time, attempted to replicate our genetic association results from the AUS/NZL families in a Finnish case–control (FINNPEC) cohort. We could not replicate our rs2322659 finding in the FINNPEC cohort even though we did observe near identical or identical minor allele frequencies for this SNP in both the AUS/NZL families (MAF = 0.24) and Australian case–control cohort (MAF = 0.27), respectively. Moreover, the FINNPEC cohort was in HWEE, whereas the Australian case–control cohort was not, leading us to speculate the possibility of a spurious association in these Australian singletons.

We had excellent power (>86%) to detect the LCT SNP association as a function of allele frequency across several estimated disease prevalence estimates in the Australian replication cohort (Supplementary data, Table SIV) and we are confident that our Australian replication cohort exhibits negligible population stratification at a genome-wide (global) level (Johnson et al., 2012). However, we cannot rule out the possibility that the LCT locus-specific (local) region may harbor subtle evidence of population stratification (Kang et al., 2009). Local regions more prone to subtle population stratification effects may be attributable to functional variants residing within chromosomal regions under selective pressures, as evident for the LCT gene region (Bersagliari et al., 2004; Ingram et al., 2009). Furthermore, random fluctuations of admixture more specific to these local regions under strong selective pressures, for example, Northern and Southern European ancestry, may exacerbate these subtle population stratification effects and thus, lead to a spurious association (Campbell et al., 2005). However, the results attained from our Australian pre-eclampsia populations, the SAFHS participants and review of the literature support LCT as a plausible candidate gene for pre-eclampsia and CVD risk. We must also point out that whilst the measured genotype test and subsequent probit-based regression coefficient is reflective of the rare rs2322659-A allele (Asparagine), it is the major rs2322659-G allele (Serine) that is predicted to damage the focal protein, according to the in silico algorithms used in this study. There is a preponderance of the rs2322659-G allele in the Australian pre-eclamptic women over the non-pre-eclamptic women and the genotype/allele frequency data for this SNP is similar to HapMap and 1kGP data of Caucasian (North Western European) ancestry (Supplementary data, Table SV).

The LCT gene encodes the lactase-phlorizin hydrolase enzyme that is responsible for hydrolyzing lactose into glucose and galactose. Lactose malabsorption, consistent with lactase non-persistence, has previously been associated with reduced triglyceride and cholesterol levels which may suggest a role for LCT with CVD risk (Sahi et al., 1977; Russo et al., 1997). This CVD risk hypothesis is further supported by more recent studies identifying LCT specific variants (Silander et al., 2008; Ma et al., 2010), and variants within an expansive region of LD encompassing LCT (Teslovich et al., 2010), to be associated with TC. The FINNRSK cohort (Silander et al., 2007), independent to the FINNPEC cohort used in this study, identified an association between TC and rs6719488, an intronic SNP in our AUS/NZL families nominally associated with pre-eclampsia susceptibility and in strong LD (r² ≒ 0.90) with rs2322659. Data from the Framingham Heart Study identified another intronic LCT SNP (rs2322660) to exhibit a strong association with TC levels (Ma et al., 2010). The rs2322660 SNP was not typed in this current study but the assessment of 1kGP data indicates this SNP to be in strong LD (r² = 0.90) with our pre-eclampsia-associated (rs2322659) SNP (Supplementary data, Table SIII). A comprehensive GWAS of blood lipid traits in more than 100,000 individuals identified rs7570971 as the SNP most significantly associated with TC levels (Teslovich et al., 2010). However, rs7570971 ($\sim$718 kb telomeric to rs2322659) exhibits only modest LD with rs2322659 (1kGP r² = 0.54) (Supplementary data, Table SIII) and based on our recent pre-eclampsia GWAS (Johnson et al., 2012), we expect the 1kGP LD data to be consistent with our data in this current study. Given that TC has been shown to be elevated in women with pre-eclampsia (Staff et al., 1999), data from our Australian pre-eclampsia cohorts, other CVD-related studies and the 1kGP may then suggest SNPs within a smaller region flanking our rs2322659 pre-eclampsia SNP ($\sim$± 200 kb) may exhibit a comorbid role with TC (a CVD risk factor) and pre-eclampsia genetic susceptibility.

Our association results with rs2322659 in the SAFHS participants also lends support to the LCT gene region having a potential role...
with other key pathophysiological features affiliated with pre-eclampsia and CVD risk, namely, oxidative stress indicators (Hubel, 1999; Schmidt et al., 1999; Vanderlee et al., 2005; Chekir et al., 2006; Heistad et al., 2009; Patil et al., 2009) and inflammatory biomarkers (Greer et al., 1994; Austsgulen et al., 1997; Fiotti et al., 1999; Simon et al., 2000; Galen 2002; Chen et al., 2009; Chen et al., 2010). Knowing that oxidative stress and circulating inflammatory biomarkers (e.g. IL-6) are perturbed in obesity, a significant risk factor for pre-eclampsia and CVD (Roberts et al., 2011), it is also of interest to note that rs2322659 and seven other LCT variants displayed nominal evidence for association with waist circumference, a sound obesity indicator (Heard-Costa et al., 2009).

The evidence for the rs4988235 LCT regulatory SNP as a proxy for the pre-eclampsia rs2322659 LCT SNP is moderate at best, therefore, it remains to be elucidated whether rs2322659 perturbs the function of LCT in vitro or in vivo. However, in light of the data from this and other LCT gene studies, this gene region is likely to harbor variants attributable to shared genetic risk factors for pre-eclampsia and CVD. Furthermore, the rs2322659 missense SNP may serve as a prioritized proxy for follow-up and additional pre-eclampsia and other CVD-related studies.

The focus of this study was an exon-centric sequencing strategy that targeted 52 prioritized positional candidate (validated protein coding) genes out of a total of 87. Thus, a sufficient amount of genomic DNA sequence variants in the AUS/NZL families remains unexplored. Given that the LCT (2.3%), LRP1B (3.6%), RND3 (2.4%) and GCA (2.4%) SNPs account for ~10.7% of the total genetic liability (i.e. heritability) in these families, a considerable amount of heritability remains unaccounted for. We are now pursuing next-generation (exome and genome) sequencing strategies to rapidly interrogate unexplored sequence variation in this and other pre-eclampsia susceptibility regions. It is also likely that our observed associated SNPs are in LD with an as yet identified/genotyped variant(s) (± 500 kb) of greater FS with a more pertinent role in pre-eclampsia, CVD-related risk factors or both. This LD may also extend beyond a ± 500 kb window which may be more relevant to variants within the positively selected LCT gene region or the LRP1B gene itself, given its size (~1.9 Mb). Our conservative gene-centric multiple hypothesis testing criteria may also imply that our observed association results are spurious. With the exception of the LCT rs2322659 SNP, it is also possible that the observed SNP associations within LRP1B, RND3 and GCA in our AUS/NZL families are specific only to that cohort and not in the population at large, akin to the STOX1 pre-eclampsia candidate gene in a Dutch familial cohort (Berends et al., 2007).

In conclusion, we herein report the localization of four novel pre-eclampsia candidate genes at the Australian chromosome 2q22 locus. Furthermore, we have shown the pre-eclampsia-associated SNPs within these novel positional candidate genes to exhibit evidence for pleiotropy with several quantitative CVD-related traits, supporting the emerging view that pre-eclampsia and CVD share underlying genetic mechanisms. The putative in silico functional evidence for these SNPs is theoretical at best and formal molecular studies (beyond the scope of this study) are required to confirm or refute their biological significance. However, given the complexity and heterogeneity of the pre-eclampsia phenotype, our associated SNPs, or their proxies, are likely to play a role in an intricate biological network housing numerous variants contributing to the genetic architecture of pre-eclampsia.

**Supplementary data**

Supplementary data are available at http://molehr.oxfordjournals.org/.

**Acknowledgements**

We thank Mr Hao-Chang Lan, Mrs Yvonne Garcia, Ms Janelle Bentz (Texas Biomed) and Mr Anthony Borg (RWH, Melbourne) for technical assistance. We sincerely acknowledge the support of the clinical research mid-wives Karen Reidy, Sue Duggan, Sue Nisbett and Moira Stewart (RWH, Melbourne) who contributed to this study. We also acknowledge the efforts of Ms Divya Neelam (Sequenom) for Sequenom assay design assistance and Mr Augustin Luna for technical assistance with snp.plotter. The HUNT study is a collaborative effort between the 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, Finnish and Mexican American individuals whose participation made this work possible.

**Authors’ roles**

M.P.J.: study design, data generation, data analysis, interpretation of results, wrote the manuscript and approved final version of the manuscript. S.P.B. responsible for the Australian samples, conceived the study, study design, reviewed Australian sample medical records and approved final version of manuscript. C.E.E.: reviewed Australian sample medical records and approved final version of the manuscript. T.D.D., J.M.P and P.E.M. contributed to data analysis, interpretation of results and approved final version of the manuscript. L.T.R. contributed to data generation, aided in manuscript writing and approved final version of the manuscript. T.A.-V. contributed to data generation, aided in manuscript writing and approved final version of the manuscript. M.H.F. contributed to data generation, FINNPEC Study Core Investigator and approved final version of the manuscript. K.M. contributed to data generation, FINNPEC Study Core Investigator, approved final version of the manuscript. S.H.: FINNPEC Study Core Investigator, approved final version of the manuscript. E.K. conceived FINNPEC Study, FINNPEC Study Core Investigator, approved final version of the manuscript. J.K. conceived FINNPEC Study, FINNPEC Study Core Investigator, approved final version of the manuscript. H.L. responsible for the Finnish samples, conceived FINNPEC Study, FINNPEC Study Core Investigator, approved final version of the manuscript. J.B. responsible for the Mexican American samples, conceived the study, study design, supervised all data analysis, interpretation of results, aided in manuscript writing and approved final version of manuscript. E.K.M. responsible for the Australian samples, conceived the study, study design, interpretation of results, supervised all elements of the study, aided in manuscript writing and approved final version of the manuscript.
Funding

M.P.J. is supported, in part, by an American Heart Association National Scientist Development grant [095DG23500008]. This work was supported by the National Institutes of Health [HD049847 to E.K.M., S.P.B. and J.B., MH059490 to J.B., S10 RR029392 to J.B.]; the Semp Russ Foundation of the San Antonio Area Foundation (to M.P.J.); and the Texas Biomed Forum (to M.P.J.). The AT&T Genomics Computing Center at Texas Biomed is supported by the AT&T Foundation. The Norwegian cohort study was supported by the Liaison Committee of Central Norway Regional Health Authority and NTNU (to L.T.R. and M.H.F.) and the Functional Genomic Programme (FUGE) of the Norwegian Research Council (to R.A.). The FINNPEC study was supported by Jane and Aatos Erkko Foundation, Paiviikki and Sakari Sohlberg Foundation, Academy of Finland, Research Funds of the University of Helsinki, Government Special Subsidy 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. This investigation was conducted, in part, in facilities constructed with support from Research Facilities Improvement Program [grants C06 RR13556 and C06 RR017515] from the NIH.

Conflict of interest

None declared.

References

Løset M, Mundal SB, Johnson MP, Fenstad MH, Freed KA, Lian IA, Lee PH, Shatkay H. An integrative scoring system for ranking SNPs by their functional importance.

Lee PH, Shatkay H. F-SNP: computationally predicted functional SNPs.


Moses EK, Fitzpatrick E, Freed KA, Dyer TD, Forrest S, Elliott K, Johnson MP, Blangero J, Brennecke SP. Objective prioritization of positional candidate genes at a quantitative trait locus for pre-eclampsia on 2q22.

Moses EK, Fitzpatrick E, Freed KA, Dyer TD, Forrest S, Elliott K, Johnson MP, Blangero J, Brennecke SP. Objective prioritization of positional candidate genes at a quantitative trait locus for pre-eclampsia on 2q22.


Moskvina V, Schmidt KM. On multiple-testing correction in genome-wide association studies.


Roberts JM, Cooper DW. Pathogenesis and genetics of pre-eclampsia. Lancet 2001; 357:53–56.


---

**Appendix**

**The Finnish Genetics of Pre-eclampsia Consortium (FINNPEC):** We appreciate the collaboration with the following members of the FINNPEC Study Group: Anneli Pouta (National Institute for Health and Welfare, Oulu, Finland), Katja Kivinen (Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK), Eeva Ekholm (Turku University Central Hospital, Turku, Finland), Reija Hietala, Susanna Sainio and Terhi Saisto (Helsinki University Central Hospital, Helsinki, Finland), Jukka Uotila, (Tampere University Hospital, Tampere, Finland), 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.