January 2015

Identification Of Fetal Genomic Copy Number Deletions Associated With Preeclamptic Mothers In The Caucasian Population

Tze-An Yuan
Yale University, joannyuan@gmail.com

Identification of Fetal Genomic Copy Number Deletions Associated with Preeclamptic Mothers in the Caucasian Population

Tze-An Yuan

Master of Public Health in Chronic Disease Epidemiology

Yale University, School of Public Health

Advisor: Andrew Thomas DeWan PhD, MPH

Associate Professor of Chronic Disease Epidemiology

Co-Advisor: Josephine Hoh PhD

Associate Professor of Environmental Health Science and Ophthalmology and Visual Science

May 2015

Abstract

Genetic determinants that are associated with preeclampsia still remain elusive nowadays, which might have resulted from limited understanding of the underlying pathogenesis mechanisms and the fetal contributions. By using genome-wide association study approach, we identified 2 candidate copy number deletions of the highest odds ratio ranking on chromosome 1 and 19, from 283 US white babies born by the mothers with preeclampsia. Based on the molecular and biological functions of the genes entailed the deleted regions, which contain CTSK, ARNT, and ACTN4, we proposed that these two candidate genomic deletions might be associated with preeclampsia. In a further attempt to replicate our findings in an independent dataset that contains 1200 white European babies, we did not find genes matched with our highest ranking candidates; however, one specific gene RYR1 deletion was identified in both datasets, and it can also be found in an established genetic database of preeclampsia genes. Our study was aimed to better understand the etiology of preeclampsia and in hopes to develop better screening tools based on genetic variants carried by the predisposed mother or fetus in the coming future.

Acknowledgments

My sincerest thanks to those people who took time from their own busy schedules to assist me in developing this MPH thesis: Andrew Thomas DeWan PhD, MPH; Josephine Hoh PhD; Leyao Wang PhD, MPH; Linlu Zhao PhD; Kai Wang PhD; Stephan Sanders PhD; William Murk MPH; and Yasmmyn Salinas MPH.

Table of Contents

Abstract----------**i**

Acknowledgements----------**ii**

Introduction----------**1**

Materials and Methods----------**3**

Ethics statement----------3

Study population----------3

Definition of preeclampsia----------4

GWAS genotyping----------5

SNP analysis----------5

Detection of copy number variants----------6

Prioritization of copy number variants----------7

Results----------**8**

SNP Association----------8

CNV Detection----------8

Tables and Figures----------**10**

Discussion----------**21**

Reference----------**27**

List of Tables

Table 1 Association (p-value) of the top 2 SNP candidates with children of preeclampic mothers----------11

Table 2 Prioritized copy-number duplicated regions in children of PE case and control mothers----------14

Table 3 The most interested copy-number deleted candidates in children of PE case and control mothers----------15

List of Figures

Figure 1 Flow chart of the SOPHIA study subject recruitment process----------10

Figure 2 Process of drawing eligible SOPHIA study subjects----------12

Figure 3 Prioritizing copy-number variable regions----------13

Figure 4 The probe fluorescent intensity of a copy-number deleted region from a child of PE case mother----------16

Figure 5 UCSC Genome Browser plot of the copy-number deleted region (Chr1:149039031-149248927)----------17

Figure 6 UCSC Genome Browser plot of the copy-number deleted region (Chr19:43835770-43835935)----------18

Figure 7 Process of drawing eligible HAPO study subjects----------19

Figure 8 Process of generating CNVs in the HAPO study to find possible replicates of the SOPHIA study----------20

Introduction

Among pregnant women in the US, approximately 6-8% develop high blood pressure annually; the underlying hypertensive disorders are the top cause of maternal death during pregnancy[1]. Preeclampsia (PE) is one of the serious hypertensive disorders that are characterized by newly onset hypertension and proteinuria, which usually starts after 20 weeks of gestation, and occurs in about 5-8% of pregnancies worldwide[2]. When preeclampsia further progresses to eclampsia (seizure), the condition is highly dangerous and can be fatal to both mother and the baby. There is still no way to predict and prevent preeclampsia, and the only known cure so far is delivery of the baby and complete removal of the conception products, which also turns out to be the leading cause of preterm birth nowadays[3].

Preeclampsia has been shown in several studies with familial heredity[4-6]. However, the underlying genetic contributors remained elusive, and how these vary among racial groups is poorly understood. A number of candidate genes that have been indicated in the pathogenesis pathways involved in preeclampsia, including those related to immune dysfunction[7], placenta ischemia, hypoxia or oxidative stress[8, 9]. Even though many candidate gene approach studies have been performed, the findings are inconclusive[10], probably because of an still incomplete understanding of the underlying biological mechanisms involved in preeclampsia pathogenesis pathways to date[11]. Moreover, genome-wide association studies today mainly focus on single nucleotide polymorphisms (SNPs), which are the most frequent point mutations but only represented one part in the complicated genetic variations in associated with diverse diseases. Considerations that are specifically in the maternal genetic contributions but limited studies from the fetus side created barriers in understanding of preeclampsia heritability as well. There is now accumulated evidence from both epidemiological and genetic studies suggest a fetal role in the

etiology of preeclampsia[5, 6]; family study models also pointed out fetal determinants in susceptibility to preeclampsia[7, 12]. A large epidemiological study from Sweden unraveled role of genes expressed by the fetus in preeclampsia[13]. In the results they concluded that while 35% of the genetic variability in preeclampsia was attributable to maternal genetics alone, 20% could be attributed to fetal genetics alone, with the remainder unaccounted for. Nevertheless, we are still lacking published candidate gene association studies that have examined the role of specific fetal genetic variants on risk of preeclampsia.

DNA structural variation—the copy number variations (CNVs)— is responsible for human evolution and is crucial for creating genetic diversity between individuals[14]. Although copy number variations are less common than SNP point mutations, the structural variation can lead to larger functional effects and make it susceptible to disease traits. Copy number variations involved deletions or duplications at distinct regions on the genome. They can functionally alter gene rearrangement by several mechanisms, including gene interruption, gene fusion, or alteration at the regulatory regions, which may lead to dose-sensitive genomic disorders. Several diseases have been found to be associated with copy number variations, such as sporadic chromosomal microdeletion Prader-Willi and DiGeorge syndromes[15], and therefore copy number variations might be the missing puzzle that can help us dissecting the complex heritability of preeclampsia.

Genome-wide association study (GWAS) is a hypothesis-free approach for uncovering the genetics of preeclampsia. Built upon our preliminary genome-wide study of maternal copy number variants in associated with preeclampsia[16], our current study is conducted to identify potential fetal SNPs and copy number variations that are associated with preeclampsia in the mother using a case-control study design.

Materials and Methods

Ethics statement

We received approval from the University of Iowa Institutional Review Board and the Yale University Human Investigation Committee.

Study population

In order to maintain homogeneity in the preeclampsia phenotype, all of the children enrolled in our study were delivered by white Caucasian mothers of primiparous, without history of pre-existing hypertension, and delivered singleton live births. The mothers of the children were originally recruited for the Study of Pregnancy Hypertension in Iowa (SOPHIA), which was a case-control approach designed to examine the effects of maternal-fetal HLA sharing on risk of preeclampsia. These mothers were initially identified for the SOPHIA study through birth certificates provided by the Iowa Department of Public Health within 9 months of delivery, from August 2002 to May 2005. All mothers who were recorded positive on the birth certificate for pregnancy-induced hypertension or eclampsia were considered potential mothers of the cases, and a random sample of the mothers who were recorded negative were considered as potential mothers of the controls. Those who consented were further determined for initial eligibility. For both cases and controls, exclusion criteria included mothers' age less than 18 years at delivery; non-English-speaking; history of an autoimmune disease (systemic lupus erythematosus, type 1 diabetes mellitus, rheumatoid arthritis, etc.); recurrent spontaneous abortion (more than 3 sequential pregnancy losses); chronic hypertension; multiplegestation; major congenital anomalies; and infant death. Mothers who met initial eligibility criteria were invited to participate in the SOPHIA study and asked to complete an extensive telephone interview collecting information on family history of cardiovascular and renal diseases and preeclampsia,

and a detailed sexual history with the children's father. A package containing buccal cell collection kits for the mother and her child, the medical record release forms and consent forms were mailed to the mothers, and they were asked to return the signed forms with buccal cell samples as soon as possible in the pre-paid postage envelope. DNA was extracted from the buccal samples and stored for analysis. Medical records were abstracted using a structured abstraction form collecting detailed information on blood pressure and urinary protein values during the prenatal, intrapartum, and immediate postpartum periods, along with information on potential exclusion criteria (diabetes, SLE, preexisting hypertension, etc.) for both mothers of cases and controls. To identify cases' mothers, preeclampsia was defined according to the National Heart, Lung and Blood Institute (NHLBI) guidelines. Based on chart review, 288 study subjects, including 183 children of preeclamptic mothers and 105 children of non-preeclamptic mothers met the strict case and control criteria (Figure 1). Regarding to replicating our findings discovered in the SOPHIA study, we used white European subjects from the HAPO (Hyperglycemia and Adverse Pregnancy Outcome) study as to confirm our results. Detailed PE phenotypic definition and how we drawn our eligible study subjects (Figure 7) have been described in our previous publication[17].

Definition of preeclampsia

Criteria of preeclampsia were defined by the National Heart, Lung and Blood Institute (NHLBI) guidelines. The criteria required: 1) newly onset hypertension (>140mmHg systolic or >90mmHg diastolic on two or more occasions at least six hours apart beginning after the 20th week of gestation; 2) accompanied with proteinuria, defined as urinary protein concentrations of 30mg/dL or greater (equivalent to dipstick value of 1+ from two or more specimens collected at least four hours apart, or one or more urinary dipstick values of 2+ near the end or pregnancy, or

one or more catheterized dipstick values of 1+ during delivery hospitalization, or 24-hour urine collection with protein of ≧300mg. We excluded those children whose mothers for whom pre-existing hypertension could not be ruled out (ie. no readings available prior to 20th weeks of gestation; medical charts indicating chronic hypertension in the patient); or whom only met partial criteria for preeclampsia (pregnancy-induced hypertension; proteinuria without hypertension); or whom had incomplete information so a definitive diagnosis could not be made.

For those mothers of the control children, status was defined as a healthy, term delivery without evidence of maternal high blood pressure in the prenatal, intrapartum, or postpartum periods, no chronic hypertension, and no proteinuria (1+ on dipstick on two or more occasions).

GWAS genotyping

DNA samples, which previously isolated from buccal cells and stored at -20℃, were arrayed on 96-well plates, 1μg of DNA at 50ng/μL, submitted to the Yale Center for Genome Analysis for processing, using the Illumina Human OmniExpress microarray for a total genetic probes of 731,442 on each chip.

SNP analysis

PLINK[18] analysis toolset was applied for genetic marker quality control. We first excluded those individuals with genotyping call rate lower than 95% (n= 2 cases and 3 controls). SNPs were next examined and removed if they met any of the following criteria: 1) not having a genotyping call rate of 98% in either the cases or controls; 2) minor allele frequency (MAF) was less than 2%; 3) violation of the Hardy Weinberg Equilibrium (HWE) as p-values less than 1 x 10^{-7}. Q-Q plot showed no evidence of population stratification (no significant deviation, data not shown). Finally we had 283 study subjects (181 cases and 102 controls) and 647,382 SNPs to carry out the association analysis. Individual SNP was tested for both allelic and genotypic

associations by calculating Fisher's exact p-values and using a strict Bonferroni-corrected genome-wide significant threshold at 7.72×10^{-8} ($\alpha = 0.05/647,382$).

Detection of copy number variants

We applied three algorithms to insure higher quality of CNV calls (Figure 2). The PennCNV[19](June 16, 2011 version), QuantiSNP[20](version 2.3 beta) and GNOSIS (bundled in CNVision[21] version 1.73) were used for CNV calls. Hidden Markov model based transitional probabilities were behind PennCNV and QuantiSNP algorithms that integrated log R ratio (LRR: a measure of total signal intensity of probes) and B allele frequency (BAF: a measure of relative intensity ratio of allelic probes) to infer CNV calls for individual genotyping sample. GNOSIS implements a continuous distribution function in sliding windows along the chromosomes to detect CNVs by using log R ratio and B allele frequency as well. Detection algorithms' command lines were run under default settings for Illumina HumanOmniExpress-12 v1.1 B microarray; LRR and BAF raw fluorescent intensities were generated and exported from Illumina GenomeStudio software.

The large number of CNV calls may reflect low DNA sample quality and false positive discovery[22], therefore stringent quality control criteria are essential to maintain valid CNV calls. A single CNV might be detected multiple times due to fragmentation[23], so we excluded those CNVs called by less than ten consecutive SNP or CNV probes first, as PennCNV suggested default CNV call QC criterion. We then $\log_{10}$-transformed the right-skewed distribution sample calls for each algorithm to correct non-normality. A study subject was considered to have failed CNV calls quality control and removed from further analysis (n = 7 cases and 5 controls) if he/she had an extremely large number of CNV calls, defined as having CNV calls greater than mean+3SD of the $\log_{10}$-transformed number of calls, presented in at least one of the three

algorithms.

CNVision[24] was used to analyze the merged CNV calls from PennCNV, QuantiSNP and GNOSIS. The merge function of CNVision combined CNV calls made by all three algorithms that had overlaps of ≥1 base pair and determined the percentage of call overlaps from specific algorithms. Merged CNV calls were excluded (n = 43,172) if one of the following conditions were met: 1) both deletion and amplification calls were made concurrently by different algorithms at the same overlap region; 2) <50% overlap between two algorithms and <25% overlap among three algorithms (Figure 3).

Prioritization of copy number variants

The recurrent function of CNVision screened copy number varied regions to extract those overlaps across our study subjects that were presented more frequently in the cases, and those overlapped regions containing less than five consecutive SNP or CNV probes were excluded along the process (Figure 3). We used a subjective prioritization strategy to further generate targeted CNV regions that were mostly enriched among cases: CNVs with odds ratios ≥ 2.25 comparing children of preeclamptic mothers to children of non-preeclamptic mothers, or CNVs called in ≥3 cases but 0 in controls were then selected as candidates. Recurrent regions overlapping centromeric (±100kb) and telomeric (±500kb within the start or end region of a chromosome) regions were removed by using PennCNV bundled filtering script. We also checked with public available online databases to exclude non-unique copy number polymorphisms that are present in the general and healthy population. (Database of Genomic Variants: http://dgv.tcag.ca/dgv/app/home; The Copy Number Variation Project: http://cnv.chop.edu).

Results

SNP Association

A total of 283 genotyped study subjects (181 children of PE case mothers and 102 children of control mothers) passed the default quality control call rate threshold (≧95%), with the overall sample call rate reaching 99.94%. No SNPs surpassed the Bonferroni-corrected genome-wide significant threshold of 7.72×10^{-8} for the Fisher's exact genotypic or allelic tests. The top 2 SNP candidates had a genotypic p-value less than 10^{-5} (Table 1).

CNV Detection

After merging CNV calls from all three CNV calling algorithms by using CNVision, a total of 4,388 overlapped CNVs across genome among 174 children of PE case mothers and 97 children of control mothers met initial inclusion criteria. The length of these CNVs ranged from 24 bp to 24 Mb with an average of 65 kb. Of these CNVs, 3,334 showed deletions (including homozygous and heterozygous deletions), and 1,054 were duplications. The CNVs were further screened to extract those overlaps across our study subjects that were mostly enriched among the cases, and those overlapped regions containing less than five consecutive SNP or CNV probes were excluded along the process; 608 deleted regions and 279 duplicated regions were then identified.

Among the 608 deleted regions, 25 candidates surpassed odds ratio ≥ 2.25 comparing children of preeclamptic mothers to children of non-preeclamptic mothers, or CNVs called in ≥3 cases but 0 in controls, as described in the method section; correspondingly, 10 copy-number duplicated regions stood out to be candidates. We confirmed these candidate regions with the public available copy number variation databases online (http://cnv.chop.edu; http://dgv.tcag.ca/dgv/app/home) to exclude non-unique copy number polymorphisms that were

present in the general and healthy population[25]. Of the 25 deleted regions, 20 were left to be unique or not yet identified rare or low frequency CNVs; of the 10 duplicated regions, only 4 were left to be unique CNVs (Table 2).

Since loss of function of the protein products by homozygous or heterozygous deletions of certain genes are more frequent and severe than gains of function by slightly duplicated copies of the genes[26, 27], we thus made deletions as a prioritized resource for exploring the genetic and biological basis of PE phenotypic variation. Of the 20 deleted regions, we discovered two interesting regions on chromosome 1q21.2 and 19q13.2 with the highest odds ratio rankings, containing ARNT, CTSK and ACTN4 annotated genes. Based on the molecular and biological regulatory functions of these genes, we proposed their associations with preeclampsia pathogenesis (Table 3).

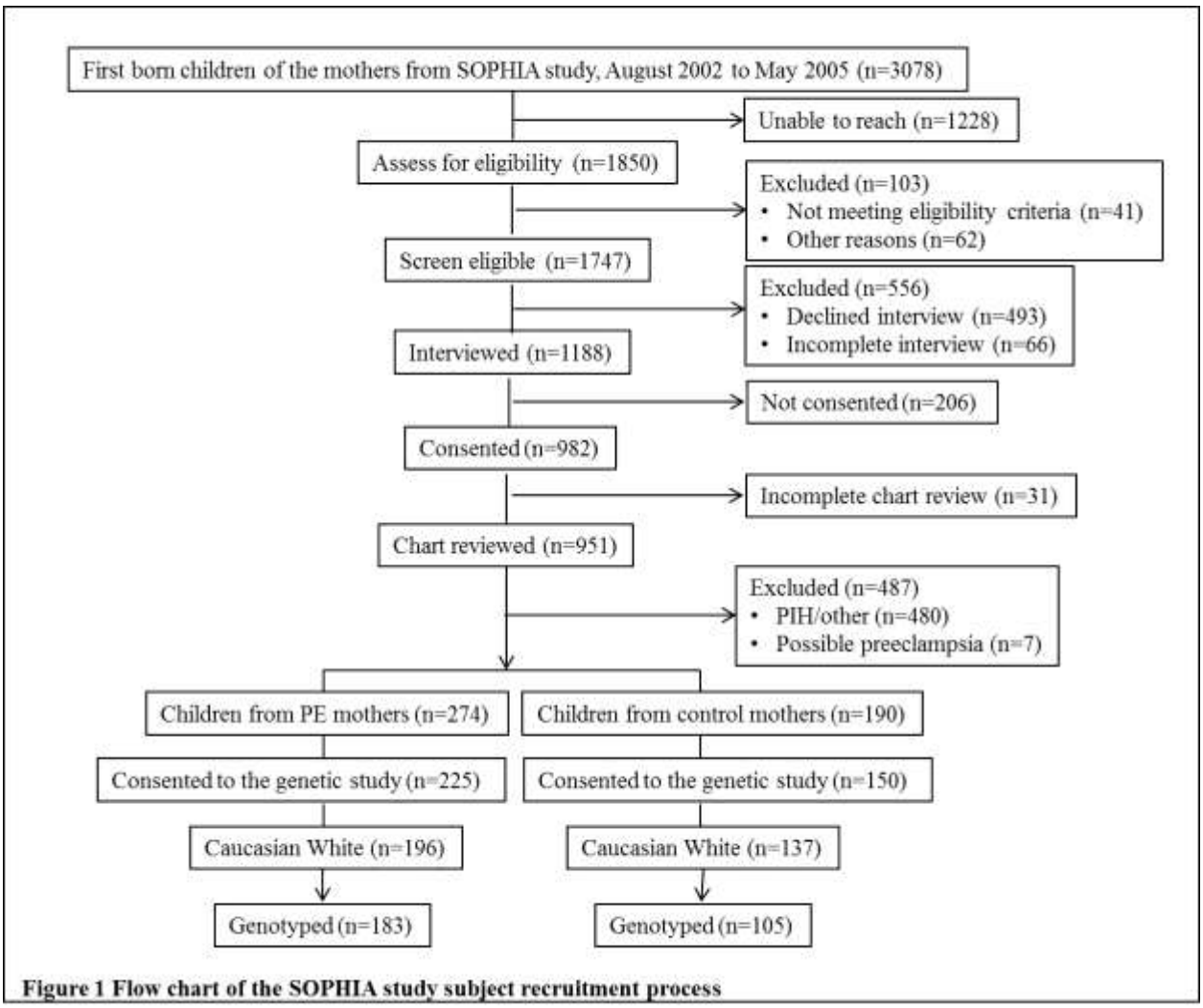

Figure 1 Flow chart of the SOPHIA study subject recruitment process

Table 1 Association (p-value) of the top 2 SNP candidates with children of preeclampic mothers[a]

SNP[b]	Minor allele frequency (%)		Fisher's exact p-value		Chr.	Position	SNP type	Closest gene	Distance to gene (kb)
	Cases (n=181)	Controls (n=102)	Genotypic	Allelic					
rs359937	24.6	20.1	4.62×10^{-6}	2.53×10^{-1}	1	89982719	intergenic	LRRC8B	7.2
rs11627903	34.8	48.0	5.04×10^{-6}	2.33×10^{-3}	14	89484667	intergenic	TRNAA17	39

[a]SNPs with genotypic p-values less than 10^{-5}.
[b]Ordered according to smallest to largest genotypic Fisher's exact p-values.

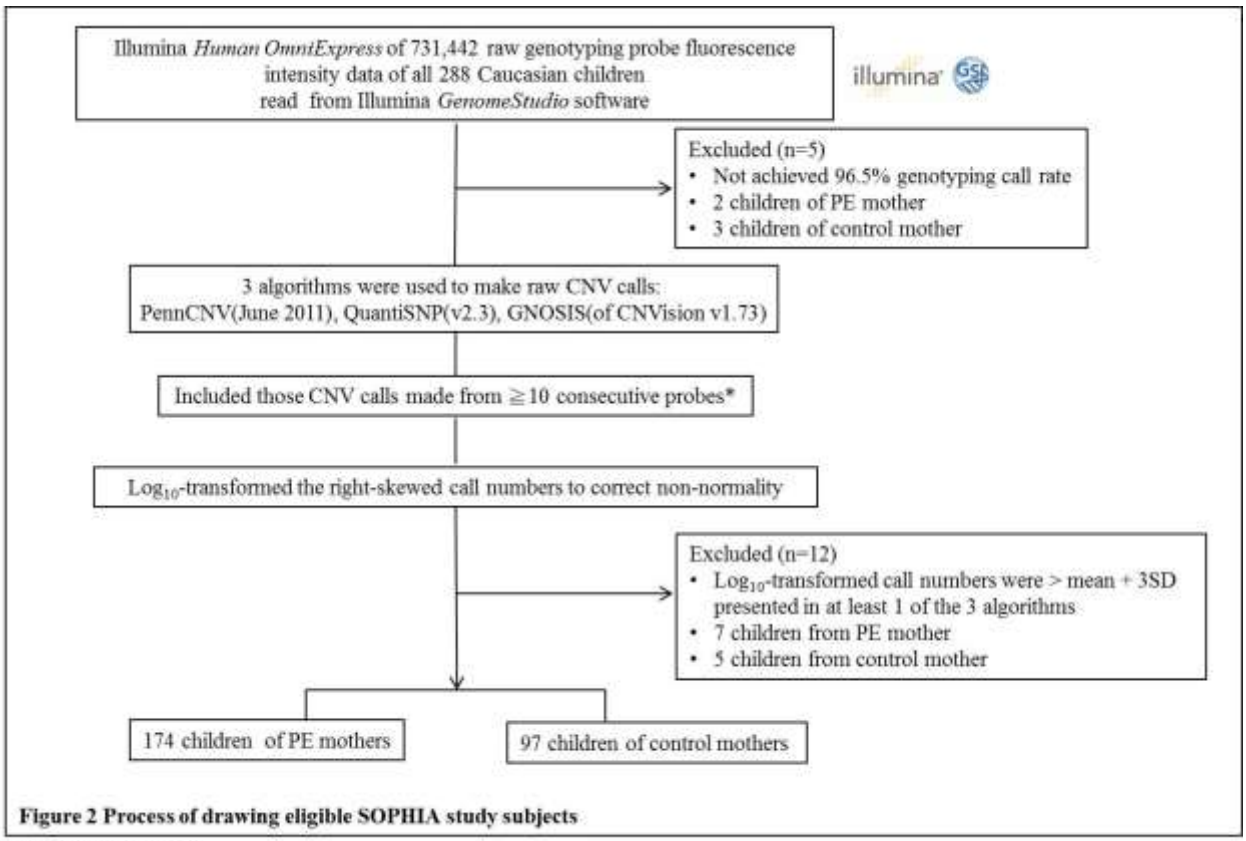

Figure 2 Process of drawing eligible SOPHIA study subjects

*PennCNV suggested QC

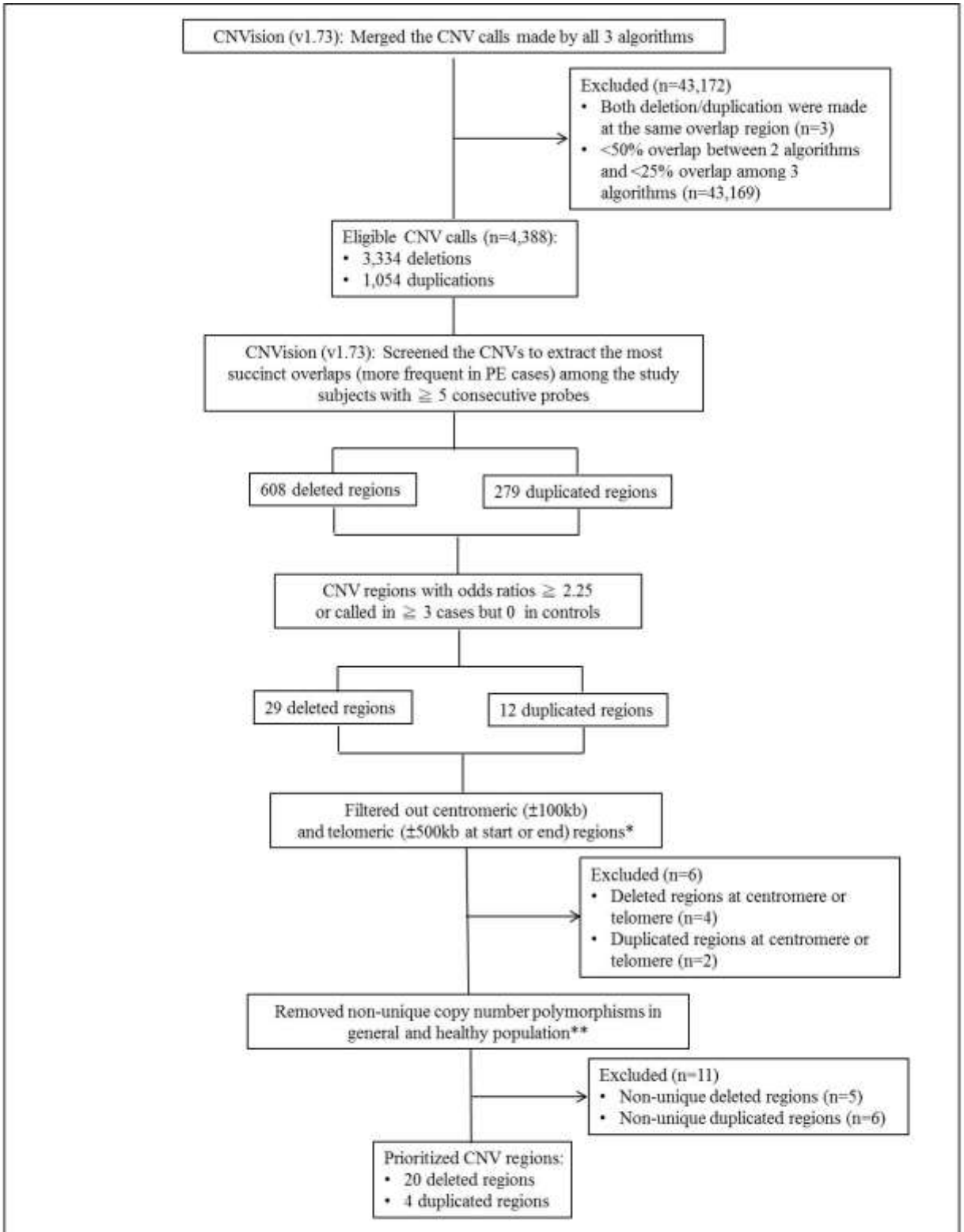

Figure 3 Prioritizing copy number variable regions

*PennCNV suggested QC

**The Copy Number Variation project at the Children's Hospital of Philadelphia : http://cnv.chop.edu; Database of Genomic Variants: http://dgv.tcag.ca/dgv/app/home

Table2 Prioritized copy-number duplicated regions in children of PE case and control mothers

CNV region[a]	CNV frequency[b]			CNV type	Gene contents	Gene full names	Gene functions	References
	Cases	Controls	Odds Ratio					
Chr12p13.31: 8102386-8114429	4/174	1/97	2.26	Exon	C3AR1	Complement component 3a receptor 1	Receptor for the chemotactic and inflammatory peptide anaphylatoxin C3a. This receptor stimulates chemotaxis, granule enzyme release and superoxide anion production.	Eur. J. Immunol. 1996; 26:1944-1950
Chr17q21.32: 44165803-44169808	4/174	1/97	2.26	Intergenic	5'gene: HOXB13	Homeobox B13	Transcription factor that plays a role in fetal skin development and cutaneous regeneration. Mutation has been implicated to associate with increased risk of breast and prostate cancer.	Tumour Biol. 2014; 35(2):1177-82. Cancer Res. 2013; 73(17):5449-58.
					3'gene: TTLL6	Tubulin tyrosine ligase-like family, member 6	Polyglutamylase which preferentially modifies alpha-tubulin. Mediates tubulin polyglutamylation in cilia. Involves in the side-chain elongation step of the polyglutamylation reaction rather than in the initiation step.	Nat Genet. 2012; 44(2):193-9
Chr17q21.32: 44169808-44285531	5/174	1/97	2.84	Exon	TTLL6	See as above		
					CALCOCO2	Calcium binding and coiled-coil domain 2	Plays an important role in innate immunity by mediating macroautophagy.	Autophagy. 2013; 9(8):1256-7. Autophagy. 2013; 9(5):784-6
Chr17q21.32: 44285531-44368212	5/174	0/97	N/A	Exon	CALCOCO2	See as above		
					ATP5G1	ATP synthase F(0) complex subunit C1	Mitochondrial ATP synthase catalyzes ATP synthesis, utilizing an electrochemical gradient of protons across the inner membrane during oxidative phosphorylation.	Cell Cycle. 2013; 12(4):674-83.
					SNF8	ESCRT-II complex subunit	Component of the endosomal sorting complex required for transport II, which is required for multivesicular body(MVB) formation and sorting of endosomal cargo proteins into MVBs. The MVB pathway mediates delivery of transmembrane proteins into the lumen of the lysosome for degradation.	BMC Cell Biol. 2012;13:33.
					UBE2Z	Ubiquitin-conjugating enzyme E2Z	Encodes an enzyme which ubiquitinates proteins that are participated in signaling pathways and apoptosis.	Mol Biol Rep. 2007; 34(3):183-8.

[a]Fused copy number deleted regions to generate the most succinct overlaps that were more frequent in PE cases among our study subjects
[b]Frequencies were derived from the genotyping microarray data

Table3 The most interested copy-number deleted candidates in children of PE case and control mothers

CNV region[a]	CNV frequency[b]			CNV type	Gene contents	Gene full names	Gene functions	References
	Cases	Controls	Odds Ratio					
Chr1q21.2: 149039031-149248927	4/174	1/97	2.59	Exon	**CTSK**	Cathepsin K	Closely involved in osteoclastic bone resorption and may participate partially in the disorder of bone remodeling. Has been found to enhance invasion of breast cancer cells.	Bone. 2014; 59:122-6. Clinical cancer research. 2008; 14:5357-67.
					ARNT	Aryl hydrocarbon receptor nuclear translocator	Heterodimer with HIF-1α ; involved in regulation of hopoxic trophoblasts.	Scientific reports. 2013; 3:2407. JBC. 2004; 279:16128-35. Epigenetics. 2013; 8:192-202.
					SETDB1	SET domain bifurcated 1	Histone methyltransferase which regulates histone methylation, gene silencing, and transcriptional repression. This gene has been identified as a target for treatment in Huntington Disease.	Genes Dev. 2002; 16(8):919-32.
					CERS2	Ceramide synthase 2	Suppresses the growth of cancer cells. May be involved in sphingolipid synthesis	PNAS. 2014; 111(15):5682-7.
					ANXA9	Annexin A9	The annexins are a family of calcium-dependent phospholipid-binding proteins.	FEBS Lett. 2012; 586(19):3090-6.
					FAM63A	Family with sequence similarity 63, member A	Diseases associated with FAM63A include retinal detachment, and kidney disease.	Nat Genet. 2010; 42(5):376-84. Hum Mol Genet. 2013; 22(15):3174-85.
					PRUNE	Prune exopolyphosphatase	Plays a role in cell proliferation; overexpression is related to advanced cancer status.	Oncogene. 2001; 20(47):6881-90. Clin Cancer Res. 2005; 11(1):199-205.
Chr19q13.2: 43835770-43835935	9/174	2/97	2.26	Intron	**ACTN4**	Alpha -actinin-4	This gene encodes a nonmuscle, alpha actinin isoform which is concentrated in the cytoplasm, and thought to be involved in metastatic processes. Mutations in this gene have been associated with focal and segmental glomerulosclerosis (scarring of blood vessels in the kidneys).	Biochim Biophys Acta. 2014; 1842(7):1028-40.
					RYR1	Ryanodine receptor 1	Family of ryanodine receptors, which form channels that transport positively charged calcium ions within cells. Channels made with the ryanodine receptor 1 protein play a critical role in skeletal muscles used for movement. Central core disease, the first autosomal dominant congenital myopathy discovered in human.	Biochem J. 2014; 460(2):261-71. Brain. 2006; 129(6): 1470-1480.
					MAP4K1	Mitogen-activated protein kinase 1	Plays a role in cell growth regulation. Downregulation inhibits cancer cell invasion.	Mol Cell Biol. 2006; 26(4):1297-306.
					EIF3K	Eukaryotic translation initiation factor 3, subunit K	Required for the initiation of protein synthesis.	FEBS Lett. 2004; 573(1-3):139-46. J Cell Sci. 2008; 121(Pt 14):2382-93

[a]Fused copy number deleted regions to generate the most succinct overlaps that were more frequent in PE cases among our study subjects
[b]Frequencies were derived from the genotyping microarray data

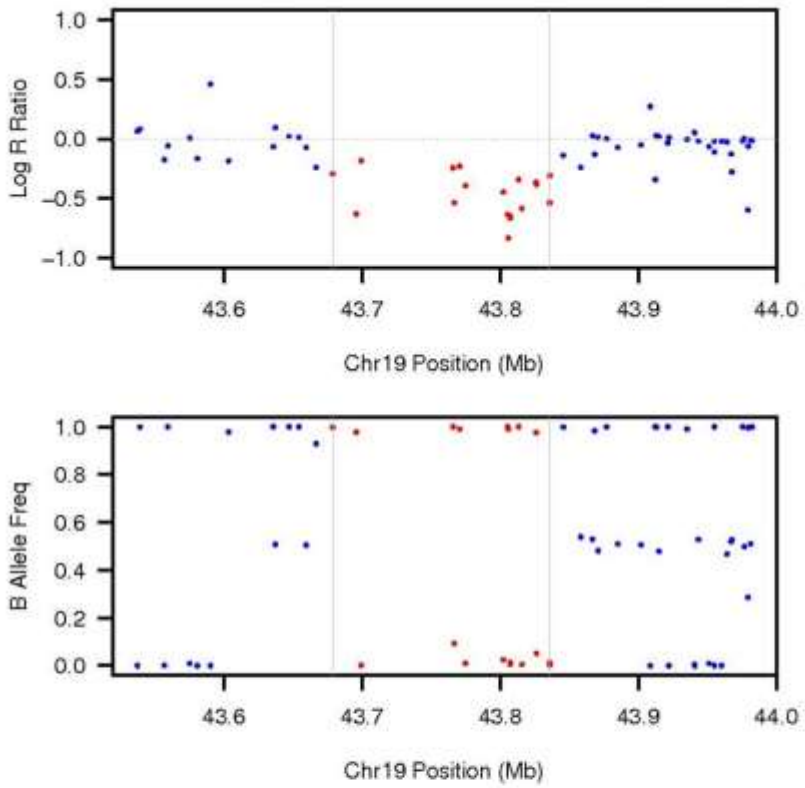

Figure 4 The probe fluorescent intensity of a copy-number deleted region from a child of PE case mother. Dots represented the genetic probes from Illumina Human OmniExpress microarray chip. The vertical gray lines circumvented the area that detected the deletion (chr19:43835770-43835935) from an algorithm. The log R ratio values for the genetic probes in this deletion (red dots) area dropped to -0.5 and the B allele frequency values for the genetic probes clustered randomly around 0 or 1. On the contrary, the adjacent normal copy regions have the log R ratio values centered at zero with three clear B allele frequency clusters (blue dots).

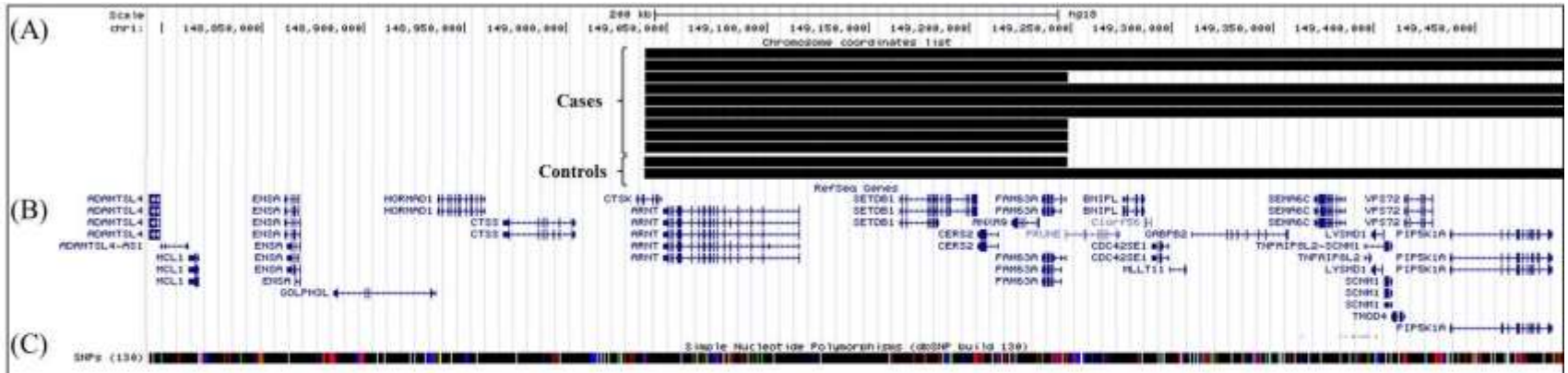

Figure 5 UCSC Genome Browser plot of the copy-number deleted region (Chr1:149039031-149248927). (A) Red rectangular delineated the most succinct overlaps across our study subjects containing the deleted region. Each black horizontal bar represented the length and breakpoints of a deletion detected in the children of PE mothers or control mothers. Exact CNV breakpoints were unknown. (B) The UCSC annotated genes within this deleted region. The red rectangular pointed out the gene that was the most interesting one in our study. (C) The top SNPs included in this deleted region.

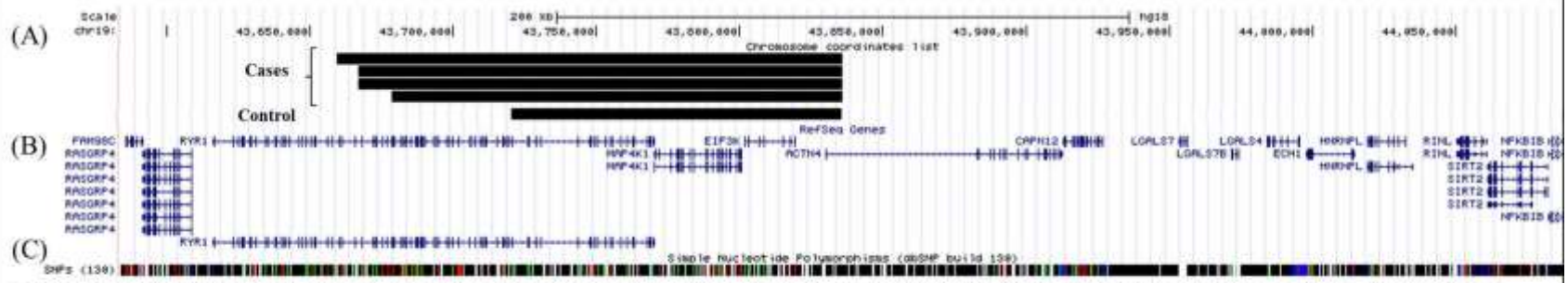

Figure 6 UCSC Genome Browser plot of the copy-number deleted region (Chr19:43835770-43835935). (A) Each black horizontal bar represented the length and breakpoints of a deletion detected in the children of PE mothers or control mothers. Exact CNV breakpoints were unknown. (B) The UCSC annotated gene within this deleted region. (C) The top SNPs included in this deleted region.

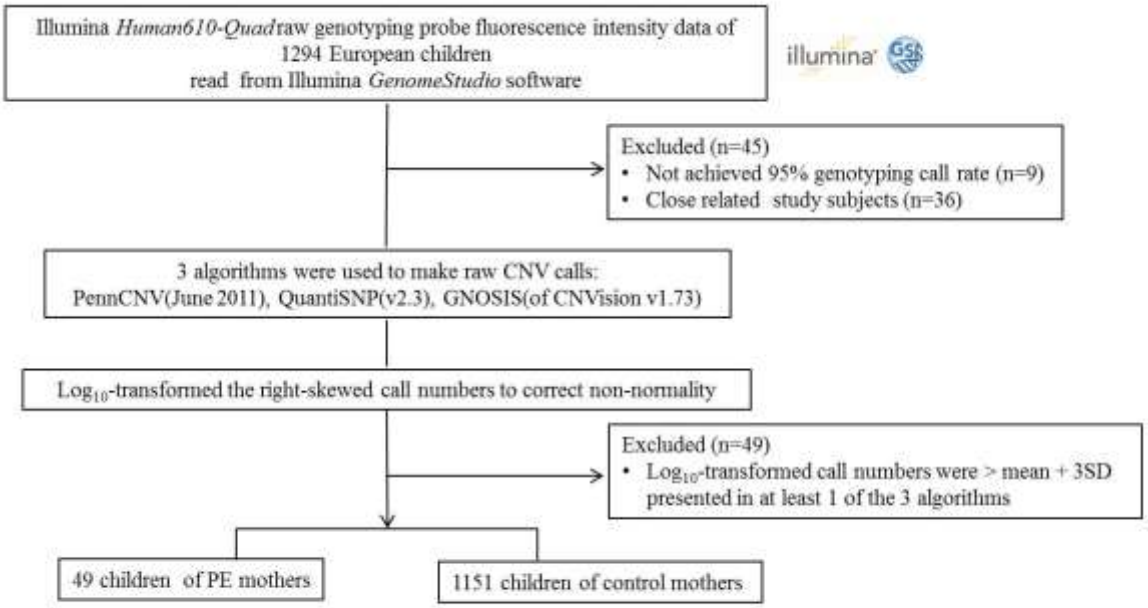

Figure 7 Process of drawing eligible HAPO study subjects

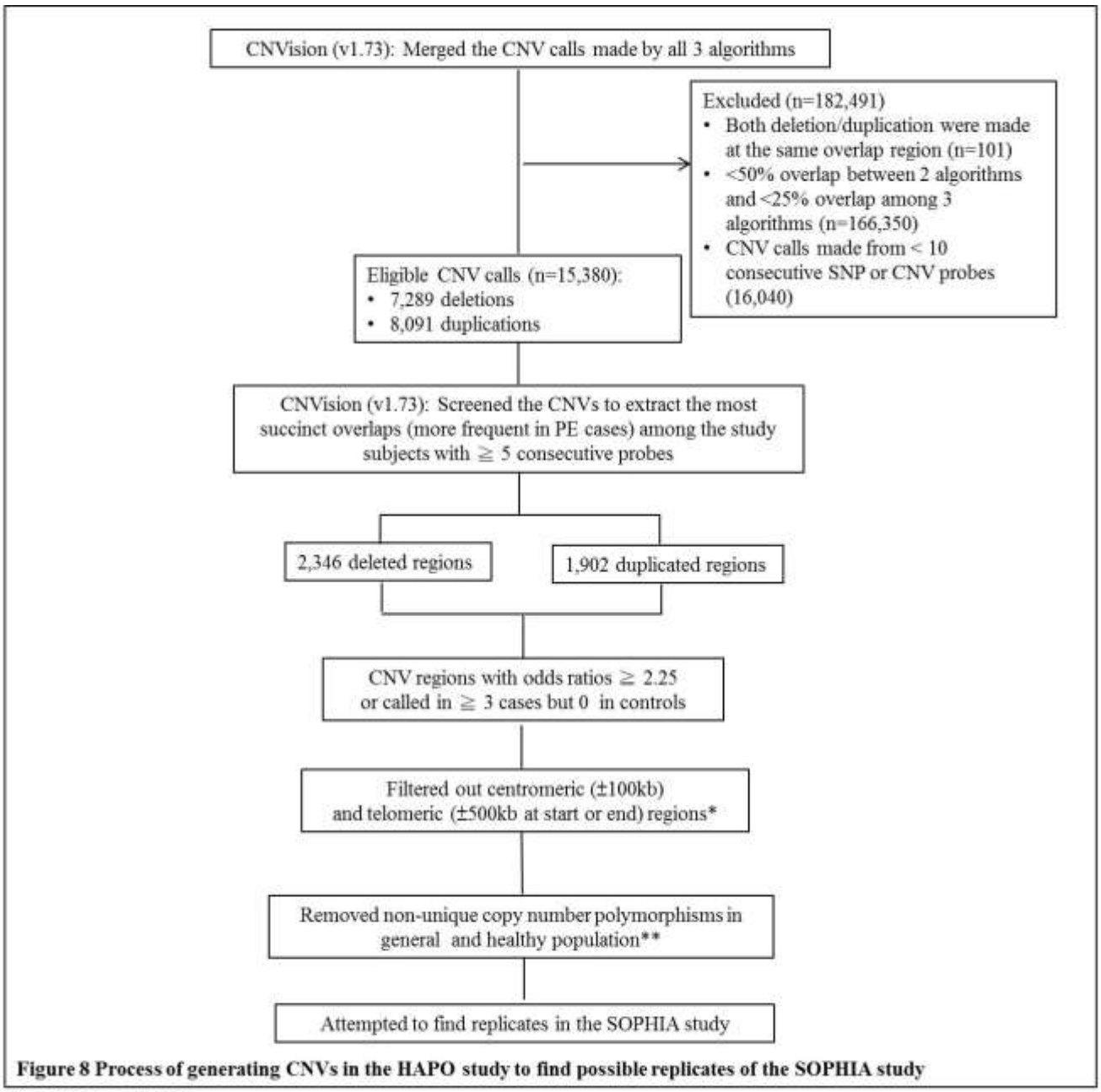

Figure 8 Process of generating CNVs in the HAPO study to find possible replicates of the SOPHIA study

*PennCNV suggested QC

**The Copy Number Variation project at the Children's Hospital of Philadelphia : http://cnv.chop.edu; Database of Genomic Variants: http://dgv.tcag.ca/dgv/app/home

Discussion

In our result, we initially identified 25 candidate copy-number deleted regions and 10 duplicated regions that mostly enriched in the children of the preeclamptic mothers. Since the more common of the CNVs, the less possible they are associated with disease traits[28], we confirmed the none uniqueness of the CNVs we detected through The Copy Number Variation Project (http://cnv.chop.edu) across healthy individuals, and the low to rare CNV frequencies in the general population through the Database of Genomic Variants (http://dgv.tcag.ca/dgv/app/home). Along the process of finding non-unique copy number variable regions, we found that most fetal CNVs were "copy number polymorphisms[29]" that might or might not have significant contributions to the preeclamptic mothers, and we ended up with 20 novel deletion candidates and 4 duplicated candidates. We decided to focus on deletions because loss of function of the protein products by homozygous or heterozygous deletions of certain genes are more frequent and severe than gains of function by slightly duplicated copies of the genes[26, 27]. The two fetal deleted regions we found most interesting on chromosome 1q21.2 and 19q13.2 respectively contain annotated genes of ARNT (aryl hydrocarbon receptor nuclear translocator), CTSK (cathepsin K) and ACTN4 (alpha-actinin-4), that might have molecular and biological effects in developing preeclampsia in the mothers. We postulated the post-translational impairment of these protein functions due to copy number deletion.

The functional transcription factor exists as a heterodimeric complex consisting of the aryl hydrocarbon receptor nuclear translocator (ARNT) and Hypoxia-inducible factor 1-alpha (HIF-1α)[30]. HIF-1α is a principal mediator of hypoxic adaptations, modulates through Wnt/β-catenin signaling in hypoxic status cells[31]. HIF-1α deletion has shown to impair hippocampal Wnt-dependent processes, including cell proliferation, differentiation and

functional maturation. A recent integrative transcriptome meta-analysis using fetal-derived histological subsections of placenta from multiple case-control preeclampsia studies pointed out that the genes clustered in the HIF-1α signaling pathway are dysregulated in the fetus of preeclamptic mothers[9]. This study also showed that the dominant genes involved in regulation of hopoxic trophoblast are HIF-1α, HIF-2α and ARNT. Another study also pointed out that hypoxia interferes with the differentiation of in vitro cultured cytotrophoblast into terminally differentiated syncytiotrophoblast, through DNA hypermythylation and suppression of associated genes[32]. The villi surfaced on syncytiotrophoblasts constitute the intervillous space which is resposible for maternal blood inflow and maternal-fetal exchange during pregnancy. Therefore, downregulated ARNT due to copy-number deletion might lead to unstable heterodimeric complex with HIF-1α, cuasing abnormal HIF-1α function in the fetus, thereby increase oxidative stress and provoke the risk of preeclampsia in the mother. CTSK mutation has been linked to a rare autosomal recessive disease pychodysotosis, which is characterized by osteosclerosis and short stature[33]. Moreover, co-culture of $CTSK^+$ fibroblast with $CTSK^-$ breast cancer cells has been found to enhance cancer cells invasion, and this enhanced invasive ability is blocked by CTSK inhibitors[34]. CTSK might thus have a putative role in regulating trophoblast invasion into maternal decidua, which is important in normal embryo implantation and the following process of vascular remodeling, maintaining stable plancetal blood inflows to the survival and growth of the fetus. CTSK downregulation might lead to defective trophoblast invasion and plancenta maldevelopment, whereas increase the possibility of ischemia in the fetus and preeclampsia in the mother. Mutation of ACTN4 has been linked to familial focal segmental glomerulosclerosis, which is a rare autosomal dominant disease characterized by increased urinary protein excretion and decreasing kidney function[35]. Reduced ACTN4 mRNA expression might dysregulate the

actin cytoskeleton of glomerular podocytes and increase protein filtration into urine. There are gradually growing statistical findings from large datasets or cohort studies pointing out the gene-environment interactions in the uterus between the fetus and the preeclamptic mother[6, 9, 13, 36], but there is still lack of direct biological and mechanistic evidence showing how the abnormal infant genes influence maternal health during pregnancy. Though nutrients and wastes exchange between the fetus and the mother, it is still unknown how extra protein excreted by the fetus into amniotic fluid would end up elevates proteinuria in the mother. We thus speculated about the possibility that the deleted fashion of ACTN4 in the infants might be passed down from the abnormality in the mothers, however, we didn't find direct evidence in reviewing the data from our previous study[16].

In a further attempt to replicate our findings in the HAPO study (Figure 8), we didn't discover matched deleted regions with those we found the most interesting regions containing ARNT, CTSK, and ACTN4 genes in the SOPHIA study. However, in the SOPHIA study, the deleted region upstream of ACTN4 gene that also resides in chromosome 19q13.2 contains a large part of RYR1 gene deletion (Figure 6). We not only found a match for it in the HAPO study, but also found a documented match that already linked fetal deletion to an association with preeclamptic mothers from an established preeclampsia gene database[37]. RYR1 is known as ryanodine receptor 1, which is part of the family of ryanodine receptors that form channels to transport positively charged calcium ions within cells[38]. One study compared mRNA and protein levels expressed by cultured primary syncytiotrophoblasts (the outer syncytial layer of the trophoblast that actively invades the uterine wall forming the outermost fetal component of the placenta) obtained from normal and preeclamptic mothers, and they found a significant decrease in calcium transport by the syncytiotrophoblast cultured from preeclamptic placentas and a

downregulation of RYR1 mRNA expression[39]. Based on systematic reviews of accumulated observational studies that found calcium supplementation may reduce the risk of preeclampsia and may help to prevent preterm birth[40, 41], WHO also recommend on populations where calcium intake is low to take calcium supplementation as part of the antenatal care for the prevention of preeclampsia in pregnant women, particularly among those at higher risk of developing hypertension (http://www.who.int/elena/titles/calcium_pregnancy/en/). Fetal RYR1 gene deletion clearly plays a role in the development of preeclampsia in the mothers, which interferes with calcium transportation during trophoblast invasion that determines a successful pregnancy. Identification of this known preeclampsia associated RYR1 gene deletion in both SOPHIA and HAPO studies further attests the appropriateness of the procedure we preformed to generate prioritized candidate regions of CNVs in our study.

Note that the strategy we implemented to determine targeted CNVs by odds ratio in the cases versus controls was completely subjective in our study; moreover, due to limited number of the study subjects, the CNV frequencies comparing cases to controls were usually low, which limited the power of our analyses, resulted in wide 95% confidence intervals (data not shown).

Genome-wide association study is a hypothesis-free approach that screens the entire genome for associations between genetic polymorphisms and a certain disease. The most important methodological issue with GWAS studies is multiple comparisons, which needs to be addressed by strictly statistical significance level correction and further replicative studies, and yet this corrected genome-wide significant level usually makes it difficult to find stringent result as can been seen in our study. The limitations of GWAS studies also includes the inability to detect associations between rare variants (prevalence < 2%) and the disease, where we assumed to be genotyping error and eliminated at the quality control step. We also excluded those SNPs at the

quality control step which violated of the Mendelian inheritance or Hardy-Weinberg equilibrium, where none in fact really existed[42, 43] due to the intrinsic that allelic or genotypic frequencies in a population will remain constant from generation to generation in the absence of other sporadic evolutionary influences. In addition, GWAS only identifies a genomic location associated with the disease, but not proving a causal polymorphism, because GWA studies are unable to provide information on biological mechanisms underlying the association, which requires further molecular biology studies to fortify our understanding.

Although our genome is born to be what it is, the difference between incidence case and prevalence case determined the genetic variations for incidences and survivors, which reflects divergent indications of the disease statuses. Recently the concept of endophenotypes, which involve multiple biological, behavioral or cognitive features to the vulnerability of a certain disorder[44], might help explain the complex mechanisms contribute to preeclampsia. Since preeclampsia is defined by both gestational hypertension and proteinuria, these two traits might be determined by different genetic variations that are closely related to each other. For example, abnormality in the vasopressin gene might affect electrolytes metabolism, cause retention of sodium in the blood and elevation of blood volume and pressure[45]; insulin resistant predisposition might be related to diabetic and hypertensive traits; long-term stimulation of sympathetic nervous system by stress might contribute to hypertension as well; hypoxia and oxidative stress has been shown to interfere with placenta development by dysregulation of the EGFR growth factor signaling pathway to blood vessel defects in placenta[9]; although rare, genotype-phenotype defects in any of the components involved in glomerular filtration system can result in proteinuria[46], etc. Although not every individual carrying certain genes will result in phenotypic expression, these people might still possess higher risk of preeclampsia during

pregnancy. Moreover, in order to prevent population stratification, we only enrolled children from white mothers in our study, therefore the preeclamptic genetic variants across multi-ethnicity are left for further investigation.

In conclusion, dysfunctional protein regulation in the fetus due to copy number deletion might or might not directly influence the mother. However, our study pointed out possible roles of ARNT, CTSK and ACTN4 gene deletions in the fetus that are associated with hypoxia modulation, postulated regulation of trophoblast invasion, and abnormal proteinuria in the fetus respectively, which need further biological experiments to substantiate the functional protein expression in associated with increased risk of preeclampsia in the mother. The long term goal of our study is to better understand the etiology of preeclampsia and in hopes to develop better screening tools based on genetic variants carried by the predisposed mother or fetus. With more thorough considerations of various genetic variations, including rare or common SNPs and CNVs, applying the concept of endophenotypes, and using integrative omics approaches, might one day help us decipher causal or etiological determinants among existing association studies, to more effectively intervene and prevent preeclampsia in susceptible individuals in a timely manner.

Reference

1. Wenger NK. Recognizing pregnancy-associated cardiovascular risk factors. The American journal of cardiology. 2014; 113:406-9.
2. Rath W, Fischer T. The diagnosis and treatment of hypertensive disorders of pregnancy: new findings for antenatal and inpatient care. Deutsches Arzteblatt international. 2009; 106:733-8.
3. Uzan J, Carbonnel M, Piconne O, Asmar R, Ayoubi JM. Pre-eclampsia: pathophysiology, diagnosis, and management. Vascular health and risk management. 2011; 7:467-74.
4. Arngrimsson R, Bjornsson S, Geirsson RT, Bjornsson H, Walker JJ, Snaedal G. Genetic and familial predisposition to eclampsia and pre-eclampsia in a defined population. British journal of obstetrics and gynaecology. 1990; 97:762-9.
5. Lie RT, Rasmussen S, Brunborg H, Gjessing HK, Lie-Nielsen E, Irgens LM. Fetal and maternal contributions to risk of pre-eclampsia: population based study. Bmj. 1998; 316:1343-7.
6. Skjaerven R, Vatten LJ, Wilcox AJ, Ronning T, Irgens LM, Lie RT. Recurrence of pre-eclampsia across generations: exploring fetal and maternal genetic components in a population based cohort. Bmj. 2005; 331:877.
7. Hiby SE, Walker JJ, O'Shaughnessy K M, Redman CW, Carrington M, Trowsdale J, et al. Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive success. The Journal of experimental medicine. 2004; 200:957-65.
8. Mutze S, Rudnik-Schoneborn S, Zerres K, Rath W. Genes and the preeclampsia syndrome. Journal of perinatal medicine. 2008; 36:38-58.
9. Moslehi R, Mills JL, Signore C, Kumar A, Ambroggio X, Dzutsev A. Integrative transcriptome analysis reveals dysregulation of canonical cancer molecular pathways in placenta leading to preeclampsia. Scientific reports. 2013; 3:2407.
10. Williamson C. Molecular biology related to pre-eclampsia. Int Congr Ser. 2005; 1279:282-9.
11. Chappell S, Morgan L. Searching for genetic clues to the causes of pre-eclampsia. Clinical science. 2006; 110:443-58.
12. Morgan L, Crawshaw S, Baker PN, Brookfield JF, Broughton Pipkin F, Kalsheker N. Distortion of maternal-fetal angiotensin II type 1 receptor allele transmission in pre-eclampsia. J Med Genet. 1998; 35:632-6.
13. Cnattingius S, Reilly M, Pawitan Y, Lichtenstein P. Maternal and fetal genetic factors account for most of familial aggregation of preeclampsia: A population-based Swedish cohort study. Am J Med Genet A. 2004; 130A:365-71.
14. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annual review of medicine. 2010; 61:437-55.
15. Fanciulli M, Petretto E, Aitman TJ. Gene copy number variation and common human disease. Clinical genetics. 2010; 77:201-13.
16. Zhao L, Triche EW, Walsh KM, Bracken MB, Saftlas AF, Hoh J, et al. Genome-wide association study identifies a maternal copy-number deletion in PSG11 enriched among preeclampsia patients. BMC pregnancy and childbirth. 2012; 12:61.
17. Zhao L, Bracken MB, DeWan AT. Genome-wide association study of pre-eclampsia detects novel maternal single nucleotide polymorphisms and copy-number variants in subsets of the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study cohort. Annals of human genetics. 2013; 77:277-87.

18. PLINK:Whole genome association analysis toolset. Available from: http://pngu.mgh.harvard.edu/~purcell/plink/.
19. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome research. 2007; 17:1665-74.
20. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, et al. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. 2007; 35:2013-25.
21. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron. 2011; 70:863-85.
22. Cooper GM, Mefford HC. Detection of copy number variation using SNP genotyping. Methods in molecular biology. 2011; 767:243-52.
23. Xu L, Hou Y, Bickhart DM, Song J, Liu GE. Comparative Analysis of CNV Calling Algorithms: Literature Survey and a Case Study Using Bovine High-Density SNP Data. Microarrays. 2013; 2:171-85.
24. Sanders S, Mason C. CNVision. State Lab at Yale University; Available from: http://futo.cs.yale.edu/mw/index.php/CNVision#CNV_Analysis.
25. Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K, et al. High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome research. 2009; 19:1682-90.
26. Jacobs PA, Browne C, Gregson N, Joyce C, White H. Estimates of the Frequency of Chromosome-Abnormalities Detectable in Unselected Newborns Using Moderate Levels of Banding. J Med Genet. 1992; 29:103-8.
27. Brewer C, Holloway S, Zawalnyski P, Schinzel A, FitzPatrick D. A chromosomal deletion map of human malformations. American journal of human genetics. 1998; 63:1153-9.
28. Yeo RA, Gangestad SW, Liu J, Calhoun VD, Hutchison KE. Rare copy number deletions predict individual variation in intelligence. PloS one. 2011; 6:e16339.
29. Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, et al. Whole population, genome-wide mapping of hidden relatedness. Genome research. 2009; 19:318-26.
30. Isaacs JS, Jung YJ, Neckers L. Aryl hydrocarbon nuclear translocator (ARNT) promotes oxygen-independent stabilization of hypoxia-inducible factor-1alpha by modulating an Hsp90-dependent regulatory pathway. The Journal of biological chemistry. 2004; 279:16128-35.
31. Mazumdar J, O'Brien WT, Johnson RS, LaManna JC, Chavez JC, Klein PS, et al. O2 regulates stem cells through Wnt/beta-catenin signalling. Nature cell biology. 2010; 12:1007-13.
32. Yuen RK, Chen B, Blair JD, Robinson WP, Nelson DM. Hypoxia alters the epigenetic profile in cultured human placental trophoblasts. Epigenetics : official journal of the DNA Methylation Society. 2013; 8:192-202.
33. Giraudeau FS, McGinnis RE, Gray IC, O'Brien EJ, Doncaster KE, Spurr NK, et al. Characterization of common genetic variants in cathepsin K and testing for association with bone mineral density in a large cohort of perimenopausal women from Scotland. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2004; 19:31-41.
34. Kleer CG, Bloushtain-Qimron N, Chen YH, Carrasco D, Hu M, Yao J, et al. Epithelial and stromal cathepsin K and CXCL14 expression in breast tumor progression. Clinical cancer research : an official journal of the American Association for Cancer Research. 2008;

14:5357-67.
35. Kaplan JM, Kim SH, North KN, Rennke H, Correia LA, Tong HQ, et al. Mutations in ACTN4, encoding alpha-actinin-4, cause familial focal segmental glomerulosclerosis. Nature genetics. 2000; 24:251-6.
36. Wu CS, Nohr EA, Bech BH, Vestergaard M, Catov JM, Olsen J. Health of children born to mothers who had preeclampsia: a population-based cohort study. Am J Obstet Gynecol. 2009; 201.
37. Triche EW, Uzun A, DeWan AT, Kurihara I, Liu J, Occhiogrosso R, et al. Bioinformatic approach to the genetics of preeclampsia. Obstetrics and gynecology. 2014; 123:1155-61.
38. Chen Y, Xue S, Zou J, Lopez JR, Yang JJ, Perez CF. Myoplasmic resting Ca2+ regulation by ryanodine receptors is under the control of a novel Ca2+-binding region of the receptor. The Biochemical journal. 2014; 460:261-71.
39. Hache S, Takser L, LeBellego F, Weiler H, Leduc L, Forest JC, et al. Alteration of calcium homeostasis in primary preeclamptic syncytiotrophoblasts: effect on calcium exchange in placenta. Journal of cellular and molecular medicine. 2011; 15:654-67.
40. Imdad A, Bhutta ZA. Effects of calcium supplementation during pregnancy on maternal, fetal and birth outcomes. Paediatric and perinatal epidemiology. 2012; 26 Suppl 1:138-52.
41. Hofmeyr GJ, Lawrie TA, Atallah AN, Duley L, Torloni MR. Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems. The Cochrane database of systematic reviews. 2014; 6:CD001059.
42. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, et al. Common deletion polymorphisms in the human genome. Nature genetics. 2006; 38:86-92.
43. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. A high-resolution survey of deletion polymorphism in the human genome. Nature genetics. 2006; 38:75-81.
44. Kendler KS, Neale MC. Endophenotype: a conceptual analysis. Molecular psychiatry. 2010; 15:789-97.
45. Johnson RJ, Feig DI, Nakagawa T, Sanchez-Lozada LG, Rodriguez-Iturbe B. Pathogenesis of essential hypertension: historical paradigms and modern insights. Journal of hypertension. 2008; 26:381-91.
46. Piscione TD, Licht C. Genetics of proteinuria: an overview of gene mutations associated with nonsyndromic proteinuric glomerulopathies. Advances in chronic kidney disease. 2011; 18:273-89.