Research Article



# Molecular Evolution of Genes Associated with Preeclampsia: Genetic Conflict, Antagonistic Coevolution and Signals of Selection

Alistair Arthur<sup>1</sup>, Ana Terman<sup>1</sup>, Chelcie Kurzontkowski<sup>2</sup>, Keith Keene<sup>1,3</sup>, Toyin Babatunde<sup>2</sup>, Kyle Summers<sup>1\*</sup>

<sup>1</sup>Department of Biology, East Carolina University, Greenville, NC 27858, USA <sup>2</sup>Department of Nutrition Science, East Carolina University, Greenville, NC 27858, USA <sup>3</sup>Center for Health Disparities, East Carolina University, Greenville, NC 27858, USA Address correspondence to Kyle Summers, SUMMERSK@ecu.edu

Received 29 September 2018; Revised 30 October 2018; Accepted 06 November 2018

Copyright © 2018 Kyle Summers et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

Parent-offspring conflict theory predicts continuing conflict between maternal and fetal interests during pregnancy. This is thought to contribute to risks of diseases like hypertension and preeclampsia during pregnancy. Genes expressed in the maternal and fetal genomes are predicted to have conflicting effects on various aspects of maternal physiology, including blood pressure. The genes are predicted to undergo continuous antagonistic coevolution, which should leave signals of positive selection in the short and long term. We tested for such signals in the FLT1 gene (previously argued to be a locus involved maternal-fetal conflict), and in several other suites of genes found to be significantly associated with preeclampsia in large-scale GWAS analyses. The FLT1 gene showed strong signals of positive selection at multiple levels of analysis. The suites of genes did not show an overall enhanced probability of positive selection (relative to a control set of genes), but a number of genes did show strong positive selection and may be good candidates for further analyses of maternal-fetal conflict.

Keywords: Preeclampsia; Antagonistic; Genes

### **1.Introduction**

Pregnancy represents one of the most intimate relationships to be found in the animal kingdom, characterized by extreme altruism by the mother toward the unborn offspring inside of her. Nevertheless, this relationship is not without conflict. Further developing theoretical principles pioneered by Hamilton et al. (parent-offspring conflict: Trivers) [1], Haig elaborated how genetic conflicts between the mother and individual embryonic offspring engender a series of struggles over resource allocation during pregnancy from control over nutrient supply to termination of pregnancy itself [2]. These genetic conflicts are a ubiquitous feature of parent-offspring interactions because offspring are more closely related to themselves than to siblings, whereas mothers are equally related to all their offspring David Haig extended this theory to explain the phenomenon of genomic imprinting, in which the paternal and maternal halves of the genome of the same individual come into conflict over resource allocation [1,3,4]. Due to historically distinct patterns of relatedness between offspring through the maternal and paternal lineages, genes from the paternal line typically favour higher levels of resource allocation to the embryo in which they reside than to other embryos, relative to genes from the maternal line. Genomic imprinting provides a mechanism whereby gene expression in a particular embryo is affected by whether or not it was inherited from the mother (maternal line) or, father (paternal line). A number of imprinted genes that affect the struggle over resource allocation between the embryo and the mother have now been identified, and show patterns of expression consistent with predictions of the kinship theory of genomic imprinting [5].

In a classic paper, Haig described several types of conflicts that could occur between the mother and the embryo during pregnancy [2]. He argued that conflict was highly likely to occur over 1) termination of pregnancy and 2) the quantity of resources (e.g. glucose) received from the mother by a particular embryo. The quantity of resources received by an embryo could be influenced by (for example) blood pressure in the maternal circulation relative to that in the fetal circulation (high blood pressure in the former and lower pressure in the latter increases access to blood-borne resources by the embryo), or by levels of glucose in the maternal circulation. Coevolutionary struggle between the mother and the embryo with regard to maternal blood pressure levels could lead to dysregulation resulting in preeclampsia, a severe condition of maternal hypertension that carries a significant risk of mortality [6]. Haig and others have argued that the coevolutionary struggle between the mother and the embryo (and between the maternal and paternal parts of the genome in the case of imprinted genes) may play out via genetic changes and counter-changes in genes that promote and/or counteract embryonic growth and survival [2,7].

From a life history evolution perspective, we expect that genetic changes that promote increased resource transfer to a particular embryo will increase the fitness (e.g. growth and survival rate) of the embryo, in spite of the increased risk of catastrophic maternal (and consequent fetal) mortality. Recent research indicates that fetal phenotypes associated with increased maternal hypertension during the first trimester of pregnancy increase embryonic survival rates and decrease later-life disease risks, consistent with the argument that genetic variants associated with pregnancy-related hypertension usually increase offspring fitness, in spite of increasing risk of preeclampsia and potential catastrophic costs such as maternal mortality [8]. At the molecular level, it has been predicted that genes involved in genetic conflicts between mother and offspring (imprinted or otherwise) will show signals of continuous conflictual coevolution [7]. Just as the immune genes of hosts and the genes of parasites that are targeted by the immune system show clear signals of antagonistic coevolution over time [9], we expect that genes involved in genetic conflicts between mothers and their offspring (and between the maternal and paternal halves of the genome within offspring) will show signals of antagonistic coevolution over the long course of molecular evolutionary history. Recent studies of the molecular evolution of some of these genes clearly support this prediction on a macro evolutionary scale. For example, Chuong et al. found that three gene families that comprise major components of genes with placentaspecific expression patterns show rapid evolution across rodents, with evidence for strong positive selection in the genus Mus [10]. Across mammals, the placenta varies with respect to invasiveness of maternal tissues, with epitheliochorial placentation being the least invasive (involving contact between the fetal chorion and the uterine epithelium), endotheliochorial (chorion in contact with maternal blood vessel endothelium) being somewhat more invasive, and hemochorial being the most invasive (in which the chorion invades the maternal circulation and is in direct contact with maternal blood) [11]. Phylogenetic analyses indicate that hemochorial placentation appeared early in mammalian evolution. Preeclampsia is associated with reduced invasion of maternal tissues by the placenta. Comparative analyses indicate that there is substantial overlap between genes affecting placental invasiveness across mammals and genes associated with risk of preeclampsia [11,12]. Specifically, a set of genes shown to be under positive

selection independently in three mammalian lineages that have evolved reduced placental invasiveness (tree shrews, kangaroo rats and lemurs), is significantly enriched for genes that affect preeclampsia. Crosley et al. carried out a genomic survey of genes under positive selection in the inferred phylogenetic point of origin of the invasive placenta in great apes, and identified several clusters of genes under positive selection that had previously been linked to risk of preeclampsia [13].

Preeclampsia in humans appears to unfold in a two stage process [14]. In the first stage, shallow invasion of the maternal circulation by the placenta leads to insufficient resource transfer to the placenta. The second stage involves widespread endothelial malfunction, including increased vascular permeability, abnormal coagulation, and altered vascular tone due to endothelial disturbances [14,15]. A number of distinct circulating factors appear to contribute to the development of generalized endothelial dysfunction during preeclampsia, although which elements have primary causal roles is difficult to determine [6]. The FLT1 locus appears to be a likely candidate for a gene with a primary causal role in the etiology of preeclampsia [6]. Recent genomewide association study (GWAS) results indicate that variation at this locus is strongly associated with risk of preeclampsia [16]. This gene produces a protein product that is expressed by the placenta but affects the physiology of the maternal circulation. A variant transcript of this gene (sFLT1) binds to VEGF (vascular endothelial growth factor) and PIGF (placenta growth factor), but (unlike the normal protein) does not anchor to the interior vascular surface. Instead, by binding to VEGF and PIGF this variant prevents these proteins from binding to the interior vascular surface, leading to endothelial damage, vasoconstriction and increased blood pressure in the maternal circulation [6]. While imposing substantial risk and cost to the mother, this variant protein also increases access to blood and associated nutrients by the embryo. Yuan et al. argue that this FLT1 variant likely evolved under selection on genes expressed in the embryo for increased access to maternal resources [6]. They speculate that *FLT1* may be just one of many loci mediating genetic conflicts over resources between the mother and her embryo(s).

Here, we first investigate the molecular evolution of the *FLT1* gene for evidence of positive selection. We predict that this locus will show evidence of continual change under positive selection, consistent with the hypothesis of continual antagonistic coevolution between genes expressed in the mother and genes expressed in the embryo due to parent-offspring conflict. We use molecular evolutionary statistical analyses to search for signals of positive selection on this locus. We then use these methods to investigate sets of genes that have been identified as likely to play a role in preeclampsia in large-scale GWAS analyses from various populations, as summarized in the GWASdb online database (http://

jjwanglab.org/gwasdb) and other resources. Our goal is to identify genes that are likely to be loci of antagonistic coevolution in the context of parent-offspring conflict. We use broad scale analyses of positive selection on these genes using the CODEML program in the PAML program suite to investigate long-term historical patterns of selection on these genes across the primates. We use branch and codon-specific models implemented in CODEML to investigate patterns of selection on these genes in the human lineage. We then use tests of recent positive selection within the human lineage to search for signals of selection acting on these genes (specifically the iHS test and the XP-EHH test [17,18]). Given the argument that these kinds of genetic conflicts will produce chronic, unending cycles of genetic change and counterresponse (i.e. antagonistic coevolution), we expect to see evidence of positive selection not only across species, but also within species, both between different populations (as different populations follow distinct coevolutionary trajectories) and within populations.

### 2.Methods

A In addition to the FLT1 gene, we analyzed seven sets of genes (54 genes in total) previously found to be associated with risk or severity of preeclampsia and one control gene set: 1) A set of five genes showing the strongest associations with preeclampsia in the GWASdb database http://jjwanglab.org/gwasdb). These genes were in both the top ten most significant hits, and also in the top 30 most strongly associated genes using the Prix Fixe method focused on functional sub-networks of genes [19]. 2) A set of seven genes with SNPs associated (via GWAS) with preeclampsia which have been found to be under positive selection on branches of the mammalian phylogeny with independent transitions from hemochorial to epitheliochorial or endotheliochorial placentation (associated with decreased placental invasiveness). 3) A set of five genes which show differential expression associated with preeclampsia (via GWAS) which has been found to be under positive selection on branches of the mammalian phylogeny with independent transitions from hemochorial to epitheliochorial or endotheliochorial (associated with increased placental invasiveness). 4) A set of 15 genes from a 2013 meta-analysis of genes differentially expressed in preeclampsia from multiple, GWAS studies [20]. 5) A set of seven imprinted genes associated with preeclampsia. 6) A set of 12 genes associated with preeclampsia that is expressed specifically by the fetal rather than maternal tissues [21]. 7) A set of nine genes with the strongest association with preeclampsia from the PESNPdb preeclampsia GWAS database (http://bejerano.stanford.edu/pesnpdb) [22]. 8) A set of 20 control genes taken from the set of genes most strongly associated with heart failure in the GWASdb database (http://jjwanglab.org/gwasdb), using the Prix Fixe method focused on functional subnetworks of genes. Note that some genes overlapped between sets. Further, due to difficulties in obtaining or aligning sequence data

for multiple taxa for some genes, the sample sizes of the gene sets for the PAML analyses are smaller than for the haplotype analyses. In selecting genes for each gene set, we removed genes that were associated with the immune response. Although it is highly likely that some aspects of genetic conflict between mothers and embryos are mediated by immune system genes, these genes may also be intimately involved in parasite-host interactions, which are well known to impose strong positive selection, making inferences about the source of the signal ambiguous.

### 4.1. Analytic methods

Sequence Alignment and Phylogenetic Arrangements: Available orthologous primate gene sequences were obtained for each candidate human gene in the following manner. First, the RefSeq protein for the gene in Homo sapiens was located in the Gene database at NCBI (https://www.ncbi.nlm.nih.gov/), and ten additional primate orthologs were then obtained using reciprocal TBLASTN (https://blast.ncbi.nlm.nih.gov/Blast.cgi: protein query to a translated nucleotide database) searches of the GenBank nucleotide database (search restricted to primates). Next, the orthologous protein and nucleotide sequences were obtained and codonbased alignments were performed. First, the protein sequences were aligned using the program MAFFT (http://mafft.cbrc.jp/alignment/server/), with default parameters. This protein sequence alignment was then imported into the PAL2NAL program (http://www.bork. embl.de/pal2nal/), and used as the guide alignment for the corresponding nucleotide sequences. This ensured that the nucleotide sequences were properly aligned by codon position in the final nucleotide alignment used in the CODEML analyses (see below). We selected primate gene sequences for use in the analyses as follows: four old world ape sequences (typically Pan troglodytes, gorilla, Pongo abelli and Nomascus leucogenys), three old world monkey sequences (typically Macaca mulata, Papio anubis and Rhinopithecus roxellana) and two new world monkey sequences (typically Cebus capuchinus and Callithrix jacchus). Phylogenetic relationships for each set of primate taxa (for each specific gene) were constructed based on a recent summary of primate phylogenetic relationships from the literature [23]. Each tree was represented in Newick format for analysis, and was represented as unrooted at the base (by forming a tritomy between the two new world primates and the old world primates), as required by the CODEML program.

### 4.2 Tests for positive selection in PAML

Rats Analyses of historical selection on specific lineage were carried out with the CODEML program in the PAML program package, which implements maximum-likelihood-based methods that test for a statistical excess of non-synonymous nucleotide changes (leading to amino acid substitutions), over synonymous changes, in sequence data of protein coding genes [24]. We used codon-specific models implemented in CODEML to investigate signals of positive selection on specific genes from the gene sets across the tree of primates used for each gene. We then used codon and branch-specific models implemented in CODEML to investigate signals of positive selection specific to the human lineage.

The following site-specific models were used to analyze the dataset for each gene region: M0 (single rate model), M1 (neutral model), M2 (basic selection model), M7 (continuous distribution model), and M8 (continuous distribution plus selection model). Codon frequencies were estimated from the average nucleotide frequencies at the three codon positions for all runs. Loglikelihood ratio tests were used to test for significant differences in the fit of the models incorporating selection relative to their (nested) counterparts that did not allow positive selection [25]. We focused on comparing model 1 (neutral model) results to model 2 (selection) results, and model 7 (continuous distribution) to model 8 (continuous plus selection), as recommended by Yang [26]. We also compared model 8 to model 8 with the value of omega fixed at one, which has been used as a stringent test in some analyses [27]. These tests provide a useful series of metrics for interpreting the significance of the results. The order of stringency of the LRTs is as follows: M7 vs. M8 261 < M1 vs. M2 < M8 vs. M8 fixed omega. We used the Bayes Empirical Bayes method implemented in CODEML to estimate posterior probabilities of selection on each codon. This method allows assessment of the probabilities that specific codon sites in the gene in question are under positive selection. For genes that showed evidence of positive selection, to check that the method had not converged on a local maximum (leaving a global maximum undetected) we carried out several runs for each set of models, using three initial values for kappa (transition/transversion ratio: 0.5, 2 and 10) in the site-specific models and omega (dN/dS: 0.1, 1 and 10) in the branch and site-specific models (see below). The final likelihoods were compared and the highest likelihood taken as the best estimate. For the branch and site-specific analyses of positive selection we used Model A as implemented in CODEML, which detects selection on specific codons along specific branches of a phylogenetic tree [27]. Model A allows four categories of selection on codon sites in a sequence: two categories  $(GD_1 \text{ and } GD_0)$  with uniform selection across all branches (( $GD_1 = 1$ , and 0) < GD<sub>0</sub> <1 (estimated from the data)), an two for which selection pressure differs in one or a few pre-specified "foreground" branches where selection is assumed to have changed. The method finds the model with the maximum likelihood given the data. If the maximum likelihood model includes a category of sites with an omega >1, then this provides evidence that positive selection has acted on those sites along the specific lineage analyzed. To test the statistical significance of the results, the log

likelihood of the maximum likelihood model is compared to similar models lacking selection using a log-likelihood ratio test (LRT). Hence, we compared the maximum likelihood model to itself, with omega constrained to equal one (i.e. a category of sites under positive selection was not allowed). Specific sites under selection were inferred using the Bayes Empirical Bayes method. Given the large number of CODEML runs required to complete these analyses, we implemented them using a PAML pipeline written in perl (LMAP: http://lmapaml. sourceforge.net/). This allowed us to organize, initiate multiple runs in succession, and consolidate and extract key results from each set of genes, without having to set up each run for each gene separately [28].

# 4.3 Tests for positive selection using the 1000 genomes project

Genes under strong positive selection sweep upward in frequency within populations, pulling linked allelic variants along with them, resulting in regions of extended haplotype homozygosity around the focal locus, the classic signature of a selective sweep [29]. Sabeti et al. developed the extended haplotype homozygosity (EHH) test for this signature using SNP data [30], and Voight et al. modified the EHH to test for recent selective sweeps in human populations using data from the human haplotype map (HapMap), which comprises the first population-based SNP genotyping study across the human genome [17]. Sabeti et al. developed a crosspopulation version of the EHH test (XP-EHH) that is designed to detect sweeps in one population relative to another [18]. This test can detect sweeps that have neared completion, whereas within-population only tests (e.g. iHS) have low power when the selected variant reaches high frequency. Such selection-generated patterns of linkage disequilibrium are transient, broken up by recombination, such that the method is unlikely to infer putative signatures of selective events older than several hundred thousand years. Currently these (and other) tests for selective sweeps have been implemented in several different online browsers that are readily available. Here we use the 1000 Genomes Project Selection Browser (http://hsb.upf.edu/), to investigate signals of selection in two human populations: Europeans (CEU) and Africans (YRI), using both the iHS test for selection within populations and the XP-EHH test for selection between populations. The results of each type of test for each population (iHS) or cross- population comparison (XP-EHH) are quantified via a ranked--score comparison in the 1000 Genomes Selection Browser, with a score  $\geq 2$ (-log 10 of the p-value) representing a significant score  $(p \le 0.01)$  on the basis of a comparison with the scores of all SNPs across the entire genome. We note that whereas PAML analyses focus on selection for amino acid change in coding regions over long-term evolutionary time, analyses using the 1000 genomes positive selection browser probe for recent signatures of selection driven by any coding or non-coding variants, exclusively in the human lineage.

## 5. Results

### 5.1. Molecular evolution of the FLT1 locus

The FLT1 gene showed strong signals of ongoing selection at multiple levels of analysis. Across the primates, Model 2 in CODEML estimated that a proportion of 0.006 of the codons were under positive selection, with an estimated dN/dS ratio of 68. This result was highly significant: the log likelihood ratio test yielded a significance level of p<0.001. The continuous model (Model 8) showed an almost identical result, with a dN/dS ratio of 65 at a proportion of 0.006 of the codons, and a significance level of p<0.001 (for either comparison: Model 8 versus Model 7, and Model 8 versus Model 8 with omega fixed at one). The Bayes Empirical Bayes method estimated four sites as having changed under positive selection with a probability greater than 95%: 656 and 658-60. For the human lineage, Model A estimated that a proportion of 0.005 of the codons were under positive selection, with an estimated dN/dS of 999 (this essentially means that virtually all changes at these codons were non-synonymous in this lineage). The log likelihood ratio test was again highly significant (p<0.001). These results are summarized in (Table 1). Similarly, the iHS and XP-EHH screens showed evidence for recent positive selection. Figure 1 shows the results of the iHS screen using the UCSC Genome Browser graphic. The green histogram shows an area of high values on the left side of the gene region window (values above 2 are significant in genome-wide comparisons). Figure 2 shows results of the XP-EHH screen using the selection browser. Allele rs4769613, found to be strongly associated with preeclampsia, has increased rapidly under selection in European populations. Figure 3 visually compares the frequencies of rs4769613 between Africa and Europe/Asia, showing the increase of this allele in Europe and Asia.

Summaries of the full results across all the gene sets, for analyses using CODEML (long-term positive selection) and linkage disequilibrium tests (iHS and XP-EHH: recent selection within and between populations) are shown in Tables S1 and S2, respectively. A number of genes from each preeclampsia gene set showed evidence for positive selection (Table 2). However, none of the sets showed a significant overall enrichment for positive selection compared to the control set of genes associated with heart failure, in either the coding sequence analyses or the haplotype homozygosity analyses. Contingency table tests (Fisher's Exact tests) did not reveal any significant differences between the control gene set and

**Table 1:** Maximum likelihood estimates of positive selection acting on the *FLT1* gene across primates, and in the human lineage, using the CODEML program implemented in PAML (Yang et al. 2007). Omega = dN/dS ratio, proportion = estimated proportion of sites under positive selection, Models (1-8 = codon-specific models, A and Afix = branch and codon-specific models): 1 = discrete purifying + neutral, 2 = discrete purifying + neutral + positive, 7 = continuous purifying + neutral, 8 = continuous purifying + neutral + positive selection, 8fix = 8 with omega fixed at one (strict control), A

Gene	Lineage	omega	proportion	Model	logLike	Comp	dlogLx2	p value	BEB
FLT1	Primates	68.07	0.006	2	-4249.57				656, 658-60
FLT1	Primates			1	-4302.31	2 vs 1	105.48	< 0.001	
FLT1	Primates	66	0.006	8	-4249.84				656, 658-60
FLT1	Primates			7	-4302.32	8 vs 7	104.95	< 0.001	
FLT1	Primates			8fix	-4296.98	8 vs 8fix	94.26	< 0.001	
FLT1	Human	999	0.005	А	-4285.80				657
FLT1	Human			Afix	-4302.24	A vs Afix	32.89	< 0.001	



**Figure 1:** Results from the 1000 Genomes Selection Browser visualized on the UCSC Genome Browser. iHS rank score (green histogram) showing evidence for selection on the *FLT1* gene in Europeans. Scores above two indicate significant evidence for positive selection in genome-wide comparisons (scale bar (left green bar) scales from 0 to 3.5). Coding regions (blocks) and intronic regions in blue.



Figure 2: Results from the 1000 Genomes Selection Browser visualized on the UCSC Genome Browser. XP-EHH rank score (green histogram) showing evidence for selection on the *FLT1* gene across populations (Europeans compared to Africans). Scores above two indicate significant evidence for positive selection in genome-wide comparisons (scale bar (left green bar) scales from 0 to 3.5). Coding regions (blocks) and intronic regions in blue.



Figure 3: Distribution of alleles at *FLT1* SNP rs4769613. Ancestral allele = blue, derived = orange. The derived allele has undergone rapid increase under selection in European populations.

Table 2: Genes sh	lowing	evidence for	or positive se	election (using	dN/dS tests	implemented	in CODEML). N.S.	= Not significant.
	<u> </u>			( U		1		0

Gene Set	Prop. Pos. Sel. (dN/dS)	Fisher's Exact Test	Pos. Sel. Genes
GWASdb	3/4	N.S.	ESSRG, INVS, RUNXI
Elliot & Crespi SNPs	5/7	N.S.	FLT1, C1QTNF6, COL4A2, F2, PLAUR
Elliot & Crespi DE	1/3	N.S.	EDN2
GWAS-meta	4/13	N.S.	FLT1, CRH, F5, HEXB
Fetal Expression	3/6	N.S.	CDKN1C, KDR, STOX1
PESNPdb	2/7	N.S.	C4orf37, LMCD1
Control	5/15		

the preeclampsia gene sets in terms of the number of genes under positive selection. This was the case when each set of preeclampsia genes was analyzed separately, or when all were pooled and analysed together. For the CODEML analyses, it could be argued that the number of genes showing evidence for positive selection (5/15) was higher than expected based on results from previous studies. We do not know why this was the case, but we

do not have any reason to believe that the control set of genes (genes associated with heart failure) should be biased toward higher levels of positive selection. The total number of control genes chosen for the analysis [20] was larger than that for any given GWAS category, but not vastly larger, to keep sample sizes similar in the comparisons. Three out of four of the GWASdb top hit set showed significant evidence for positive selection across the primates using CODEML (compared to 5/15 for the control set), but this result was not significant. The genes showing evidence for positive selection across the primates were as follows (proportion under positive selection and dataset in parentheses): ESRRG, INVS, RUNX1 (3/4: GWASdb); , C1QTNF6,COL4A2, F2, PLAUR (5/7: Crespi-SNPs); EDN2 (1/3: Crespi DE);, CRH, F5, HEXB (4/13: GWAS-Meta); CDKN1C, KDR, STOX1 (3/6: fetal); C4orf37, LMCD1 (2/7: PESNPdb). None of the imprinted genes showed significant evidence for positive selection. Significant evidence for site-specific selection on the human lineage was rare, but was found for HTR4A and LMCD1. Most genes showed some regions with significant evidence of positive selection from either the iHS test or the XP-EHH test, yet this was true of the control set of genes also.

### 6. Discussion

Our analyses of the FLT1 gene revealed abundant evidence for the action of positive selection on this gene at many levels. Over the long course of primate evolutionary history, there has apparently been strong positive selection acting at several different sites in the coding sequence. Furthermore, selection continued to affect this gene on the human branch of the primate evolutionary tree. Extended haplotype homozygosity tests reveal evidence for recent positive selection on this gene both across populations, and within the European and African populations. These results are consistent with the hypothesis that the FLT1 gene is subject to persistent positive selection in the context of antagonistic coevolution, in which new variants that provide an advantage to the embryo (in terms of acquisition of resources from the mother) are counteracted by changes in maternally expressed genes. Over evolutionary time, this can produce a signal of positive selection in both types of genes. Previous work on the FLT gene has also detected evidence for selection, particularly in the context of malaria. Muelenbachs et al. found evidence of selection against specific fetal genotypes in Tanzania at this locus in the context of placental malarial infection [32]. The SS genotype (associated with small repeat length number in the 3' UTR) showed lower fetal survival rates, possibly through an effect on inflammation mediated by interactions with the maternal immune system. The authors suggest these interactions may have influenced the evolution of sFlt1 regulatory mechanisms over the course of human evolution. These effects are likely to be different than (or in addition to) effects of antagonistic coevolution between maternal and fetal interests discussed above. The particular genetic variation associated with placental malaria involved repeat length, which would not be detected in our dN/ dS analyses. Further, the study focused on an African population, where selection pressure from malaria is likely to be more intense than within the European populations that comprised (part of) our EHH analyses.

Our broad scale comparisons of gene sets derived from GWAS analyses of association between specific genes and preeclampsia did not reveal that such gene sets are more likely to show signals of positive selection than a control set of disease related genes. In some cases, this result may have been affected by small sample sizes and the resulting low power of the statistical tests, but overall it seems reasonable to infer that genes associated with preeclampsia form a heterogenous group. Unlike immune system-related genes like those in the MHC, which tend to uniformly show strong signals of positive selection in the context of parasitehost coevolution [9], the preeclampsia-related genes are likely a mixed bag, with some being involved in coevolutionary arms races between conflicting fetal and maternal interests, and some simply associated with random contributions to the mechanical problems inevitably associated with the operation of any complex system. Nevertheless, our analyses have allowed us to identify a number of loci that are both strongly associated with preeclampsia and show signals of positive selection at several different levels. Some of these loci are, we think, excellent candidates for loci that are focal points of parent-offspring conflict between the mother and the fetus, with selection driving molecular evolutionary change at the level of the coding sequence or regulatory regions affecting expression. As noted above, a variety of genes showed evidence for both long and short-term positive selection in our analyses. The INVS locus produces inversin, and variants of this gene are associated with the development of nephronophthisis and other renal disorders [33].

Renal function is of course critical in regulating maternal blood pressure, and preeclampsia is often characterized by renal abnormalities. The F2 and F5 loci both code for factors essential to the coagulation cascade necessary for wound repair. The F2 locus codes for coagulation factor II, or thrombin. This protein plays a key role in vascular integrity, and exerts important effects on blood pressure. Previous research indicates that increased thrombin production (leading to "hypercoagulability") during pregnancy is associated with increased blood pressure and other aspects of hypertension in the maternal circulation characteristic of preeclampsia [34,35]. The F5 locus produces the Coagulation Factor V, a gene highly expressed in the placenta, with variation at this locus is known to affect several diseases related to elevated blood pressure, such as stroke [36]. The PLAUR locus encodes the receptor for the urokinase plasminogen activator, which has well-known associations with pathological circulatory conditions including cardiovascular disease (e.g. hardening of the arteries) and stroke [37]. The RUNX1 locus codes for a transcription factor involved in blood cell production and development. This transcription factor plays a key role in the development of placental hematopoietic stem and progenitor cells (HSPCs), and shows high expression in the context of the inflammatory conditions characteristic

of preeclampsia [38]. The COL4A2 locus codes for a component of type IV collagen, a key component of the endothelial basement membrane. Variation at this locus has been associated with blood vessel pathologies, such as cerebral small vessel disease [39]. The EDN2, or endothelin 2, locus codes for a vasoconstrictive protein, which has profound contractile effects on blood vessels [40]. The KDR (kinase insert domain receptor) locus codes for a receptor of vascular endothelial growth factor (VEGF), and mediates endothelial growth, development and survival. Preeclampsia is characterized by elevated VEGF levels, and exogenous administration of VEGF has been demonstrated to induce preeclampsia-like symptoms in mice [41]. The STOX1 (storkhead box 1) locus codes for a DNA binding factor. Overexpression of this gene in a mouse model causes a variety of effects (e.g. endothelial inflammation) that are characteristic of preeclampsia [42]. The CRH locus produces corticotropin releasing hormone. Excess production of this hormone in the maternal circulation can lead to earlier birth [43]. Hence this gene may be a good candidate for a gene for which higher expression levels favor maternal interests but are opposed by fetal interests. The CDKN1C (cyclin dependent kinase inhibitor 1C) is a paternally imprinted gene that regulates cell proliferation and interacts with VEGF. In the absence of CDKN1C expression, VEGF levels (which are associated with preeclampsia symptoms) are increased. Hence, expression of this gene is also likely to favor maternal interests.

We note that, according to the two-stage progression of preeclampsia, changes in the expression of embryonic genes that cause endothelial dysfunction in the maternal circulation may be triggered by the lack of normal trophoblast invasion characteristic of stage one. This can be seen as a high risk strategy on the part of the fetus in response to a lack of key nutrients [45]. Given the hypothesized long-term coevolutionary struggle between maternal and fetal gene expression, constrained by the potential harm to both parties, it is likely that patterns of gene expression by the fetus are in a finely matched balance with opposing effects from maternally expressed genes, and subject to over-expression given disturbance of the typical patterns of placental growth and maternal gene expression. In our opinion, further tests of the hypothesis of maternal-fetal conflict at the molecular level should focus on genes that show strong signals of selection over both recent and ancient evolutionary history, and those exhibit properties that are consistent with a role for the gene and/or gene products in conflict. Such properties could include specific expression by the fetus, differential effects on the maternal and fetal circulations, and allelic variant effects that directly impact key foci of conflict (as in the case of FLT1 variants and effects on blood pressure). Hence, moving forward, tests of selection can serve a useful purpose in identifying candidate genes that may be loci of antagonistic coevolution, but these genes should be further investigated from

multiple perspectives to determine if they are in fact involved in mediating conflicts between maternal and fetal interests.

### Acknowledgements

The authors would like to thank Bernard Crespi for a constructive review of the manuscript.

#### References

- [1]. R. L. Trivers, *Parent-offspring conflict*, Amer Zool, 14 (1974): 249-264.
- [2]. D. Haig, Genetic conflicts in human pregnancy. Q Rev Biol, 68: 495-532.
- [3]. D. Haig and M. Westoby, Parent specific gene expression and the triploid endosperm. Amer Nat, 134(1989): 147-155.
- [4]. D. Haig, and C. Graham, Genomic imprinting and the strange case of the insulin-like growth factor-II receptor. Cell, 64(1991): 1045-1046.
- [5]. D. Haig, Genomic imprinting: how good is the evidence? Ann Rev Genetics, 38(2004): 553-85.
- [6]. H. T. Yuan, D. Haig and S. A. Karumanchi, Angiogenic Factors in the Pathogenesi of Preeclampsia, Sem Dev Biol, 71 (2005): 297-312.
- [7]. T. Moore, Review: Parent-offspring conflict and the control of placental function. Placenta, 33 (2012): 33-36.
- [8]. B. Hollegaard, S. G. Byars, J. Lykke and J. J. Boomsma, Parent-Offspring Conflict and the Persistence of Pregnancy-Induced Hypertension in Modern Humans. PLoS ONE, 8 (2013): e56821.
- [9]. K. Summers, K. Roney, j. da Silva, G. Capraro, B. Kuthberston, G. Rosenthal et al. *Evolutionary dynamics of the DAB and DXB MHC loci in xiphophorus fishes*. Genetica, 135 (2009): 379:390.
- [10]. E. B. Chuong, W. Tong and H. E. Hoekstra, *Maternal–Fetal Conflict: Rapidly Evolving Proteins in the Rodent Placenta*. Mol Biol Evol, 27(2010): 1221–1225.
- [11]. M. E. Elliot and B. J. Crespi, Genetic recapitulation of human pre-eclampsia risk during convergent evolution of reduced placental invasiveness in eutherian mammals. Phil Trans Roy Soc B, 370(2015): 20140069.
- [12]. C. W. Redman and I. L. Sargent, *Latest advances in understanding preeclampsia*. Science, 308 (2005): 1592.
- [13]. E. J. Crosley, M. G. Elliot, J. K. Christians and B. J. Crespi, Placental invasion, preeclampsia risk and adaptive molecular evolution at the origin of the great apes: Evidence from genomewide analyses. Placenta, 34(2013): 127-132.
- [14]. S. A. Karumanchi, S. E. Maynard, I. E. Stillman, F. H. Epstein and V. P. Sukhatme, *Preeclampsia: a renal perspective*. Kidney Int, 67(2005): 2101-13.
- [15]. J. M. Roberts, *Endothelial dysfunction in preeclampsia*. Semin Reprod Endocrinol 16 (1988): 5–15.
- [16]. K. J. Gray, R. Saxena and S. A. Karumanchi, Genetic predisposition to preeclampsia is conferred by fetal DNA variants near, a gene involved in the regulation of angiogenesis. Am J Obstet Gynecol, 218(2017): 1-8.
- [17]. B. F. Voight, S. Kudaravalli, J. K. Pritchard and X. Wen, et al. A map of recent positive selection in the human genome. PLoS Biol. 4 (2008): e72.
- [18]. P. C. Sabeti, P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, C. Cotsapas et al. *International HapMap Consortium, Genome-wide*

detection and characterization of positive selection in human populations. Nature 449: (2007). 913-918 635.

- [19]. M. Taşan, G. Musso and T. Hao, Selecting causal genes from genome-wide association studies via functionally-coherent subnetworks, Nat Methods, 12(2015): 154–159.
- [20]. C. E. Kleinrouweler, M. Van Uitert, P. D. Moerland, C. Ris-Stalpers, J. A. M. Van der Post and G. B. Afink, *Differentially Expressed Genes in the Pre-Eclamptic Placenta: A Systematic Review and Meta-Analysis.* PLoS ONE, 8(2013): e68991.
- [21]. C. J. Petry, K. Beardsall and D. B. Dunger, *The potential impact of the fetal genotype on maternal blood pressure during pregnancy*. J Hypertens, 32 (2014): 1553–1561.
- [22]. G. Tuteja, E. Cheng, H. Papadakis and G. Bejerano, A comprehensive database of SNPs studied in association with preeclampsia. Placenta, 33 (2012): 1055-1057.
- [23]. M. S. Springer, R. W. Meredith, J. Gatesy, C. A. Emerling, J. Park and D. L. Rabosky, *Macroevolutionary dynamics and historical* biogeography of Primate diversification inferred from a species supermatrix. PLoS ONE, 7 (2012): e49521.
- [24]. Z. Yang. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci, 13 (1997): 555–556.
- [25]. R. Nielsen and Z. Yang, Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics, 148 (1988): 929–936.
- [26]. Z. Yang, Codon-substitution models for heteroge- neous selection pressure at amino acid sites, Genetics 155 (2000): 431-439.
- [27]. J. Z. Zhang, R. Nielsen and Z. H. Yang, Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol, 22(2005): 2472–2479.
- [28]. Petry, E. Maldonado, D. Almeida, T. Escalona, I. Khan, V. Vasconcelos and A. Antunes, *LMAP: Lightweight multigene* analyses in PAML. BMC Bioinformatics, 17 (2016): 354.
- [29]. R. Nielsen, Molecular signatures of natural selection. Ann Rev Genet, 39 (2005): 197-218.
- [30]. P.C. Sabeti, Detecting recent positive selection in the human genome from haplotype structure. Nature, 419 (2002):832-837.
- [31]. G. T. Sucak, K. Acar and A. Sucak, *Increased global fibrinolytic capacity as a clue for activated fibrinolysis in pre-eclampsia*. Blood Coagulation and Fibrinolysis, 17 (2006): 347-352.
- [32] Muelenbachs, M. Fried, J. Lachowitzer, T. K. Mutabingwa and P. E. Duffy, *Natural selection of alleles and their association with malaria resistance in utero*. Proc Natl Acad Sci, 105 (2008) :14488-14491.
- [33]. S. Bellavia, K. Dahan, S. Terryn, J. P. Cosyns, O. Devuyst and

Y. Pirson, A homozygous mutation in INVS causing juvenile nephronophthisis with abnormal reactivity of the Wnt/betacatenin pathway. Nephrol Dial Transplant, 25(2010): 4097-102.

- [34]. J. Bonnar, G. P. Mcnicol and A. S. Douglas, *Coagulation and fibrinolysis system in preeclampsia and eclampsia*. Br Med J, 113(1971): 905-910.
- [35]. Halima, A. B. Bhuiyanb, F. A. Azim, S. Khatunb, N. Kanayamaa, E. E. Maradnya, et al. *Blood Coagulation and Fibrinolysis in Eclamptic Patients and Their Correlation with the Clinical Signs*. Gynecol Obstet Invest, 39(1995): 97-102.
- [36]. A. Alhazzani, A. Kumar and M. Selim, Association between Factor V Gene Polymorphism and Risk of Ischemic Stroke: An Updated Meta-Analysis. J Stroke CerebVasc Dis, 27(2018): 1252-1261.
- [37]. M. Persson, G. Ostling, G. Smith and V.Hamrefors, Soluble urokinase plasminogen activator receptor: a risk factor for carotid plaque, stroke, and coronary artery disease. Stroke 45 (2014): 18-23.
- [38]. K. L. Ponder, A. Barcena, F. L. Bos, M. Gormley and Y. Zhou, Preeclampsia and Inflammatory Preterm Labor Alter the Human Placental Hematopoietic Niche. Reprod Sciences 23 (2016): 1179-1192.
- [39]. K. Rannikmäe, G. Davies and P. A. Thomson, Common variation in COL4A1/COL4A2 is associated with sporadic cerebral small vessel disease. Neurology 84 (2015): 918-926.
- [40]. E. Poli, C.Casoli, I. Spaggiari, R. Starcich and G. Bertaccini, Contractile effect endothelin-2 on the isolated human saphenous vein. Arch Int Pharm Ther 313 (1991): 108-119.617 (1991).
- [41]. Y. Murakami, T. Kobayashi, K. Omatsu, M. Suzuki and R.Ohashi, *Exogenous vascular endothelial growth factor can induce preeclampsia-like symptoms in pregnant mice*. Seminars in Thrombosis and Hemostasis 31 (2005): 307-313.
- [42]. A. Ducat, L Doridot and R Calicchio, Endothelial cell dysfunction and cardiac hypertrophy in the STOX1 model of preeclampsia. Sci Rep, 6 (2016): 19196.
- [43]. M. McLean, A. Bisits, J. Davies, R. Woods, P. Lowry and R. Smith, A placental clock controlling the length of human pregnancy. Nat Med, 1(1995): 460–463.
- [44]. E. Karteris, M. Vatish, E. W. Hillhouse and D. K. Grammatopoulos, Preeclampsia Is Associated with Impaired Regulation of the Placental Nitric Oxide-Cyclic Guanosine Monophosphate Pathway by Corticotropin-Releasing Hormone (CRH) and CRH-Related Peptides. J Clin Endocrinol Metab, 90(2005): 3680–3687.
- [45]. D. Haig, Intimate relations: evolutionary conflicts of pregnancy and childhood. In: Evolution in Health and Disease (S.C. Stearns & J.C. Koella, Eds.). Oxford University Press, Oxford, UK. 2008