biology and life sciences forum

MDPI

Proceeding Paper

In Silico Determination of Changes in Transcription Factor Binding Sites for the Preeclampsia Risk Haplotype in the Regulatory Region of the FLT1 Gene †

Nataliia Karpova *, Olga Dmitrenko and Ekaterina Arshinova

Institute of General Pathology and Pathophysiology (NIIOPP), 125315 Moscow, Russia

* Correspondence: nataliiakarpova.sp@gmail.com

† Presented at the 2nd International Electronic Conference on Biomolecules: Biomacromolecules and the Modern World Challenges, 1–15 November 2022; Available online: https://iecbm2022.sciforum.net/.

Abstract: Preeclampsia (PE) is one of the most common complications of pregnancy that occurs in 3–8% of pregnant women, being one of the top five causes of maternal morbidity and mortality. We found that PE-associated polymorphisms near the FLT1 gene are located in the same regulatory region. In combination, these polymorphisms can form a genetic pattern which affects the development of the pathology by forming a PE risk haplotype. When analyzing the changes in TPFS, which are characteristic only for the risk haplotype with a prevalence in the European population of 0.0825, we found that five TFBs change. The number of TFBSs for ELF1 and SPIB increases, whereas the number of TFBSs for POLR2A and KLF15 decreases (not expressed in the placenta).The newly emerged transcription factor binding site KAT5 acquires a promoter signature in the placenta only after 118 days of pregnancy, whereas before the 118th day, only the DNase signature is observed. Theoretically, the appearance of a new TFBS can increase the expression of FLT1, causing an imbalance of angiogenic/antiangiogenic factors, which is characteristic of PE.

Keywords: preeclampsia SNP; polymorphism; FLT1; risk haplotype; TFBs; TF

Citation: Karpova, N.; Dmitrenko, O.; Arshinova, E. In Silico Determination of Changes in Transcription Factor Binding Sites for the Preeclampsia Risk Haplotype in the Regulatory Region of the FLT1 Gene. *Biol. Life Sci. Forum* **2022**, *20*, 31. https://doi.org/10.3390/IECBM2022-13721

Academic Editor: Peter Nielsen

Published: 18 November 2022

1. Introduction

One of the most serious pregnancy disorders, preeclampsia (PE), affects both the mother and the fetus [1]. According to the meta-analysis conducted by Wang et al. (2021), preeclampsia affects 2–10% of pregnancies globally, with an average of 4.6% in 2021 [2,3]. No matter the patient's history of blood pressure, PE is characterized by a rise in blood pressure that typically occurs after the 20th week of pregnancy, with systolic blood pressure >140 mm Hg Art and diastolic blood pressure >90 mm Hg Art, and proteinuria or at least one other parameter indicating multiple organ failure [1,4,5].

Researchers at the Margaret Haig Maternity Hospital in New Jersey discovered the genetic predisposition to PE in the early 1960s. They found that preeclampsia is more prevalent in the sisters and daughters of preeclamptic moms [6]. Later studies by Swedish researchers revealed that the heritability of PE is estimated to be around 55% [7].However, at that time it remained unknown which particular gene or other region of the genome had undergone changes.

The fact that trisomy of the 13 chromosomes has historically been associated with PE allows us to look for the reason for preeclampsia to develop on this chromosome [8]. The causes of preeclampsia were searched for by analyzing changes in the concentration of various proteins in the blood. At 15–20 weeks of gestation, Flt-1 (which is located on 13 chromosomes), endoglin, and placental growth factor mRNA transcripts were examined by Sekizawa et al. (2010) in maternal whole blood [9]. Galaziou et al. (2021) and Lim et al. (2021) confirmed that PE patients had higher levels of sFlt-1 mRNA expression in their

blood [10,11]. Data obtained by Phupong et al. revealed that PE is the only pregnancy problem related to sFlt1/PlGF [12].

Genome-wide association studies (GWASs) carried out by McGinnis et al. (2019) originally discovered a connection between the T allele rs4769613 and preeclampsia (*p*-value of 5.4×10^{-11}), which was later supported by many case–control association studies [13,14]. T allele rs7318880 (p-value of 8×10^{-8}), in the analysis of the mother's genotype, and C allele rs4769612 (p-value of 4×10^{-14}), in the analysis of the child's genotype, were later found to be related to preeclampsia, according to Steinthorsdottir et al. (2020). Close-linkage disequilibrium exists between the variants rs4769612 and rs7318880 (r2 = 0.98) [15]. All these single-nucleotide polymorphisms (SNPs) are located in the FLT1 gene regulatory area.

According to this data, the aim of our study was to determine the regulatory potential of these SNPs, analyze the possible risk haplotype of PE and changes in transcription factor binding sites (TFBSs), and evaluate the information obtained through the FLT1 excess theory of PE etiology and pathogenesis.

2. Materials and Methods

2.1. Determination of SNPs Associated with Preeclampsia

We used the GWAS catalog [16] and PubMed to identify SNPs in FLT1 associated with preeclampsia in the maternal and child genomes.

2.2. Determination of Regulatory Areas Overlapping with SNPs Associated with Preeclampsia

Detected SNPs were mapped to the human genome using the UCSC Genome Browser to identify overlapping polymorphisms with regulatory regions [17].

2.3. Study of the Enhancer Signature of Regulatory Areas in the Placenta

For the determination of the enhancer signature of regulatory areas overlapping with SNPs, cCRE details from the ENCODE SCREEN were used [18].

2.4. Selection of SNPs

For SNPs in the regulatory region, we selected SNPs with a MAF of 1%, according to the UCSC Genome Browser Short Genetic Variants from the dbSNP release 155 track [19].

2.5. Preeclampsia Risk Haplotype Determination

Selected SNPs were identified as possible haplotypes in European populations using the LDhap Tool [20].

2.6. Obtaining the Reference DNA Sequence for the Region of Interest

For further analysis, the region chr13:28,563,912–28,564,785 (873 base pairs) was chosen, which overlaps with the regulatory regions of the FLT1 gene. The DNA sequence was obtained with the addition of 100 base pairs from the 5′ and 3′ ends using the Ensembl Genome Browser to assemble the human genome (GRCh38) [21]. To obtain sequences for all possible haplotypes, the corresponding nucleotides were changed manually using a standard text editor.

2.7. Determination of Change in Transcription Factor Binding Sites (TFBSs)

The identification of changes in TFBSs was performed using the Human TFDB for the reference genotype and for the risk genotype, with further identification of the change from the side of TFBSs only for those on a DNA strand, where the FLT1 gene was located. Additionally, a comparison was made for the genotype with reference alleles, with the exception of rs4769612 and rs4769613 (for which the reference allele is a risk allele) [22].

2.8. Determination of Transcription Factor Expression (TF)

Determination of each TF expression was performed using The Human Protein Atlas [23].

3. Results and Discussion

In the GWAS catalog, we found two polymorphisms in FLT1 associated with preeclampsia: rs4769612-C (p-value of 4×10^{-14}) and rs7318880-T (*p*-value of 8×10^{-8}), with rs4769612 associated with preeclampsia in the analysis of the child's genotype, and rs7318880 in the analysis of the mother's genotype. PubMed was able to detect one more SNP in the FLT1 gene regulatory area (rs4769613), which is associated with PE as a consequence of a genome-wide search for a relationship with PE. Additionally, the pathophysiology is more affected by fetal polymorphism than by maternal genetic variables [24].

The mapping of these SNPs was performed according to the latest assembly of the human genome (GRCh38/hg38) in the UCSC Genome Browser. All three SNPs (rs7318880, rs4769612, and rs4769613) are located in the regulatory region of the FLT1 gene, according to the cCREs ENCODE project and oRegAnno. According to the UCSC Genome Browser (oRegAnno) data, there are four regulatory elements overlapped with SNPs: OREG1191996, OREG1658246, OREG1688336, and OREG1537828. According to the cCRE details from the ENCODE SCREEN, this region contains the putative regulatory element EH38E1663332, of which the largest distal enhancer signature sharply increases at 16 weeks of gestation in the placenta and embryonic tissues, which can lead to changes in FLT1 expression [18]. In addition to rs4769612 and rs7318880, the EH38E1663332 region contains seven more polymorphisms with an MAF > 1%: rs7320190, rs12867370, rs4769613, rs74623647, rs7321138, rs76592233, and rs9579193 (Table 1). Of these, only rs7320190 and rs4769613 are mentioned in scientific articles and rs4769612 and rs7318880 in the GWAS database, whereas rs12867370 is associated with the risk of developing schizophrenia in offspring born to mothers with PE [25].

Table 1. Description for selected SNPs in FLT1 gene regulation areas (OREG1191996, OREG1658246, OREG1688336, OREG1537828, and EH38E1663332).

Gene	Variant (SNP)	Chr Location (GRCh38.p13)	MAF *	Consequence	Allele Substitution
FLT1	rs7320190	chr13:28564119	0.20121	None	T > C
	rs7318880	chr13:28564148	0.50888	None	C > T
	rs12867370	chr13:28564261	0.06396	None	G > A
	rs4769612	chr13:28564361	0.45778	None	C > T
	rs4769613	chr13:28564472	0.475826	None	C > A, C > T
	rs74623647	chr13:28564495	0.00021	None	G > A, G > T
	rs7321138	chr13:28564568	0.187966	None	T > C, T > G
	rs76592233	chr13:28564624	0.00021 **	None	C > A, C > G, C > T
	rs9579193	chr13:28564631	0.19905	None	G > A, G > T

* Minor allele frequency (MAF) for total population (release version: 20201027095038), according to dbSNP. ** MAF for rs76592233 prevalence varies by source. According to 1000Genomes, for global population of MAF of rs76592233 is 0.0104.

Searching for possible haplotypes in European populations by applying the LDhap Tool allowed us to identify four possible haplotypes for given SNPs (Figure 1).

A potential preeclampsia risk haplotype with a prevalence of 8.25% (rs7320190-C, rs7318880-T, rs12867370-A, rs4769612-C, rs4769613-C, rs74623647-G, rs7321138-C, and rs76592233-C) was identified after examining the prevalence of four different haplotypes (considering the risk alleles of maternal and fetal PE). Additionally, the frequency of this risk haplotype in the homozygous condition is 0.68%, which is comparable to the prevalence of early-onset preeclampsia, which is 0.38% [26].

Theoretically, the appearance of a new TFBS can increase the expression of FLT1, causing an imbalance in the angiogenic/antiangiogenic factors, which is characteristic of PE. Therefore, we identified changes in 83 transcription factor binding sites (TFBs), minus DNA strands, in the analysis of all four possible haplotypes using the Human TFDB. According to the Proteomic DB, only 40 transcription factors (TFs) are expressed in the placenta. When analyzing the changes in TFBSs, which are characteristic only for the risk

haplotype with a prevalence in the European population of 0.0825, we found that five TFBSs were changed (Table 2).

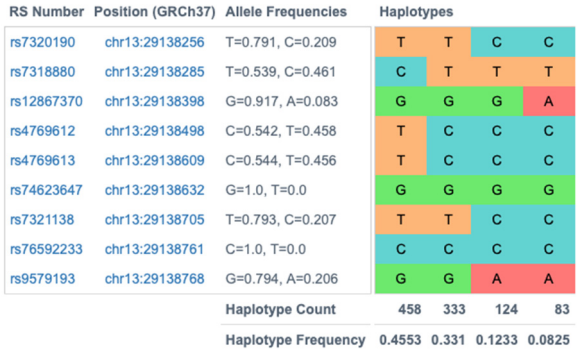

Figure 1. For polymorphisms rs7320190, rs7318880, rs12867370, rs4769612, rs4769613, rs74623647, rs7321138, rs76592233, and rs9579193, the prevalence of potential haplotypes was determined for EUR populations (SEU, TSI, FIN, GBR, and IBS). In addition, the risk haplotype (C T A C C G C C A) occurs at 8.25%.

Table 2. TFBS changes for each haplotype and its expression according to The Human Protein Atlas.

TFBS	TFBS Changes for Each Haplotype (Hap.) with Frequency				Type of TF	RNA Expression in Placenta (The Human Protein Atlas; nTPM)	Protein Expression in Placenta (The Human Protein Atlas; Types of Cells)
	Hap. 1 (0.4553) *	Hap. 2 (0.331)	Hap. 3 (0.1233)	Hap. 4 (0.0825)			
KAT5	0	0	0	1	Activator, acyltransferase, chromatin regulator, transferase	29.9	Decidual cells: medium Trophoblastic cells: high
ELF1	2	2	2	3	Activator, DNA-binding	48.6	Trophoblastic cells: medium
POLR2A	12	12	12	11		0.2	NA
KLF15	3	3	3	2	Activator, DNA-binding	1.7	Not detected
SPIB	2	2	2	3	Activator, DNA-binding	0.2	Decidual cells: medium Trophoblastic cells: medium

* Haplotypes correspond to those indicated in Figure 1.

The number of TFBSs for ELF1 and SPIB increases, whereas the number of TFBSs for POLR2A and KLF15 decreases (not expressed in the placenta).The newly emerged transcription factor binding site KAT5 acquires a promoter signature in the placenta only after 118 days of pregnancy, whereas before the 118th day, only a DNase signature is observed. Theoretically, the appearance of a new TFBS can increase the expression of FLT1, causing an imbalance in angiogenic/antiangiogenic factors, which is characteristic of PE.

4. Conclusions

As a result of this study, we were able to identify a potential preeclampsia risk haplotype (C T A C C G C C A), which has a prevalence of 0.68% for homozygotes and a rate of 0.38% for the start of preeclampsia in its early stages.

Additionally, we discovered that the most critical event is the formation of a novel TFBS KAT5, for which only a DNase signature is seen in the placenta up to day 118 of pregnancy, after which it gains a promoter signature. According to theory, the emergence

of a new TFBS can boost FLT1 expression, leading to an imbalance in angiogenic and antiangiogenic factors that is typical of PE.

Supplementary Materials: The presentation material of this work is available online at https://www.mdpi.com/article/10.3390/IECBM2022-13721/s1.

Author Contributions: Conceptualization, N.K.; methodology, N.K.; software, N.K.; validation, N.K., O.D. and E.A.; formal analysis, N.K.; investigation, N.K.; resources, N.K.; data curation, N.K; writing—original draft preparation, N.K; writing—review and editing, N.K., O.D. and E.A.; visualization, N.K.; supervision, N.K., O.D. and E.A.; project administration, N.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors express their gratitude to the supervisors of their PhD study, Nurbekova Malik Kubanychbekov and Budykina Tatiana Sergeevna, for their help with research and comprehensive support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Preeclampsia. *Eclampsia. Edema, Proteinuria and Hypertensive Disorders during Pregnancy, Childbirth and the Postpartum Period*; Clinical Guidelines: Moscow, Russia, 2021. (In Russian)
2. Osungbade, K.O.; Ige, O.K. Public health perspectives of preeclampsia in developing countries: Implication for health system strengthening. *J. Pregnancy* **2011**, *2011*, 481095. [CrossRef]
3. Wang, W.; Xie, X.; Yuan, T.; Wang, Y.; Zhao, F.; Zhou, Z.; Zhang, H. Epidemiological trends of maternal hypertensive disorders of pregnancy at the global, regional, and national levels: A population-based study. *BMC Pregnancy Childbirth* **2021**, *21*, 364. [CrossRef]
4. Hypertension in pregnancy. Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy. *Obs. Gynecol.* **2013**, *122*, 1122–1131. [CrossRef]
5. ACOG Practice Bulletin No. 202 Summary: Gestational Hypertension and Preeclampsia. *Obs. Gynecol.* **2019**, *133*, 1. [CrossRef]
6. Chesley, L.C.; Cosgrove, R.A.; Annitto, J.E. Pregnancies in the sisters and daughters of eclamptic women. *Obstet Gynecol.* **1962**, *20*, 39–46. [CrossRef] [PubMed]
7. Lie, R.T.; Rasmussen, S.; Brunborg, H.; Gjessing, H.K.; Lie-Nielsen, E.; Irgens, L.M. Fetal and maternal contributions to risk of pre-eclampsia: Population-based study. *BMJ* **1998**, *316*, 1343–1347. [CrossRef]
8. Maynard, S.E.; Venkatesha, S.; Thadhani, R.; Karumanchi, S.A. Soluble Fms-like Tyrosine Kinase 1 and Endothelial Dysfunction in the Pathogenesis of Preeclampsia. *Pediatr. Res.* **2005**, *57*, 1–7. [CrossRef] [PubMed]
9. Sekizawa, A.; Purwosunu, Y.; Farina, A.; Shimizu, H.; Nakamura, M.; Wibowo, N.; Rizzo, N.; Okai, T. Prediction of pre-eclampsia by an analysis of placenta-derived cellular mRNA in the blood of pregnant women at 15-20 weeks of gestation. *BJOG* **2010**, *117*, 557–564. [CrossRef]
10. Galaziou, A.; Filidou, E.; Spathakis, M.; Arvanitidis, K.; Arzou, B.C.; Galazios, G.; Koutlaki, N.; Nikolettos, N.; Kolios, G. Imbalance of growth factors mRNA expression associated with oxidative stress in the early pregnancy loss. *J. Matern. Fetal Neonatal Med.* **2021**, *35*, 1–7. [CrossRef]
11. Lim, S.; Li, W.; Kemper, J.; Nguyen, A.; Mol, B.W.; Reddy, M. Biomarkers and the Prediction of Adverse Outcomes in Preeclampsia: A Systematic Review and Meta-analysis. *Obs. Gynecol.* **2021**, *137*, 72–81. [CrossRef]
12. Phupong, V.; Areeruk, W.; Tantbirojn, P.; Lertkhachonsuk, R. Soluble fms-like tyrosine kinase 1 and placental growth factor ratio for predicting preeclampsia in elderly gravida. *Hypertens Pregnancy* **2020**, *39*, 139–144. [CrossRef]
13. Ratnik, K.; Rull, K.; Aasmets, O.; Kikas, T.; Hanson, E.; Kisand, K.; Fischer, K.; Laan, M. Novel Early Pregnancy Multimarker Screening Test for Preeclampsia Risk Prediction. *Front. Cardiovasc. Med.* **2022**, *9*, 932480. [CrossRef]
14. Kikas, T.; Inno, R.; Ratnik, K.; Rull, K.; Laan, M. C-allele of rs4769613 Near FLT1 Represents a High-Confidence Placental Risk Factor for Preeclampsia. *Hypertension* **2020**, *76*, 884–891. [CrossRef] [PubMed]
15. Steinthorsdottir, V.; McGinnis, R.; Williams, N.O.; Stefansdottir, L.; Thorleifsson, G.; Shooter, S.; Fadista, J.; Sigurdsson, J.K.; Auro, K.M.; Berezina, G.; et al. Genetic predisposition to hypertension is associated with preeclampsia in European and Central Asian women. *Nat. Commun.* **2020**, *11*, 1–4. [CrossRef] [PubMed]

16. Buniello, A.; MacArthur, J.A.L.; Cerezo, M.; Harris, L.W.; Hayhurst, J.; Malangone, C.; McMahon, A.; Morales, J.; Mountjoy, E.; Sollis, E.; et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **2019**, *47*, D1005–D1012. [CrossRef] [PubMed]
17. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, D. The human genome browser at UCSC. *Genome Res.* **2002**, *12*, 996–1006. [CrossRef]
18. Moore, J.E.; Purcaro, M.J.; Pratt, H.E.; Epstein, C.B.; Shoresh, N.; Adrian, J.; Kawli, T.; Davis, C.A.; Dobin, A.; Kaul, R.; et al. Expanded Encyclopaedias of DNA Elements in the Human and Mouse Genomes. *Nature* **2020**, *583*, 699–710.
19. Sherry, S.T.; Ward, M.H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E.M.; Sirotkin, K. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **2001**, *29*, 308–311. [CrossRef]
20. Machiela, M.J.; Chanock, S.J. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **2015**, *31*, 3555–3557. [CrossRef]
21. Rigden, D.J.; Fernández, X.M. The 2022 Nucleic Acids Research database issue and the online molecular biology database collection. *Nucleic Acids Res.* **2022**, *50*, D1–D10. [CrossRef]
22. Hu, H.; Miao, Y.R.; Jia, L.H.; Yu, Q.Y.; Zhang, Q.; Guo, A.Y. AnimalTFDB 3.0: A comprehensive resource for annotation and prediction of animal transcription factors. *Nucl. Acids Res.* **2019**, *47*, D33–D38. [CrossRef] [PubMed]
23. Karlsson, M.; Zhang, C.; Méar, L.; Zhong, W.; Digre, A.; Katona, B.; Sjöstedt, E.; Butler, L.; Odeberg, J.; Dusart, P.; et al. A single-cell type transcriptomics map of human tissues. *Sci Adv.* **2021**, *7*, eabh2169. [CrossRef]
24. McGinnis, R.; Steinthorsdottir, V.; Williams, N.O.; Thorleifsson, G.; Shooter, S.; Hjartardottir, S.; Bumpstead, S.; Stefansdottir, L.; Hildyard, L.; Sigurdsson, J.K.; et al. Variants in the fetal genome near FLT1 are associated with risk of preeclampsia. *Nat. Genet.* **2017**, *49*, 1255–1260. [CrossRef] [PubMed]
25. Malaspina, D.; Kranz, T.; Kleinhaus, K.; Daboul, S.; Rothman, K.; Gilman, C.; Getz, M.; Harlap, S.; Friedlander, Y. Short duration of marriage at conception as an independent risk factor for schizophrenia. *Schizophr. Res.* **2019**, *208*, 190–195. [CrossRef] [PubMed]
26. Lisonkova, S.; Joseph, K.S. Incidence of preeclampsia: Risk factors and outcomes associated with early- versus late-onset disease. *Am. J. Obstet. Gynecol.* **2013**, *209*, 544.e1–544.e12. [CrossRef]