

Profiling Gene Specific Changes to RNA Editome in Preeclampsia

Saam Hasan^{1,2}, Shahriyar Mahdi Robbani^{1,2}, Tamanna Afroze², Gias U Ahsan³, and Muhammad Maqsud

Hossain^{1,2*}

¹Department of Biochemistry and Microbiology, North South University, Dhaka, Bangladesh

²NSU Genome Research Institute, North South University, Dhaka, Bangladesh

³Department of Public Health, North South University, Dhaka, Bangladesh

*Corresponding author:

Dr. Muhammad Maqsud Hossain

Mailing address: Plot 15, Block B, Bashundhara, Dhaka 1229, Bangladesh

Tel.: +[02-55668200](tel:02-55668200)

Mobile phone: +88-01829153284

E-mail: muhammad.maqsud@northsouth.edu

Abstract

RNA editing is a form of post-transcriptional modification that results in changes to the messenger RNA sequence. At the onset of the study we focused on detecting the changes in RNA editing patterns in cell lines exposed to hypoxic conditions followed by the detection of changes in RNA editing patterns in the fetuses of preeclamptic mothers using publicly available RNA sequence data from the NCBI SRA database. The results showed an increase in RNA editing activity in hypoxic cell lines and a decrease in RNA editing activity in the fetuses with preeclamptic mothers. A total of 85 genes common in the cell lines and 33 in the fetus disease models and not present in controls were identified as harboring editing sites in exonic, downstream, upstream or splicing regions. Subsequently we focused on unique editing sites in genes and categorized in order of relevance to Preeclampsia as A, B and C (A being most closely related to the disease and C the least). The genes implicated ones involved in respiration chains, blood cell growth, cytokine and complement activation. Among the most significant of the genes were CTSB, GSR, CASP10, and MAPK13. Total number of common editing sites were found in different conditions and these were 667 for cell lines and 23 for fetuses. Validation of these variations in a larger samples size determines refined editing sites which could be used as potential diagnostic markers for intervention.

Keywords: RNA Editing, Preeclampsia, Post Transcriptional Modification

Introduction

Statistically, the significance and severity of Preeclampsia as a condition is illustrated by its considerably high incidence rate among both women who enter a normal pregnancy versus women who enter pregnancy with prior hypertension. These numbers stand at 5 percent of normal pregnancy and 25 of pregnant women with prior hypertension suffering from preeclampsia. The disease itself as evident from its variety of clinical manifestations affects multiple organs, a trait not common to most other gestational diseases. It is one of a myriad of pregnancy associated hypertensive disorders. These range from simple elevations of blood pressure in chronic hypertension prior to the 20 week mark of pregnancy and gestational hypertension subsequent to the 20 week mark, to the more clinically worse conditions such as preeclampsia, its culminated state, eclampsia, and the Hemolysis, Elevated Liver Enzymes and Low Platelets or HELLP syndrome (5). Preeclampsia in general is taken as a collection of conditions and is classified based on differences in symptoms such as whether or not there is profound hypertension (6), early vs late gestational hypertension (7), signs of the onset of HELLP and others. In terms of origin, preeclampsia is defined as either placental or maternal in origin. An area that is the subject of intense scrutiny and study at this point in time. The major and most clearly observable clinical and pathological manifestation of preeclampsia is elevated proteinuria. This has led to greater focus dedicated to the renal component of the pathological state brought about by the disease (5). RNA editing was first discovered in the protozoa *Trypanosoma brucei* where the insertion of additional cytidines were observed in the mRNA for the cytochrome c oxidase II gene. This was an example of non template nucleotide insertion (23). Another example of RNA editing occurs in the fungus *Physarum polycephalum* where base insertions are known to occur at characteristic editing sites (1). Examples of RNA editing in humans include the transcript for the Glutamate B receptor gene that encodes the glutamate gated membrane channels which contains adenosine to inosine editing sites (2). The adenosine to inosine conversions, which are the predominant editing mechanism in mammals (3), holds special significance aside from altering the amino acid sequence of the final protein. These changes have a variety of other effects on the transcript even before the translation process begins. These modifications alter mRNA stability, alter or direct localization of the messenger RNAs, influence the splicing pattern and may even alter splice sites and carry an effect on the efficiency with which the transcript is translated(4). Prior research has already implicated a possible role for RNA editing in a number of disease models such as Glioblastoma Multiforme, Uterine Corpus, Lung Adenocarcinoma, etc (24). Our aims here were to identify potential

mimicked the Preeclampsia disease state which may reflect the kind of differential editing we would observe in actual Preeclampsia patients. The second group is data from fetuses born to Preeclamptic mothers. Here we looked for differential editing that may predispose these individuals to Preeclampsia later in life. The goal being to identify a hereditary component of Preeclampsia mediated by RNA editing events.

Methodology

RNA Seq Datasets

Here we chose the human hg38 build or GRCh38 which was published on 2013 by the Genome Research Consortium (10). We used patch 11 of the build, published in June 2017. Next, the NCBI SRA database was used for obtaining experimental RNA seq data associated with Preeclampsia or Preeclampsia patients (Table 1) (20). The fastq files were obtained from the SRA database using the fastq dump command available on SRA toolkit(11). Data from two separate experiments was used. The first was an experiment involving RNA sequencing of fetuses from Preeclampsia patients. This we used to gauge the possible hereditary component of RNA editing mediated influence on Preeclampsia (see table 1). We chose data from female fetuses whose mothers were affected by Preeclampsia (20). The second experimental data we used was an experiment involving two separate cell lines that were stimulated in hypoxic conditions to mimic Preeclampsia (table 1). These involved two placental cell lines, JEU 3 and BeWo. Data available at NCBI BioProject under accession number PRJNA294057 and NCBI SRA under respective accession number for individual runs. Details about the runs and their SRA accessions can be found in supplementary Table 3.

Table 1: Accession Numbers of each dataset used in analysis and model each belonged to (control or Preeclampsia/hypoxic)

| Run | Accession number | Experiment | Sample |
|--------|------------------|------------|--------------|
| Run 1 | SRR3317125 | Fetus | Preeclampsia |
| Run 2 | SRR3317136 | Fetus | Preeclampsia |
| Run 3 | SRR3317145 | Fetus | Control |
| Run 4 | SRR2187123 | Cell Line | Hypoxic |
| Run 5 | SRR2187133 | Cell Line | Hypoxic |
| Run 6 | SRR2187128 | Cell Line | Control |
| Run 7 | SRR2187138 | Cell Line | Control |
| Run 8 | SRR2187122 | Cell Line | Hypoxic |
| Run 9 | SRR2187132 | Cell Line | Hypoxic |
| Run 10 | SRR2187127 | Cell Line | Control |

SNPs using less stringent mismatch filters and detection of more potential positives at this stage. We also set the program to directly generate an indexed bam and coordinate sort it from leftmost coordinate as needed by the downstream GATK tools used. All other parameters were set as prescribed by Dobin *et al*(8) for mapping of RNA sequence data.

Calling RNA Editing Sites

A modified version of the GATK RNA Seq best practices pipeline was used as defined by BROAD Institute for calling our variants (9). The pipeline, in sequence of steps, included adding read groups using Picard, splitting N Cigar reads using GATK's SplitNCigarReads function, this step also involved reassigning a mapping quality score of 60 to good calls. Reason being STAR assigns a default quality score of 255 to all good alignments which GATK cannot read. Then we re-calibrated base calls using GATK's variant recalibrator before finally calling variants. The realigning indels step was skipped as HaplotypeCaller no longer requires this step. Variants were called using HaplotypeCaller which is GATK's first choice variant caller, after it replaced UnifiedGenotyper. The variants were filtered using hard filters for finding groups of 3 SNPs that occur within 35 base pairs of each other, Fisher Strand Bias of 30 or above and Qual by Depth scores of below 2.

Common and rare SNPs were filtered using GATK and Annovar respectively. GATK's VariantAnnotator was used for marking common SnpS using an Hg38 dbsnp reference as a soft filter. These were later removed altogether. Annovar's Avsnp [1] soft filter was used to again apply a soft filter to rare and hence not previously undiscovered variants. These too were later removed. The variant calls were annotated as per their genomic and gene based location using Annovar's geanno annotation algorithm (22).

Results

Genomic and Functional Location of Editing Sites-Cell Lines

The cell line analysis involved two different placental cell lines, the BeWo and Jeu 3 cell lines. Each cell line was exposed to hypoxic conditions for a period of either 16 hours or 24 hours and a corresponding control existed for each of those conditions making a total of 8 cell line runs. The total number of RNA editing events in each run is summarized in Table 2. Very little difference exists between the number of edits found in the hypoxic and control samples for the 16 hour samples. However, the 24 hour samples show a drastic increase in the number of edits for the hypoxic runs in both cell lines. This is possible indication of increased RNA editing activity in cells experiencing extended periods of hypoxic conditions.

Table 2: Number of edits for each run analyzed. Number of edits increases for both 16 and 24 hour hypoxic runs compared to corresponding controls for both cell lines. No common temporal pattern can be established for both cell lines from 16 to 24 hour results.

| Cell Line | Time | Sample | Frequency of Edits |
|-----------|------|---------|--------------------|
| BeWo | 16h | Hypoxic | 28,950 |
| BeWo | 16h | Control | 26,008 |
| Jeu | 16h | Hypoxic | 40,445 |
| Jeu | 16h | Control | 41,635 |
| BeWo | 24h | Hypoxic | 40,669 |
| BeWo | 24h | Control | 24,318 |
| Jeu | 24h | Hypoxic | 29,508 |
| Jeu | 24h | Control | 18,231 |

After identification of RNA editing sites in all the runs, the common editing sites from control run was subtracted from its corresponding hypoxic run in order to identify the types of RNA editing variants unique to the cell lines exposed to hypoxic conditions. Figure 1 summarizes the common editing sites between the various runs. The vast majority of editing sites were unique to each run although we found a total of 667 common RNA editing sites between all four hypoxic runs. The two 16 hour runs had 968 common RNA editing sites which increased to 1,118 common sites for the 24 hour runs which lends to further indication of increased RNA editing activity with increased exposure to hypoxia.

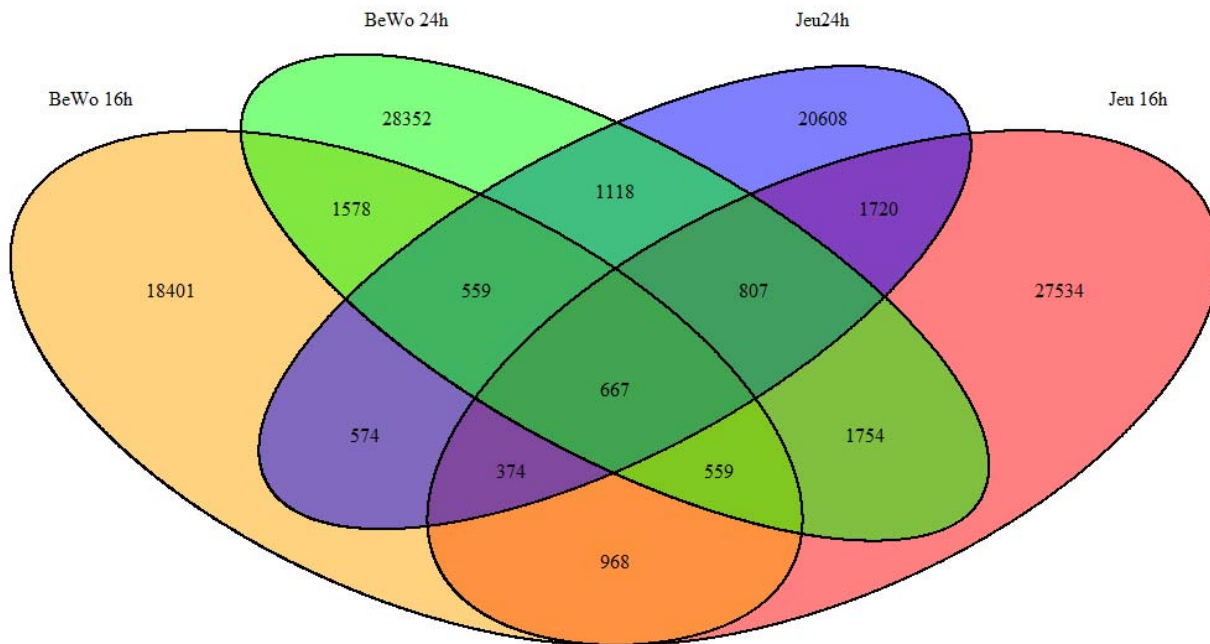


Fig 1: Frequency of common RNA editing sites among the hypoxic cell line runs.

The vast majority of the unique RNA editing sites uncovered were in the intronic region of all four runs, with the BeWo 24 hour run being the highest at 25,109 edits and Jeu 24 hour run being the lowest at 17,185 edits. A large number of editing sites were also found in the intergenic region which amounted to roughly 5,000 edits in each run with the exception of the Jeu 16 hour run where it was 7,141. All further edits were found in small numbers spread throughout the rest of the categories, with most significant being the UTR3 region with approximately 1,200 edits. This is particularly

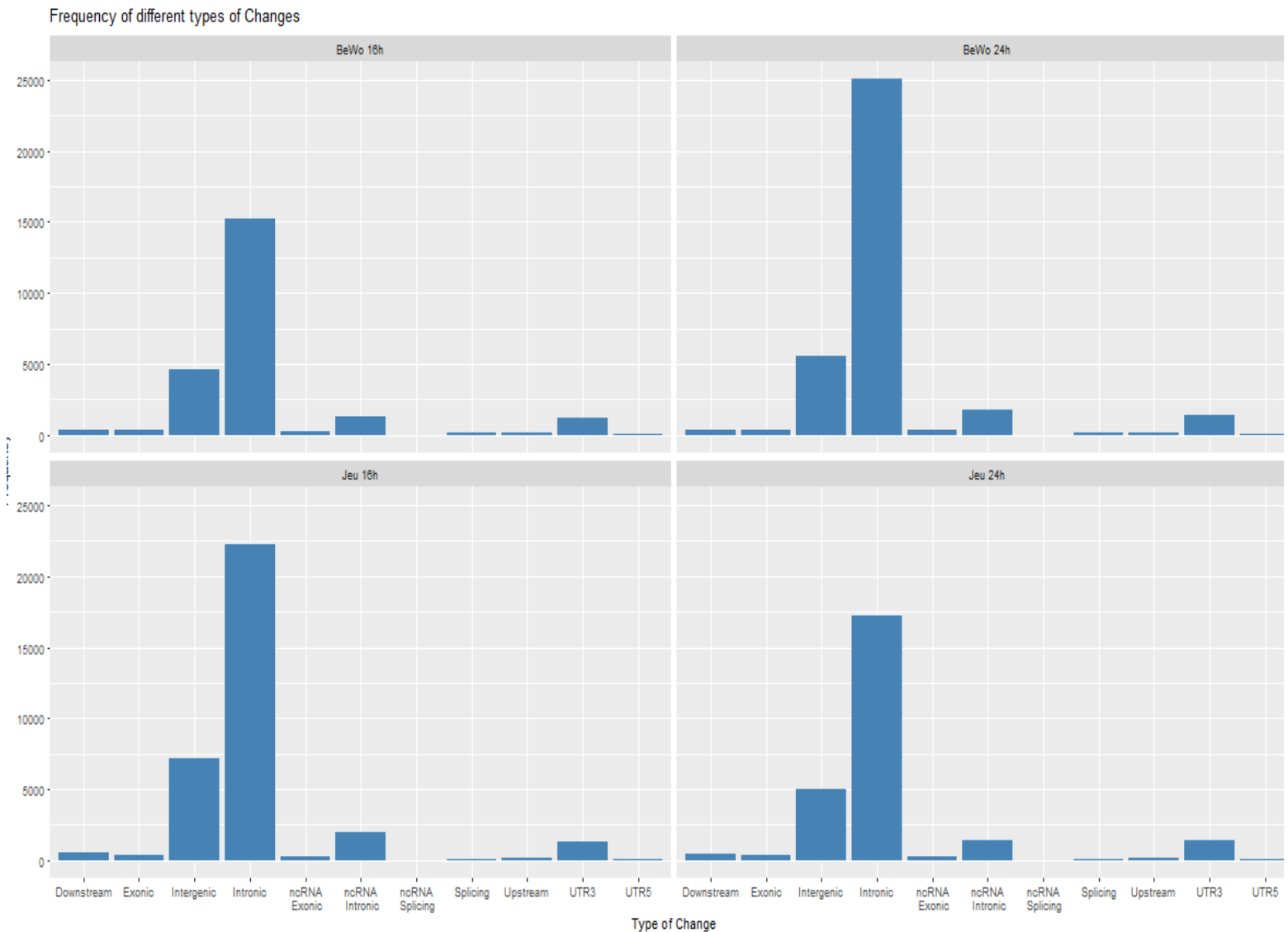


Fig 2: Frequency of different types of changes in the cell line runs.

The percentage of various types of base changes are shown in Figure 3. As expected, the number of A-to-G variants were the highest in each run ranging from 38.2% (BeWo 16h) to 40.4% (BeWo 24h). A very large number of T-to-C variants were also observed and ranged from 37.5% (BeWo 16h) to 40.1% (Jeu 16h). Roughly 5% of the changes in each run were attributed to C-to-T and G-to-A variants. All other changes accounted for less than 2% of the total variants in each run. All runs showed similar base change patterns, although a slight increase in A-to-G variants and T-to-C variants with

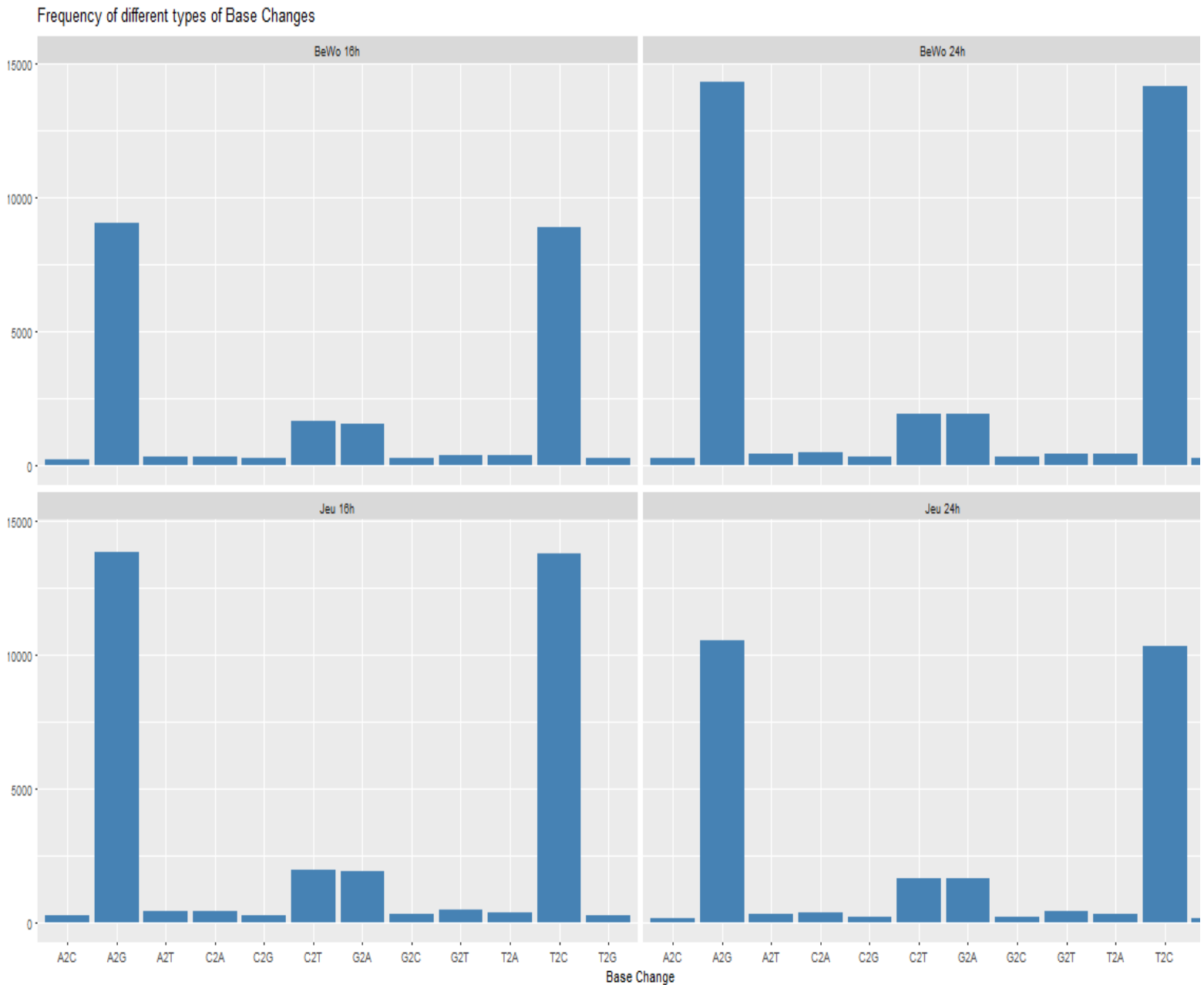


Fig 3: Frequency of different base changes in the cell line runs.

Our study also involved finding the number of RNA variants per chromosome. The results are displayed in Figure 4. The number of RNA variants for each chromosome remained nearly identical for both 16 hour runs. However, there is a significant increase in the number of RNA editing events in the cell lines exposed to hypoxic conditions for 24 hours. Chromosome 19 and 17 showed the highest peak in both cell lines which could indicate a functional relevance to preeclampsia. Since the underlying pattern of editing remains constant in both the hypoxic and control runs, another possibility is that the discrepancy in the number of edits occurred due to different sequencing depths of each run.

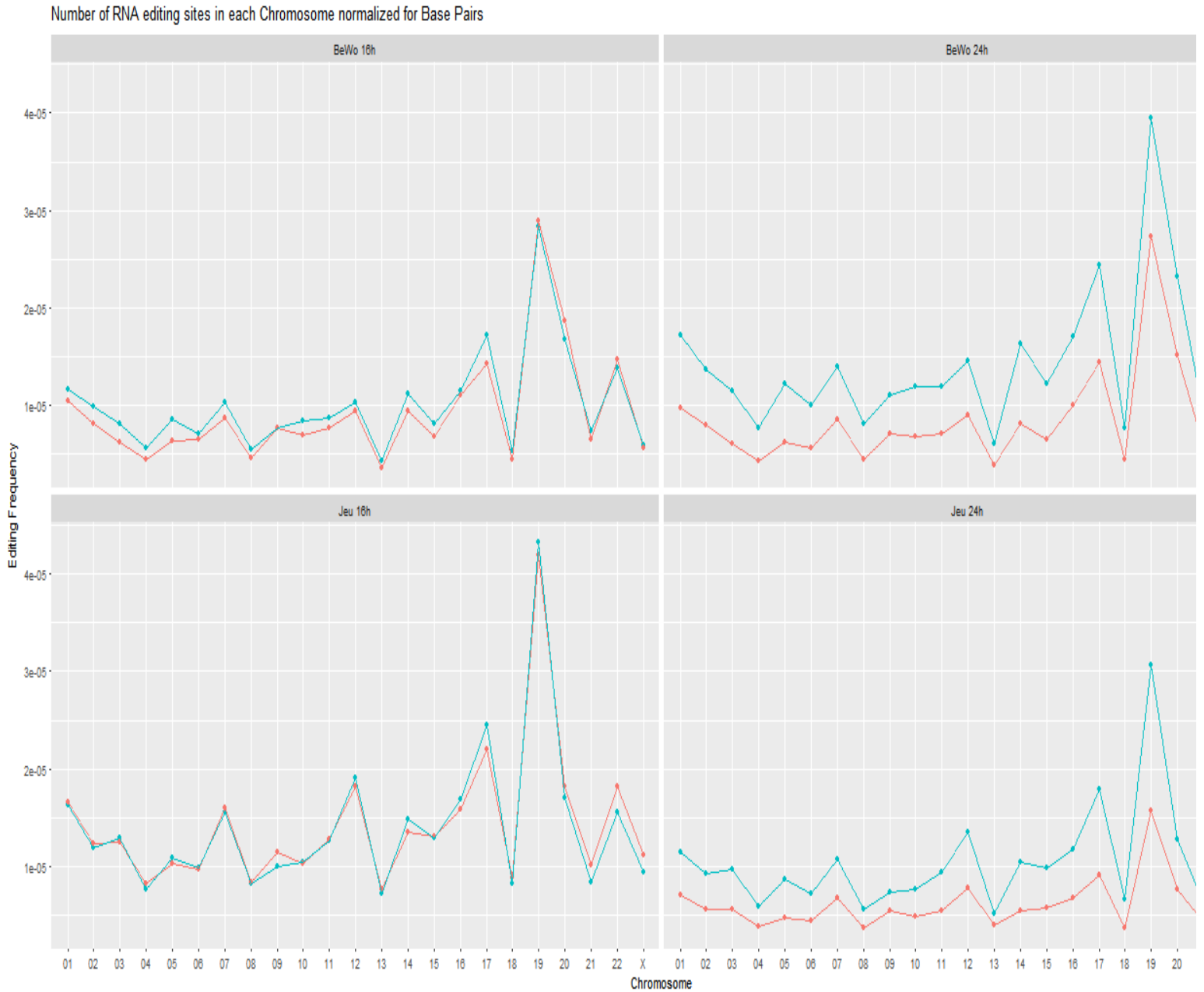


Fig 4: Number of RNA editing sites in each chromosome normalized for base pairs.

Another facet of our study involved analyzing the expression data for the ADAR and ADARB1 genes. The expression data of the genes from the study, as available on NCBI GEO(21) was compared to the number of editing sites found per run. Figure 6 shows a positive correlation between the expression of the ADAR, ADARB1, ADARB2 and Total ADAR genes with the number of RNA edits for both the hypoxic and control runs. The majority of the hypoxic runs showed a higher number of editing sites compared to the controls but a lower expression of the ADAR gene and Total ADAR

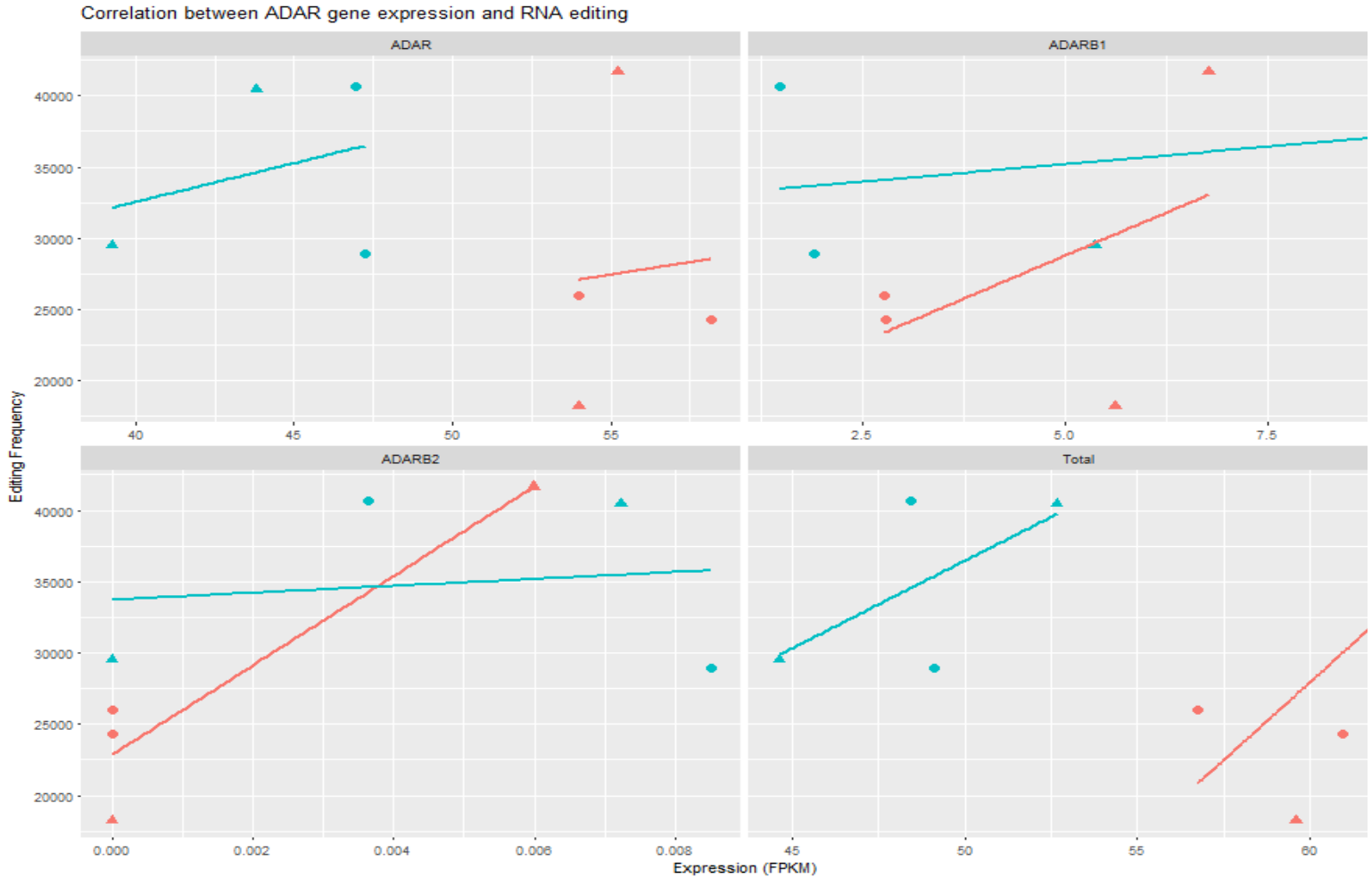


Fig 5: Correlation between ADAR gene expression and number of RNA editing events.

As our final goal, a list of genes were identified as potentially edited in Preeclampsia. These genes all possessed common RNA editing sites across all the hypoxic runs but none of the control runs. Table 3 shows these in detail. They were classified into categories based on how relevant they are to preeclampsia, with Category A being the highest and Category C being the lowest (Table 4).

Table 3A: List of genes containing RNA editing sites with relevance to preeclampsia in the cell line experiments.

Category A being most relevant and category C the least.

| Gene Symbol | Gene Name | Category |
|-------------|----------------------------------------------------------------------------------------------|----------|
| GSR | Glutathione reductase | A |
| MAPK13 | Mitogen-activated protein kinase 13 | A |
| PLEKHA2 | Pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 2 | A |
| SLC9A7 | Solute carrier family 9, subfamily A (NHE7, cation proton antiporter 7), member 7 | A |
| ATP5S | ATP synthase, H ⁺ transporting, mitochondrial Fo complex, subunit s (factor B) | B |
| CASP10 | Caspase 10, apoptosis-related cysteine peptidase | B |
| ENAH | Enabled homolog (Drosophila) | B |
| HSD17B12 | Hydroxysteroid (17-beta) dehydrogenase 12 | B |
| NDUFV3 | NADH dehydrogenase (ubiquinone) flavoprotein 3, 10kDa | B |
| NOA1 | Nitric oxide associated 1 | B |
| TOR1AIP1 | Torsin A interacting protein 1 | B |
| ALDH6A1 | Aldehyde dehydrogenase 6 family, member A1 | C |
| CFLAR | CASP8 and FADD-like apoptosis regulator | C |
| SZT2 | Seizure threshold 2 homolog (mouse) | C |

Genomic and Functional Location of Editing Sites-Fetus

The second part of our experiment involved identifying the RNA editing events in female fetuses of preeclamptic mothers. We performed a total of 4 fetal runs using 2 fetuses from preeclamptic mothers and 2 controls. A total of 2,571 RNA editing sites were found in the preeclamptic fetuses while a total of 8,207 editing sites were found in the control fetuses. A possible reason for this significant increase in the number of RNA editing sites in the fetus controls is that one of the control runs had a higher sequencing depth which lead to the discovery of more editing sites. The editing sites found in the control runs were then subtracted from the 2,571 RNA editing sites found in the preeclamptic runs to reveal 1,994 RNA editing sites unique to the preeclamptic runs. Out of these sites, only 23 editing sites were common between the preeclamptic runs as shown in Figure 6.

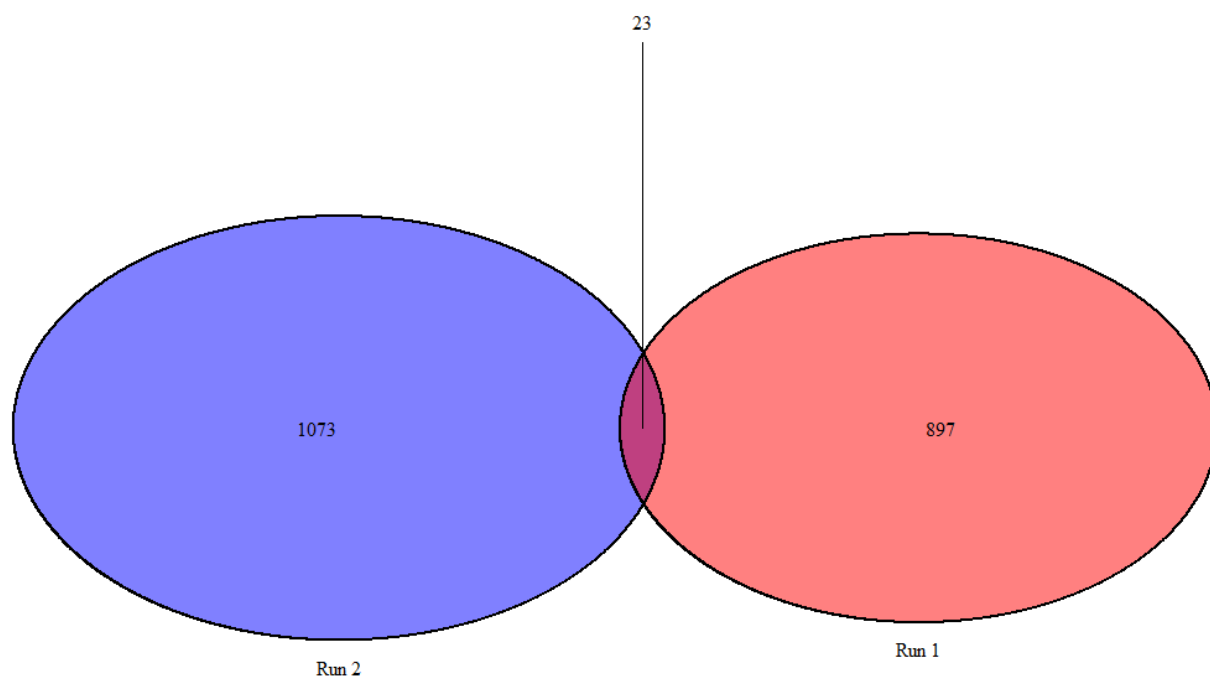


Fig 6: Number of common RNA editing sites between the preeclamptic fetal runs.

The different types of base changes are summarized in Figure 7. Once again, the intronic region has the highest number of unique RNA edits at 709, however a significantly large number of exonic RNA editing sites were also present. This could imply that the preeclamptic conditions of the mother increased RNA editing events in the fetus. Significant numbers of edits were also seen in the UTR3 region which may lead to change in gene expression as well as the Intergenic region.

Frequency of different types of Changes in Fetal Runs

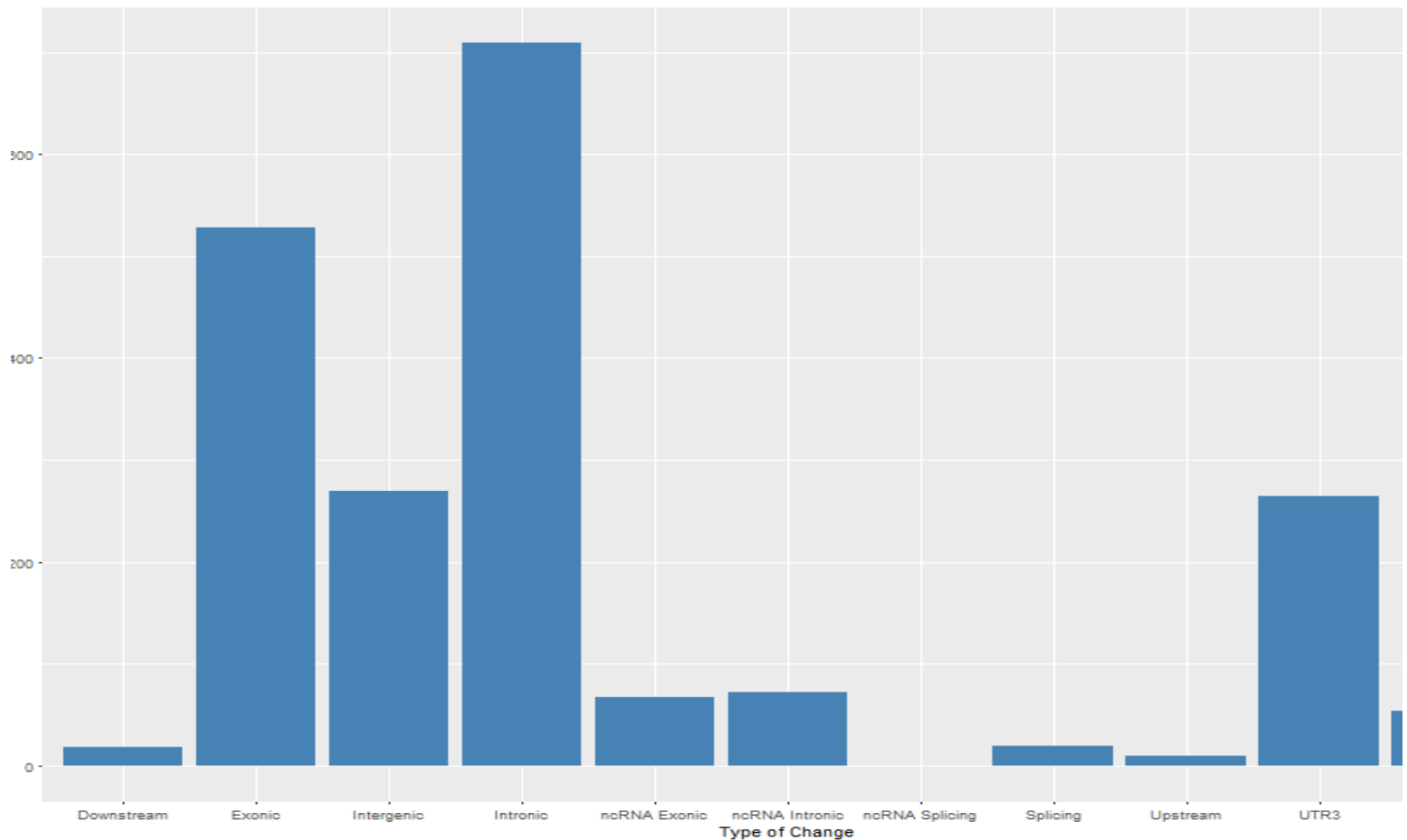


Fig 7: Frequency of different types of changes in the fetal runs.

The fetal runs showed an unexpectedly small percentage of A-to-G variants at only 18.7%. The percentage of T-to-C edits were also reduced but were still larger than A-to-G edits at 19.5%. All other types of base changes showed a large increase and most were above 5%. T-to-A variants showed the largest increase at 11.6%. This data is summarized in Figure 8.

Frequency of different types of Base Changes in Fetal runs

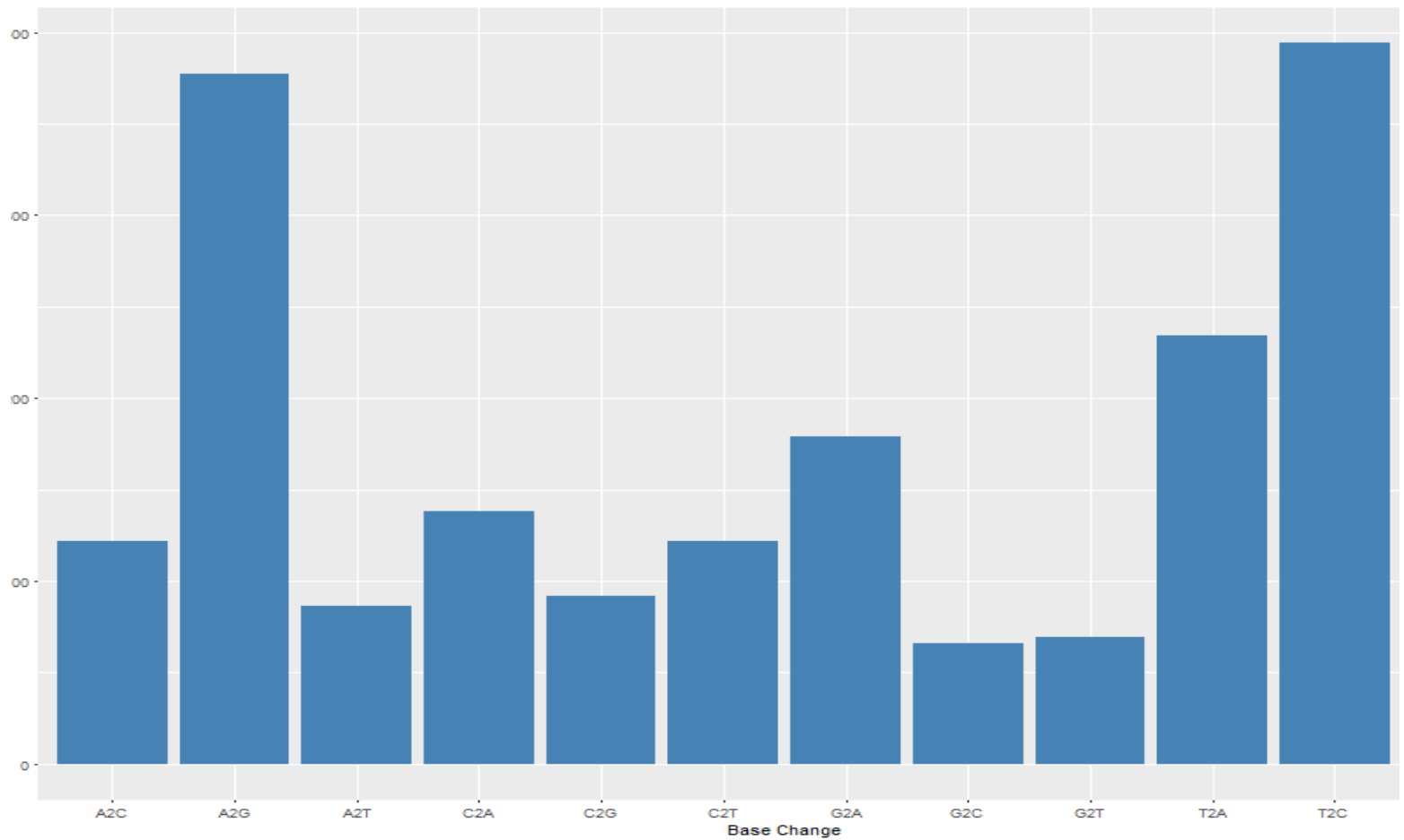


Fig 8: Frequency of different base changes in the fetal runs.

The RNA editing events in each chromosome showed a peculiar difference in the fetal runs. The number of edits in the preclamptic fetuses was actually lower than those of the control runs. As mentioned earlier, a possible cause for this is difference in sequencing depth. The overall pattern for editing events in each chromosome remained similar with peaks chromosome 17 and 19 but with an additional peak at chromosome 10. This can be seen in Figure 9.

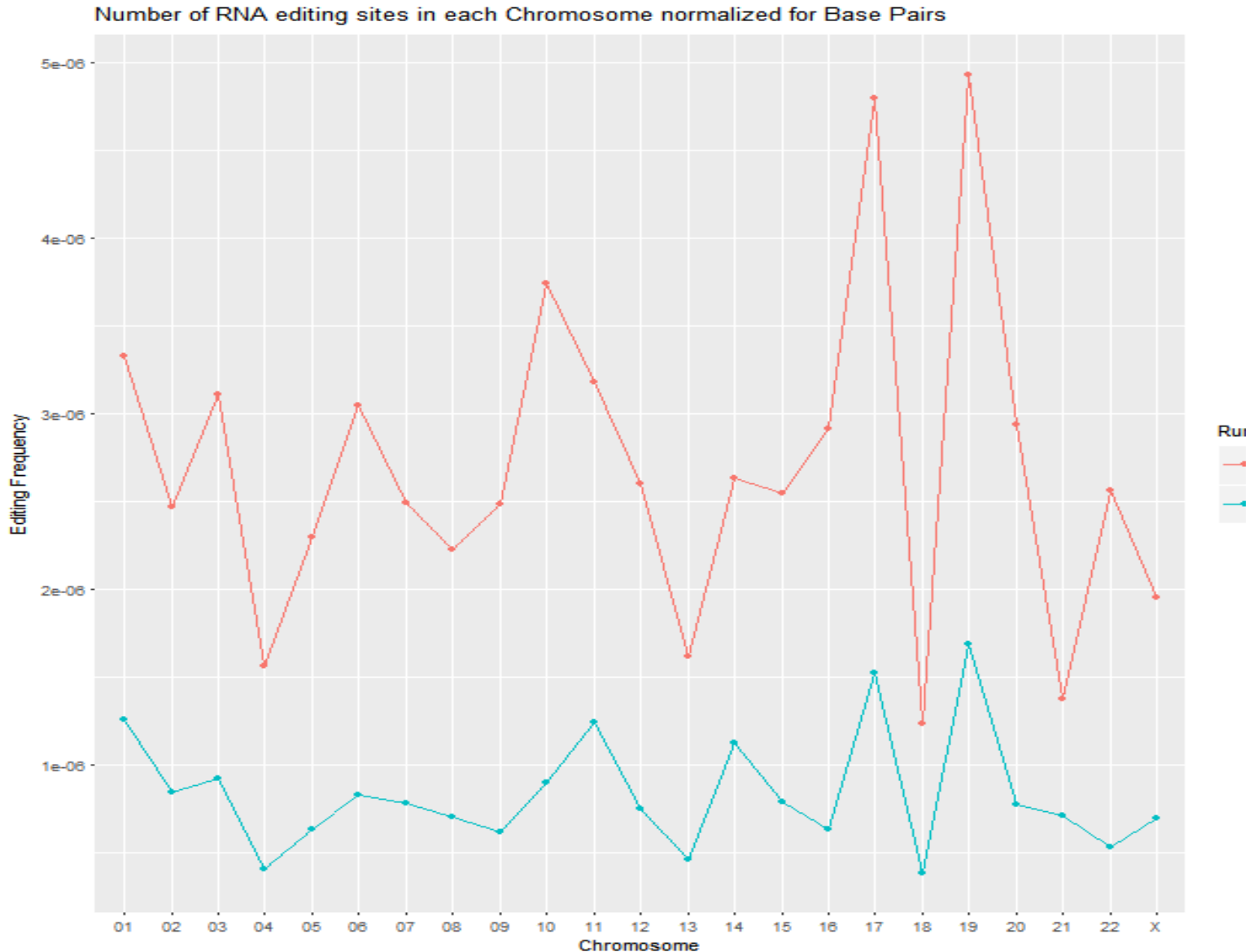


Fig 9: Number of RNA editing sites in each chromosome normalized for base pairs.

The relationship between ADAR expression and the editing frequency can be seen in Figure 10. One of the control runs had a significantly larger ADAR expression and RNA editing frequency than the preeclampsia runs while the other had lower expression and editing frequency. Surprisingly, the fetuses born to preeclamptic mothers actually showed a decrease in RNA editing events with increased ADAR expression while the control runs showed the opposite trend. While this could mean that Preeclampsia in mothers could lead to decreased RNA editing activity in fetuses, this change could also be linked with the sequencing depth of all the runs used in this experiment.

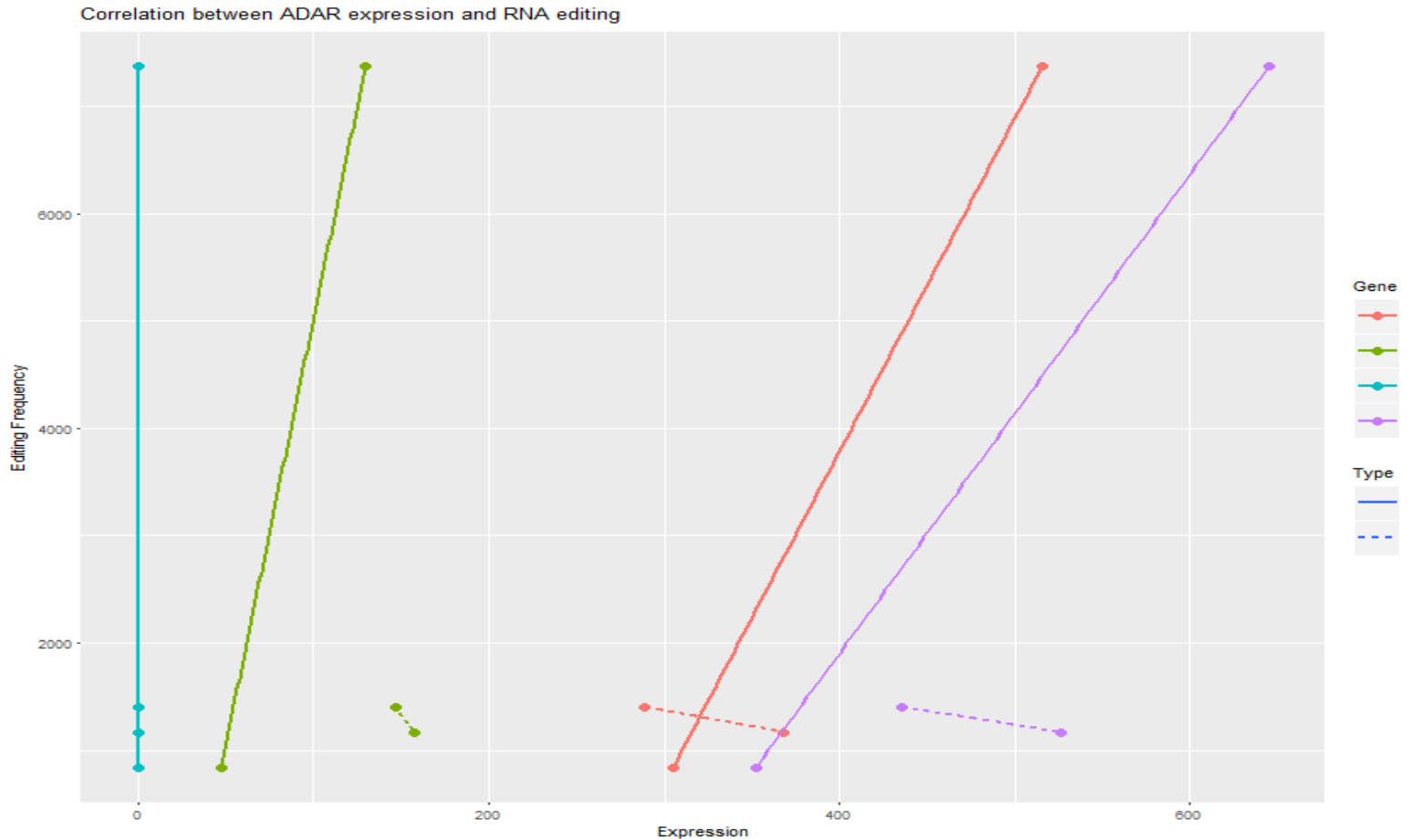


Fig 10: Correlation between ADAR gene expression and number of RNA editing events.

A list of genes were identified in the fetal runs that may have relevance to preeclampsia. These genes all possessed common RNA editing sites across the two hypoxic runs, but not in either control run. They were classified into categories based on how relevant they are to preeclampsia, with Category A being the highest and Category C being the lowest. Table 3 is shown below.

Table 3B: List of genes containing RNA editing sites with possible relevance to preeclampsia in the fetal experiments.

| Gene Symbol | Gene Name | Class |
|-------------|--------------------------------------------------------------------------|-------|
| FLNB | Filamin B, Beta | A |
| CLDN4 | Claudin 4 | A |
| CTSB | Cathepsin B | A |
| RHOC | Ras homolog family member C | B |
| MYOF | Myoferlin | B |
| ATP1B3 | ATPase, Na ⁺ /K ⁺ transporting, beta 3 polypeptide | B |
| APP | Amyloid beta (A4) precursor protein | C |

Exonic Changes

Two editing sites from each category of data drew our interest the most. Two common editing sites in the cell line data caused changes in exonic regions. One of them caused an amino acid change. This being a lysine changing to a glutamic acid in the product of the EIF4E gene. Similarly we found two common editing sites in the fetal data that caused exonic changes. One of these caused an amino acid change from Isoleucine to Valine in the COPA gene. Table 4 shows these and their positions.

Table 4: Common editing sites in the cell line and fetal data respectively that result in exonic changes. Two of these (one each for cell line and fetus) result in amino acid changes.

| Data | Position | Chromosome | Gene | Codon Change | Amino Acid Change |
|-----------|-----------|------------|-------|---------------------|-------------------------------|
| Cell Line | 910526 | 16 | NOMO1 | Thymine to Cytosine | No change (Remains Histidine) |
| Cell Line | 98886471 | 4 | EIF4E | Adenine to Guanine | Lysine to Glutamic Acid |
| Fetus | 160332454 | 1 | COPA | Adenine to Guanine | Isoleucine to Valine |
| Fetus | 53617406 | X | HUWE1 | Adenine to Thymine | No change (Remains Serine) |

Discussion

In this study we, for the first time as per our knowledge, identified potential biomarkers for Preeclampsia in the form of genes carrying specific RNA editing sites. These sites owing to either their functional significance or the functional significance of the gene itself, may increase likelihood of developing or severity of Preeclampsia. Our research focused primarily on the discovery of RNA editing sites in genes relevant to Preeclampsia. By that rationale, we shortlisted the genes we did find in our study in accordance to proximity to pathways implicated and dysregulated in Preeclampsia. Aside from these, we focused more closely on the two genes that showed changes in amino acid sequence of the encoded protein due to these editing sites. One gene from each category of data contained such a change. These EIF4E for cell lines and COPA for fetuses. EIF4E codes for elongation factor 4E. This is involved in translation of numerous proteins including those involved in immune and hypoxia influenced pathways (12). Something with relevance to Preeclampsia owing to functions involving these pathways being dysregulated in the disease (17, 18). COPA codes for Coatmer Protein Complex Subunit Alpha. This protein plays a part in transport from endoplasmic reticulum to the Golgi (12). In the future we would hope to validate the changes in these genes in the wet lab using available samples.

The categories A, B, C list the genes in decreasing order of importance (A most important). The two groups of datasets we had also conveniently corresponded to the broad categories of Preeclampsia, maternal and placental. The fetal data being associated with the former and the cell line with the latter. Although this distinction is an oversimplification, both variants of Preeclampsia maybe connected to our finds in either sets of data. But given the fetuses were unexposed to Preeclampsia and their relevance to the study is through association with their disease bearing mothers, they in that regard qualify as more likely candidates for use in search of markers for maternal Preeclampsia. Similarly the cell lines were placental cell lines exposed to hypoxia to replicate the conditions of the organ during the disease and hence are more likely to display markers for placental Preeclampsia.

First we scrutinize the findings from the cell line data. The category A genes included Gluathione Reductase, Pleckstrin Homology Domain Containing, Map Kinase 13 and Solute Linked Carrier Family 9. The most significant of these by our rationale was Map Kinase 13. It is linked to complement activation pathways(12) which is a target for pathological dysfunction during Preeclampsia(13). Giving weight to this as a possible candidate for RNA editing mediated influence

plasma membrane (12). Fibronectin elevation is another highlighted pathology observed in Preeclampsia. More specifically this elevation leads to impairment of Fibronectin function. The PLEKHA2 protein is involved in recruiting proteins to the plasma membrane. Thus any alteration in translation and consequently function may possibly contribute the process (16).

Category B included five genes. These genes encompassed three key biological processes. ATP5S, NOA1 and NDUFB1 code for an ATP synthase, Nitric Oxide Synthase and NADH Dehydrogenase respectively. All are involved in the respiration chain (12). Hypoxia in placental cells is prevalent in Preeclampsia (17). Finding genes involved in the utilization of oxygen and generation of ATP as harboring unique editing sites gives us reason to speculate the hypoxia like effects of the disease may include impairment in oxygen utilization in addition to oxygen transport. SFXN1 is relevant in the same disease process. Its product is an iron transporter and is involved in erythrocyte differentiation (12). Impairment of RBC development can theoretically contribute to dysfunctional oxygen transport, leading to hypoxia. Another process implicated here and by prior research of Preeclampsia is the immune response. Immune dysfunctions are another key feature of the disease (18). Here we found ENAH and CASP10 harboring editing sites. The former is involved in T cell signaling, the latter in programmed cell death of infected cells (12). CASP10 becomes more interesting in light of the fact that we found several genes that are involved in viral infection processes (full list of genes in supplementary table 1). The third process implicated by our category B genes is lipid metabolism. Here we found editing sites in HSD17B12 whose product is involved in biosynthesis of very long chain fatty acids, as well as conversion of estrone to estradiol in the ovary, cell adhesion and fibronectin binding (12). The ovarian function is the reason we included in category B and not lower. As lipid distribution is dysregulated in Preeclampsia (19), the function of this gene may make it another candidate for RNA edited mediated influence pathological manifestations. Category C included three genes. The first of these CFLAR and its encoded protein is involved in apoptosis pathways in support of caspase and was considered relevant similarly in an immune system context. The second gene SZT2 encodes a protein involved in nervous development and regulation of superoxide dismutase (12). The latter is an important antioxidant enzyme and may theoretically contribute to the oxidative damage in Preeclampsia if dysregulated. The third gene, ATG14 codes a protein involved in cellular response to glucose starvation (12). The reason for its inclusion here being, we believe that since the end result of oxygen starvation and glucose starvation are similar, i.e energy shortage, this may theoretically

gene, APP. Its product is involved in heparin binding (12) and maybe important in context of the blood clotting disorder in Preeclampsia. Category B had 3 genes. ATB1B3, whose product is involved sodium, potassium transport and blood coagulation (important in context of blood disorders), MYOF, whose product is involved in heart muscle contraction and VEGF signalling (both dysregulated in Preeclampsia) and RHOC, whose product is involved in apical junction formation, axonal guidance and NF- κ B signalling (important perhaps for epithelial damage, CNS disorder and immune dysregulation seen in Preeclampsia) (12). Finally there were 3 category A genes that we deemed the most important. These were Filamin B, involved in cytokine signalling and cytoskeletal anchoring (the former being more relevant in a Preeclampsia context), Claudin 4, which plays several key roles such as tight junction formation and regulation of response to Progesterone during pregnancy (the former particularly relevant for the epithelial damage observed in Preeclampsia) and lastly Cathepsin B, which is an enzyme used for degrading pathogenic matter by the immune system and whose dysregulation can contribute to excess inflammation (another pathology observed in Preeclampsia) (12).

Moving away from the focus on genes, this study offered several interesting findings. The genes shortlisted here, mostly harbored editing sites in 3' UTR regions. Whether or not these regions have a higher frequency of RNA editing is something future studies could explore. Furthermore, overall we found editing sites in a large variety of genes with multifarious functions such as cell adhesion, structural support, respiration chain, metabolism, various biosynthetic pathways, etc. Previously RNA editing was believed to be directed to mostly transcription and growth regulation genes. Something which is also consistent with our study given the large of number of similar genes we found, but the additional genes with more diverse functions offers perhaps a promising ground for further investigation. For example, we found editing sites in genes such as Collagen, nucleoporins, F11 receptor, etc (supplementary File 1).

Our major breakthrough from this study was the identification of common RNA editing sites in data from Preeclampsia disease models. The discovery of 667 common sites from cell lines and 23 from fetuses from mothers with Preeclampsia offer a promising avenue for further research. The two genes we identified from the two data pools as containing an amino acid change due to editing also serve as an important find. In particular EIF4E and several of the other mentioned genes owing to their involvement in Preeclampsia effected pathways and functions, may become important biomarkers for diagnosis and prevention of the disease. Subject to further validation of course.

In future efforts, we would look to validate the identified potential markers using PCR and sanger sequencing. At present we have the access to more than 10,000 blood samples from pregnant women with different outcome including preeclampsia, the validation of these markers would provide useful information for intervention and diagnosis.

Conflict of Interest

The authors declare no conflict of interest

References

1. Bundschuh R, Altmüller J, Becker C, Nürnberg P, Gott J. Complete characterization of the edited transcriptome of the mitochondrion of *Physarum polycephalum* using deep sequencing of RNA. *Nucleic Acids Research*. 2011;39(14):6044-6055.
2. Sommer B, Köhler M, Sprengel R, Seeburg P. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell*. 1991;67(1):11-19.
3. Nishikura K. Functions and Regulation of RNA Editing by ADAR Deaminases. *Annual Review of Biochemistry*. 2010;79(1):321-349.
4. Rueter S, Dawson T, Emeson R. Regulation of alternative splicing by RNA editing. *Nature*. 1999;399(6731):75-80.
5. Craici I, Wagner S, Weissgerber T, Grande J, Garovic V. Advances in the pathophysiology of pre-eclampsia and related podocyte injury. *Kidney International*. 2014;86(2):275-285.
6. Report of the National High Blood Pressure Education Program Working Group on High Blood Pressure in Pregnancy. *American Journal of Obstetrics and Gynecology*. 2000;183(1):s1-s22.
7. Raymond D, Peterson E. A Critical Review of Early-Onset and Late-Onset Preeclampsia. *Obstetrical & Gynecologic Survey*. 2011;66(8):497-506.
8. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.
9. DePristo M, Banks E, Poplin R, Garimella K, Maguire J, Hartl C et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*. 2011;43(5):491-498.
10. correction: Initial sequencing and analysis of the human genome. *Nature*. 2001;412(6846):565-566.
11. Leinonen R, Sugawara H, Shumway M. The Sequence Read Archive. *Nucleic Acids Research*. 2010;39(Database):D19-D21.
12. UniProt Consortium T. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*. 2018;46(5):2699-2699.
13. Regal J, Burwick R, Fleming S. The Complement System and Preeclampsia. *Current Hypertension Reports*. 2017;19(11).

17. Tal R. The Role of Hypoxia and Hypoxia-Inducible Factor-1Alpha in Preeclampsia Pathogenesis. *Biology of Reproduction*. 2012;87(6).
18. Bounds K, Newell-Rogers M, Mitchell B. Four Pathways Involving Innate Immunity in the Pathogenesis of Preeclampsia. *Frontiers in Cardiovascular Medicine*. 2015;2.
19. Wetzka B, Winkler K, Kinner M, Friedrich I, März W, Zahradnik H. Altered Lipid Metabolism in Preeclampsia and HELLP Syndrome: Links to Enhanced Platelet Reactivity and Fetal Growth. *Seminars in Thrombosis and Hemostasis*. 1999;25(05):455-462.
20. Suzuki M, Maekawa R, Patterson N, Reynolds D, Calder B, Reznik S et al. Amnion as a surrogate tissue reporter of the effects of maternal preeclampsia on the fetus. *Clinical Epigenetics*. 2016;8(1).
21. Barrett T, Wilhite S, Ledoux P, Evangelista C, Kim I, Tomashevsky M et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research*. 2012;41(D1):D991-D995.
22. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*. 2010;38(16):e164-e164.
23. Benne R, Van Den Burg J, Brakenhoff J, Sloof P, Van Boom J, Tromp M. Major transcript of the frameshifted *coxII* gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell*. 1986;46(6):819-826.
24. Han L, Diao L, Yu S, Xu X, Li J, Zhang R et al. The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. *Cancer Cell*. 2015;28(4):515-528.

Supplementary Material

Supplementary Table 1: List of Genes with unique common editing sites (common between disease models and not found in controls) for Cell line data.

Supplementary Table 2: List of Genes with unique common editing sites (common between disease models and not found in controls) for fetal data.

Supplementary File 1: List of unique editing sites in fetal data for Preeclampsia model (Run 1)

Supplementary File 2: List of unique editing sites in fetal data for Preeclampsia model (Run 2)

Supplementary File 3: List of unique editing sites in fetal data for Control (Run 3)

Supplementary File 4: List of unique editing sites in BeWo cell line, hypoxic model sampled at 24 hours (Run 4)

Supplementary File 5: List of unique editing sites in Jeu cell line, hypoxic model sampled at 24 hours (Run 5)

Supplementary File 6: List of unique editing sites in BeWo cell line, control sampled at 24 hours (Run 6)

Supplementary File 7: List of unique editing sites in Jeu cell line, control sampled at 24 hours (Run 7)

Supplementary File 8: List of unique editing sites in BeWo cell line, hypoxic model sampled at 16 hours (Run 8)

Supplementary File 9: List of unique editing sites in Jeu cell line, hypoxic model sampled at 16 hours (Run 9)

Supplementary File 10: List of unique editing sites in BeWo cell line, control sampled at 16 hours (Run 10)

Supplementary File 11: List of unique editing sites in Jeu cell line, control sampled at 16 hours (Run 11)

Supplementary File 12: List of unique editing sites in fetal data for Control (Run 12)