# A Multi-Omic Analysis of Pregnancy-Associated Variances in Endometrial Carcinoma

Erika Li, Tomas Manea, Nathan Yoon

## Introduction

Endometrial carcinoma has been increasing both in mortality and morbidity over the past 10 years disproportionately compared to other cancers (International Journal of Gynecological Cancer). It is the most common gynecological cancer in high-income countries, and its incidence rate is rising globally: up 132% in the last 30 years (Crosbie et. al, 2023). Risk for endometrial carcinoma increases with age, especially after menopause. In addition, obesity, diabetes, age, early menarche, late menopause, and tamoxifen use (a breast cancer drug) are also all risk factors for developing the disease (Ali, 2013). Importantly, it has been found that nulliparity (having had no live births) is a risk factor for endometrial cancer. In 2015, Wu et. al found that there is a significant inverse association between parity (number of live births) and risk of endometrial cancer. Subsequent studies have characterized a decreasing risk of endometrial carcinoma for multiple births; ze pop et. al (2022) found that 3 live births compared to none attributed to 50% decrease in endometrial cancer risk. In addition, it was found that multiparous patients had higher survival probability compared to nulliparous patients in Hachisuga et. al (2002), and that cumulative survival rates for late-stage endometrial carcinoma patients were highest in patients with parity of 3 or more. Using this baseline, our analysis aimed to explore pregnancy-related differences across the molecular landscape of endometrial cancer, specifically related to parity. To this end, our study designed two groups, based on the significance of 3 live births found in D'Urso et. al and Hachisuga et. al: a "low-medium" pregnancy group of endometrial carcinoma

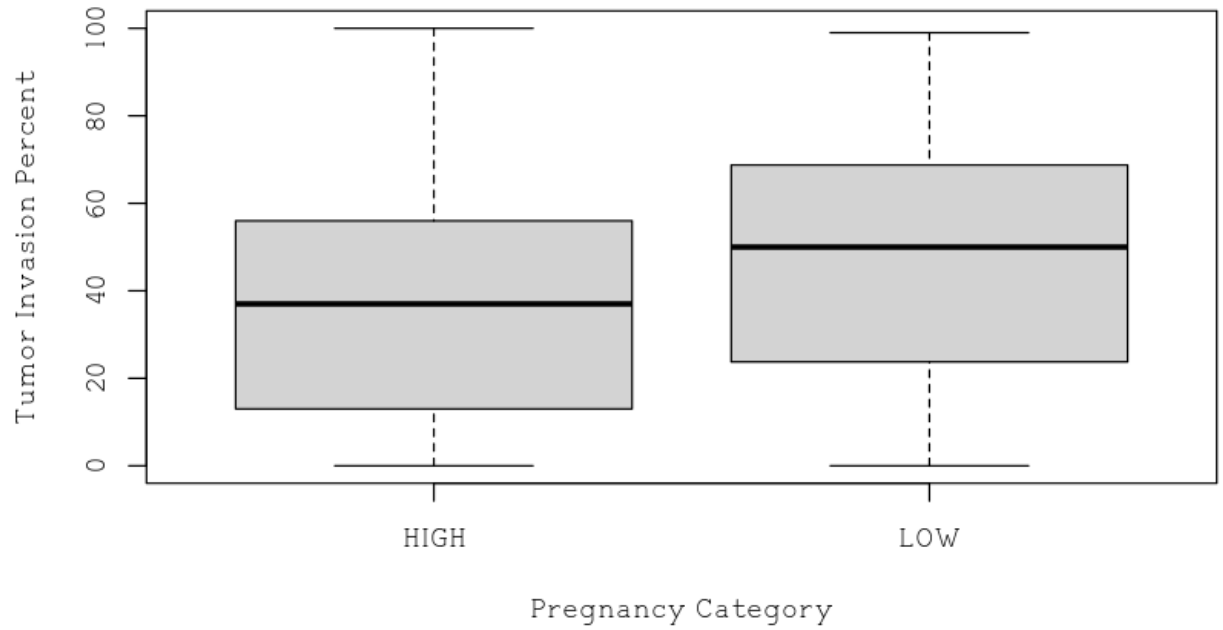
patients with fewer than three pregnancies, and a “high” pregnancy group with at least three pregnancies. This analysis utilized multi-omic analysis, combining and analyzing diverse types of biological information, such as genomics, transcriptomics, and proteomics, to gain a comprehensive understanding of the biological basis of endometrial carcinoma and differences between the two pregnancy groups. In this analysis, we aimed to examine how molecular alterations across various -omic levels contribute to the difference in risk and survival based on parity for endometrial carcinoma.

**Methods**

In order to conduct this analysis, we utilized both The Cancer Genome Atlas (TCGA) and the Clinical Proteomic Tumor Analysis Consortium (CPTAC), free, open-access databases containing genomic, transcriptomic, and proteomic data across dozens of cancer types. For the TCGA analysis, we utilized RStudio in order to conduct genomic and transcriptomic analysis. Endometrial cancer clinical and Mutation Annotation Format (MAF) data were accessed from TCGA using the accession code “TCGA - UCEC”. Initial Kaplan-Meier survival plots were generated using various clinical factors (race, age, BMI) as part of initial exploratory analysis. Subsequently, the clinical data was divided into the low-medium pregnancy ($< 3$) and the high pregnancy ($\geq 3$) groups. A Welch two sample t-test was performed on tumor invasion percentage data between these two groups. The R Bioconductor package maftools was used to create oncoplots/cooncoplots, somaticMutations plots, and mutation type data tables. Differential gene expression was conducted using the R Bioconductor package DESeq2. Finally, these genes were inputted into the G:Profiler web server for functional enrichment analysis to yield significant biological pathways and functions associated with the differentially expressed genes. For the

second part of the analysis, CPTAC data was analyzed within Python in order to conduct proteomic analysis. Proteins with the absolute mean greatest differential expression between low-medium pregnancy and high pregnancy patient groups were identified, along with the most highly expressed proteins in both groups.

**Results**




```
        Welch Two Sample t-test

data:  low_preg_pt$tumor_invasion_percent and high_preg_pt$tumor_invasion_percent
t = 1.3451, df = 35.48, p-value = 0.1871
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.446363 31.796522
sample estimates:
mean of x mean of y
 49.64286  36.96778
```

**Figure 1:** Boxplot and two sample t-tests of tumor invasion percent between high and low-medium pregnancy groups.

Tumor invasion percent data was split between high pregnancy versus low-medium pregnancy patients. A Welch two sample t-test was conducted, revealing that the high pregnancy group had a mean tumor invasion percentage of 36.97% and the low-medium group had a mean tumor invasion percentage of 49.64%, with a p-value of 0.1871.

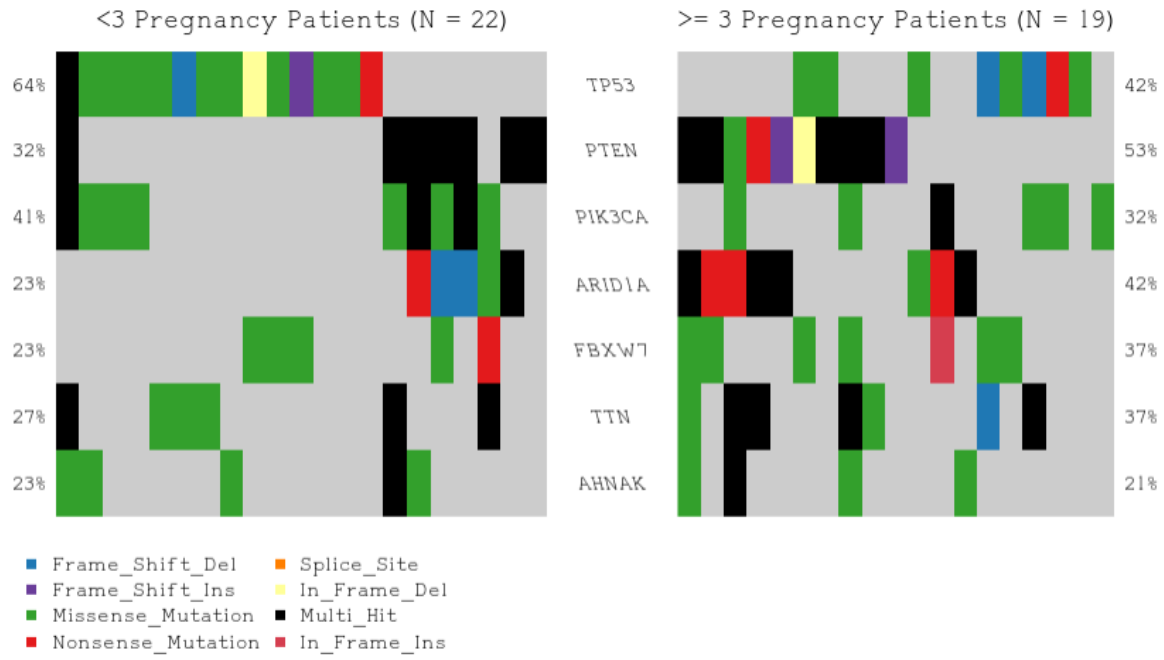



**Figure 2:** Co-oncoplot of low-medium pregnancy group and high pregnancy group.

The genes TP53, PIK3CA, and AHNAK were more highly mutated in the low-medium pregnancy group. PTEN, ARID1A, FBXW7, and TTN were more highly mutated in the high pregnancy group. TP53 and PTEN were observed to have a mutually-exclusive relationship - patients with one mutation typically did not have the other.

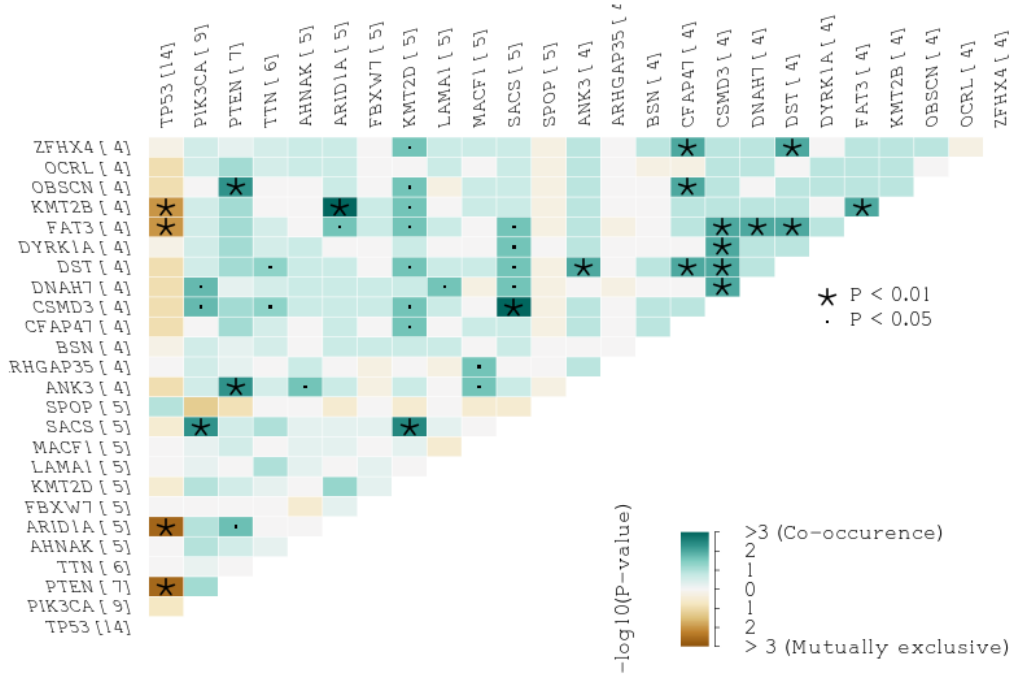



**Figure 3:** Somatic interactions plot for the low-medium pregnancy group.

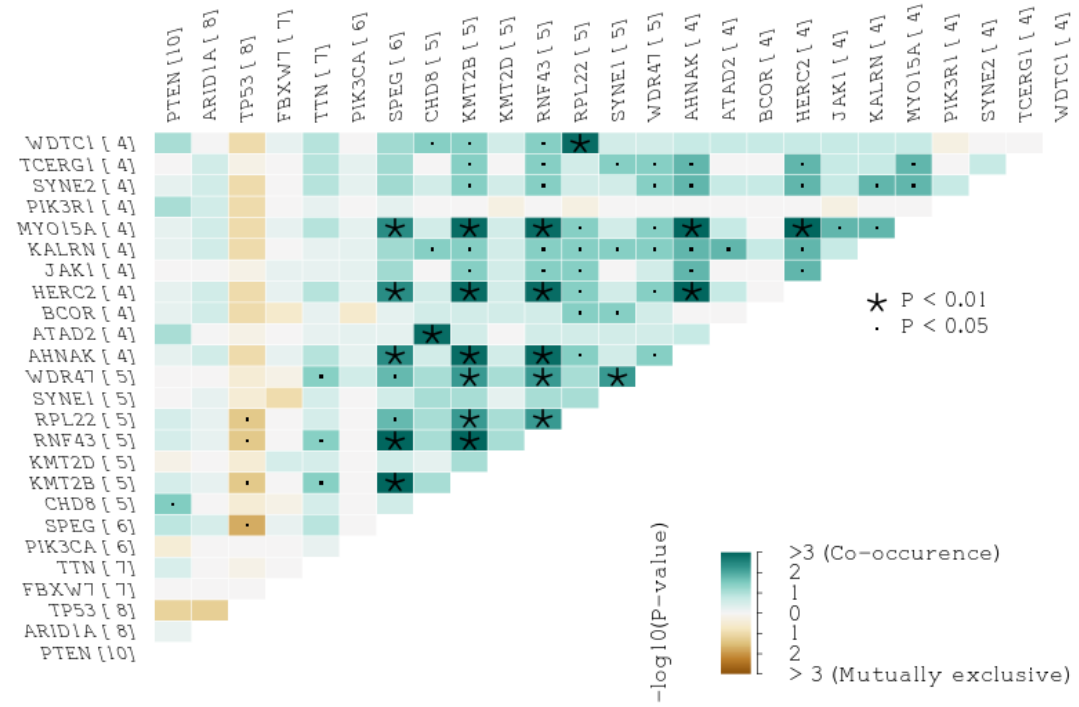



**Figure 4:** Somatic interactions plot for the high pregnancy group.

For the low-medium pregnancy group, there was a mutually exclusive relationship between the presence of mutations of TP53 and several other genes, most notably KMT2B, FAT3, ARID1A, and PTEN. This trend with TP53 was not as significant in the high pregnancy group. However, the high pregnancy group seemed to have more significant co-occurrence of mutations with SPEG, KMT2B, RNF43 with MYO15A, HERC2, and AHNAK.

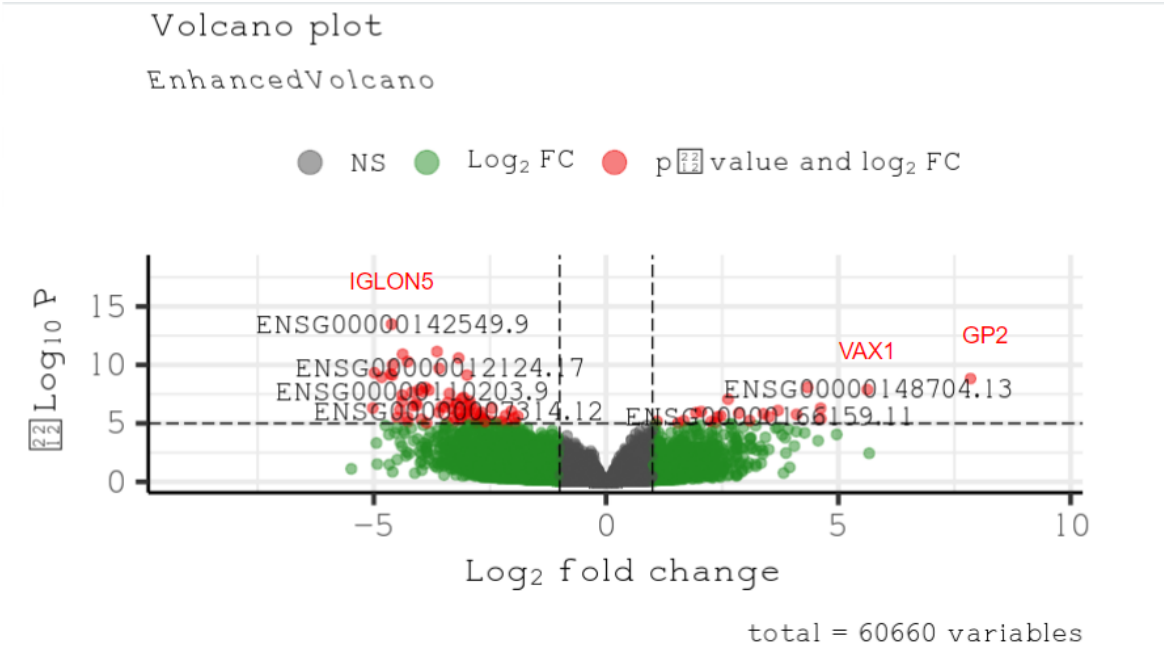



**Figure 5 :** Volcano plot showing most differentially expressed proteins in the low-medium pregnancy group in comparison to the high-pregnancy group.

The volcano plot shows that IGLON5 was significantly downregulated in the low-medium pregnancy group, while proteins such as GP2 and VAX1 were significantly upregulated. The most differentially expressed proteins did not include TP53 or PTEN (not significant), suggesting that the most differentially mutated genes are different from the most differentially expressed proteins between the low- and high-pregnancy groups.

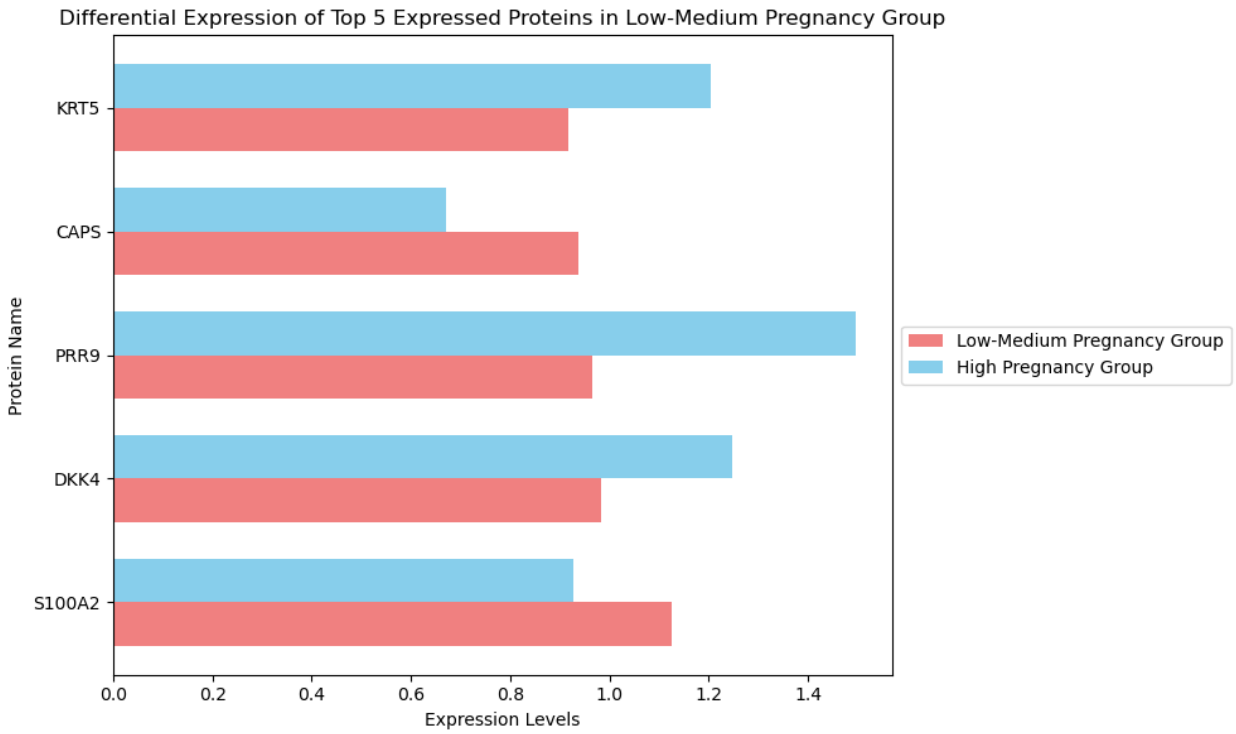



**Figure 6 :** Differences in protein expression for the top five expressed proteins in the low-medium pregnancy group.

KRT5, PRR9, and DKK4 proteins were more highly expressed in the high pregnancy group, despite these proteins being the most highly expressed five proteins in the low-medium group. However, CAPS and S100A2 were both more highly expressed in the low-medium pregnancy group.

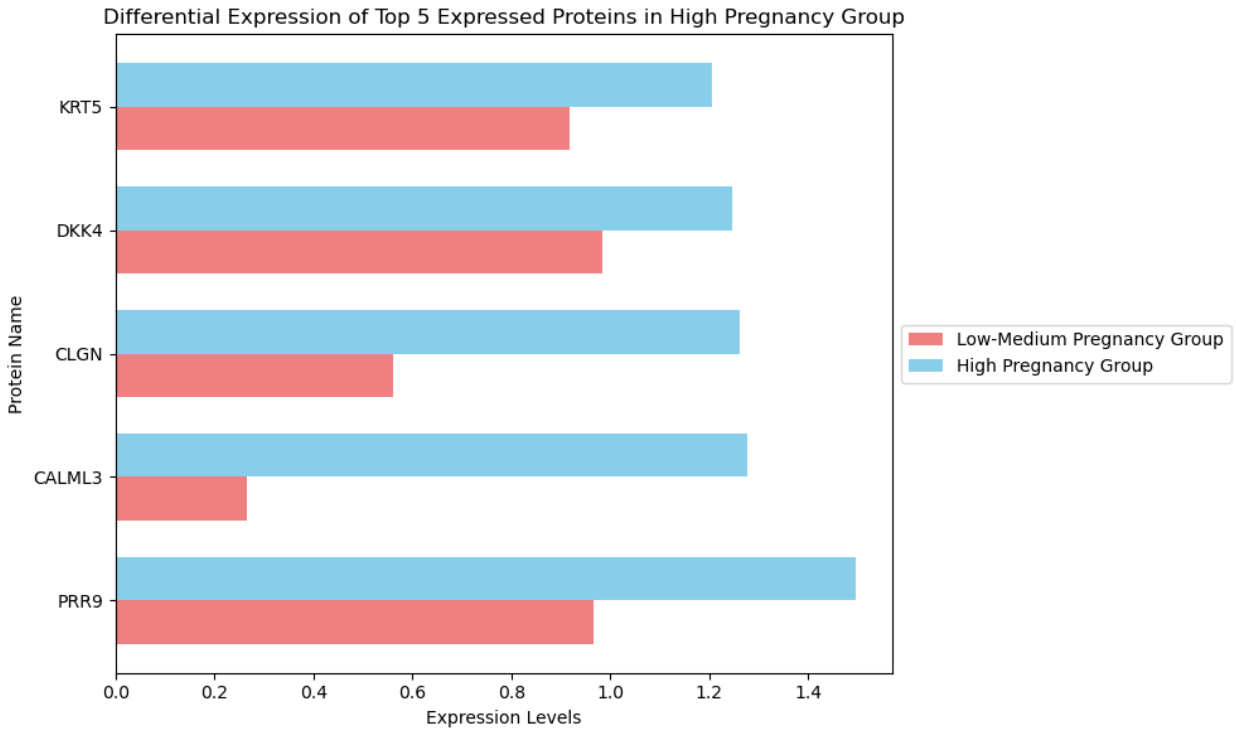



**Figure 7 :** Differences in protein expression for the top five expressed proteins in the high pregnancy group.

All of the top five expressed proteins for the high pregnancy group were expressed at higher levels than those of the low-medium pregnancy group. The CALML3 protein in particular had very low expression in the low-medium group compared to the high pregnancy group.

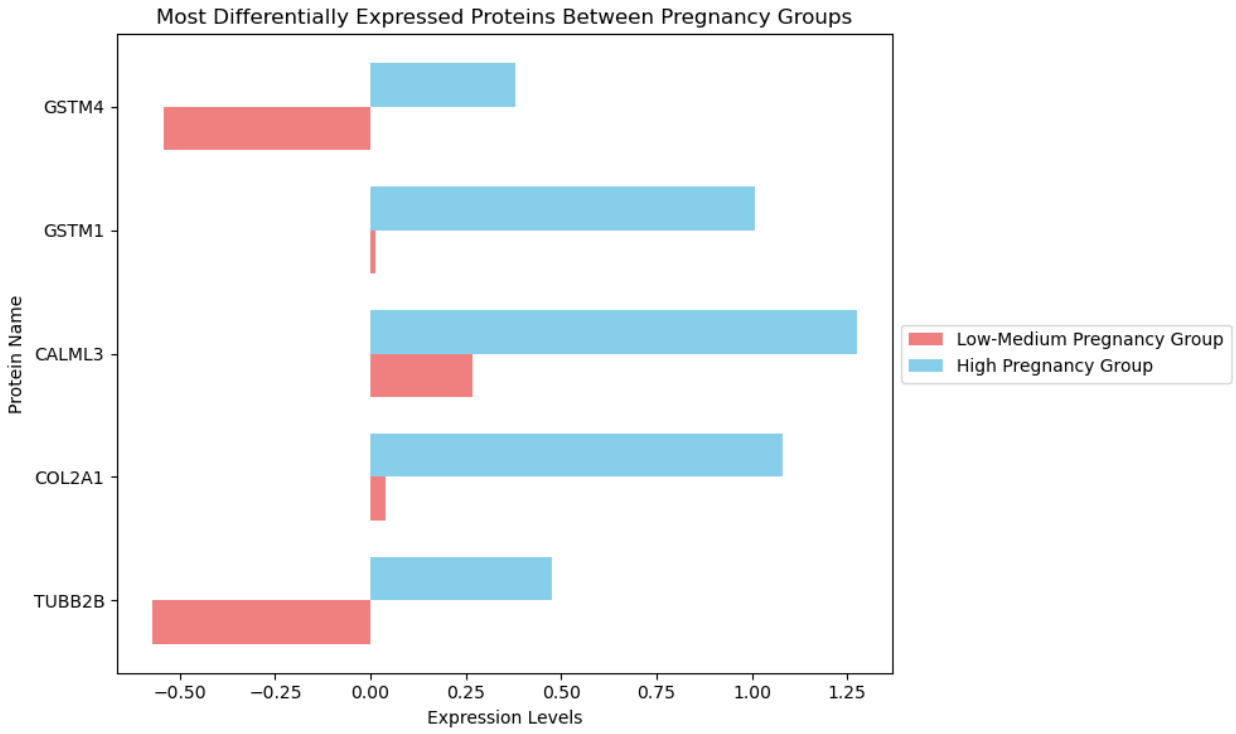



**Figure 8:** Top five differentially expressed proteins between the low-medium pregnancy group and the high pregnancy group.

The GSTM4 and TUBB2B proteins were downregulated in the low-medium pregnancy group but upregulated in the high pregnancy group. All the most differentially expressed genes were more highly expressed in the high pregnancy group than in the low-medium group.

**Discussion**

Pregnancies, in existing literature, are believed to have a protective effect on endometrial cancer. Previous studies have suggested there exists a strong correlation with parity and a decreased risk of endometrial cancer (Raglan et al., 2018; D'Urso et al., 2022). Proposed by Jordan et al., a possible hypothesis is that the increase in progesterone during pregnancy exerts a protective effect on the risk of endometrial cancer, however, there remain several variables that need further exploration. Within our initial analysis within TCGA (Figure 1), we observed that—on average—our high pregnancy group exhibited less tumor invasion percent relative to our low-medium pregnancy group, echoing the existing literature that the amount of pregnancies serve as a protective effect against endometrial cancer (Chen et al., 2015; Murali et al., 2014; Main et al., 2022).

Furthermore, in Figure 3 and 4, we can observe in the high pregnancy group that there are several significant co-occurrences for genes, SPEG and RNF43, respectively. SPEG was noted to co-occur with MYO15A, HERC2, AHNAK, RNF43, and KMT2B, while RNF43 co-occurred with MYO15A, HERC2, AHNAK, WDR47, and RPL22. Specifically, MYO15A encodes a myosin protein, akin to how SPEG encodes for a striated muscle enrichment protein kinase, which is involved in myocyte protein development (NIH et., 2023). Furthermore, our findings corroborate with Suda et al.'s study, where it has been noted that SPEG is a frequently mutated gene in endometrial epithelium cells. On the other hand, we observe that RPL22 has been noted to be frequently mutated in endometrioid endometrial cancers: which is a gene that is significantly co-occurrence with RNF43, a gene has been noted to be a common gene mutation in in the mucinous tumours of the ovary and within endometrial cancer (Novetsky et al., 2012;

Giannakis et al., 2014). With the myriad of co-occurrence genes, we begin to draw together a narrative demonstrating how genes involved in myosin development may be involved with localized proteins in the endometrial region—in patients with endometrial cancer.

Within our proteomic analysis, we found that CALML3 was detected as one of the most top expressed proteins in the high pregnancy group. CALML3 encodes for proteins that enable calcium ion binding activity and enzyme regulator activity. Furthermore, we see that CALML3 has been noted to serve as a biomarker for the presence of endometrial cancer (Nojuku et al., 2023). Furthermore, CALML3 has been noted to be upregulated by the presence of estrogen—with estrogen signaling being hypothesized as potentially being responsible for mutations in *PTEN*, *ARID1A*, *PIK3CA*, *PIK3R1* (Rodriguez et al., 2019).

Our study design contains a number of inherent limitations that arise when analyzing aggregate, large public datasets. For one, there are discrepancies between the classification between TCGA's "number of full term pregnancies" versus CPTAC's "number of pregnancies," differences that may omit pregnancies that were not full term pregnancies. This nuance presents itself as a confounding variable as we conducted our analysis. Parity, as defined in the literature, refers to the number of *live births*, not just pregnancies. We don't know due to limited information and de-identification of these datasets whether all these pregnancies resulted in live births. However, literature also shows protective effects of incomplete pregnancies and stillbirths as well, which is why we still centered our analysis on pregnancy—despite incomplete information about the status of patients' pregnancies.

For future studies, researchers should look to expand on nuances present in the measurement of parity: factoring in variables of race, Type 1 versus Type 2, age at pregnancy, and types of pregnancies—so as to better understand the factors that comprise parity. Furthermore, we should refine our results through pinpointing specific genes that were noted above and conduct gene specific studies, allowing us to target key biochemical pathways that have been noted to be involved in progesterone, estrogen, and keratin production.

## References

Ali A. T. (2013). Risk factors for endometrial cancer. Ceska gynekologie, 78(5), 448–459.

Chavez, T. F., Snyder, R., Lee, R. K., et al. (2023). Rising endometrial cancer rates and potential for screening. International Journal of Gynecologic Cancer. https://doi.org/10.1136/ijgc-2023-004701.

Chen, Q., Tong, M., Guo, F., Lau, S., & Zhao, M. (2015). Parity Correlates with the Timing of Developing Endometrial Cancer, But Not Subtype of Endometrial Cancer. Journal of Cancer, 6(11), 1087–1092. https://doi.org/10.7150/jca.12736

Corpus uteri - International Agency for Research on Cancer. The Global Cancer Observatory. (2020). https://gco.iarc.fr/today/data/factsheets/cancers/24-Corpus-uteri-fact-sheet.pdf

Crosbie, E. J., Kitson, S. J., McAlpine, J. N., Mukhopadhyay, A., Powell, M. E., & Singh, N. (2022). Endometrial cancer. Lancet (London, England), 399(10333), 1412–1428. https://doi.org/10.1016/S0140-6736(22)00323-3

D'Urso, S., Arumugam, P., Weider, T., et al. (2022). Mendelian randomization analysis of factors related to ovulation and reproductive function and endometrial cancer risk. BMC Med, 20, 419. https://doi.org/10.1186/s12916-022-02585-w

Giannakis, M., Hodis, E., Jasmine Mu, X., Yamauchi, M., Rosenbluh, J., Cibulskis, K., Saksena, G., Lawrence, M. S., Qian, Z. R., Nishihara, R., Van Allen, E. M., Hahn, W. C., Gabriel, S. B., Lander, E. S., Getz, G., Ogino, S., Fuchs, C. S., & Garraway, L. A. (2014). RNF43

is frequently mutated in colorectal and endometrial cancers. Nature genetics, 46(12), 1264–1266. https://doi.org/10.1038/ng.3127

Hachisuga, T., Fukuda, K., Hirakawa, T., & Kawarabayashi, T. (2001). The Effect of Nulliparity on Survival in Endometrial Cancer at Different Ages. Gynecologic Oncology, 82(1), 122-126. https://doi.org/10.1006/gyno.2001.6220

Jordan, S. J., Na, R., Weiderpass, E., Adami, H. O., et al. (2021). Pregnancy outcomes and risk of endometrial cancer: A pooled analysis of individual participant data in the Epidemiology of Endometrial Cancer Consortium. International Journal of Cancer, 148(9), 2068–2078. https://doi.org/10.1002/ijc.33360

Main, C., Chen, X., Zhao, M., Chamley, L. W., Chen, Q. (2022). Understanding How Pregnancy Protects Against Ovarian and Endometrial Cancer Development: Fetal Antigens May Be Involved. Endocrinology, 163(11), bqac141. https://doi.org/10.1210/endocr/bqac141

Murali, R., Soslow, R. A., & Weigelt, B. (2014). Classification of endometrial carcinoma: more than two types. The Lancet. Oncology, 15(7), e268–e278. https://doi.org/10.1016/S1470-2045(13)70591-6

Novetsky, A. P., Zighelboim, I., Thompson, D. M., Jr, Powell, M. A., Mutch, D. G., & Goodfellow, P. J. (2013). Frequent mutations in the RPL22 gene and its clinical and functional implications. Gynecologic oncology, 128(3), 470–474. https://doi.org/10.1016/j.ygyno.2012.10.026

Salvesen, H. B., Akslen, L. A., Albrektsen, G., & Iversen, O. E. (1998). Poorer survival of nulliparous women with endometrial carcinoma. Cancer, 82(7), 1328-1333.

https://doi.org/10.1002/(SICI)1097-0142(19980401)82:7<1328::AID-CNCR16>3.0.CO;2-8

Suda, K., Nakaoka, H., Yoshihara, K., Ishiguro, T., Tamura, R., Mori, Y., Yamawaki, K., Adachi, S., Takahashi, T., Kase, H., Tanaka, K., Yamamoto, T., Motoyama, T., Inoue, I., & Enomoto, T. (2018). Clonal Expansion and Diversification of Cancer-Associated Mutations in Endometriosis and Normal Endometrium. Cell reports, 24(7), 1777–1789. https://doi.org/10.1016/j.celrep.2018.07.037

Raglan, O., Kalliala, I., Markozannes, G., Cividini, S., Gunter, M. J., Nautiyal, J., Gabra, H., Paraskevaidis, E., Martin-Hirsch, P., Tsilidis, K. K., & Kyrgiou, M. (2019). Risk factors for endometrial cancer: An umbrella review of the literature. International Journal of Cancer, 145, 1719-1730. https://doi.org/10.1002/ijc.31961

Survival Rates for Endometrial Cancer. American Cancer Society. https://www.cancer.org/cancer/types/endometrial-cancer/detection-diagnosis-staging/survival-rates.html

Wu, Q. J., Li, Y. Y., Tu, C., et al. (2015). Parity and endometrial cancer risk: a meta-analysis of epidemiological studies. Sci Rep, 5, 14243. https://doi.org/10.1038/srep14243