UCLA
UCLA Electronic Theses and Dissertations

Title

A Multilayered and Clinically-Informed Integration of the Transcriptome, Phenome, and Radiome in Multifactorial Disorder Assessment

Permalink

https://escholarship.org/uc/item/4bh5416t

Author

Katrib, Amal

Publication Date

2018

Supplemental Material

https://escholarship.org/uc/item/4bh5416t#supplemental

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

A Multilayered and Clinically-Informed Integration of the Transcriptome,

Phenome, and Radiome in Multifactorial Disorder Assessment

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biomedical Physics

by

Amal Katrib

2018

ABSTRACT OF THE DISSERTATION

A Multilayered and Clinically-Informed Integration of the Transcriptome,

Phenome, and Radiome in Multifactorial Disorder Assessment

by

Amal Katrib

Doctor of Philosophy in Biomedical Physics

University of California, Los Angeles, 2018

Professor Yi Xing, Chair

Researchers continue to struggle in deciphering the underlying molecular machinery of complex, multifactorial, and comorbid medical disorders. Integrating multiple layers of data –from genomic to exposomic – and evaluating their combinatorial effect on the phenome can mitigate limitations of simple differential analyses and ultimately help uncover causal factors.

In my dissertation work, I specifically focus on the integration of transcriptomic data with other data types that have a high clinical translatability such as phenomic and radiomic characteristics. I apply a multi-layered transcriptome-phenome-radiome integrative framework to two use case scenarios to demonstrate its benefits and drawbacks.

For use case scenario 1, I perform a multi-level analysis of RNA sequencing collected from in-house human placental decidual samples of various modes of parturition in late-stage pregnancy. I highlight differences in gene expression, co-expression, and alternative splicing and

identify tissue- and labor-specific enrichment. I then incorporate dense prognostic and maternal and fetal phenomic information to derive genes and biological processes associated with premature and ceased labor. I demonstrate how an integrative framework successfully allows us to extract biologically relevant information that would have otherwise been missed through hypothesis-driven or monolayer differential analysis. For use case scenario. 2, I generate isoform-level information from RNA sequencing collected from The Cancer Genome Atlas (TCGA) GBM tumors. Using additional layers of the transcriptome, I filter for tumor-enriched genes to subtract microenvironment effects. I then incorporate 2 forms of quantitative morphologic radiomic features to extract exon inclusion-radiophenotype correlates. Through functional annotation, I highlight the underlying biological differences between tumor phenotypes. I demonstrate how an integrative framework provides exploratory insights into the biology of a GBM tumor yet fails to reveal significant associations due to data quality and analytical limitations.

The potential applications of a multi-layered and clinically-informed integration of the transcriptome, phenome, and radiome extend far beyond the immediate rejoice of joining systems biology efforts in the integration of "big data". Through a synergistic coupling of functional molecular indexes, phenotypic characterization, and dense prognostic traits, it enables an in-depth and comprehensive investigation of multifactorial disorders. In the process, it uses a converged data- and hypothesis-mediated approach to balance the benefit of a comprehensive analysis approach and an elaborate mechanistic depiction of etiology. By incorporating individual-level information (from phenomic and radiomic traits) into population-level findings (from transcriptomic analyses), it poses as a promising contributor to the personalized and precision medicine initiatives of modern medicine.

The dissertation of Amal Katrib is approved.

Michael McNitt-Gray

William Hsu

Alex Anh-Tuan Bui

Yi Xing, Committee Chair

University of California, Los Angeles

2018

To Leen, Mama, Baba, and the rest of my family

for their unconditional and unwavering love and support

and to Georges St. Laurent

for believing in me and introducing me to the world of Systems Biology

TABLE OF CONTENTS

`

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my family for loving me unconditionally and for learning to accept me for the person I am despite the unconventional path I have chosen to take in life. Mama and Baba, I am forever grateful for the sacrifices you have made for us, for moving to the United States to give us the chance at a better life –one that can provide us with opportunities we could not otherwise attain and enable the pursuit of our wildest dreams. I now realize the pain you had to endure when you sent me, on my own and at the age of sixteen, to college in Washington, DC; when you later packed everything and moved away from loved ones to stay close to us; when you started again from scratch, unable to work in your professional fields yet still persistently providing for us through less ideal career options; and when you had to deal with my inner turmoil throughout, always reaching out to guide me back. I would not have made it to this stage in my life, facing my fears and getting up every time I got defeated, without you. I promise you to always give it my best and to never give up on things I genuinely care about. Leen, despite our vastly different personalities and lifestyles, we always managed to be there for one another. You are a major source of inspiration to many people and certainly to me. But most importantly, you are my one true love. You were always by my side, even when you disagreed with my actions. You also gave me the stability I desperately needed and the motivation to go above and beyond. I am beyond lucky to have you as my baby sister. Khalto Hala, Huda, Maha, and Rima and Khalo Abed, I cannot even begin to explain how fortunate I feel to have you as my close and loving family. Leaving Dubai was difficult, because it meant moving away from you –a major source of my happiness. I carry your love and the purity of your souls everywhere I go.

Graduate school was a period of self-reflection and growth, both on a personal and professional level. I would like to thank my closest friends for being a big part of this journey. Stephanie Delgado, I am so glad to have met you during my first year at UCLA. I was going

through a rough patch and needed your strong and genuine friendship to make it through. I am excited to celebrate your wedding to Sina Kalbasi in a few months! Alice Lai, even though we did not get the chance to meet often because of my fear of the 405, I am forever grateful for our friendship that has lasted since the first day of undergrad. I am happy to know I have you as a friend for life! And Howie Stern, thank you for always showing me the beautiful side of humanity. You and the bears will always have a special place in my heart.

Academically, I would like to first and foremost thank my advisor, Yi Xing, for showing interest in my research proposal and offering me a chance at his lab even though I was already in my third year. Lan Lin, I am grateful for all the support you have provided throughout the course of my PhD. You helped me tremendously with the pregnancy project and made me realize how passionate I was about it. Emad Samani, thank you for always cheering me up whenever I needed it. To my doctoral committee, Michael McNitt-Gray, William Hsu, and Alex Bui, I greatly appreciate your guidance and continued support of this dissertation. David Nagel, I am fortunate to have had you as a Master's advisor and a friend. Thank you for showing me how exciting research can be and for helping me get on the academic path. And of course, Alex Moser, thank you for also being a wonderful research mentor and friend. I am fortunate to have someone as caring as you in my life.

VITA

EDUCATION

'18   **University of California, Los Angeles, Ph.D. Candidate**, Biomedical Physics

'11   **George Washington University, M.Sc.,** Electrical Engineering (Signal Processing)

'09   **George Washington University, B.Sc.,** Biomedical Engineering (Biophysics minor)

'17   **NeuroBridges,** Computational Neuroscience

'17   **Systems Genetics of Neurodegeneration,** Systems Medicine

'16   **University of Washington,** Statistics for Big Data

'15   **Institute for Systems Biology,** Systems Biology of Disease

ACADEMIC EMPLOYMENT

'15- '18   **University of California, Los Angeles, Dept. of Microbiology, Immunology, and Molecular Genetics,** Graduate Student Researcher, Dr. Yi Xing

'15- '16   **Johns Hopkins University Center for Talented Youth,** Cogito Research Mentor; "Principles of Engineering Design" Instructor

'14- '15   **American Film Institute, Alfred P Sloan Foundation,** Science Advisor

'15   **University of California, Los Angeles, Dept. of Chemistry and Biochemistry,** Graduate Teaching Assistant, Dr. Eric Scerri

'14- '15   **University of California, Los Angeles, Dept. of Neurosurgery,** Graduate Student Researcher, Dr. Itzhak Fried

'13- '14   **University of California, Los Angeles, Crump Institute for Molecular Imaging,** Graduate Student Researcher, Dr. Thomas Graeber

'09- '10   **US Naval Research Laboratory, Dept. of Material Science and Technology,** Research Assistant, Dr. David Nagel and Dr. Alex Moser

'07- '08   **Johannes Guteberg Universitat / National Laboratory for Synchrotron Radiation, A2 "Real Photons" Collaboration,** Engineering Intern, Dr. William Briscoe

PUBLICATIONS

'18   ***Katrib A.**, Kim, J., Lin L., Chong, A., Murray, J., Xing Y. A Multi-Level Transcriptomic Analysis of Human Decidua in Abnormal Parturition. Awaiting Submission.

'18   ***Katrib A.**, Jeong H., Fransen, N., Henzel, K., Miller, J. An Inflammatory Landscape for Glioblastoma-Associated Neurological Deficit. *Frontiers in Genetics.* Under Review.

'17   Nagel, D., **Katrib A.**, Feltrin V. LENR, Energy, and Water. *Infinite Energy*, 134:43-50.

'16   **Katrib A.,** Xing Y., Bui A., Hsu W. "Radiotranscriptomics" - A Synergy of Transcriptomics and Imaging in Clinical Assessment. *Quantitative Biology,* 4(1): 1-12. *(Cover Page)*

'11   **Katrib, A.** Analysis of the Factors Contributing to the Heat Observed in Electrochemical Cells Used in Condensed Matter Nuclear Science (CMNS) (Vol. 49-03). ProQuest Dissertations and Theses.

'09   **Katrib, A.,** Nagel, D. Can Water Be the Origin of Excess Energy. Proceedings of 15th International Conference on Condensed Matter Nuclear Science, The "Anomalous" Energy Production - Section 1: Electrochemical Experiments, 65-66

TALKS AND PRESENTATIONS

'18   **Speaker**, "Human Placental Transcriptome During Parturition", DOHaD 6th International Symposium on Metabolic Programming and Microbiome, Cancun, Mexico.

'18   **Speaker**, "Deciphering the Pregnancy Paradox", Physics and Biology in Medicine 19th Annual Research Colloquium, UCLA David Geffen School of Medicine, Los Angeles, CA.

'17   **Speaker**, "Telemedicine: SAHA Platform", The Syrian American Medical Society (SAMS) 6th National Conference, Huntington Beach, CA.

'17   **Katrib A.**, Smedley N., Xing Y., Hsu W., El-Saden S., Radiotranscriptomic Analysis of Glioblastoma. *2nd Penn Symposium on Mathematical and Computational Biology*. **Poster Presentation.**

'16   **Speaker**, "Radiotranscriptomic" Analysis of Glioblastoma Multiforme (GBM) – Preliminary Findings", The UCLA Institute for Quantitative and Computational Biosciences 2nd Annual Retreat, Skirball Cultural Center, Los Angeles, CA.

'16   **Katrib A.,** Xing Y., Transcriptomic Analysis of Human Placenta. *17th Annual UCLA Physics and Biology in Medicine Research Colloquium.* **Poster Presentation.**

'16   **Speaker**, "Keys to Success and Survival in Graduate School", Southern California Forum for Diversity in Graduate Education, Loyola Marymount University, Los Angeles, USA.

'09   **Katrib, A.,** Nagel, D. Cold Fusion – Potential Green Nuclear Energy. *13th Annual Green Chemistry and Engineering Conference.* College Park, MD. **Poster Presentation.**

'08   **Katrib, A.**, Briscoe, W., Hulsing, T. Developing an Efficient Read-Out System for the A2 Pair Spectrometer at Mainz. *2008 Annual Meeting of the Division of Nuclear Physics.* Oakland, CA. **Poster Presentation.**

'07   **Katrib, A.**, Briscoe, W. Real Photon Physics at MAMI. *2007 Annual Meeting of the Division of Nuclear Physics.* Newport News, VA. **Poster Presentation.**

SELECT ACADEMIC AWARDS

• Eugene V. Cota-Robles fellowship, UCLA • "Neurobridges" scholarship, French National Center for Scientific Research • "Systems Medicine" scholarship, German Ministry of Education & Research • Big Data to Knowledge (BD2K) Biomedical Big Data training grant, NIH-NCI • Summer Institute in Statistics for Big Data scholarship • Neuroimaging Training Program grant, UCLA/NIH (declined) • Competitive Edge fellowship, UCLA • Naval Research Enterprise Intern Program (NREIP) fellowship, Office of Naval Research • Full-Tuition Undergraduate Presidential Academic Scholarship, GWU • International Research Experience for Students Program at the Mainz Microtron award, NSF/DoE

CHAPTER I

Transcriptomic Data Integration in Multi-Omic Studies

1.1 THE RICH AND EXPANDING LANDSCAPE OF THE TRANSCRIPTOME

1.1.1 Why study the transcriptome?

Given its central role in numerous aspects of cellular functioning, the transcriptome has become a major candidate for in-depth molecular investigations of disease states, readily competing with other "omic" tools (Figure 1.1) Profiling of the transcriptome, which encompasses the entire set of assorted ribonucleic acid (RNA) molecules, is extensively used by researchers to evaluate pathogenic disruptions in normal cellular functions. As encapsulated in the central dogma of molecular biology, the transcriptome lies as an intermediate between the genome (consisting of the biological units of inheritance) and the proteome (consisting of the final set of functional products). It offers a snapshot of information transcribed by the genome, within a specific cell type
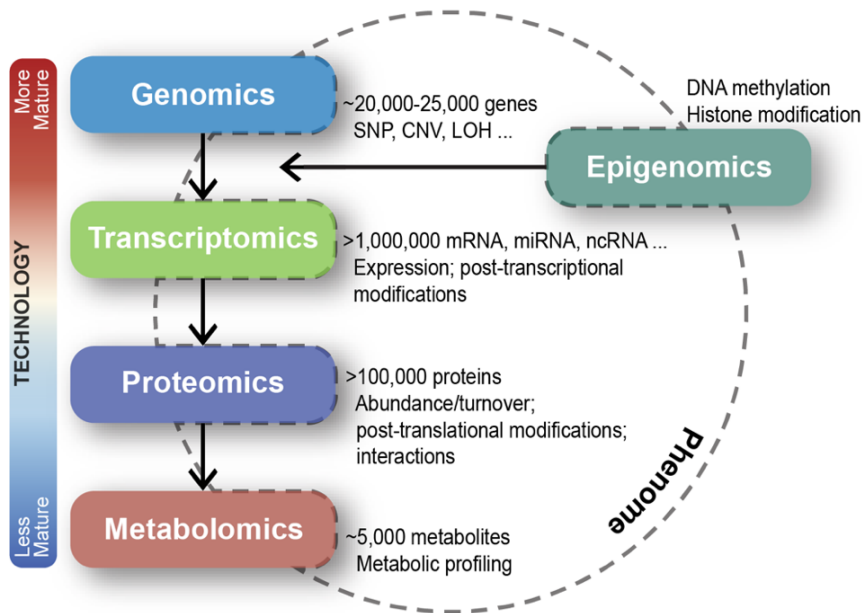
Figure 1.1 | Omic analysis in phenotype assessment

Flow of omic data within the central dogma of biology and in the context of phenome characterization. Abbreviations include: SNP – single nucleotide polymorphism; CNV – copy number variation; LOH – loss of heterozygosity; mRNA – messenger RNA; miRNA – microRNA; and ncRNA – non-coding RNA.

and tissue, during a precise stage of development, and under specific physiologic conditions. While the genome only provides a static characterization of genetic variations that define a phenotype, the transcriptome reflects dynamically-evolving gene activity. Quantification of protein-coding (1-4% messenger) and non-coding (>95% - ribosomal, small nuclear, micro-, long non-coding, small nuclear, small interfering, etc). RNA expression allows us to evaluate the influence of environmental factors and varying physiological demands on genetic information as it relates to the manifestation of a phenotype (Manzoni *et al*, 2018). Transcriptomic analysis also informs gene structure and function as well as gene expression plasticity and regulation. By reflecting cellular state at the transcription level and exposing cellular dynamics, it therefore offers a unique look into the inner workings of a cell –one that can be missed using other omic analyses. With each newly characterized class of RNA molecules, the dominant agency of the transcriptome continues to be reinforced, necessitating its incorporation in biomedical investigations.

From a technology perspective, transcriptomics is impermeable to the limitations of novel proteomic-based approaches. This includes bias towards highly abundant proteins, an input with a large domain size and a dynamic nature, and coverage that depends on sample preparation and separation methods (Dove, 1999; Smaczniak *et al*, 2012). And albeit its inability to relay cellular biochemical activity, it surpasses metabolomics in its simplicity, maturity, and sensitivity, and reproducibility. The lead of transcriptomics is further enhanced by the advent of sequencing technologies that have facilitated rapid, high-throughput, and robust investigations of the transcriptome at the single nucleotide resolution (Anderson & Schrijver, 2010).

Transcriptomics holds a great potential for prognostic and diagnostic biomarker research, patient stratification, and therapeutic drug application. Nevertheless, its translatability to the clinic is limited by the need for invasive and risky extractions of less readily available tissue biopsies as well as the slow turnaround times and the high costs of operation (Damodaran *et al*, 2015). "Liquid biopsies" have gained tremendous traction in the last few years, promising to detect diseases

during the early stages and with minimal invasiveness. However, due to their novelty, low detection rates, and unclear biology, There is also the concern of limited harmonization across platforms and standardized benchmarking to ensure accuracy and reproducibility (Damodaran *et al*, 2015; Van Keuren-Jensen *et al*, 2014). Finally, inherent to transcriptomic analysis, there is a lack of established metrics that can minimize artifacts and false discoveries and rigorous protocols that can correct for biases and low-quality and -abundance specimens (Damodaran *et al*, 2015; Van Keuren-Jensen *et al*, 2014).

1.1.2 Eukaryotic isoform complexity and diversity
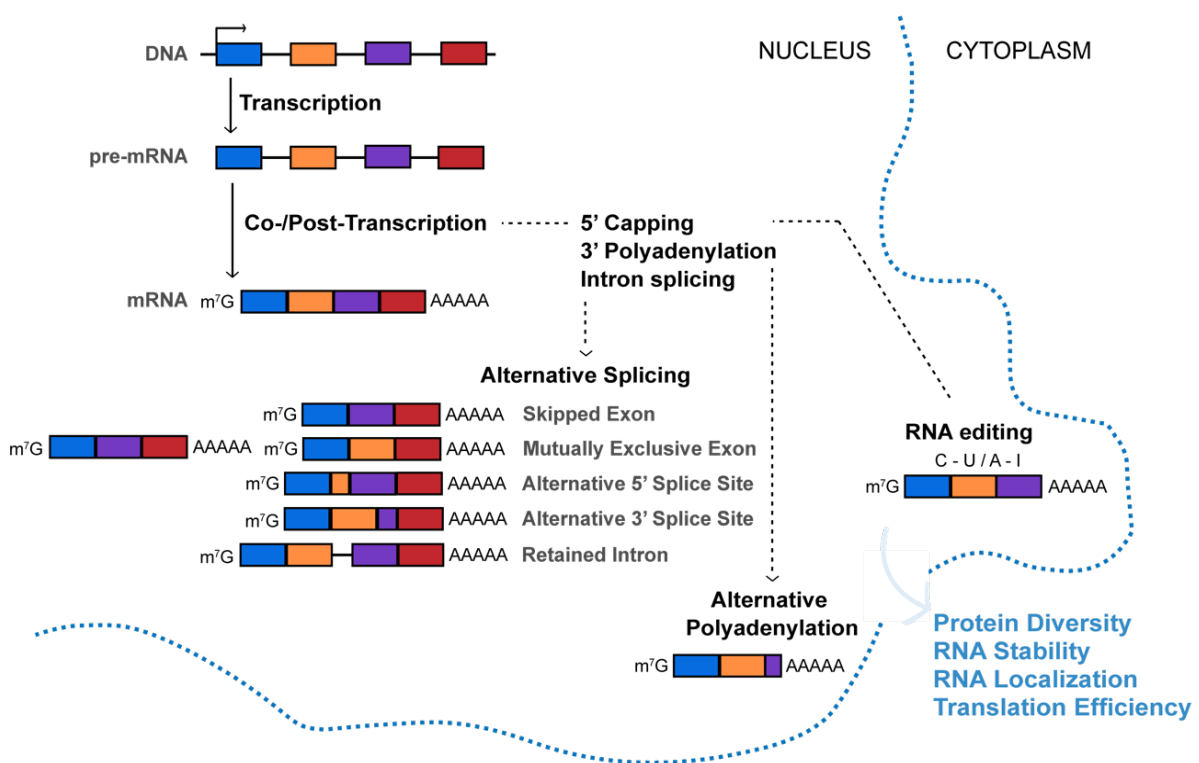
Figure 1.2 | Eukaryotic transcription and gene expression

The evaluation of messenger RNA (mRNA) gene expression pattern differences between normal and pathogenic (or abnormal) states is commonly pursued to identify genes involved in disease etiology. In eukaryotes, protein-coding genes are transcribed as a nascent pre-mRNA and then further processed through 5' 7-methyl guanosine capping, 3' polyadenylation, and intron splicing

to generate the final mature mRNA (Figure 1.2). Modifications in the transcription and RNA processing stages often occur in a spatiotemporal and cell type-, stage-, and condition-specific manner, contributing to a diverse set of protein products and a high degree of variability in RNA translation, stability, and localization (Figure 1.2). Such modifications can yield consequences of substantial functional magnitude, including the acquisition of pathogenic traits. By studying the mRNA, we can explore these stages and learn about changes in **(a)** transcribed products – through the quantification of gene- and isoform-level abundances and **(b)** co- and post-transcriptional regulatory mechanisms – through the evaluation of alternative splicing, alternative polyadenylation, and RNA editing.

Dysregulation in transcription and alternative splicing (AS) has been extensively linked to the production of altered disease-driving transcript variants and isoforms (Scotti & Swanson, 2016). Variations at the transcriptional level can be monitored via differential gene expression profiling and at the AS co-/post-transcriptional level via differential isoform-specific gene expression profiling or direct quantification of AS events. AS is a highly prevalent and versatile regulatory RNA processing regulatory mechanism, undergone by more than 90-95% of multi-exonic human genes (Shen *et al*, 2014). It is a major contributor to the estimated 5-10 fold discrepancy between the number of protein-coding genes and generated proteins (Wang *et al*, 2014; Scotti & Swanson, 2016). Changes in AS patterns can result in either a complete switch in protein isoforms or more commonly a deviation in relative splice isoform abundances. Even a small shift in splice isoform proportions is sufficient to induce pathogenesis (Buchner *et al*, 2003). This highlights the importance of AS in promoting disease-driving traits. AS involves coordinated *cis* and *trans* splicing mechanisms that are catalyzed by the spliceosome ribonucleoprotein enzymatic complex. In addition to constitutive splicing, AS can occur through the following basic modes: **(a)** skipped exon (the most common form); **(b)** mutually exclusive exon; **(c)** alternative 5' splice site; **(d)** alternative 3' splice site; and **(e)** retained intron (Figure 1.2). In human cells,

skipped exon, or cassette exon skipping, represents the most common form of AS, in which exons are either included or skipped from the final transcript.

1.2   NEXT-GENERATION RNA SEQUENCING

1.2.1 Why use RNA sequencing?

A variety of hybridization- and sequencing-based tools have been developed to quantify the transcriptome. Notwithstanding the popularity of microarray technology, high-throughput next generation sequencing (NGS) is surpassing Moore's law predictions, rapidly becoming the platform of choice for transcriptional profiling (Hitzemann *et al*, 2013). Hybridization-based methods are limited by their dependence on a predefined transcriptome and focused measurement of mRNA with corresponding homologous printed probes (Steger *et al*, 2011; Zhang, 2005). This has resulted in high noise and saturation background levels and cross-hybridization in microarrays, constraining their dynamic range of detection and inflating error due to the nonlinear dye response (Koltai & Weingarten-Baror, 2008). RNA sequencing (RNA-seq) has become the preferred choice for transcript quantification, despite a higher cost and analytical complexity (Hitzemann *et al*, 2013). With the exponentially declining sequencing costs and advances in sequence detection methods, library preparation protocols, and multiplexing capabilities, RNA-seq has revolutionized gene transcriptomic analysis, translating into countless novel discoveries of pathogenic molecular signatures.

1.2.2 RNA sequencing workflow

RNA-seq can be performed either at the cell-population level (bulk RNA-seq) or the single-cell level (single-cell RNA-seq). Investigations of polyadenylated (polyA$^+$) mRNAs from bulk tissue involves the following sample-to-insight workflow:

Table 1.1 | Bulk RNA-seq workflow

RNA-SEQ		
(1)	Data Generation	RNA extraction; RNA fragmentation; library construction and sequencing
(2)	Data Pre-processing	Raw data quality checking; adapter sequence trimming
(3)	Data Processing	Read mapping to reference genome/ transcriptome/ *de novo* assembly; indexing to coding regions/ splice junctions; annotation extraction
PROCESSED RNA-SEQ		
(4)	Abundance Quantification	Gene expression levels
(5)	Data Manipulation	Normalization; transformation; missing data imputation
(6)	Differential Analysis	Differential gene / isoform expression; differential alternative splicing
(7)	Data Visualization	Pattern recognition; knowledge extraction
(8)	Pathway Analysis	Functional enrichment
(9)	Network Analysis	Multi-dimensional network representation; regulatory interactions
PROCESSED RNA-SEQ		
(10)	Data Integration	Phenotypic correlation; feature space reduction
(11)	Enrichment Analysis	Functional interpretation; biological insights

Gene expression levels are estimated by mapping RNA-seq reads or k-mers (read sequences of length k; used by pseudo-aligners) against a reference genome or transcriptome. Expression units are then processed through a differential expression analysis pipeline to extract key differential genes that can be investigated for functional enrichment or embed in a network framework for a systems evaluation of interactions. Pre-mRNA alternative splicing differences can then be analyzed to explore the role of co- and post-transcriptional events in modulating gene expression and rendering multiple gene isoforms. Alternative splicing analysis can be achieved through the evaluation of either differential isoform expression (using transcript-based tools) or differential splicing (using event-based tools) (Park *et al*, 2018). For direct differential splicing analyses, exon inclusion levels of alternatively spliced cassette exons are estimated using the percent spliced in (PSI or Ψ) metric that represents the percentage of gene isoforms including the exon of interest. The various modes of alternative splicing (Figure 1.2) can help investigators

incorporate information on the dynamic regulation of gene function, thus spatially and temporally accounting for environmental impacts on pathogenic development.

1.3   MULTIOMICS IN MULTIFACTORIAL DISORDER ASSESSMENT

Despite the expansive development of computational methods and high-throughput technologies, scientists struggle in identifying causal factors in multifactorial disorders and medical conditions. This is notably a result of their composite (having many causes), multifaceted (having many features), and comorbid (co-occurring with other disorders) nature. A simple differential analysis of one type of biomolecule is unable to account for the inherent multilevel structure of biological mechanisms. It also misses the collective involvement of various types of biomolecules and extrinsic factors in disease etiopathogenesis. Furthermore, the variation in clinical manifestation and treatment response between patients has fostered the dissolution of a "one-size-fits-all" attitude in evaluation in favor of personalized and individually-tailored options. Such variability has necessitated careful patient stratification, which requires the inclusion of individual-level risk and prognostic information to complement population-level findings. We propose that a multilayered profiling of the transcriptome and a clinically-informed integration with other omic data can help mitigate some of those hurdles. For this work, we focus on the incorporation of phenomic and radiomic data due to their value in precision medicine endeavors. Those two forms of omic data are already collected in the clinic and offer extensive coverage of patient-specific traits. The transcriptome provides rich molecular information at a cell-to-tissue resolution. The radiome provides non-invasive and longitudinal radiophenotyping at the tissue/organ resolution. And the phenome provides an eclectic repertoire of extrinsic contributing factors.

1.3.1 Why incorporate phenomic data?

The digital revolution has accelerated the collection of rich patient-specific electronic health data with measurements gathered from clinical practice, biomedical research studies, independent health surveys, and patient-reported outcomes. Such records often include a large selection of

data sources: **(a)** demographics – such as age, gender, race, and socioeconomic status; **(b)** health risks – such as behavioral and lifestyle factors; **(c)** medical history – such as prior procedures, pregnancies, and family history; **(d)** management of current condition – such as lab tests, medications, allergies, diagnostics, and therapeutics; and **(e)** self-reported measures – such as sleep, exercise, and diet (Nathanson, 1994).

With the push towards personalized and precision health care, researchers are now recognizing the value in incorporating detailed high-dimensional and longitudinal phenotypic data –referred to as the **phenome** – into omic studies. The value of phenome-level information is especially evident when studying multifactorial disorders. Such disorders often arise from the combinatorial role of many elements, some of which are not easily identified, or are unaccounted for, in mere molecular evaluations. By providing an extra layer of information and facilitating further sample stratification, linking phenomic traits can therefore improve the clinical significance of transcriptome-level findings.

1.3.2 Why incorporate radiomic data?

Medical imaging has long been a standard in the diagnostic and therapeutic assessment of diseases and the inference of clinical outcome. Physicians use medical imaging on a regular basis to monitor morphologic, functional, molecular, metabolic, and microenvironmental changes in patients *in vivo*. Structural medical imaging modalities, such as X-ray, magnetic resonance imaging (MRI), and computed tomography (CT), are used to observe anatomical abnormalities (Mahesh, 2013). On the other hand, functional imaging modalities, such as functional-MRI (fMRI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT), are used to assess physiologic activity (Mahesh, 2013). By melding complementary imaging modalities, modern multimodal devices have expedited the correlation of anatomy to pathogenic function and molecular composition (Martí-Bonmatí *et al*, 2010; Padhani & Miles, 2010).

The standardization of imaging protocols, along with automated image registration, alignment, and segmentation and the availability of reference atlases, has propelled imaging from a largely qualitative tool to a robust quantitative measure. This transition has been supported by abundant open-source and user-friendly image analysis software. Using those tools, investigators can quickly perform quality control steps to alleviate acquisition scheme- and operator-induced errors as well as normalize imaging data across normal variations (Filippi *et al*, 1998). With a variety of available manual, semi-automatic, and automatic methods, they can then proceed to generate quantitative imaging features of interest –referred to as the **radiome.** A region-of-interest (ROI) can be used to extract signal intensity information within pre-defined anatomical areas, albeit with a limited precision in the evaluation of smaller regions (Poldrack, 2007). Voxel-based methods can also be employed to carry statistical tests across image voxels and identify correlates to preselect covariates of interest (Smith *et al*, 2006). Voxel-based results, nevertheless, are prone to misinterpretation due to misalignment, imperfect registration to standard space, and arbitrary spatial smoothing. Tract-based spatial statistics (TBSS) can be introduced to mitigate those concerns, carefully tuning registration and projecting onto an alignment-invariant tract representation (Smith *et al*, 2006).

The non-invasive yet inclusive nature of the radiome has made it a valuable data point in biomedical investigations and an ideal biomarker for clinical diagnoses and treatment responses. Radiomic features offer a quick yet comprehensive characterization of many diseases. They can be generated on a recurrent basis to monitor progress and at a relatively lower cost relative to biopsy-based options. On its own, however, imaging can fail to capture early disease onset or multifaceted disorders that lack a standard trend.

1.4 OVERVIEW, SPECIFIC AIMS, AND USE CASES

This dissertation leverages a multiomic framework –integrating transcriptomic, phenomic, and radiomic data – to investigate multifactorial disorders.

OVERALL AIM: To evaluate the ability of a transcriptome-phenome-radiome integrative framework to uncover clinically-relevant findings otherwise undetected through hypothesis-driven or monolayer differential analysis

AIM 1
Mine the transcriptomic underpinnings of the system of interest, using a variety of analytical tools to profile gene expression, co-expression, and alternative splicing as well as measure tissue-, condition-, and function-specific enrichment

AIM 2
Incorporate complementary phenomic and radiomic features, correlating them with transcriptome-level information

AIM 3
Derive candidate genes and biological processes that correspond to a specific phenotype

(transcriptome ↔ clinical trait / transcriptome ↔ radiophenotype)

AIM 4
Apply this framework to <u>two use case scenarios</u> to demonstrate its strengths and weaknesses

1.4.1 USE CASE SCENARIO 1 – Examine the transcriptomic and phenotypic landscape of various modes of parturition in late-stage pregnancy

<u>Objective</u>: Can the integration of decidual transcriptome and maternal / fetal phenome inform normal and abnormal parturition in pregnancy?

We explore this objective in **Chapter II**, in which we collect the following data points: **(1)** in-house raw RNA sequences from decidual placental tissue at birth; **(2)** in-house clinical data highlighting birth and pregnancy complications and outcomes, interview information, maternal chart abstraction, maternal labs and self-report health information, and infant labs; and **(3)** publicly-accessible lists of placental tissue- (Human Protein Atlas) and labor-enriched genes (peer-

reviewed publication). We demonstrate that a multilayered analysis of the transcriptome allows us to identify genes that repeatedly show significant differential expression, co-expression, and splicing. We then show how integrating phenotypic data can help us narrow down our list of genes to those with significant association with pregnancy risk and prognostic factors. Finally, we functionally annotate our findings and incorporate prior knowledge of mechanisms involved in late-stage pregnancy to delineate the molecular story of abnormal and normal parturition. Overall, we note the strength of a transcriptome-phenome integrative framework in **(a)** extracting biologically meaningful information that would have otherwise been missed through simple differential analysis and **(b)** overcoming limitations of analyzing clinical samples that have been collected from different center, under different protocols, and over a span of a few years and that are inherently heterogeneous. This use case demonstrates the strengths of our proposed framework.

1.4.2 USE CASE SCENARIO 2 – Examine the heterogeneous transcriptomic and radiomic landscape of glioblastoma tumors

Objective: Can the integration of tumor transcriptome and MRI radiome reveal unique mRNA isoform signatures for glioblastoma (GBM) morphologic radiophenotypes?

We explore this objective in **Chapter III**, in which we collect the following data points: **(1)** The Cancer Genome Atlas (TCGA) RNA sequencing read counts for GBM tumor samples; **(2)** corresponding TCGA raw RNA sequences; **(3)** corresponding The Cancer Imaging Archive (TCIA) T1- and T2-weighted MRI images; and **(4)** publicly-accessible list of glioma-intrinsic genes (peer-reviewed publication). We demonstrate that the incorporation of multiple layers of the transcriptome (gene expression and cell type-specific enrichment) can be used to focus the analysis on glioma-enriched genes. We show how this approach can help mitigate the heterogeneity between transcriptome and radiome-level data, allowing us to evaluate the relationship between tumor-specific transcriptome and tumor-specific morphologic radiome. We

also show how this approach can be employed to safely reduce the high-dimensional molecular feature space that often limits correlation-based investigations in the fields of radiogenomics and "radiotranscriptomics". We then extract exon inclusion – radiophenotype correlates and functionally annotate them to highlight transcriptome-level molecular differences between the different GBM tumor phenotypes. Overall, we note the ability of a transcriptome-radiome integrative framework to provide exploratory insights into the biological state of GBM tumor phenotypes. We also note the framework's weakness in revealing significant associations as a result of several data quality and analytical limitations. These include: **(a)** inherent intratumoral and inter-patient heterogeneity, **(b)** variability in image acquisition, tumor segmentation, and image feature extraction, **(c)** limited evaluation of image feature robustness, **(d)** lack of textural image features to capture the inherent heterogeneity, and **(e)** high degree of missing values in exon inclusion estimates. This use case demonstrates the weaknesses of our proposed framework.

CHAPTER II

Integrated Transcriptomic and Phenomic Analysis of Various Modes of Parturition

2.1   ABSTRACT

The multifactorial and paradoxical nature of pregnancy has rendered it difficult to decipher its intricacies. A comprehensive analysis of placental organization –from early stages of gestation to parturition – can provide a keener insight into the underlying molecular machinery. We examine the molecular signature of different forms of parturition to explore the pregnancy paradox. We perform an exhaustive investigation of transcriptomic profiles of human placental decidual samples from normal spontaneous vaginal (N=16), cesarean section (N=18), and preterm labor (N=16) deliveries. We use RNA sequencing to characterize the gene expression and splicing dysregulation signature of abnormal parturition. We identify significant differences in the expression and exon inclusion levels of genes involved in inflammatory and stress response, extracellular matrix remodeling, neovascularization, and lipid metabolism. We also portray the decidual co-expression architecture during parturition and its relationship with prognostic maternal, fetal, and pregnancy characteristics. This study provides the first multilevel survey of the transcriptional and post-transcriptional landscape of human decidua under various modes of parturition. Through a systems integration of transcriptomic and rich phenomic information, it expands on previous findings by highlighting the orchestrated and synergistic interplay of key factors in pregnancy.

2.2   INTRODUCTION

2.2.1 The pregnancy paradoxical complexity

Pregnancy is paradoxical in its nature. The host not only tolerates an invading foreign body but also nurtures and sustains its growth irrespective of its overconsumption of available resources.

This is a testament to the well-maintained and adaptive nature of the human body in response to changing environmental cues. A major shift in the inherent biological state of the mother, from expulsion to tolerance, is necessary to ensure the successful establishment and maintenance of pregnancy. A growing body of evidence has shown that an immunosuppressive phenotype is set forth within the decidua - the immunologically distinct maternal component of the placenta that directly interfaces with the feto-placental unit (Mori *et al*, 2016; Salker *et al*, 2010). Prompted by elevated ovarian steroid hormone levels, decidualization of the endometrium commences following embryonic implantation into the uterine luminal epithelium (Lei *et al*, 2012). This process regulates placentation and initiates fetomaternal cross-talk to prepare the mother's immune system for the semi-allogeneic fetus (Cartwright *et al*, 2010; Faas *et al*, 2014; Lei *et al*, 2012). Decidualization plays a central role in pregnancy – from conception to parturition – by orchestrating maternal-fetal immunologic interactions and signaling cascades that mediate cell-to-cell and cell-to-extracellular matrix (ECM) communication and vasculature remodeling (Kim *et al*, 2012; Vinketova *et al*, 2016). Leukocytes, including natural killer cells (NKs), macrophages, T-cells, and dendritic cells (DCs), are recruited to support local immune function, through protection against infection and immunomodulation (Faas *et al*, 2014; Pavličev *et al*, 2017). The hemodynamic network within the uterus concurrently expands through a dynamic series of vasodilation, cellular hypertrophy, and hyperplasia (Osol & Moore, 2014; Sipos *et al*, 2013). The decidual stromal matrix undergoes extensive reorganization to support the rapidly proliferating vessels and to ensure an undisrupted supply of nutrients and oxygen to the fetus, (Smith *et al*, 2016). These events evolve synergistically in a spatial and temporal manner throughout the course of gestation, with varying degrees of involvement per pregnancy stage-specific needs. During the final stage of pregnancy, a fetal rejection program is invoked to initiate labor. Physiological changes progress in a coordinated fashion: from uterine quiescence and cervical softening prior to parturition to uterine activation and cervical ripening in preparation for labor-associated myometrial contractions and cervical dilation (Timmons *et al*, 2010). It is believed that

a successful onset of active labor is triggered by a precise combination of attenuation of earlier processes in pregnancy and pro-inflammatory signaling during later stages (El-Azzamy *et al*, 2017; Tan *et al*, 2012). The same hormonal, humoral, vascular, and adhesion bioactive mediators that initially helped sustain the fetus now contribute to the inflammation burst and the proceeding degradation of the ECM in gestational tissues-that is needed to expel it (Marcellin *et al*, 2017). By comprehensively and systemically investigating the underlying dynamically evolving molecular signature of pregnancy, we can decipher the varying contribution of different biological components and ultimately isolate key players in pregnancy-associated pathologies.

2.2.2   Normal and abnormal parturition

In an attempt to delve into the pregnancy paradox, we examine the transcriptomic landscape of placental deciduae during parturition in late-stage pregnancy. We compare normal labor cases to two forms of abnormal parturition: cesarean section (C/S) and preterm labor (PTL). Cesarean section (C/S) –the surgical delivery of a baby via incisions in the abdominal wall and uterus of the mother – is performed when there is failure or fear of induction of labor. Given its invasive nature, it has been associated with several intra- and post-operative maternal complications. This includes infection, post-partum hemorrhage, organ injury, uterine rupture, placental anomalies, and infertility (Mylonas & Friese, 2015). C/S has also been shown to increase neonatal risk for
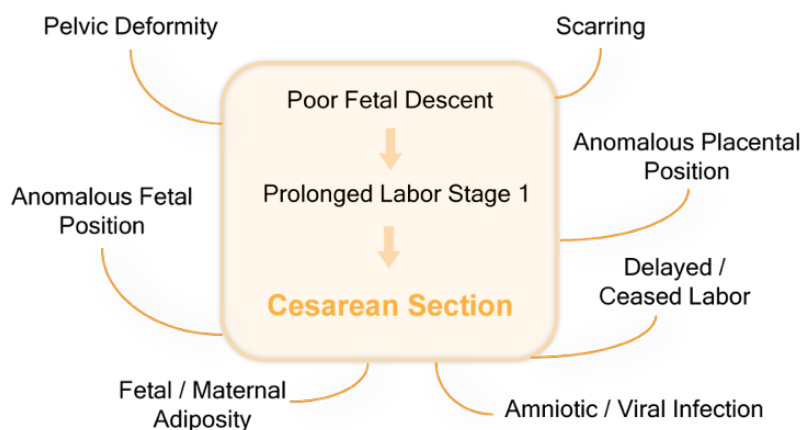
Figure 2.1 | Indications of Cesarean section deliveries

respiratory disease, type 1 diabetes mellitus, and even mortality (Mylonas & Friese, 2015). With only a few elective cases, C/S procedures are typically scheduled per doctor's recommendation due to obstetrical indications (Mylonas & Friese, 2015). Such indications include pelvic deformity, anomalous placental or fetal position, amniotic or viral infection, delayed or ceased labor, and prior C/S delivery (Mylonas & Friese, 2015). PTL –labor prior to 37 completed weeks of gestation – is a leading cause of perinatal and neonatal morbidity and mortality (Mwaniki *et al*, 2012). Health implications of PTL have been shown to extend into adulthood, with higher incidences of neurodevelopmental and growth deficits, chronic medical problems, and recurrent hospitalization (Blencowe *et al*, 2013). A number of pregnancy-related diseases are risk factors for PTL. These include multiple pregnancies, preeclampsia, gestational diabetes, short cervical length, uterine over-distension, amniotic fluid leak, infections, and cervical disorders (Asl *et al*, 2017; Hermans *et al*, 2015; Koucký *et al*, 2014). While direct causes of PTL remain unclear, numerous studies have proposed the involvement of T-regulatory lymphocytes (Treg), ECM glycoproteins, oxidative stress, progesterone dysfunction, decidual senescence and hemorrhage, and abnormal uterine vascular remodeling within the "decidual clock" program (Asl *et al*, 2017; Hermans *et al*, 2015;
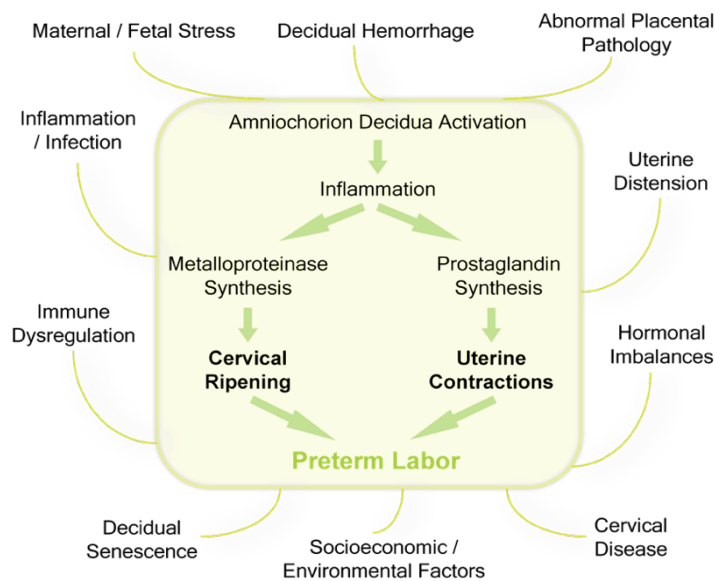
Figure 2.2 | The speculated heterogeneous etiology of preterm labor

Koucký *et al*, 2014; Norwitz *et al*, 2015) (Figure 2.2). Studying C/S and PTL offers a preview into the ramifications of imbalances in the homeostatic state of the mother. It thus holds the potential to get us a step closer to deciphering the complex pregnancy paradox.

2.2.3 Why use a multi-layered transcriptome-phenome framework in pregnancy?

The high prevalence and cost –physically, emotionally, and economically – of pregnancy complications, including prematurity, have rendered them a major public health priority. Premature birth, on its own, accounts for over 1 million deaths worldwide and over $26 billion in societal costs in the US every year (Blencowe *et al*, 2013; Liu *et al*, 2016; Boardman, 2008). Health implications of prematurity extend far beyond the neonatal period, with numerous reports on lifelong effects on neurodevelopment and growth and higher rates of chronic medical problems and recurrent hospitalization (Blencowe *et al*, 2013). Global-scale efforts to devise innovative and actionable plans to address this growing medical urgency have significantly increased in the recent years (Blencowe *et al*, 2013). Nonetheless, researchers continue to lag in unraveling the underpinnings of pregnancy, resulting in a paucity of robust predictors of associated pathologies in clinical practice. In Table 2.1, we briefly highlight elements that we believe have hindered translational progress in pregnancy research.

A multidimensional examination of human placentae undergoing different forms of labor and delivery can help mitigate some those of barriers, ultimately cultivating our understanding of

Paradoxical:	• A host immune system that needs to tolerate a foreign body yet equally defend against pathogens • A host vascular system that needs to increase the blood supply of nutrients to support a growing fetus yet maintain a safe level of vascular organization to prevent bleeding
Transient:	• Typically lasting 40 weeks and involving the placenta, a temporary yet critical organ in pregnancy

Multifactorial:	• An orchestra of immune; pro-inflammatory; neuro-endocrine; extracellular matrix; cytoskeletal; metabolic; and growth biological factors as well as genetic and environmental factors
Synergistic:	• An orchestra composed of biological factors that synergistically interact in a controlled fashion to maintain harmony amidst a less controllable external environment
Multifaceted:	• Discordance in the harmonious interaction of involved factors can yield a wide range of complications including preterm labor; (pre)eclampsia; placental abruption and previa; intrauterine growth restriction; and miscarriage
Dynamic and Temporal:	• Biological processes that evolve throughout the course of gestation, with significant variation across the 1st, 2nd, and 3rd trimesters and parturition due to varying stage-specific demands
A 2-in-1 scenario:	• Mediated via the placenta, which is composed of both maternal (decidua) and fetal (chorion and amnion) components • Mis-alignment in maternal and fetal goals for optimal survival
Ethically challenging:	• Knowledge-based investigations of early stages of pregnancy necessitate invasive measures that pose major risks to both the mother and the fetus • Pregnant women cohorts are typically excluded from clinical drug trials, yielding a significant knowledge gap in potential points of intervention

Table 2.1 | Elements that hamper pregnancy research

the impact of molecular perturbations in late-stage pregnancy. While such method arguably only provides a static capture of parturition-specific factors, when viewed in the context of a system, it can also offer a snapshot into mechanisms generally involved in the unidirectional flow of pregnancy.

2.2.4 Study design

We apply a systems approach –as depicted in Figure 2.3 – to evaluate transcriptomic changes associated with normal parturition (normal spontaneous vaginal delivery; NSVD) and abnormal parturition (cesarean section; C/S and preterm labor; PTL). We analyze the transcriptomic landscape of the placenta given its relevance to placental health and pregnancy complications

(Kleinrouweler *et al*, 2013; Struwe *et al*, 2010). We compare gene expression and alternative splicing profiles of samples collected from human placentae, to derive molecular processes that manifest in labor and aberrant parturition timing. We specifically evaluate the decidual tissue to explore the role of maternal factors and endometrial immunological priming in abnormal parturition. We mine multiple levels of data to account for the heterogeneous molecular and phenomic makeup of patients. By integrating findings from the various analyses, we aim to highlight term- and labor-specific transcriptomic and phenomic signatures as well as elucidate molecular mechanisms and key players potentially involved in abnormal parturition in general.

2.3   METHODS

2.3.1 Sample collection

Human placentae are obtained within one hour of delivery with signed informed consent under the protocols approved by the University of Iowa Institutional Review Board (201411731). Samples are obtained from 16 term vaginal deliveries (NSVD), 16 preterm deliveries (PTL), and 18 cesarean deliveries (C/S). Placenta decidual tissue samples are macroscopically isolated from the maternal-facing surface of the placenta. Samples are cut into small pieces and placed in RNAlater® solution (Applied Biosystems, Foster City, CA). Biopsies are obtained after written informed consent, according to a protocol approved by the University of Iowa and in accordance with the Department of Health and Human Services regulations at 45 CFR 46. Maternal and infant chart abstraction, maternal and infant labs, medical history, and social history are available in Appendix Table 2.1.
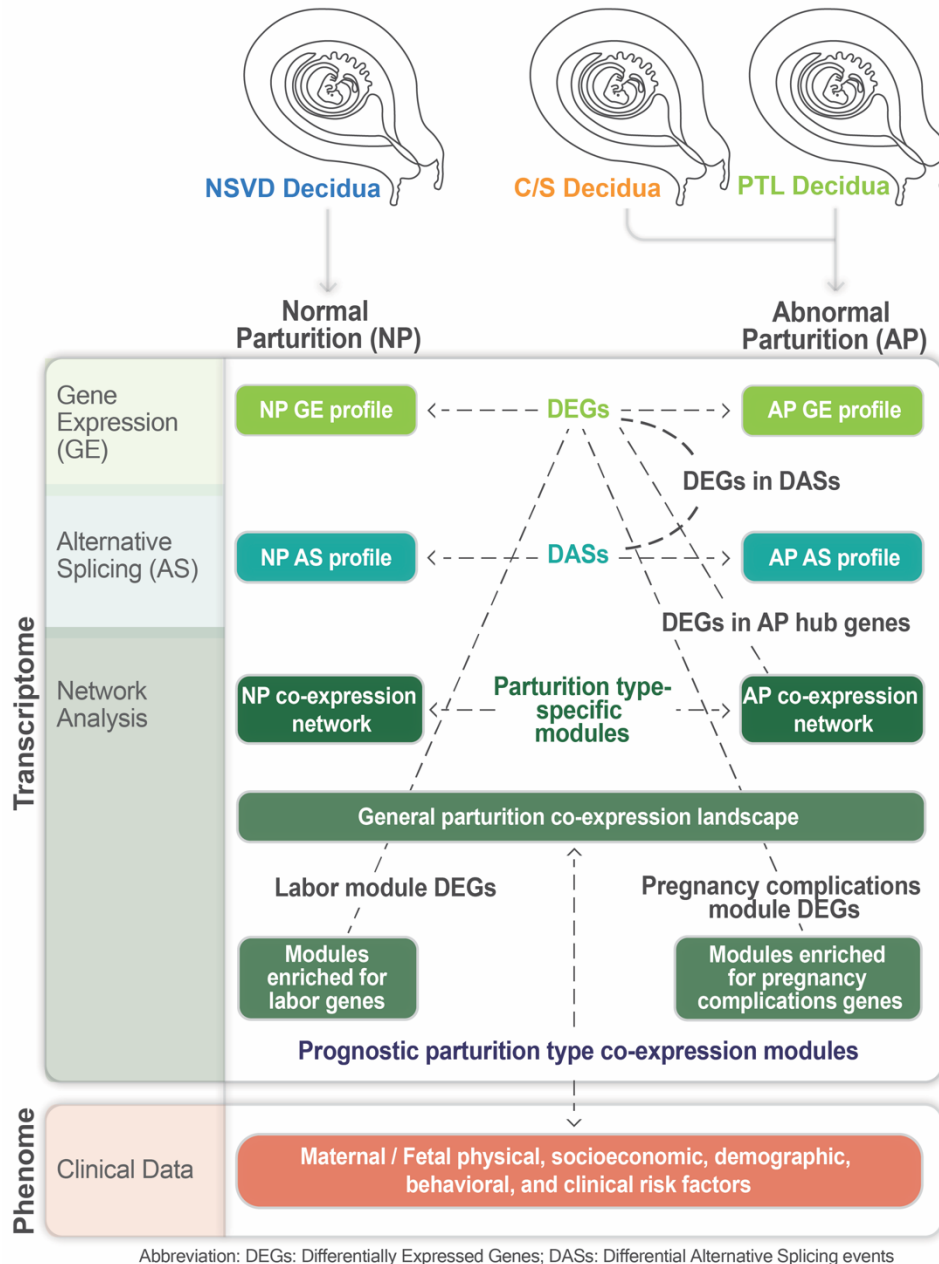
Figure 2.3 | Overview of analysis workflow

Flowchart describing the multiple levels of analysis in the study. RNA-seq data from normal spontaneous vaginal delivery (NSVD) deciduae are used to generate gene expression (GE) and alternative splicing (AS) profiles and a co-expression network for normal parturition (NP). RNA-seq data from cesarean section (C/S) and preterm labor (PTL) deciduae are used to derive GE and AS profiles and a co-expression network for abnormal parturition (AP). Pairwise comparison is used to identify significantly differentially expressed genes (DEGs), differential alternative splicing events (DASs), and gene clusters specific to each parturition type. RNA-seq data from all samples are used to create a co-expression landscape for general parturition, which in turn is correlated with clinical data to identify gene clusters that are prognostic indicators of parturition type. Within the general parturition co-expression network, we also highlight modules that are enriched for labor and pregnancy complications genes. Finally, we integrate findings from individual analyses to investigate DEGs in DASs; DEGs in abnormal parturition co-expression modules; DEGs in general parturition co-expression modules that are enriched for labor-associated genes; and DEGs in general parturition co-expression modules that are enriched for pregnancy complications genes.

20

2.3.2 RNA extraction and library preparation and sequencing

Total RNA is extracted from each tissue using the TRIzol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instruction and stored at -80˚C until used. RNA integrity is measured using an Agilent Bioanalyzer. RNA samples included in the analysis had an RNA integrity number (RIN) > 6. RNA-seq libraries are prepared using TruSeq RNA Sample Prep Kit v2 (Illumina) and sequenced on an Illumina HiSeq 2000 to produce 27-49 million 2x100nt pair-end reads per sample. The entire RNA-seq dataset, including raw and processed files for Gene Expression Omnibus (GEO) submission and sample metadata, is available at Dataset 2.1.

2.3.3 RNA-seq data processing

Raw RNA-seq reads are filtered for adapter sequences and mapped against the Ensembl human genome (hg19; GRCh37.75) using the software STAR v2.4.1c (Dobin *et al*, 2013) with default parameters. The percentage of mapped reads is 88.6%±2.9% (mean±s.d). Transcript abundances are quantified using htseq-count v0.6.1(Anders *et al*, 2015). Raw count data are then normalized with respect to library size using DESeq2 (Love *et al*, 2014) and log-transformed to minimize differences between samples with low counts. Outlier samples are identified by performing pairwise correlation of the normalized expression data of samples and by principal component analysis (PCA). Unsupervised hierarchical clustering of normalized expression data using squared Euclidean distances confirms identified outliers. Samples CS101 and CS105 are removed from downstream analysis (Appendix Figure 2.1).

2.3.4 Differential Gene Expression analysis

Differential Gene Expression (DGE) analysis is performed pairwise (NSVD versus CS; NSVD versus PTL; CS versus PTL) using the DESeq2 R package (Love *et al*, 2014). Genes with significant differential expression show a Benjamini-Hochberg FDR-adjusted Wald test *p*-value ≤ 0.05 and pass an extra filter of max raw read count > 10 to disqualify genes with very low expression levels across samples. Double filtration using a maximum likelihood estimate (MLE)

of log2 fold change $\geq$ 1.5 is performed to limit the number of significant C/S vs PTL gene hits to those with greatest expression difference between the 2 conditions when used as input into functional enrichment analysis.

2.3.5 Network construction

Co-expression networks for each sample group and for the entire dataset are generated using normalized, log-transformed and filtered (max read count > 10) read counts as input into the weighted gene co-expression network analysis (WGCNA) R package (Langfelder & Horvath, 2008). An adjacency matrix is calculated for the signed networks using the pairwise Pearson correlation of each set of genes. The adjacency matrix is then raised to a soft-thresholding power to increase weight of strong correlations while avoiding limitations of a hard "discontinuous" threshold. Per the scale-free topology criterion and a scale-free fit R^2 cutoff of 0.75, a power of β = 14 is chosen for the NSVD network, β = 12 for the CS network, β = 14 for the PTL network, and β = 20 for all samples. Using the blockwiseModules function for automatic network construction, the topological overlap measure (for interconnectedness) is then calculated and used as input for average linkage hierarchical clustering. The dynamic tree cut method is applied to identify dendrogram branches and extract network modules. Genes are grouped into modules with a minimum size of 40 genes and minimum height for merging modules at 0.25. Only genes that exhibit a high module membership to the module (absolute correlation of their expression values with module eigengene - kME) $\geq$ 0.7 are included within a module. Hub genes that exhibit highest number of connections within a module are extracted using kME $\geq$ 0.9.

2.3.6 Network analysis

The conservation of complimentary modules across the three condition-specific co-expression networks is evaluated using the hypergeometric probability of intersecting genes, adjusted using Bonferroni correction for multiple comparisons. Complimentary modules are designated as modules with the maximum number of overlapping genes without repetition. To test for significant

associations between modules in the whole dataset co-expression network and clinical data, we correlate module eigengenes with select patient phenotypic and socio-economic information previously associated with pregnancy complications. We use median-based biweight midcorrelation to limit the effect of outliers. Missing clinical input is imputed separately for each labor subtype by mean. Gene-gene interaction maps for "interesting" modules are created by exporting the WGCNA adjacency matrix, filtered for module genes, into Cytoscape version 3.5.1 (Shannon *et al*, 2003).

2.3.7 Differential Alternative Splicing

Alternative Splicing (AS) events corresponding to all five basic AS patterns (skipped exon – SE; alternative 5' splice site – A5SS; alternative 3' splice site – A3SS; mutually exclusive exons – MXE; and retained intron – RI) are quantified using the percent spliced in (PSI; Ψ) metric using replicate Multivariate Analysis of Transcript Splicing (rMATS) version turbo (Shen *et al*, 2014). rMATS is an exon-centroid method that uses a modified version of the generalized linear mixed model to detect differential AS from RNA-seq data with replicates. For each AS event, we use both reads that span splicing junctions and reads on target (defined as reads that are fully contained within the alternatively spliced region) as rMATS input. Significant differential AS between two sample groups were identified using a cutoff of Benjamini-Hochberg FDR adjusted *p*-value < 0.05 and $|\Delta\Psi| \geq 0.05$.

2.3.8 Enrichment of Gene Ontology (GO) functional categories, tissue-specific genes, and cell types

Functional enrichment of significantly upregulated/downregulated, co-expressed, and differentially spliced genes is assessed using Enrichr (Chen *et al*, 2013) at default settings. Select gene ontology (GO) terms and pathways are extracted with a Benjamini-Hochberg adjusted *p*-value < 0.1 and a combined score ≥ 10, unless otherwise noted. Enrichment analysis for genes with elevated expression levels within placental tissues (with at least five-fold higher mRNA levels

in the placenta compared to other tissues) is performed by assessing the overlap with The Human Protein Atlas placenta-specific gene set (available from www.proteinatlas.org) (Uhlén *et al*, 2015). For enrichment for labor genes, we use meta-analysis results for differentially expressed genes associated with labor (Lee *et al*, 2010). For enrichment for preterm genes, we assess the overlap with public curated lists of genes from The Database for Preterm Birth (Uzun *et al*, 2012) and The Comparative Toxicogenomics Database (CTD) for premature birth (Rouillard *et al*, 2016). For enrichment for genes associated with pregnancy complications, we use the overlap with the list of genes from CTD for pregnancy complications.

2.4 RESULTS

2.4.1 Transcriptome-wide differential gene expression analysis in premature and caesarean deliveries

We conduct deep RNA sequencing analysis of 50 human decidua samples from term vaginal delivery (NSVD; n=16), preterm delivery (PTL; n=16), and cesarean delivery (C/S; n=18) placentae (Figure 2.4A). After processing sequencing data to remove outliers (Appendix Figure 2.1), we calculate gene expression levels represented by read counts and normalize using the

A

	NSVD (n = 16)	C/S (n = 18)	PTL (n = 16)
Maternal Age (year)*	28.9 ± 5.8	30.6 ± 5.5	26.7 ± 5.0
Labor Type	Term spontaneous (10) Term induced (6)	no labor (18)	Preterm spontaneous (15) Preterm induced (1)
Delivery Type	Vaginal vertex (16)	C/S without labor (18)	Vaginal vertex (13) Vaginal breech (1) C/S after labor (2)
Gestational Age (week)*	39.2 ± 1.2	38.9 ± 0.9	31.8 ± 3.4
Rupture of Membrane	S(8) : A(8)	S(1) : A(17)	S(9) : A(7)
Number of Pregnancies Prior to Index Pregnancy*	2.1 ± 1.1	3.1 ± 1.2	2.3 ± 1.7
Prenatal Steroid Treatment	Yes(0) : No(16)	Yes(0) : No(18)	Yes(13) : No(3)
Fetal Gender	F(6) : M(10)	F(6) : M(12)	F(7) : M(9)
Fetal Weight at Birth (g)*	3433.8 ± 439.3	3783.8 ± 578.7	2038.4 ± 739.9
Fetal Head Circumference (cm)*	34.5 ± 1.0	35.8 ± 1.2	29.8 ± 4.1
Fetal Length at Birth	51.0 ± 2.7	51.0 ± 2.8	44.0 ± 6.1

Abbreviation: NSVD, normal spontaneous vaginal delivery; C/S, cesarean section; PTL, preterm labor; F, female; M, male; S, spontaneous; A, artificial
* Values expressed as mean ± standard deviation

B
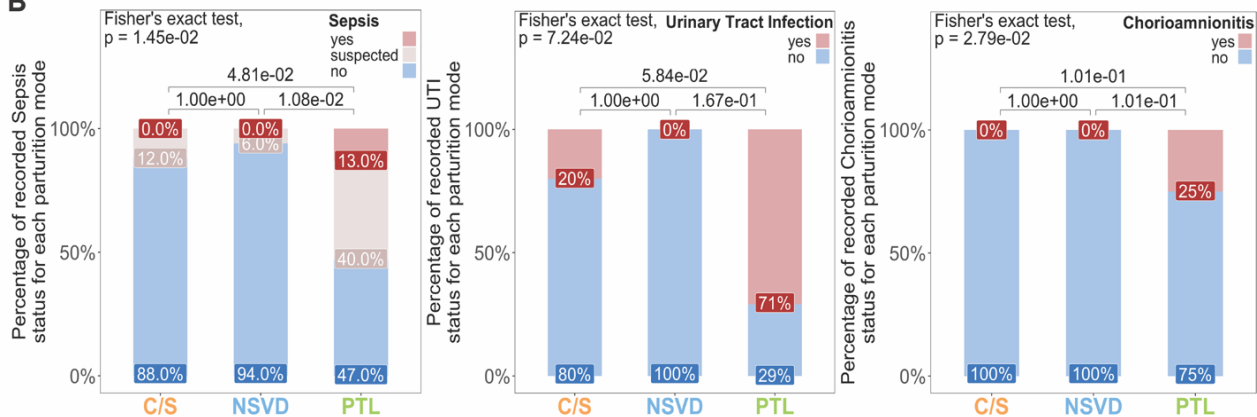
Figure 2.4 | Patient clinical demographics

(A) Select maternal / fetal clinical characteristics. Additional clinical data is found under Appendix Table 2.1 **(B)** Histograms depicting the difference in rates of infection (sepsis; urinary tract infection; and chorioamnionitis) across the 3 parturition groups. Fisher's exact test of independence *p*-values are indicated.

default relative log expression (RLE) method in DESeq2 ([Methods](#)). Using a strict Benjamini-Hochberg adjusted Wald test *p*-value ≤ 0.05, we identify a total of 1028 differentially expressed genes (DEGs) out of 21,465 expressed genes ([Methods](#)), 32 of which have significant altered expression between NSVD and C/S; 14 between NSVD and PTL; and 1010 between C/S and PTL (Figure 2.5A; [Supplementary Table 2.2A-C](#)).
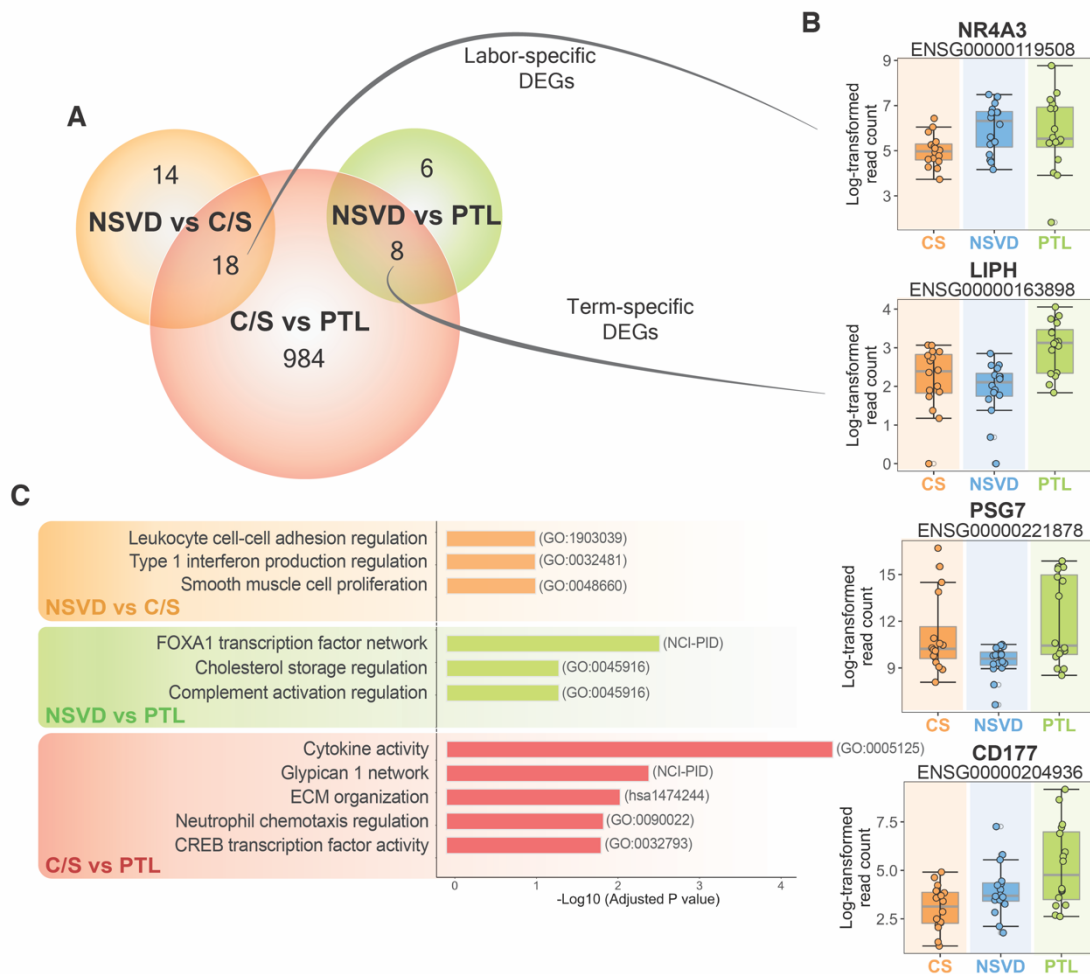
Figure 2.5 | Differences in gene expression profiles of NSVD, C/S, and PTL decidual samples

(A) Venn diagrams displaying the intersection of DEGs detected in pairwise comparison. Genes overlapping in NSVD vs PTL and C/S PTL are designated as term-specific DEGs. Genes overlapping in NSVD vs C/S and C/S vs PTL are designated as labor-specific genes. **(B)** Boxplots displaying the log-transformed read count profile of *NR4A3* (labor-specific DEG overlapping in NSVD vs C/S and C/S vs PTL); *LIPH* (term-specific DEG overlapping in NSVD vs PTL and C/S vs PTL); *PSG7* (NSVD vs PTL DEG with highest $|\log_2$ (fold change)$|$); and *CD177* (C/S vs PTL DEG with highest $|\log_2$ (fold change)$|$. **(C)** List of gene ontology GO) terms and pathways that are significantly enriched in NSVD vs C/S (orange); NSVD vs PTL (green); and C/S vs PTL (red). Significance is set as FDR < 0.1 and a combined score $\geq$ 10 in Enrichr.

We identify 18 DEGs that overlap between NSVD vs C/S and C/S vs PTL and label them labor-specific due to their consistent varied expression signature in non-laboring decidual samples (Figure 2.5A; Supplementary Table S2.2D). *NR4A3* (nuclear receptor subfamily 4 group A member 3) is a labor-specific DEG that exhibits over 2-fold decrease in expression in C/S samples (Figure 2.5B; one-tailed T-test; *p*-value $= 2.57 \times 10^{-3}$). Downregulation of NR4A3 transcription factor has been documented in cases of impaired decidualization (Jiang *et al*, 2016). We perform functional enrichment analysis of NSVD vs C/S DEGs to explore the biological property of genes with significantly altered expression levels in non-laboring deciduae (Materials and Methods). Using Enrichr (Methods), we find enriched gene ontology (GO) terms associated with inflammation, leukocyte adhesion, and smooth muscle cell proliferation (Figure 2.5C, orange bars; Supplementary Table S2.3A). These results are in line with our expectations of an absent pro-inflammatory and contractile state in non-laboring cases.

We then identify 8 DEGs that overlap between NSVD vs PTL and C/S vs PTL and designate them term-specific due to their significantly altered expression in preterm decidual samples (Figure 2.5A; Supplementary Table 2.2D). *LIPH* (lipase H) is a term-specific DEG that exhibits over 2-fold decrease in expression in PTL samples (Figure 2.5B; one-tailed T-test; *p*-value $= 5.95 \times 10^{-5}$). The lipid mediator is heavily involved in smooth muscle contraction and platelet aggregation (Thiriet, 2013). Amongst the 14 DEGs significantly differentially expressed between spontaneous term and preterm, *PSG7* (pregnancy specific beta-1 glycoprotein 7) exhibits the highest fold-increase (21-fold-increase in expression) labor (Figure 2.5B). *PSG7* is highly enriched in the syncytiotrophoblast of the placenta, with a prominent – albeit elusive – immunoregulatory, angiogenic, and anti-platelet role in pregnancy (Aleksic *et al*, 2016; Sanborn *et al*, 1997; Snyder *et al*, 2001). Previous studies have reported significant differences in *PSG7* expression between preterm and term placentae, with a potential role in aberrant trophoblast differentiation and uteroplacental vascular insufficiency (Brockway *et al*, 2017). While mean *PSG7*

expression levels in our study do not drastically differ between the NSVD and PTL sample groups, we note a significantly larger variance and the presence of two subpopulations within PTL (F-test; p-value = 2.49×10^{-4}). The presence of 2 subpopulations within the PTL group appears to correlate with birth order (Spearman $\rho = -0.66$; p-value = 5.38×10^{-2}), in which high *PSG7* expression is observed in nulliparous (first pregnancy) PTL cases (n=8). In fact, mean *PSG7* expression in firstborn PTL placentae is significantly higher than that of firstborn NSVD placentae (unequal variance Welch's T-test; p-value = 1.25×10^{-2}) and overall NSVD placentae (unequal variance Welch's T-test; p-value = 1.28×10^{-2}). We note a similar dichotomy in *PSG7* expression within the C/S group (F-test; p-value = 8.22×10^{-4}), suggesting a general involvement of PSG7 deregulation in abnormal parturition. Functional enrichment analysis of NSVD vs PTL DEGs reveals FOXA1 transcription factor network, a regulator of embryonic development and lipid metabolism (Bochkis *et al*, 2012) and complement pathway activation, a crucial mediator of inflammatory response to infection in preterm labor (Vaisbuch *et al*, 2010) (Figure 2.5C, green bar; Supplementary Table 2.3B).

We observe the greatest distinction in gene expression levels between our C/S and PTL cohorts (n=1010). *CD177* exhibits 2-fold increased expression in C/S compared to PTL (one-tailed T-test; p-value =7.55×10^{-4}; Figure 2.5B). Neutrophil-specific *CD177* encodes a glycosyl-phosphatidylinositol-anchored glycoprotein reported to be upregulated in inflammatory conditions and involved in activated platelet adhesion to endothelial cells (Sachs *et al*, 2007). We then filter for absolute maximum likelihood estimate (MLE) of log2 fold change $\geq$ 1.5, calculated by DESeq2 (Methods), to limit the input into Enrichr to genes with the greatest variation between C/S and PTL. We note significant over-representation of processes with importance in peripartum adaptations of the placenta. This includes cytokine, neutrophil chemotaxis, glycosylated ECM proteoglycan, and cyclic AMP-dependent CREB transcription factor activity (Figure 2.5C, red bars). Reasons behind the drastic distinction in transcriptomic signatures between the two modes
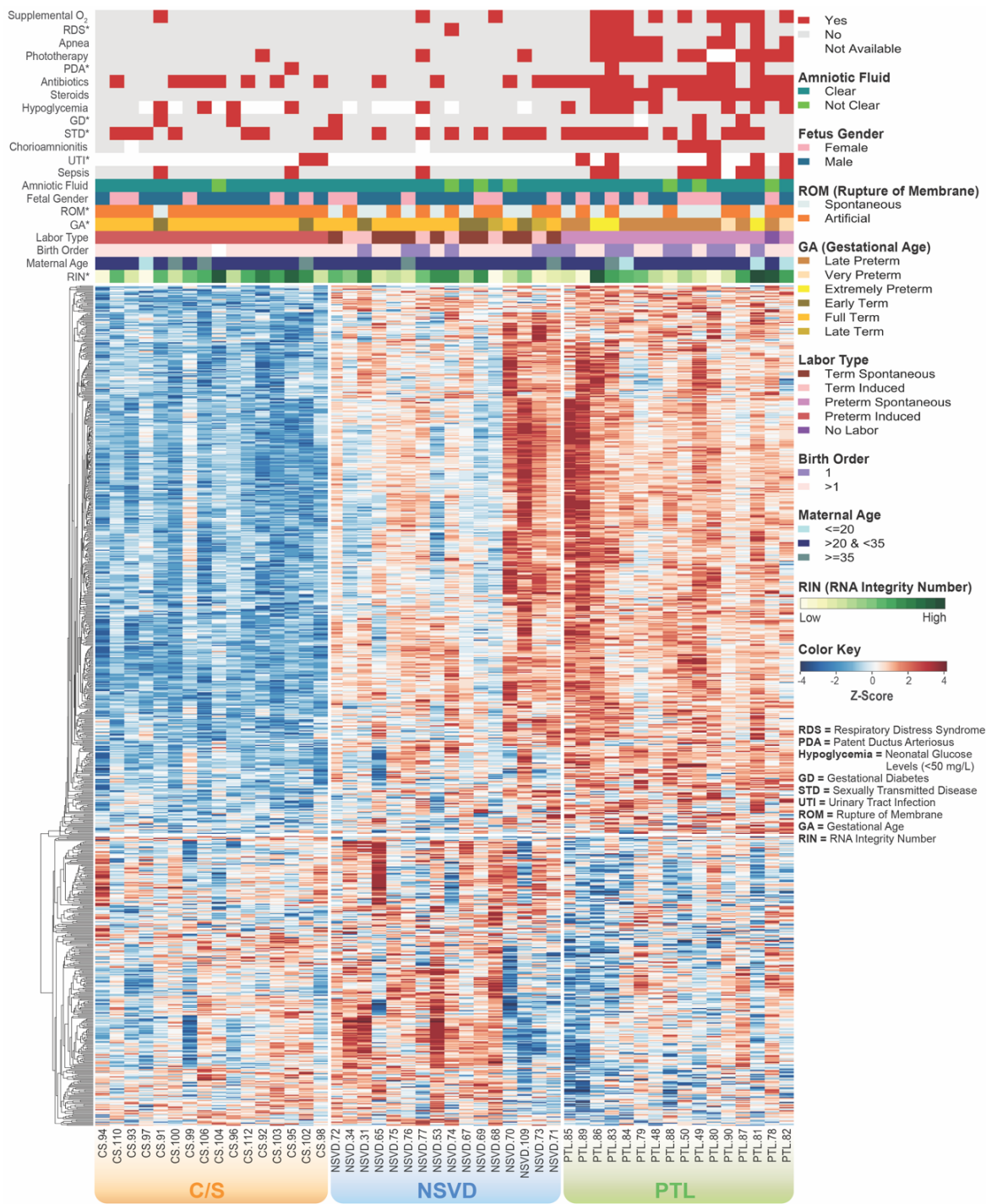
Figure 2.6 | Differential gene expression clustering of parturition groups

Heatmap of 956 genes identified from the union of DEGs from each pairwise comparison (NSVD vs C/S – 33 genes; NSVD vs PTL – 4 genes; C/S vs PTL – 942 genes) using DESeq2 for differential gene expression analysis. Genes are filtered for maximum read count > 10 and adjusted p-value (FDR) ≤ 0.05. Rows represent the z-transformed distribution of gene expression values and are color-coded according to the color key on the right. Columns represent patient samples, which are separately clustered for each condition using average-linkage hierarchical clustering and accordingly ordered in the heatmap. The bottom bars indicate the 3 parturition groups (C/S; NSVD; and PTL). The top bars indicate select maternal and fetal metadata that are color-coded according to the legend on the right.

of abnormal parturition remain unclear. Such observation, however, leads us to believe in the possible involvement of similar molecular factors but with divergent activity in all modes of abnormal parturition.

Using the comprehensive list of DEGs (n=1030), we then perform supervised hierarchical clustering of normalized and log-transformed counts to stratify the 3 parturition groups (Figure 2.6). Overall, we do not observe a strong distinction in DEG signature between NSVD and PTL. We attribute this to within-group variation in clinical characteristics (Figure 2.4A; Appendix Table 2.1) and transcriptomic profiles (Appendix Figure 2.2) as well as to the study's limited power by virtue of its clinical nature. In spite of the general heterogeneity in those factors, documented infections appear to covary with parturition mode (Figure 2.4B; Figure 2.6). The PTL cohort showcases a higher percentage of sepsis (Fisher's exact test; p-value =1.08×10^{-2} relative to NSVD and p-value =4.81×10^{-2} relative to C/S); urinary tract infection (UTI) (Fisher's exact test; p-value = 1.67×10^{-1} relative to NSVD and p-value = 5.84×10^{-2} relative to C/S), and chorioamnionitis (Fisher's exact test; p-value =1.01×10^{-1} relative to NSVD and C/S) (Figure 2.4B). This agrees with our DEG findings, in which PTL exhibits a prevalence of processes related to immune reactions. While a clear segregation of gene expression profiles is not achieved, the C/S cohort appears to be the primary driver of any differences (Figure 2.6; Appendix Figure 2.2). This highlights the dominance of a labor-specific transcriptomic state of the decidua during parturition.

2.4.2 Parturition mode-specific gene co-expression patterns

We next pursue a network-based approach to integrate the coordinated expression differences between parturition types into a higher-order systems level framework. We run weighted gene co-expression network analysis (WGCNA) to identify biologically relevant patterns in the decidual transcriptome and functionally classify groups of genes with unknown connections to parturition. We perform this in two steps (Methods).

First, we construct a separate co-expression network for each parturition condition (Appendix Figure 2.3A-C; Appendix Table 2.3). We use this to assess the preservation of topologically distinct modules across the NSVD, C/S and PTL networks and develop a systems-level understanding of transcriptomic differences. Using complete-linkage hierarchical clustering, we identify module gene clusters within each network and perform functional analysis of their hub genes (Figure 2.7). Modules that are conserved across the 3 parturition conditions are enriched
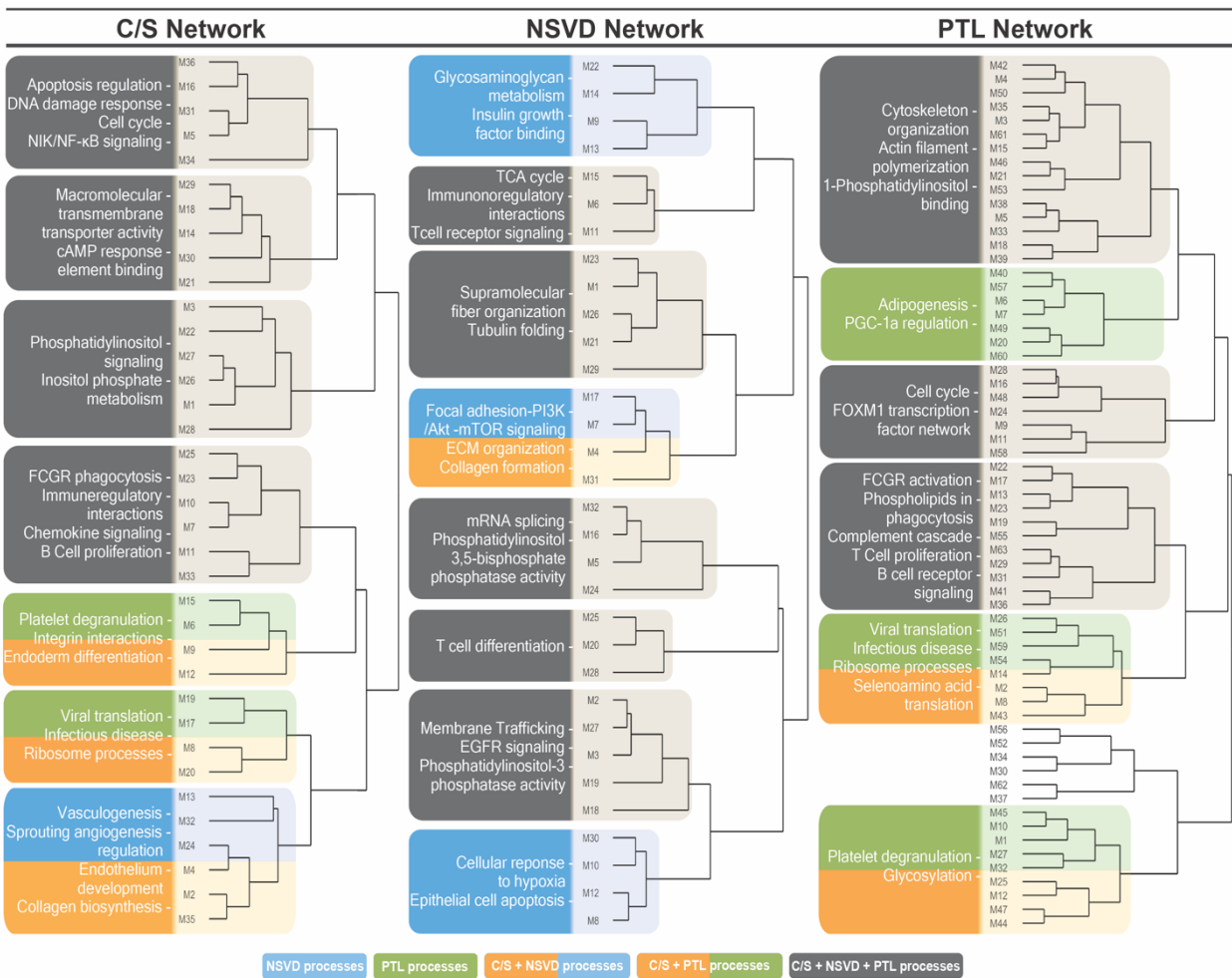
Figure 2.7 | Clustering of NSVD, C/S, and PTL co-expression network modules

Complete-linkage hierarchical clustering of WGCNA co-expression modules for each parturition group, generated using module eigengene measures. Clusters of modules are categorized according to the legend on the bottom. For each cluster, module hub genes (genes with module interconnectedness kME ≥ 0.9) are input into Enrichr to identify significant enrichment for biological processes (FDR < 0.1 and a combined score ≥ 10). Blue boxes reflect processes more common in NSVD (spontaneous term parturition). Green boxes reflect processes more common in PTL (spontaneous preterm labor). Orange and blue boxes reflect processes common in NSVD and C/S (term parturition). Orange and green boxes reflect processes common in C/S and PTL (abnormal parturition). And grey boxes reflect processes common across NSVD, C/S and PTL (general parturition).

for the common cellular processes of splicing, apoptosis evasion, protein transport, and cytoskeletal assembly. They are also involved in reproductive functions including phosphatidylinositol signaling and immunoregulation. We highlight sets of modules that are unique to each network and thus reflect processes specific to the mode of parturition. The NSVD network modules (Figure 2.7, blue boxes) are enriched for glycosaminoglycan metabolism, insulin growth factor binding, and oxidative stress response. These processes are heavily employed during the early phases of pregnancy to mediate placental and fetal tissue growth. They are also involved in regulating cervical softening in normal parturition. Both NSVD and C/S networks exhibit a higher incidence of modules associated with ECM modification, collagen tensile strength maintenance, and vascular development (Figure 2.7, blue and orange boxes). The assembly and distribution of collagen fibril within the decidual endometrial ECM structure of the uterus influences reproductive efficiency. It also dictates the contractile capacity of the myometrium in preparation for labor. The PTL network modules, on the other hand, show a stronger affiliation with lipoprotein metabolism and homeostasis regulation (Figure 2.7, blue borders). Peroxisome proliferator-activated receptor gamma (PPAR-γ) is a key regulator of energy metabolism and adipogenesis. Upon activation within intrauterine tissues, it develops inflammatory properties that impact labor onset in term parturition (Dunn-Albanese *et al*, 2004; St. Louis *et al*, 2016). Distinct from the NSVD network, the C/S and PTL networks share modules with pathogen-associated processes (Figure 2.7, green and boxes). It will be interesting to further explore the role of microbial-induced platelet activation in abnormal parturition.

Next, we perform a hypergeometric test to evaluate our qualitative observations and quantitatively assess the extent of overlap within complimentary modules in the 3 networks. There is a substantial overlap, with moderate to high levels of confidence, between modules that are
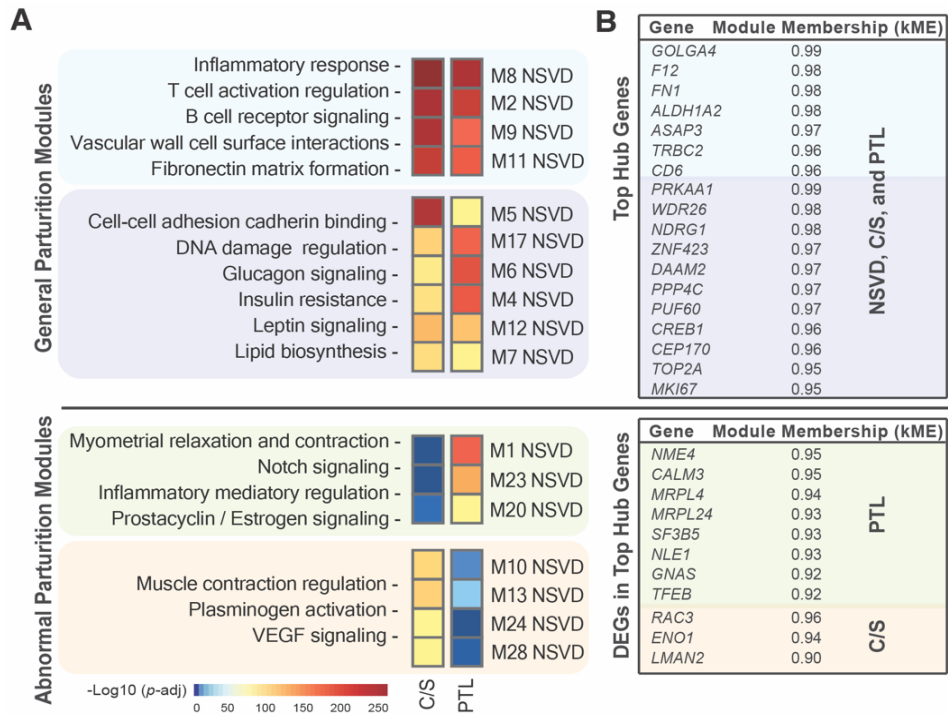
Figure 2.8 | Overlap assessment of NSVD, C/S, and PTL network modules

(A) Heatmap representing the conservation of WGCNA modules across the NSVD, C/S, and PTL co-expression networks. NSVD modules are used as a reference to compare with complimentary modules in C/S and PTL. Modules are designated as complimentary across the 3 networks based on a high percentage of overlap of similar genes. Adjusted p-values denote the hypergeometric probability of overlap between modules, adjusted for multiple comparisons using Bonferroni correction and color-coded per the color key on the bottom left. Modules are labeled as General Parturition Modules upon exhibiting high conservation across the 3 networks. General Parturition Modules are segregated into 2 groups – (1) modules with a very high degree of conservation and (2) modules with a lower degree of conservation. Modules are labeled as Abnormal Parturition upon exhibiting limited to no conservation across the 3 networks. Abnormal Parturition Modules are segregated into 2 groups – (3) modules observed in PTL and NSVD but not in C/S and (4) modules observed in C/S and NSVD but not in PTL. For each group, module genes are input into Enrichr to reveal significant gene ontology (GO) terms (FDR < 0.1 and a combined score ≥ 10). **(B)** List of top module hub genes within General and Abnormal Parturition Modules, along with their degree of module interconnectedness. In General Parturition Modules, we highlight all top hub genes, whereas in Abnormal Parturition Modules, we restrict the list of top hub genes to ones that are also found in our list of DEGs (n=1030).

complimentary across the three networks (Figure 2.8A; Appendix Table 2.4). This suggests that the organization of the human placental transcriptome, from a co-expression purview, is generally preserved across parturition modes. Overlapping modules are comprised of genes involved in core placental functions. The NSVD modules 2, 8, 9, and 11 show the greatest conversation across the 3 conditions and are thus designated as general parturition modules (Figure 2.8A, blue box). They include a total of 657 overlapping genes that are linked to immunity and vascular

development. These reproductive processes are heavily involved in fetomaternal tolerance and support of fetal growth. The NSVD modules 4, 5, 6, 7, 12, and 17 are conserved across the 3 conditions but to a lesser extent (Figure 2.8A, purple box). The shared 87 genes are associated with major maternal endocrine and metabolic adaptions in pregnancy and parturition. Hormones in late stage gestation, such as leptin and glucagon, regulate energy homeostasis by altering glucose and lipid metabolism within the placenta. This ensures continuous ample supply of nutrients to a rapidly growing fetus. We explore top hub genes within the general parturition modules (Figure 2.8B). *FN1* (fibronectin 1) exhibits high module interconnectedness (kME = 0.98) and the highest average expression level across the 3 parturition groups (>200,000 mean normalized read count; Appendix Table 2.5A). The NSVD modules 10, 13, 24, and 28 overlap with the C/S network and are therefore labeled as abnormal parturition modules potentially implicated in labor (Figure 2.8A, orange box). Th e 74 overlapping module genes are involved in plasminogen (Plg) activation and VEGF (vascular endothelial growth factor) signaling, with prominent roles in ECM degradation, vascular remodeling, and inflammation (Mehra *et al*, 2016). Aberrations in these fibrinolytic and pro-inflammatory processes can hinder the rhythmic contractions necessary for labor and that are lacking in C-section deliveries. *ENO1* (enolase 1) is a DEG within our list of top labor module hub genes (Figure 2.8B). It exhibits high module interconnectedness (kME = 0.94) and the highest average expression level in C/S (>2,500 mean normalized C/S read count; Appendix Table 2.5C). *ENO1* encodes for a multifunctional plasminogen-binding protein receptor that can promote ECM degradation through local focalization of plasmin activity and recruitment of inflammatory cells (Plow & Das, 2009). Separately, the NSVD modules 1, 20, and 23 are defined as abnormal parturition modules potentially implicated in term timing (Figure 2.8A, green box). The 411 genes uniquely shared with the PTL network are significantly enriched for immune-modulating estrogen and NOTCH signaling. These processes are crucial to the mediation of macrophage plasticity and polarization within gestational tissue, as we expect to see in prematurity (Jaiswal *et al*, 2015; Pazos *et al*,

2012). We observe a few DEGs within our list of top term module hub genes (Figure 2.8B). This includes *CALM3* (calmodulin 3) that has high module interconnectedness (kME = 0.95) and the second highest average expression level in PTL (>2,000 mean normalized PTL read count; Appendix Table 2.5B).

2.4.3 Global transcriptional architecture of human placenta during parturition

In an attempt to portray the global transcriptional architecture of human placenta during parturition notwithstanding delivery method, we construct a co-expression network using all 48 samples. We stratify 19 discrete gene clusters (Appendix Figure 2.3D; Appendix Table 2.3D) and test their association with patient traits. This allows us to detect gene connectivity patterns that align with clinical phenotypic, demographic, and socio-economic data. We correlate each module eigengene (first principal component, or effectively the average gene expression of a module) with patient information previously shown to either influence placental gene expression or
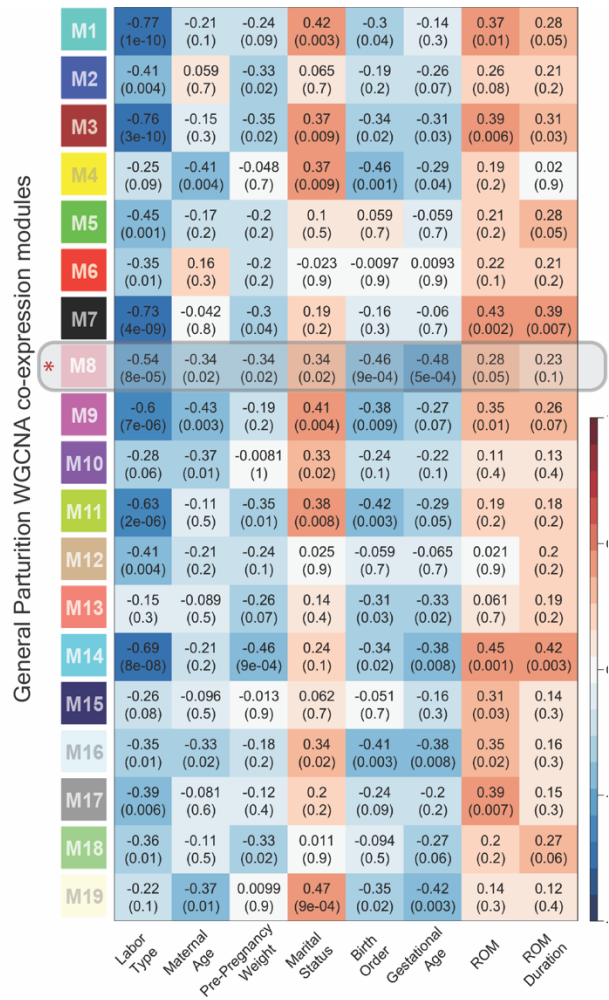
Figure 2.9 | General parturition co-expression module correlation with clinical data

Heatmap illustrating significant correlations (biweight midcorrelation coefficient ≥ 0.45 and asymptotic p-value ≤ 0.01) of general parturition module eigengenes with patient information. Modules are extracted from the WGCNA co-expression network generated using all parturition samples. The x axis shows significantly correlated patient information (n=8). This includes labor type (term; preterm; or no labor); maternal age (continuous); pre-pregnancy weight (in lbs; continuous); marital status (married/engaged or single); birth order (first or higher order pregnancy); gestational age (in wks; continuous); rupture of membrane (ROM – spontaneous or artificial); and ROM duration (in hrs; continuous). The y axis shows all network modules (n=19; color-coded using the default module color scheme of WGCNA). Each cell contains a measure of correlation and asymptotic p-value, color-coded according to the legend on the right. * Module M8 exhibits the greatest number of significant correlations (biweight midcorrelation coefficient ≥ 0.45 and asymptotic p-value ≤0.01).

accompany abnormal parturition ([Appendix Table 2.6](#)). We note significant correlation (biweight midcorrelation coefficient ≥ 0.45 and student asymptotic *p*-value ≤ 0.01) between co-expression modules and the following parturition indicator traits: gestational age (GA) category, as established by the World Health Organization (WHO), and rupture of membrane (ROM) mode (Figure 2.9; [Appendix Figure 2.4](#)). We also observe significant correlation with abnormal parturition risk and prognostic factors such as maternal age category; marital status; birth order; and pre-pregnancy weight (Figure 2.9; Figure 2.10). Amongst our PTL patients, we observe a higher prevalence of young maternal age (≤ 20), single marital status, and low pre-pregnancy maternal weight (Figure 2.10). Our C/S cohort, on the other hand, shows a greater occurrence of advanced maternal age (≥ 35), married/ engaged marital status, and high pre-pregnancy weight (Figure 2.10). Maternal age at the time of delivery and pre-pregnancy weight, at both ends of the spectrum and after adjusting for other confounding variables, have been extensively linked to a higher risk of pregnancy complications

Figure 2.10 | Abnormal parturition risk and prognostic factors

Histograms and bar plot depicting differences in clinical traits often portrayed as prognostic of abnormal parturition. Factors are correlated with co-expression network modules to identify relevant gene clusters.

(Cavazos-Rehg *et al*, 2015; Cnattingius, 1998; Da Silva *et al*, 2003; Hauger *et al*, 2008; McDonald *et al*, 2010). And maternal marital status, through its proposed association with sufficient availability of social support and care services, has been frequently highlighted as a major and indirect influencer of pregnancy outcome (Shah *et al*, 2011).

Module M8 presents the highest number of significant correlations with the aforementioned clinical information, including labor type (biweight midweight correlation coefficient = 0.54; student asymptotic *p*-value = 8.00×10^{-5}; FDR *p*-value $= 1.52 \times 10^{-3}$;Figure 2.9). Taking a closer look into the M8 co-expression module, we note the presence of two distinct sub-module clusters associated with cytoskeletal reorganization and vascular ECM remodeling processes (Figure 2.11). This highlights the co-operative interplay of the two processes, with varying degrees and levels of aberrations in parturition and labor complications. GO analysis of DEGs; hub genes (exhibiting module interconnectedness kME ≥ 0.9; Methods); and placenta-enriched genes in sub-module 1 reveals significant enrichment of cell morphological and physiological processes commonly involved in premature labor (Figure 2.11). This includes fibroblast growth factor (FGF) and phosphatidylinositol signaling and actin filament reorganization. In sub-module 2, we see uteroplacental insufficiency and vascular smooth muscle processes commonly involved in labor complications (Figure 2.11). This includes collagen and fibronectin fibril organization, ECM disassembly, vasoconstriction, and negative regulation of hypoxia-inducible factor 1α signaling. We also note the central role *of LAMC1* within the M8 module (Figure 2.11). *LAMC1* (laminin γ1 chain), a placenta-enriched gene with kME = 0.88, is one of several laminin (LAM) genes that encode for large ECM basement membrane glycoproteins that are extensively localized within the decidua and around newly differentiated blood vessels within the fetomaternal junction in the later stages of pregnancy (Kaloglu & Onarlioglu, 2010). *LAMC1* has been reported as a core placental gene, mediating cell-cell interactions (Armstrong *et al*, 2017). A reduction in *LAMC1* expression has been observed in term non-laboring unripe cervices (Hassan *et al*, 2009). We observe a similar trend in our study, in which the C/S cohort exhibits presents lower *LAMC1* mRNA levels relative to the other groups (one-tailed T-test; *p*-value =5.51×10^{-2}).
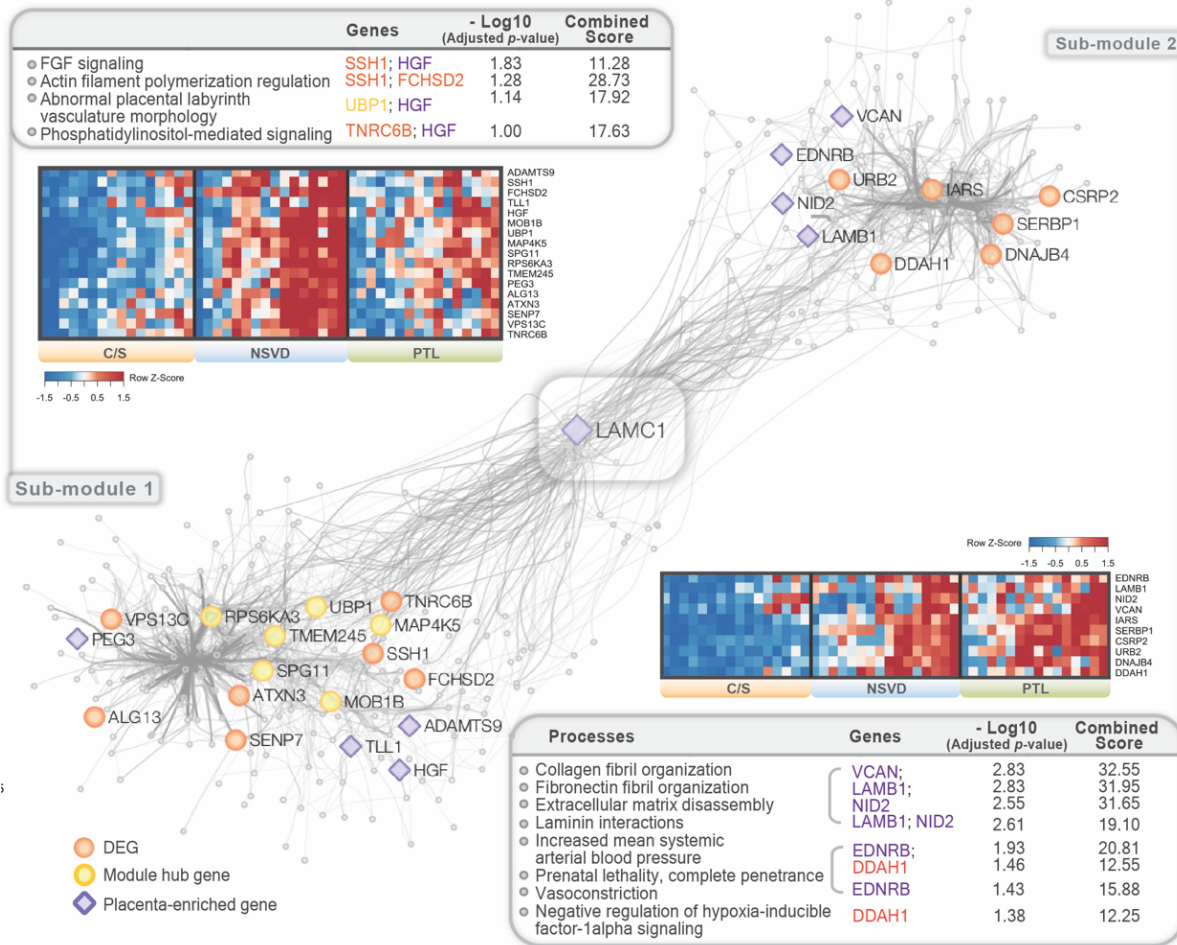
Figure 2.11 | Labor type co-expression module

M8 co-expression module Cytoscape gene-gene interaction plot, highlighting the relationship of DEGs (orange); hub genes (yellow); and placenta-enriched genes (purple) within the two sub-module gene clusters. Enrichr gene ontology (GO) analysis of highlighted genes in sub-module 1 reveals significant enrichment for cytoskeletal and cellular dynamics / motility processes, whereas it shows significant enrichment for vascular and ECM remodeling processes in sub-module 2. Significance is indicated as FDR adjusted p-value < 0.1 and combined score ≥ 10.

Next, we characterize gene co-expression modules that underpin parturition and pregnancy complications using a publicly-accessible curated list of associated genes (Materials and Methods). We observe significant enrichment for labor genes in modules M1 (FDR p-value $= 5.89 \times 10^{-11}$) and M7 (FDR p-value $= 1.41 \times 10^{-6}$; Figure 2.12A). Functional analysis of labor module genes reveals a strong involvement in selenocysteine biosynthesis, nuclear factor NF-κB and phosphatidylinositol 3-kinase (PI3K) activity, and adhesion glycoprotein expression (Figure 2.12B). The aforementioned inflammation and cohesion processes are commonly engaged in
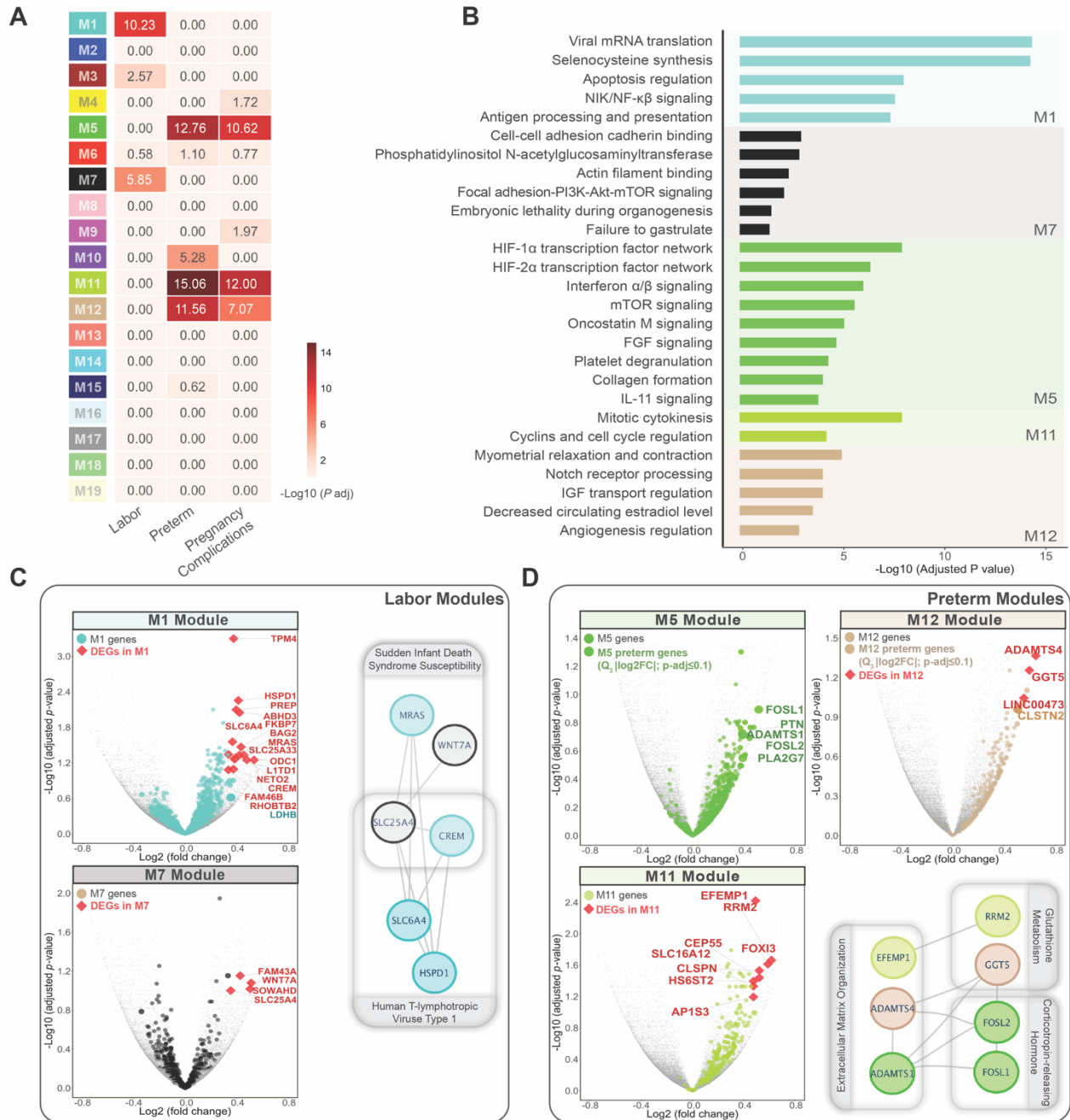
Figure 2.12 | Enrichment for labor, preterm and pregnancy complications genes within global parturition network modules

(A) Hypergeometric test of the over-representation of labor, preterm, and pregnancy complications genes (gathered from publicly-accessible curated lists) within general parturition network modules. P-values are adjusted for multiple comparisons using Benjamini-Hochberg correction. **(B)** Functional annotation of parturition-specific genes within select modules: M1 in turquoise and M7 in black for labor-enriched modules; M5 in green, M11 in green yellow, and M12 in tan for preterm-enriched modules. **(C)** Overview of labor- and preterm-enriched modules. Volcano plots show gene expression changes within each module, emphasizing module genes, DEGs within module, and parturition-specific DEGs within module. Cytoscape interaction diagrams depict the biological function of genes with statistically significant differential expression patterns within each module.

general parturition. Amongst the significantly differentially expressed genes within the labor modules (Figure 2.12C), we pinpoint the implication of *WNT7A* (Wnt family member 7A); *MRAS* (muscle RAS oncogene homolog); *CREM* (cAMP responsive element modulator); and *SLC25A4* (solute carrier family 25 member 4) in Sudden Infant Death Syndrome (SIDS) (Figure 2.12C). SIDS is the leading cause of post-neonatal mortality within the first year of life. Despite its unclear pathophysiology, the predisposition for SIDS correlates with gestational age, exhibiting an increased occurrence in both premature (Malloy, 2013) and post-term (Karagas *et al*, 1993) labor cases. Conversely, modules M5 (FDR *p*-value $= 1.74 \times 10^{-13}$; Figure 2.12A), M11 (FDR *p*-value $= 8.71 \times 10^{-16}$; Figure 2.12A), and M12 (FDR *p*-value $= 2.75 \times 10^{-12}$; Figure 2.12A) are significantly enriched for preterm genes. Module genes have a combinatorial role in immunity, vasculature, placental morphogenesis, maternal-fetal substrate transfer, and hormone secretion (Figure 2.12B). These processes are extensively evaluated in the context of abnormal parturition. Significantly differentially expressed genes and preterm genes within the preterm modules include *ADAMTS1* and *4* (a disintegrin metalloproteinase with thrombospondin type 1 motif 1 and 4) and *EFEMP1* (epidermal growth factor containing fibulin extracellular matrix protein 1). These genes encode for metalloproteases and fibulin glycoproteins that have been shown to interact and potentiate ECM remodeling (Figure 2.12D) (Fontanil *et al*, 2014). We also note the involvement of *RRM2* (ribonucleotide reductase regulatory subunit M2) and *GGT5* (gamma glutamyl transferase 5) in glutathione metabolism (Figure 2.12D). Glutathione metabolism is known to modulate the production of macrophage-induced ROS in the myometrium and subsequently contribute to fetal distress and premature birth (Hadi *et al*, 2015). We finally observe *FOSL1* and *2* (FOS-like antigen 1 and 2) as significantly differentially expressed preterm genes engaged in corticotropin-releasing hormone (CRH) signaling (Figure 2.12D). CRH is a major contributor to myometrial activation and the ensuing heightened sensitivity to prostaglandins that induce the synchronous contractions of labor (Fetalvero *et al*, 2008).

Dysregulation of decidual alternative splicing in abnormal parturition

To further mine the transcriptomic complexity of the placenta, we go beyond analysis at the gene-level and evaluate isoform-level differences as a result of pre-mRNA alternative splicing. We employ the replicate Multivariate Analysis of Transcript Splicing (rMATS) computational pipeline to detect splicing alterations between each pair of parturition mode (Methods). We identify a total of 1334 significant differential alternative splicing (DAS) events in NSVD vs C/S; 1327 events in NSVD vs PTL; and 1056 events in C/S vs PTL covering the major subtypes of AS patterns (Table 2.2). Cassette skipped exons (SE) represent the majority of AS events, accounting for ~70% of all splicing events in each comparison. Mutually exclusive exon (MXE) events dominate in terms of significant hits. We, however, believe this finding to be an overestimation.

Using significant pairwise DAS SE events (Appendix Table 2.7), we are able to achieve a reasonable segregation between parturition types (Figure 2.13A-B). We again note a more distinct AS signature for C/S patients, with a larger proportion of variance captured by the first principal component (67.34% versus 2.97%) (Figure 2.13B). This is in accordance with our DEG results and thus further supports our hypotheses that labor has the greatest impact on the transcriptomic state of the decidua. We explore our list of significant DAS SE events in abnormal parturition (NSVD vs C/S and NSVD vs PTL) and extract

Table 2.2 | Alternative splicing program in NSVD, C/S, and PTL deciduae

rMATS Summary of alternative splicing events that are significantly differentiated between each pair of parturition mode.

NSVD vs C/S

Event Type		# Events (JC only)	Sig Events (JC only)
SE		132,065	426
MXE		32,827	639
A5SS		6,640	69
A3SS		10,211	90
RI		6,902	110

FDR ≤ 0.05 and |ψ| > 0.05

NSVD vs PTL

Event Type		# Events (JC only)	Sig Events (JC only)
SE		131,218	479
MXE		32,153	546
A5SS		6,615	64
A3SS		10,209	123
RI		6,912	115

FDR ≤ 0.05 and |ψ| > 0.05

C/S vs PTL

Event Type		# Events (JC only)	Sig Events (JC only)
SE		107,044	490
MXE		21,494	406
A5SS		6,350	47
A3SS		9,853	77
RI		6,870	36

FDR ≤ 0.05 and |ψ| > 0.05

Abbreviation: JC, junction count; SE, skipped exon; MXE, mutually exclusive exon; A5SS, alternative 5' splice site; A3SS, alternative 3' splice site; RI, retained intron
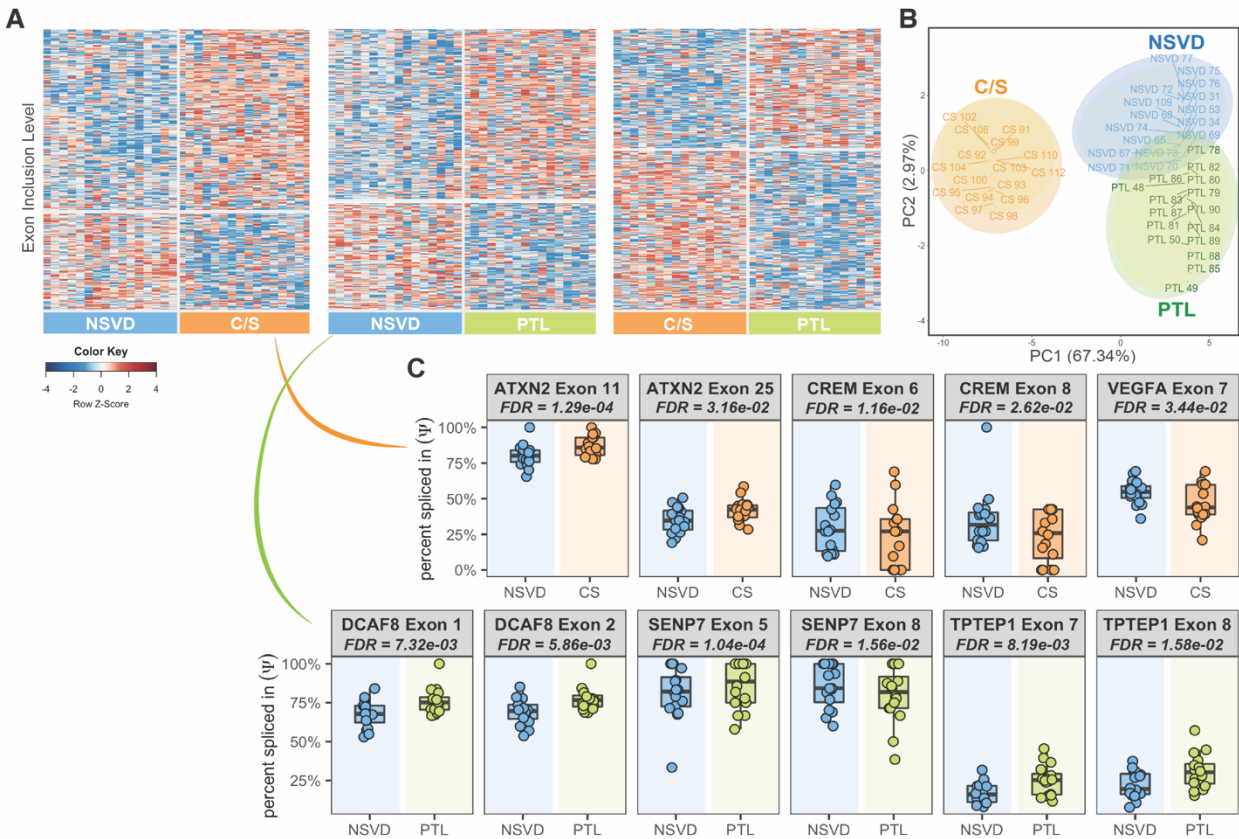
Figure 2.13 | Skipped exon differences between parturition groups

(A) Heatmap showing significant differentially skipped exons (SE)s identified from pairwise DAS analyses using rMATS (NSVD vs C/S – 426 events; NSVD vs PTL – 479 events; C/S vs PTL – 490 events). Significance is determined as $|\bar{\Psi}_{condition1} - \bar{\Psi}_{condition2}| \geq 0.05$ (where Ψ represents exon percent spliced in value) and adjusted p-value (FDR) ≤ 0.05. Rows represent the z-transformed distribution of $\Psi_{exon\_a} - \bar{\Psi}_{exons}$ values for each gene and are color-coded according to the indicated color key. Columns represent decidual samples. **(B)** Principal component analysis (PCA) of NSVD, C/S, and PTL samples using the union of significant SE DAS events. **(C)** Boxplots displaying the distribution of Ψ values for gene variants with multiple significant SE event hits and within the list of DEGs in relation to abnormal parturition. The top panel showcases isoforms in NSVD vs CS while the bottom panel showcases isoforms in NSVD vs PTL.

genes with multiple isoforms and within our list of DEGs (Figure 2.13C). *ATXN2* (ataxin 2), *CREM* (cAMP responsive element modulator), and *VEGFA* (vascular endothelial growth factor A) are seen in NSVD vs C/S. We previously noted *CREM* as a gene that is significantly differentially expressed within labor-enriched co-expression modules (Figure 2.12C) and that, along with angiogenic factor *VEGFA*, is involved in the SIDS pathway. Through alternative splicing, *CREM* yields an array of basic leucine-zipper (bZIP) transcription activator/repressor protein isoforms that regulate the expression of CRE-containing target genes (Sanborn *et al*, 1997). The switch in

CREM isoform expression is believed to modulate myometrial gene expression during gestation, with a prominent reduction observed in the later stages and a potential contribution to the quiescence-to-contractility uterine state shift during parturition (Bailey *et al*, 2000). This matches our observations, in which *CREM* exhibits a reduction in overall gene expression level (one-tailed T-test; *p*-value =1.50×10^{-7}; Figure 2.14) and the inclusion level of two alternatively spliced exons (Figure 2.13C) in the C/S samples. Exclusion of exon 6 within the glutamine-rich trans-activation domain of *CREM* (chr10: 35,477,128-35,477,316) has been shown to render a *CREM* repressor splice variant (Inada *et al*, 1999) that can bind to palindromic CRE motifs (TGACGTCA) of target genes to halt their activation in laboring myometrial tissue (Bailey *et al*, 2005). *CREM* repressor target genes in the human myometrium include *NRP1* (neuropilin-1) and *ID4* (inhibitor of DNA
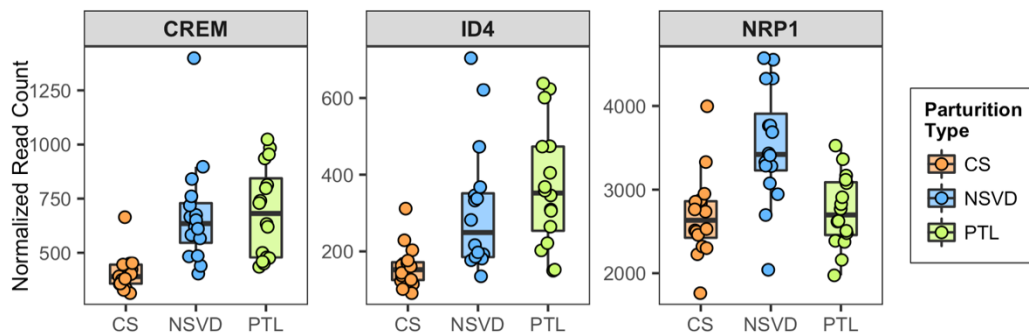
Figure 2.14 | CRE-target gene expression across parturition types

binding 4) (Bailey et al, 2005), two of our DEGs that have reduced expression in C/S placentae (Figure 2.14). NRP1 is a VEGFA receptor that is downregulated in fetal growth restriction and is believed to play a role in malformed fetoplacental vascular branching in abnormal pregnancies (Maulik *et al*, 2015). And *ID 4* has been shown to exhibit differential expression in relation to delivery mode and labor onset (Sõber *et al*, 2015). Within the list of NSVD vs PTL DAS genes, we identify the following DEGs with multiple isoforms: *DCAF8* (DDB1 and CUL4 associated factor 8), *SENP7* (SUMO1/sentrin specific peptidase 7), and *TPTEP1* (transmembrane phosphatase with tensin homology pseudogene 1). *DCAF8* encodes a WD40 repeat-containing substrate

receptor that binds to damage-specific DNA binding protein 1 adaptor to form the DDB1-Cullin4-based E3 (CUL4) ubiquitin ligase complex (Sang *et al*, 2015). The ligase complex is involved in regulating DNA damage checkpoint response and protein turnover, with deficient CUL4 and DDB1 levels linked to abnormal placental development and embryonic lethality (Jiang *et al*, 2012). While neither identified DAS DCAF8 isoforms involve amino acid deletions within the DDB1-binding WD40 motifs, its cellular function within the ligase complex has been shown to be extensively subverted by pathogenic viruses to degrade host anti-viral response factors and allow undisrupted viral replication (Li *et al*, 2010). DCAF variants with mutations outside WD repeats have reduced interaction with DDB1 and can be competed off by viral protein motifs (Li *et al*, 2010). Viral infection of the placenta has been previously demonstrated to stimulate a fetal inflammatory response, sensitize the mother host to bacterial infection, and ultimately promote preterm labor (Cardenas *et al*, 2010).

We next assess the overlap of genes corresponding to significant DAS SE events from each pairwise analysis (Figure 2.15A). We identify 82 intersecting genes in NSVD vs C/S and C/S vs PTL and label them as labor-specific DAS genes (Figure 2.15A). We investigate those containing alternatively spliced exons with the highest absolute percent spliced in values (PSI; Ψ). *IMMP1L* and *AGBL5* yield isoforms with significantly upregulated exon inclusion levels ($\Delta\Psi$) in C/S relative to NSVD (FDR = 4.47×10^{-2} and FDR = 1.04×10^{-2} respectively) (Figure 2.15B-C). *IMMP1L* (inner mitochondrial membrane peptidase 1-like) encodes for catalytic subunit 1 of the inner membrane peptidase (IMP) complex that is responsible for cleaving signal peptide sequences after reaching the inner mitochondrial membrane. *AGBL5* (ATP/GTP binding protein like 5) encodes for a matrix metalloproteinase (MMP) enzyme with increased proteolytic cleavage activity in the presence of pro-inflammatory cytokines in the uterine cervix during parturition (Dubicke *et al*, 2010; Wang & Stjernholm, 2007). Within the 84 genes that overlap between NSVD vs PTL and C/S vs PTL – designated as term-specific DAS genes (Figure 2.15A,) we observe

significantly upregulated exon inclusion levels for *ST5* in PTL (FDR = 6.61×10^{-4}) (Figure 2.15B-

C). *ST5* (suppression of tumorigenicity 5) encodes for a protein regulator of MAPK1/ERK2

signaling, whose activity in human uterine cervix has been shown to covary with gestational period

and labor onset (Ruzycky, 1998; Wang & Stjernholm, 2007). We finally note 25 genes that

intersect across the three pairwise DAS analyses (Figure 2.15A). Amongst those genes are

PLA2G6 and *EEF1D*, with a total of 3 isoforms exhibiting significantly downregulated exon

inclusion levels in PTL (Figure 2.15B-C). *PLA2G6* (phospholipase A2 group VI) encodes for a
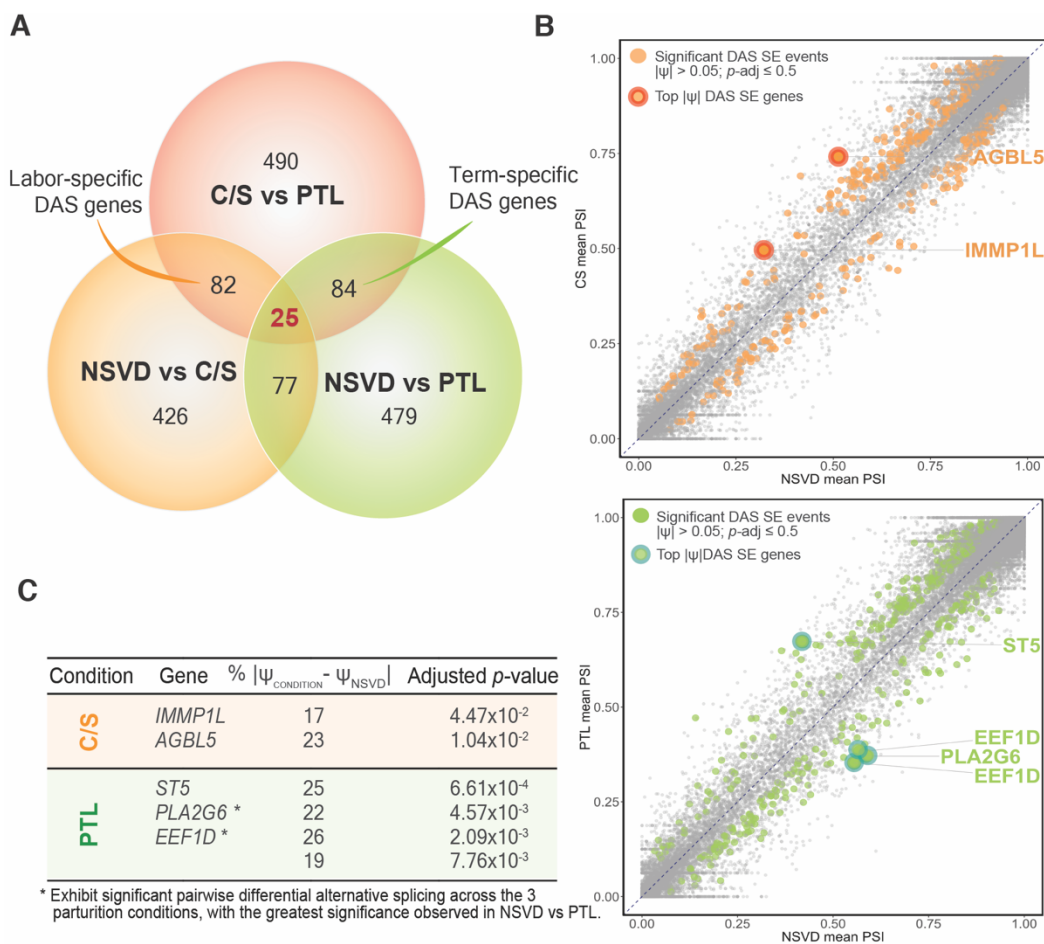
Figure 2.15 | Genes of differentially skipped exons in abnormal parturition

(A) Venn diagram displaying the overlap of DAS SE genes from each pairwise analysis. Intersecting genes in NSVD vs PTL and C/S PTL are labeled term-specific genes. Intersecting genes in NSVD vs C/S and C/S vs PTL are labeled labor-specific genes. (B) Scatter plots showing skipped exons with significantly different exon inclusion levels between NSVD and C/S (top plot) and NSVD and PTL (bottom plot). Significance is determined as $|\bar{\Psi}_{condition1} - \bar{\Psi}_{condition2}| \geq 0.05$ and FDR ≤ 0.05. Genes corresponding to exons with the highest absolute expression fold change are highlighted. (C) List of genes associated with top significant DAS SE events in C/S and PTL along with their percentage exon inclusion relative to NSVD and FDR adjusted p-values.

PLA2 enzyme involved in glycerophospholipid homeostasis and prostaglandin synthesis (Gilbert & Harmon, 2004; Sun *et al*, 2010). PLA2 enzyme levels in human placenta have been linked to premature rupture of membranes (PROM) (Lappas *et al*, 2001) and fetal rejection due to perturbations in the pro-/anti-inflammatory mediators ratio at the fetomaternal interface (Mosher *et al*, 2014). *EEF1D* (eukaryotic translation elongation factor 1 delta) encodes for a subunit of the eEF1 protein complex. A known translational factor member in protein biosynthesis, the eEF1 complex has acquired various non-canonical and moonlight functions. This includes pro-inflammatory signal transduction, viral replication, response to oxidative stress, and cytoskeletal remodeling (Gross & Kinzy, 2005; Li *et al*, 2013; Sasikumar *et al*, 2012; Schulz *et al*, 2014) – processes we previously saw as associated with PTL.

2.5   DISCUSSION

Disruption in the seemingly Sisyphean task of pregnancy, from early through late stages of gestation, can result in a wide range of life-threatening complications that impact the fetus and the mother. Such disruption is believed to be brought on by intrinsic genetic and extrinsic environmental stimuli or stressors that interact in a dynamic, temporally-varying, and many-to-many correspondence manner. The complex gene-environment interplay in pregnancy has long hindered scientists' ability to isolate precise trigger points for pregnancy complications. It is thus imperative to pursue novel methodologies to unveil the underlying etiology and ultimately address a major unmet medical need for better prognostic and diagnostic markers. We perform a comprehensive and multilayered analysis of RNA sequencing and clinical data from 3 parturition groups to explore the underpinnings of late-stage pregnancy. In doing so, we capture the biological detail transcribed within the transcriptomic state of maternal uterine tissue and use it to convey the narrative for general and abnormal parturition.
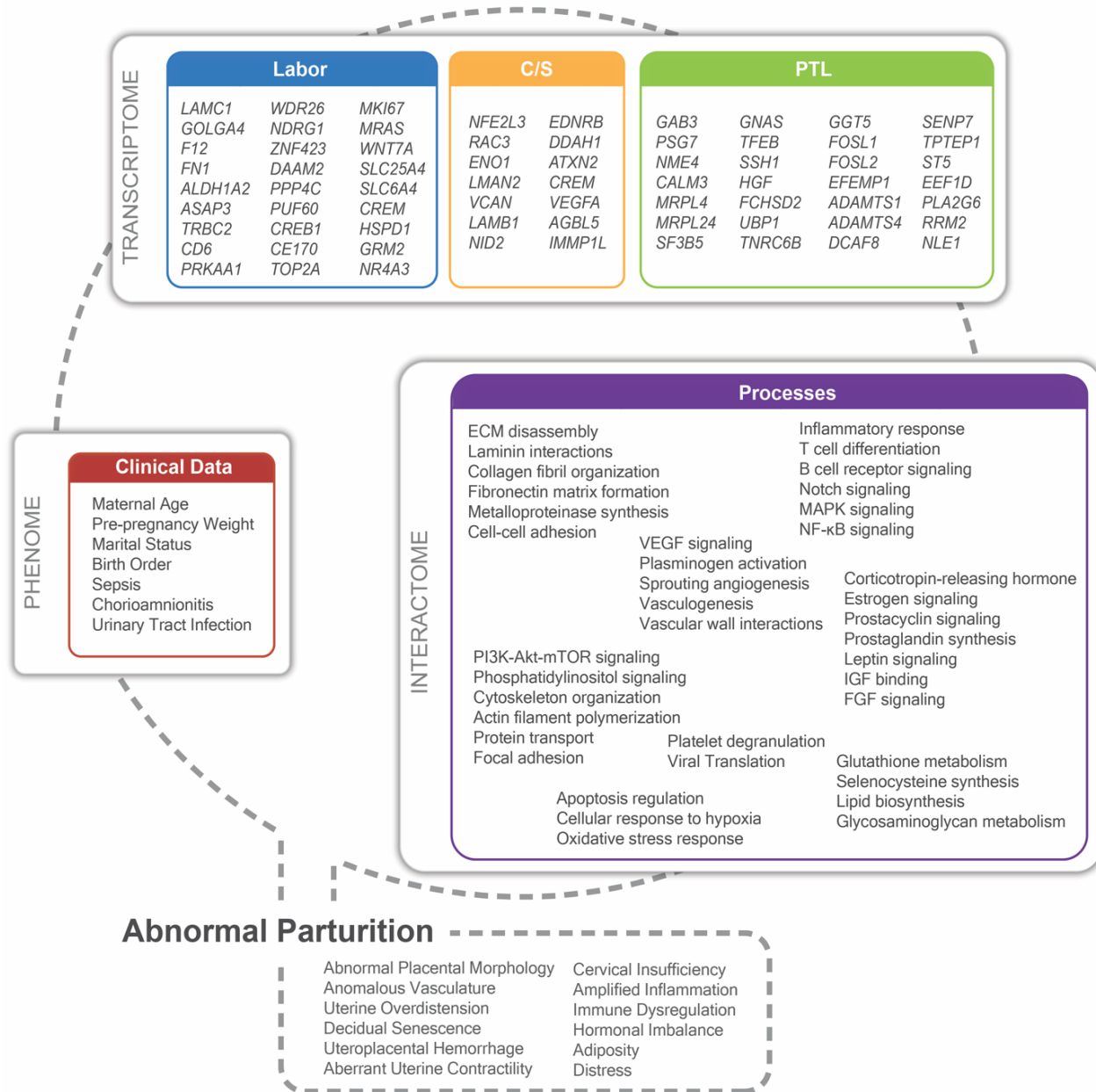
Figure 2.16 | Summary of Findings

Summary of findings from transcriptomic, interactomic, and phenomic analyses of intrapartum decidual specimens. We highlight potential key player genes, molecular processes, and clinical traits that frequently appear to associate labor, C-section, and preterm labor.

2.5.1 A proposed molecular depiction of labor processes in the decidua

Our findings lead us to recognize the decidua, with its unique immunomodulatory and cancer-like properties, as an important mediator of the rapid physiological changes invoked prior to labor.

Functional analysis of its transcriptomic landscape reveals the composite inflammatory nature of parturition in its various forms. While the exact mechanisms that trigger parturition remain unclear, we are able to systemically integrate results from individual analyses to paint a rough picture of involved factors. We posit that the growing gestation demands in late-stage pregnancy activate the NOTCH transmembrane receptor to reprogram mitochondrial metabolism (Eun & Jeong, 2016; Xu *et al*, 2015). The increased respiration rate and oxidative burst stimulates master transcription factor nuclear factor NF-κB activity, stabilizes the transcription factor hypoxia-inducible factor (HIF), and promotes a transient and parturition-specific hypoxic episode within the uterine myometrium. Hypoxic stress, in turn, increases the production of chemokines to attract leukocytes and induce a higher degree of pro-inflammatory M1-polarized phenotype expression (Xu *et al*, 2015). Leukocytes successively infiltrate the inflammatory cell-enriched cervix and release proteases to disrupt its rigid extracellular matrix and initiate cervical effacement. The pro-inflammatory and neuroendocrine feedback loop concomitantly stimulates pro-labor uterotonic signaling and prostaglandin synthesis in intra-uterine tissues. This process is modulated by selenoprotein levels and the phosphatidylinositol turnover-mediated release of arachidonic acid from membrane glycerophospholipids (Folkert *et al*, 1984; Huang *et al*, 2012; Walsh, 2011). The surge in vasoactive lipid autacoids then elicits a biphasic response, manifested as an alternating constriction and dilation rhythm, in the endometrial endothelial vascular smooth muscles. The contractile signal propagates through gap junctions within the cytoskeletal framework of myometrial cells and culminates in the strong phasic uterine contractions of labor. Cervical dilation advances as a result of the highly excitable uterine state and facilitates transportation of the fetus down the cervical canal. The amniochorionic fetal membranes continue to weaken and ultimately rupture to commence birth. We pinpoint several humoral, adhesion, vascular, endocrine, and cytoskeletal gene factors that we believe can cooperatively converge to instigate the coordinated series of labor events (Figure 2.16; blue box). Amongst those are *FN1* and *LAMC1,* which are highly expressed in human placental tissue (Figure 2.17). *FN1* and *LAMC1* encode major

structural glycoproteins that characteristically reside at the chorionic-decidual interface. They actively participate in extracellular matrix remodeling of the cervix in preparation for labor (Goossens *et al*, 2009; Hassan *et al*, 2009).
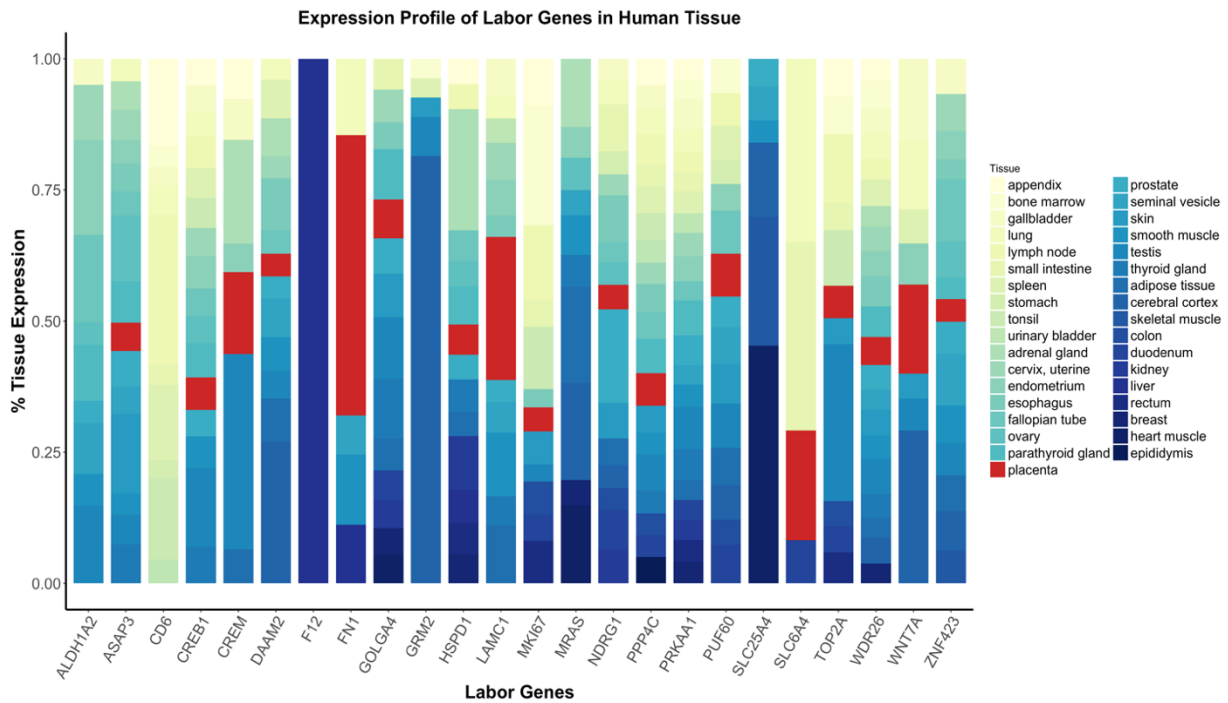
Figure 2.17 | Expression profile of labor genes in human tissue

Tissue-specific enrichment for labor genes that were identified using the integration of analyses in the study. RNA levels in 37 human tissue samples were downloaded from the Human Protein Atlas (www.proteinatlas.org). % Tissue Expression represents tissue-specific gene expression level divided by total gene expression level across all tissues (in TMP - Transcripts per Million values). In red we denote placenta-specific enrichment.

2.5.2 A proposed molecular depiction of C-section processes in the decidua

The higher number of altered gene and isoform expression hits we observe in our C/S population underlines the extensive molecular and structural modifications associated with labor. Cesarean deliveries are performed as a result of labor-specific obstetric complications that include delayed, prolonged, dystotic, or absent labor. We typically foresee the identification of a large selection of pathophysiologic factors that can contribute to C/S from various gestational periods. By analyzing only non-laboring C/S placentae, we are able to isolate those specific to halting proper parturition. In addition to the previously conveyed processes of parturition, we notice significantly higher

enrichment for eccentric vascular ECM mechanics, angiogenesis, and smooth muscle contraction in C/S. In agreement with previously published studies (Bogaerts *et al*, 2013; Ecker *et al*, 2001), C/S patients also exhibit a more advanced age, greater pre-pregnancy weights, and higher parity. We extrapolate that older age, overweight, and altered reproductive tract morphology impair the utero-placental hemodynamic system of parturition. Reduction in vascular permeability and thickening of the amniotic basement membrane can be conducive to prostaglandin insensitivity, delays in cervical activation, inefficient uterine contractile function, and the cessation of fetal membrane rupture. This manifests in labor difficulties including an increase in gestation duration and inadequate preparation for parturition. Overweight pregnant women are also more likely to deliver macrocosmic newborns that necessitate C-section deliveries. Furthermore, diminished plasminogen activity within the vascular smooth muscle cells of certain C/S placentae can influence the ability of MMPs to degrade its extracellular matrix at term. *NFE2L3, EDRNB,* and *NID2* show the highest enrichment in placental tissue (Figure 2.18) within the list of gene factors we present as associated with C/S (Figure 2.16; orange box). *NFE2L3* (Nuclear Factor Erythroid 2 Like 3) has been characterized as the "Cinderella" of Cap'n'Collar (CNC) transcription factors due to its under-studied and elusive functionality (Chevillard & Blank, 2011). Recent studies have revealed its involvement in placental vascularization and fetal growth during the late stages of pregnancy (Kashif *et al*, 2012). *EDNRB* (Endothelial Receptor Type B) encodes a G protein-coupled receptor protein that binds endothelin-1 (ET-1) within vascular smooth muscles. In addition to its vasomodulatory properties, ET-1 has been shown to influence the peripartum endocrinological cascade and moderate the detachment of fetal membranes (GRAM *et al*, 2017; Takagi *et al*, 2008). *NID2* (Nidogen 2) encodes a peptide within vascular and epithelial basement membranes of decidual tissue. By binding nidogen, a known MMP substrate, it can modulate ECM composition and limit its degradability at term (Strauss, 2013).

50

Expression Profile of CS Genes in Human Tissue
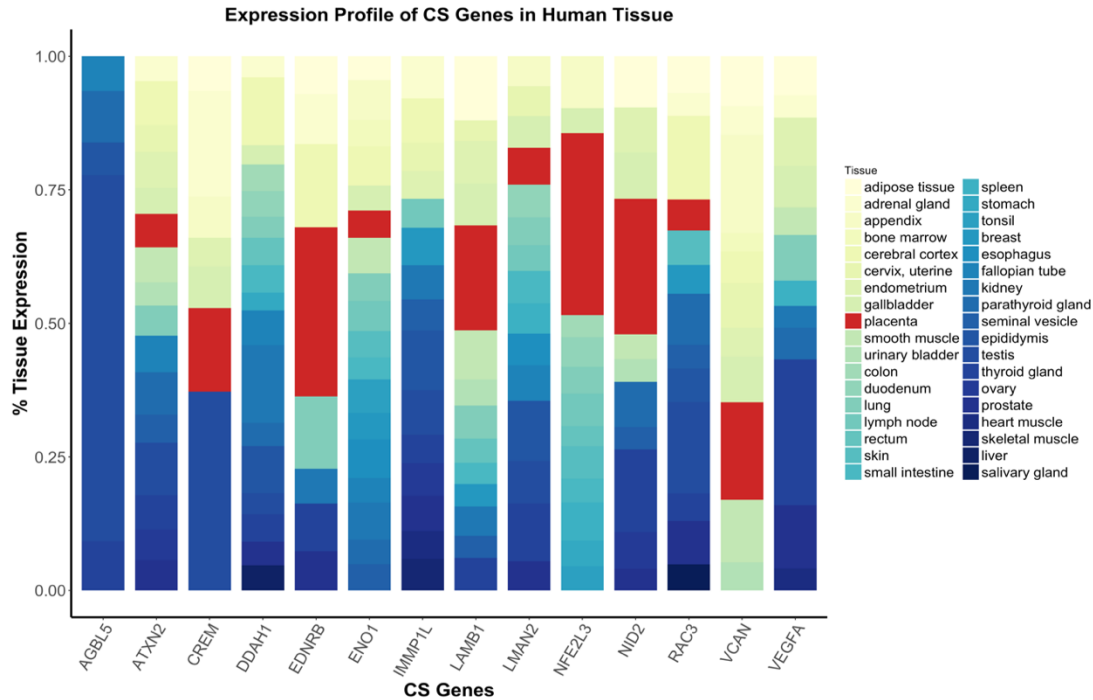
Figure 2.18 | Expression profile of C/S genes in human tissue

Tissue-specific enrichment for C/S genes that were identified using the integration of analyses in the study. RNA levels in 37 human tissue samples were downloaded from the Human Protein Atlas (www.proteinatlas.org). % Tissue Expression represents tissue-specific gene expression level divided by total gene expression level across all tissues (in TMP - Transcripts per Million values). In red we denote placenta-specific enrichment.

2.5.3 A proposed molecular depiction of PTL processes in the decidua

Preterm labor, on the other hand, is a term-specific obstetric complication prompted by disruption in the scheduling of parturition. While exhibiting diverse pathophysiology, it is generally associated with instability in fetal microenvironment, from an immunologic and sufficiency purview, and defective timing regulation for uterine quiescence-to-contractility state transition. The transcriptomic profile of PTL appears to be less distinctive, as it incorporates components from across parturition modes. We do, however, note a higher level of immunologic response to infectious agents and an imbalance in the attack/tolerate homeostatic state of pregnancy in PTL. This matches our clinical findings, in which infections, such as sepsis, UTI, and chorioamnionitis, are more common in the PTL cohort. We propose that sterile and microbial-induced inflammation

contribute to prematurity in the following manner: Physical trauma, including systemic and intra-amniotic infection and decidual hemorrhaging and senescence, as well as psychological and psychosocial stress provide stimulus for the premature release of CRH – of varying maternal, fetal, hypothalamic, and placental origins. Levels of circulating pro-inflammatory mediators, such as leukocytes and MMPs, amplify in response. This disturbs the finely-tuned local inflammatory environment and elicits a shift towards a M1 inflammatory predominant state. Both hemostatic coagulation and complementary systems get activated to generate thrombin at a higher rate and recruit activated platelets to further modulate immune response. The overstimulation of prostaglandin and estrogen production and MMP activity then consecutively prompts an intrapartum switch in myometrial contractility. It also effectuates an increase in collagen solubility, a disruption in rigid cervical matrix, and softening of fetal membranes. Additionally, we detect significantly lower maternal pre-pregnancy weights, higher rates of neonatal hypoglycemia (Figure 2.19), and enriched lipid metabolic processes in PTL. This makes us speculate that underweight mothers are more prone to nutritional deficiencies that can impact fetal and placental development as well as limit adequate placental-fetal exchange. Nutritional stress, due to unmet metabolic needs by the growing fetus, results in increased CRH and glucocorticoid production and glucose shortage. The PI3K/AkT/mTOR signaling network is then engaged to modulate glucose metabolism by influencing forkhead (FOXO1) and peroxisome proliferator-activated

Figure 2.19 | Neonatal hypoglycemia records across parturition types

receptor (PPAR) transcription factor levels in the placenta. PPARs successively enhance the production of cytokines, creating a heightened state of inflammation and prompting early labor. Alternatively, a caloric deficit can increase the mother's susceptibility to infections through

glucocorticoid hindrance of immune response. We identify *PSG7, HGF,* and *EFEMP1* as PTL genes (Figure 2.16; green box) with greatest placental expression (Figure 2.20) *PSG7* (Pregnancy Specific Beta-1-Glycoprotein 7) encodes for one of the most abundant trophoblastic proteins in maternal serum (Lin *et al*, 1974). With a demonstrated role in immunomodulation, pregnancy-specific glycoproteins (PSGs) regulate cytokine production and immune activation phenotypes within the decidual microenvironment (Martinez *et al*, 2013; Snyder *et al*, 2001). High PSG concentration levels are typically detected near term (Blois *et al*, 2014), whereas low PSG concentration levels have been reported in pregnancy complications and abortion (Grudzinskas *et al*, 1983; Tamsen, 1984). *HGF* (Hepatocyte Growth Factor) encodes for a pluripotent cytokine with established roles in placental development (Somerset *et al*, 1997), placental glucose uptake (Visiedo *et al*, 2015), and placental inflammation. It has also been linked to inflammation-mediated uterine distensibility and premature fetal membrane rupture in preterm labor (Waldorf *et al*, 2015).

Figure 2.20 | Expression profile of PTL genes in human tissue

Tissue-specific enrichment for PTL genes that were identified using the integration of analyses in the study. RNA levels in 37 human tissue samples were downloaded from the Human Protein Atlas (www.proteinatlas.org). % Tissue Expression represents tissue-specific gene expression level divided by total gene expression level across all tissues (in TMP - Transcripts per Million values). In red we denote placenta-specific enrichment.

Lastly, *EFEMP1* (EGF Containing Fibulin Extracellular Matrix Protein 1) encodes an ECM glycoprotein that strongly binds laminins, fibronectins, and nidogens and modulates MMP activity (Timpl *et al*, 2003). It is highly expressed at the site of spontaneous rupture of membranes (Nhan-Chang *et al*, 2010).

2.5.4 Study limitations

Our study faces the following limitations that can reduce the reproducibility and translatability of reported findings:

2.5.4.1 Abnormal parturition heterogeneity

The identification of a limited number of DEGs in PTL highlights its inherent heterogeneous nature (Figure 2.2). PTL has several subtypes including: **(a) medically indicated PTL**, brought on by maternal and fetal complications; **(b) PTL associated with preterm premature rupture of membranes (PPROM)**, typically due to infection; and **(c) spontaneously PTL**, occurring without evident risk factors (Moutquin, 2003). PTL is therefore a condition of multiple etiologies and worthy of further sub-classification in future assessments of larger sample sizes. Prior classification of C/S cases into elective or emergency C/S groups in future studies is also necessary Figure 2.1). C/S can be performed either as a result of abnormally slow or failed progress of labor (referred to as dystocia) or pre-/co-existing life-threatening complications and thus involving varying pathophysiology.

2.5.4.2 Confounding factors

Patients within each parturition group exhibit a wide range of characteristics, raising the concern of confounding factors that can diminish or over-exaggerate our findings. For example, PTL placental samples are collected over a span of 3 years from 3 centers. While all cases were filtered for non-singleton births, preeclampsia, and gestational diabetes, many factors remain variable such as gravidity and parity, infection, pre-pregnancy maternal weight, and prenatal steroid treatment. The majority of women that deliver prematurely in our study (13/16) are

administered glucocorticoids (GCs) to promote fetal maturation (Figure 2.4). Synthetic GCs can significantly impact the gene expression make-up of the placenta (Ozmen *et al*, 2017; Robinson *et al*, 1988) and therefore need to be controlled for in forthcoming studies. Additionally, while an ideal model to study maternal-fetal interface anomalies, the decidua is a mélange of maternal stromal cells, glandular epithelial and immune cells, and fetal trophoblast cells (Fu & Wei, 2016). It is therefore difficult to extract clean maternal decidual cells and isolate biological events specific to the mother. We note minor fetal contamination in our dataset by observing 21 genes that align to the Y chromosome ([Appendix Figure 2.5](#)). Samples collected from patients with male fetuses exhibit much higher expression levels for those genes relative to their female counterparts thus confirming the presence of fetal transcriptome.

2.5.4.3 Study design weaknesses

An averaged total gene expression overview of the placenta conceals the underlying biological heterogeneity. Bulk RNA-seq analysis limits our ability to assess the extrinsic and intrinsic variations and interactions between cell types. This is especially of crucial interest when studying a microenvironment heavily enriched with and regulated by inflammatory cells. And finally, the switch from pregnancy to labor can be viewed as a dynamic system of highly interactive feedback loops – both maternal and fetal – with a tissue-specific and time-dependent nature. Research remains unclear on whether abnormal parturition reflects late-stage aberrations or is rather rooted in early placental development defects. In order to extract the point of disruption within such tightly controlled system, we need to assess the joint role of the mother and the fetus as well as evaluate chronological changes throughout the course of gestation. This can be achieved through longitudinal and comprehensive analysis of both fetal and maternal membranes, which will enable the design of therapeutic agents that can provide precise, targeted dosages at the appropriate time window.

2.6 CONCLUSION

Through a high-dimensional examination of RNA sequencing and clinical data, we provide a comprehensive description of the transcriptomic landscape of human decidua under term, preterm, and non-laboring conditions. By integrating differential expression and alternative splicing, co-expression, and functional enrichment analyses findings, we show that labor has the most significant influence on the underlying molecular composition. We associate the C/S form of abnormal parturition with compromised utero-placental ECM organization and hemodynamics, advanced maternal age, and higher pre-pregnancy weight. On the other hand, we postulate an integral function for CRH – the "placental clock" – in birth timing and stress-induced humoral response in the PTL form of abnormal parturition. We also uncover the association of infection, younger maternal age, and lower pre-pregnancy weight with PTL. We generate a list of transcriptomic, phenomic, and interactomic factors (Figure 2.16) that we believe have potential contribution to parturition-specific complications, paving the way for future experimental validation and predictive biomarker identification.

2.7 FUTURE DIRECTIONS

Our findings offer a strong indication of factors involved in various modes of parturition. The goal is to harness the knowledge acquired about mechanisms in late-stage pregnancy, moving past the exploratory stage to ultimately extract determinants of a successful pregnancy from earlier stages. I intend to further pursue this work as a postdoctoral fellow, with the goal of depicting the "central dogma" of pregnancy. I will do so by participating in efforts to **(a)** create a comprehensive database of longitudinal pregnancy-associated data from whole genome sequences, epigenomes, proteomes, metabolomes, microbiomes, clinical tests, medical imaging, and patient traits; **(b)** use the high dimensional landscape of pregnancy at its various stages to identify points of perturbation, and **(c)** ultimately identify biomarkers that improve on the sensitivity of predictors

currently used in the clinical assessment of life-threatening pregnancy complications such as preterm labor.

CHAPTER III

Isoform-Level Profiling of Glioblastoma Radiophenotypes

3.1 ABSTRACT

Glioblastoma (GBM), used to refer to Grade-IV astrocytoma, is a malignant type of solid brain tumors that continues to exhibit a high degree of resistance to treatment despite the abundance of newly discovered molecular targets and pathways. This is notably the result of GBM tumor's heterogeneous molecular structure as well as the undefined involvement of its microenvironment. The push for precision and personalized management in medicine has fostered highly integrative efforts in GBM assessment, with a heavy focus on incorporating a wide range of imaging- and molecular-based profiling techniques to improve clinical phenotypic stratification. Using a "radiotranscritpomic" framework, we evaluate the relationship between glioma-intrinsic exon inclusion levels and quantitative, morphological MRI image features for 38 GBM patients from The Cancer Genome Atlas (TCGA) and The Cancer Imaging Archive (TCIA). Functional enrichment analysis of significant exon correlates reveals the overrepresentation of motility-enhancing signaling cascades, extracellular matrix remodeling, and ion channel activity, highlighting the proliferative and invasive property of GBM tumors. Identified exon genes exhibit a high prognostic power, yielding significant associations with radiographic edema. We posit that an invasion-specific "radiotranscriptomic" profile of GBM tumors serves as a robust prognostic tool that can be used to derive multimodal treatment options that can specifically target motile, invading tumor cells.

3.2   INTRODUCTION

3.2.1 The heterogeneous landscape of glioblastoma

Glioblastoma (GBM), used to refer to Grade-IV astrocytoma, is the most common, aggressive, and malignant form of primary brain tumors in adults. It can either directly manifest as a Grade-IV tumor –referred to as primary GBM – or evolve from lower grade gliomas –referred to as secondary GBM (Ozawa *et al*, 2014). Upon suspicion of glioblastoma (GBM), patients undergo a regimen of gadolinium-enhanced magnetic resonance imaging (MRI) and histopathologic tissue analysis for a definitive diagnosis (Itakura *et al*, 2015). From an imaging perspective, GBM tumors lack a characteristic regional manifestation. They exhibit lesions with varying sizes, contrast enhancing margins, necrotic masses, edema, neoplastic cell infiltration, hemorrhaging, and microvasculature (Jain *et al*, 2014; Hammoud, 1996). Histological evaluations have also revealed fluctuations in the proportion of GBM tumor anatomic structures with increased cellularity, pseudopalisading necrosis, nuclear pleomorphism, vascular endothelial proliferation, and mitotic figures (Lawton, 2009). The eclectic mixture of cell types, with genetically distinct clonal populations and varying levels of cellular and nuclear polymorphisms, is believed to contribute to a high degree of GBM intratumoral heterogeneity (Louis *et al*, 2016). This has spurred numerous efforts to decipher the molecular underpinnings of GBM to develop personalized treatment strategies that better accommodate the variability between patients.

3.2.2 Existing imaging- and molecular-based efforts to decode GBM tumor heterogeneity

3.2.2.1 GBM radiomic features

Several studies have explored image-based biomarkers of distinct GBM tumor phenotypes, with the goal of non-invasively stratifying patients into subgroups for targeted therapy (Itakura *et al*, 2015). Image-based features to characterize GBM tumors are currently

extracted from a wide range of imaging modalities –such as MRI, CT, and PET – and sequences –such as T1-weighted (T1W), T2-weighted (T2W), fluid attenuation inversion recovery (FLAIR), and diffusion weighted imaging (DWI).

In an attempt to ensure uniformity between radiographic measurements, the cancer research community has generated the Visually AcceSAble Rembrandt Images (VASARI) MRI feature set schema (https://wiki.nci.nih.gov/display/ CIP/VASARI) for neuroradiologists to use in their assessments. The well-defined and standardized lexicon, comprising thirty semantic features, has been widely adopted by domain experts to provide a comprehensive depiction of features associated with GBM (Gutman *et al*, 2013). It consists of a combination of quantitative/continuous, categorical (ordinal), and qualitative grading systems for feature rating. VASARI features have provided significant radiophenotypic information about GBM tumors and have been extensively investigated in relation to survival and tumor molecular profile (Gutman *et al*, 2013). With the increased interest in further elucidating GBM tumor characteristics, researchers have also independently derived computational features. These include intensity, morphologic, volumetric, histogram-based and texture analysis-derived features using 2D/3D tumor regions (Narang *et al*, 2016). Others have also collected diffusion properties and spatial information using models of glioma growth (Bakas *et al*, 2017). PET imaging has also been used to derive semantic and computational features due to its ability to differentiate residual GBM tumors from scar tissue and edema (Narang *et al*, 2016). MRI, however, continues to be the preferred modality for radiomic evaluations of GBM due to its high sensitivity and specificity. The development of Diffusion Weighted Imaging and Perfusion MRI analysis methods now offers the potential to depict tumor phenotypic information at the tissue-level including tumor cell density and invasion (Padhani *et al*, 2009).

3.2.2.2 GBM molecular profiling

Molecular profiling-based classification methods have also been extensively explored to elucidate the heterogeneous molecular structure of GBM tumors and risk stratify GBM patients. In a study led by The Cancer Genome Atlas (TCGA), researchers identified four distinct, clinically-relevant, and non-transitioning GBM patient subtypes: **(1)** classical (CL); **(2)** mesenchymal (MES); **(3)** proneural (PN); and **(4)** neural (NL) (Verhaak *et al*, 2010). This was achieved through validated differential analysis of transcript abundances of 600 TCGA GBM tumors. They were able to demonstrate that distinct gene expression profiles between GBM subgroups with varying somatic mutations, DNA copy numbers, treatment response, and survival (Verhaak *et al*, 2010). In the process, they generated a list of 840 predictive and subtype-specific genes that has been a standard in GBM molecular classification (https://tcga-data.nci.nih.gov/docs/publications/gbm_exp/). In an attempt to segregate the glioma-intrinsic transcriptomic landscape from that of tumor microenvironment, another study leveraged single cell RNA sequencing to identify genes uniquely expressed by tumor cells (Wang *et al*, 2017). They generated a list of 150 subtype-predictive genes and demonstrate a level of overlap with Verhaak *et al.*'s list of MES, PN, and CL genes. Such efforts have provided a tremendous amount of knowledge regarding the molecular landscape of GBM.

3.2.3 Radio-'omic' analyses of GBM tumors

Matching the molecular landscape of GBM tumors, at the microscopic level, with their radiographic manifestation, at the macroscopic level, can provide a more comprehensive overview of the underlying heterogeneity. Radiological imaging can provide a detailed representation of GBM tumor phenotype, while molecular profiling can better isolate sources of its heterogeneous infrastructure. Radiogenomic (alternatively image genomic or "radiotranscriptomic") efforts are actively pursued in GBM research, with the goal of integrating quantitative radiomic features with biomolecular information to identify more robust prognostic

and predictive tumorigenesis biomarkers, stratify patients, and inform clinical decisions (Jaffe, 2012).

3.2.4 Why use a multi-layered transcriptome-radiome framework in GBM?

Patients with GBM face a poor prognosis, with a median survival rate of less than 2 years (Ohgaki & Kleihues, 2005). GBM heterogeneity has limited clinician's ability to predict treatment response and recurrence following the standard course of surgical re-sectioning, fractionated radiotherapy, and chemotherapy (Martinez *et al*, 2010). A robust classification of GBM tumors is therefore necessary to improve our understanding of primary brain tumor etiology and identify determinants of GBM subtypes for tailored therapies.

Molecular profiling of GBM, on its own, has failed to decode intratumoral heterogeneity. This is due to its inability to differentiate between tumor, tumor microenvironment, disease progression, and treatment contributions and its bias towards effects specific to the tumor region from which a sample is collected. Furthermore, molecular investigations of biopsied tumor samples can only be performed following surgical resectioning. There is, therefore, a higher likelihood of falsely capturing a pro-inflammatory signature that is separate from GBM. And image-based biomarkers, by themselves, are unable to

Researchers have previously evaluated the relationship between GBM tumor copy number variation, DNA methylation, and gene expression profiles and MRI image traits. We further expand on such findings by incorporating glioma-intrinsic alternative splicing information. The goal is to use a "radiotranscriptomic" approach to identify isoform-level molecular correlates of morphologic images features of GBM tumors. By integrating complementary radiophenotypic, transcriptomic, and clinical information, we hope to provide insight into the heterogeneous landscape of GBM tumors and GBM-associated gene isoforms and biological processes.

3.2.5 Study design

In this study, we use a multi-layered transcriptome-radiome framework to investigate the relationship between the transcriptome (exon-level molecular information) and the radiome (pre-operative MRI quantitative anatomical tumor radiophenotypes) (Figure 3.1). We posit that the phenotypic diversity of GBM tumors that is captured via radiographic imaging also reflects the underlying molecular heterogeneity.
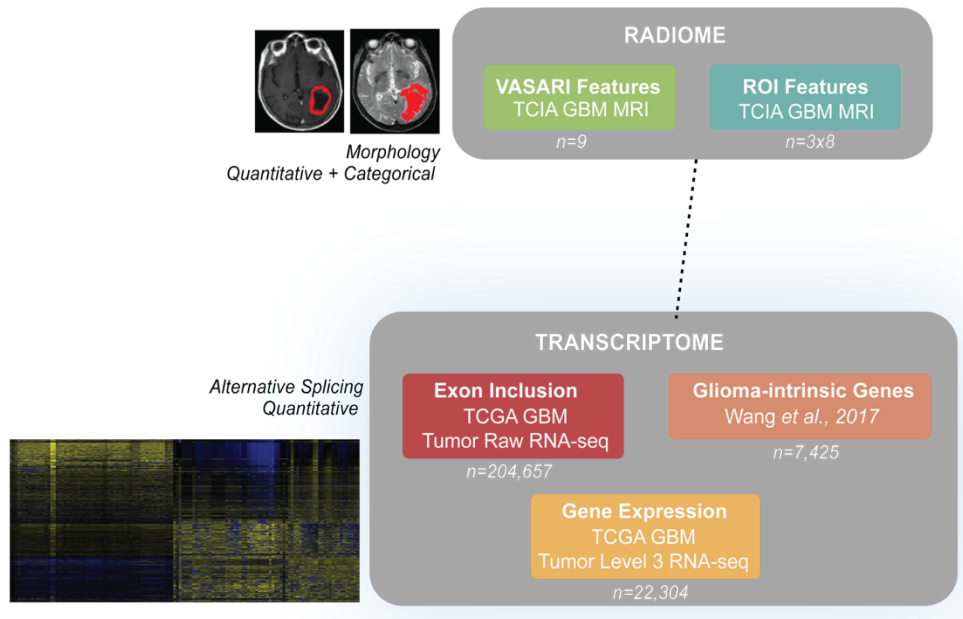
Figure 3.1 | An integrative transcriptome-radiome assessment of GBM

3.3   METHODS

3.3.1   Selection of GBM MRI images

GBM MRI images from The Cancer Imaging Archive (TCIA) are selected using the following criteria: **(a)** includes pre-operative baseline images; **(b)** includes pre- and post-contrast T1-weighted and T2-weighted fluid-attenuated inversion recovery (FLAIR) MR images; and **(c)** has a timestamp that relatively matches that of RNA sample acquisition. Image features are extracted in 2 parts: **(1)** supervised manner, using Visually AcceSAble Rembrandt Images (VASARI) feature

annotations, and **(2)** semi-supervised manner, using Region of Interest (ROI) segmentation and morphologic features.

3.3.2 VASARI feature extraction

Image features characterizing the tumor and its microenvironment are manually extracted in Osirix using the VASARI MR feature set schema (https://wiki.nci.nih.gov/display/CIP/VASARI). The following subset of prognostic VASARI features are generated: **(1)** F1 - Tumor Location; **(2)** F9 - Multifocal or Multicentric; **(3)** F5 - Proportion Enhancing; **(4)** F6 - Proportion Non-Enhancing; **(5)** F7 - Proportion Necrosis; **(6)** F10 - T1/FLAIR ratio; **(7)** F14 - Proportion Edema; **(8)** F29 - Largest Cross-Sectional Length; and **(9)** F30 - Perpendicular Length to the Largest Cross-Sectional Length. The latter two features are quantitative. Extracted VASARI features can be found in Appendix Table 3.1.

3.3.3 Quantitative ROI feature extraction

A semi-supervised segmentation pipeline is used to regions of interest (ROIs). Images are first pre-processed for correction and smoothing, image re-slicing, and cleaning). Three ROI subtypes are then extracted: **(1)** Contrast Enhancement; **(2)** Edema (FLAIR); **(3)** Total (Enhancement + Necrosis) (Figure 3.2). Eight morphological features that capture tumor intensity and shape are used to describe the ROI subtypes:

- **Surface Area (A)** = Surface Area of the Tumor
- **Volume (V)** = # of image voxels
- **Compactness 1** = $V/\sqrt{\pi}A^{2/3}$
- **Compactness 2** = $36\pi V^2/A^3$

- **Maximum Diameter** = Largest distance between 2 image voxels
- **Spherical Disproportion** = $A/4\pi R^2$
- **Sphericity** = $\pi^{\frac{1}{3}}(6V)^{2/3}/A$
- **Surface to Volume Ratio** = A/V

These features are calculated using the three-dimensional size and shape of tumor region in the following manner (Aerts *et al,* 2014). For patients with multiple tumors, ROI features are filtered

to represent only primary and contiguous tumors with the largest volume. Extracted ROI features can be found in Appendix Table 3.2.
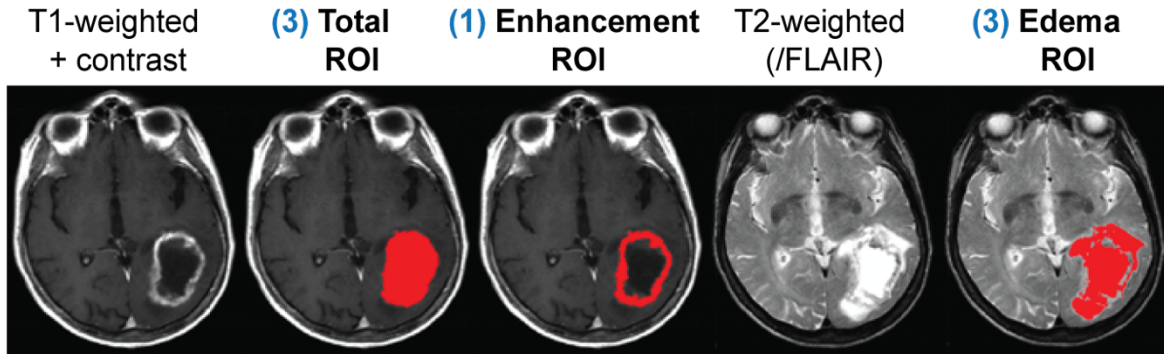
Figure 3.2 | ROI GBM tumor segments

3.3.4 RNA-seq data download and pre-processing

Processed TCGA Level 3 RSEM normalized RNA-seq count data is downloaded from The Cancer Genome Atlas (TCGA) for 38 patients with available MRI imaging data from TCIA. The aggregated gene expression dataset, which includes 20,531 genes, is then filtered for coefficient of variation > 1.0, to remove genes with low distribution dispersion, and max > 1 TPM, to remove genes with a low read count. Raw RNA sequences are also downloaded for further downstream analysis. Reads are filtered for adapter sequences and mapped against the Ensembl human genome (hg19; GRCh37.75) using the software STAR v2.4.1c (Dobin *et al*, 2013) with default parameters.

3.3.5 Exon inclusion calculation and filtering

We use an in-house statistical tool, replicate multivariate analysis of transcript splicing (rMATs; http://rnaseq-mats.sourceforge.net – version turbo) to generate percent spliced in (Ψ ; psi) estimates of gene exon inclusion. Exon inclusion estimates are calculated as $\widehat{\Psi} = \frac{I/l_I}{I/l_I + S/l_S}$, in which I represents read count of inclusion isoforms connecting upstream splice junction, alternative exon, and downstream splice junction; S represents read count of skipping isoforms

connecting upstream and downstream splice junctions; and l_I and l_S represent the effective lengths of inclusion and skipping isoforms, respectively. Using the raw RNA sequences, we generate a list of 204,657 exons. To focus our analysis on tumor-intrinsic heterogeneity, we filter the list of exons to include those within the 7,425 genes identified by Wang *et al.* as uniquely expressed by glioma cells instead of tumor-associated host cells (Wang *et al*, 2017). For the 27.4% of exons missing inclusion values, ones with >50% missing values are discarded and the remaining exons are given median-imputed values. Exon inclusion levels are then filtered for coefficient of variation > 1.0 and corresponding gene expression values > 1 TPM.

3.3.6 Exon inclusion-image feature correlation

Exon inclusion levels are correlated with quantitative VASARI and ROI image features (n = 26) using median-based biweight midcorrelation. Permutation resampling with 1000 permutations is then applied to adjust *p*-values for multiple testing.

3.3.7 Functional enrichment and gene-gene interaction analysis

We use Enrichr (Kuleshov *et al*, 2016), at default settings, to identify functional annotations of identified genes. Significant gene ontology (GO) terms and pathways are extracted using a Benjamini-Hochberg adjusted *p*-value < 0.1 and a combined score ≥ 10. GeneMANIA (Warde-Farley *et al*, 2010) is also used to generate a functionally-relevant biological network of identified genes. GeneMANIA is run with maximum resultant attributes = 10 and an automatically-selected weighting method for co-expression; co-localization; genetic interactions; pathway; physical interactions; predicted; and shared protein domains networks.

3.4   RESULTS

3.4.1 Quantitative MRI image feature characterization of GBM

Using TCIA MRI images and the VASARI lexicon and ROI metrics, we generate a variety of quantitative features to comprehensively characterize GBM tumor morphology (Methods; Appendix Table 3.1; Appendix Table 3.2; Table 3.1). For the ROI features, we obtain good interrater agreement amongst 4 imaging experts –Contrast Enhancement (median $\kappa = 0.82$); Edema (median $\kappa = 0.88$); and Total (median $\kappa = 0.89$). We produce a total of 26 quantitative image features for 38 patients.

Table 3.1 | VASARI and ROI feature description and reason for selection

Tumor Location (categorical / nominal metric)	Lesion geographic epicenter location A full resectioning of a GBM tumor is extremely difficult due to its highly infiltrative nature. Tumor location therefore greatly affects the outcome of the surgery and onset of recurrence
Multifocal or Multicentric (categorical / nominal metric)	**Whether or not the tumor is multifocal (with at least one region that is not contiguous with the dominant lesion as a result of dissemination, spread via commissural pathways, or local metastasis but with clear communication) or multicentric (with lesions in separate lobes or hemispheres and with no communication)** The presence of more than one focus is less common than single focal tumors in GBM patients, but has been implicated in even worse prognosis due to the spreading and involvement of many key structures in the brain (Lasocki *et al*, 2016)
Proportion Enhancing (categorical / ordinal metric)	**The proportion of the tumor that exhibits a high concentration of contrast enhancement, producing significantly higher signal on post-contrast compared to pre-contrast T1-weighted images** To assess active lesions with a clear breakdown of the blood-brain barrier
Proportion Non-Enhancing (categorical / ordinal metric)	**Region of intermediate hyper-intensity associated with mass effect and architectural distortion** To assess active lesions with a clear blood-brain barrier breakdown

Proportion Necrosis (categorical / ordinal metric)	**Proportion of the tumor that exhibits a high percentage of pre-mature, autolytic tumor cell death, often as a result of growth beyond vasculature supply and overconsumption** Greatly implicated in poor prognosis
T1/FLAIR (categorical / nominal metric)	**Size of pre-contrast T1 abnormality in relation to the size of FLAIR abnormality** Used to assess whether tumor is expansive, infiltrative, or both
Proportion Edema (categorical / ordinal metric)	**Proportion of the tumor that exhibits the presence of fluid within the brain extracellular space** A result of tumor-induced capillary endothelial tight junctions and has pseudopods extending up to the subcortical white matter
Largest Cross-Sectional Length and the Length Perpendicular	**Lesion size major and minor axes** Used to characterize tumor longitudinal and transversal lengths and calculate total volume
Contrast Enhancement (ROI)	**Represented as a ring pattern in an MRI image, is used the infer the extent of microvasculature and blood brain barrier disruption.** Complementary to the VASARI Proportion Enhancement metric.
Edema (ROI)	**Can arise either as a result of pivotal molecular changes or as a response mechanism to tumor development, leading to both peritumoral and intratumoral abnormal accumulation of water** Associated with a disrupted blood-tumor barrier and excess tissue spacing and thus swelling. It has also been speculated to have the biggest impact on patient cognition, which renders it an interesting target for studies of cancer-associated (or cancer-induced) cognitive decline (Lin, 2013) This ROI feature complementary to the VASARI Proportion Edema metric.
Total (ROI)	**Measures both enhancement and necrosis** Can be used to assess the extent of the necrotic core of a GBM tumor

3.4.2 Exon inclusion – quantitative image feature correlation

We proceed to identify exon inclusion-based transcriptomic signatures for GBM tumor radiophenotypes. Using rMATS, we calculate exon inclusion levels (n = 204,657) from raw RNA sequences collected from the corresponding 38 TCGA patient tumor samples. We further filter the list of exons to include only those that are glioma-intrinsic, highly-expressed, and highly-variable (Methods). This ensures that the transcriptomic features reflect information specific to the tumor and not its microenvironment, thus safely reducing our high-dimensional feature space. Using biweight midcorrelation, we then correlate filtered exon inclusion estimates (n = 6599) with quantitative VASARI and ROI MRI image features (n = 26). We apply a permutation test to obtain 1000 realizations of the null permutation distribution of correlation coefficients. We then extract exons that are associated with tumor radiophenotypes in general. For each exon, significant correlations are determined via a minimum permutation-adjusted p-value < 0.05 across the 26 quantitative image features. Significant Fisher z-transformed correlations for 93 exons, arranged into 3 clusters via unsupervised complete hierarchical clustering, are shown in Figure 3.3A. Enrichr analysis of exon cluster genes reveals significant enrichment for neuronal growth, axonal guidance, extracellular matrix (ECM) organization, focal adhesion, and channel activity – processes that exhibit an association with tumor morphology (Figure 3.3B; Methods).

In order to identify exon-level signatures for each ROI metric, we extract exons that instead significantly correlate (permutation-adjusted p-value < 0.1) with at least 6 out of the 8 morphologic features used to characterize Enhancement, Total, and Edema ROIs and both quantitative VASARI features. We identify 5 exon genes associated with the Enhancement ROI that are enriched for high voltage-gated calcium channel activity and cytoskeletal reorganization; 3 exon genes associated with the Total ROI that are enriched for PDGF (platelet-derived growth factor) signaling; and 11 exon genes associated with VASARI tumor length that are enriched for extracellular matrix (ECM) organization and proteoglycans in cancer (Methods). We observe
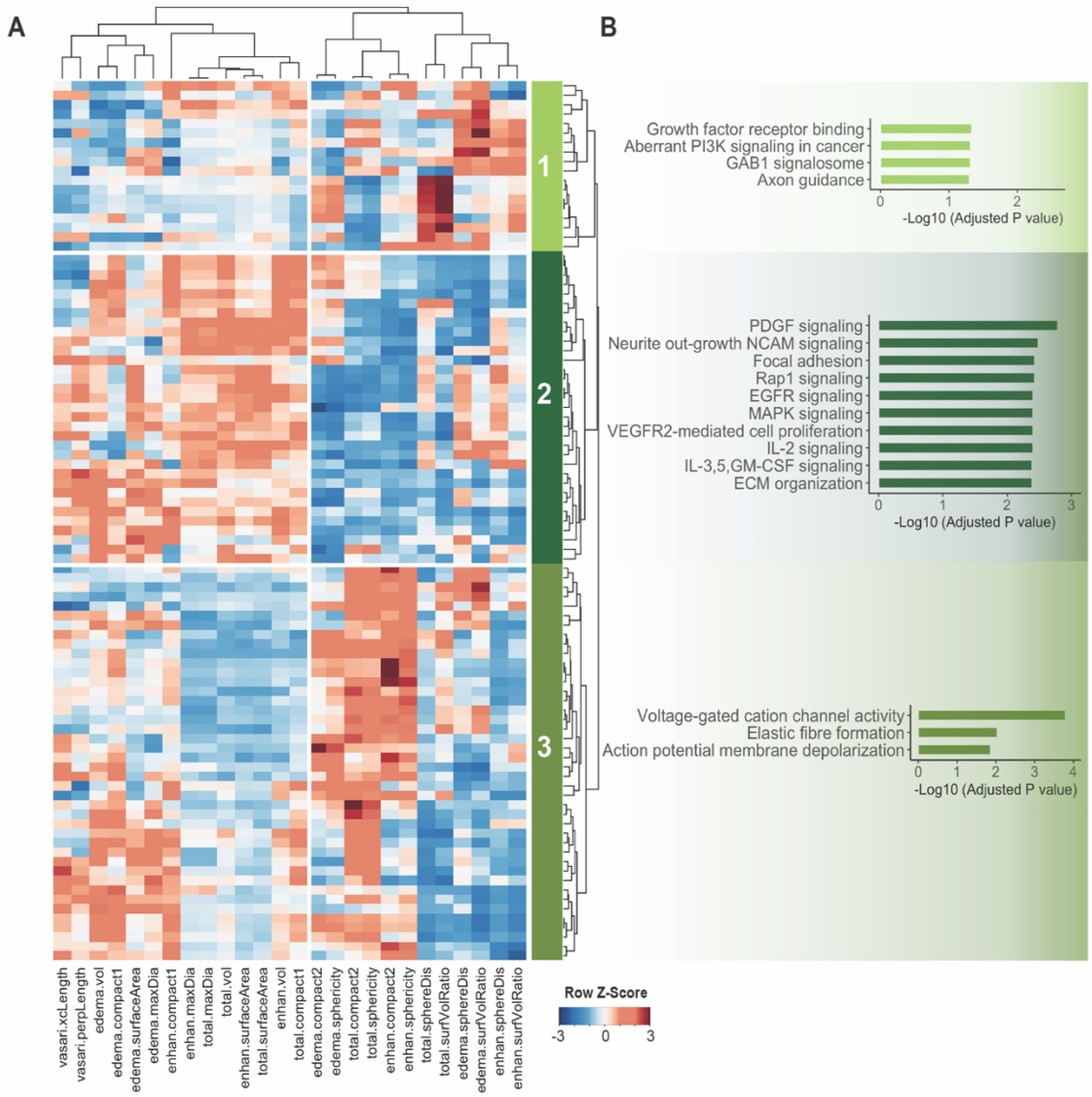
Figure 3.3 | Significant exon inclusion-image feature correlations

(A) Unsupervised hierarchical clustering of 93 significant exon inclusion – quantitative image feature correlations, as determined by a row minimum permutation-adjusted *p*-value < 0.05. Rows represent Fisher's z-transformed biweight midcorrelation coefficients that are color coded per key on the bottom right. Columns represent quantitative VASARI and ROI image features. Bar on the right indicates exon clusters.
(B) Significantly enriched functions for exon clusters, identified using FDR < 0.1 and a combined score ≥ 10.

CACNA1A (calcium voltage-gated channel subunit alpha 1A) in both Enhancement and Total ROIs. We note that the Total ROI also measures contrast enhancement and therefore an overlap in correlated exon genes is expected.

70

	Gene Name	SurfaceArea	Vol	Compact1	Compact2	MaxDia	SphereDis	Sphericity	SurfVolRatio
ENHAN	TAC3	0.01	0.01	0.16	0.01	<0.009	<0.009	<0.009	0.50
	SLC26A10	<0.009	<0.009	0.29	0.04	<0.009	0.07	0.08	0.47
	AGBL2	0.07	0.44	0.30	<0.009	<0.009	<0.009	0.02	0.06
	CACNA1A	<0.009	0.02	0.71	<0.009	0.01	<0.009	<0.009	0.13
	THBS4	0.02	0.02	0.27	0.04	0.03	0.06	0.06	0.36
	Gene Name	SurfaceArea	Vol	Compact1	Compact2	MaxDia	SphereDis	Sphericity	SurfVolRatio
TOTAL	FRS2	0.11	0.02	<0.009	0.04	0.19	0.05	0.04	0.05
	CACNA1A	<0.009	<0.009	0.04	0.03	0.01	0.05	0.05	0.42
	THBS4	0.02	0.05	0.1	0.02	0.06	0.01	0.03	0.61

	Gene Name	Largest Cross Sectional Length	Perpendicular Length
VASARI	FN1	0.09	0.07
	COLEC11	0.04	<0.009
	CLGN	0.07	0.08
	FXYD3	<0.009	0.02
	FGFR4	0.05	0.02
	FRS2	<0.009	0.04
	FRS2	0.01	0.04
	MFAP5	0.01	0.03
	RPH3A	<0.009	0.09
	GALNT14	<0.009	0.07
	KCNG1	0.09	0.01

Figure 3.4 | Significant exon-radiophenotype correlations

Cells indicate permutation-adjusted *p*-values from biweight midcorrelation of all exons with image features. Highlighted cells represent significant values.

We input those genes into GeneMANIA to generate a molecular functional interaction network involving associated genes that are predicted through documented physical interactions and co-expression (Methods; Figure 3.5). We highlight the overlapping contribution of genes in ECM organization and calcium channel activity.

3.5 DISCUSSION

A transcriptomic-radiomic integrative framework allows us to explore the transcriptomic underpinnings (at the exon-level) of quantitative MRI radiomic phenotypes of GBM. This information can be used to better understand the molecular mechanisms that regulate the development of tumor phenotype and ultimately untangle intratumoral heterogeneity for future sub-classification of patients.
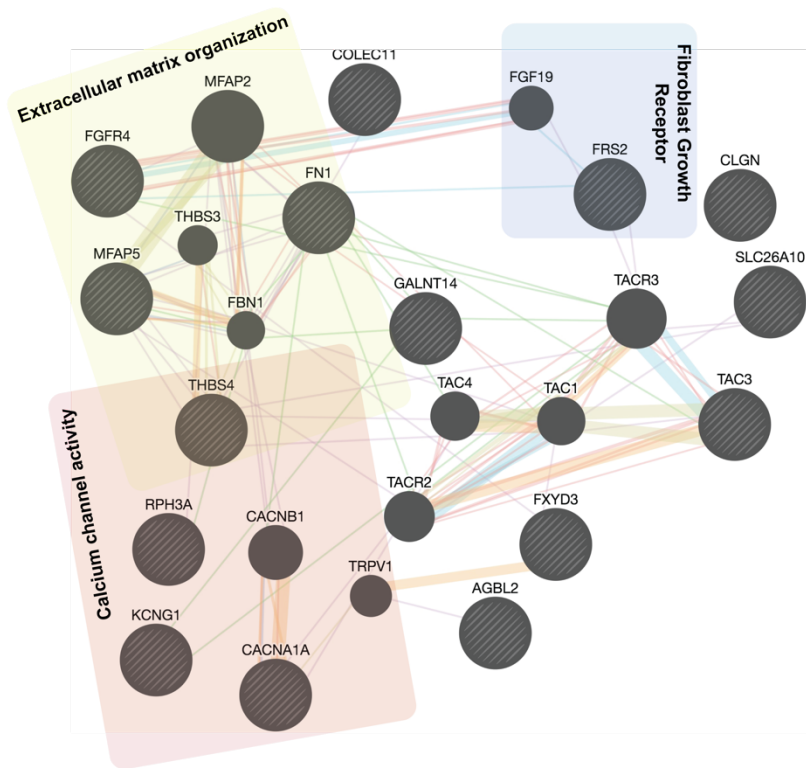
Figure 3.5 | GeneMANIA functional interaction network of genes with significant exon-radiophenotype correlation

We identify exon inclusion measures that are significantly correlated with Enhancement and Total ROIs. Our results indicate that the Enhancement ROI, which has been shown to inform blood brain barrier disruption through neovascularization, angiogenesis, and cellular proliferation, is associated with calcium signaling. Calcium signaling is a known orchestrator of GBM development and is involved in various tumor characteristics including invasion and sustained angiogenesis (Leclerc *et al*, 2016; Morrone *et al*, 2016). As for the Total ROI, which contains overlapping information with the Enhancement radiophenotype, we observe a role for PDGF signaling. PDGF has an established involvement in cell cycle regulation an cellular death (Westermark, 2014). This matches our findings, as the Total radiophenotype is also used to infer the extent of necrosis in GBM tumors.

Functional enrichment and interaction network mapping of morphology gene correlates reveal a molecular signature for glioma invasion and motility, a defining hallmark of GBM. We

observe the involvement of motility-enhancing signaling cascades involving PI3K, tyrosine kinase receptors, angiogenesis regulators, and neurotrophin in addition to ECM remodeling, ion channel activity, cytoskeletal reorganization and neural excitability. There is a growing body of evidence showing that neutrotrophin growth factors and disrupted cell-ECM interactions are major contributors to paracrine-mediated tumor invasiveness (Roux & Barker, 2002; Denkins, 2004; Wolf *et al*, 2018). Neurocentric processes, including altered ion channel activity, have also been shown to enable glioma call invasion into perivascular space (Cuddapah *et al*, 2014).

Tumor microenvironment is believed to primarily drive the excessive proliferation, diffusive parenchymal infiltration, and perivascular invasion behavior in GBM (Wolf *et al*, 2018). It is therefore heavily investigated to clarify the variegated treatment resistance, recurrence, and failure patterns across GBM patients. Given that our analysis only reflects glioma-intrinsic contributions, we are able to highlight the equally important role of core tumor cells in dictating the degree of invasiveness. GBM cells with an invasive phenotype can engage developmental migration mechanisms to increase proliferation. This agrees with our imaging findings, in which we note a significant association with edema levels. Edema is an important prognostic variable in GBM and a strong measure of invasiveness. We speculate that the introduction of new diagnostic measures of cellular invasiveness, through a concurrent monitoring of molecular correlates and the extent of edema, can help predict disease progression and better tailor therapies.

3.5.1 Study limitations

However, it is critical to note that findings from this study are purely exploratory and cannot be readily translated to clinical practice.

3.5.1.1 "Low" data quality and heterogeneity

We believe an integrative transcriptome-radiome framework has the potential to reveal unique molecular profiles for GBM tumor phenotypes. However, it is unable to yield reproducible findings

when applied to a dataset with two forms of inherently heterogeneous data and a small sample size. Bulk RNA sequencing provides a single molecular representation of the entire tumor. It therefore has an admixture of different types of cells with varying degrees of mutations and infiltration of other molecules. This can be mitigated by using a dataset with either histologically-defined tumor compartments (such as the Ivy GAP dataset; http://ivygap.swedish.org/) or single-cell RNA-seq.

Both ROI and VASARI radiomic features are also heterogeneous in that they are influenced by artifacts introduced at the image acquisition, image registration and processing, tumor segmentation, and feature extraction stages. They also highly depend on proper contouring of tumor regions, which varies between radiologists and image processing engineers/scientists.

3.5.1.2   Limited imaging descriptors

The radiomic features used in this analysis only reflect *a priori*-defined classic tumor compartments (contrast enhancements; edema; and necrosis) and their morphological characteristics (volume; surface area; etc). They provide useful information on tumor appearance but might fail to fully capture its structural complexity. For a more comprehensive assessment of GBM transcriptome-radiome relationships, it will be useful to include information from other tumor subregions as well as textural, kinetic, and functional descriptors that can inform tumor tissue microstructure and composition.

3.5.1.3   Incorporate categorical VASARI features

This study only incorporates quantitative image features and does not take advantage of the remaining rich semantic information provided through VASARI. We can evaluate the accuracy of the 24 extracted ROI features by correlating them to comparable VASARI radiophenotypes: proportion enhancing (F5); proportion necrosis (F7); and proportion edema (F8). We can also explore the relationship between ROI measures (such as enhancement ROI) and associated VASARI characteristics (such as and enhancement quality (F4); thickness of enhancing margin

(F11); definition of enhancing margin (F12); and CET cross midline (F23)). Finally, since we are interested in highlighting the molecular underpinnings of radiophenotypes, it will be beneficial to include VASARI features that are prognostic of genomic variation, recurrence and survival such as eloquent brain (F3); multifocal or multicentric (F9); ependymal extension (F19); etc.

3.5.1.4   Exon inclusion estimate distribution and missing values

There is a large number of exon inclusion estimates that have missing values. Median imputation is used to replace those values. Multiple imputation, however, is a better approach as it uses the distribution of observed data to estimate missing values that account for the uncertainty around the real value.

3.5.1.5   Confounding factors

Correlation-based investigations are notorious for their inability to distinguish true biological signals from confounding factors incurred by technical variability and stochastic events/ noise. Technical variability can occur throughout the various stages of this study's workflow. This is further discussed in the following chapter.

3.5.1.6   Association versus causation

This study assesses significance using a resampling approach to calculate adjusted *p*-values, specifying measurements equal to or less than a set threshold. On the basis of such ubiquity of assessment of significance, it is important to note that *p*-values are a measure of likelihood and not inference, a very common misinterpretation amongst scientific studies. P-values are indeed useful to assess the strength of evidence against a null hypothesis (no association). However, they are not indicative of biological inference, which in the context of Bayesian networks can be assessed by measuring the posterior probability of the hidden variable (imaging biomarker(s) in a phenotype-to-transcriptome relationship or gene expression/ exon inclusion pattern(s) in a transcriptome-to-phenotype relationship) after accounting for observations. Statistical methods are incapable of determining causality, as they only handle behavior under uncertain, static

conditions. Causality testing on candidate genes or image surrogates requires changing experimental conditions. Given their ability to account for changes in complex data, Bayesian-approaches can be employed for multidimensional association studies; however, they require knowledge of prior probabilities. Using uninformative prior probability distributions can result in spurious results and should therefore be avoided at all costs in radiotranscriptomic quantitative studies. Instead, an informed priory can be obtained by performing a separate controlled experiment, arbitrarily changing a set, underlying condition and observing potential modifications in observed response. We can therefore use molecular findings from such studies and conduct longitudinal studies or bench experiments to better determine causality/inference instead of association.

3.6 CONCLUSION

Here, we present the first exon inclusion-based "radiotranscriptomic" evaluation of GBM tumors. We demonstrate the transcriptome-radiome framework's ability to identify an exon-level signature for the Enhancement and Total (Necrotic) radiophenotypes. The high degree of heterogeneity in both feature spaces of the correlation analysis, as well as mismatches in dimensionality (high number of features with a small sample size) and lack of comprehensive radiomic features, has limited the framework's ability to reveal biologically- and clinically-relevant information.

3.7 FUTURE DIRECTIONS

For further "radiotranscriptomic" investigations of GBM tumor heterogeneity, we propose the following series of analyses for future studies: **(a)** explore the relationship between radiological and/or transcriptomic characteristics and other clinical attributes; **(b)** validate findings using the Ivy GAP dataset, which includes RNA sequencing from histologically-distinct tumor compartments as well as various longitudinal medical imaging modalities; **(c)** integrate additional omic information, including additional transcriptomic and radiomic features, for a comprehensive characterization of tumor heterogeneity; and **(d)** incorporate molecular subtype information.

CHAPTER IV

Systems Integration of Multiomics
in Translational Biomedical Research

4.1   BOTTLENECKS, PRECAUTIONS, AND CONSIDERATIONS

A multilayered and clinically-informed approach to integrating multiomic data enables the systematic assessment of relationships between different layers of rich data. Accordingly, it has all the ingredients fitted to investigate biomedical questions that have not been answered through hypothesis-driven or monolayer differential analysis methodologies. Nonetheless, this approach presents numerous shortcomings that necessitate a precautious attitude in analysis. The biggest challenge lies in the need to integrate terabytes of heterogeneous data from distinct sources. This can lead to the accumulation of errors throughout the various stages of the study, potentially biasing results and reducing their reproducibility and translatability. It also raises the concern of how to adequately link the various data types. Here, we highlight some of the challenges associated with this integrative approach, specifically with regards to the integration of transcriptomic, phenomic, and radiomic data, and provide workarounds to moderate them.

4.1.1   General shortcomings and considerations

4.1.1.1   Data quality

The number of existing, well-documented datasets that include concurrent RNA-seq, clinical, and medical imaging data is limited. The majority of those datasets often involve small cohorts; multi-institutional collection centers and platforms; variable protocols and study parameters; and missing information. While this limitation is not specific to a study, it will determine what biological question we can ask and what additional information (or datasets) we can incorporate to alleviate data quality concerns. In _use case scenario 2_, we highlight such limitation and the need to explore additional datasets to improve the power of the study and achieve statistically significant results.

4.1.1.2 Choice of data integration methodology

Multidimensional data integration can be achieved through either the summation of results from single omic layer analyses or the formation of mechanistic models that recapitulate true biological topology. While the former approach is simpler, the latter approach accounts for the interdependency between different biological layers, facilitating the inference of regulatory relationships between different biological elements (for example: transcription factor activation and repression of gene expression). To reflect the complex biology involved, it is beneficial to elucidate the structure and connection between biological layers and their dynamic evolution in response to the varying developmental and environmental demands and infections. Various tools have been developed to integrate heterogeneous multiomic data that exhibit differences in dynamic range, temporal information, and annotation. It is imperative to evaluate the costs and benefits of such integration methods, early on, as they can conceal biologically-important relationships.

We also highlight the need to normalize data (using log-transformation and variance-stabilization, for example) to ensure it conforms to normality prior to integration using statistical methods with prior assumptions of data distribution. Untransformed raw data can be skewed, dominated by extreme values from outliers, and exhibit heteroscedasticity through the dependence of mean on variance. In *use case scenario 1*, for example, we demonstrate the
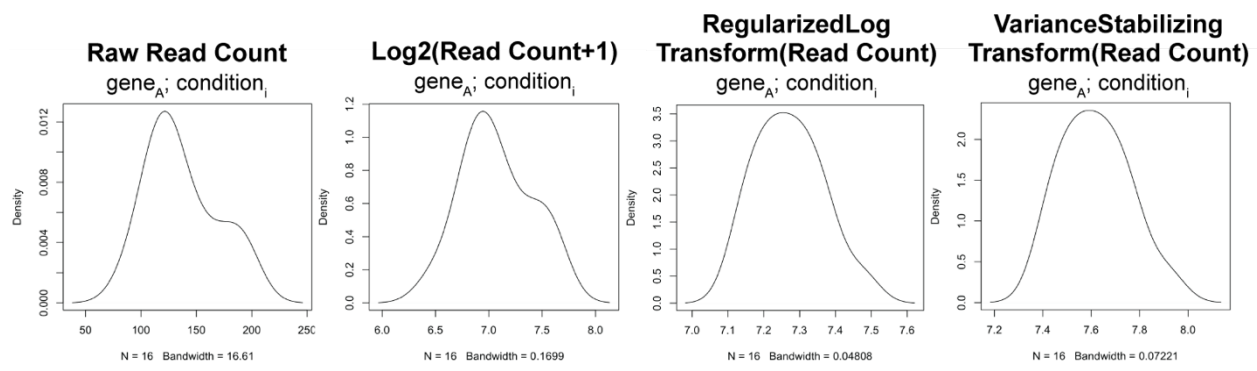
Figure 4.1 | Data transformation effects use gene expression read counts from use case scenario 1

effects of various transformation methods (log2 transformation; regularized log-transformation that also normalizes with respect to library size; and variance stabilizing transformation that output constant-variance data while also normalizing with respect to library size) on raw read count distribution for a single gene across a condition.

4.1.1.3 Exploratory nature

Studies in this arena predominantly use correlation / linear regression-based approaches to identify feature-phenotype relationships. To transition from an exploratory "guess and pray" attitude to a "predict, test, and validate" plan, we highlight the need for datasets with larger sample sizes. These datasets are better equipped to generate machine learning models for validation or computational models for cohesive and mechanistic representation of disease etiology. Furthermore, to reduce the number of false positives in association studies and ensure reproducibility and clinical applicability, further validation is necessary. This can be done experimentally, in which either another research group repeats the experiment using independent biological replicate samples or the same group attempts to reproduce the results using similar or different techniques such as quantitative reverse-transcription (qPCR). We can apply this form of validation to our findings in *use case scenario 1* to confirm their effects in a biological context. we can This can also be done computationally, through cross-validation using an independent dataset and measurement of concordance. We can use this in *use case scenario 2* to increase the study's power and ensure the reproducibility of results.

4.1.1.4 Confounding effects

Confounding effects influence results by either falsely demonstrating an association or masking a true relationship. Technical artifacts can be introduced during the following stages: study design and protocol selection; data collection; RNA-seq library preparation and analysis; image acquisition and processing; and statistical/mathematical integration models. Biological artifacts can also occur due to transient molecular changes as a result of unrelated environmental changes

and inherent variations between subjects. Such stochastic effects can be compounded in multiomic endeavors. The contribution of these effects on reported findings can be mitigated through quality control measures (to control for batch effect, for example) and emphasis on findings that recurrently demonstrate significance from each layer of analysis. We further discuss data type-specific confounding effects in later sections.

4.1.1.5 Determination of significance

Careful determination of the significance of reported findings is especially important in studies involving high-dimensional data. The arbitrary selection of cut-off thresholds for significance lacks the scientific rigor necessary to differentiate between true associations and chance. For example, in *use case scenario 2*, we correlate 6,599 exons with 26 radiomic features. If we were to use a standard p-value cuff-off of 0.05, we would expect to identify 330 "significant" genes due to chance alone. Multiple testing adjustment of p-values is commonly used to address this concern, with a variety of methods available to correct for type II or type II errors. A Benjamini-Hochberg false discovery rate (FDR) correction is often used in transcriptomic studies to control the proportion of false positive amongst a set of rejected hypotheses. Unlike family-wise error rate (FWER) methods, such as Bonferroni, it reduces the number of false positive discoveries without unnecessarily sacrificing power (probability of significance for tests of the false null hypotheses). On the other hand, re-sampling approaches (such as permutation testing) are able to control for error without heavily relying on prior assumption of underlying data distribution and p-value dependency structure.

Accordingly, it is important to select a multiple testing adjustment procedure that helps formulate biological hypotheses and prioritize statistically significant findings for future investigations. We emphasize that such procedure must be able to account for:

(a) Significant mismatches in data dimensionality

(b) Failure of the central limit theorem when working with small sample sizes, as asymptotic and exact *p*-values can drastically differ. We demonstrate this in *use case scenario 2*, in which a resampling-based technique is more appropriate when analyzing a 6,599 x 26 feature matrix for 38 patients.

(c) Correlation between features (gene-gene interactions, for example)

(d) The inherent proportion of features with significant effects (a percentage of genes are known to have a differential effect, for example). We demonstrate this in *use case scenario 1*, in which less a stringent FDR vs FWER technique is better at uncovering genes with differential expression patterns. However, with only 48 patients analyzed, a re-sampling approach might be a better fit.

4.1.1.6 Dimensionality mismatches

The number of investigated features often far exceeds experimental sample size. This necessitates the careful adoption of feature space reduction strategies to limit the number of hypotheses to be tested.

(a) One can remove uninteresting hypotheses –ones that do not pertain to the biological question or depend on the results of the analysis – prior to testing. We show this in *use case scenario 2*, in which we are interested in correlating information specific to the tumor and not the microenvironment. By filtering for expressed (read count > 1 TPM) glioma-intrinsic genes (n=7,425), we are able to reduce our exon inclusion feature space from 204,657 to 6,599 exon inclusion estimates without sacrificing information pertinent to the question in hand.

(b) Another approach is to use network-level information of features for correlation. For example, *use case scenario 1*, we correlate 19 co-expression gene clusters instead of 25,531 genes with maternal traits.

(c) Redundant features can also be filtered by removing those that highly correlate with one another. This approach can be applied in *use case scenario 2* to reduce the number of image features in the correlation matrix.

4.1.2 Shortcomings associated with RNA-seq

4.1.2.1 Inherent data collection and generation errors

Pre-sequencing chemistry (enzymatic reactions during RNA purification and extraction) can significantly influence findings from downstream analyses. Artifacts can also be introduced during library preparation as a result of amplification, inconsistent trimming of the RNA, and low "quality" RNA with a high level of degradation and contamination. Furthermore, insufficient gene coverage, which is related to low sequencing depth, can result in higher technical variation. As such, it is critical to carefully consider the biases and trade-offs involved in the selection of sequencing protocols when preparing an RNA-seq experiment.

4.1.2.2 Bulk RNA-seq limitations

Bulk tumor RNA-seq only reveals information at the cell-population level. It is unable to fully dissect intratumoral heterogeneity, as seemingly identical cells can harbor varying degrees of mutation and as such promote an inconsistent contribution to disease progression. With the expansion of single-cell RNA-seq, there is now the possibility of exploring the transcriptomic landscape at a single-cell resolution and deconvolving bulk RNA-seq samples using cell type proportion estimates.

4.1.2.3 Bias in differential gene expression analysis

The wealth of available algorithms and statistical methods raises the concern of methodology selection bias on reported findings. RNA-seq computational analysis tools vary in their statistical frameworks; assumptions of underlying data distribution; read-counting procedure; and identification of significant events. We explore this in *use case scenario 2*, in which we opt to use a count-based approach (HTSeq → DESeq2) to identify DEGs for the following reasons:

(1) It uses negative binomial regression for generalized linear modeling, therefore eliminating the dependency on a normally-distributed data.

(2) It calculates total expression, across gene isoforms, instead of using a probabilistic measure for transcript abundance like fragments per kilobase of transcript per million (FPKM)-based approaches.

(3) It corrects for library size and RNA composition bias by internally normalizing the data using the ratio of gene count to gene geometric mean across samples.

(4) It applies a shrinkage estimation for dispersion to account for variance within conditions.

Additionally, in differential gene expression analysis, a common approach to narrowing down the list of DEGs is to apply a simultaneous fold-change and multiple testing-adjusted *p*-value cutoff. This approach, however, favors genes with low expression values and misses biologically-important genes that typically exhibit small changes in RNA expression. We account for this in *use case scenario 1,* in which we elect to first apply a more stringent FDR cut-off of 0.05 to extract DEGs and then rank them by fold change.

4.1.2.4   Bias in differential alternative splicing analysis

Detection of alternative splicing is prone to error. Human exons are small, with ~80% measuring at less than 200 base pairs in length. Human introns, on the other hand, span much larger sequence lengths. The confidence interval of exon inclusion estimates is also highly dependent on RNA-seq read coverage. As dictated by ENCODE guidelines and Illumina, a minimum of 30 million paired-end reads > 30 base pairs are needed for gene expression estimation and a minimum of 30-60 million are needed for splicing analysis. Furthermore, exons of highly expressed genes have high read counts that can introduce false positives.

4.1.3 Shortcomings specific to radiomics

4.1.3.1 Inconsistent image acquisition protocols

Medical imaging acquisition parameters are constrained by local settings and scanner capabilities, hindering the standardization of generated images. Pre-processing techniques, such as image registration, intensity normalization, and re-slicing, can be applying to ensure images are comparable across various acquisitions (Narang *et al*, 2016).

4.1.3.2 Image segmentation and feature extraction variability

Variability can also arise during image alignment and segmentation and feature extraction and quantification. It is important to ensure that radiomic features are not only independent of collection site but also of the radiologist/annotator (to limit inter- and intra-observer uncertainty) and extraction approach and protocol (to limited inter-software uncertainty).

There are various ways to validate the accuracy of tumor segmentation. A common approach is to achieve group-consensus contouring, by taking the overlap of contours generated by field experts or repeated segmentations and calculating a metric of overlap similarity such as the Dice coefficient (as we do *use case scenario 2*) or distance similarity. Since the calculation of image features highly depends on accurate tumor contouring, it is important to evaluate image feature robustness against tumor delineation uncertainties to maintain features with high intraclass correlation scores and ensure reproducibility. It is equally important to employ a suitable image segmentation evaluation metric that is sensitive to the properties of the segmented image, accounting for size, boundary, and density.

Another approach is to refer to a controlled and standardized vocabulary system for image features (such as the VASARI lexicon used in *use case scenario 2*). By normalizing invariant feature measurements, the rank-ordering form of scoring used in VASARI is able to reduce the qualitative nature of manual segmentation. Nonetheless, supervised extraction of VASARI

features remains subjective and as a result, quantitative measurements can vary between different radiologists and different extraction dates for the same radiologist. Inter-observer variability can also arise due to the limited precision when assessing certain features. This can introduce undesired noise in association analysis. To moderate inhomogeneity and subsequent obfuscation in quantitative analysis, it is common to consult the expertise of only one specialist, to ensure measurements are done in a narrow time frame, and validate VASARI semantic findings with computational features such as the ROI measures obtained in *use case scenario 2*.

4.1.4 Shortcomings associated with phenomics

Phenomic data can also face the limitation of incomplete and inaccurate documentation of information. We observe this in *use case scenario 1*, in which many patient clinical variable entries are missing or are designated as "unknown", thus reducing our ability to expand on our findings with additional phenotypic relationships. And in some cases, infection records are populated under "Notes" instead of the appropriate column, which can drastically influence our results for significant differences in infection rates between parturition groups. Digitization is becoming more widely adopted to streamline the process of data acquisition and limit human error as a result of manual entry

Furthermore, there is a concern for inconsistencies and mismatches in data entry due to the use of different metrics and criterion for assessment and the limited sharing between institutions. This can be mitigated through effective data governance, in which standardized terminology and a uniform data structure can be implemented to ensure consistency across investigation efforts and facilitate cross-institutional communication.

4.2 SYSTEMS BIOLOGY MEET MEDICINE

4.2.1 A reductionist, hypothesis-driven theme in "old" medicine

The "Oslerian paradigm", which uses a top-down hypothesis-driven approach to emphasize a direct link between symptoms and etiology, heavily predominates the current practice in Western modern medicine. The correlation of clinical manifestations with pathological findings has bestowed upon clinicals a clearly-defined list of syndromic patterns to look for. Using this list, they are able to effectively address acute or monofactorial disorders. A reductionist lens is elegant in its focus on select elements. However, it fails to account for the emergent properties that arise from the dynamic and interdependent interactions within biological systems. It also assumes an averaged contribution of individual components. As such, it undervalues the effect of unpredictable outliers within heterogeneous populations and the multifunctionality of certain components. The shortcomings of this approach become especially evident when evaluating chronic and multifactorial health conditions.

4.2.2 A holistic, systems theme in biomedical research

The early 21st century witnessed a paradigm shift in biomedical research, driven by biology's transition from a qualitative, descriptive discipline into a quantitative, numerical field. The recent emergence of high-throughput and wet-lab technologies, advanced imaging modalities, and powerful computational capabilities has rendered a wealth of biomedical "big data". This newfound capacity has facilitated the adoption of a bottom-up data-driven approach, in which knowledge is gained from data integration rather than conventional hypotheses testing. By applying sophisticated algorithms and statistical tools, investigators are now able to rapidly examine massive amounts of data and unravel novel biology at an unprecedented level.

Left unguided, a hypothesis-limited methodology, however, can render findings that derail from biological truth and lack clinical significance. With the conviction that *"every object that*

biology studies is a system of systems" (Trewavas, 2006), the field of systems biology emerged to comprehensively evaluate time- and space-sensitive communication within and across systems, subsystems, and individual components. It emphasizes the incorporation of cross-disciplinary techniques to showcase the joint involvement of multimodal data points in driving a phenotype within a defined biological context. In doing so, it attempts to extract order from an inherently chaotic data pool. A systems data-driven approach has helped inform numerous novel hypotheses for follow-up studies and identify key biomolecules in health and disease states.

4.2.3 A reconciliation of hypothesis- and data-driven approaches in "new" medicine

A data-driven approach offers an all-inclusive depiction of the underlying structure, accounting for the genetic, phenotypic, and environmental diversity of human populations. Coupled with a systems perspective, it can provide elaborate mechanistic models of disease etiology. On its own, such approach can have limited clinical utility. The development of systems models is highly iterative, requiring infinite rounds of fine-tuning to hone in on key biological functions while accommodating newly-discovered influences of individual variability. And the level of detail in reported systems biology findings does not match the capabilities of medical practice. This has resulted in a "work-in-progress" style in data-driven research, where results make their way to publications but not the clinic.

In a 1933 lecture at Oxford, Einstein argued that *"it can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience".* A hypothesis or *a priori* information can safely guide systems data-driven models while alleviating some of the complexity that has hampered their adoption in the clinic. In a highly integrative process, both hypothesis- and data-driven tactics can synergistically come in play at various stages and throughout the course of a study. To better illustrate this, let us assume the

role of a detective who has reported to a murder scene and is trying to solve the case. Using her past experience, she knows what areas to target to look for traces left behind by the criminal. She begins to gather all available pieces of evidence directly surrounding the body, in the vicinity, and from relevant areas. She starts analyzing those pieces, looking for clues and patterns that deviate from the norm. Using her knowledge of human nature and criminal mentality, she then tries to build a narrative to put together the pieces of evidence. She continues to gather additional pieces of information as she goes along. She studies the victim's background and interrogates involved bystanders, constantly tweaking and strengthening the narrative. She has now narrowed down her list of potential criminals to a much smaller pool of players with strong motives for committing the crime. She proceeds to confirm her highly informed exploratory model by testing which hypotheses hold true. With the help of other members of the police and judicial authority, she is now able to isolate the true criminal.

We believe such converged data- and hypothesis-mediated framework can lay the groundwork for translational research and can help identify true determinants of health and
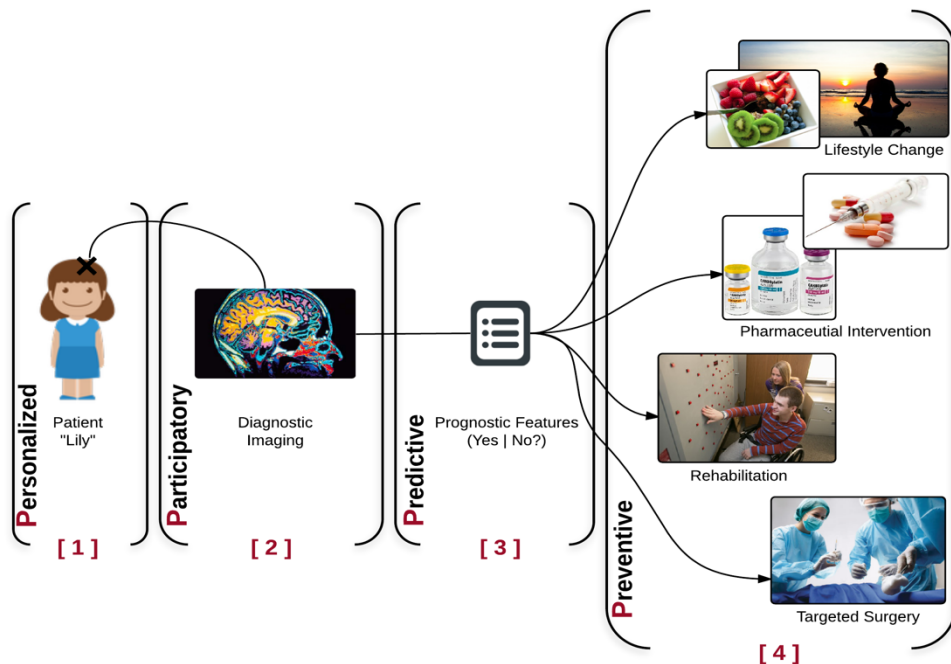
Figure 4.2 | Converged hypothesis- and data-mediated research approaches in P4 medicine

disease. By maintaining an ongoing and multidirectional line of communication between basic biological and computational sciences, population-centric biomedical research, and patient-centric clinical practice, it can ultimately help us make the leap towards modern P4 (personalized, participatory, predictive and preventive) medicine (Figure 4.2).

.

REFERENCES

Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ & Lambin P (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.*

Aleksic D, Blaschke L, Mißbach S, Hänske J, Weiß W, Handler J, Zimmermann W, Cabrera-Sharp V, Read JE, De Mestre AM, O'Riordan R, Moore T & Kammerer R (2016) Convergent evolution of pregnancy-specific glycoproteins in human and horse. *Reproduction*

Anders S, Pyl PT & Huber W (2015) HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*

Anderson MW & Schrijver I (2010) Next generation DNA sequencing and the future of genomic medicine. *Genes (Basel).*

Armstrong DL, McGowen MR, Weckle A, Pantham P, Caravas J, Agnew D, Benirschke K, Savage-Rumbaugh S, Nevo E, Kim CJ, Wagner GP, Romero R & Wildman DE (2017) The core transcriptome of mammalian placentas and the divergence of expression with placental shape. *Placenta*

Asl AAH, Safari S & Hamrah MP (2017) Epidemiology and Related Risk Factors of Preterm Labor as an obstetrics emergency. *Emerg. (Tehran, Iran)*

Bailey J, Sparey C, Phillips RJ, Gilmore K, Robson SC, Dunlop W & Europe-Finner GN (2000) Expression of the cyclic AMP-dependent transcription factors, CREB, CREM and ATF2, in the human myometrium during pregnancy and labour. *Mol Hum Reprod*

Bailey J, Tyson-Capper AJ, Gilmore K, Robson SC & Europe-Finner GN (2005) Identification of human myometrial target genes of the cAMP pathway: The role of cAMP-response element binding (CREB) and modulator (CREMα and CREMτ2α) proteins. *J. Mol. Endocrinol.*

Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K & Davatzikos C (2017) Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data*

Blencowe H, Cousens S, Chou D, Oestergaard M, Say L, Moller AB, Kinney M & Lawn J (2013) Born Too Soon: The global epidemiology of 15 million preterm births. *Reprod. Health*

Blois SM, Sulkowski G, Tirado-González I, Warren J, Freitag N, Klapp BF, Rifkin D, Fuss I, Strober W & Dveksler GS (2014) Pregnancy-specific glycoprotein 1 (PSG1) activates TGF-β and prevents dextran sodium sulfate (DSS)-induced colitis in mice. *Br. Dent. J.*

Boardman JP (2008) Preterm Birth: Causes, Consequences and Prevention. *J. Obstet. Gynaecol. (Lahore).*

Bochkis IM, Schug J, Ye DZ, Kurinna S, Stratton SA, Barton MC & Kaestner KH (2012) Genome-wide location analysis reveals distinct transcriptional circuitry by paralogous

regulators Foxa1 and Foxa2. *PLoS Genet.*

Bogaerts A, Witters I, Van den Bergh BRH, Jans G & Devlieger R (2013) Obesity in pregnancy: Altered onset and progression of labour. *Midwifery*

Brockway H, Jones H & Muglia L (2017) Pregnancy specific glycoproteins are disrupted in idiopathic preterm birth and impact cytotrophoblast differentiation. *Placenta* **57:** 256 Available at: https://www.sciencedirect.com/science/article/pii/S0143400417307634 [Accessed October 17, 2018]

Buchner DA, Trudeau M & Meisler MH (2003) SCNM1, a putative RNA splicing factor that modifies disease severity in mice. *Science (80-.).*

Cardenas I, Means RE, Aldo P, Koga K, Lang SM, Booth C, Manzur A, Oyarzun E, Romero R & Mor G (2010) Viral Infection of the Placenta Leads to Fetal Inflammation and Sensitization to Bacterial Products Predisposing to Preterm Labor. *J. Immunol.*

Cartwright JE, Fraser R, Leslie K, Wallace AE & James JL (2010) Remodelling at the maternal-fetal interface: Relevance to human pregnancy disorders. *Reproduction*

Cavazos-Rehg PA, Krauss MJ, Spitznagel EL, Bommarito K, Madden T, Olsen MA, Subramaniam H, Peipert JF & Bierut LJ (2015) Maternal Age and Risk of Labor and Delivery Complications. *Matern. Child Health J.*

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G V., Clark NR & Ma'ayan A (2013) Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*

Chevillard G & Blank V (2011) NFE2L3 (NRF3): The Cinderella of the Cap'n'Collar transcription factors. *Cell. Mol. Life Sci.*

Cnattingius S (1998) Prepregnancy Weight and the Risk of Adverse Pregnancy Outcomes. *N. Engl. J. Med.*

Cuddapah VA, Robel S, Watkins S & Sontheimer H (2014) A neurocentric perspective on glioma invasion. *Nat. Rev. Neurosci.*

Damodaran S, Berger MF & Roychowdhury S (2015) Clinical Tumor Sequencing: Opportunities and Challenges for Precision Cancer Medicine. *Am. Soc. Clin. Oncol. Educ. B.*

Denkins Y (2004) Brain metastases in melanoma: Roles of neurotrophins. *Neuro. Oncol.*

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M & Gingeras TR (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*

Dove A (1999) Proteomics: Translating genomics into products? *Nat. Biotechnol.*

Dubicke A, Fransson E, Centini G, Andersson E, Byström B, Malmström A, Petraglia F, Sverremark-Ekström E & Ekman-Ordeberg G (2010) Pro-inflammatory and anti-inflammatory cytokines in human preterm and term cervical ripening. *J. Reprod. Immunol.*

Dunn-Albanese LR, Ackerman IV WE, Xie Y, Iams JD & Kniss DA (2004) Reciprocal expression

of peroxisome proliferator-activated receptor-γ and cyclooxygenase-2 in human term parturition. *Am. J. Obstet. Gynecol.*

Ecker JL, Chen KT, Cohen AP, Riley LE & Lieberman ES (2001) Increased risk of cesarean delivery with advancing maternal age: Indications and associated factors in nulliparous women. *Am. J. Obstet. Gynecol.*

El-Azzamy H, Balogh A, Romero R, Xu Y, LaJeunesse C, Plazyo O, Xu Z, Price TG, Dong Z, Tarca AL, Papp Z, Hassan SS, Chaiworapongsa T, Kim CJ, Gomez-Lopez N & Than NG (2017) Characteristic Changes in Decidual Gene Expression Signature in Spontaneous Term Parturition. *J. Pathol. Transl. Med.*

Eun HS & Jeong W II (2016) Dual Notch signaling in proinflammatory macrophage activation. *Hepatology*

Faas MM, Spaans F & De Vos P (2014) Monocytes and macrophages in pregnancy and pre-eclampsia. *Front. Immunol.*

Fetalvero KM, Zhang P, Shyu M, Young BT, Hwa J, Young RC & Martin KA (2008) Prostacyclin primes pregnant human myometrium for an enhanced contractile response in parturition. *J. Clin. Invest.*

Filippi M, Horsfield MA, Adèr HJ, Barkhof F, Bruzzi P, Evans A, Frank JA, Grossman RI, McFarland HF, Molyneux P, Paty DW, Simon J, Tofts PS, Wolinsky JS & Miller DH (1998) Guidelines for using quantitative measures of brain magnetic resonance imaging abnormalities in monitoring the treatment of multiple sclerosis. *Ann. Neurol.*

Folkert VW, Yunis M & Schlondorff D (1984) Prostaglandin synthesis linked to phosphatidylinositol turnover in isolated rat glomeruli. *Biochim. Biophys. Acta (BBA)/Lipids Lipid Metab.*

Fontanil T, Rúa S, Llamazares M, Moncada-Pazos A, Quirós PM, García-Suárez O, Vega JA, Sasaki T, Mohamedi Y, Esteban MM, Obaya AJ & Cal S (2014) Interaction between the ADAMTS-12 metalloprotease and fibulin-2 induces tumor-suppressive effects in breast cancer cells. *Oncotarget*

Fu B & Wei H (2016) Decidual natural killer cells and the immune microenvironment at the maternal-fetal interface. *Sci. China Life Sci.*

Gilbert ES & Harmon JS (2004) Manual of high risk pregnancy and delivery, 3rd edition. *St Louis Mosby*

Goossens K, Van Soom A, Van Zeveren A, Favoreel H & Peelman LJ (2009) Quantification of Fibronectin 1 (FN1) splice variants, including two novel ones, and analysis of integrins as candidate FN1 receptors in bovine preimplantation embryos. *BMC Dev. Biol.*

GRAM A, BOOS A & KOWALEWSKI MP (2017) Cellular localization, expression and functional implications of the utero-placental endothelin system during maintenance and termination of canine gestation. *J. Reprod. Dev.*

Gross SR & Kinzy TG (2005) Translation elongation factor 1A is essential for regulation of the

actin cytoskeleton and cell morphology. *Nat. Struct. Mol. Biol.*

Grudzinskas JG, Gordon YB, Menabawey M, Lee JN, Wadsworth J & Chard T (1983) Identification of high-risk pregnancy by the routine measurement of pregnancy-specific β1-glycoprotein. *Am. J. Obstet. Gynecol.*

Gutman DA, Cooper LAD, Hwang SN, Holder CA, Gao J, Aurora TD, Dunn WD, Scarpace L, Mikkelsen T, Jain R, Wintermark M, Jilwan M, Raghavan P, Huang E, Clifford RJ, Mongkolwat P, Kleper V, Freymann J, Kirby J, Zinn PO, et al (2013) MR Imaging Predictors of Molecular Profile and Survival: Multi-institutional Study of the TCGA Glioblastoma Data Set. *Radiology*

Hadi T, Bardou M, Mace G, Sicard P, Wendremaire M, Barrichon M, Richaud S, Demidov O, Sagot P, Garrido C & Lirussi F (2015) Glutathione prevents preterm parturition and fetal death by targeting macrophage-induced reactive oxygen species production in the myometrium. *FASEB J.*

Hammoud MA (1996) Prognostic significance of preoperative MRI scans in glioblastoma multiforme. *J. Neurooncol.*

Hassan SS, Romero R, Tarca AL, Nhan-Chang CL, Vaisbuch E, Erez O, Mittal P, Kusanovic JP, Mazaki-Tovi S, Yeo L, Draghici S, Kim JS, Uldbjerg N & Kim CJ (2009) The transcriptome of cervical ripening in human pregnancy before the onset of labor at term: Identification of novel molecular functions involved in this process. *J. Matern. Neonatal Med.*

Hauger MS, Gibbons L, Vik T & Belizán JM (2008) Prepregnancy weight status and the risk of adverse pregnancy outcome. *Acta Obstet. Gynecol. Scand.*

Hermans FJ, Bruijn MM, Vis JY, Wilms FF, Oudijk MA, Porath MM, Scheepers HC, Bloemenkamp KW, Bax CJ, Cornette JM, Nij Bijvanck BW, Franssen MT, Vandenbussche FP, Kok M, Grobman WA, Van Der Post JA, Bossuyt PM, Opmeer BC, Mol BW, Schuit E, et al (2015) Risk stratification with cervical length and fetal fibronectin in women with threatened preterm labor before 34 weeks and not delivering within 7 days. *Acta Obs. Gynecol Scand*

Hitzemann R, Bottomly D, Darakjian P, Walter N, Iancu O, Searles R, Wilmot B & Mcweeney S (2013) Genes, behavior and next-generation RNA sequencing. *Genes, Brain Behav.*

Huang Z, Rose AH & Hoffmann PR (2012) The Role of Selenium in Inflammation and Immunity: From Molecular Mechanisms to Therapeutic Opportunities. *Antioxid. Redox Signal.*

Inada A, Someya Y, Yamada Y, Ihara Y, Kubota A, Ban N, Watanabe R, Tsuda K & Seino Y (1999) The cyclic AMP response element modulator family regulates the insulin gene transcription by interacting with transcription factor IID. *J. Biol. Chem.*

Itakura H, Achrol AS, Mitchell LA, Loya JJ, Liu T, Westbroek EM, Feroze AH, Rodriguez S, Echegaray S, Azad TD, Yeom KW, Napel S, Rubin DL, Chang SD, Harsh GR & Gevaert O (2015) Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. *Sci. Transl. Med.*

Jaffe CC (2012) Imaging and Genomics: Is There a Synergy? *Radiology*

Jain R, Poisson LM, Gutman D, Scarpace L, Hwang SN, Holder CA, Wintermark M, Rao A, Colen RR, Kirby J, Freymann J, Jaffe CC, Mikkelsen T & Flanders A (2014) Outcome Prediction in Patients with Glioblastoma by Using Imaging, Clinical, and Genomic Biomarkers: Focus on the Nonenhancing Component of the Tumor. *Radiology*

Jaiswal MK, Agrawal V, Pamarthy S, Katara GK, Kulshrestha A, Gilman-Sachs A, Beaman KD & Hirsch E (2015) Notch Signaling in Inflammation-Induced Preterm Labor. *Sci. Rep.*

Jiang B, Zhao W, Yuan J, Qian Y, Sun W, Zou Y, Guo C, Chen B, Shao C & Gong Y (2012) Lack of Cul4b, an E3 ubiquitin ligase component, leads to embryonic lethality and abnormal placental development. *PLoS One*

Jiang Y, Jiang R, Cheng X, Zhang Q, Hu Y, Zhang H, Cao Y, Zhang M, Wang J, Ding L, Diao Z, Sun H & Yan G (2016) Decreased expression of NR4A nuclear receptors in adenomyosis impairs endometrial decidualization. *Mol. Hum. Reprod.*

Kaloglu C & Onarlioglu B (2010) Extracellular matrix remodelling in rat endometrium during early pregnancy: The role of fibronectin and laminin. *Tissue Cell*

Karagas MR, Hollenbach K a, Hickok DE & Daling JR (1993) Induction of labor and risk of sudden infant death syndrome. *Obstet. Gynecol.*

Kashif M, Hellwig A, Hashemolhosseini S, Kumar V, Bock F, Wang H, Shahzad K, Ranjan S, Wolter J, Madhusudhan T, Bierhaus A, Nawroth P & Isermann B (2012) Nuclear factor erythroid-derived 2 (Nfe2) regulates junD DNA-binding activity via acetylation: A novel mechanism regulating trophoblast differentiation. *J. Biol. Chem.*

Van Keuren-Jensen K, Keats JJ & Craig DW (2014) Bringing RNA-seq closer to the clinic. *Nat. Biotechnol.*

Kim J, Zhao K, Jiang P, Lu Z xiang, Wang J, Murray JC & Xing Y (2012) Transcriptome landscape of the human placenta. *BMC Genomics*

Kleinrouweler CE, van Uitert M, Moerland PD, Ris-Stalpers C, van der Post JAM & Afink GB (2013) Differentially Expressed Genes in the Pre-Eclamptic Placenta: A Systematic Review and Meta-Analysis. *PLoS One*

Koltai H & Weingarten-Baror C (2008) Specificity of DNA microarray hybridization: Characterization, effectors and approaches for data correction. *Nucleic Acids Res.*

Koucký M, Malíčková K, Cindrová-Davies T, Germanová A, Pařízek A, Kalousová M, Hájek Z & Zima T (2014) Low levels of circulating T-regulatory lymphocytes and short cervical length are associated with preterm labor. *J. Reprod. Immunol.*

Kuleshov M V., Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW & Ma'ayan A (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44:** W90–W97

Langfelder P & Horvath S (2008) WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*

Lappas M, Permezel M, Georgiou HM & Rice GE (2001) Type II phospholipase A2in preterm human gestational tissues. *Placenta*

Lasocki A, Gaillard F, Tacey M, Drummond K & Stuckey S (2016) Multifocal and multicentric glioblastoma: Improved characterisation with FLAIR imaging and prognostic implications. *J. Clin. Neurosci.*

Lawton CA (2009) Malignant Gliomas in Adults. *Yearb. Oncol.*

Leclerc C, Haeich J, Aulestia FJ, Kilhoffer MC, Miller AL, Néant I, Webb SE, Schaeffer E, Junier MP, Chneiweiss H & Moreau M (2016) Calcium signaling orchestrates glioblastoma development: Facts and conjunctures. *Biochim. Biophys. Acta - Mol. Cell Res.*

Lee KJ, Shim SH, Kang KM, Kang JH, Park DY, Kim SH, Farina A, Shim SS & Cha DH (2010) Global gene expression changes induced in the human placenta during labor. *Placenta*

Lei W, Feng XH, Deng WB, Ni H, Zhang ZR, Jia B, Yang XL, Wang TS, Liu JL, Su RW, Liang XH, Qi AR & Yang ZM (2012) Progesterone and DNA damage encourage uterine cell proliferation and decidualization through up-regulating ribonucleotide reductase 2 expression during early pregnancy in mice. *J. Biol. Chem.*

Li D, Wei T, Abbott CM & Harrich D (2013) The Unexpected Roles of Eukaryotic Translation Elongation Factors in RNA Virus Replication and Pathogenesis. *Microbiol. Mol. Biol. Rev.*

Li T, Robert EI, van Breugel PC, Strubin M & Zheng N (2010) A promiscuous α-helical motif anchors viral hijackers and substrate receptors to the CUL4–DDB1 ubiquitin ligase machinery. *Nat. Struct. Mol. Biol.*

Lin TM, Halbert SP & Spellacy WN (1974) Measurement of pregnancy-associated plasma proteins during human gestation. *J. Clin. Invest.*

Lin ZX (2013) Glioma-related edema: New insight into molecular mechanisms and their clinical implications. *Chin. J. Cancer*

Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, Lawn JE, Cousens S, Mathers C & Black RE (2016) Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet*

St. Louis D, Romero R, Plazyo O, Arenas-Hernandez M, Panaitescu B, Xu Y, Milovic T, Xu Z, Bhatti G, Mi Q-S, Drewlo S, Tarca AL, Hassan SS & Gomez-Lopez N (2016) Invariant NKT Cell Activation Induces Late Preterm Birth That Is Attenuated by Rosiglitazone. *J. Immunol.*

Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P & Ellison DW (2016) The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.*

Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*

Mahesh M (2013) The Essential Physics of Medical Imaging, Third Edition. *Med. Phys.*

Malloy MH (2013) Prematurity and sudden infant death syndrome: United States 2005-2007. *J. Perinatol.*

Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA & Ferrari R (2018) Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. *Brief. Bioinform.*

Marcellin L, Schmitz T, Messaoudene M, Chader D, Parizot C, Jacques S, Delaire J, Gogusev J, Schmitt A, Lesaffre C, Breuiller-Fouché M, Caignard A, Vaiman D, Goffinet F, Cabrol D, Gorochov G & Méhats C (2017) Immune Modifications in Fetal Membranes Overlying the Cervix Precede Parturition in Humans. *J. Immunol.*

Martí-Bonmatí L, Sopena R, Bartumeus P & Sopena P (2010) Multimodality imaging techniques. *Contrast Media Mol. Imaging*

Martinez FF, Cervi L, Knubel CP, Panzetta-Dutari GM & Motran CC (2013) The Role of Pregnancy-Specific Glycoprotein 1a (PSG1a) in Regulating the Innate and Adaptive Immune Response. *Am. J. Reprod. Immunol.*

Martinez R, Rohde V & Schackert G (2010) Different molecular patterns in glioblastoma multiforme subtypes upon recurrence. *J. Neurooncol.*

Maulik D, De A, Ragolia L, Evans J, Grigoryev D, Lankachandra K, Mundy D, Muscat J, Gerkovich MM & Ye SQ (2015) Downregulation of Placental Neuropilin-1 in Fetal Growth Restriction. *Am.J.Obstet.Gynecol.*

McDonald SD, Han Z, Mulla S & Beyene J (2010) Overweight and obesity in mothers and risk of preterm birth and low birth weight infants: Systematic review and meta-analyses. *BMJ*

Mehra A, Ali C, Parcq J, Vivien D & Docagne F (2016) The plasminogen activation system in neuroinflammation. *Biochim. Biophys. Acta - Mol. Basis Dis.*

Mori M, Bogdan A, Balassa T, Csabai T & Szekeres-Bartho J (2016) The decidua—the maternal bed embracing the embryo—maintains the pregnancy. *Semin. Immunopathol.*

Morrone FB, Gehring MP & Nicoletti NF (2016) Calcium Channels and Associated Receptors in Malignant Brain Tumor Therapy. *Mol. Pharmacol.*

Mosher AA, Rainey KJ, Riley B, Levinson HS, Vinturache AE, Wood SL & Slater DM (2014) Regulation of sPLA2-IID in human decidua: Insights into the complexity of the prostaglandin pathway in labor. *Reprod. Sci.*

Moutquin JM (2003) Classification and heterogeneity of preterm birth. In *BJOG: An International Journal of Obstetrics and Gynaecology*

Mwaniki MK, Atieno M, Lawn JE & Newton CRJC (2012) Long-term neurodevelopmental outcomes after intrauterine and neonatal insults: A systematic review. *Lancet*

Mylonas I & Friese K (2015) The indications for and risks of elective cesarean section. *Dtsch.*

Ärzteblatt Int.

Narang S, Lehrer M, Yang D, Lee J & Rao A (2016) Radiomics in glioblastoma: current status, challenges and potential opportunities. *Transl. Cancer Res.*

Nathanson V (1994) Health Data in the Information Age: Use, Disclosure, and Privacy. *BMJ*

Nhan-Chang CL, Romero R, Tarca AL, Mittal P, Kusanovic JP, Erez O, Mazaki-Tovi S, Chaiworapongsa T, Hotra J, Than NG, Kim JS, Hassan SS & Kim CJ (2010) Characterization of the transcriptome of chorioamniotic membranes at the site of rupture in spontaneous labor at term. *Am. J. Obstet. Gynecol.*

Norwitz ER, Bonney EA, Snegovskikh V V., Williams MA, Phillippe M, Park JS & Abrahams VM (2015) Molecular regulation of parturition: The role of the decidual clock. *Cold Spring Harb. Perspect. Med.*

Ohgaki H & Kleihues P (2005) Epidemiology and etiology of gliomas. *Acta Neuropathol.*

Osol G & Moore LG (2014) Maternal Uterine Vascular Remodeling During Pregnancy. *Microcirculation*

Ozawa T, Riester M, Cheng YK, Huse JT, Squatrito M, Helmy K, Charles N, Michor F & Holland EC (2014) Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. *Cancer Cell*

Ozmen A, Unek G & Korgun ET (2017) Effect of glucocorticoids on mechanisms of placental angiogenesis. *Placenta*

Padhani AR, Liu G, Mu-Koh D, Chenevert TL, Thoeny HC, Takahara T, Dzik-Jurasz A, Ross BD, Van Cauteren M, Collins D, Hammoud DA, Rustin GJS, Taouli B & Choyke PL (2009) Diffusion-Weighted Magnetic Resonance Imaging as a Cancer Biomarker: Consensus and Recommendations. *Neoplasia*

Padhani AR & Miles KA (2010) Multiparametric Imaging of Tumor Response to Therapy. *Radiology*

Park E, Pan Z, Zhang Z, Lin L & Xing Y (2018) The Expanding Landscape of Alternative Splicing Variation in Human Populations. *Am. J. Hum. Genet.*

Pavličev M, Wagner GPGP, Chavan AR, Owens K, Maziarz J, Dunn-Fletcher C, Kallapur SG, Muglia L, Jones H, Pavli??ev M, Wagner GPGP, Chavan AR, Owens K, Maziarz J, Dunn-Fletcher C, Kallapur SG, Muglia L & Jones H (2017) Single-cell transcriptomics of the human placenta: Inferring the cell communication network of the maternal-fetal interface. *Genome Res.*

Pazos MA, Kraus TA, Muñoz-Fontela C & Moran TM (2012) Estrogen mediates innate and adaptive immune alterations to influenza infection in pregnant mice. *PLoS One*

Plow EF & Das R (2009) Enolase-1 as a plasminogen receptor. *Blood*

Poldrack RA (2007) Region of interest analysis for fMRI. *Soc. Cogn. Affect. Neurosci.*

Robinson BG, Emanuel RL, Frimt DM, Majzoub JA & Frim DM (1988) Glucocorticoid stimulates expression of corticotropin-releasing hormone gene in human placenta. *Proc. Natl. Acad. Sci. U. S. A.*

Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG & Ma'ayan A (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford).*

Roux PP & Barker PA (2002) Neurotrophin signaling through the p75 neurotrophin receptor. *Prog. Neurobiol.*

Ruzycky AL (1998) Down-regulation of the mitogen-activated protein kinase cascade immediately before parturition in the rat myometrium. *J. Soc. Gynecol. Investig.*

Sachs UJH, Andrei-Selmer CL, Maniar A, Weiss T, Paddock C, Orlova V V., Eun YC, Newman PJ, Preissner KT, Chavakis T & Santoso S (2007) The neutrophil-specific antigen CD177 is a counter-receptor for platelet endothelial cell adhesion molecule-1 (CD31). *J. Biol. Chem.*

Salker M, Teklenburg G, Molokhia M, Lavery S, Trew G, Aojanepong T, Mardon HJ, Lokugamage AU, Rai R, Landles C, Roelen BAJ, Quenby S, Kuijk EW, Kavelaars A, Heijnen CJ, Regan L, Macklon NS & Brosens JJ (2010) Natural selection of human embryos: Impaired decidualization of endometrium disables embryo-maternal interactions and causes recurrent pregnancy loss. *PLoS One*

Sanborn BM, Millan JL, Meistrich ML & Moore LC (1997) Alternative splicing of CREB and CREM mRNAs in an immortalized germ cell line. *J. Androl.*

Sang Y, Yan F & Ren X (2015) The role and mechanism of CRL4 E3 ubiquitin ligase in cancer and its potential therapy implications. *Oncotarget*

Sasikumar AN, Perez WB & Kinzy TG (2012) The many roles of the eukaryotic elongation factor 1 complex. *Wiley Interdiscip. Rev. RNA*

Schulz I, Engel C, Niestroj AJ, Kehlen A, Rahfeld JU, Kleinschmidt M, Lehmann K, Roßner S & Demuth HU (2014) A non-canonical function of eukaryotic elongation factor 1A1: Regulation of interleukin-6 expression. *Biochim. Biophys. Acta - Mol. Cell Res.*

Scotti MM & Swanson MS (2016) RNA mis-splicing in disease. *Nat. Rev. Genet.*

Shah PS, Zao J & Ali S (2011) Maternal marital status and birth outcomes: A systematic review and meta-analyses. *Matern. Child Health J.*

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B & Ideker T (2003) Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.*

Shen S, Park JW, Lu Z, Lin L, Henry MD, Wu YN, Zhou Q & Xing Y (2014) rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci.*

Da Silva AAM, Simões VMF, Barbieri MA, Bettiol H, Lamy-Filho F, Coimbra LC & Alves MTSSB

(2003) Young maternal age and preterm birth. *Paediatr. Perinat. Epidemiol.*

Sipos PI, Rens W, Schlecht HL, Fan X, Wareing M, Hayward C, Hubel CA, Bourque S, Baker PN, Davidge ST, Sibley CP & Crocker IP (2013) Uterine vasculature remodeling in human pregnancy involves functional macrochimerism by endothelial colony forming cells of fetal origin. *Stem Cells*

Smaczniak C, Li N, Boeren S, America T, Van Dongen W, Goerdayal SS, De Vries S, Angenent GC & Kaufmann K (2012) Proteomics-based identification of low-abundance signaling and regulatory protein complexes in native plant tissues. *Nat. Protoc.*

Smith SD, Choudhury RH, Matos P, Horn JA, Lye SJ, Dunk CE, Aplin JD, Jones RL & Harris LK (2016) Changes in vascular extracellular matrix composition during decidual spiral arteriole remodeling in early human pregnancy. *Histol. Histopathol.*

Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM & Behrens TEJ (2006) {T}ract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*

Snyder SK, Wessner DH, Wessells JL, Waterhouse RM, Wahl LM, Zimmermann W & Dveksler GS (2001) Pregnancy-specific glycoproteins function as immunomodulators by inducing secretion of IL-10, IL-6 and TGF-β1 by human monocytes. *Am. J. Reprod. Immunol.*

Sõber S, Reiman M, Kikas T, Rull K, Inno R, Vaas P, Teesalu P, Marti JML, Mattila P & Laan M (2015) Extensive shift in placental transcriptome profile in preeclampsia and placental origin of adverse pregnancy outcomes. *Sci. Rep.*

Somerset DA, Afford SC, Strain AJ & Kilby MD (1997) Fetal growth restriction and hepatocyte growth factor. *Arch. Dis. Child. Fetal Neonatal Ed.*

Steger D, Berry D, Haider S, Horn M, Wagner M, Stocker R & Loy A (2011) Systematic spatial bias in DNA microarray hybridization is caused by probe spot position-dependent variability in lateral diffusion. *PLoS One*

Strauss JF (2013) Extracellular matrix dynamics and fetal membrane rupture. *Reprod. Sci.*

Struwe E, Berzl G, Schild R, Blessing H, Drexel L, Hauck B, Tzschoppe A, Weidinger M, Sachs M, Scheler C, Schleussner E & Dötsch J (2010) Microarray analysis of placental tissue in intrauterine growth restriction. *Clin. Endocrinol. (Oxf).*

Sun GY, Shelat PB, Jensen MB, He Y, Sun AY & Simonyi A (2010) Phospholipases A2 and inflammatory responses in the central nervous system. *NeuroMolecular Med.*

Takagi M, Yamamoto D, Ogawa S, Otoi T, Ohtani M & Miyamoto A (2008) Messenger RNA expression of angiotensin-converting enzyme, endothelin, cyclooxygenase-2 and prostaglandin synthases in bovine placentomes during gestation and the postpartum period. *Vet. J.*

Tamsen L (1984) Pregnancy???Specific ??1???Glycoprotein (Spo Levels Measured by Nephelometry in Serum from Women with Vaginal Bleeding in the First Half of Pregnancy. *Acta Obstet. Gynecol. Scand.*

Tan H, Yi L, Rote NS, Hurd WW & Mesiano S (2012) Progesterone receptor-A and -B have opposite effects on proinflammatory gene expression in human myometrial cells: Implications for progesterone actions in human pregnancy and parturition. *J. Clin. Endocrinol. Metab.*

Thiriet M (2013) Preamble to Cytoplasmic Protein Kinases. In *Intracellular Signaling Mediators in the Circulatory and Ventilatory Systems*

Timmons B, Akins M & Mahendroo M (2010) Cervical remodeling during pregnancy and parturition. *Trends Endocrinol. Metab.*

Timpl R, Sasaki T, Kostka G & Chu ML (2003) Fibulins: A versatile family of extracellular matrix proteins. *Nat. Rev. Mol. Cell Biol.*

Trewavas A (2006) A brief history of systems biology. 'Every object that biology studies is a system of systems.' Francois Jacob (1974). *Plant Cell* **18:** 2420–2430 Available at: https://www.ncbi.nlm.nih.gov/pubmed/17088606

Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson IM, Edlund K, Lundberg E, Navani S, Szigyarto CAK, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, et al (2015) Tissue-based map of the human proteome. *Science (80-.).*

Uzun A, Laliberte A, Parker J, Andrew C, Winterrowd E, Sharma S, Istrail S & Padbury JF (2012) DbPTB: A database for preterm birth. *Database*

Vaisbuch E, Romero R, Erez O, Mazaki-Tovi S, Kusanovic JP, Soto E, Dong Z, Chaiworapongsa T, Kim SK, Ogge G, Pacora P, Yeo L & Hassan SS (2010) Activation of the alternative pathway of complement is a feature of pre-term parturition but not of spontaneous labor at term. *Am. J. Reprod. Immunol.*

Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, et al (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17:** 98–110

Vinketova K, Mourdjeva M & Oreshkova T (2016) Human Decidual Stromal Cells as a Component of the Implantation Niche and a Modulator of Maternal Immunity. *J. Pregnancy*

Visiedo F, Bugatto F, Carrasco-Fernández C, Sáez-Benito A, Mateos RM, Cózar-Castellano I, Bartha JL & Perdomo G (2015) Hepatocyte growth factor is elevated in amniotic fluid from obese women and regulates placental glucose and fatty acid metabolism. *Placenta*

Waldorf KMA, Singh N, Mohan AR, Young RC, Ngo L, Das A, Tsai J, Bansal A, Paolella L, Herbert BR, Sooranna SR, Gough GM, Astley C, Vogel K, Baldessari AE, Bammler TK, MacDonald J, Gravett MG, Rajagopal L & Johnson MR (2015) Uterine overdistention induces preterm labor mediated by inflammation: Observations in pregnant women and nonhuman primates. *Am. J. Obstet. Gynecol.*

Walsh S (2011) Prostaglandins in pregnancy. *Glob. Libr. Women's Med.*

Wang H & Stjernholm YV (2007) Plasma membrane receptor mediated MAPK signaling pathways are activated in human uterine cervix at parturition. *Reprod. Biol. Endocrinol.*

Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, DeCarvalho AC, Lyu S, Li P, Li Y, Barthel F, Cho HJ, Lin Y-H, Satani N, Martinez-Ledesma E, Zheng S, Chang E, Sauvé C-EG, Olar A, Lan ZD, et al (2017) Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell*

Wang Y, Liu J, Huang B, Xu Y, Li J, Huang L, Lin J, Zhang J, Min Q, Yang W & Wang X (2014) Mechanism of alternative splicing and its regulation (Review). *Biomed. Reports*

Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD & Morris Q (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38:** W214-20

Westermark B (2014) Platelet-derived growth factor in glioblastoma-driver or biomarker? *Ups. J. Med. Sci.*

Wolf KJ, Lee S & Kumar S (2018) A 3D topographical model of parenchymal infiltration and perivascular invasion in glioblastoma. *APL Bioeng.*

Xu J, Chi F, Guo T, Punj V, Paul Lee WN, French SW & Tsukamoto H (2015) NOTCH reprograms mitochondrial metabolism for proinflammatory macrophage activation. *J. Clin. Invest.*

Zhang W (2005) Genetics: Analysis of Genes and Genomes, Sixth Edition. *Neuro. Oncol.*