Human Gene Expression Variability and its Dependence on Methylation and Aging

Nasser Bashkeel

Thesis submitted to the Faculty of Graduate
and Postdoctoral Studies in partial fulfillment
of the requirements for the Master of Science
in Biochemistry

Department of Biochemistry, Microbiology, and Immunology
Faculty of Medicine
University of Ottawa

© Nasser Bashkeel, Ottawa, Canada, 2019

Abstract

The phenotypic variability in human populations is partly the result of gene polymorphisms and differential gene expression. Studying the variability of gene expression across human populations is essential to understanding the molecular basis for diversity. However, key issues remain unanswered with respect to human expression variability. For example, the role of gene methylation in expression variability is uncertain, nor is it clear what role tissue-specific factors may have. Moreover, the contribution that expression variability has in aging and development is unknown. Here we classified human genes based on their expression variability in normal human breast and brain samples and identified functional aspects associated with high and low expression variability. Interestingly, both high variability and low variability gene sets are enriched for developmentally essential genes. There is limited overlap between the variably expressed genes of different tissues, indicating that tissue-specific rather than individual-specific factors are at work. We also find that methylation likely has a key role in controlling expression variability insofar as genes with low expression variability are likely to be non-methylated. Importantly, we find that genes with high population expression variability are likely to have age-, but not sex-dependent expression. Taken together, our work indicates that gene expression variability is tissue-specific, methylation-dependent, and is an important component of the natural aging process.

Acknowledgements

I would like to thank Dr. Jonathan Lee for providing me the opportunity to work on such a promising and exciting project. I am grateful for his invaluable counsel that enabled me to think and solve many problems I was faced with during my research. His diligence, honesty, and approachability inspired me to mature not only as a researcher, but also as a person.

I had the pleasure of working with some incredible people in the Lee lab. To lab-mates, both past and present, thank you for the support, advice, and friendship over the years. I'd also like to thank my thesis committee, Dr. Mads Kaern and Dr. Theodore Perkins. Their valuable comments, guidance, and advice over the years greatly shaped my work. Finally, I would like to thank all my friends and family who have supported me throughout my graduate studies.

My work was supported by a scholarship from the UAE Ministry of Higher Education and Scientific Research.

TABLE OF CONTENTS

Abstract ii
Acknowledgements iii
List of Abbreviations vi
List of Figures viii
List of Tables ix
List of Supplementary Figures ix
1: Introduction 1
1.1: Biological Variability as an Informative Metric 1
1.1.1: Phenotypic Variation 1
1.1.2: Natural Selection and Coupling Between Variability and Plasticity 2
1.1.3: Disease Susceptibility, Development, and Progression 3
1.1.4: Embryonic development 6
1.2: Potential Sources of Expression Variability (EV) 8
1.2.1: Tissue specificity 8
1.2.2: Gene structural features 8
1.2.3: Gene Function & Essentiality 9
1.2.4: DNA Methylation 10
1.2.5: Aging 11
1.2.6: Sex Differences in Gene Expression 12
1.3: Measures of Variability 13
1.3.1: Intrinsic, Extrinsic, and Technical Sources of Variability 13
1.3.2: Differential Expression Analyses 14
1.3.3: Common Measures of Statistical Variability 16
1.4: Rationale & Objectives 18
2: Materials and Methods 20
2.1: Gene Expression & Methylation Microarray Datasets 20
2.2: Preprocessing the Datasets 21
2.3: Identification and Removal of Bimodal Expression Genes 22
2.4: Estimating Expression Variability 24
2.5: EV Gene Set Classification 25
2.6: Bootstrapping EV Gene Set Classifications 25
2.7: Structural Analysis of EV Genes 26
2.8: Gene Set Functional Enrichment Analysis 26

2.9: Enrichment Analysis ... 26
2.10: Hierarchical Clustering of Age-Dependent Hypervariable Genes ... 27
3: Results ... 28
3.1: Estimating Expression Variability ... 28
3.1.1: Estimating EV ... 28
3.1.2: Classification of EV Gene Sets ... 30
3.1.3: Cross-Validation of EV Classifications ... 34
3.1.3: Statistical Nature of Hypervariability ... 37
3.1.5: Tissue Specificity of EV ... 37
3.1.6: EV and Gene Structural Characteristics ... 39
3.1.7: Functional Enrichment of EV Gene Sets ... 47
3.2: Essentiality Enrichment in Variable Genes ... 49
3.3: DNA Methylation and Expression Variability ... 53
3.3.1: Gene Methylation Clustering ... 53
3.3.2: Methylation Enrichment in EV Classes ... 58
3.4: Effects of Age, Sex, and PMI on Variability ... 61
3.4.1: Linear Regression of Sex, PMI, and Aging (and EV) ... 61
3.4.2: Heatmap Clustering ... 63
3.4.3: Functional Enrichments of Age-Regulated Hypervariable Genes ... 65
4: Discussion ... 68
4.1: Characteristics of EV ... 70
4.2: Essentiality & Methylation ... 73
4.3: Effects of Aging on EV ... 76
Conclusion ... 81
References ... 83
Appendices ... 95

List of Abbreviations

AU	Approximately Unbiased
BIC	Bayesian Information Criterion
CCDS	Consensus Coding DNA Sequence Project
CLL	Chronic Lymphocytic Leukemia
CRISPR	Clustered regularly interspaced short palindromic repeats
CV	Coefficient of Variation
EGA	European Genome-Phenome Archive
EM	Expectation-maximization algorithm
EV	Expression variability
FDR	False discovery rate
GEO	Gene Expression Omnibus
GO	Gene Ontology
HIV	Human immunodeficiency virus
MAD	Median absolute deviation
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
NABEC	North American Brain Expression Consortium
PDF	Probability density function
PMI	Post-mortem interval
PPI	Protein-protein interactions
scRNA-Seq	Single-cell RNA sequencing
SD	Standard Deviation

TBI	Traumatic brain injury
TF	Transcription factors
UKBEC	UK Human Brain Expression Consortium
WSS	Within-clusters sum of squares
β-gal	Beta-galactosidase
β-value	Beta-value

List of Figures

Figure 1. Overall analysis pipeline. 19

Figure 2. Hierarchical clustering of brain dataset samples 23

Figure 3. EV in human breast, cerebellum, and frontal cortex tissue 31

Figure 4. Relationship between EV and expression........ 32

Figure 5. EV probability density functions........ 33

Figure 6. EV gene set classification accuracy 35

Figure 7. Number of genes in EV gene sets 36

Figure 8. Modality of Hypervariable gene expression distribution 38

Figure 9. Tissue-specificity of EV........ 40

Figure 10. EV gene set overlap 41

Figure 11. Effect of genomic position on EV 44

Figure 12. Chromosome combinations cosine similarity 45

Figure 13. Paired chromosome cosine similarity between tissues........ 46

Figure 14. Expression and methylation correlation plot 54

Figure 15. Gaussian mixture model of gene methylation........ 56

Figure 16. PDF of gene methylation 57

Figure 17. PDF of average gene methylation by EV classification........ 59

Figure 18. Optimal number of K-means age clusters 64

Figure 19. Cerebellum Hypervariable heatmap........ 66

Figure 20. Frontal Cortex Hypervariable heatmap........ 67

List of Tables

Table 1. Structural analysis of genes as a function of EV gene classification 42

Table 2. Common GO annotations of EV gene sets 48

Table 3. Unique GO annotations of EV gene sets 48

Table 4. Essentiality enrichment analysis 51

Table 5. Common GO Annotations of essential genes 51

Table 6. Unique GO annotations of essentiality genes 52

Table 7. Methylation enrichment analysis 60

Table 8. Gene-wise multiple linear regression of Hypervariable EV and PMI, sex, and age 62

List of Supplementary Figures

Supplementary Figure 1. Structural analysis of genes as a function of EV 95

Supplementary Figure 2. Complete list of GO terms for all genes 105

Supplementary Figure 3. Complete list of GO terms for essential genes 112

Supplementary Figure 4. Complete list of GO terms for age-regulated Hypervariable genes ... 115

Supplementary Figure 5. Complete list of GO terms for pseudo-Hypervariable and pseudo-Hypovariable genes 121

1: Introduction

1.1: Biological Variability as an Informative Metric

1.1.1: Phenotypic Variation

Phenotypes can be defined as the set of observable traits of an individual organism, resulting from an organism's genotype, environment, and the specific interactions between the two. Due to the existence of these dynamic factors, there is a large amount of variability in phenotypes in any population. These phenotypes can occur in a continuous range or gradient, such as height or skin colour. Phenotypes can also occur discretely, as is the case for blood types.

Phenotypic variability in human populations is partly the result of differential gene expression. These patterns of gene expression vary between individuals in a population, across different cell types in a single individual (Alemu et al., 2014), and even within cells in a genetically homogenous population (Roberfroid et al., 2016). As such, many studies have been conducted to understand the molecular basis of phenotypic diversity. In particular, mapping out how gene expression and gene regulation vary across individuals provides a deeper insight into the function of healthy tissues and the molecular origins of diseases (Consortium et al., 2017).

The sources of phenotypic variation can be divided into several subcategories. One such source of gene expression differences is phenotypic plasticity, whereby an organism adjusts its' phenotype to a variable environment (Chen et al., 2017a). Two commonly cited examples include coat color changes in agouti gene mice caused by dietary modifications (Cooney et al., 2002; Feinberg and Irizarry, 2010; Waterland and Jirtle, 2003), and methylation of axin-fused allele in kinked tail mice (Rakyan et al., 2003; Waterland and Jirtle, 2003). These changes in the genome

are then transmitted intergenerationally, suggesting that these changes in phenotype act to redirect natural selection pressures (Cooke, 1995). Another source of phenotypic variability is stochastic fluctuations, or noise, in gene expression and can occur in genetically identical cells in a constant environment (Newman et al., 2006; Silander et al., 2012; Singh, 2013; Taniguchi et al., 2010). The level of variability, or noise, differs across genes, whereby some genes are noisier than others. This variability in gene expression contributes to phenotypic variability in both clonal and heterogeneous populations.

Gene expression variability has been emerging as an informative metric of a phenotypic state, particularly as it relates to human disease (Ecker et al., 2015; Ho et al., 2008; Li et al., 2010a; Mar et al., 2011). These studies have correlated gene expression variability with a wide variety of biological and disease-specific phenomena including evolutionary fitness (Feinberg and Irizarry, 2010), reduced penetrance (Eichler et al., 2010; Manolio et al., 2009), disease susceptibility (Li et al., 2010a), and embryonic development (Hasegawa et al., 2015; Kalmar et al., 2009; Piras et al., 2014).

1.1.2: Natural Selection and Coupling Between Variability and Plasticity

Biological systems are subject to both internal and external factors referred to as noise or stochastic fluctuations. Without these biological variations, cells cannot react to a changing environment and therefore perform their regular function. As such, there is an optimal signal-to-noise ratio inherent in biological systems (Kemkemer et al., 2002).

While changes in expression levels of some genes may not cause apparent phenotypic effects, variability in expression of other genes may reduce organismal fitness. These detrimental

changes in expression levels are bidirectional, whereby decreases (Deutschbauer et al., 2005) and increases (Giaever et al., 2002; Sopko et al., 2006) in gene expression can both be potentially harmful.

In the context of natural selection and evolution, if a new expression level of a gene is advantageous, natural selection will push the population expression levels towards the new optimal level. This causes a significant reduction in expression variability for the gene. Due to selective pressures, expression variability of dosage-sensitive genes is expected to be minimized (Lehner, 2008; Li et al., 2010a).

In a dynamic, changing environment, increased variability in gene expression substantially increases the fitness of a population, even with no change in mean phenotype. However, we do not know how much variability can be tolerated in dosage-sensitive genes. Therefore, selection would be determined largely by the ability to vary around the optimal level, or 'setpoint', rather than by the setpoint itself (Feinberg and Irizarry, 2010).

1.1.3: Disease Susceptibility, Development, and Progression

The role of genes in disease susceptibility is traditionally viewed from a variant- or sequence-specific perspective. While extensive studies have been done to characterize the effects of sequence variants on disease phenotypes, not all variability in disease susceptibility is attributable to genetic factors (Feinberg and Irizarry, 2010). As more evidence is uncovered regarding the importance of gene expression variability on phenotypic variability in a population, mapping variance profiles of populations may reveal aberrant genetic events that contribute to disease phenotypes (Mar et al., 2011).

The vast majority of human genes exhibit low within-population variability, minimizing differential expression of genes between populations. In the case of genes that are differentially expressed, a vast majority exhibit a large shift in mean expression levels without affecting the expression variability in each population. Of the genes that do exhibit differential variability and differential expression, many of them are associated with a number of human diseases, including heart disease, hepatocellular carcinoma, and psoriasis (Li et al., 2010a).

Recent studies of the monocyte transcriptome suggest that expression variability of disease-associated genes plays a greater role on disease susceptibility than sequence variants, regardless of the source of the variability. The correlation between transcriptome variability and disease susceptibility can be viewed from differing perspectives. The first is that these variations in expression levels reflect the cumulative effects of genomic and environmental interactions over the organism's life and simply report on the state of biological processes associated with diseases. The second perspective is that gene expression variability is pathophysiologically relevant and plays a direct role in susceptibility to disease (Zeller et al., 2010). Many new studies support the second perspective, suggesting that gene expression variability plays a crucial role in disease susceptibility across populations directly (Feinberg and Irizarry, 2010; Hansen et al., 2011; Landau et al., 2014), including human immunodeficiency virus (HIV) susceptibility (Li et al., 2010a), neurological disorders (Li et al., 2010a; Mar et al., 2011), and oncogenesis (Afsari et al., 2014; Ecker et al., 2015).

In spite of the complex and heterogeneous nature of cancers, there are six overarching biological capabilities that are acquired during the development of human tumours. These hallmarks include sustaining proliferative signalling, evading growth suppressors, resisting cell death, enabling replicative immortality, and activating invasion and metastasis (Hanahan and

Weinberg, 2000). At the core of all these hallmarks is genome instability caused by acquisition of somatic alterations and mutations at random locations in the genome that fuels cancer plasticity and phenotypic variability. (Hanahan and Weinberg, 2011). During the course of the disease, cancers generally become more heterogenous and display cells with distinct morphological and phenotypic profiles within a single tumour. These increases in intra-tumoral heterogeneity play a crucial role in tumour progression and evolution, and has a significant impact on clinical outcomes (Landau et al., 2014).

The effects of expression variability on disease susceptibility stems from both host and pathogen heterogeneity. Studies of λ phage, a bacterial virus, have demonstrated that the control of the phage's lysis/lysogeny choice is regulated through stochastic thermal fluctuations (Arkin et al., 1998). Common gene regulatory motifs, such as positive feedback loops, can amplify these fluctuations to drive phenotypic variability (Weinberger et al., 2005). In addition to viral latency, host phenotypic variability also affects disease susceptibility. While host factors involving immune system remain the focus of HIV research, non-immunity genes also interact with viral genes. These non-immunity host-viral gene interactions control viral entry, progression, and replication cycles, and exhibit substantially elevated expression variability (Li et al., 2010a). In addition to identifying sequence variants, studying the expression variability of genes will broaden the scope of candidate disease genes, thereby reducing susceptibility to diseases such as HIV.

In addition to identifying novel disease-associated genes, expression variability can be used to correctly identify distinct disease subtypes, including chronic lymphocytic leukemia (CLL) (Ecker et al., 2015). More aggressive types of CLL exhibit higher gene expression variability that is associated with cancer heterogeneity, conferring a high level of tumour adaptability and drug resistance.

While different diseases, such as schizophrenia and Parkinson's disease, exhibit vastly different expression variability profiles, both global disease states are considered abnormal relative to normal tissue profiles. Parkinson's disease is associated with an increase in expression variability of core signalling pathways while schizophrenia displays decreases in variability of similar signalling pathways (Mar et al., 2011). Interestingly, these highly constrained pathways in schizophrenic patients are largely involved in cancer. Although speculative, the reduction in expression variability in these oncogenes is thought to confer protection against lung and colorectal cancers, despite increased smoking (Zammit et al., 2003) and drinking habits (Fowler et al., 1998) in schizophrenic patients. In addition, mouse models of schizophrenia have shown a reduced tumour growth (Asada et al., 2008), supporting the hypothesis that a reduction in expression variability of these cancer-associated pathways plays a role in inhibiting cancer development.

1.1.4: Embryonic development

A key intrinsic feature of embryonic stem cells is their phenotypic heterogeneity at each stage of embryonic development. Through the use of single cell technologies such as scRNA-Seq, it has become possible to examine the transcriptome of a developing embryo as it transitions through each developmental stage. By examining inter-cellular gene expression variability during development, functional insights into stage-specific markers have been uncovered that show different degrees of transcriptional regulatory control (Hasegawa et al., 2015). One example of stage-specific markers in embryonic stem cell populations is the distribution of Nanog expression states. Nanog null embryos are not viable beyond implantation as the cells do not progress into pluripotency (Blinka and Rao, 2017; Darr et al., 2006; Mitsui et al., 2003; Silva et al., 2009). However, embryonic stem cells with low levels of Nanog expression exhibit increased global gene

expression variability and are prone to differentiation (Chambers et al., 2007; Kalmar et al., 2009). This suggests that pluripotency of embryonic stem cells is a state of dynamic cellular heterogeneity driven by transcriptional variability of key regulators, including Nanog and GATA6.

As the embryo develops past the 2-cell stage, global expression variability drastically increases with each phase, generating significant gene expression heterogeneity between individual cells (Hasegawa et al., 2015; Piras et al., 2014). This global increase in gene expression variability overlaps with a phenomenon described by other studies as "waves of transcriptional activation" (Vassena et al., 2011). These series of successive waves involve a hierarchical activation of genes associated with the regulation of pluripotency, further supporting the fact that gene expression variability drives pluripotency and is necessary for cell fate diversification (Chang et al., 2008; Maamar et al., 2007).

These studies suggest that gene expression variability in pluripotent embryonic stem cells is both predictable and essential. The range of expression values for specific genes associated with regulatory pathways are directly linked to their potential for differentiation, allowing a sub-population of embryonic stem cells continuously primed for differentiation without being pre-committed to a particular fate for a long period of time (Kalmar et al., 2009; Mar et al., 2011). These changes in expression variability in differentiating pluripotent cells lay the foundations for tissue-specific gene expression and tissue identity.

1.2: Potential Sources of Expression Variability (EV)

1.2.1: Tissue specificity

Although all tissue types of an organism carry out common processes that are essential for survival, each tissue type displays a unique gene expression profile that defines their tissue-specific phenotype. These tissue-specific gene expression differences are the result of distinct regulatory networks that control tissue specificity, identity, and function (Sonawane et al., 2017). While many studies have been conducted to characterize these tissue-specific expression profiles, expression variability profiles of different human tissue types are still largely undefined. By developing baseline expression variability profiles of healthy tissue, we can deepen our understanding of the dysregulation of gene networks that occurs during aberrant phenotypic states, including disease development and disease progression in individual tissue types.

1.2.2: Gene structural features

Similar to tissue-specific expression levels, expression variability is likely an inherent and heritable property of a gene. As such, EV is subject to selective pressures and is linked with evolvability of complex organisms (Alemu et al., 2014; Feinberg and Irizarry, 2010; Lehner, 2008).

In a study conducted by Alemu et al., a comprehensive investigation of genetic and epigenetic correlates of EV across 41 human tissue types showed significant relationships between various structural genetic features and expression variability. These genetic features included gene length, number of exons, transcript length, number of transcripts, copy number variation, and promoter composition and conservation. Interestingly, they reported significant monotonic

relationships between EV and genetic features, suggesting that the physical structure of a gene is directly involved in the regulation of both highly and lowly constrained genes.

1.2.3: Gene Function & Essentiality

Expression variability has also been shown to be associated with the functional role and physical location of a gene product through specific transcriptional constraints, or lack thereof. Generally, genes that respond to extracellular cues exhibit high expression variability and are located at the cell periphery. On the other hand, genes whose expression is tightly regulated are generally involved in cellular house-keeping processes and are located in cell interior (Alemu et al., 2014; Mar et al., 2011; Newman et al., 2006).

These housekeeping genes are typically constitutively expressed across all tissue types and are required for basic cellular function (Butte et al., 2001; Eisenberg and Levanon, 2003; Zhu et al., 2008). Housekeeping genes tend to form a core in human protein-protein interaction (PPI) networks, while genes with tissue-specific functions occupy more peripheral positions in a PPI network (Lin et al., 2009). The connectivity of a gene in a PPI network determines the functional constraints and inversely affects expression variability (Li et al., 2010b; Mar et al., 2011).

In addition to housekeeping genes, any gene required for the survival of an organism is considered "essential". Fluctuations in the abundance of essential proteins can be highly deleterious if the dosage declines too low relative to the optimal level. As such, we expect all genes that are essential to be tightly regulated and exhibit lower noise than non-essential genes (Batada and Hurst, 2007; Lehner, 2008; Newman et al., 2006). This noise minimization process prevents harmful stochastic variation in gene expression levels, but also limits the evolvability of complex

biological systems by constraining the ability of gene expression to respond to non-stochastic variation. This cost-benefit conflict between noise and plasticity is thought to be an evolvable trait, suggesting that gene-specific noise-plasticity coupling is likely constrained by selection (Lehner, 2010; Singh, 2013).

However, essentiality of a gene is not always a static or intrinsic property of a gene. It is highly dependent of a variety of factors including gene function, host genome, the cellular environment, and other settings. Genes with variable essentiality statuses under different cellular conditions are referred to as "conditionally essential genes" (Chen et al., 2017b). Consequently, the essentiality phenotype should be viewed contextually based on the type of experiment (D'Elia et al., 2009). As such, we limited our analysis to genes with consistent essentiality status by eliminating conditionally essential genes.

1.2.4: DNA Methylation

Over the last two decades, it has become clear that epigenetic modifications of DNA and histones can play a primary role in determining phenotypic outcomes, including human disease (Feinberg and Irizarry, 2010; Kaminsky et al., 2009). These epigenetic modifications include DNA methylation which plays a complex role in regulating gene expression. Methylation of DNA alters the transcriptional activity of DNA through the addition of a methyl group without changing the DNA sequence itself. Similar to gene expression, DNA methylation is highly variable at the cell, tissue, and individual level (Zhang et al., 2013). Distinct DNA methylation patterns can arise even within clonal populations and monozygotic twins (Kaminsky et al., 2009). Genes that are differentially methylated in genetically identical organisms are either silenced or activated, leading to phenotypic variation with a given genotype.

While its exact role in gene expression remains unclear, precise DNA methylation is essential for cell differentiation and embryonic development. DNA methylation patterns exhibit partial stability that can change rapidly based on a number of factors, including developmental programs, environmental factors, hormones, and stochastic events (Jaenisch and Bird, 2003; Jirtle and Skinner, 2007; Riggs and Xiong, 2004; Ushijima et al., 2003).

Phenotypic variation can also be affected by DNA methylation without affecting the mean phenotype of a population. Specifically, genetic variants that do not change the mean phenotype can still affect phenotypic variability through epigenetic mechanisms. New models suggest that heritable genetic variations affect stochastic phenotypic variation by contributing to gene expression variance but not mean phenotype (Feinberg and Irizarry, 2010). This presents an interesting departure from neo-Darwinian and classical population genetics principles that postulates that heritable phenotypic variation is due entirely to the cumulative effects of individual trait loci. Feinberg and Irizarry posit that the phenotypic variability is the heritable component, and selection is determined by the ability of genes to vary their expression around a setpoint, rather than by the setpoint itself.

1.2.5: Aging

Aging can be characterized by the progressive functional decline of tissues, and is associated with changes in cellular morphology and cell count (Zahn et al., 2007). While specific genes are uniformly affected by age across a range of tissues, aging is largely a tissue-specific process with unique aging signature profiles (Glass et al., 2013). One potential source of age-specific expression changes is the accumulation of somatic DNA damage throughout an organism's lifetime, whereby higher levels of accumulated genome damage results in increased transcriptional variability (Bahar

et al., 2006). Cells in skin tissue are exposed to high levels of exogenous sources of DNA damage, such as UV light and various chemical agents, and exhibit extensive age-related gene expression changes (Glass et al., 2013). While many studies are being conducted to pinpoint the effects of aging on specific genes and pathways, the role of aging on gene expression variability remains unclear. Since aging can be described as a form of cellular functional decline, we expect to see a tissue-specific dysregulation of highly-constrained pathways resulting in increased gene expression variability.

1.2.6: Sex Differences in Gene Expression

The basis of mammalian sexual differentiation is the XY sex-determination system and is mediated by the effects of sex hormones. There is strong evidence that demonstrates sex differences in brain structure, neurochemistry, behaviour, and susceptibility to neurodegenerative and neuropsychiatric disease in humans and many other animals (Cahill, 2006; Cosgrove et al., 2007; Jazin and Cahill, 2010). Sexually dimorphic gene expression is highly tissue specific, and can range from tens to thousands of genes in mouse liver, kidney, blastocysts, lacrimal gland, prenatal brain, and adult brain substructures (Amador-Noguez et al., 2005; Clodfelter et al., 2006; Dewing et al., 2003; Kobayashi et al., 2006; Lauber et al., 1991; Richards et al., 2006; Rinn et al., 2004; Vawter et al., 2004; Xu et al., 2006).

These sex differences in gene expression are also widespread in humans and likely to have functional consequences relevant to human disease (Trabzuni et al., 2013). While the exact molecular basis remains unclear, these sex differences stem from genetic and transcriptional regulatory mechanisms of gene expression (Yang et al., 2006), and may contribute to differences in expression variability and phenotypic variability.

1.3: Measures of Variability

1.3.1: Intrinsic, Extrinsic, and Technical Sources of Variability

Often thought of in non-biological terms, variability in gene expression measurements is typically attributed to technical or background noise. Indeed, there are many non-biological or technical sources of variability, including experimental equipment, fluorescence quality, probe hybridization strength, and batch effects (Bryant et al., 2011; Liggett, 2006; McCall et al., 2016). If technical noise was indeed the primary driver of variability, we would expect the equal effects across all experimental samples. This is not the case as different samples exhibit different gene expression variability profiles that correlate to distinct phenotypic states.

To minimize the effects of technical variability, internal replication are built into the experimental design, and include several practices such as multiple probes per gene and replicated experiments for each sample (Liu and Rattray, 2010).

After taking precautions to minimize and account for technical variability, the remaining variability in gene expression can be attributed to biological sources. The first source of biological gene expression variability is directly related to extrinsic factors affecting the cell sample. In a large multi-cellular organism such as a human, these factors include cellular age and the cell cycle stage (Newman et al., 2006), the physical environment of the cell, such as nutrient availability, endocrine signalling, immune responses, and systemic inflammation (Elsamanoudy et al., 2016; Gil et al., 2007; Pritchard et al., 2001), and organelle distributions, such as the number and functionality of mitochondria (Johnston et al., 2012).

Stochastic, or random and probabilistic, noise is the second source of variability in gene expression (Ansel et al., 2008; Kærn et al., 2005; Raser and O'Shea, 2004). This phenomenon has been observed as early as 1957, when Novick and Weiner showed that expression of beta-galactosidase (β-gal) in cells is random and highly variable (Novick and Weiner, 1957). Their study showed that induction of β-gal through the addition of lactose resulted in a higher proportion of cells expressing the β-gal enzyme rather than a uniform increase in expression levels of β-gal across the cell population. Many cellular processes, including gene expression, rely on collision between the biochemical reactants to initiate the chemical reactions; also known as the collision theory. Given the diffusive nature of cells, it is clear that gene expression is stochastic and inherently highly variable.

Despite the complex nature of gene expression variability, many studies have provided evidence that biological sources of variability play an important role in cellular and organismal phenotypes. Rather than associating gene expression variability with technical artefacts, it should be examined as an intrinsic property reflecting phenotypic heterogeneity. In recent years, studies have started to focus on variability in gene expression within populations from a biological perspective. By modelling changes in gene expression variability, new regulatory or differential control pathways have been discovered that would have otherwise been missed using traditional analyses of average gene expression (Ho et al., 2008; Yu et al., 2008).

1.3.2: Differential Expression Analyses

With the advancements in high throughput sequencing technology, thousands of genes can be analyzed from a single sample simultaneously using microarrays and provide qualitative and quantitative profiles of gene expression. In a typical gene expression profiling experiment,

expression data is generated from a population of samples in two (or more) phenotypic states, for example a cancerous and non-cancerous sample of a tissue. Various differential gene expression algorithms, such as edgeR and DESeq, are then used to identify genes whose expression means differ significantly between the phenotypic states. In addition, expression of certain genes appeared to be correlated, sparking the examination of gene co-expression. In these differential co-expression analyses, the goal is to identity gene sets whose expression is always co-expressed in one phenotypic state but not another (Choi et al., 2005; Ho et al., 2008; Kostka and Spang, 2004; Watson, 2006).

There still remains some challenges in using microarray technologies to study complex, multi-cellular organisms. First, different tissues are comprised of different cell types at different proportions in different samples. Second, the environment of a tissue sample cannot be fully controlled. This also applies to genetically identical samples treated under identical conditions, as their hormonal profile, state of the immune system, and degree of inflammation will differ between individual samples (Pritchard et al., 2001). These factors concertedly cause inherent biological variability in gene expression across a population, homogenous or otherwise.

While both differential expression and differential co-expression analyses have successfully been used to identify novel disease-related genes and underlying pathogenic molecular events, the identifiable genes of interest are limited to those with significant changes in mean expression levels between two phenotypic states. The statistical methods used in both analyses assume the gene expression variance across the population as a measure of experimental reproducibility and statistical significance. However, gene expression levels tend to be similar between technical replicates but vary between individuals in a population. The existence of these gene expression

variability patterns in clonal populations suggests that variability is more than a technical artefact, but rather an important area of study in human diseases.

1.3.3: Common Measures of Statistical Variability

Transcriptional heterogeneity measured across a population can arise from multiple origins. Inter-individual variability can be attributed to individual differences including genetic makeup, age, sex, and overall lifestyle. On the other hand, intra-individual variability originates in response to stochastic processes as well as in response to external stimuli, such as cell activation and communication (Choi and Kim, 2009; Dong et al., 2011; Lehner and Kaneko, 2011).

While RNA-seq and other next generation sequencing platforms are rising in popularity and use in recent years, microarrays remain the most commonly used sequencing tool as they provide a robust and relatively inexpensive method of sequencing as well as an abundance of established and easy-to-use software packages. Despite the multitude of microarray studies examining the effects of gene expression variability, the choice for the best statistical practices involving differential variability are still evolving (Hasegawa et al., 2015). While different studies use different measures of variability and statistical methods to determine differentially variable genes, there are several commonly used metrics. The first classic measure of variability is the standard deviation (SD or σ). This statistic measures the dispersion of a dataset relative to its mean and is calculated by determining the variation between each data point relative to the mean. The formula for the sample standard deviation is given by

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}} \tag{1}$$

Standard deviation is usually the best choice for assessing dispersion for normal data. However, there is several limitations when using SD as a measure of gene expression variability. First, many gene expression profiles across a population are highly non-normal. Second, SD does not take expression magnitude into account. This becomes problematic when dealing with genomic data, as statistically significant differences in dispersion may not yield biological changes within a cell population. Lastly, it is sensitive to outliers as the distance from the mean is squared, causing large deviations being weighed more heavily.

The second commonly used measure of variability is the coefficient of variation (CV). Also known as the relative standard deviation, the CV is defined as the ratio of the standard deviation to the mean; where the formula for which is

$$CV = \frac{\sigma}{\mu} \tag{2}$$

This method presents the degree of variability in relation to the mean of the population. While the CV attempts to overcome the limitations of SD by taking the expression magnitude into account, it is subject to zero-inflation for genes with very low levels of average expression.

Rather than measuring gene expression variability using one of the aforementioned measures of variability, we developed our own method of calculating variability in gene expression and a statistical method of classifying genes into differentially variable gene sets based on their expression variability (EV). This modifed EV method utilizes the median absolute deviation (MAD) as a robust measure of variability in gene expression; the formula for which is

$$MAD = median(|X_i - \tilde{X}|) \tag{3}$$

As previously mentioned, SD is a suitable dispersion measure for normally-distributed gene expression data. However, in the presence of non-normal distributions of gene expression, MAD becomes the more suitable alternative to measure dispersion. In addition, a number of outliers become irrelevant to the statistic by virtue of taking the median value of the distances from the median. This results in a robust statistic that is more resistant to outliers than SD or CV. The use of MAD as a measure of variability establishes the basis of the expression variability metric that allows us to examine the effects of EV across the human population. Using MAD as the underlying measure of variability, expression variability can be defined as

$$EV = MAD_{Bootstrapped} - MAD_{Expected} \quad (4)$$

The main feature of expression variability is its ability to account for the relationship between mean expression and variability, and present, on average, no relationship to overall expression. Figure 1 illustrates the overall pipeline of calculating EV.

1.4: Rationale & Objectives

1. With the popularity of microarray experiments and the lack of standardized methods of measuring variability, the first objective is to develop a robust and generalizable method of measuring variability in gene expression. This will be done using a modified method of calculating gene EV and novel EV classification method to detect differentially variable gene sets.
2. The second objective is to investigate potential correlates of gene EV and the effects of gene expression variability on phenotypic variation in a human population.

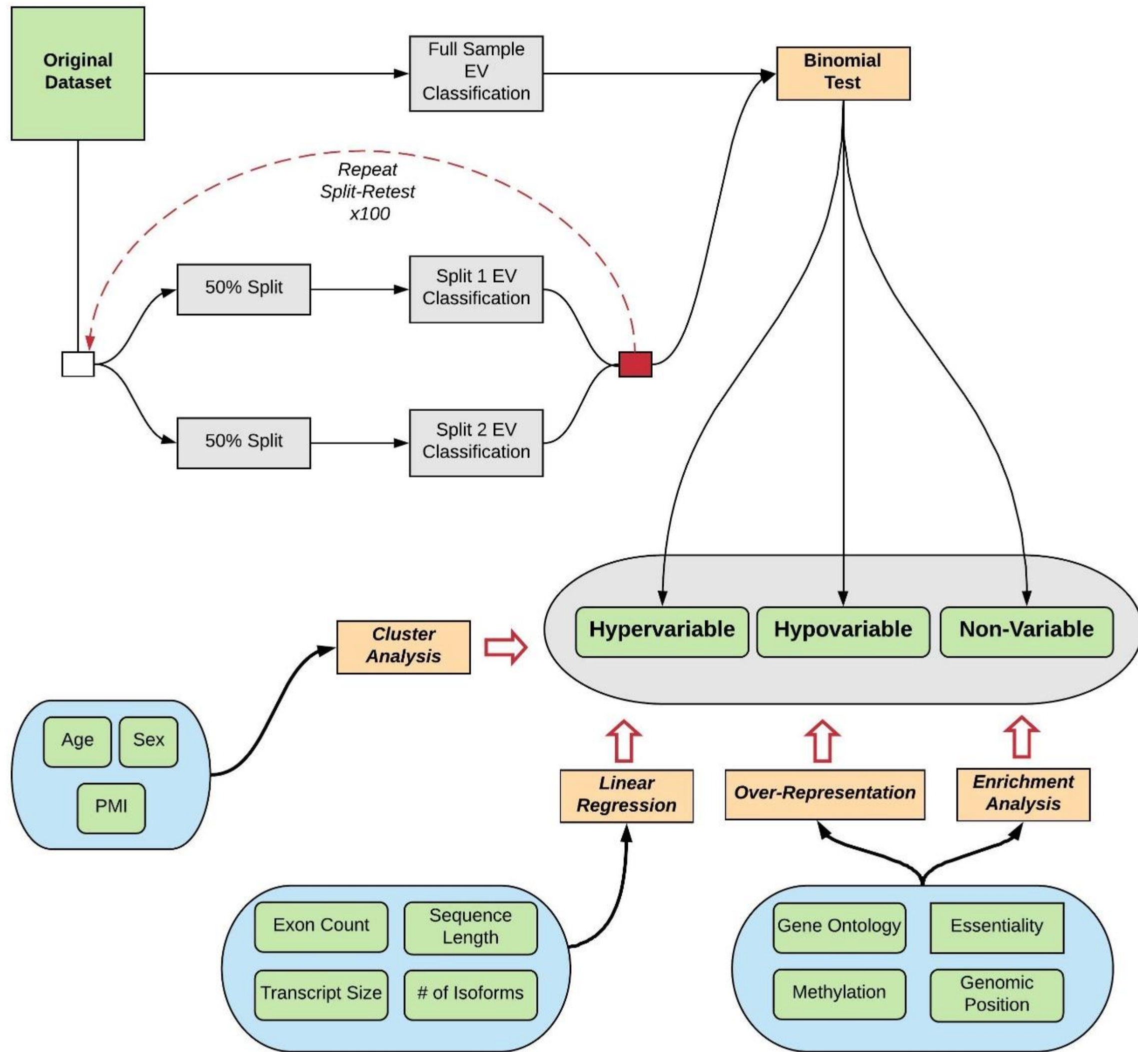


Figure 1. Overall analysis pipeline.

Our goal is to investigate correlates of gene expression variability and the effects of gene expression variability on phenotypic variation in a human population. We obtained and processed gene expression profiles from healthy breast and brain tissue, measured gene EV and classified them into three gene sets. The original dataset was then divided into two equally sized subsets (50% split) and the EV analysis in conducted on each set. The split-retest procedure is conducted 100 times. The resulting 100 EV classifications are then compared to the full sample EV classifications using a binomial test. Finally, we evaluated the relationship between EV and various potential causal factors.

2: Materials and Methods

2.1: Gene Expression & Methylation Microarray Datasets

The EV analyses were conducted on two separate datasets from three different tissue types. Both datasets utilized the Illumina HumanHT-12 V3.0 expression beadchip platform to measure the expression levels of 48,804 probes. In the first expression dataset, the samples were extracted, analyzed, and pre-processed by the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (Curtis et al., 2012). The downloaded expression dataset provided high quality RNA-derived transcriptional profiles of healthy breast tissue from 144 samples. The associated genotype and expression data are deposited at the European Genome-Phenome Archive (EGA, http://www.ebi.ac.uk/ega/), which is hosted by the European Bioinformatics Institute, under accession number EGAS00000000083. The microarray readings were preprocessed using the METABRIC group's own custom script based on existing functionality within the beadarray package in R (Dunning et al., 2007) and were reported as a log2 intensity. This dataset is referred to as breast tissue.

The second expression dataset is catalogued by the North American Brain Expression Consortium (NABEC) and the UK Human Brain Expression Consortium (UKBEC). The expression data was downloaded from the NCBI Gene Expression Omnibus (GEO) database (Barrett et al., 2013) under the accession number GSE36192. A total of 991 tissue samples were analyzed from frozen post-mortem brain tissue from the cerebellum and the frontal cortex. The microarray readings were preprocessed by NABEC/UKBEC using a cubic spline normalization method in Illumina Genome Studio Gene Expression Module v3.2.7 and were reported as a log2

intensity. This dataset is separated into two subsets and referred to as cerebellum tissue and frontal cortex tissue.

In addition to the expression data, DNA methylation data was also available for 724 of the brain tissue samples measured by NABEC/UKBEC. The data was measured using the Illumina HumanMethylation BeadChip and was processed using the Beadstudio Methylation Module v3.2.0 with no normalization. The dataset is catalogued in the GEO database under the accession number GSE36194. The methylation profiles generated through the BeadChip quantify the average methylation at each CpG site using the Beta-value (β-value). Defined as the ratio of the methylated probe intensity and the overall intensity (sum of methylated and non-methylated probe intensities), the β-values range from 0 to 1 and follow a Beta distribution (Bibikova et al., 2006; Du et al., 2010; Weinhold et al., 2016).

2.2: Preprocessing the Datasets

The expression microarray datasets are comprised of both coding and non-coding transcript probe targets. We excluded the non-coding transcripts as well as transcript with no known gene coordinates, reducing the expression datasets to 42,084 probes, down from 48,804 probes.

Since the brain expression and methylation datasets were individually processed by different tissue banks and in several batches, we corrected for the batch effect using the limma (Ritchie et al., 2015) package in R. The breast tissue dataset was previously batch corrected by the METABRIC group.

Next, we subset the data into groups based on the available clinical annotations provided by the NABEC/UKBEC database. These annotations included tissue type (Cerebellum and

Frontal Cortex), sex (Male and Female), and age (ranging from 0 to 102 years old). We clustered the age annotations into groups using a K-Means clustering algorithm, whereby the optimal number of clusters was determined using the elbow method. After four clusters, the change in total within-clusters sum of squares did not explain a significant amount of additional variance, therefore $k = 3$ was chosen as the optimal number of clusters for the age annotation. We then converted the continuous, numeric age annotation into three categorical age groups (0-21 years, 22-73 years, 74+ years).

We then compared the 12 possible clinical annotation permutations to determine the optimal method to subset the brain samples. For each of the 12 groups, we calculated the median expression for each probe and performed a hierarchical clustering via multiscale bootstrap resampling using the pvclust package in R (Suzuki and Shimodaira, 2006). Using an AU (Approximately Unbiased) p-value of 0.99, analogous to a p-value significance level of 0.01, the ideal clustering method was to subset the data solely by tissue type. Thus, we divided the brain dataset into the cerebellum tissue and frontal cortex tissue datasets. Due to the paired nature of the methylation and expression data, the methylation brain dataset was also subset into cerebellum and frontal cortex tissue subsets.

2.3: Identification and Removal of Bimodal Expression Genes

Gene whose expression exhibited a bimodal distribution can be thought of as having two exclusive phenotypic states. However, our focus in this analysis was to examine the factors affecting the tightly regulated, or conversely the highly variable gene expression genes in a single phenotypic state. In order to identify if a gene's expression was unimodal or bimodal, we modeled

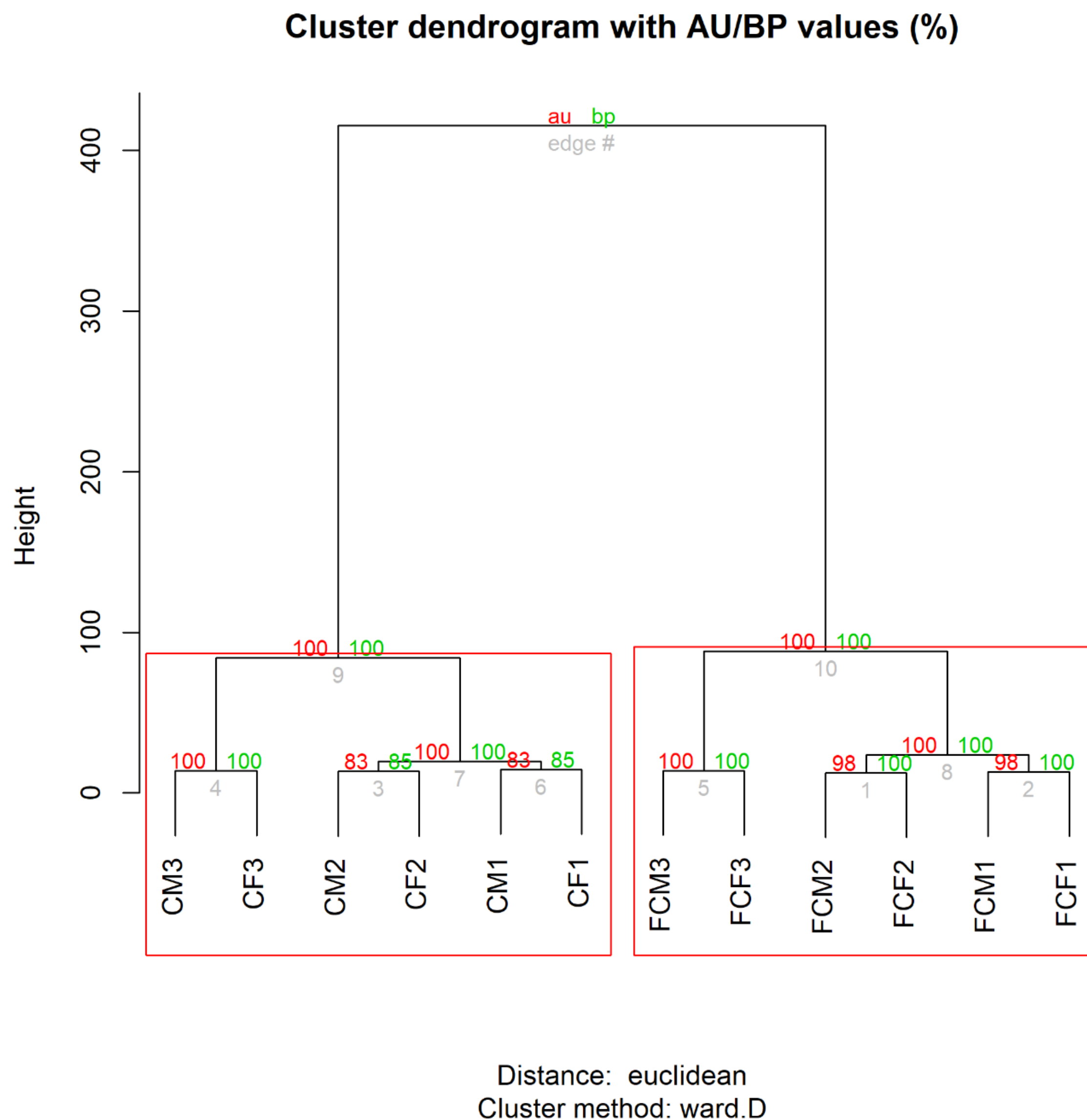


Figure 2. Hierarchical clustering of brain dataset samples

The brain samples were subset into 12 groups based on the possible permutations of three clinical annotations: tissue type (Cerebellum and Frontal Cortex), sex (Male and Female), and age (ranging from 0 to 102 years old). Age was converted to a categorical variable consisting of three age groups (1: 0-21 years, 2: 22-73 years, 3: 74+ years). Next, the groups were hierarchically clustered via multiscale bootstrap resampling using the expression data across 42,084 probes. The Euclidean distance was used to measure the similarity between clusters. The approximately unbiased (AU) p-values are shown in red text while the red rectangles represent clustering methods with a threshold AU of 0.99, analogous to a significance level of 0.01.

each gene expression as a mixture of two gaussian distributions using the mixtools package in R (Benaglia et al., 2009).

We first identified the peaks of the probability density functions (PDF) for each gaussian distribution. Next, we compared the distance between the peaks as well as the ratio of peak density heights. Genes with peaks that were greater than one MAD apart and displayed a peak ratio greater than 0.1 were considered to have a bimodal expression and subsequently removed from the analysis. Only 15/3453 breast tissue genes, 6/3487 cerebellum genes, and 6/2980 frontal cortex genes were excluded for being bimodal. Genes that did not meet these criteria were considered to have a unimodal distribution and were kept for further analysis.

2.4: Estimating Expression Variability

To calculate a measure of variability for expression that accounts for the relationship between mean expression and variability while being, on average, unaffected by overall expression, we modified an existing EV method (Alemu et al., 2014) to determine each gene's expression variability (EV). We first calculated a bootstrapped estimate of the median absolute deviation of each gene using 1000 bootstrap replicates, referred to as the observed MAD. Next, a local polynomial regression curve (loess function with default parameters on R version 3.4.2) was used to determine the expected MAD of gene expression as a function of the median value. No additional smoothing was used for the regression curve. Lastly, we calculated each gene's EV as the difference between the bootstrapped MAD and the expected MAD at the gene's median expression level (Equation 4).

2.5: EV Gene Set Classification

We classified the genes into three distinct gene sets based on their expression variability:

$$\tilde{x}_{EV} \pm 3 * MAD_{Bootstrapped} \quad (5)$$

where $\tilde{x}_{EV}$ is the median of EV values for each dataset, and $MAD_{Bootstrapped}$ is the median absolute deviation of EV from 1000 bootstrapped replicates. Genes whose EV fell within the aforementioned EV range were considered Non-Variable. Genes above this range were termed Hypervariable, and the remaining genes were considered Hypovariable.

2.6: Bootstrapping EV Gene Set Classifications

To statistically validate our EV classifications, we employed an iterative split-half reliability test for each dataset. First, we randomly split the data into two equally-sized sample subsets. The EV estimation and gene set classification analyses were repeated to reclassify the genes using the new subsets. This 50-50 split-retest procedure was repeated 100 times per tissue.

Next, the accuracy of the original EV classifications was assessed by comparing the original classification of each gene with the newly determined 50-50 split-retest classifications. To this end, a binomial test was used with a probability of success greater than 0.5. For each split-retest replicate, an accurate EV classification, or "success", is defined as a consistent EV classification of a gene between the original dataset and both split datasets. Gene classifications were considered significant with a p-value < 0.05 using the binomial test.

2.7: Structural Analysis of EV Genes

Data regarding the structural features of the genes was obtained from the GRCh38/hg38 assembly of UCSC Table Browser (Karolchik et al., 2004). Linear regression n analyses were conducted to find any correlation between gene EV and their structural features. If a single gene had multiple transcript sizes, we conducted a separate linear regression analysis for the largest and the smaller transcripts individually. The sequence lengths excluded introns, 3' and 5' UTR exons, and any upstream or downstream regions.

2.8: Gene Set Functional Enrichment Analysis

The Gene Ontology (GO) term enrichment analyses were conducted using ConsensusPathDB gene set over-representation analysis (Pruitt et al., 2009). The truncated list of unique Illumina HumanHT-12 V3.0 expression BeadChip genes that corresponded to coding transcripts with known gene coordinates was used as a background list of genes. The resulting GO terms were filtered using p-value cutoff of 0.01 and limited to gene ontology level 5 categories.

Common and unique GO terms were then summarized using REVIGO (Supek et al., 2011) and visualized through treemaps by the provided R scripts. The parameters used were a medium allowed similarity (0.7) using Homo sapiens database of GO terms and SimRel as the semantic similarity measure between ontology terms.

2.9: Enrichment Analysis

Using the Pearson's chi-square test, we tested for enrichment of essential genes in each gene set relative to the total number of essential genes in the Illumina HumanHT-12 V3.0 expression

BeadChip. A list of 20,029 protein coding genes from the Consensus Coding DNA Sequence Project (CCDS) database was used to test for essentiality enrichment (Pruitt et al., 2009). Only genes that are solely classified as essential are considered, while non-essential and conditionally essential genes were excluded from the analysis. The resulting list of protein coding genes consisted of 2377 essential genes present in the Illumina dataset. Once the number of annotated genes and gene sets were deemed dependent variables through the Pearson's chi-square test, we determined the enrichment of annotated genes in each gene set using standardized Pearson residuals.

The Pearson's chi-square test and standardized residuals were also used to test the enrichment of methylation clusters across the Hypervariable, Hypovariable, and Non-Variable probe sets.

2.10: Hierarchical Clustering of Age-Dependent Hypervariable Genes

The hierarchical clustering of 12 clinical annotation permutations resulted in the clustering of groups with the same age group but the opposite sex. While the p-values of the sex and age groupings during the hierarchical clustering were too high to warrant further subsetting of the brain dataset samples into distinct groups, the p-values were trending towards significance. This suggested that inspecting the Hypervariable gene sets on a gene-by-gene basis might identify a specific subset of genes affected by these clinical annotations. We used a multiple linear regression model to measure the changes in expression of the Hypervariable genes as a function of age, sex, and post-mortem interval (PMI):

$$Y_i = \beta_0 + \beta_1 Age + \beta_2 Sex + \beta_3 PMI \tag{6}$$

where Y_i is the expression level of a probe and β_n is the coefficient for each term. The p-values were calculated using a type III sum of squares regression and adjusted for multiple comparisons using the Benjamini-Hochberg method. Genes that exhibit a false discovery rate (FDR) < 0.01 were considered significant for the specific coefficient, and the sign of the coefficient determines if the probe is positively or negatively correlated with the factor.

To examine the changes in EV as a function of age, we subset the samples into distinct age clusters. The choice to use three age clusters for Hypervariable genes as the optimal number of clusters was determined using an expectation-maximization (EM) algorithm initialized by hierarchical clustering for parameterized Gaussian mixture models in the mclust package of R (Scrucca et al., 2016). The Bayesian information criterion for each hierarchical clustering model was determined, and both the cerebellum and frontal cortex displayed identical optimal numbers of age clusters ($k = 3$). Once the samples were correctly clustered by age, the gene clusters were selected by manually cutting the gene dendrograms. The gene expressions were then visualized as heatmaps using the gplots package in R (Warnes et al., 2015).

3: Results

3.1: Estimating Expression Variability

3.1.1: Estimating EV

Gene expression variability within populations is usually quantified by standard deviation or the coefficient of variation. However, these measures do not adequately or accurately account

for the effects of expression magnitude. The standard deviation is the most common measure of variability or dispersion of expression values that excludes any measure of expression magnitude. As a result, genes with high standard deviation are often those with high average expression while genes with low standard deviation often exhibit low average expression. The coefficient of variation, on the other hand, is a measure of relative variability, also known as relative standard deviation. It is defined as the ratio of the standard deviation to the mean. However, using CV as a measure of expression variability has many drawbacks. For example, genes with high coefficient of variation values are often genes with low mean expression. In addition, similar amounts of variation in expression values at different average expression values result in drastically different CV values and may not reflect changes in biologically-relevant processes in a cell. As such, a metric of gene expression variability cannot be calculated independent nor relative to expression magnitude, but rather be derived as a function of expression magnitude. To this end, we opted to use expression variability, or EV (Alemu et al., 2014), as a measure of variability in gene expression across populations. EV defines gene expression variability as a function of a gene's average expression in a non-biased and robust manner.

In order to study EV in a human population, we analyzed gene expression variability in non-diseased human tissue samples. We studied 911 post-mortem brain tissue samples (Hernandez et al., 2012; Trabzuni et al., 2013) (456 cerebellum samples and 455 frontal cortex samples) as well as 144 biopsied breast tissue samples (Curtis et al., 2012). Each dataset used the Illumina HumanHT-12 v3 Expression BeadChip platform to quantify gene expression of 48,803 unique Illumina probes. We excluded probes corresponding to non-coding transcripts as well as those with missing probe coordinates, resulting in a list of 42,084 genes.

To estimate a gene's EV, we modified an existing EV method (Alemu et al., 2014). We used the median absolute deviation as a measure of variability due to its robustness and reduced sensitivity to outliers. In order to estimate the population MAD of each gene, we calculated the bootstrap estimate of the MAD for each probe using 1000 bootstrap replicates. Next, we estimated the expected MAD for individual genes as a locally weighted polynomial regression curve of the bootstrapped MAD as a function of median expression of all the genes (Fig. 3, solid red line). Lastly, we calculated each gene's EV as the difference between the bootstrapped MAD and the expected MAD at the gene's median expression level, as shown in Equation (4).

The expected MAD regression curves exhibit a flat, negative parabolic shape, whereby the lowest and highest gene expression levels represent the troughs of the curve. Variability in gene expression levels has previously been shown to decrease as expression approaches either extrema (Carey et al., 2013; Newman et al., 2006; Taniguchi et al., 2010). In order to confirm that is EV was not a relative measure of variability, as is the case for CV, we conducted a linear regression analysis between expression magnitude and EV (Fig. 4). Based on the poor adjusted R^2 values ($2x10^{-4}$, $8x10^{-4}$, and $5x10^{-3}$ for breast, cerebellum, and frontal cortex respectively) and the flat slopes, we concluded that EV accounts for the relationship between mean expression and variability with no direct relationship to overall expression.

3.1.2: Classification of EV Gene Sets

To begin investigating the statistical characteristics of the newly calculated gene EV values, we first examined the sample space of the resulting values. To do so, we plotted the probability density function of EV for each tissue (Fig. 5). The EV tissue types exhibit large peaks around the zero mean and a long tail for positive EV genes. Breast tissue exhibited a larger shoulder of the

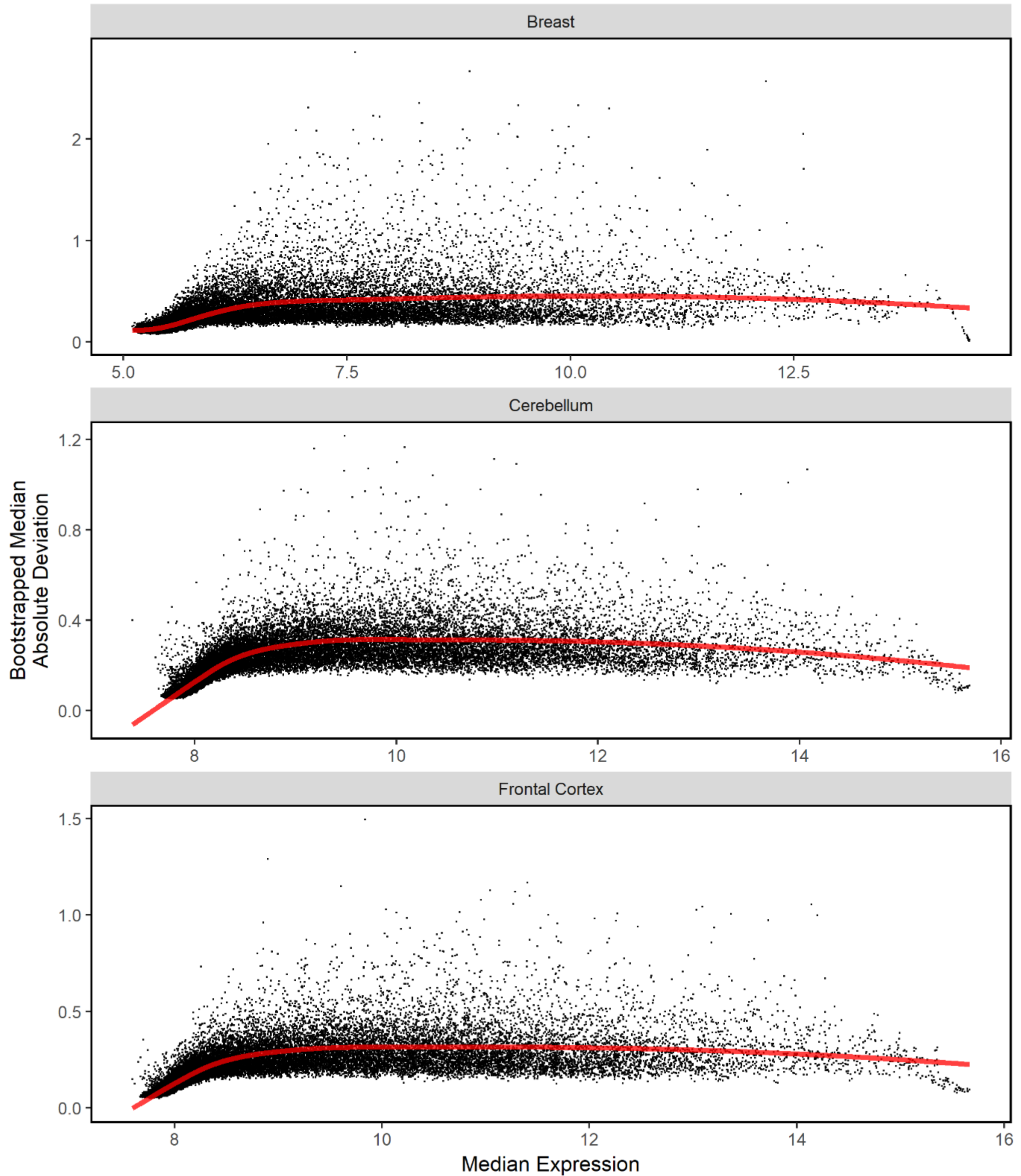


Figure 3. EV in human breast, cerebellum, and frontal cortex tissue

The expected median absolute deviation, shown as a solid red line, is calculated as the function of median expression using a local polynomial regression (LOESS) curve. EV is calculated as the difference between the observed MAD and the expected MAD values.

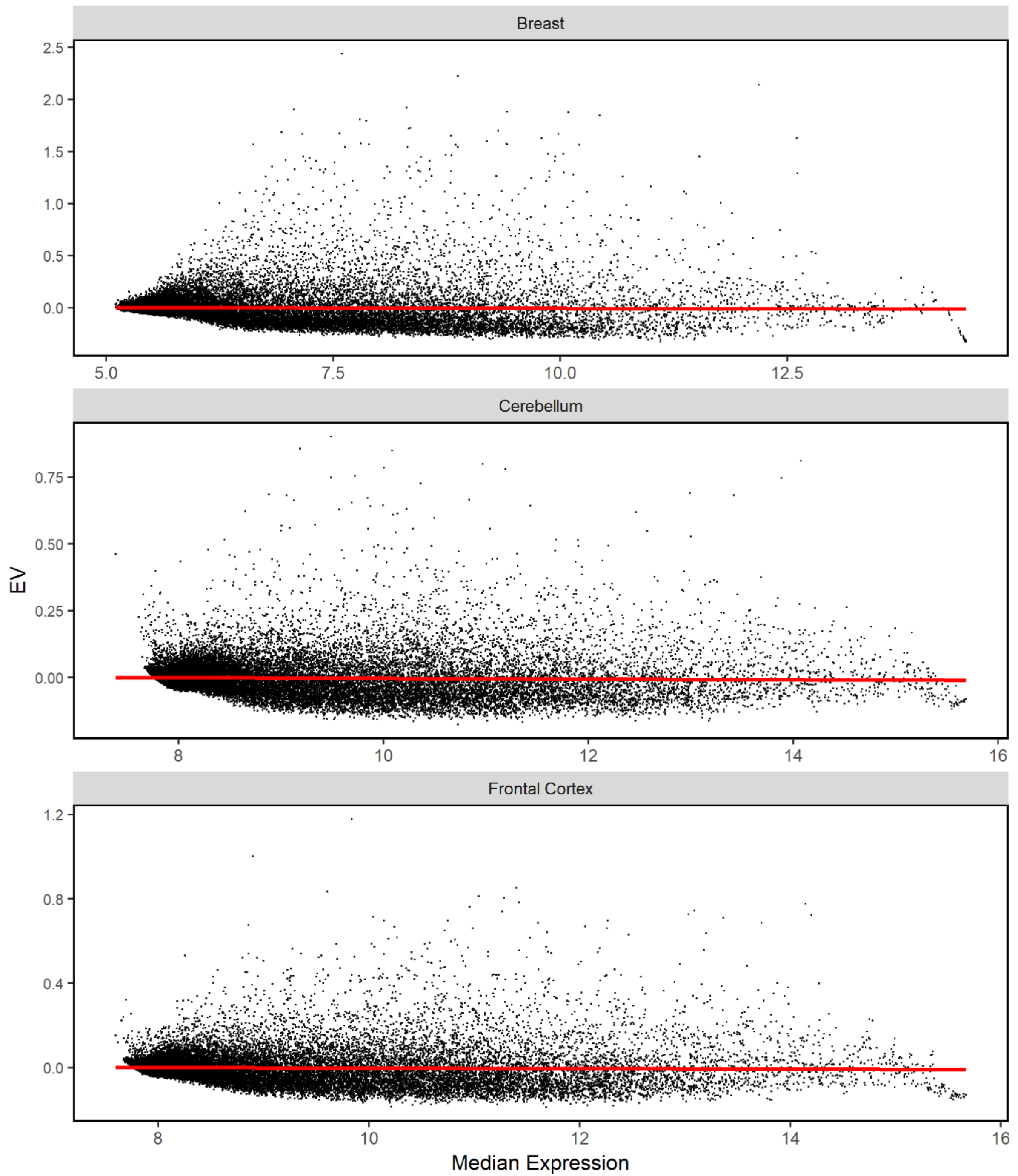


Figure 4. Relationship between EV and expression

Adjusted R^2 values for the linear regression model shown in red were 0.0002, 0.0008, and 0.005 for breast, cerebellum, and frontal cortex tissues respectively.

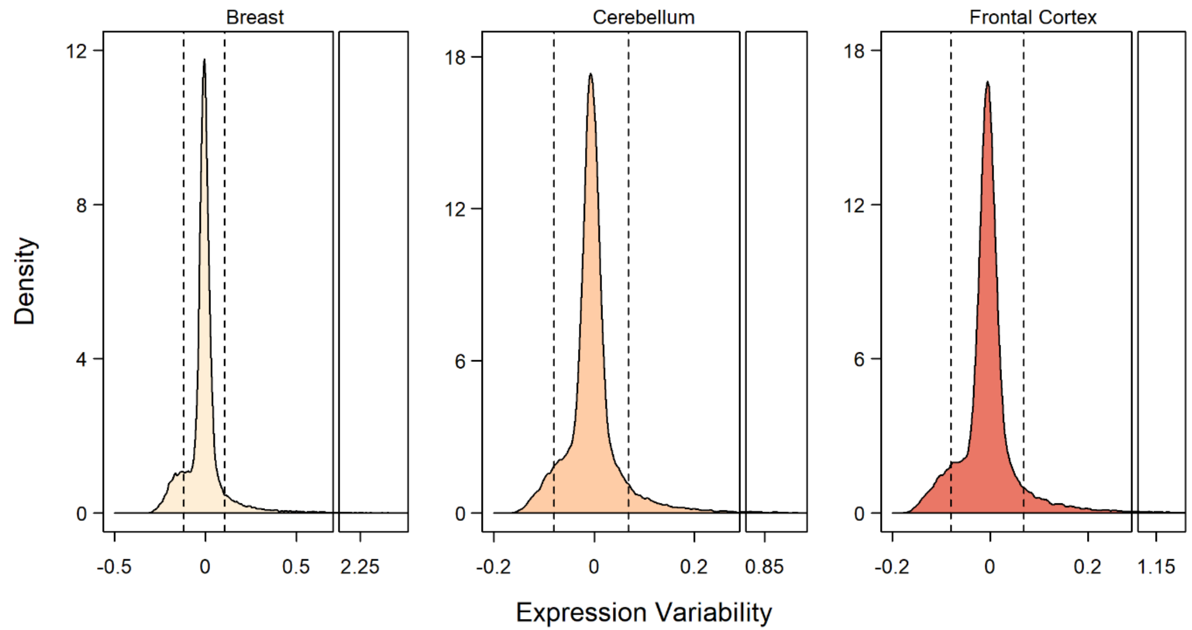


Figure 5. EV probability density functions

The dashed vertical black lines represent the EV range of Non-Variable genes. EVs greater than the Non-Variable EV range represent Hypervariable genes while EVs lower than the EV range are classified as Hypovariable.

negative EV genes compared to cerebellum and frontal cortex tissues. This is likely attributable to the lower number of breast samples (144 compared to 456 and 455 samples respectively). We then classified each probe into three categories based on their EV relative to the EV of other genes. First, we calculated the bootstrap estimate of the median absolute deviation of the EV for each dataset using 1000 bootstrap replicates. Genes whose EV fell outside the range of $\tilde{x}_{EV} \pm 3 * MAD_{EV}$ were considered significantly variable. We used the term "Hypervariable" to describe genes whose EV was greater than $\tilde{x}_{EV} + 3 * MAD_{EV}$. Genes with an EV less than $\tilde{x}_{EV} - 3 * MAD_{EV}$ were deemed "Hypovariable". The remaining genes that fell within the range of $\tilde{x}_{EV} \pm 3 * MAD_{EV}$ were considered "Non-Variable". We propose that these three distinct gene groups, categorized based on EV, correspond to distinct functional and phenotypic gene groups.

3.1.3: Cross-Validation of EV Classifications

After classifying genes based on their expression variability as either Hypervariable, Hypovariable, or Non-Variable, we then accounted for the effects of sampling error on EV. To minimize the effects of sampling errors and to increase the accuracy of our EV classification method, we cross-validated our EV classification method using a split-retest classification method (Fig 1). Briefly, we divided our tissue samples into two equally sized subsets and repeated the EV analysis and gene set classification in each sample subset with 100 replicates. Figure 6 shows the relative frequencies of correct EV classifications across the three subsets for each EV class and each tissue type, and Figure 7 shows the resulting number of genes before and after the retest method. By conducting the 50-50 split-retest procedure, we increased the robustness and accuracy of our EV classification method for each gene by increasing the specificity of the classification model.

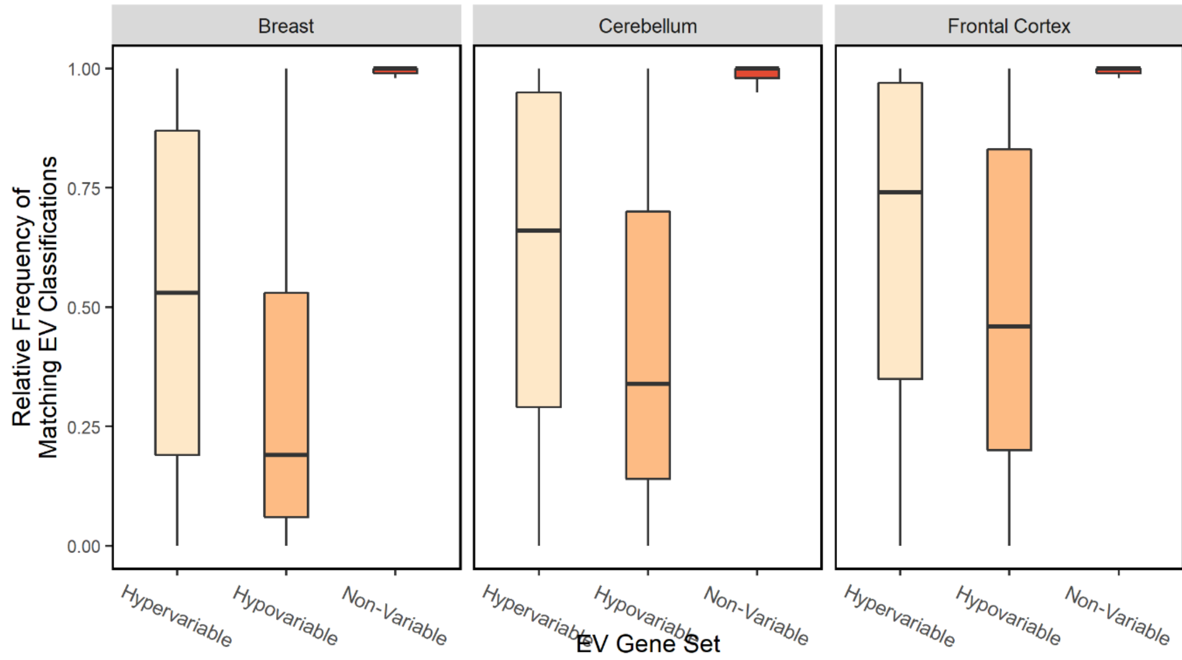


Figure 6. EV gene set classification accuracy

Relative frequency of gene EV classification accuracy between original distribution and 50-50 split-retest replicates (n=100).

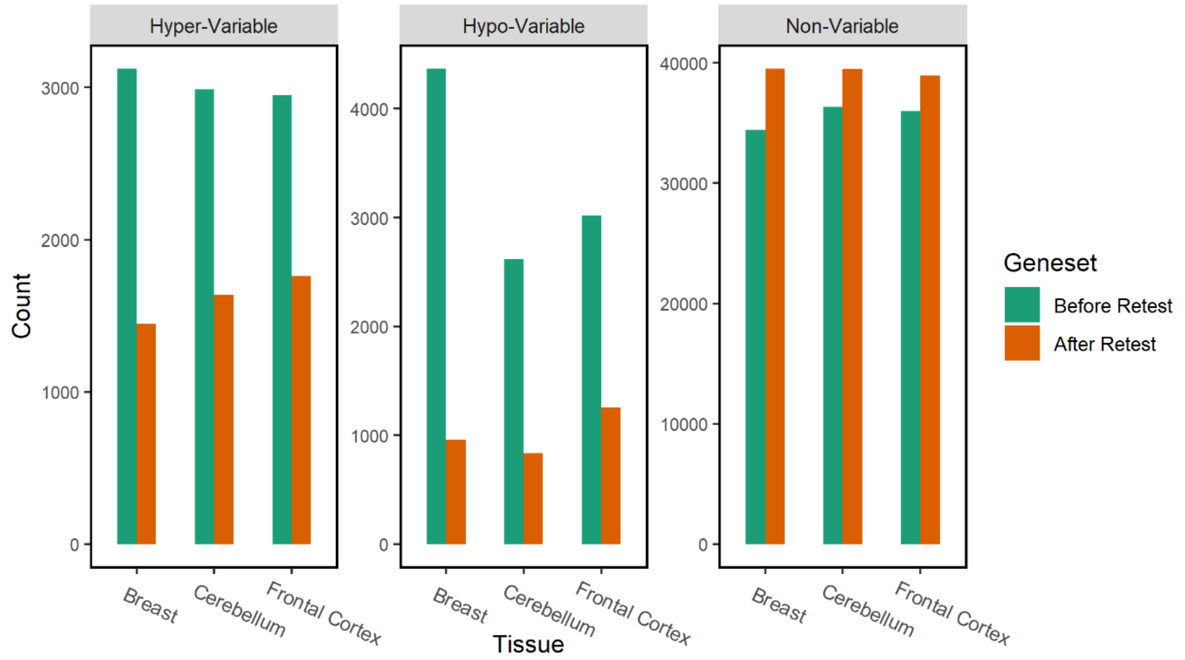


Figure 7. Number of genes in EV gene sets

Barplot showing the number of genes in each gene set before and after the 50-50 split-retest procedure.

3.1.3: Statistical Nature of Hypervariability

An important aspect of Hypervariability to consider is the statistical distribution of the wide range of gene expression levels across the population. Specifically, high EV might result from a broadening of expression values around a unimodal mean value. Alternatively, high EV could be the result of a multimodal distribution of gene expression with two or more distinct expression patterns associated with different phenotypic states.

In order to distinguish between the two possibilities, we modeled each gene expression as a mixture of two Gaussian distributions (Fig. 8). Next, we identified the peaks of the probability density function for each Gaussian distribution and compared the distance between the peaks as well as the ratio of peak heights. Genes with peaks that were greater than one median absolute deviation apart and displayed a peak ratio greater than 0.1 were classified as having a bimodal expression distribution. Genes that did not satisfy both criteria were considered to have a unimodal distribution. Only a small minority of the Hypervariable (high EV) genes (15/3453 breast tissue genes, 6/2980 cerebellum genes, and 6/3487 frontal cortex genes) showed a bimodal distribution of gene expression. The remaining majority of Hypervariable genes had a unimodal distribution. This indicates that high expression variability is a result of a widening of possible expression values across a single mean rather than the gene expression existing in two or more discrete states.

3.1.5: Tissue Specificity of EV

An important issue to address in the study of gene expression variability is the extent to which tissue-specific effects impact EV. The cerebellum and frontal cortex, both tissues located

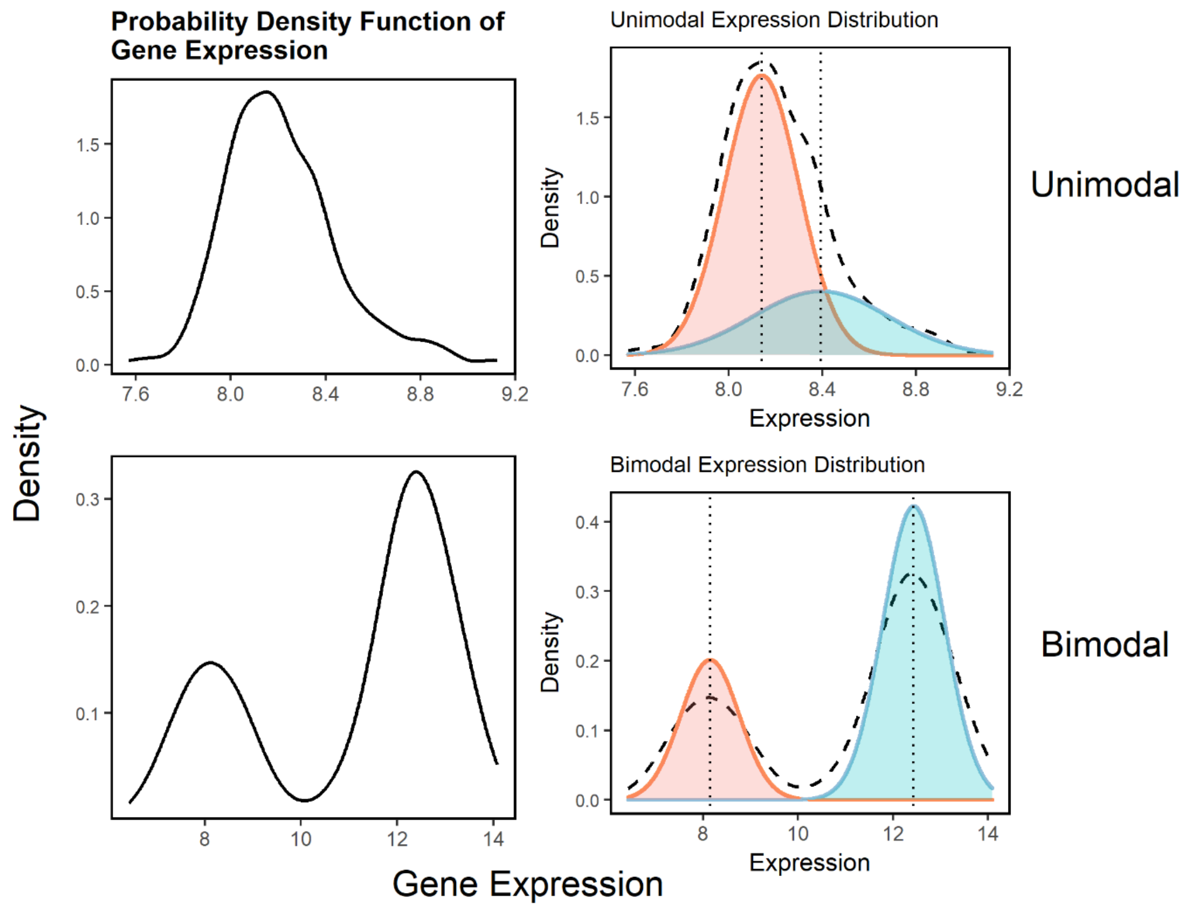


Figure 8. Modality of Hypervariable gene expression distribution

Gaussian mixture modelling method of detecting bimodal genes. The plots in the left column represent the original gene expression distributions. The dotted vertical lines represent the means of the each of the Gaussian distributions. The plots in the right column represent the Gaussian mixture model of gene expression composed to two Gaussian distributions. An example of a unimodal distribution is shown in the top row while the bottom row shows a bimodal distribution.

in the brain, exhibit a higher correlation between gene EV relative to the breast tissue, suggesting a level of tissue specificity of EV (Fig. 9). As shown in Figure 10, there is considerable overlap in the Non-Variable gene sets, with over 71% of the measured genes commonly classified as NV in all three tissue types. The Hypervariable and Hypovariable gene sets however exhibit limited overlap between the three different tissues. 13-16% of the Hypervariable genes were classified as such in the three tissues and 18-26% of the Hypovariable were so classified. The poorly overlapping nature of the variable EV classification suggests that both highly constrained and lowly constrained gene expression variability is largely determined by tissue-specific pathways.

3.1.6: EV and Gene Structural Characteristics

To understand the possible genomic mechanisms by which population EV occurs, we explored the relationship between EV and various structural features of the genes. Expression variability has previously been reported to be associated with gene size, gene structure, and surrounding regulatory elements (Alemu et al., 2014). However, we found no significant linear correlation between EV and a gene's exon count, sequence length, transcript size, or number of isoforms (Table 1). While certain linear models, such as the number of transcripts in breast tissue, exhibited statistical significance ($p < 0.05$), the fit of the model and subsequent comparison of the linear model against a local polynomial regression curve showed that the correlation was either not correctly defined by a linear model or was simply too small to draw a conclusion. The complete linear regression analysis plots and summary tables can be found in Supplementary Figure 1.

While we did not find that the physical gene characteristics were correlated to EV, previous studies have shown that the position of a gene on a chromosome has considerable effects on stochastic gene expression variability, independent of gene- and promoter-specific variables

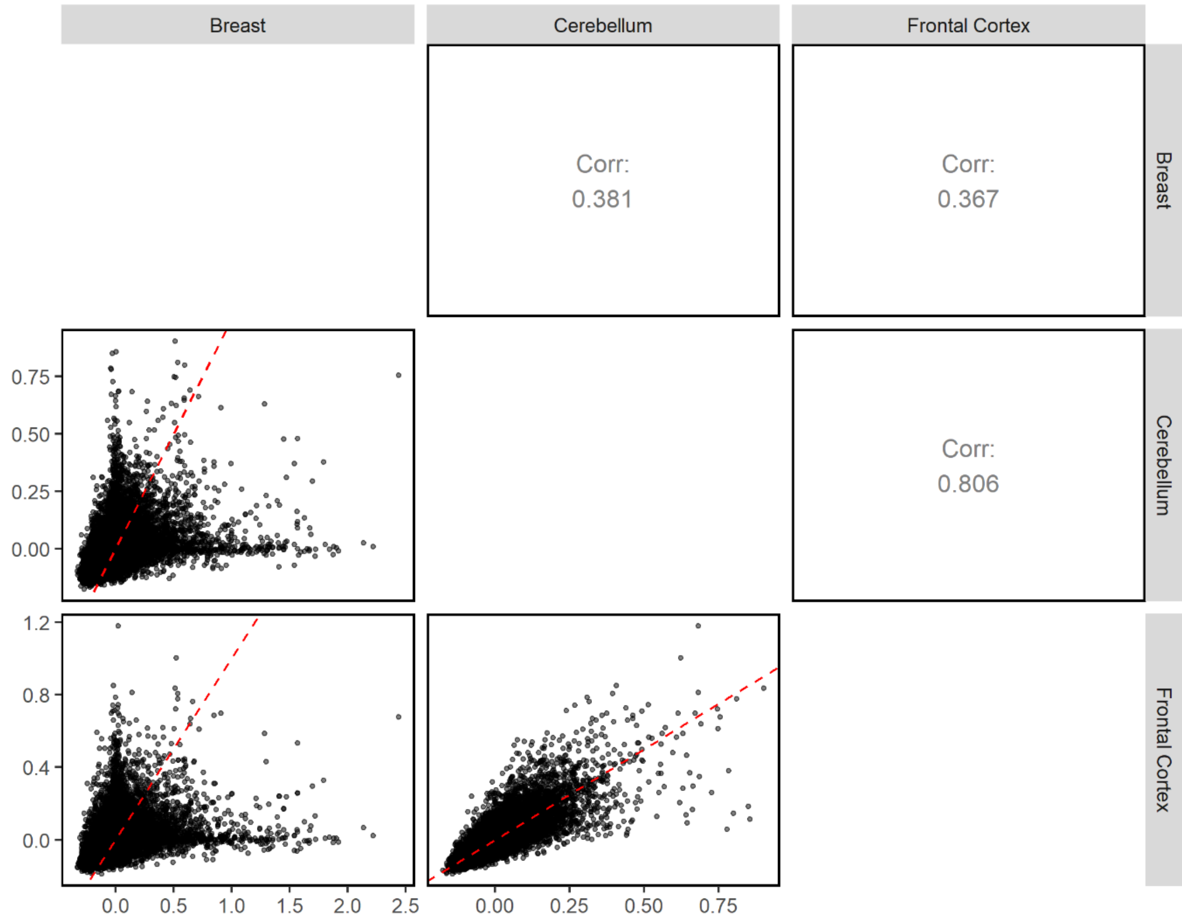


Figure 9. Tissue-specificity of EV

Cross-comparison of gene-specific expression variability between breast, cerebellum, and frontal cortex tissues.

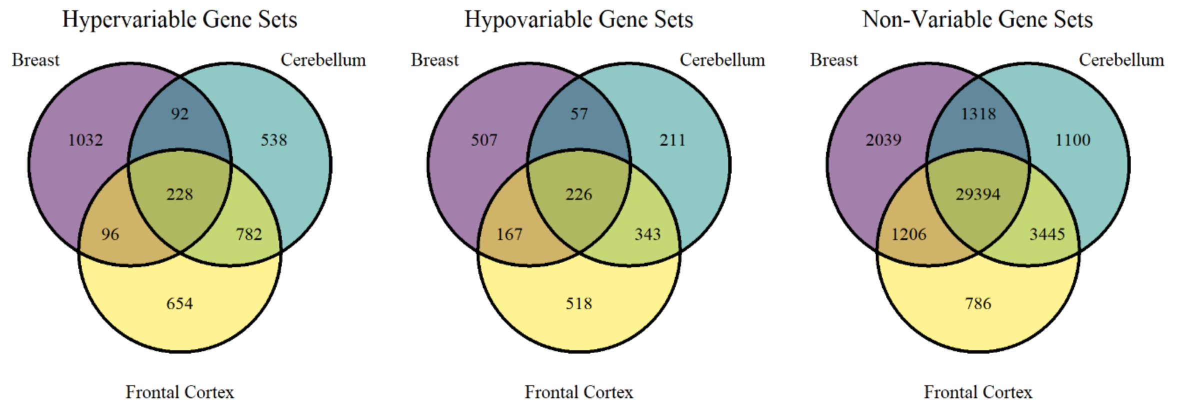


Figure 10. EV gene set overlap

Venn diagrams comparing EV gene set classifications between breast, cerebellum, and frontal cortex tissues.

Table 1. Structural analysis of genes as a function of EV gene classification
Linear regression p-values of EV as a function of each structural feature

	Breast		Cerebellum		Frontal Cortex	
	Hyper[a]	Hypo[b]	Hyper[a]	Hypo[b]	Hyper[a]	Hypo[b]
Largest Transcript	3.90E-01	7.09E-01	1.03E-01	2.17E-02	6.40E-01	5.73E-01
Smallest Transcript	9.05E-01	7.19E-01	1.25E-01	5.00E-02	7.66E-01	6.68E-01
Number of Transcripts	2.48E-02	3.07E-02	1.11E-01	7.15E-01	5.88E-01	4.01E-01
Sequence Length	8.50E-02	4.61E-01	3.19E-02	9.52E-02	4.05E-02	1.83E-01
Exon Count	7.05E-01	4.69E-01	2.87E-01	4.81E-02	1.67E-01	6.89E-02

[a] Hypervariable genes

[b] Hypovariable genes

(Batenchuk et al., 2011). We next tested if there is a relationship between expression variability and chromosomal position (Fig. 11). To this end, each chromosome was divided into 100 bins and the average EV all the genes within each bin was determined. If there was no relationship between the two, we would expect to see EV to be uniformly distributed throughout the genome. We found that EV is not uniformly distributed across the genome, and individual regions of chromosomes exhibited peaks of high expression variability or troughs of low expression variability.

To further confirm this hypothesis of non-uniform EV distribution across the genome, we tested the cosine similarities of each bin between the chromosomes in each tissue type separately (Fig. 12) as well as comparing EV similarities between the tissue types (Fig. 13). The chromosome combination similarity matrices exhibited a relatively low cosine similarity in all three tissue types and may partly be caused by the varying sizes of the chromosome bins or the number of genes within each bin. The paired chromosome analysis, on the other hand, exhibited a low similarity between either breast pairwise combination while the cerebellum-frontal cortex pairwise comparison resulted in higher cosine similarities. This is likely due to the fact that both the cerebellum and the frontal cortex tissues are located in the brain, and therefore share similar tissue compositions, and metabolic pathways. This supports the hypothesis that EV plays a role in tissue specificity and tissue identity. In addition, the chromosome combination similarity matrices exhibit a low cosine similarity in all three tissue types, suggesting that each chromosome also exhibits a unique EV signature that is distinct from other chromosomes.

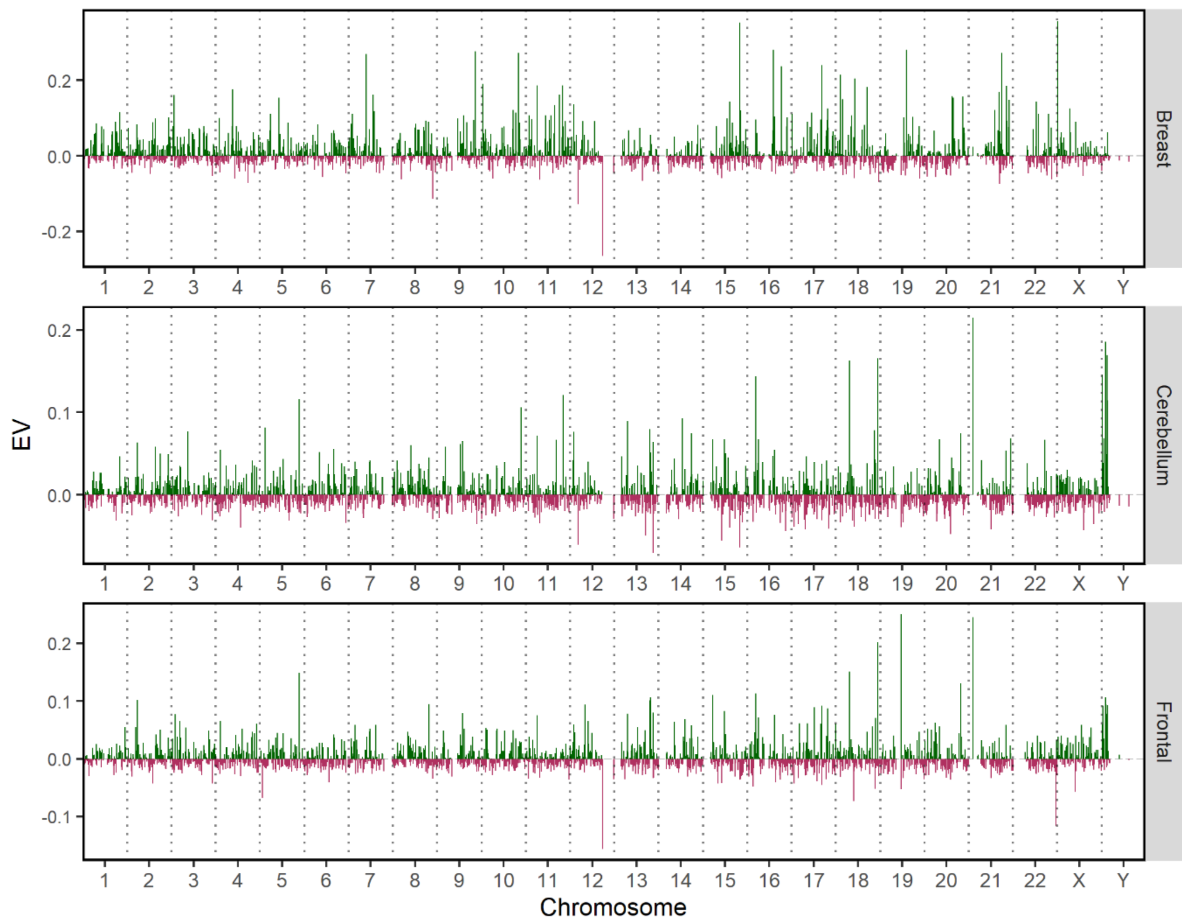


Figure 11. Effect of genomic position on EV

Each chromosome is divided into 100 bins (x-axis) based on the maximum gene coordinate annotation and the average EV in each bin in measured (y-axis). The green vertical lines represent bins with a positive average EV, while the red vertical lines represent negative average EV.

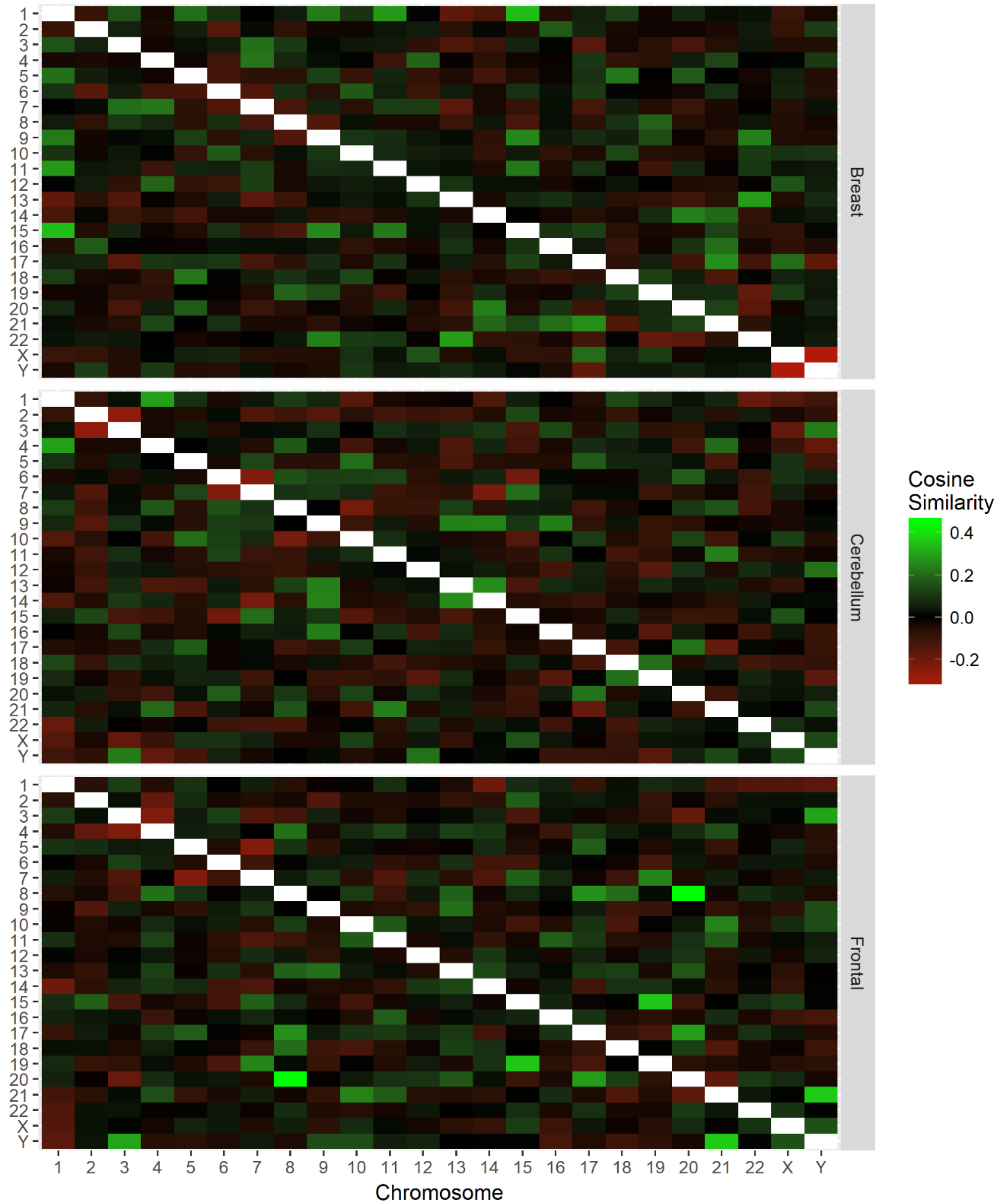


Figure 12. Chromosome combinations cosine similarity

Cosine similarity between pairs of chromosomes in the same tissue type. Each chromosome represents a feature vector, with 100 features corresponding to the average EV of the 100 chromosomal bins.

Figure 13. Paired chromosome cosine similarity between tissues

Cosine similarity between the same chromosomes in different tissue type. Each chromosome represents a feature vector, with 100 features corresponding to the average EV of the 100 chromosomal bins.

3.1.7: Functional Enrichment of EV Gene Sets

In order to understand the overall biological significance of gene EV, we examined the enriched functional annotations in the Hypervariable, Hypovariable, and Non-Variable gene sets by conducting a gene set enrichment analysis for each gene set. We determined the over-represented GO terms that were unique in each tissue type, as well as GO terms that were common in all three tissue types. The resulting GO annotations were simplified and visualized using a REVIGO treemap. The top five GO annotations that are commonly enriched in all three tissue types can be found in Table 2, while the top 5 unique GO annotations in each tissue type can be found in Table 3.

The breast Hypervariable gene set was uniquely enriched for epithelial cell differentiation, primary alcohol metabolism, and positive regulation of cellular component movement. The cerebellum Hypervariable gene set was uniquely enriched for regulation of nervous system development, transmembrane transport, and neuron death. The frontal cortex Hypervariable gene set was enriched for histamine secretion, regulation of cell morphogenesis, and trans-synaptic signalling. The breast, cerebellum, and frontal cortex Hypervariable gene sets were commonly enriched for regulation of tissue remodeling, inflammatory responses, and responses to inorganic substances. Of note, many of the enriched GO annotations of the Hypervariable genes are involved in signalling pathways. These pathways require dynamic control based on internal and external stimuli and their high EV likely represents differing environmental or hormonal conditions amongst the individuals.

In the case of the Hypovariable gene sets, all three tissue types were enriched for protein catabolism and metabolism, ribonucleoprotein complexes, and negative regulation of autophagy.

Table 2. Common GO annotations of EV gene sets

Top five common enriched GO terms between breast, cerebellum, and frontal cortex within the Hypervariable and Hypovariable gene sets

Common Hypervariable Annotations	Common Hypovariable Annotations
Regulation of bone remodeling	Proteolysis involved in cellular protein catabolism
Regulation of inflammatory response	Ribonucleoprotein complex assembly
Response to zinc ion	Regulation of cellular amino acid metabolism
Carboxylic acid biosynthesis	Innate immune response activating cell surface receptor signaling pathway
Regulation of ion transport	Negative regulation of autophagy

Table 3. Unique GO annotations of EV gene sets

Top five unique enriched GO terms of Hypervariable and Hypovariable genes unique to breast, cerebellum, and frontal cortex tissues.

	Unique Breast Annotations	Unique Cerebellum Annotations	Unique Frontal Cortex Annotations
Hypervariable	Epithelial cell differentiation	Regulation of nervous system development	Histamine secretion
	Primary alcohol metabolism	Regulation of transmembrane transport	Regulation of cell morphogenesis
	Positive regulation of cellular component movement	Regulation of neuron death	Trans-synaptic signaling
	Response to corticosteroid	Negative regulation of response to external stimulus	Regulation of neurological system process
	Transmembrane receptor protein tyrosine kinase signaling pathway	Response to calcium ion	Dephosphorylation
Hypovariable	Golgi vesicle transport	DNA conformation change	ncRNA metabolism
	Nucleoside monophosphate metabolism	Modification-dependent macromolecule catabolism	Response to interleukin-1
	Proteolysis involved in cellular protein catabolism	Response to camptothecin	Regulation of enter of bacterium into host cell
	Cellular response to nitrogen starvation	Retrograde transport, endosome to Golgi	
	Mitochondrial respiratory chain complex I assembly	Regulation of ubiquitin-protein transferase activity	

The breast Hypovariable gene set was enriched for Golgi vesicle transport, nucleoside metabolism, and protein catabolism. The cerebellum Hypovariable gene set was enriched for DNA conformation change, modification-dependent macromolecule catabolism, and retrograde transport. Lastly, the frontal cortex Hypovariable gene set was enriched for ncRNA metabolism, response to interleukin-1, and regulation of enter of bacterium into host cell.

Genes with higher expression variability have previously been shown to be functionally and physically involved with the cell periphery, localizing in the membrane, transmembrane, or extracellular matrix regions (Alemu et al., 2014; Mar et al., 2011; Newman et al., 2006). Our results corroborated these findings as the Hypervariable gene sets from breast, cerebellum, and frontal cortex were enriched for GO annotations associated with cell surface signalling pathways, as well as cellular component ontologies enriched at the plasma membrane. In contrast, genes with low expression variability genes tended to regulate nucleic acid and metabolic pathways, localizing in the cell interior. These Hypovariable genes are likely involved in complex, dose-sensitive gene networks and require tight regulation of their expression to function correctly.

3.2: Essentiality Enrichment in Variable Genes

Previous studies in yeast and *E. coli* have shown that gene expression variability is reduced in genes that are essential for survival. It is believed that evolution has selected for transcriptional networks that limit stochastic expression variation of essential genes (Lehner, 2008; Silander et al., 2012). If this were true for humans, we would expect a significant number of essential genes to exhibit Hypovariable expression. Conversely, we expect a depletion of essential genes within the Hypervariable probe sets.

In order to examine a potential correlation between expression variability and essentiality in human tissues, we first tested the independence between EV classification and annotation of human essentiality (Table 4). Essentiality annotations were obtained from the CCDS (Pruitt et al., 2009) and MGD (Blake et al., 2017) databases. Here, direct human orthologs of genes essential for prenatal, perinatal, or postnatal survival of mice were classified as essential. Using the Pearson's chi-square test for the number of essential genes in each probe set, we find that that the Hypovariable gene set in breast, cerebellum, and frontal cortex tissues were significantly enriched for genes with essentiality annotation (p-value = 1.50×10^{-34}, 8.27×10^{-63}, and 1.52×10^{-92}, respectively). Thus, expression variability for many essential genes is constrained in humans, likely reflecting a similar biology to essential yeast genes. However, we also observe a significant enrichment of essential genes within the Hypervariable gene sets. This was a surprise to us since essential genes are thought of as being dose-sensitive and changes in the level of gene expression would be predicted to be deleterious or lethal.

To better understand the implications of high variability in essential genes, we examined the functional annotations associated with both Hypervariable and Hypovariable essential genes (Tables 5-6). The breast essential Hypervariable gene set was enriched for embryonic development, responses to growth factors, cell-substrate junction assembly, regulation of epithelial cell proliferation, and positive regulation of cellular component movement. The cerebellum essential Hypervariable gene set was enriched for cell differentiation, anion transport, trans-synaptic signaling, response to growth factors, and cell projection organization. Lastly, the frontal cortex essential Hypervariable gene set was enriched for cell differentiation, secretion, tyrosine kinase signaling, cell projection organization, and heart contractions. Overall, the Hypervariable essential gene sets tended to be enriched for morphogenic, tissue, and organ system

Table 4. Essentiality enrichment analysis

Pearson's Chi-squared test for essentiality in Hypervariable, Hypovariable, and Non-Variable gene sets in breast, cerebellum, and frontal cortex tissue

Tissue	Probe Set	Total Gene Count	Essential Gene Counts	Standardized Residuals	P-Value
Breast	Hyper	1448	180	12.22	1.50×10^{-34}
	Hypo	957	108	8.27	
	NV	33957	1657	-14.94	
Cerebellum	Hyper	1640	170	8.66	8.27×10^{-63}
	Hypo	837	83	5.54	
	NV	35257	1849	-10.42	
Frontal Cortex	Hyper	1760	196	10.6	1.52×10^{-92}
	Hypo	1254	125	7.04	
	NV	34831	1764	-12.89	

Table 5. Common GO Annotations of essential genes

Top five overlapping enriched GO terms of essential genes between breast, cerebellum, and frontal cortex tissues.

Common Hypervariable Annotations	Common Hypovariable Annotations
Positive regulation of cell differentiation	Lysosomal transport
Cyclic-nucleotide-mediated signalling	Negative regulation of NF-kappaB transcription factor activity
Regulation of blood vessel size	Viral release from host cell
Regulation of epithelial cell proliferation	Mitochondrial genome maintenance
Terpenoid metabolism	Proteolysis involved in cellular protein catabolism

Table 6. Unique GO annotations of essentiality genes

Top five unique enriched GO terms of essential genes unique to breast, cerebellum, and frontal cortex tissues.

	Unique Breast Annotations	Unique Cerebellum Annotations	Unique Frontal Cortex Annotations
Hypervariable	Embryo development ending in birth or egg hatching	Positive regulation of cell differentiation	Positive regulation of cell differentiation
	Cellular response to growth factor stimulus	Anion transport	Regulation of secretion
	Cell-substrate junction assembly	Trans-synaptic signalling	Transmembrane receptor protein tyrosine kinase signaling pathway
	Regulation of epithelial cell proliferation	Cellular response to growth factor stimulus	Regulation of cell projection organization
	Positive regulation of cellular component movement	Regulation of cell projection organization	Heart contraction
Hypovariable	Negative regulation of cellular component organization	DNA repair	DNA repair
	DNA repair	Negative regulation of cellular component organization	Covalent chromatic modification
	Regulation of cellular protein localization	Positive regulation of viral release from host cell	mRNA transport
	Embryo development ending in birth or egg hatching	Regulation of cellular protein localization	Progesterone receptor signaling pathway
	Apoptotic process	Regulation of cell cycle process	Regulation of cell cycle process

development. This suggests that tight regulation of certain essential genes may only be required for embryonic and morphogenic development and dose-sensitivity is lost in adults, allowing for high expression variability.

3.3: DNA Methylation and Expression Variability

3.3.1: Gene Methylation Clustering

While the relationship between methylation and gene expression is complex, low promoter methylation is often associated with high levels of gene expression (Cedar, 1988; Irvine et al., 2002; Moore et al., 2013; Wagner et al., 2014). Like gene expression, DNA methylation is highly variable at the cell, tissue, and individual level (Zhang et al., 2013), suggesting that gene expression variability could result from variations in gene methylation. To explore this idea, we used DNA methylation annotations that were available in 724 out of 911 brain tissue samples.

We first investigated the possible impact of methylation changes on gene expression. Figure 14 shows the histogram distribution of Pearson correlation coefficients between the sample-specific gene expression and gene methylation. We observe no strong correlation between expression and methylation, suggesting that methylation state does not play a large direct role in affecting expression magnitude.

While we did not see a strong correlation between methylation and expression magnitude, DNA methylation has been postulated to regulate variability in gene expression (Feinberg and Irizarry, 2010). CpG sites are discretely methylated or non-methylated at a single site, however methylation microarrays provide the methylation status of CpG sites of numerous cells in a sample simultaneously. Therefore, methylation is commonly reported as a ratio of methylated probes to

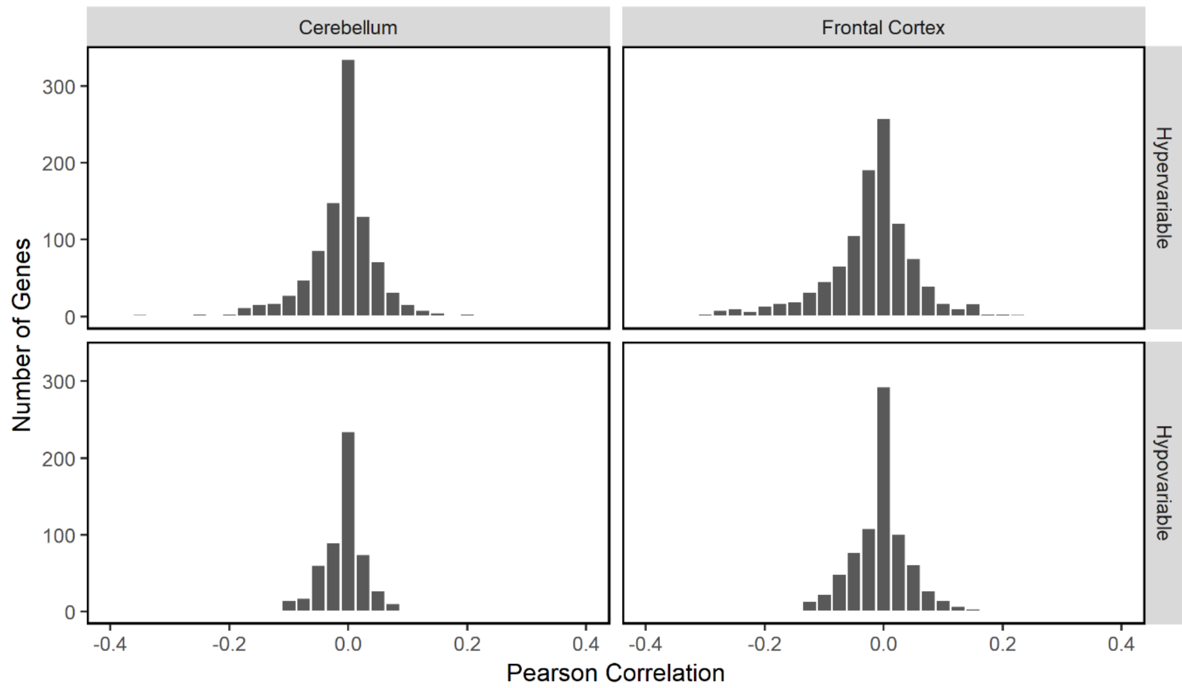


Figure 14. Expression and methylation correlation plot

Histogram of Pearson correlation coefficient between paired gene expression and gene methylation levels in Hypervariable and Hypovariable genes.

the overall number of probes in a sample, referred to as the β-value. In order to examine the relationship between methylation and gene expression, many analyses dichotomize the β-value using binary classifications (Bibikova et al., 2006; Wagner et al., 2014). If the β-value of a CpG site is low, then it is assumed to be hypomethylated. Conversely, the CpG site is hypermethylated if the β-value is large. However, a β-value around 0.5 indicates an equally likely chance that the site is methylated or non-methylated in different cells within a single sample. Rather than dichotomously classify these ambiguously-methylated sites as discretely methylated or non-methylated, we opted to use a tripartite classification system: non-methylated, medium methylation, and highly methylated states.

In order to differentiate between low, medium, and high methylation states in our samples, we modelled gene methylation using Gaussian mixture models for the mean methylation for each gene (Fig. 15). The distribution of gene methylation in both cerebellum and frontal cortex tissue was best modelled as a three-component system. The first component was composed of a sub-population Gaussian mixture while the second and third components were modelled as single Gaussian distributions. Genes whose methylation fell within the first component were classified as Non-Methylated genes. Genes were classified as Medium Methylated for those in the second component and Highly Methylated if they were in third. The distribution of methylation amongst the genes is predominantly bimodal with only a minority of genes being Medium Methylated (Fig. 16). In contrast, over 62% of cerebellum genes are non-methylated and 23% highly methylated. Similarly, 58% of frontal cortex genes are non-methylated and 22% are highly methylated.

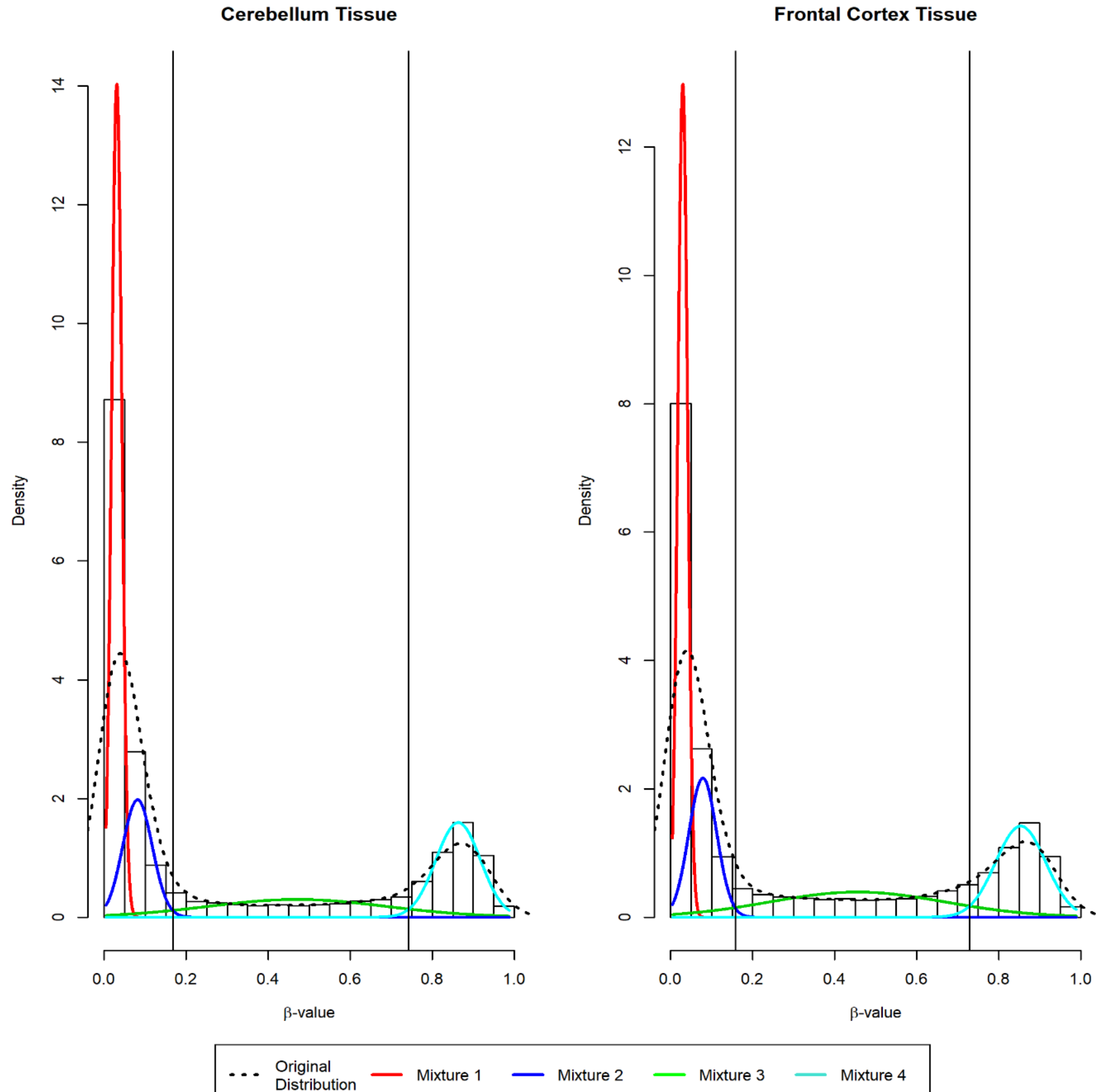


Figure 15. Gaussian mixture model of gene methylation

The vertical black lines represent the β-value threshold for the distinct methylation component. The first component state is composed of a sub-population Gaussian mixture model of mixture 1 and 2, while the second and third components were modelled as single Gaussian distributions, plotted as mixtures 3 and 4, respectively.

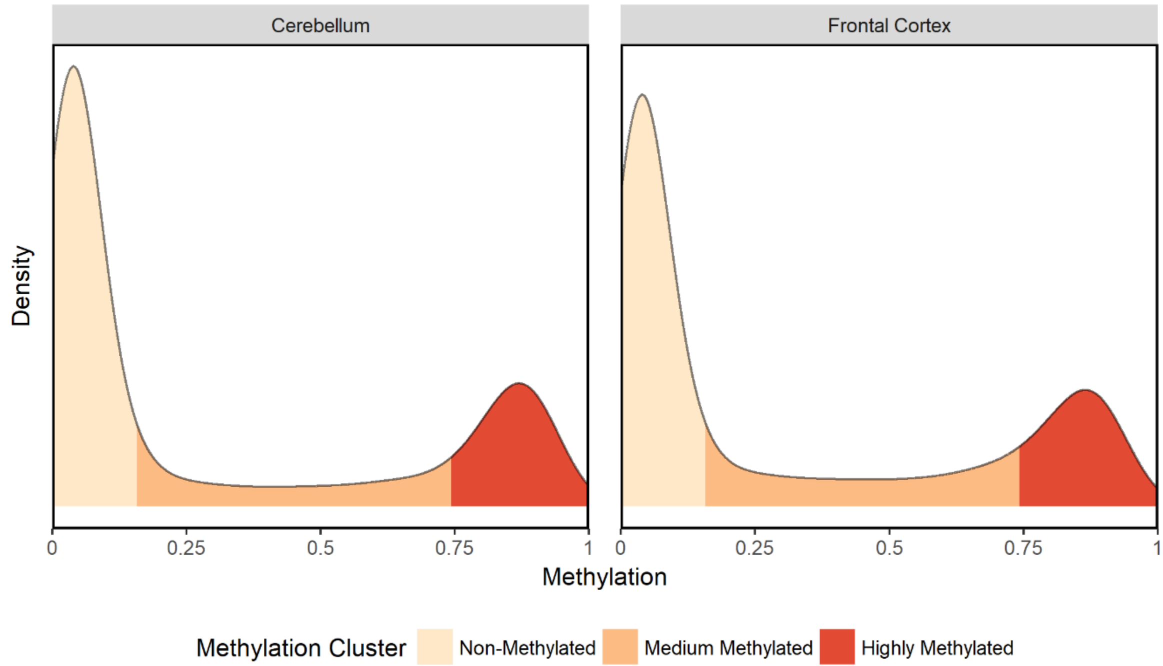


Figure 16. PDF of gene methylation

Gaussian mixture models were used to classify the genes into Non-, Medium-, and Highly methylated gene clusters.

3.3.2: Methylation Enrichment in EV Classes

Next, we explored the correlation between methylation and expression based on their tissue-specific EV classification. When we subset the methylation distribution by EV classification (Fig. 17), we observe that Hypovariable genes have a visibly different methylation pattern than Hypervariable or Non-Variable genes insofar as Hypovariable genes are visibly overrepresented in the Non-Methylated gene group compared to both the Hypervariable and Non-Variable genes.

To further quantify the overrepresentation of Hypovariable genes in the Non-Methylated gene group, we conducted a chi-squared test of independence between the methylation state clusters and the EV classifications (Table 7). Both the cerebellum and frontal cortex tissues exhibited a significant relationship between the methylation clusters and EV classifications ($p = 7.57 \times 10^{-36}$ and $p = 1.58 \times 10^{-59}$, respectively).

By examining the standardized residuals of the chi-square test of independence, we quantitatively confirmed the enrichment of Non-Methylated genes within the Hypovariable gene set. We also observe a significant enrichment of Highly Methylated genes in the Non-Variable gene set as well as an enrichment of Medium Methylated genes in the Hypervariable gene set. The high significance and non-overlapping enrichments across each of the three groups suggests that there is a strong relationship between methylation and EV classifications: low gene methylation is important for the tight expression constraint in Hypovariable genes while high gene methylation contributes to Non-variable expression.

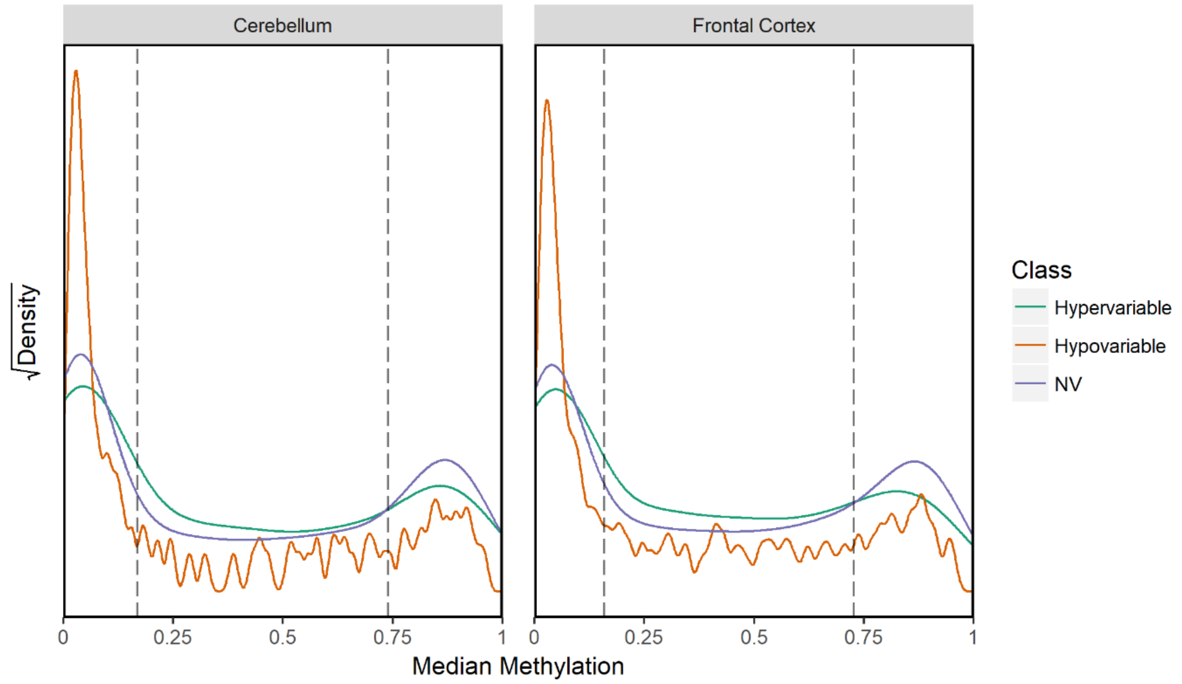


Figure 17. PDF of average gene methylation by EV classification

The dashed vertical lines represent the methylation state cluster cut-offs generated by the Gaussian mixture model. The y-axis is scaled by the square root of the methylation density.

Table 7. Methylation enrichment analysis

Pearson's Chi-squared test for methylation status in Hypervariable, Hypovariable, and Non-Variable gene sets. There is significant relationship between the two variables in both the cerebellum and the frontal cortex tissues ($\boldsymbol{p = 7.57 * 10^{-36}}$ and $\boldsymbol{p = 1.58 * 10^{-59}}$, respectively).

	Cerebellum Tissue			Frontal Cortex Tissue		
	Non-Methylated	Medium Methylated	Highly Methylated	Non-Methylated	Medium Methylated	Highly Methylated
Hypovariable	11.98	-5.69	-9.04	14.84	-7.11	-10.79
Non-Variable	-7.52	0.06	8.59	-10.00	-0.04	11.73
Hypervariable	0.07	4.21	-3.58	-0.23	6.23	-5.47

3.4: Effects of Age, Sex, and PMI on Variability

3.4.1: Linear Regression of Sex, PMI, and Aging (and EV)

To further understand the biological relevance of EV, we focused on the Hypervariable genes to identify potential mechanisms and downstream processes affected by decreased constraint on gene expression across the samples. We systematically analyzed EV as a function of sex, age, and PMI. The breast tissue dataset lacked these clinical annotations and was excluded from this analysis. We employed a probe-wise linear regression analysis to model the relationship between Hypervariable gene expression and age, sex, and PMI (Table 8). The resulting p-values were adjusted for multiple comparisons using the Benjamini-Hochberg procedure and considered significant when the adjusted p-value was less than 0.01.

Extended post-mortem intervals compromises sample RNA integrity and lead to degradation of labile RNA (Birdsill et al., 2011), suggesting that PMI might be a source of expression variability. Brain samples had PMI times ranging from 1 hour to 94 hours (mean = 36.14 hr), but we observe a negligible number of genes whose expression magnitude correlated strongly with PMI. This suggests that sample integrity is unlikely to be a source of EV changes. Somewhat more surprisingly, however, is the low number of genes that are correlated with sex. Only 25 out of 1640 Hypervariable cerebellum genes and 23 out of 1760 Hypervariable frontal cortex genes showed sex-dependent differences in EV.

While other studies have shown widespread sex differences in post-mortem adult brain gene expression (Clodfelter et al., 2006; Rinn et al., 2004; Trabzuni et al., 2013; Vawter et al., 2004; Xu et al., 2006), we conclude that in our analysis, Hypervariability is overwhelmingly sex-independent.

Table 8. Gene-wise multiple linear regression of Hypervariable EV and PMI, sex, and age

	PMI			Sex			Age		
	Up	Down	Total	Up	Down	Total	Up	Down	Total
Cerebellum	12	10	22	16	9	25	247	267	514
Frontal Cortex	8	15	23	18	5	23	373	354	727

This finding is supported by several studies, including a meta-analysis of 293 human and mouse microarray datasets that reported no significant difference phenotypic variability between males and females (Itoh and Arnold, 2015). This lack of differential variability is true for behavioral, electrophysiological, histological, and neurochemical measures in rats, even when the estrous cycle is not accounted for (Becker et al., 2016).

However, we observe that age has a substantial effect on gene expression variability. Age is correlated with over 31% of Hypervariable cerebellum genes and over 41% of Hypervariable frontal cortex genes, indicating that the expression of these genes becomes either more or less constrained during aging. In the cerebellum, there were 247 Hypervariable genes whose expression increased as a function of age and 267 genes with decreased expression. Similarly, the frontal cortex contained 373 genes with increased expression and 354 genes with reduced expression.

3.4.2: Heatmap Clustering

Given that age is correlated with a considerable number of Hypervariable genes, we classified the age of the samples in the cerebellum and frontal cortex tissues into three age clusters (Fig. 18) according to the total within-clusters sum of squares (WSS) and the Bayesian Information Criterion (BIC) using the mclust package in R (Scrucca et al., 2016). The oldest age cluster contained samples whose ages were between 58 and 98 ($\bar{x}_1 = 79$). The second cluster ranged between 32 and 57 years ($\bar{x}_2 = 45$), while the youngest age cluster contained samples aged 1 through 31 ($\bar{x}_3 = 17$). To further explore this age-dependent effect on EV, we examined the age-dependent changes in expression of the Hypervariable genes across the three clusters. In each tissue type, we labeled genes whose expression was positively correlated with age as "upregulated",

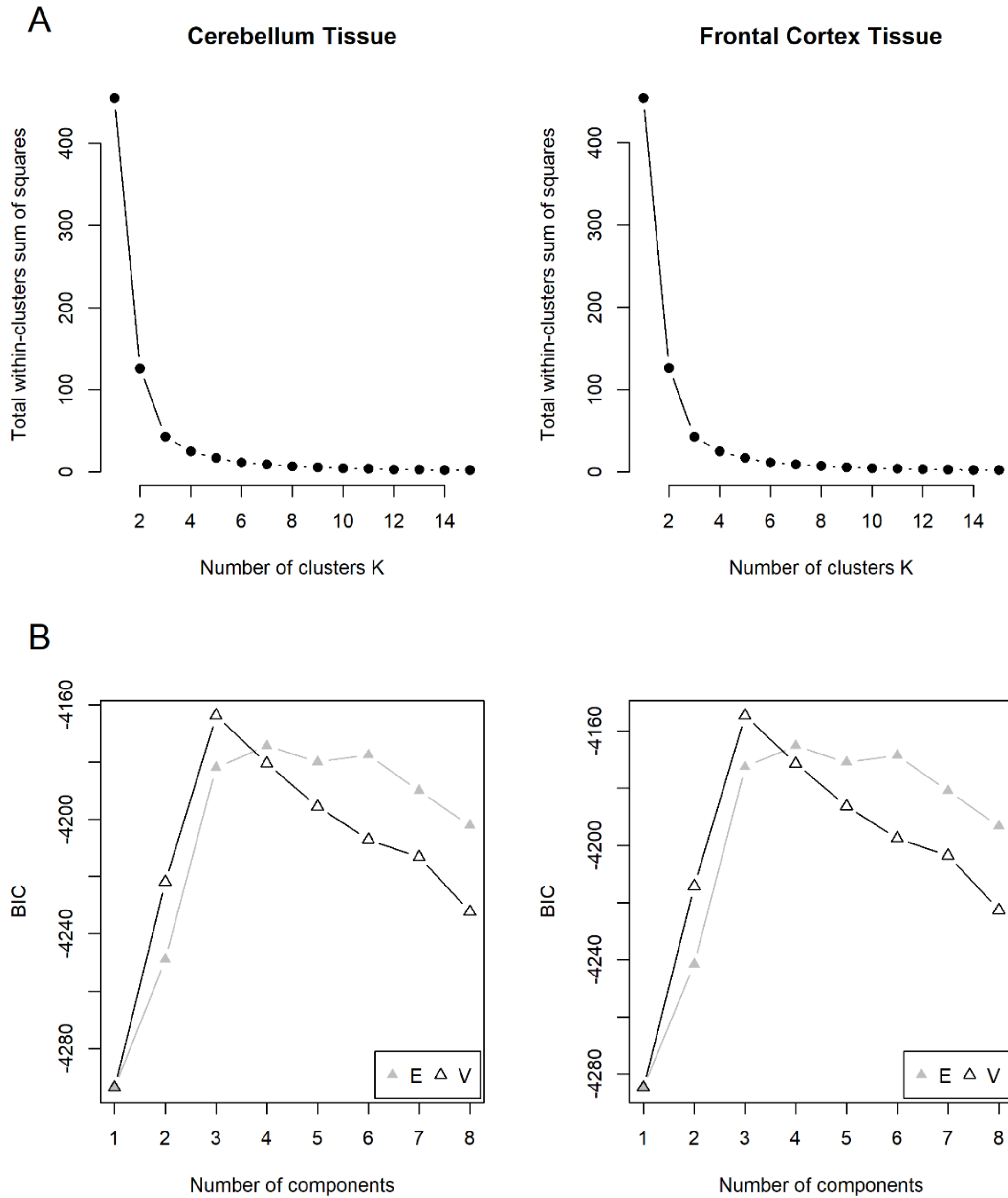


Figure 18. Optimal number of K-means age clusters

The choice of k clusters to segment sample age was identified using two different methods, both indicating three as the optimal number. (A) The elbow method choses the k value at which the decrease in total within-clusters sum of squares from each additional cluster begins to drop, leading to marginal gains in performance. (B) Finding the local maxima of the Bayesian Information Criterion (BIC) using the mclust package.

while the negatively correlated genes were termed “downregulated”. Then, we used a hierarchical clustering method with an expression heatmap to visualize how these upregulated and downregulated genes are expressed throughout the age clusters (Fig. 19-20). The resulting gene hierarchical trees were clustered into gene sets via manual hierarchical tree cutting.

3.4.3: Functional Enrichments of Age-Regulated Hypervariable Genes

While the cerebellum is generally considered a regulator of motor processes, it is also implicated in cognitive and non-motor functions (Harada et al., 2013). Many of these age-dependent upregulated Hypervariable genes corroborate previous studies exploring the relationship between brain aging and changes in gene expression, including cellular responses to chemical stimuli (gold cluster). In particular, reactive oxygen and nitrogen species have been shown to change ion transport channel activity, and serve as an important mechanism in brain aging (Annunziato et al., 2002). While all the genes selected were age-regulated, some genes exhibit outlier samples whose expression remains high across all genes in the dark orange cluster, regardless of age. These genes are more likely to be overexpressed in the samples as age increases and are enriched for peripheral nervous system neuron development and neuron apoptotic pathways. Similar enrichments of neurogenic and chemical stimuli response pathways are seen in the upregulated frontal cortex genes (gold cluster). The dark orange cluster in the upregulated frontal cortex age-dependent genes exhibited an entirely sample-specific over- or under-expression of genes. These bimodally expressed genes are enriched for glial cell differentiation, adenosine receptor signaling pathways, and antigen processing. Lastly, we see a random scattering of expression in the yellow cluster of the frontal cortex heatmap that steadily increases with age. These genes are enriched for glial cell differentiation, cellular response to alcohol, and defense responses to fungus.

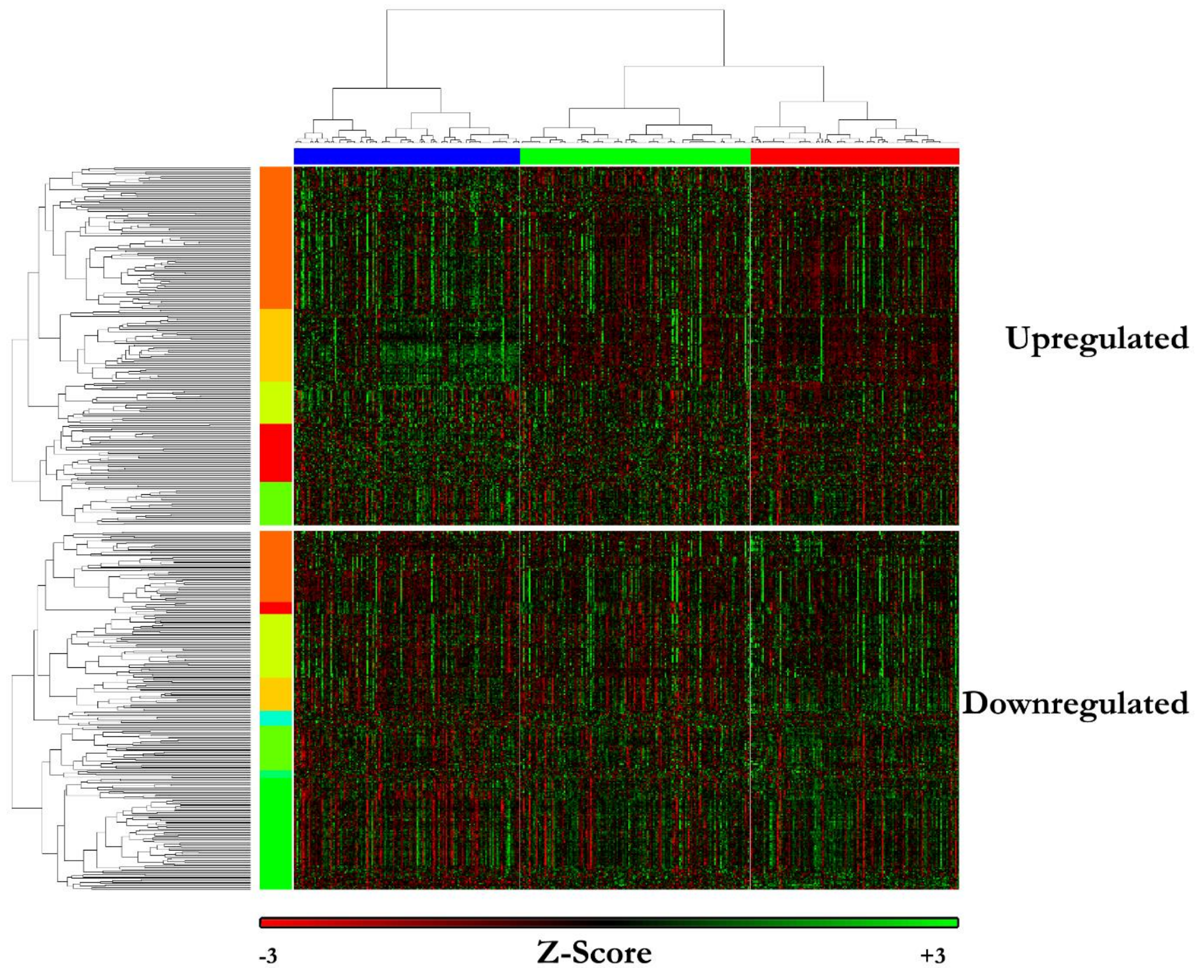


Figure 19. Cerebellum Hypervariable heatmap

The vertical axis represents the age-regulated Hypervariable genes while the samples were clustered by age and plotted on the horizontal axis. The top heatmaps represent the positively correlated age-regulated genes while the bottom heatmaps represent the negatively correlated age-regulated genes. The age clusters decrease in age from left to right and correspond to the following age ranges: $\bar{x}_1 = 79\ [58,98]$, $\bar{x}_2 = 45\ [32,57]$, and $\bar{x}_3 = 17\ [1,31]$.

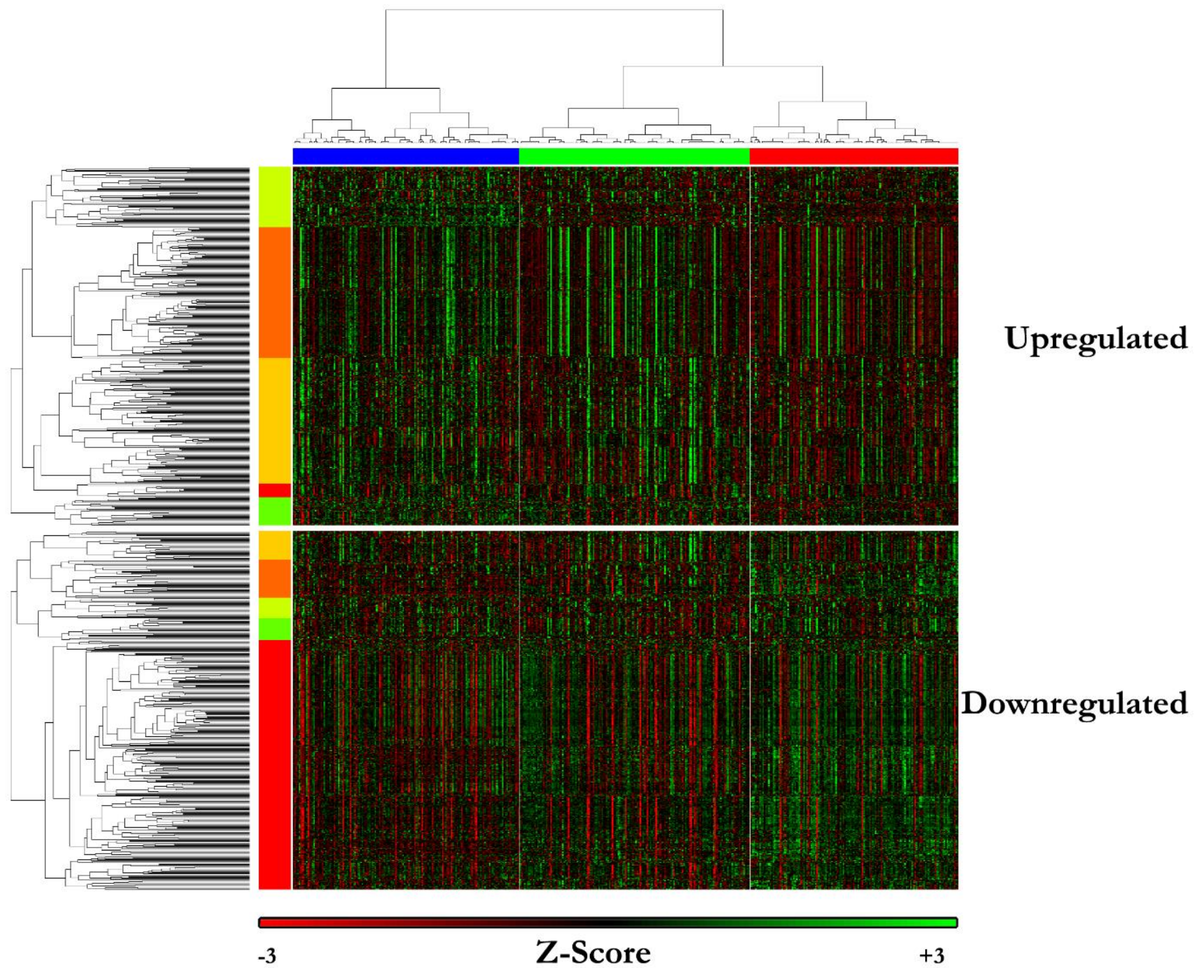


Figure 20. Frontal Cortex Hypervariable heatmap

The vertical axis represents the age-regulated Hypervariable genes while the samples were clustered by age and plotted on the horizontal axis. The top heatmaps represent the positively correlated age-regulated genes while the bottom heatmaps represent the negatively correlated age-regulated genes. The age clusters decrease in age from left to right and correspond to the following age ranges: $\bar{x}_1 = 79\ [58{,}98]$, $\bar{x}_2 = 45\ [32{,}57]$, and $\bar{x}_3 = 17\ [1{,}31]$.

Most of the downregulated age-dependent Hypervariable genes in the cerebellum fall into the green cluster, whereby certain individuals exhibit reduced expression across genes in the cluster and increases in sample frequency as age increases. These genes are involved in leukocyte-mediated immunity and defense responses to other organisms, which is expression is downregulated by age is supported by previous studies (Montecino-Rodriguez et al., 2013). Interestingly, the yellow cluster exhibited U-shaped expression levels, whereby the lowest expression is seen in the middle age cluster. These genes are enriched for optic nerve development, response to interferon-gamma, and synaptic signalling. In the frontal cortex, the majority of downregulated age-dependent genes fall in the red cluster, and are enriched for ion transport, cell morphogenesis, and trans-synaptic signalling. Overall, the functional annotations of the age-regulated Hypervariable gene clusters suggest that population EV is one outcome of age-dependent gene expression changes.

4: Discussion

Gene expression variability across a population is the cumulative result of extrinsic environmental factors and intrinsic stochastic factors. A fundamental issue in biology is understanding the causal factors of gene expression variability within an individual organism and between isogenic and genetically dissimilar individuals within a given population. This variability in gene expression reflects phenotypic heterogeneity and is thought to be under driven by natural selective pressures. Regulation of EV has been shown to be directly involved in many biological processes, including dosage sensitivity, disease susceptibility, and embryonic development. In this report, I investigated potential sources of gene expression variability in human breast, cerebellum, and frontal cortex tissues and the effects of EV on phenotypic variation in a human population.

In particular, the effects of methylation and aging were examined as correlates of EV in human populations.

Prior to examining the potential correlates of EV, I first developed a robust and generalizable method of measuring expression variability in gene expression profiles generated through microarray platforms. Despite numerous available methods to measure variability, including SD and CV, there is no standardized method in the field. In addition, each method has their unique disadvantages as well as a common drawback of being sensitive to outliers and opt to remove outliers entirely. As such, these methods are not suitable for examining large-scale studies of variability and noise. The new EV method used in this work quantified gene expression variability in a manner suitable for large-scale microarray studies and is resistant to outlier samples. These outlier measurements can originate from technical errors during the DNA microarray measurement procedures, or from specific individuals in a population with aberrant gene expression levels compared to the rest of the population. Regardless of the source of noise associated with these outliers, they should not be excluded from any analysis pertaining to the exploration of transcriptional noise or expression variability. Once gene expression variability could be quantified in a non-biased manner, I could thus perform a comprehensive analysis of potential causal factors that contribute to changes in expression variability in humans in different tissue types. I first examined the overarching EV differences between the three tissue types to establish baseline expression patterns specific to each tissue type. Next, I investigated the functional role, cellular localization, essentiality status, and methylation levels associated with the genes with low, medium, and high levels of EV. Lastly, I explored the inter-individual causes of high variability using clinical annotations available in the samples, including the age of the person, their sex, and the post-mortem interval time prior to sample collection.

4.1: Characteristics of EV

A crucial consideration to take into account when exploring expression variability in a population is the probability distribution of gene expression magnitude. The expression levels of the large majority of genes in the microarray datasets do not follow a normal distribution. Conducting a statistical test requires assumptions about the population distribution parameters. If the data meets the assumptions of the parametric test (ie. normality of the data), then it is safe to use a parametric statistical test. Otherwise, a non-parametric statistical test is required. Using a parametric statistic on non-normal data results in less powerful statistical tests, increasing the probability of making a type II error, whereby the test wrongly fails to reject the null hypothesis (Columb and Atkinson, 2016). Thus, we cannot use standard parametric statistical tests or non-robust measures of variability.

Rather than using the median absolute deviation, many papers investigating expression variability erroneously use the coefficient of variation, a non-robust measure of variability. The use of MAD as a robust statistic of variability ensures an outlier-resistant method of calculating EV. This eliminates the need to remove the outliers, while ensuring that the resulting measure of variability is not unduly affected by extreme values. Taking these measures to include the outliers in a statistically correct manner ensures that gene expression variability is representation of a population-wide trend of gene expression levels. This is of particular importance when investigating genes with high expression variability as the resulting variability values may reflect a statistical artefact of a few outlier samples, rather than a wide range of permissible expression levels.

The genes in the datasets were classified into three distinct EV classes, however, there are no "true" EV class labels of EV we can compare our results against. As a result of this

unsupervised classification method, we cannot assign a numeric p-value for the classification validity or accuracy. While this accuracy cannot be quantified, the split-retest classification method nevertheless considerably improves the model by controlling the trade-off between specificity and sensitivity (Fig. 6). By adjusting the binomial test p-value to a higher threshold, the model's specificity increases at the cost of decreased sensitivity. In the context of EV classification, high specificity/low sensitivity results in fewer genes being classified as Hypervariable or Hypovariable but with much higher confidence and higher theoretical positive predictive value. In other words, we reduce the number of false positive EV classification of Hyper- or Hypovariability (type I error) with the assumption of increased false negative (type II error) classifications. Conversely, a lower p-value threshold results in higher sensitivity but lower specificity, with a higher number of false positives and less false negatives. Typically, favoring one over the other done at the discretion of the researcher, and the level of uncertainty they are willing to assume. Our choice in binomial test p-value of 0.5 for this analysis was specifically chosen to balance these two antagonistic measures and did not bias towards specificity or sensitivity. It inherently accepts a certain amount of false positive Hypervariable and Hypovariable EV classifications but is still stringent enough to test hypotheses regarding the biological importance of expression variability.

While the robust statistics used in this analysis (i.e. MAD and EV) are outlier-resistant and are suitable for non-normal distributions, they cannot be used to accurately model multimodal distributions. In the case of Hypervariable genes, the wide range of expression levels expressed in a population could manifest itself as either unimodal or multimodal distributions. Unimodal distributions represent a continuum of expression levels in a population, whereby the probability density function has a single peak. On the other hand, multimodal distributions represent two or more distinct expression states that an individual cell may switch between (Balaban et al., 2004). This bimodality in gene expression is represented by a probability distribution containing multiple

distinct peaks. It can be modelled accurately using a mixture model of several distributions, whereby each distribution within the mixture model is representative of a unique phenotypic state (Figure 8).

Since Hypervariability is predominantly unimodal, we can infer that high EV is the result of a wide range of permissible expression levels in a population rather than the existence of discrete gene expression or phenotypic subpopulations. In the case of organismal phenotypes, the benefits of a wide unimodal phenotypic distribution is highest in environments with very rapid or very noisy environmental changes (Garcia-Bernardo and Dunlop, 2016). As such, genes functionally associated with signalling pathways exhibit high expression variability in order to respond to a dynamic and rapidly changing environment at the cell periphery.

While there is limited overlap of specific genes within the corresponding EV gene sets of different tissues, the Hypovariable and Hypervariable gene sets across the three tissue types have similar functional enrichments and cellular protein localizations. Specifically, the Hypovariable gene sets were enriched for nucleic acid and metabolic regulatory pathways. As such, these complex and dosage-sensitive gene networks require tight regulation of their expression to function correctly, regardless of tissue or cell type. In addition to performing house-keeping processes, the proteins associated with Hypovariable genes are also localized to the cell interior, validating previous studies of Hypovariability (Alemu et al., 2014; Mar et al., 2011; Newman et al., 2006).

Conversely, genes with Hypervariable EV expression tend to localize at the cell periphery and at membranes. These lowly constrained genes are enriched for cell surface signalling pathways and tissue development, including tissue remodeling and ion transport. These results corroborate previous findings that genes with high expression variability are functionally and physically

localized in membrane, transmembrane, and extracellular matrix domains (Alemu et al., 2014; Mar et al., 2011; Newman et al., 2006).

The split-retest methodology used in the analysis increased the specificity of the classification model, resulting in a large number of genes initially classified as variable to be re-classified as Non-variable, referred to as "pseudo-variable" genes. These pseudo-Hypervariable and pseudo-Hypovariable genes exhibit highly similar functional enrichments to the true Hypervariable and Hypovariable GO terms enrichment (Supplementary Figure 5). This is likely a result of the specificity/sensitivity trade-off of the novel classification method. There is an inherent understanding that reducing the sensitivity results in more false negatives, which are represented by pseudo-variable genes whose associated pathways are identical to the "true" variable gene sets.

Comparing gene-specific EV between the tissue types shows a larger similarity between the cerebellum and frontal cortex tissues, both of which are located within the brain. We see a much smaller similarity between either brain tissue (cerebellum or frontal cortex) compared to breast tissue, suggesting a highly tissue-specific component of EV. Given the limited overlap of specific genes in corresponding gene sets and similar overarching functional processes across the three tissue types, we therefore propose that gene expression variability is highly associated with tissue-specific developmental pathways and establishing tissue identity.

4.2: Essentiality & Methylation

Single-cell studies of yeast and *E. coli* have shown that essential genes tend to have lower expression variability than non-essential genes. This tight regulation of gene expression levels is likely the result of selective pressures to minimise noise and fluctuations in protein levels of dosage-sensitive genes (Lehner, 2008, 2010; Newman et al., 2006; Singh, 2013). These fluctuations

can be highly lethal to the cell if the dosage of protein falls below an optimal and heritable threshold. In addition, noise-sensitive non-essential genes tend to cluster with noise-sensitive essential genes in persistently open, low-noise chromatin domains (Batada and Hurst, 2007).

Our results corroborate these findings as we see a significant enrichment of essential genes in the Hypovariable gene sets in breast, cerebellum, and frontal cortex tissues. In addition, the Hypovariable gene sets were significantly non-methylated, a common characteristic of low-noise chromatin domains. By remaining in an open chromatin configuration, the genes that occupy the low-noise domains are unaffected by transcriptional bursting due to the switching of chromatin between open and closed configurations (Becskei et al., 2005; Raj et al., 2006; Raser and O'Shea, 2005). While chromatin configuration is correlated with changes in expression variability and fluctuations in protein abundance, the open chromatin configuration does not guarantee gene expression. Previous studies showed no significant differences in the mean expression levels of essential and non-essential genes in and out of low-noise chromatin domain clusters (Batada and Hurst, 2007). Thus, epigenetic changes in methylation are able to affect the expression variability of a gene independent of the magnitude of gene expression. This uncoupling between expression variability and expression magnitude is precisely identical to the considerations taken when developing the new EV method. By ensuring that EV exhibits no relationship with overall expression magnitude (Figure 4), we were able to examine the effects of methylation solely as a correlate of EV without the confounding effects of changes in gene expression magnitude.

While the enrichment of essential genes in the Hypovariable gene sets is in agreement with previous findings in yeast and *E. coli* showing that essential genes are likely to have low expression variability (Batada and Hurst, 2007; Lehner, 2008, 2010; Newman et al., 2006; Singh, 2013), we were surprised to detect a significant number of essential genes amongst the Hypervariable genes

in breast, cerebellum, and frontal cortex tissue. Generally, fluctuations in protein levels associated with high expression variability may inactivate gene activity by reducing the protein level below a required threshold. This inactivation of dose-sensitive essential genes during embryonic development leads to pre- or neonatal fatality in mice and humans (Georgi et al., 2013). Functional enrichment analyses of Hypervariable and Hypovariable essential genes in breast, cerebellum, and frontal cortex tissues indicate that the essential genes in both gene sets are indeed involved in developmental pathways. It is not fully clear as to why highly variable expression could be tolerated in obligate developmental pathways since *a priori* expression of developmental genes should be tightly regulated to ensure proper development.

One possibility is that these "essential" genes are required for embryonic development but have different post-embryonic roles and may not be essential post-natally. Mounting evidence suggests that gene essentiality is not a static or binary property of a gene. Instead, it is both context-dependent and highly evolvable (Rancati et al., 2018). In particular, it is dependent on the environment and/or genetic context. Therefore, a gene is that is considered essential in a pre-natal embryonic context may not be essential in a post-natal aging adult. Studies involving CRISPR screens are already underway to develop an atlas of gene essentiality, or an "Essentialome", to better understand the key aspects of cellular essentiality and pluripotency (Yilmaz et al., 2018).

Alternatively, it is possible that the essential genes in this analysis are dose-sensitive in mice specifically, and not dose-sensitive in humans. The list of essential protein-coding genes provided by the CCDS database consists of human orthologs of essential genes found in mouse models (Georgi et al., 2013). If the human orthologs are not dose-sensitive, only a certain level of baseline expression is required and expression above this baseline level might be well tolerated. While generally incomplete, the curators of the CCDS database posit that the set of human orthologs

exhibit characteristics fully consistent with known essential human genes. These claims are supported by other studies claim that genes that are essential in one species strongly tend to be essential in other species (Amsterdam et al., 2004; Batada and Hurst, 2007; Kamath et al., 2003).

One additional possibility is that protein abundance of essential genes could be regulated translationally rather than transcriptionally. Inefficient translation of certain genes may have been selected for during evolution to prevent fluctuations in protein concentrations (Kærn et al., 2005; Ozbudak et al., 2002). Perhaps a combination of these factors is at play. Regardless, tight regulation of these essential pathways associated with Hypervariable genes is not required in fully developed adults, allowing for a wider range of expression values within the human population, resulting in high expression variability.

We propose a model for methylation-dependent expression variability where the highly constrained levels of Hypovariable gene expression require genes to be non-methylated. We speculate that the lack of methylation allows transcriptional regulators that require non-methylated DNA for binding are able to tightly regulate gene expression, resulting in low expression variability.

4.3: Effects of Aging on EV

One key factor to consider when examining expression variability in a population is the differences between the individuals. In addition to environmental and intrinsic factors, high variability can arise from specific inter-individual differences, including the persons' sex and age.

While previous studies have shown clear evidence regarding sex-specific neurophysiological gene differences including brain structure and neurochemistry (Cahill, 2006; Cosgrove et al., 2007; Jazin and Cahill, 2010), little is known about the sex-specific differences in expression variability. Despite the widespread sexually dimorphic gene expression found in both humans and mice, we

did not observe any substantial sex-dependent differences associated with global Hypervariability. While significant differences in gene expression between males and females have been previously identified, these differentially expressed genes did not exhibit differential variability. The majority of sex-dependent genes that were detected through the gene-wise multiple linear regression analysis (Table 8) are located on the Y-chromosome. Despite the lack of Y chromosomes in females, the non-zero expression in these samples can be attributed to background intensity during the microarray scanning. Alternatively, these Y-linked genes may be located in X transposed or degenerative regions of the chromosome, which exhibit a high homology to the X chromosome. These homologs may incorrectly hybridize to the Y-linked gene probes during the microarray process (Johansson et al., 2015). The genes located on the X-degenerate region are mostly associated with housekeeping functions (Bachtrog, 2013; Lahn and Page, 1997; Skaletsky et al., 2003), a characteristic of Hypovariability rather than the Hypervariability. Lastly, several Y-linked genes, including KDM5D and EIF1AY, whose mean expression was significantly larger than the female background expression levels were removed from the analysis due to their bimodal distribution.

Another cause of intra- and inter-individual differences in cellular morphology and gene expression is aging. Since Hypovariable genes exhibit low expression variability, their expression levels are assumed to be age-independent. However, an important conclusion of this study is that many Hypervariable genes have age-dependent expression variability: that is, their significantly expression increases or decreases during natural aging.

In particular, there are three gene clusters whose expression is highly age-dependent and correlated with significant neurological diseases. The first gene cluster is enriched for neuroinflammation and immunological functions in the brain. Inflammatory mechanisms are

linked to functional and mental impairments and pathogenesis in the elderly, and play a key role in neurodegenerative diseases, including Alzheimer's disease (CHEN et al., 2016; Heneka et al., 2015). The second gene cluster is functionally involved in synaptic dysfunction and neuroendocrine signalling. The enriched functional annotations in the last gene cluster is associated with a compensatory neurogenesis mechanism associated with aging.

One main cause of accelerated brain aging and a causal factor of neurodegeneration is neuroinflammation and immunosenescence, a reduction in immunological functions in the brain (Lucin and Wyss-Coray, 2009; Streit and Xue, 2010). We see evidence of downregulated immune responses in the cerebellum, specifically leukocyte-mediated immunity, defense responses to other organisms, and interferon-gamma response pathways. Many studies also suggest that aging is associated with the upregulation of inflammatory responses (Singh et al., 2011), which is a pathogenic mechanism implicated in many age-related diseases, including cardiovascular disease, Alzheimer's disease, and Parkinson's disease (Wu and Meydani, 2008). Consistent with this idea, we see an enrichment of acute inflammatory response in the cerebellum gold cluster. Normal aging is commonly accompanied by a chronic, low-grade inflammation called 'inflammaging". This systemic increase in inflammatory factors occurs in both the immune and central nervous system, and is a major contributor of mortality and morbidity in the elderly (Di Benedetto et al., 2017; Fagiolo et al., 1992, 1993).

Another mechanism that has been implicated with age-related diseases is "synaptopathy", whereby disruptions in synaptic structure and function are major causes of brain disorders. The concept of synaptopathies as brain diseases is a fairly recent concept, but has been looked at as a potential mechanism of many common neurological diseases, including Alzheimer's disease, Huntington's disease, Parkinson's disease, schizophrenia, and autism (Brose et al., 2010; Lepeta et

al., 2016). In addition to increased inflammatory responses, synaptopathy is also associated with changes in neuroendocrine signalling (Azpurua and Eaton, 2015; Hebert et al., 2001; Levy et al., 2002). We see a downregulation of ion transport and trans-synaptic signaling in the frontal cortex, which are key components of neurotransmission and membrane excitability, and whose downregulation likely causes deficiencies in these complex processes.

Furthermore, we see an upregulation of genes associated with glial cell differentiation in the frontal cortex across multiple gene clusters. Initially thought of as cells that merely support neurons, emerging research shows that neuron-astrocyte-microglia interactions are crucial for the functional organization of the brain (Cerbai et al., 2012). Astrocytes, a type of glial cell, are the most abundant cell in the brain and play a role in many processes including maintaining osmotic and ionic balance for neurons, recycling of neurotransmitters, and metabolite homeostasis (Buffo et al., 2010; Kimelberg, 2005; Kimelberg and Nedergaard, 2010; Sajja et al., 2016). Both astrocytes and oligodendrocytes, another type of glial cell, have been shown to shift regional expression patterns upon aging, and are better predictors of biological age than neuronal-specific genes (Soreq et al., 2017). While astrocytes confer a neuroprotective effect with the neuronal synapses they associate with, they can also display a neurotoxic effect. This effect is dependent on the environmental enrichment of cytokines present within the brain and can be exasperated by chronic neuroinflammation. Inflammaging in particular has been shown to transform glial cells towards a highly pro-inflammatory state, resulting in a loss of neuroprotective function, neuronal dysfunction, and eventually cognitive impairments and neurodegeneration (Di Benedetto et al., 2017; Singhal et al., 2014).

Another potential mechanism for synaptic dysfunction and neuroendocrine signalling disruption is the cellular and molecular changes and injury mechanisms that occur after traumatic

brain injury (TBI). TBI is a complex, and progressive condition that consists of a primary injury mechanism, often due to direct impact, as well as a secondary molecular cascade injury mechanism (Sajja et al., 2016). In both experimental and clinical settings, the synaptic structures have been shown to be highly vulnerable to damage due to direct or indirect concussion attack following TBI (Logue et al., 2016; Wen et al., 2017; Zhang et al., 2014). While the exact mechanism is unknown, it is clear that the effects of glial dysfunction due to TBI affects synaptic plasticity and neuroendocrine signalling, resulting in memory loss (Croft et al., 2015; Gundersen et al., 2015; Walker and Tesco, 2013), neuroinflammation (Hsieh and Yang, 2013), and neuronal impairment (D'Ambrosio et al., 1999; Pietrobon and Moskowitz, 2014).

Without examining the mechanistic role of individuals genes, it is difficult to determine if changes in gene expression, either upregulation or downregulation, result in repression or activation of their associated pathways. For example, we see an upregulation in neurogenesis-associated genes during aging in both the cerebellum and the frontal cortex. Despite the common theory that neurodegeneration is a ubiquitous effect of normal brain aging, an emerging concept in neuroscience is that homeostatic plasticity of neurons is maintained through local adjustments of neural activities (Braegelmann et al., 2017). This overexpression of genes in neurogenesis pathways whose function is known to decline over time may be a compensatory mechanism for an inefficient, aging system. Within the cerebellum, a decline in neuronal function that occurs with aging may cause an upregulation of genes associated with neurogenesis pathways. In addition to mitigating neuronal dysfunction, localized increases in neurogenesis may be induced in response to cerebral diseases or acute injuries for self-repair (Galvan and Jin, 2007). These acute injuries include injuries resulting in TBI, whereby the severity of the injury is correlated with increased levels of neural stem cells and neurogenesis (Wang et al., 2016). Taken all these factors in consideration, we can infer that age-regulated synaptic dysfunction and disruptions of

neuroendocrine signalling is largely caused by cumulative damage to the brain tissue throughout a person' life.

One confounding variable associated with our hypothesis of compensatory neurogenesis is the neurogenesis-depression hypothesis. This hypothesis posits that neurogenesis is negatively regulated by stressful experiences and positively regulated by treatment with antidepressant drugs, and that the rate of neurogenesis plays a key role in the treatment of depression (Hanson et al., 2011). Previous studies of neurogenesis research has shown that chronic antidepressant usage can result in an increase in neurogenesis (Malberg et al., 2000; Petrik et al., 2012), suggesting that psychopharmaceuticals can alter neurochemistry and mimic compensatory anti-aging responses.

Overall, Hypervariability is significantly correlated with aging, and play an important role in neuroinflammation, synaptopathy, and neurogenesis in the brain. In particular, the increases in low-grade chronic inflammation, known as inflammaging, coupled with a gradual age-associated deterioration of the immune system leads to a shift of glial cells to pro-inflammatory states and a deficiency in neurotransmission and membrane excitability. These age-dependent expression changes in Hypervariable genes could reflect a loss of regulatory control or a response to a deteriorating or inefficient biological pathway.

Conclusion

In summary, our work shows that gene expression variability in the human population is likely to be important in tissue development and identity, methylation, and in natural biological aging. As such, the EV of a gene is an important functional characteristic of the gene itself. Therefore, the classification of a gene as one with Hypervariability or Hypovariability in a human

population or in a specific tissue should be useful in the identification of important genes that functionally regulate development or disease.

References

Afsari, B., Geman, D., and Fertig, E.J. (2014). Learning Dysregulated Pathways in Cancers from Differential Variability Analysis. Cancer Inform. *13*, 61–67.

Alemu, E.Y., Carl, J.W., Corrada Bravo, H., and Hannenhalli, S. (2014). Determinants of expression variability. Nucleic Acids Res. *42*, 3503–3514.

Amador-Noguez, D., Zimmerman, J., Venable, S., and Darlington, G. (2005). Gender-specific alterations in gene expression and loss of liver sexual dimorphism in the long-lived Ames dwarf mice. Biochem. Biophys. Res. Commun. *332*, 1086–1100.

Amsterdam, A., Nissen, R.M., Sun, Z., Swindell, E.C., Farrington, S., and Hopkins, N. (2004). Identification of 315 genes essential for early zebrafish development. Proc. Natl. Acad. Sci. U. S. A. *101*, 12792–12797.

Annunziato, L., Pannaccione, A., Cataldi, M., Secondo, A., Castaldo, P., Di Renzo, G., and Taglialatela, M. (2002). Modulation of ion channels by reactive oxygen and nitrogen species: a pathophysiological role in brain aging? Neurobiol. Aging *23*, 819–834.

Ansel, J., Bottin, H., Rodriguez-Beltran, C., Damon, C., Nagarajan, M., Fehrmann, S., François, J., and Yvert, G. (2008). Cell-to-Cell Stochastic Variation in Gene Expression Is a Complex Genetic Trait. PLoS Genet. *4*.

Arkin, A., Ross, J., and McAdams, H.H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. Genetics *149*, 1633–1648.

Asada, M., Ebihara, S., Numachi, Y., Okazaki, T., Yamanda, S., Ikeda, K., Yasuda, H., Sora, I., and Arai, H. (2008). Reduced tumor growth in a mouse model of schizophrenia, lacking the dopamine transporter. Int. J. Cancer *123*, 511–518.

Azpurua, J., and Eaton, B.A. (2015). Neuronal epigenetics and the aging synapse. Front. Cell. Neurosci. *9*.

Bachtrog, D. (2013). Y chromosome evolution: emerging insights into processes of Y chromosome degeneration. Nat. Rev. Genet. *14*, 113–124.

Bahar, R., Hartmann, C.H., Rodriguez, K.A., Denny, A.D., Busuttil, R.A., Dollé, M.E.T., Calder, R.B., Chisholm, G.B., Pollock, B.H., Klein, C.A., et al. (2006). Increased cell-to-cell variation in gene expression in ageing mouse heart. Nature *441*, 1011–1014.

Balaban, N.Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004). Bacterial Persistence as a Phenotypic Switch. Science *305*, 1622–1625.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. *41*, D991–D995.

Batada, N.N., and Hurst, L.D. (2007). Evolution of chromosome organization driven by selection for reduced gene expression noise. Nat. Genet. *39*, 945–949.

Batenchuk, C., St-Pierre, S., Tepliakova, L., Adiga, S., Szuto, A., Kabbani, N., Bell, J.C., Baetz, K., and Kærn, M. (2011). Chromosomal Position Effects Are Linked to Sir2-Mediated Variation in Transcriptional Burst Size. Biophys. J. *100*, L56–L58.

Becker, J.B., Prendergast, B.J., and Liang, J.W. (2016). Female rats are not more variable than male rats: a meta-analysis of neuroscience studies. Biol. Sex Differ. *7*.

Becskei, A., Kaufmann, B.B., and Oudenaarden, A. van (2005). Contributions of low molecule number and chromosomal positioning to stochastic gene expression. Nat. Genet. *37*, 937–944.

Benaglia, T., Chauveau, D., Hunter, D.R., and Young, D.S. (2009). mixtools: An R Package for Analyzing Mixture Models. J. Stat. Softw. *32*.

Bibikova, M., Lin, Z., Zhou, L., Chudin, E., Garcia, E.W., Wu, B., Doucet, D., Thomas, N.J., Wang, Y., Vollmer, E., et al. (2006). High-throughput DNA methylation profiling using universal bead arrays. Genome Res. *16*, 383–393.

Birdsill, A.C., Walker, D.G., Lue, L., Sue, L.I., and Beach, T.G. (2011). POSTMORTEM INTERVAL EFFECT ON RNA AND GENE EXPRESSION IN HUMAN BRAIN TISSUE. Cell Tissue Bank. *12*, 311–318.

Blake, J.A., Eppig, J.T., Kadin, J.A., Richardson, J.E., Smith, C.L., and Bult, C.J. (2017). Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. Nucleic Acids Res. *45*, D723–D729.

Blinka, S., and Rao, S. (2017). Nanog Expression in Embryonic Stem Cells – An Ideal Model System to Dissect Enhancer Function. BioEssays News Rev. Mol. Cell. Dev. Biol. *39*.

Braegelmann, K., Streeter, K., Fields, D., and Baker, T. (2017). Plasticity in respiratory motor neurons in response to reduced synaptic inputs: a form of homeostatic plasticity in respiratory control? Exp. Neurol. *287*, 225–234.

Brose, N., O'Connor, V., and Skehel, P. (2010). Synaptopathy: dysfunction of synaptic function? Biochem. Soc. Trans. *38*, 443–444.

Bryant, P.A., Smyth, G.K., Robins-Browne, R., and Curtis, N. (2011). Technical Variability Is Greater than Biological Variability in a Microarray Experiment but Both Are Outweighed by Changes Induced by Stimulation. PLOS ONE *6*, e19556.

Buffo, A., Rolando, C., and Ceruti, S. (2010). Astrocytes in the damaged brain: Molecular and cellular insights into their reactive response and healing potential. Biochem. Pharmacol. *79*, 77–89.

Butte, A.J., Dzau, V.J., and Glueck, S.B. (2001). Further defining housekeeping, or "maintenance," genes Focus on "A compendium of gene expression in normal human tissues." Physiol. Genomics *7*, 95–96.

Cahill, L. (2006). Why sex matters for neuroscience. Nat. Rev. Neurosci. *7*, 477–484.

Carey, L.B., van Dijk, D., Sloot, P.M.A., Kaandorp, J.A., and Segal, E. (2013). Promoter Sequence Determines the Relationship between Expression Level and Noise. PLoS Biol. *11*.

Cedar, H. (1988). DNA methylation and gene activity. Cell *53*, 3–4.

Cerbai, F., Lana, D., Nosi, D., Petkova-Kirova, P., Zecchi, S., Brothers, H.M., Wenk, G.L., and Giovannini, M.G. (2012). The Neuron-Astrocyte-Microglia Triad in Normal Brain Ageing and in a Model of Neuroinflammation in the Rat Hippocampus. PLOS ONE *7*, e45250.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nat. Lond. *450*, 1230–1234.

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature *453*, 544–547.

Chen, E.-H., Hou, Q.-L., Wei, D.-D., Jiang, H.-B., and Wang, J.-J. (2017a). Phenotypic plasticity, trade-offs and gene expression changes accompanying dietary restriction and switches in Bactrocera dorsalis (Hendel) (Diptera: Tephritidae). Sci. Rep. *7*.

Chen, W.-H., Lu, G., Chen, X., Zhao, X.-M., and Bork, P. (2017b). OGEE v2: an update of the online gene essentiality database with special focus on differentially essential genes in human cancer cell lines. Nucleic Acids Res. *45*, D940–D944.

CHEN, W.-W., ZHANG, X., and HUANG, W.-J. (2016). Role of neuroinflammation in neurodegenerative diseases (Review). Mol. Med. Rep. *13*, 3391–3396.

Choi, J.K., and Kim, Y.-J. (2009). Intrinsic variability of gene expression encoded in nucleosome positioning sequences. Nat. Genet. *41*, 498–503.

Choi, J.K., Yu, U., Yoo, O.J., and Kim, S. (2005). Differential coexpression analysis using microarray data and its application to human cancer. Bioinformatics *21*, 4348–4355.

Clodfelter, K.H., Holloway, M.G., Hodor, P., Park, S.-H., Ray, W.J., and Waxman, D.J. (2006). Sex-Dependent Liver Gene Expression Is Extensive and Largely Dependent upon Signal Transducer and Activator of Transcription 5b (STAT5b): STAT5b-Dependent Activation of Male Genes and Repression of Female Genes Revealed by Microarray Analysis. Mol. Endocrinol. *20*, 1333–1351.

Columb, M.O., and Atkinson, M.S. (2016). Statistical analysis: sample size and power estimations. BJA Educ. *16*, 159–161.

Consortium, Gte., Analysts, L., Aguet, F., Brown, A.A., Castel, S.E., Davis, J.R., He, Y., Jo, B., Mohammadi, P., Park, Y., et al. (2017). Genetic effects on gene expression across human tissues. Nature *550*, nature24277.

Cooke, J. (1995). Beyond natural selection? -- Epigenetic Inheritance and Evolution: The Lamarckian Dimension by Eva Jablonka and Marion J. Lamb. Nat. Lond. *375*, 744.

Cooney, C.A., Dave, A.A., and Wolff, G.L. (2002). Maternal Methyl Supplements in Mice Affect Epigenetic Variation and DNA Methylation of Offspring. J. Nutr. *132*, 2393S-2400S.

Cosgrove, K.P., Mazure, C.M., and Staley, J.K. (2007). Evolving Knowledge of Sex Differences in Brain Structure, Function, and Chemistry. Biol. Psychiatry *62*, 847–855.

Croft, W., Dobson, K.L., and Bellamy, T.C. (2015). Plasticity of Neuron-Glial Transmission: Equipping Glia for Long-Term Integration of Network Activity. Neural Plast. *2015*.

Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.

D'Ambrosio, R., Maris, D.O., Grady, M.S., Winn, H.R., and Janigro, D. (1999). Impaired K+ Homeostasis and Altered Electrophysiological Properties of Post-Traumatic Hippocampal Glia. J. Neurosci. Off. J. Soc. Neurosci. *19*, 8152–8162.

Darr, H., Mayshar, Y., and Benvenisty, N. (2006). Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. Development *133*, 1193–1201.

D'Elia, M.A., Pereira, M.P., and Brown, E.D. (2009). Are essential genes really essential? Trends Microbiol. *17*, 433–438.

Deutschbauer, A.M., Jaramillo, D.F., Proctor, M., Kumm, J., Hillenmeyer, M.E., Davis, R.W., Nislow, C., and Giaever, G. (2005). Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in Yeast. Genetics *169*, 1915–1925.

Dewing, P., Shi, T., Horvath, S., and Vilain, E. (2003). Sexually dimorphic gene expression in mouse brain precedes gonadal differentiation. Mol. Brain Res. *118*, 82–90.

Di Benedetto, S., Müller, L., Wenger, E., Düzel, S., and Pawelec, G. (2017). Contribution of neuroinflammation and immunity to brain aging and the mitigating effects of physical and cognitive interventions. Neurosci. Biobehav. Rev. *75*, 114–128.

Dong, D., Shao, X., Deng, N., and Zhang, Z. (2011). Gene expression variations are predictive for stochastic noise. Nucleic Acids Res. *39*, 403–413.

Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W.A., Hou, L., and Lin, S.M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics *11*, 587.

Dunning, M.J., Smith, M.L., Ritchie, M.E., and Tavaré, S. (2007). beadarray: R classes and methods for Illumina bead-based data. Bioinformatics *23*, 2183–2184.

Ecker, S., Pancaldi, V., Rico, D., and Valencia, A. (2015). Higher gene expression variability in the more aggressive subtype of chronic lymphocytic leukemia. Genome Med. *7*, 8.

Eichler, E.E., Flint, J., Gibson, G., Kong, A., Leal, S.M., Moore, J.H., and Nadeau, J.H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. Nat. Rev. Genet. *11*, 446–450.

Eisenberg, E., and Levanon, E.Y. (2003). Human housekeeping genes are compact. Trends Genet. *19*, 362–365.

Elsamanoudy, A.Z., Neamat-Allah, M.A.M., Mohammad, F.A.H., Hassanien, M., and Nada, H.A. (2016). The role of nutrition related genes and nutrigenetics in understanding the pathogenesis of cancer. J. Microsc. Ultrastruct. *4*, 115–122.

Fagiolo, U., Cossarizza, A., Santacaterina, S., Ortolani, C., Monti, D., Paganelli, R., and Franceschi, C. (1992). Increased Cytokine Production by Peripheral Blood Mononuclear Cells from Healthy Elderly Peoplea. Ann. N. Y. Acad. Sci. *663*, 490–493.

Fagiolo, U., Amadori, A., Cozzi, E., Bendo, R., Lama, M., Douglas, A., and Palù, G. (1993). Humoral and cellular immune response to influenza virus vaccination in aged humans. Aging Clin. Exp. Res. *5*, 451–458.

Feinberg, A.P., and Irizarry, R.A. (2010). Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proc. Natl. Acad. Sci. U. S. A. *107*, 1757–1764.

Fowler, I.L., Carr, V.J., Carter, N.T., and Lewin, T.J. (1998). Patterns of current and lifetime substance use in schizophrenia. Schizophr. Bull. Oxf. *24*, 443–455.

Galvan, V., and Jin, K. (2007). Neurogenesis in the aging brain. Clin. Interv. Aging *2*, 605–610.

Garcia-Bernardo, J., and Dunlop, M.J. (2016). Phenotypic Diversity Using Bimodal and Unimodal Expression of Stress Response Proteins. Biophys. J. *110*, 2278–2287.

Georgi, B., Voight, B.F., and Bućan, M. (2013). From Mouse to Human: Evolutionary Genomics Analysis of Human Orthologs of Essential Genes. PLoS Genet. *9*.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. Nature *418*, 387–391.

Gil, A., Aguilera, C.M., Gil-Campos, M., and Cañete, R. (2007). Altered signalling and gene expression associated with the immune system and the inflammatory response in obesity. Br. J. Nutr. *98*, S121–S126.

Glass, D., Viñuela, A., Davies, M.N., Ramasamy, A., Parts, L., Knowles, D., Brown, A.A., Hedman, Å.K., Small, K.S., Buil, A., et al. (2013). Gene expression changes with age in skin, adipose tissue, blood and brain. Genome Biol. *14*, R75.

Gundersen, V., Storm-Mathisen, J., and Bergersen, L.H. (2015). Neuroglial Transmission. Physiol. Rev. *95*, 695–726.

Hanahan, D., and Weinberg, R.A. (2000). The Hallmarks of Cancer. Cell *100*, 57–70.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. Cell *144*, 646–674.

Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D., et al. (2011). Increased methylation variation in epigenetic domains across cancer types. Nat. Genet. *43*, 768–775.

Hanson, N.D., Owens, M.J., and Nemeroff, C.B. (2011). Depression, Antidepressants, and Neurogenesis: A Critical Reappraisal. Neuropsychopharmacology *36*, 2589–2602.

Harada, C.N., Natelson Love, M.C., and Triebel, K. (2013). Normal Cognitive Aging. Clin. Geriatr. Med. *29*, 737–752.

Hasegawa, Y., Taylor, D., Ovchinnikov, D.A., Wolvetang, E.J., de Torrenté, L., and Mar, J.C. (2015). Variability of Gene Expression Identifies Transcriptional Regulators of Early Human Embryonic Development. PLoS Genet. *11*.

Hebert, L.E., Beckett, L.A., Scherr, P.A., and Evans, D.A. (2001). Annual Incidence of Alzheimer Disease in the United States Projected to the Years 2000 Through 2050. Alzheimer Dis. Assoc. Disord. *15*, 169–173.

Heneka, M.T., Carson, M.J., El Khoury, J., Landreth, G.E., Brosseron, F., Feinstein, D.L., Jacobs, A.H., Wyss-Coray, T., Vitorica, J., Ransohoff, R.M., et al. (2015). Neuroinflammation in Alzheimer's Disease. Lancet Neurol. *14*, 388–405.

Hernandez, D.G., Nalls, M.A., Moore, M., Chong, S., Dillman, A., Trabzuni, D., Gibbs, J.R., Ryten, M., Arepalli, S., Weale, M.E., et al. (2012). Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. Neurobiol. Dis. *47*, 20–28.

Ho, J.W.K., Stefani, M., dos Remedios, C.G., and Charleston, M.A. (2008). Differential variability analysis of gene expression and its application to human diseases. Bioinformatics *24*, i390–i398.

Hsieh, H.-L., and Yang, C.-M. (2013). Role of Redox Signaling in Neuroinflammation and Neurodegenerative Diseases. BioMed Res. Int. *2013*.

Irvine, R.A., Lin, I.G., and Hsieh, C.-L. (2002). DNA Methylation Has a Local Effect on Transcription and Histone Acetylation. Mol. Cell. Biol. *22*, 6689–6696.

Itoh, Y., and Arnold, A.P. (2015). Are females more variable than males in gene expression? Meta-analysis of microarray datasets. Biol. Sex Differ. *6*.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. *33*, 245–254.

Jazin, E., and Cahill, L. (2010). Sex differences in molecular neuroscience: from fruit flies to humans. Nat. Rev. Neurosci. *11*, 9–17.

Jirtle, R.L., and Skinner, M.K. (2007). Environmental epigenomics and disease susceptibility. Nat. Rev. Genet. *8*, 253–262.

Johansson, M.M., Van Geystelen, A., Larmuseau, M.H.D., Djurovic, S., Andreassen, O.A., Agartz, I., and Jazin, E. (2015). Microarray Analysis of Copy Number Variants on the Human Y Chromosome Reveals Novel and Frequent Duplications Overrepresented in Specific Haplogroups. PLoS ONE *10*.

Johnston, I.G., Gaal, B., Neves, R.P. das, Enver, T., Iborra, F.J., and Jones, N.S. (2012). Mitochondrial Variability as a Source of Extrinsic Cellular Noise. PLoS Comput. Biol. *8*.

Kærn, M., Elston, T.C., Blake, W.J., and Collins, J.J. (2005). Stochasticity in gene expression: from theories to phenotypes. Nat. Rev. Genet. *6*, nrg1615.

Kalmar, T., Lim, C., Hayward, P., Muñoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Arias, A.M. (2009). Regulated Fluctuations in Nanog Expression Mediate Cell Fate Decisions in Embryonic Stem Cells. PLOS Biol. *7*, e1000149.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Bot, N.L., Moreno, S., Sohrmann, M., et al. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature *421*, 231–237.

Kaminsky, Z.A., Tang, T., Wang, S.-C., Ptak, C., Oh, G.H.T., Wong, A.H.C., Feldcamp, L.A., Virtanen, C., Halfvarson, J., Tysk, C., et al. (2009). DNA methylation profiles in monozygotic and dizygotic twins. Nat. Genet. *41*, 240–245.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493–D496.

Kemkemer, R., Schrank, S., Vogel, W., Gruler, H., and Kaufmann, D. (2002). Increased noise as an effect of haploinsufficiency of the tumor-suppressor gene neurofibromatosis type 1 in vitro. Proc. Natl. Acad. Sci. U. S. A. *99*, 13783–13788.

Kimelberg, H.K. (2005). Astrocytic swelling in cerebral ischemia as a possible cause of injury and target for therapy. Glia *50*, 389–397.

Kimelberg, H.K., and Nedergaard, M. (2010). Functions of astrocytes and their potential as therapeutic targets. Neurotherapeutics *7*, 338–353.

Kobayashi, S., Isotani, A., Mise, N., Yamamoto, M., Fujihara, Y., Kaseda, K., Nakanishi, T., Ikawa, M., Hamada, H., Abe, K., et al. (2006). Comparison of Gene Expression in Male and Female Mouse Blastocysts Revealed Imprinting of the X-Linked Gene, Rhox5/Pem, at Preimplantation Stages. Curr. Biol. *16*, 166–172.

Kostka, D., and Spang, R. (2004). Finding disease specific alterations in the co-expression of genes. Bioinformatics *20*, i194–i199.

Lahn, B.T., and Page, D.C. (1997). Functional Coherence of the Human Y Chromosome. Science *278*, 675–680.

Landau, D.A., Clement, K., Ziller, M.J., Boyle, P., Fan, J., Gu, H., Stevenson, K., Sougnez, C., Wang, L., Li, S., et al. (2014). Locally Disordered Methylation Forms the Basis of Intratumor Methylome Variation in Chronic Lymphocytic Leukemia. Cancer Cell *26*, 813–825.

Lauber, A.H., Romano, G.J., and Pfaff, D.W. (1991). Gene expression for estrogen and progesterone receptor mRNAS in rat brain and possible relations to sexually dimorphic functions. J. Steroid Biochem. Mol. Biol. *40*, 53–62.

Lehner, B. (2008). Selection to minimise noise in living systems and its implications for the evolution of gene expression. Mol. Syst. Biol. *4*, 170.

Lehner, B. (2010). Conflict between Noise and Plasticity in Yeast. PLoS Genet. *6*.

Lehner, B., and Kaneko, K. (2011). Fluctuation and response in biology. Cell. Mol. Life Sci. *68*, 1005–1010.

Lepeta, K., Lourenco, M.V., Schweitzer, B.C., Martino Adami, P.V., Banerjee, P., Catuara-Solarz, S., de La Fuente Revenga, M., Guillem, A.M., Haidar, M., Ijomone, O.M., et al. (2016). Synaptopathies: synaptic dysfunction in neurological disorders – A review from students to students. J. Neurochem. *138*, 785–805.

Levy, G., Schupf, N., Tang, M.-X., Cote, L.J., Louis, E.D., Mejia, H., Stern, Y., and Marder, K. (2002). Combined effect of age and severity on the risk of dementia in Parkinson's disease. Ann. Neurol. *51*, 722–729.

Li, J., Liu, Y., Kim, T., Min, R., and Zhang, Z. (2010a). Gene Expression Variability within and between Human Populations and Implications toward Disease Susceptibility. PLOS Comput. Biol. *6*, e1000910.

Li, J., Min, R., Vizeacoumar, F.J., Jin, K., Xin, X., and Zhang, Z. (2010b). Exploiting the determinants of stochastic gene expression in Saccharomyces cerevisiae for genome-wide prediction of expression noise. Proc. Natl. Acad. Sci. U. S. A. *107*, 10472–10477.

Liggett, W. (2006). Normalization and Technical Variation in Gene Expression Measurements. J. Res. Natl. Inst. Stand. Technol. *111*, 361–372.

Lin, W., Liu, W., and Hwang, M. (2009). Topological and organizational properties of the products of house-keeping and tissue-specific genes in protein-protein interaction networks. BMC Syst. Biol. *3*, 32.

Liu, X., and Rattray, M. (2010). Including Probe-Level Measurement Error in Robust Mixture Clustering of Replicated Microarray Gene Expression. Stat. Appl. Genet. Mol. Biol. *9*.

Logue, O.C., Cramer, N.P., Xu, X., Perl, D.P., and Galdzicki, Z. (2016). Alterations of functional properties of hippocampal networks following repetitive closed-head injury. Exp. Neurol. *277*, 227–243.

Lucin, K.M., and Wyss-Coray, T. (2009). Immune activation in brain aging and neurodegeneration: too much or too little? Neuron *64*, 110–122.

Maamar, H., Raj, A., and Dubnau, D. (2007). Noise in Gene Expression Determines Cell Fate in Bacillus subtilis. Science *317*, 526–529.

Malberg, J.E., Eisch, A.J., Nestler, E.J., and Duman, R.S. (2000). Chronic Antidepressant Treatment Increases Neurogenesis in Adult Rat Hippocampus. J. Neurosci. *20*, 9104–9110.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.

Mar, J.C., Matigian, N.A., Mackay-Sim, A., Mellick, G.D., Sue, C.M., Silburn, P.A., McGrath, J.J., Quackenbush, J., and Wells, C.A. (2011). Variance of Gene Expression Identifies Altered Network Constraints in Neurological Disease. PLOS Genet. *7*, e1002207.

McCall, M.N., Illei, P.B., and Halushka, M.K. (2016). Complex Sources of Variation in Tissue Expression Data: Analysis of the GTEx Lung Transcriptome. Am. J. Hum. Genet. *99*, 624–635.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. Cell *113*, 631–642.

Montecino-Rodriguez, E., Berent-Maoz, B., and Dorshkind, K. (2013). Causes, consequences, and reversal of immune system aging. J. Clin. Invest. *123*, 958–965.

Moore, L.D., Le, T., and Fan, G. (2013). DNA Methylation and Its Basic Function. Neuropsychopharmacology *38*, 23–38.

Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840–846.

Novick, A., and Weiner, M. (1957). ENZYME INDUCTION AS AN ALL-OR-NONE PHENOMENON*. Proc. Natl. Acad. Sci. U. S. A. *43*, 553–566.

Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., and Oudenaarden, A. van (2002). Regulation of noise in the expression of a single gene. Nat. Genet. *31*, 69–73.

Petrik, D., Lagace, D.C., and Eisch, A.J. (2012). The neurogenesis hypothesis of affective and anxiety disorders: are we mistaking the scaffolding for the building? Neuropharmacology *62*, 21–34.

Pietrobon, D., and Moskowitz, M.A. (2014). Chaos and commotion in the wake of cortical spreading depression and spreading depolarizations. Nat. Rev. Neurosci. *15*, 379–393.

Piras, V., Tomita, M., and Selvarajoo, K. (2014). Transcriptome-wide Variability in Single Embryonic Development Cells. Sci. Rep. *4*, 7137.

Pritchard, C.C., Hsu, L., Delrow, J., and Nelson, P.S. (2001). Project normal: Defining normal variance in mouse gene expression. Proc. Natl. Acad. Sci. U. S. A. *98*, 13266–13271.

Pruitt, K.D., Harrow, J., Harte, R.A., Wallin, C., Diekhans, M., Maglott, D.R., Searle, S., Farrell, C.M., Loveland, J.E., Ruef, B.J., et al. (2009). The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. Genome Res. *19*, 1316–1323.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. PLoS Biol. *4*.

Rakyan, V.K., Chong, S., Champ, M.E., Cuthbert, P.C., Morgan, H.D., Luu, K.V.K., and Whitelaw, E. (2003). Transgenerational inheritance of epigenetic states at the murine AxinFu allele occurs after maternal and paternal transmission. Proc. Natl. Acad. Sci. U. S. A. *100*, 2538–2543.

Rancati, G., Moffat, J., Typas, A., and Pavelka, N. (2018). Emerging and evolving concepts in gene essentiality. Nat. Rev. Genet. *19*, 34–49.

Raser, J.M., and O'Shea, E.K. (2004). Control of Stochasticity in Eukaryotic Gene Expression. Science *304*, 1811–1814.

Raser, J.M., and O'Shea, E.K. (2005). Noise in Gene Expression: Origins, Consequences, and Control. Science *309*, 2010–2013.

Richards, S.M., Jensen, R.V., Liu, M., Sullivan, B.D., Lombardi, M.J., Rowley, P., Schirra, F., Treister, N.S., Suzuki, T., Steagall, R.J., et al. (2006). Influence of sex on gene expression in the mouse lacrimal gland. Exp. Eye Res. *82*, 13–23.

Riggs, A.D., and Xiong, Z. (2004). Methylation and epigenetic fidelity. Proc. Natl. Acad. Sci. U. S. A. *101*, 4–5.

Rinn, J.L., Rozowsky, J.S., Laurenzi, I.J., Petersen, P.H., Zou, K., Zhong, W., Gerstein, M., and Snyder, M. (2004). Major Molecular Differences between Mammalian Sexes Are Involved in Drug Metabolism and Renal Function. Dev. Cell *6*, 791–800.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Roberfroid, S., Vanderleyden, J., and Steenackers, H. (2016). Gene expression variability in clonal populations: Causes and consequences. Crit. Rev. Microbiol. *42*, 969–984.

Sajja, V.S.S.S., Hlavac, N., and VandeVord, P.J. (2016). Role of Glia in Memory Deficits Following Traumatic Brain Injury: Biomarkers of Glia Dysfunction. Front. Integr. Neurosci. *10*.

Scrucca, L., Fop, M., Murphy, T.B., and Raftery, A.E. (2016). mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. R J. *8*, 289–317.

Silander, O.K., Nikolic, N., Zaslaver, A., Bren, A., Kikoin, I., Alon, U., and Ackermann, M. (2012). A Genome-Wide Analysis of Promoter-Mediated Phenotypic Noise in Escherichia coli. PLoS Genet. *8*.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog Is the Gateway to the Pluripotent Ground State. Cell *138*, 722–737.

Singh, G.P. (2013). Coupling Between Noise and Plasticity in E. coli. G3 Genes Genomes Genet. *3*, 2115–2120.

Singh, P., Goode, T., Dean, A., Awad, S.S., and Darlington, G.J. (2011). Elevated Interferon Gamma Signaling Contributes to Impaired Regeneration in the Aged Liver. J. Gerontol. A. Biol. Sci. Med. Sci. *66A*, 944–956.

Singhal, G., Jaehne, E.J., Corrigan, F., and Baune, B.T. (2014). Cellular and molecular mechanisms of immunomodulation in the brain through environmental enrichment. Front. Cell. Neurosci. *8*.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature *423*, 825–837.

Sonawane, A.R., Platig, J., Fagny, M., Chen, C.-Y., Paulson, J.N., Lopes-Ramos, C.M., DeMeo, D.L., Quackenbush, J., Glass, K., and Kuijjer, M.L. (2017). Understanding Tissue-Specific Gene Regulation. Cell Rep. *21*, 1077–1088.

Sopko, R., Huang, D., Preston, N., Chua, G., Papp, B., Kafadar, K., Snyder, M., Oliver, S.G., Cyert, M., Hughes, T.R., et al. (2006). Mapping Pathways and Phenotypes by Systematic Gene Overexpression. Mol. Cell *21*, 319–330.

Soreq, L., Rose, J., Soreq, E., Hardy, J., Trabzuni, D., Cookson, M.R., Smith, C., Ryten, M., Patani, R., and Ule, J. (2017). Major Shifts in Glial Regional Identity Are a Transcriptional Hallmark of Human Brain Aging. Cell Rep. *18*, 557–570.

Streit, W.J., and Xue, Q.-S. (2010). The Brain's Aging Immune System. Aging Dis. *1*, 254–261.

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLOS ONE *6*, e21800.

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics *22*, 1540–1542.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Science *329*, 533–538.

Trabzuni, D., Ramasamy, A., Imran, S., Walker, R., Smith, C., Weale, M.E., Hardy, J., Ryten, M., and Consortium, N.A.B.E. (2013). Widespread sex differences in gene expression and splicing in the adult human brain. Nat. Commun. *4*, ncomms3771.

Ushijima, T., Watanabe, N., Okochi, E., Kaneda, A., Sugimura, T., and Miyamoto, K. (2003). Fidelity of the Methylation Pattern and Its Variation in the Genome. Genome Res. *13*, 868–874.

Vassena, R., Boué, S., González-Roca, E., Aran, B., Auer, H., Veiga, A., and Belmonte, J.C.I. (2011). Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. Development *138*, 3699–3709.

Vawter, M.P., Evans, S., Choudary, P., Tomita, H., Meador-Woodruff, J., Molnar, M., Li, J., Lopez, J.F., Myers, R., Cox, D., et al. (2004). Gender-Specific Gene Expression in Post-Mortem Human Brain: Localization to Sex Chromosomes. Neuropsychopharmacology *29*, 373–384.

Wagner, J.R., Busche, S., Ge, B., Kwan, T., Pastinen, T., and Blanchette, M. (2014). The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol. *15*, R37.

Walker, K.R., and Tesco, G. (2013). Molecular mechanisms of cognitive dysfunction following traumatic brain injury. Front. Aging Neurosci. *5*.

Wang, X., Gao, X., Michalski, S., Zhao, S., and Chen, J. (2016). Traumatic Brain Injury Severity Affects Neurogenesis in Adult Mouse Hippocampus. J. Neurotrauma *33*, 721–733.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., et al. (2015). gplots: Various R Programming Tools for Plotting Data.

Waterland, R.A., and Jirtle, R.L. (2003). Transposable Elements: Targets for Early Nutritional Effects on Epigenetic Gene Regulation. Mol. Cell. Biol. *23*, 5293–5300.

Watson, M. (2006). CoXpress: differential co-expression in gene expression data. BMC Bioinformatics *7*, 509.

Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P., and Schaffer, D.V. (2005). Stochastic Gene Expression in a Lentiviral Positive-Feedback Loop: HIV-1 Tat Fluctuations Drive Phenotypic Diversity. Cell *122*, 169–182.

Weinhold, L., Wahl, S., Pechlivanis, S., Hoffmann, P., and Schmid, M. (2016). A statistical model for the analysis of beta values in DNA methylation studies. BMC Bioinformatics *17*.

Wen, Z., Li, D., Shen, M., and Chen, G. (2017). Therapeutic Potentials of Synapses after Traumatic Brain Injury: A Comprehensive Review. Neural Plast. *2017*.

Wu, D., and Meydani, S.N. (2008). Age-associated changes in immune and inflammatory responses: impact of vitamin E intervention. J. Leukoc. Biol. *84*, 900–914.

Xu, J., Watkins, R., and Arnold, A.P. (2006). Sexually dimorphic expression of the X-linked gene Eif2s3x mRNA but not protein in mouse brain. Gene Expr. Patterns *6*, 146–155.

Yang, X., Schadt, E.E., Wang, S., Wang, H., Arnold, A.P., Ingram-Drake, L., Drake, T.A., and Lusis, A.J. (2006). Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res. *16*, 995–1004.

Yilmaz, A., Peretz, M., Aharony, A., Sagi, I., and Benvenisty, N. (2018). Defining essential genes for human pluripotent stem cells by CRISPR–Cas9 screening in haploid cells. Nat. Cell Biol. *20*, 610.

Yu, K., Ganesan, K., Tan, L.K., Laban, M., Wu, J., Zhao, X.D., Li, H., Leung, C.H.W., Zhu, Y., Wei, C.L., et al. (2008). A Precisely Regulated Gene Expression Cassette Potently Modulates Metastasis and Survival in Multiple Solid Cancers. PLoS Genet. *4*.

Zahn, J.M., Poosala, S., Owen, A.B., Ingram, D.K., Lustig, A., Carter, A., Weeraratna, A.T., Taub, D.D., Gorospe, M., Mazan-Mamczarz, K., et al. (2007). AGEMAP: A Gene Expression Database for Aging in Mice. PLOS Genet. *3*, e201.

Zammit, S., Allebeck, P., Dalman, C., Lundberg, I., Hemmingsson, T., and Lewis, G. (2003). Investigating the Association Between Cigarette Smoking and Schizophrenia in a Cohort Study. Am. J. Psychiatry *160*, 2216–2221.

Zeller, T., Wild, P., Szymczak, S., Rotival, M., Schillert, A., Castagne, R., Maouche, S., Germain, M., Lackner, K., Rossmann, H., et al. (2010). Genetics and Beyond – The Transcriptome of Human Monocytes and Disease Susceptibility. PLoS ONE *5*.

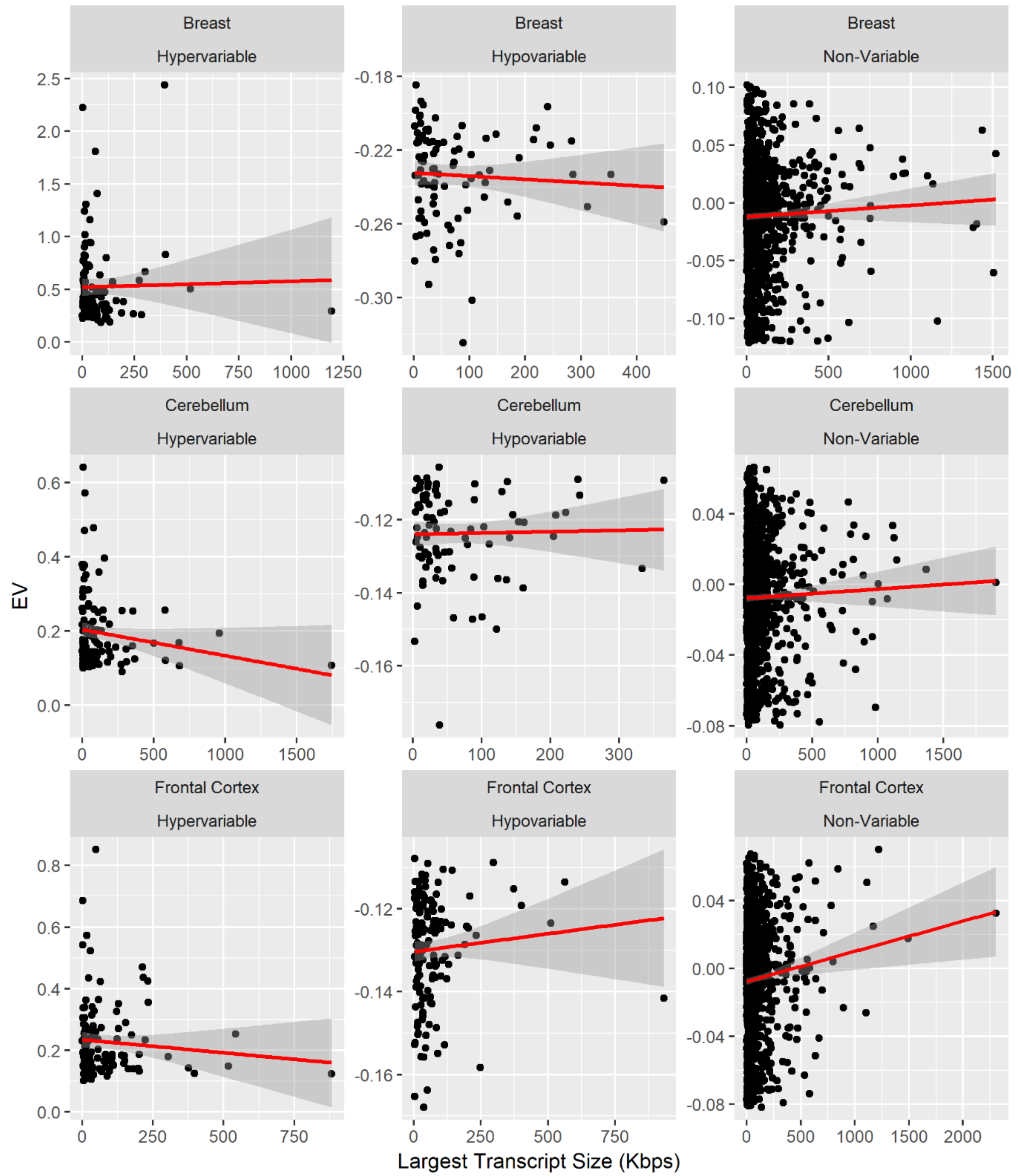
Zhang, B., Zhou, Y., Lin, N., Lowdon, R.F., Hong, C., Nagarajan, R.P., Cheng, J.B., Li, D., Stevens, M., Lee, H.J., et al. (2013). Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. Genome Res. *23*, 1522–1540.

Zhang, Y.P., Cai, J., Shields, L.B.E., Liu, N., Xu, X.-M., and Shields, C.B. (2014). Traumatic Brain Injury Using Mouse Models. Transl. Stroke Res. *5*, 454–471.

Zhu, J., He, F., Hu, S., and Yu, J. (2008). On the nature of human housekeeping genes. Trends Genet. *24*, 481–484.

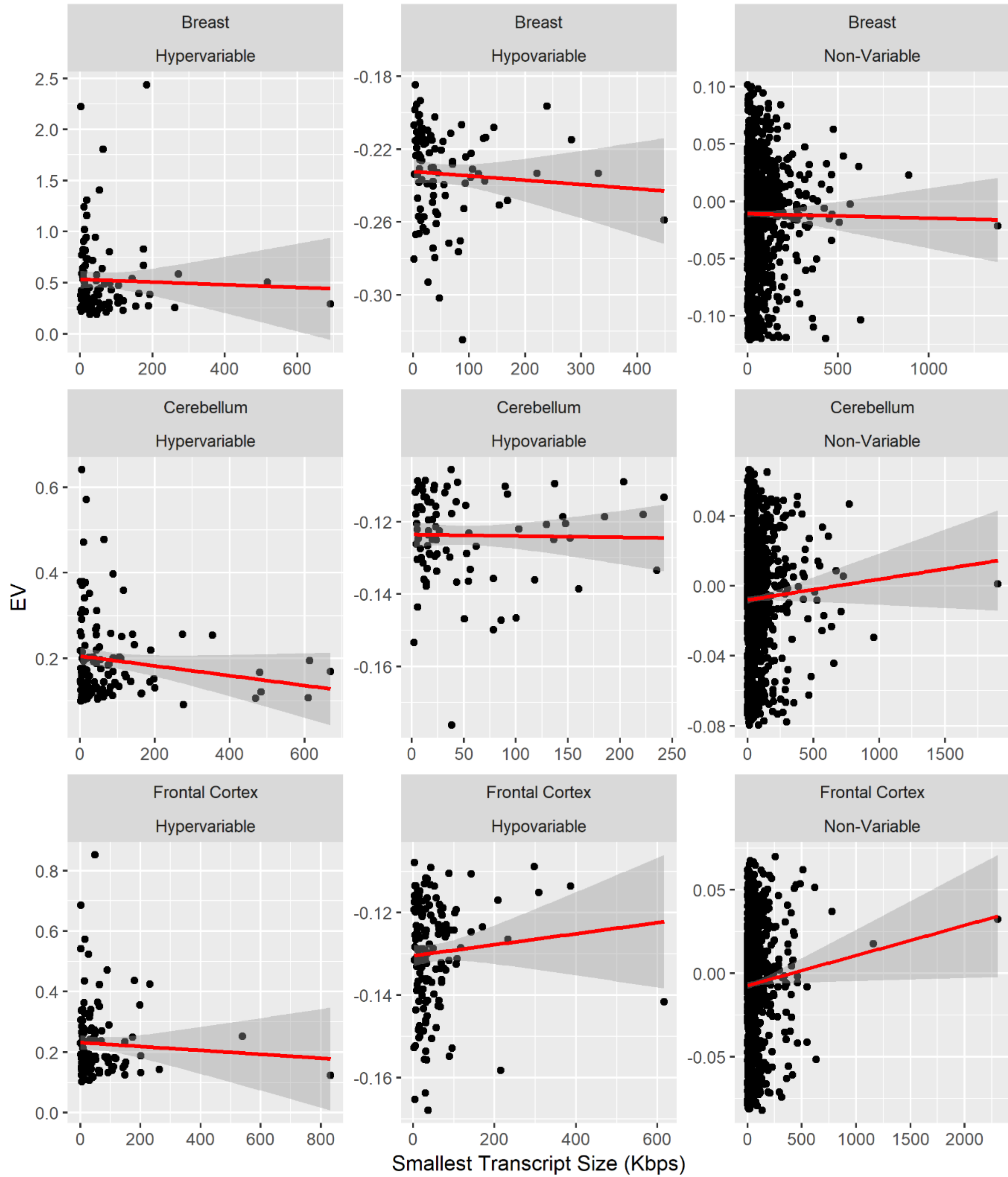
Appendices

Supplementary Figure 1. Structural analysis of genes as a function of EV



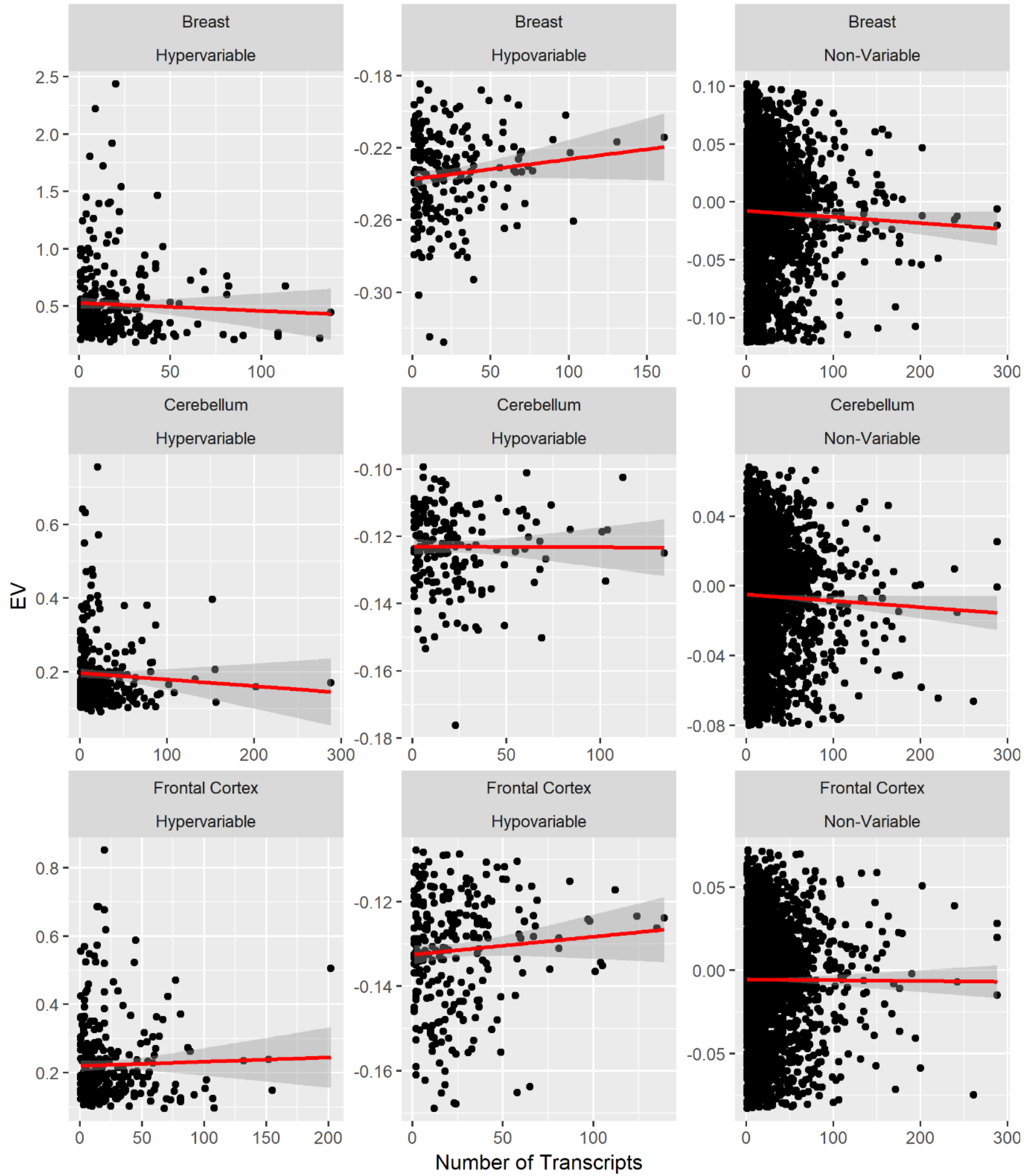
Largest Transcript Size Linear Regression Analysis

	Estimate	Std. Error	t value	Pr(>|t|)
Breast Intercept	7.94e+01	1.52e+01	5.22e+00	2.08e-07
Breast Hypovariable	-8.24e+00	2.21e+01	-3.73e-01	7.09e-01
Breast Hypervariable	1.36e+01	1.58e+01	8.60e-01	3.90e-01
Cerebellum Intercept	1.15e+02	1.43e+01	8.04e+00	1.77e-15
Cerebellum Hypovariable	-5.07e+01	2.21e+01	-2.30e+00	2.17e-02
Cerebellum Hypervariable	-2.43e+01	1.49e+01	-1.63e+00	1.03e-01
Frontal Cortex Intercept	8.03e+01	1.31e+01	6.15e+00	9.96e-10
Frontal Cortex Hypovariable	-9.89e+00	1.76e+01	-5.63e-01	5.73e-01
Frontal Cortex Hypervariable	6.38e+00	1.36e+01	4.68e-01	6.40e-01


Breast
Hypervariable
Breast
Hypovariable
Breast
Non-Variable
Cerebellum
Hypervariable
Cerebellum
Hypovariable
Cerebellum
Non-Variable
Frontal Cortex
Hypervariable
Frontal Cortex
Hypovariable
Frontal Cortex
Non-Variable
EV
Smallest Transcript Size (Kbps)

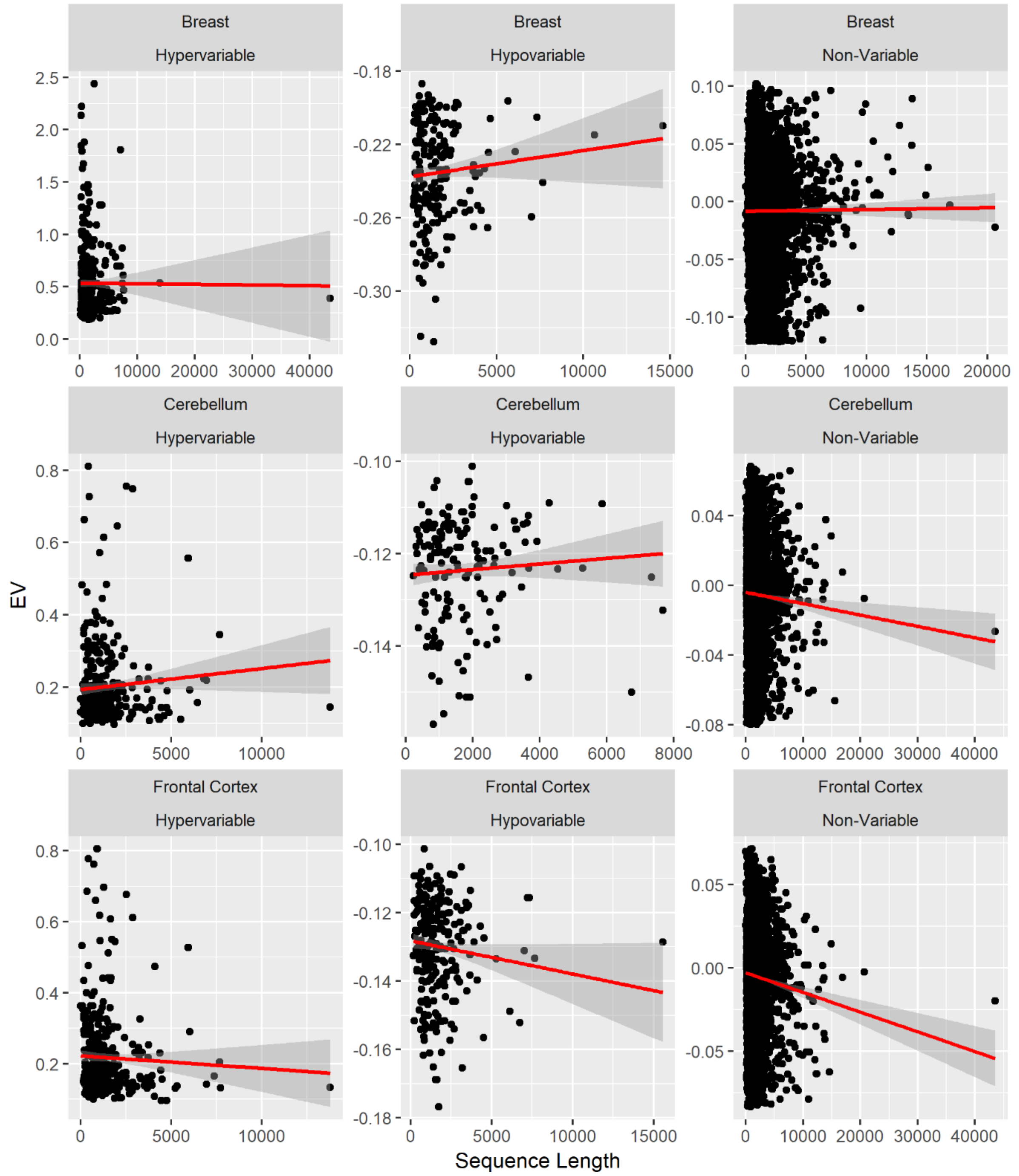
Smallest Transcript Size Linear Regression Analysis

	Estimate	Std. Error	t value	Pr(>|t|)
Breast Intercept	5.98e+01	8.94e+00	6.69e+00	3.14e-11
Breast Hypovariable	-4.67e+00	1.30e+01	-3.60e-01	7.19e-01
Breast Hypervariable	1.11e+00	9.31e+00	1.19e-01	9.05e-01
Cerebellum Intercept	7.85e+01	9.66e+00	8.13e+00	8.25e-16
Cerebellum Hypovariable	-2.93e+01	1.49e+01	-1.96e+00	5.00e-02
Cerebellum Hypervariable	-1.54e+01	1.01e+01	-1.53e+00	1.25e-01
Frontal Cortex Intercept	5.85e+01	9.56e+00	6.12e+00	1.20e-09
Frontal Cortex Hypovariable	-5.52e+00	1.29e+01	-4.29e-01	6.68e-01
Frontal Cortex Hypervariable	2.97e+00	9.99e+00	2.98e-01	7.66e-01


Breast
Hypervariable
Breast
Hypovariable
Breast
Non-Variable
Cerebellum
Hypervariable
Cerebellum
Hypovariable
Cerebellum
Non-Variable
Frontal Cortex
Hypervariable
Frontal Cortex
Hypovariable
Frontal Cortex
Non-Variable
EV
Number of Transcripts

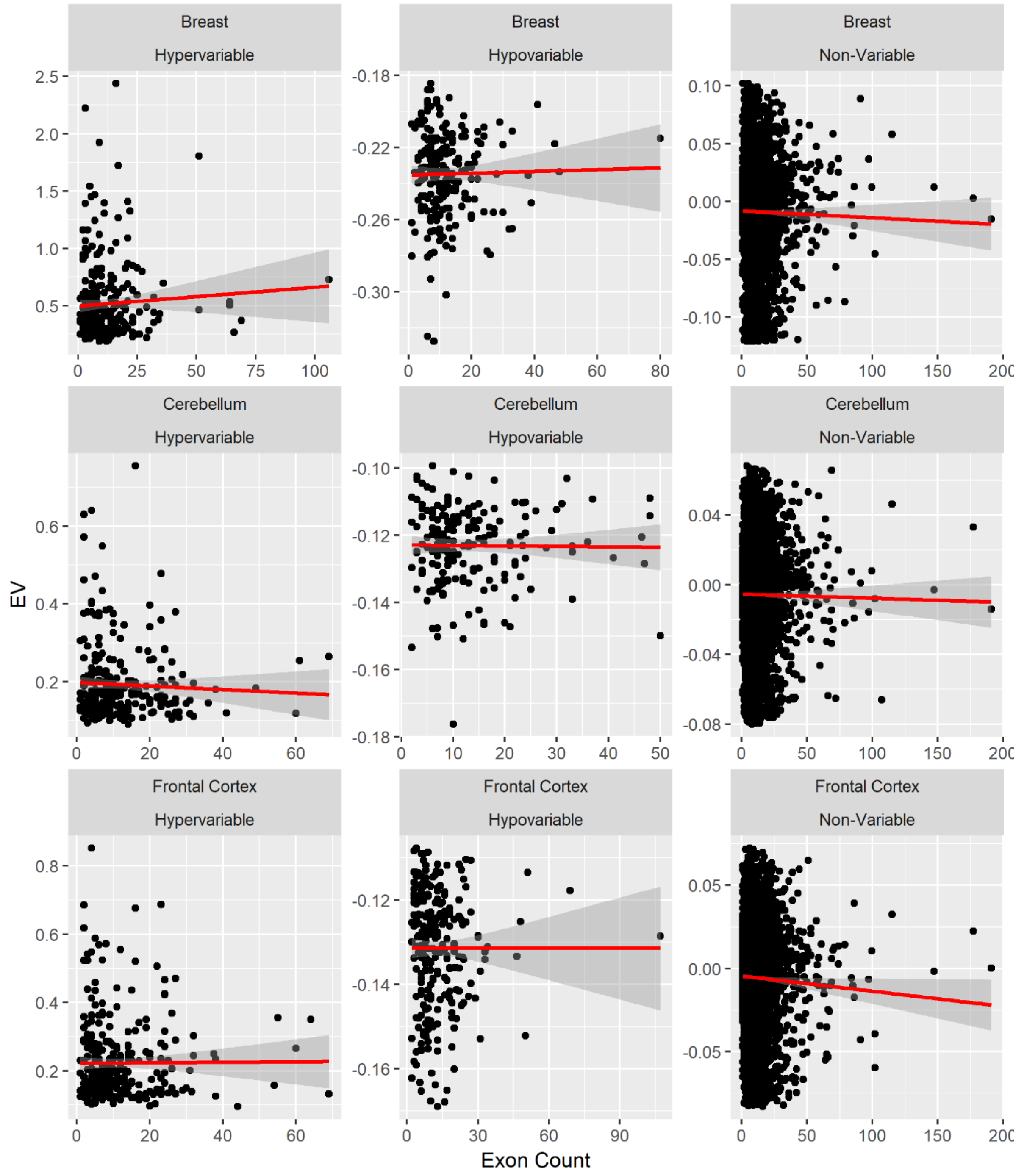
Number of Transcripts Linear Regression Analysis

	Estimate	Std. Error	t value	Pr(>|t|)
Breast Intercept	1.84e+01	1.59e+00	1.15e+01	2.62e-30
Breast Hypovariable	5.28e+00	2.44e+00	2.16e+00	3.07e-02
Breast Hypervariable	3.72e+00	1.65e+00	2.25e+00	2.48e-02
Cerebellum Intercept	2.38e+01	1.55e+00	1.54e+01	6.27e-52
Cerebellum Hypovariable	-9.10e-01	2.49e+00	-3.65e-01	7.15e-01
Cerebellum Hypervariable	-2.55e+00	1.60e+00	-1.59e+00	1.11e-01
Frontal Cortex Intercept	2.24e+01	1.51e+00	1.48e+01	1.43e-48
Frontal Cortex Hypovariable	1.87e+00	2.23e+00	8.40e-01	4.01e-01
Frontal Cortex Hypervariable	-8.51e-01	1.57e+00	-5.42e-01	5.88e-01


Breast
Hypervariable
Breast
Hypovariable
Breast
Non-Variable
Cerebellum
Hypervariable
Cerebellum
Hypovariable
Cerebellum
Non-Variable
Frontal Cortex
Hypervariable
Frontal Cortex
Hypovariable
Frontal Cortex
Non-Variable
EV
Sequence Length

Sequence Length Linear Regression Analysis

	Estimate	Std. Error	t value	Pr(>|t|)
Breast Intercept	1.78e+03	6.94e+01	2.56e+01	7.24e-139
Breast Hypovariable	-8.73e+01	1.18e+02	-7.37e-01	4.61e-01
Breast Hypervariable	-1.24e+02	7.20e+01	-1.72e+00	8.50e-02
Cerebellum Intercept	1.51e+03	7.80e+01	1.93e+01	3.37e-81
Cerebellum Hypovariable	2.26e+02	1.36e+02	1.67e+00	9.52e-02
Cerebellum Hypervariable	1.72e+02	8.03e+01	2.15e+00	3.19e-02
Frontal Cortex Intercept	1.51e+03	7.35e+01	2.06e+01	1.57e-91
Frontal Cortex Hypovariable	1.50e+02	1.13e+02	1.33e+00	1.83e-01
Frontal Cortex Hypervariable	1.56e+02	7.60e+01	2.05e+00	4.05e-02


Breast
Hypervariable
Breast
Hypovariable
Breast
Non-Variable
Cerebellum
Hypervariable
Cerebellum
Hypovariable
Cerebellum
Non-Variable
Frontal Cortex
Hypervariable
Frontal Cortex
Hypovariable
Frontal Cortex
Non-Variable
EV
Exon Count

Exon Count Linear Regression Analysis

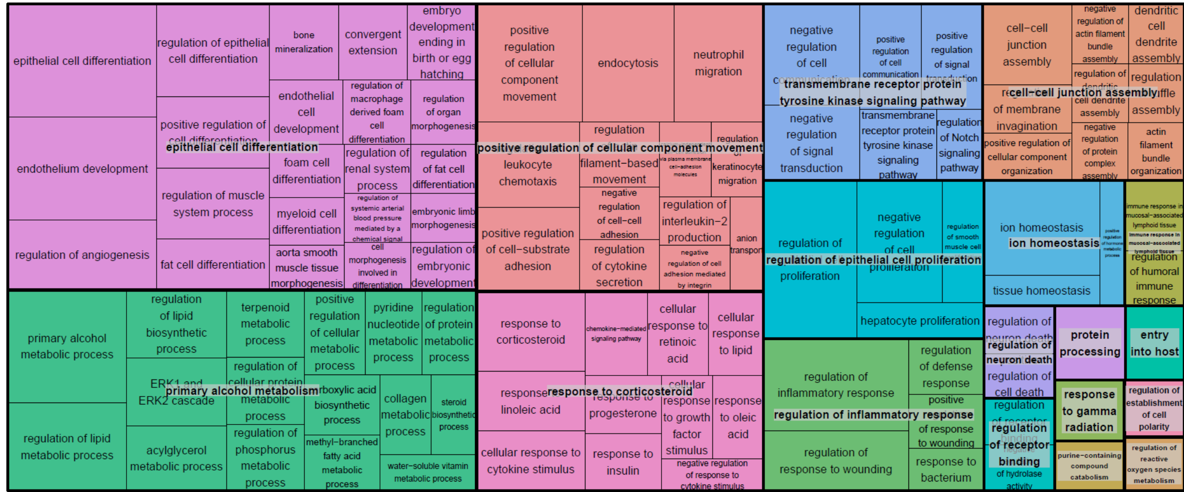
	Estimate	Std. Error	t value	Pr(>|t|)
Breast Intercept	1.16e+01	6.79e-01	1.71e+01	1.34e-63
Breast Hypovariable	7.55e-01	1.04e+00	7.24e-01	4.69e-01
Breast Hypervariable	2.67e-01	7.06e-01	3.78e-01	7.05e-01
Cerebellum Intercept	1.12e+01	6.55e-01	1.71e+01	8.57e-64
Cerebellum Hypovariable	2.09e+00	1.06e+00	1.98e+00	4.81e-02
Cerebellum Hypervariable	7.25e-01	6.80e-01	1.07e+00	2.87e-01
Frontal Cortex Intercept	1.11e+01	6.41e-01	1.73e+01	9.71e-65
Frontal Cortex Hypovariable	1.72e+00	9.46e-01	1.82e+00	6.89e-02
Frontal Cortex Hypervariable	9.22e-01	6.68e-01	1.38e+00	1.67e-01

Supplementary Figure 2. Complete list of GO terms for all genes

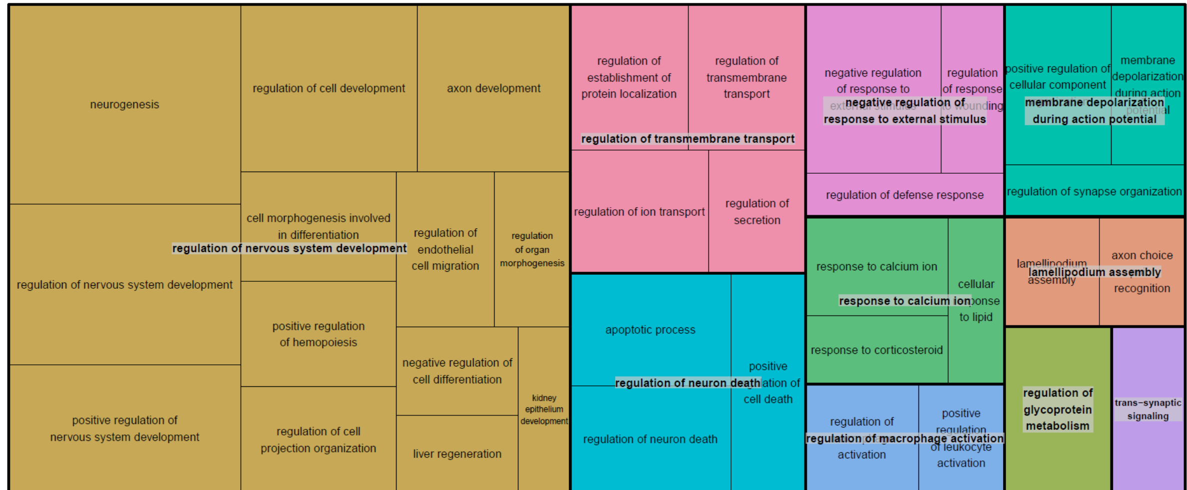
Common Hypervariable Biological Processes

Common Hypovariable Biological Processes

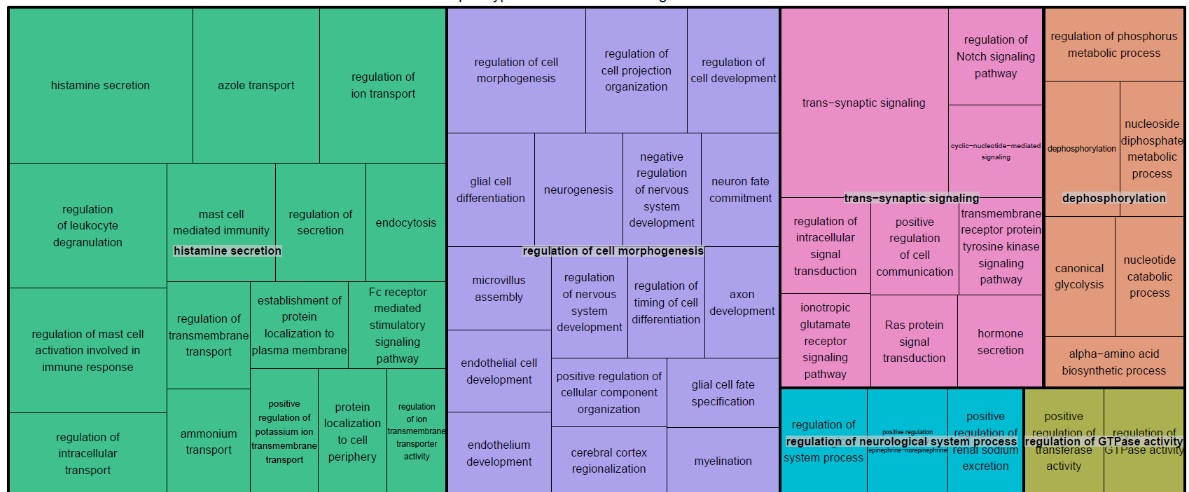
Unique Hypervariable Breast Biological Processes

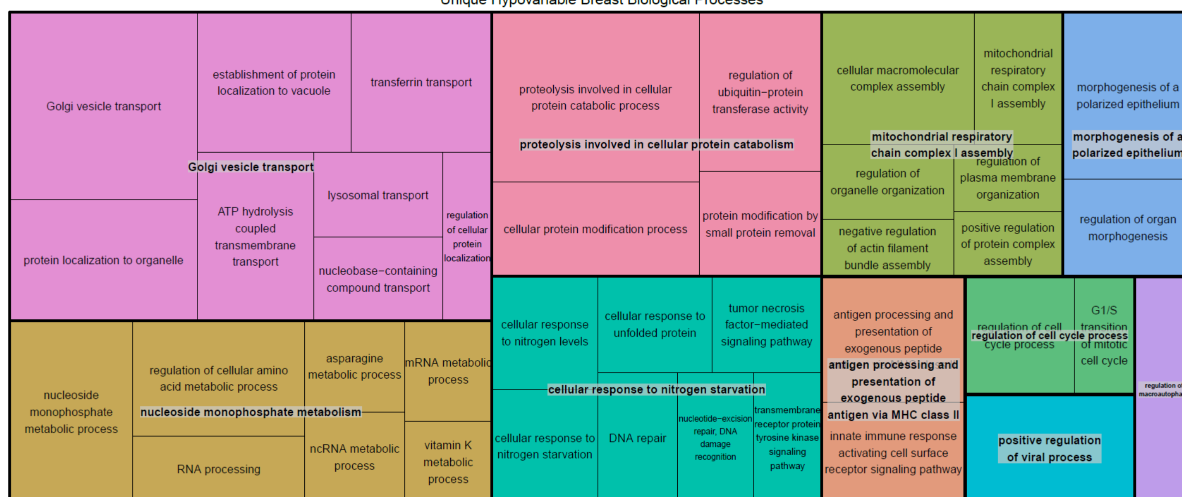


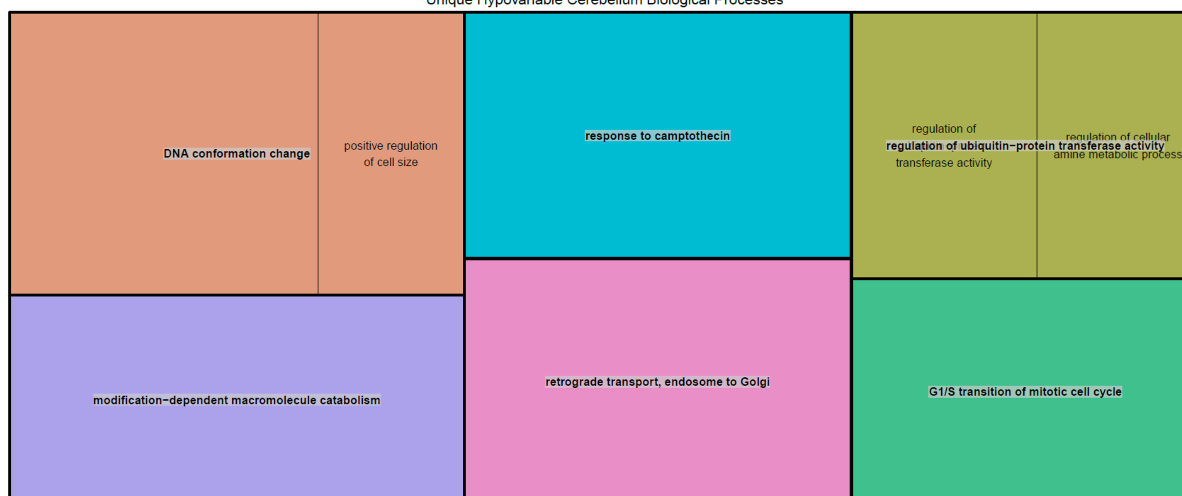
Unique Hypervariable Cerebellum Biological Processes



Unique Hypervariable Frontal Biological Processes




Unique Hypovariable Breast Biological Processes
Golgi vesicle transport
establishment of protein localization to vacuole
transferrin transport
Golgi vesicle transport
protein localization to organelle
ATP hydrolysis coupled transmembrane transport
lysosomal transport
regulation of cellular protein localization
nucleobase−containing compound transport
proteolysis involved in cellular protein catabolic process
regulation of ubiquitin−protein transferase activity
proteolysis involved in cellular protein catabolism
cellular protein modification process
protein modification by small protein removal
cellular macromolecular complex assembly
mitochondrial respiratory chain complex I assembly
mitochondrial respiratory chain complex I assembly
regulation of organelle organization
regulation of plasma membrane organization
negative regulation of actin filament bundle assembly
positive regulation of protein complex assembly
morphogenesis of a polarized epithelium
morphogenesis of a polarized epithelium
regulation of organ morphogenesis
nucleoside monophosphate metabolic process
regulation of cellular amino acid metabolic process
asparagine metabolic process
mRNA metabolic process
nucleoside monophosphate metabolism
RNA processing
ncRNA metabolic process
vitamin K metabolic process
cellular response to nitrogen levels
cellular response to unfolded protein
tumor necrosis factor−mediated signaling pathway
cellular response to nitrogen starvation
cellular response to nitrogen starvation
DNA repair
nucleotide−excision repair, DNA damage recognition
transmembrane receptor protein tyrosine kinase signaling pathway
antigen processing and presentation of exogenous peptide
antigen processing and presentation of exogenous peptide antigen via MHC class II
innate immune response activating cell surface receptor signaling pathway
regulation of cell cycle process
regulation of cell cycle process
G1/S transition of mitotic cell cycle
positive regulation of viral process
regulation of macroautophagy


Unique Hypovariable Cerebellum Biological Processes
DNA conformation change
positive regulation of cell size
response to camptothecin
regulation of ubiquitin−protein transferase activity
regulation of ubiquitin−protein transferase activity
regulation of cellular amine metabolic process
modification−dependent macromolecule catabolism
retrograde transport, endosome to Golgi
G1/S transition of mitotic cell cycle

Unique Hypovariable Frontal Biological Processes

RNA processing

ncRNA metabolic process

mRNA metabolic process

amide biosynthetic process

regulation of gene expression

ncRNA metabolism

cap−independent translational initiation

DNA repair

regulation of macromolecule biosynthetic process

RNA modification

covalent chromatin modification

ribonucleoprotein complex assembly

tRNA aminoacylation

base−excision repair, AP site formation

RNA (guanine−N7)−methylation

RNA biosynthetic process

transcription, DNA−templated

RNA secondary structure unwinding

2'−deoxyribonucleotide metabolic process

regulation of nucleobase−containing compound metabolic process

regulation of RNA metabolic process

nucleic acid phosphodiester bond hydrolysis

DNA biosynthetic process

DNA modification

cellular hyperosmotic response

response to interleukin−15

regulation of entry of bacterium into host cell

Common Hypervariable Cellular Component

anchored component of external side of plasma membrane

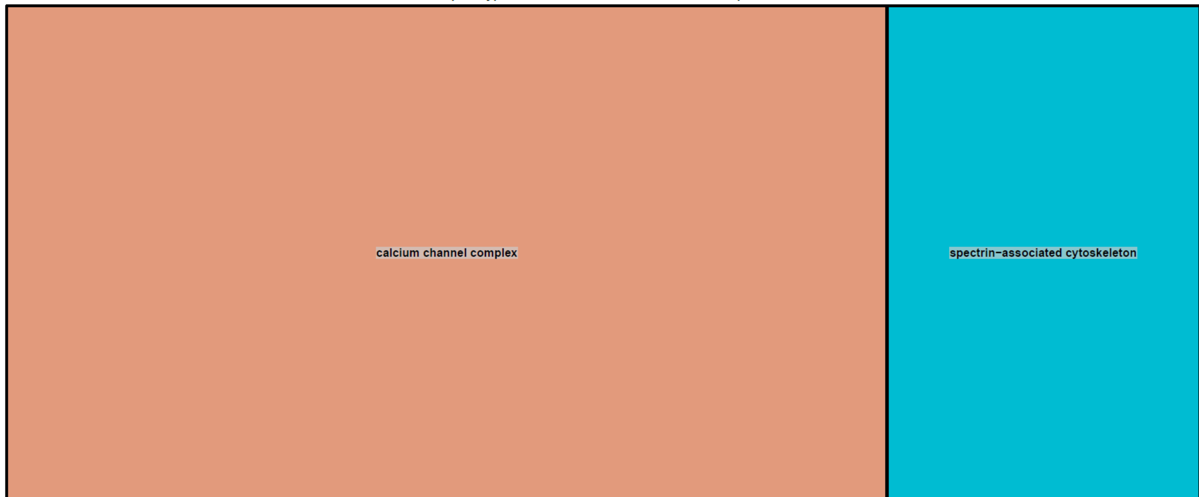
cytosolic ribosome

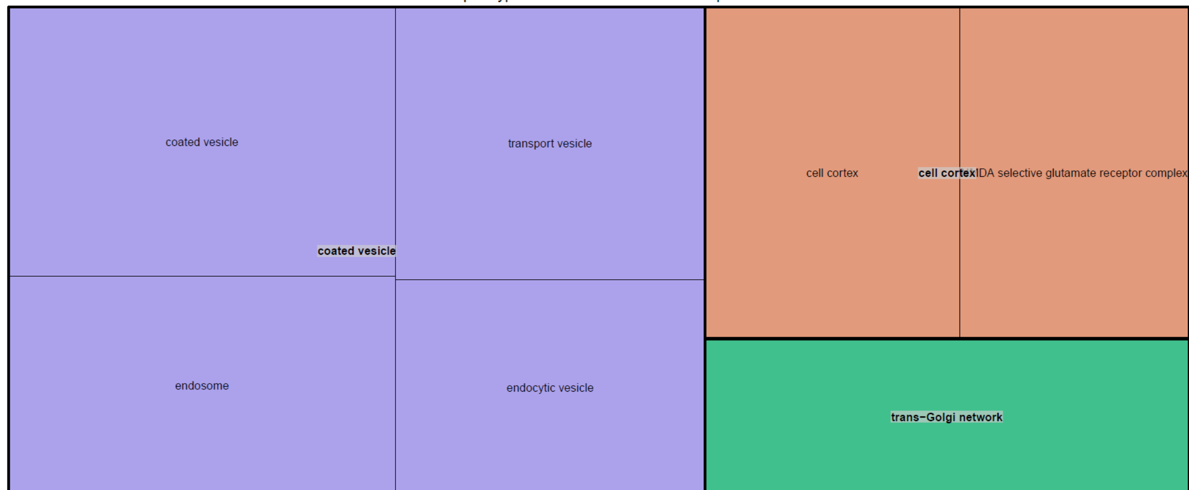
Common Hypovariable Cellular Component

endosome

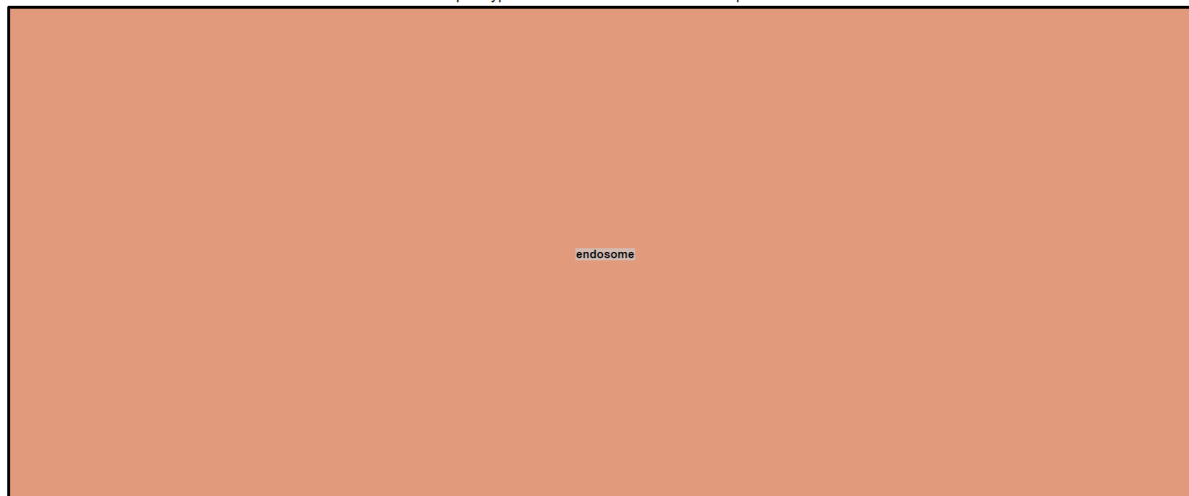
IkappaB kinase complex

Unique Hypervariable Breast Cellular Component
contractile actin filament bundle
contractile actin filament bundle
intermediate filament cytoskeleton
secretory vesicle


Unique Hypervariable Cerebellum Cellular Component
calcium channel complex
spectrin−associated cytoskeleton


Unique Hypervariable Frontal Cellular Component
coated vesicle
transport vesicle
cell cortex
cell cortex
IDA selective glutamate receptor complex
coated vesicle
endosome
endocytic vesicle
trans−Golgi network

Unique Hypovariable Breast Cellular Component
endosome
pigment granule
chromosome, centromeric region
protein kinase cor
protein kinase complex
Cul4−RING E3 ubiquitin ligase complex
endosome
chromosome, centromeric region
trans−Golgi network
transport vesicle
telosome
lysosomal membrane


Unique Hypovariable Cerebellum Cellular Component
endosome

Unique Hypovariable Frontal Cellular Components

H4 histone acetyltransferase complex

eukaryotic 43S preinitiation complex

H4 histone acetyltransferase complex

nuclear body

sex chromatin

Set1C/COMPASS complex

U12-type spliceosomal complex

spliceosomal snRNP complex

Supplementary Figure 3. Complete list of GO terms for essential genes

Hypervariable Essential GO Annotations in Breast Tissue

Hypervariable Essential GO Annotations in Frontal Cortex Tissue

Hypervariable Essential GO Annotations in Cerebellum Tissue

positive regulation of cell differentiation
embryo development ending in birth or egg hatching
regulation of nervous system development
muscle cell differentiation
regulation of gene expression
positive regulation of cell proliferation
negative regulation of cell proliferation
neurogenesis
striated muscle tissue development
regulation of cellular biosynthetic process
positive regulation of cell death
regulation of angiogenesis
regulation of protein metabolic process
mesenchymal cell differentiation
proximal/distal pattern formation
positive regulation of cellular metabolic process
regulation of intracellular signal transduction
glial cell fate specification
regulation of RNA metabolic process
regulation of macrophage derived foam cell differentiation
transcription, DNA-templated
foam cell differentiation
pigment cell differentiation
glial cell differentiation
RNA biosynthetic process
regulation of neuron death
myelination
positive regulation of cell differentiation
positive regulation of cell communication
ear morphogenesis
embryonic limb morphogenesis
ureter urothelium development
regulation of lipid metabolic process
neuron fate commitment
epithelial cell differentiation
salivary gland morphogenesis
fat cell differentiation
spinal cord patterning
regulation of cell morphogenesis
negative regulation of cell communication
endodermal cell differentiation
regulation of osteoblast proliferation
leukocyte differentiation
forebrain regionalization
neuron fate determination
positive regulation of signal transduction
stem cell differentiation
cerebral cortex regionalization
positive regulation of lipid metabolic process
epithelial cell fate commitment
lung epithelium development
kidney epithelium development
skin epidermis development
regulation of stem cell proliferation
cellular protein modification process
regulation of secretion
peptide secretion
regulation of protein transport
acid secretion
histamine secretion
endocytosis
hormone secretion
organic anion transport
regulation of intracellular transport
azole transport
regulation of secretion
epithelial cell migration
positive regulation of cellular component movement
regulation of actin filament-based movement
amine transport
ammonium transport
regulation of lipid transport
regulation of amine transport
anion transport
regulation of ion transport
regulation of transmembrane transport
neutrophil chemotaxis
import across plasma membrane
transmembrane receptor protein tyrosine kinase signaling pathway
regulation of phosphorus metabolic process
cellular response to organic cyclic compound
ERK1 and ERK2 cascade
response to progesterone
regulation of Notch signaling
transmembrane receptor protein tyrosine kinase signaling pathway
growth factor stimulus
phosphorylation
trans-synaptic signaling
Fc receptor mediated stimulatory signaling pathway
response to monosaccharide
cellular response to lipid
response to nicotine
positive regulation of immune response
positive regulation of cellular component organization
actomyosin structure organization
axon development
regulation of cell projection organization
regulation of cell projection organization
protein complex assembly
negative regulation of signal transduction
heart contraction
relaxation of cardiac muscle
negative regulation of response to external stimulus
heart contraction
smooth muscle contraction
sensory perception of pain
response to bacterium
cellular ion homeostasis
positive regulation of transferase activity
cellular ion homeostasis
intestinal epithelial structure maintenance
mast cell mediated immunity
chronic inflammatory response

Hypovariable Essential GO Annotations in Breast Tissue

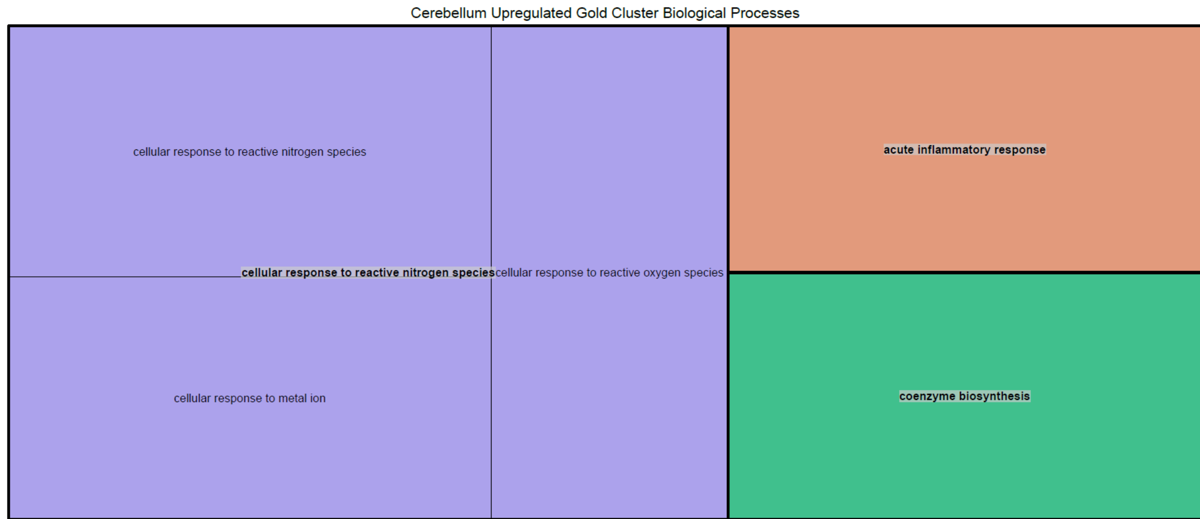
cellular protein modification process
regulation of RNA metabolic process
RNA biosynthetic process
transcription, DNA-templated
positive regulation of macromolecule metabolic process
negative regulation of cellular component organization
regulation of protein metabolic process
positive regulation of protein complex assembly
protein-DNA complex assembly
negative regulation of cellular component organization
stress granule assembly
negative regulation of cell proliferation
covalent chromatin modification
regulation of gene expression
regulation of macromolecule biosynthetic process
multivesicular body assembly
positive regulation of exosomal secretion
cytoplasmic mRNA processing body assembly
cellular macromolecular complex assembly
negative regulation of NF-kappaB transcription factor activity
regulation of nucleobase-containing compound metabolic process
regulation of cellular protein metabolic process
regulation of organelle organization
endosome organization
pigment granule organization
DNA repair
proteolysis involved in cellular protein catabolic process
stress-activated MAPK cascade
nucleotide-excision repair, DNA damage recognition
innate immune response activating cell surface receptor signaling pathway
positive regulation of cell communication
Wnt signaling pathway
cellular response to DNA repair
antigen processing and presentation of exogenous peptide antigen via MHC class II
tumor necrosis factor-mediated signaling pathway
regulation of intracellular signal transduction
mitochondrial DNA metabolic process
positive regulation of signal transduction
cellular response to nitrogen levels
lagging strand elongation
cellular response to light stimulus
beta-amyloid metabolic process
protein localization to organelle
regulation of cellular protein localization
establishment of protein localization to vacuole
lysosomal transport
Golgi vesicle transport
embryo development ending in birth
morphogenesis of a polarized epithelium
embryo development ending in birth or egg hatching
regulation of organ morphogenesis
regulation of macroautophagy
apoptotic process
phosphorylation
positive regulation of viral process
regulation of cell cycle process

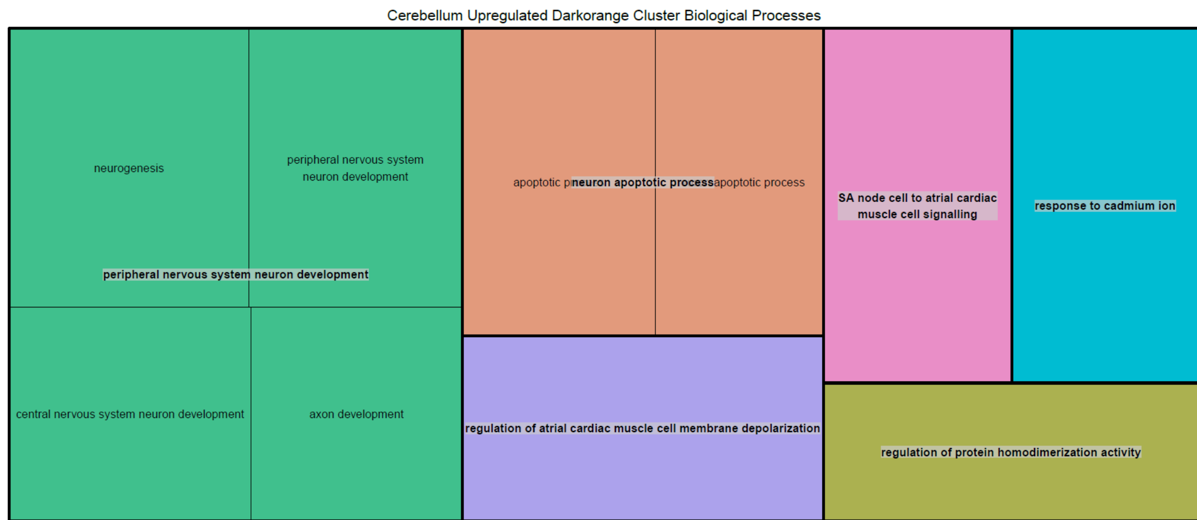
Hypovariable Essential GO Annotations in Cerebellum Tissue

DNA repair
cellular protein modification process
regulation of cellular protein metabolic process
regulation of protein metabolic process
fatty acid catabolic process
proteolysis involved in cellular protein catabolic process
DNA conformation change
mitochondrial DNA metabolic process
protein modification by small protein removal
DNA-dependent DNA replication maintenance of fidelity
DNA repair
DNA ligation
nucleic acid phosphodiester bond hydrolysis
very long-chain fatty acid metabolic process
short-chain fatty acid metabolic process
adenylate cyclase-modulating G-protein coupled receptor signaling pathway
nucleoside monophosphate biosynthetic process
Wnt signaling pathway
nucleotide-excision repair, DNA damage recognition
cellular response to ionizing radiation
strand displacement
regulation of cellular catabolic process
protein complex assembly
positive regulation of protein complex assembly
nuclear division
autophagosome organization
protein-DNA complex assembly
negative regulation of cellular component organization
mitochondrial genome maintenance
negative regulation of cellular component organization
microtubule nucleation
negative regulation of NF-kappaB transcription factor activity
negative regulation of signal transduction
multivesicular body assembly
positive regulation of viral release from host cell
regulation of nucleobase-containing compound metabolic process
regulation of gene expression
RNA biosynthetic process
positive regulation of viral release from host cell
positive regulation of cellular metabolic process
regulation of cellular biosynthetic process
regulation of RNA metabolic process
transcription, DNA-templated
protein localization to organelle
regulation of cellular protein localization
regulation of cellular protein localization
regulation of cell cycle process
G1/S transition of mitotic cell cycle
regulation of cell cycle process
embryo development ending in birth or egg hatching
embryo development ending in birth or egg hatching
camera-type eye development
apoptotic process

Hypovariable Essential GO Annotations in Frontal Cortex Tissue

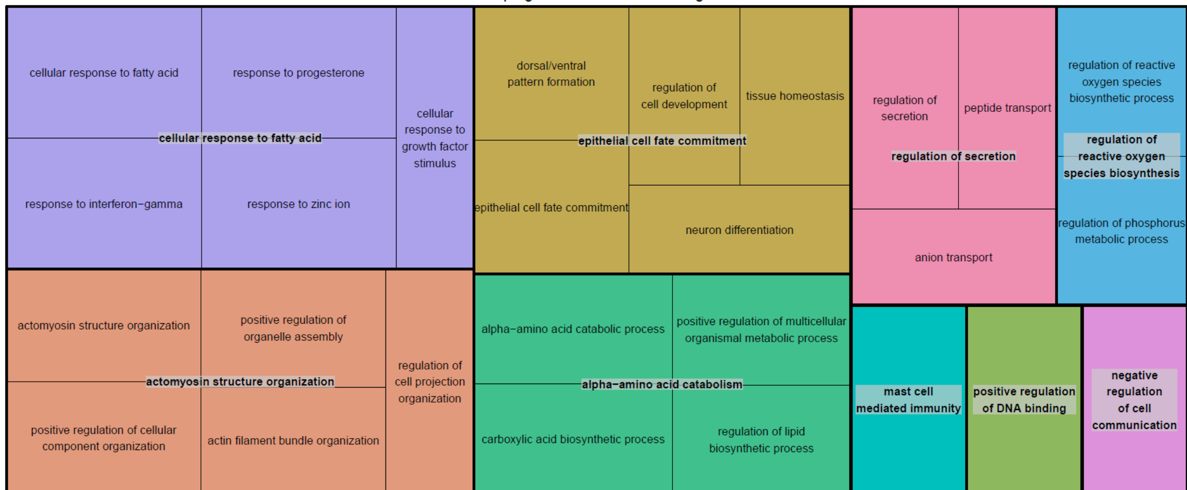
Supplementary Figure 4. Complete list of GO terms for age-regulated Hypervariable genes





Cerebellum Downregulated Yellow Cluster Biological Processes
optic nerve development
cell morphogenesis involved in differentiation
optic nerve development
neurogenesis
regulation of substrate adhesion−dependent cell spreading
oligodendrocyte development
response to interferon−gamma
antigen processing and presentation of exogenous peptide antigen via MHC class II
positive regulation of leukocyte activation
signal release from synapse
positive regulation of calcium ion transport into cytosol
bleb assembly
positive regulation of actin filament bundle assembly
negative regulation of cell proliferation
phenol−containing compound biosynthesis

Cerebellum Downregulated Green Cluster Biological Processes
establishment of protein localization to chromosome
regulation of establishment of protein localization
telomerase RNA localization to Cajal body
telomerase RNA localization to Cajal body
regulation of secretion
mast cell mediated immunity
regulation of mast cell activation involved in immune response
defense response to protozoan
positive regulation of cellular component organization
positive regulation of tissue remodeling
positive regulation of tissue remodeling
ionotropic glutamate receptor signaling pathway
trans−synaptic signaling
trans−synaptic signaling
adherens junction assembly

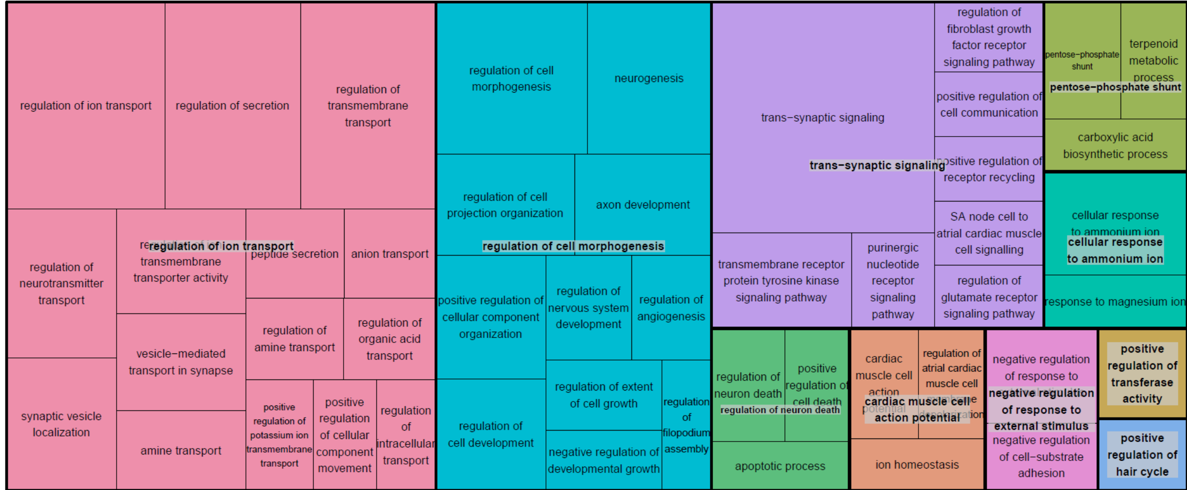


Frontal Cortex Upregulated Gold Cluster Biological Processes
cellular response to fatty acid
response to progesterone
cellular response to fatty acid
cellular response to growth factor stimulus
response to interferon−gamma
response to zinc ion
dorsal/ventral pattern formation
regulation of cell development
tissue homeostasis
epithelial cell fate commitment
epithelial cell fate commitment
neuron differentiation
regulation of secretion
peptide transport
regulation of secretion
anion transport
regulation of reactive oxygen species biosynthetic process
regulation of reactive oxygen species biosynthesis
regulation of phosphorus metabolic process
actomyosin structure organization
positive regulation of organelle assembly
actomyosin structure organization
regulation of cell projection organization
positive regulation of cellular component organization
actin filament bundle organization
alpha−amino acid catabolic process
positive regulation of multicellular organismal metabolic process
alpha−amino acid catabolism
carboxylic acid biosynthetic process
regulation of lipid biosynthetic process
mast cell mediated immunity
positive regulation of DNA binding
negative regulation of cell communication

Frontal Cortex Upregulated Dark Orange Cluster Biological Processes

Frontal Cortex Upregulated Yellow Cluster Biological Processes

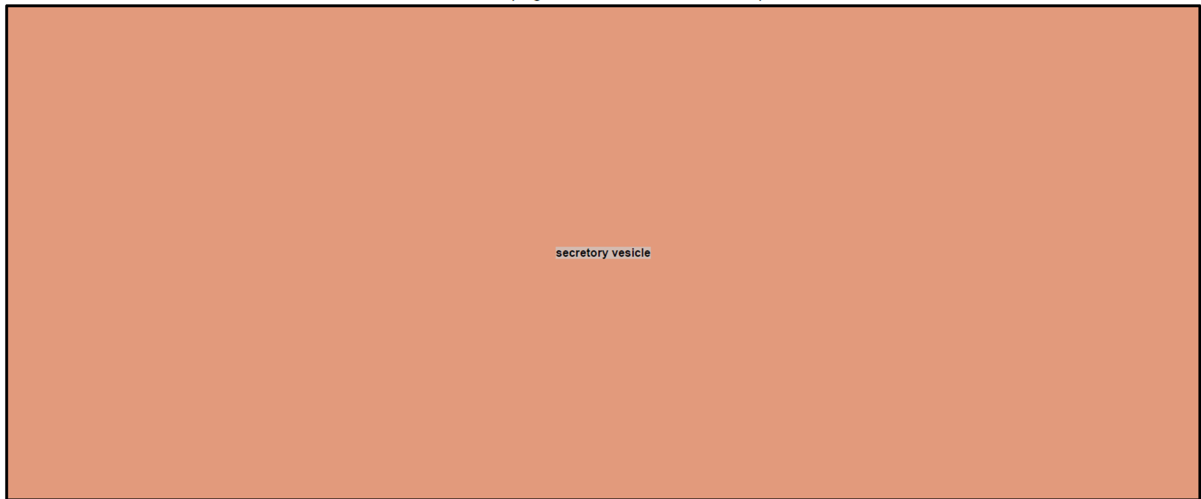
Frontal Cortex Downregulated Red Cluster Biological Processes

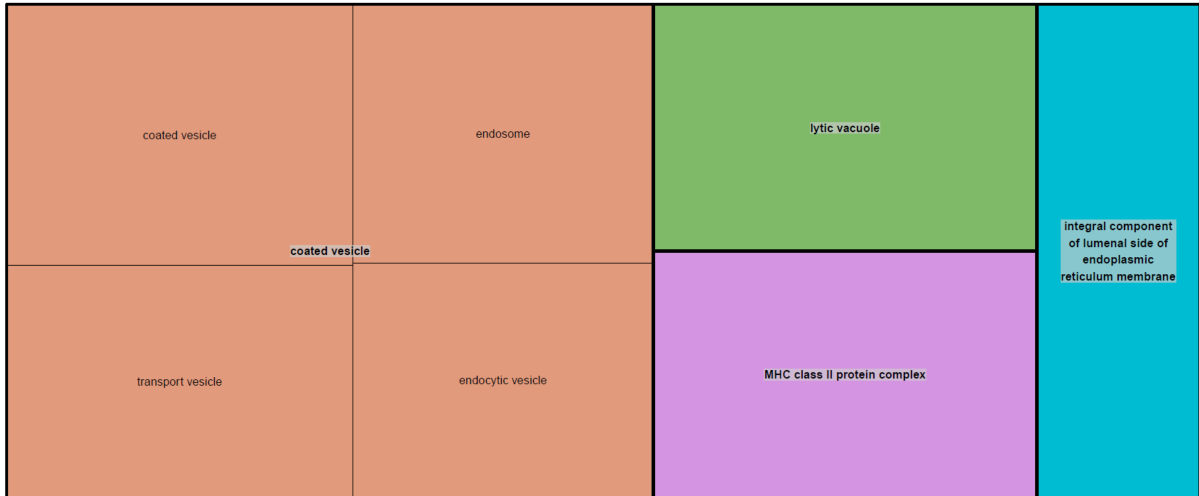


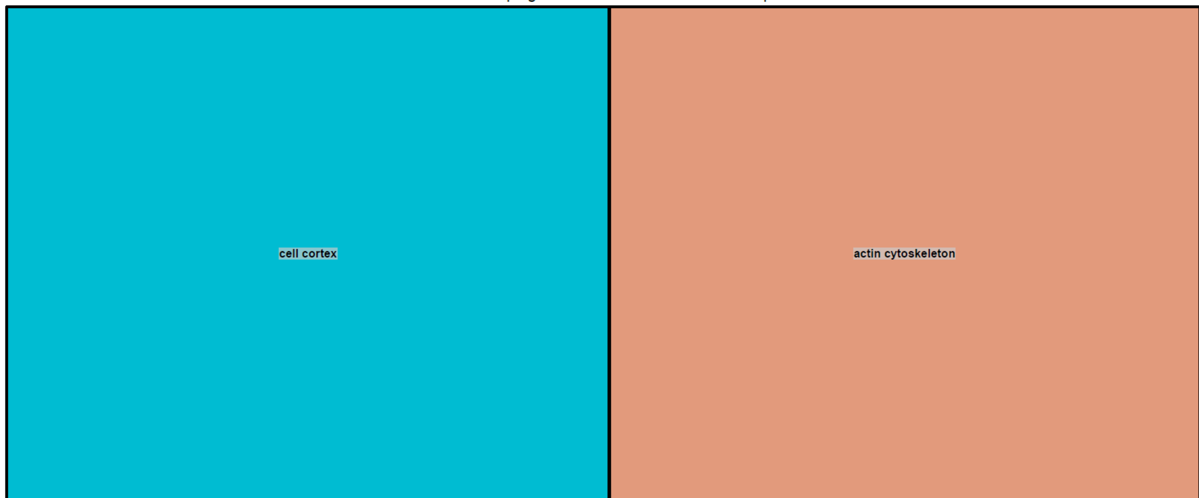
Cerebellum Upregulated Darkorange Cluster Cellular Component

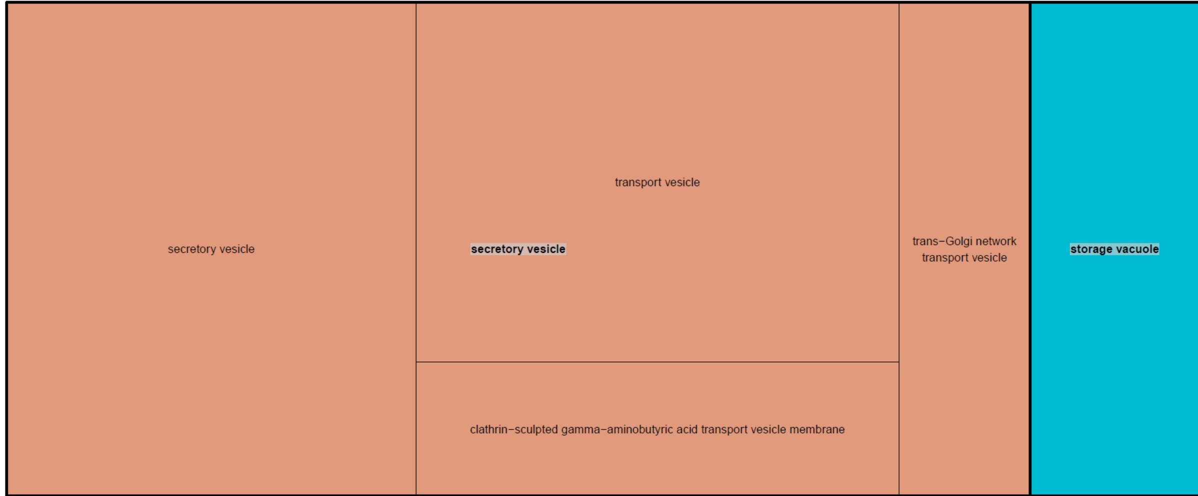
Cerebellum Downregulated Yellow Cluster Cellular Component

Cerebellum Downregulated Green Cluster Cellular Component
spliceosomal snRNP complex
spliceosomal snRNP complex
U2−type spliceosomal complex

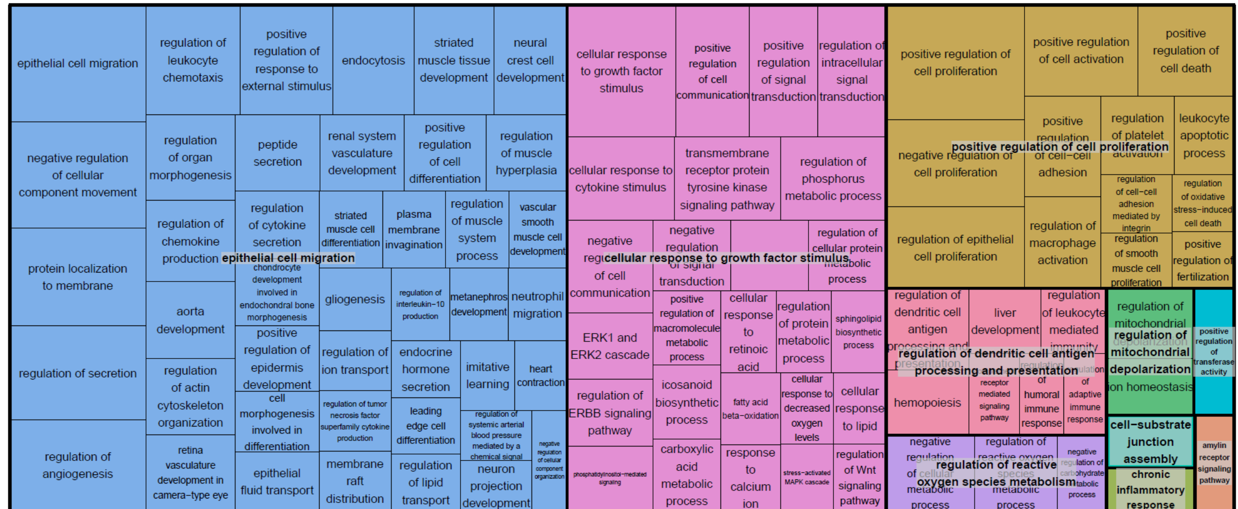

Frontal Cortex Upregulated Gold Cluster Cellular Component
secretory vesicle

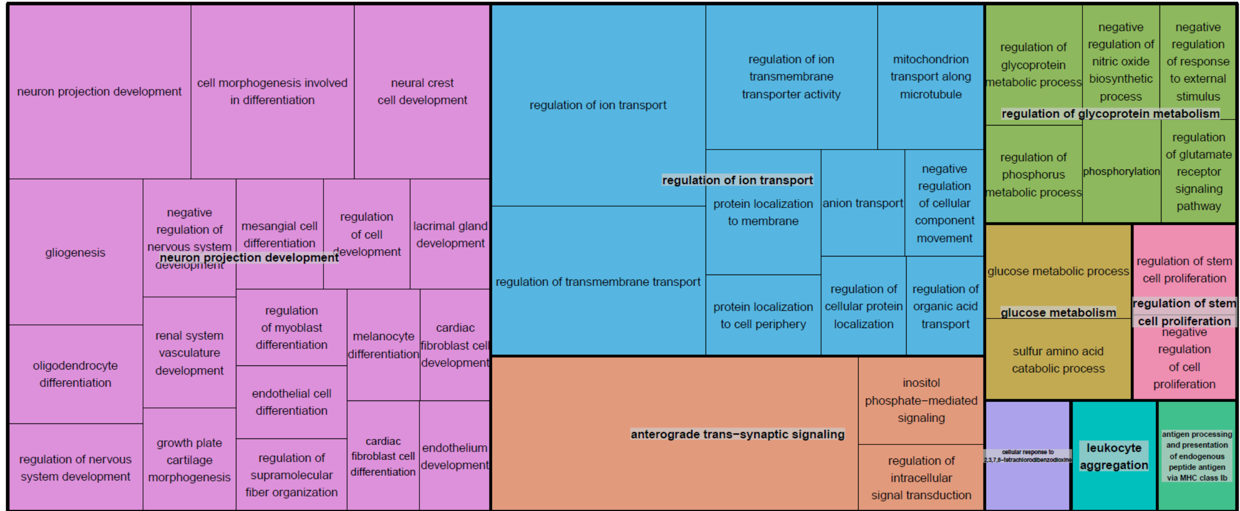

Frontal Cortex Upregulated Darkorange Cluster Cellular Component
coated vesicle
endosome
coated vesicle
transport vesicle
endocytic vesicle
lytic vacuole
MHC class II protein complex
integral component of lumenal side of endoplasmic reticulum membrane


Frontal Cortex Upregulated Yellow Cluster Cellular Component
cell cortex
actin cytoskeleton

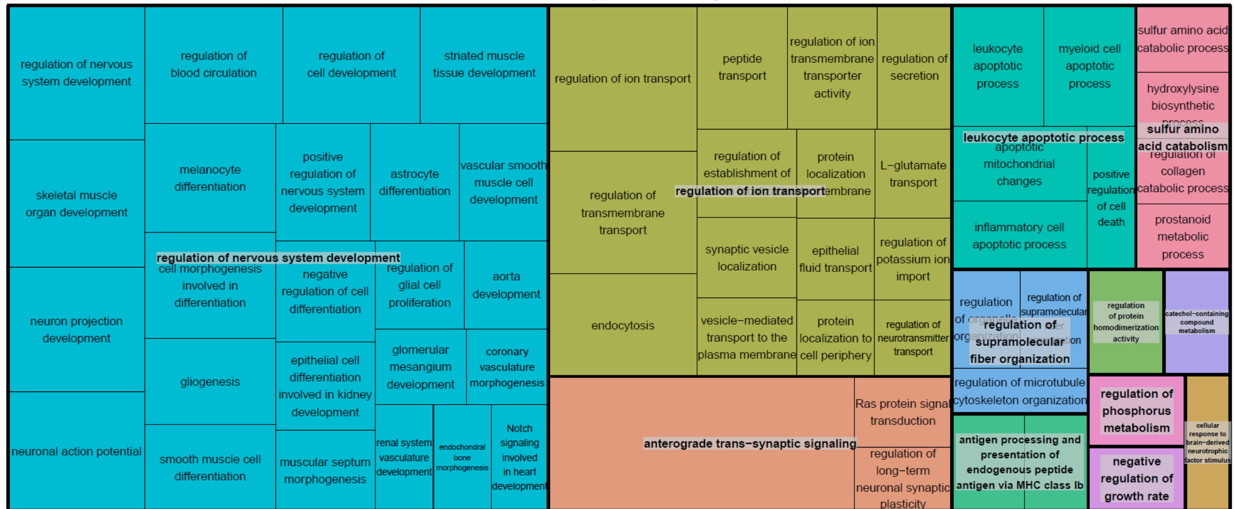


Supplementary Figure 5. Complete list of GO terms for pseudo-Hypervariable and pseudo-Hypovariable genes

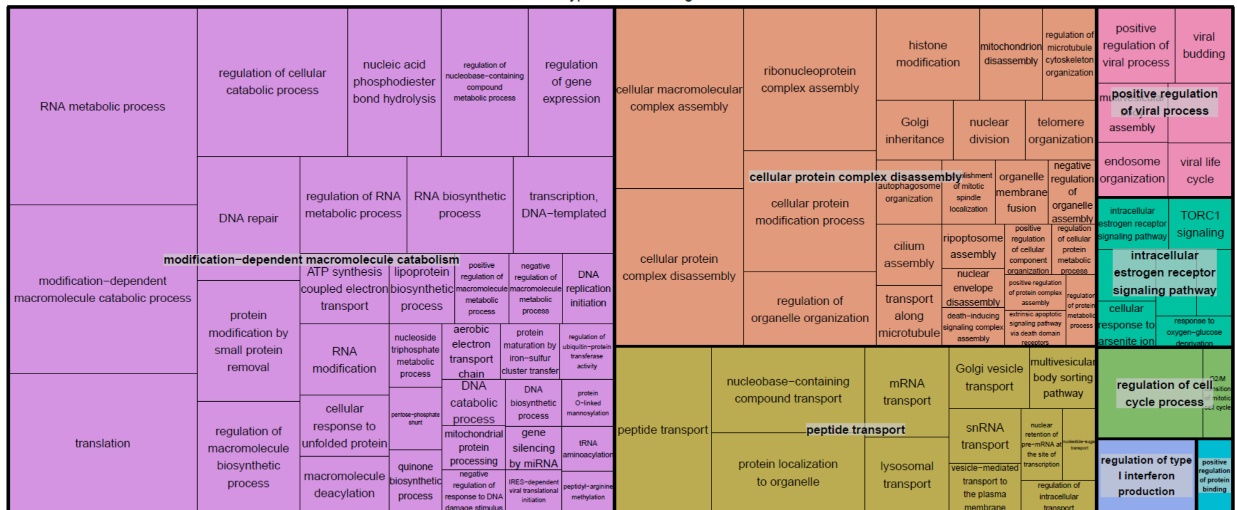




Cerebellum Pseudo−Hypervariable Biological Processes
neuron projection development
cell morphogenesis involved in differentiation
neural crest cell development
gliogenesis
negative regulation of nervous system development
mesangial cell differentiation
regulation of cell development
lacrimal gland development
neuron projection development
oligodendrocyte differentiation
renal system vasculature development
regulation of myoblast differentiation
melanocyte differentiation
cardiac fibroblast cell development
endothelial cell differentiation
regulation of nervous system development
growth plate cartilage morphogenesis
regulation of supramolecular fiber organization
cardiac fibroblast cell differentiation
endothelium development
regulation of ion transport
regulation of ion transmembrane transporter activity
mitochondrion transport along microtubule
regulation of ion transport
protein localization to membrane
anion transport
negative regulation of cellular component movement
regulation of transmembrane transport
protein localization to cell periphery
regulation of cellular protein localization
regulation of organic acid transport
anterograde trans−synaptic signaling
inositol phosphate−mediated signaling
regulation of intracellular signal transduction
regulation of glycoprotein metabolic process
negative regulation of nitric oxide biosynthetic process
negative regulation of response to external stimulus
regulation of glycoprotein metabolism
regulation of phosphorus metabolic process
phosphorylation
regulation of glutamate receptor signaling pathway
glucose metabolic process
glucose metabolism
regulation of stem cell proliferation
regulation of stem cell proliferation
sulfur amino acid catabolic process
negative regulation of cell proliferation
leukocyte aggregation
antigen processing and presentation of endogenous peptide antigen via MHC class Ib



Frontal Cortex Pseudo−Hypervariable Biological Processes
regulation of nervous system development
regulation of blood circulation
regulation of cell development
striated muscle tissue development
skeletal muscle organ development
melanocyte differentiation
positive regulation of nervous system development
astrocyte differentiation
vascular smooth muscle cell development
regulation of nervous system development
cell morphogenesis involved in differentiation
negative regulation of cell differentiation
regulation of glial cell proliferation
aorta development
neuron projection development
gliogenesis
epithelial cell differentiation involved in kidney development
glomerular mesangium development
coronary vasculature morphogenesis
neuronal action potential
smooth muscle cell differentiation
muscular septum morphogenesis
renal system vasculature development
endochondral bone morphogenesis
Notch signaling involved in heart development
regulation of ion transport
peptide transport
regulation of ion transmembrane transporter activity
regulation of secretion
regulation of transmembrane transport
regulation of establishment of
protein localization
L−glutamate transport
regulation of ion transport
synaptic vesicle localization
epithelial fluid transport
regulation of potassium ion import
endocytosis
vesicle−mediated transport to the plasma membrane
protein localization to cell periphery
regulation of neurotransmitter transport
anterograde trans−synaptic signaling
Ras protein signal transduction
regulation of long−term neuronal synaptic plasticity
leukocyte apoptotic process
myeloid cell apoptotic process
sulfur amino acid catabolic process
hydroxylysine biosynthetic process
leukocyte apoptotic process
sulfur amino acid catabolism
apoptotic mitochondrial changes
positive regulation of cell death
regulation of collagen catabolic process
inflammatory cell apoptotic process
prostanoid metabolic process
regulation of supramolecular fiber organization
regulation of protein homodimerization activity
catechol−containing compound metabolism
regulation of microtubule cytoskeleton organization
regulation of phosphorus metabolism
antigen processing and presentation of endogenous peptide antigen via MHC class Ib
negative regulation of growth rate



Breast Pseudo−Hypovariable Biological Processes
RNA metabolic process
regulation of cellular catabolic process
nucleic acid phosphodiester bond hydrolysis
regulation of nucleobase−containing compound metabolic process
regulation of gene expression
DNA repair
regulation of RNA metabolic process
RNA biosynthetic process
transcription, DNA−templated
modification−dependent macromolecule catabolic process
modification−dependent macromolecule catabolism
ATP synthesis coupled electron transport
lipoprotein biosynthetic process
DNA replication initiation
protein modification by small protein removal
RNA modification
aerobic electron transport chain
translation
regulation of macromolecule biosynthetic process
cellular response to unfolded protein
DNA catabolic process
DNA biosynthetic process
mitochondrial protein processing
gene silencing by miRNA
tRNA aminoacylation
macromolecule deacylation
quinone biosynthetic process
cellular macromolecular complex assembly
ribonucleoprotein complex assembly
histone modification
mitochondrion disassembly
Golgi inheritance
nuclear division
telomere organization
cellular protein complex disassembly
cellular protein complex disassembly
cellular protein modification process
organelle membrane fusion
cilium assembly
ripoptosome assembly
nuclear envelope disassembly
regulation of organelle organization
transport along microtubule
positive regulation of viral process
viral budding
positive regulation of viral process
endosome organization
viral life cycle
TORC1 signaling
intracellular estrogen receptor signaling pathway
nucleobase−containing compound transport
mRNA transport
Golgi vesicle transport
multivesicular body sorting pathway
peptide transport
peptide transport
snRNA transport
protein localization to organelle
lysosomal transport
regulation of cell cycle process
regulation of type I interferon production

Cerebellum Pseudo−Hypovariable Biological Processes

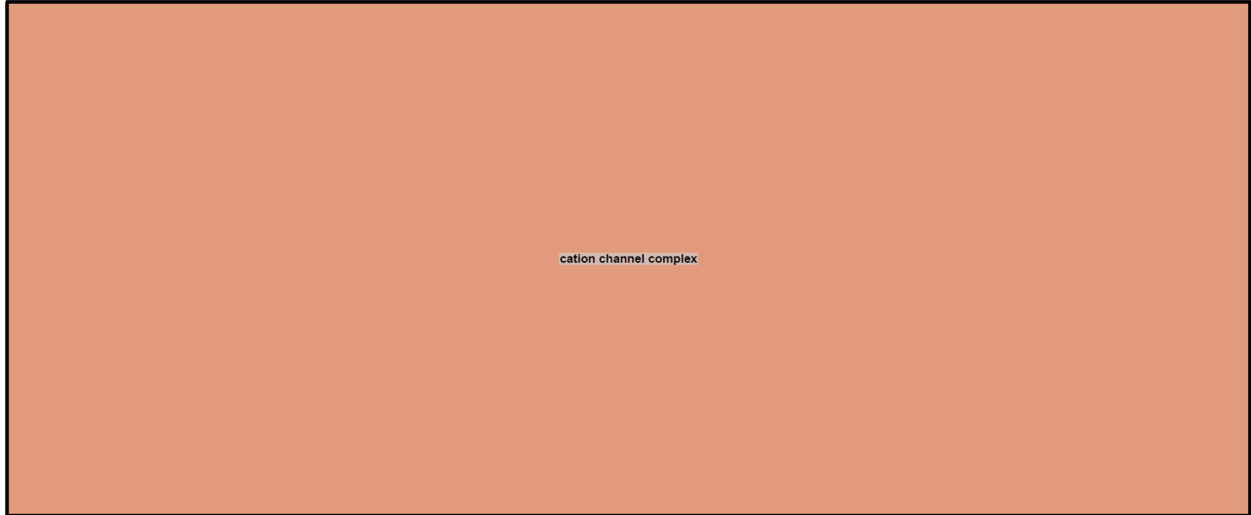
RNA metabolic process
modification−dependent macromolecule catabolic process
nucleic acid phosphodiester bond hydrolysis
protein modification by small protein removal
translation
DNA repair
regulation of nucleobase−containing compound metabolic process
regulation of macromolecule biosynthetic process
cellular response to unfolded protein
RNA modification
translation
viral translation
IRES−dependent viral translational initiation
coenzyme A metabolic process
nucleoside phosphate biosynthetic process
regulation of RNA metabolic process
ubiquinone metabolic process
transcription, DNA−templated
2−deoxyribonucleotide metabolic process
regulation of gene expression
regulation of cellular catabolic process
RNA biosynthetic process
negative regulation of macromolecule metabolic process
protoporphyrinogen IX metabolic process
nucleobase−containing compound transport
protein localization to organelle
peptide transport
vesicle coating
nucleobase−containing compound transport
Golgi vesicle transport
mRNA transport
plasma membrane to endosome transport
endosome organization
regulation of cellular protein localization
vesicle−mediated transport to the plasma membrane
ribonucleoprotein complex assembly
cilium assembly
regulation of organelle organization
cellular macromolecular complex assembly
ribonucleoprotein complex assembly
organization
cellular protein complex disassembly
regulation of microtubule cytoskeleton organization
histone modification
cellular protein modification process
telomere organization
regulation of protein metabolic process
rRNA methylation
regulation of cell cycle process
regulation of cell cycle process
G2/M transition of mitotic cell cycle
exit from mitosis
regulation of cell cycle checkpoint
androgen receptor signaling pathway
neural tube closure

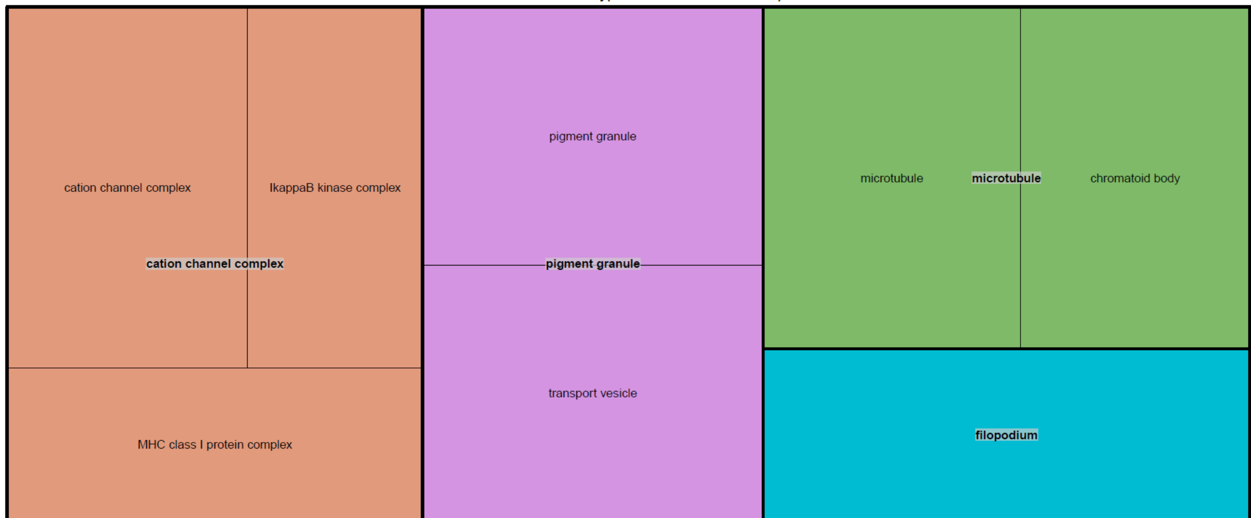
Frontal Cortex Pseudo−Hypovariable Biological Processes

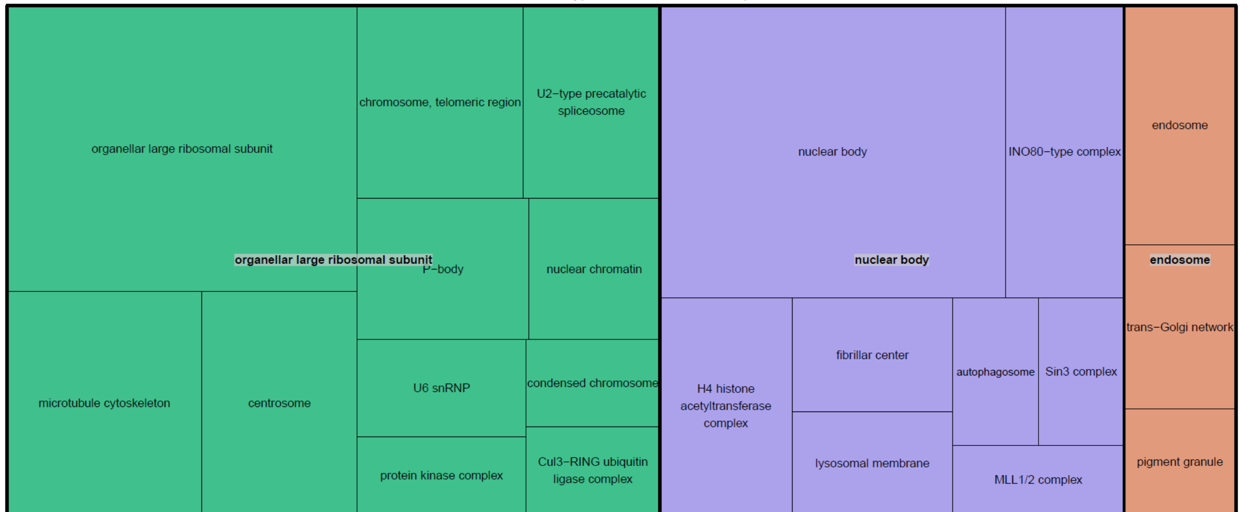
RNA metabolic process
RNA modification
regulation of gene expression
tRNA methylation
DNA repair
peptidyl−arginine methylation
RNA biosynthetic process
regulation of nucleobase−containing compound metabolic process
protein modification by small protein removal
modification−dependent macromolecule catabolism
modification−dependent macromolecule catabolic process
regulation of RNA metabolic process
regulation of cellular catabolic process
regulation of viral−induced cytoplasmic pattern recognition receptor signaling pathway
gene silencing by miRNA
coenzyme A metabolic process
protein deglycosylation
oxidative single−stranded DNA demethylation
sterol biosynthetic process
fatty acid catabolic process
translation
regulation of macromolecule biosynthetic process
transcription, DNA−templated
protein repair
DNA modification
nucleic acid phosphodiester bond hydrolysis
glycolipid biosynthetic process
protein localization to organelle
Golgi vesicle transport
peptide transport
nucleobase−containing compound transport
regulation of cellular protein localization
transport along microtubule
nucleobase−containing compound transport
mRNA transport
regulation of microtubule cytoskeleton organization
microtubule anchoring at microtubule organizing center
ribonucleoprotein complex assembly
regulation of organelle organization
histone modification
ribonucleoprotein complex assembly
cellular protein modification process
telomere organization
cellular macromolecular complex assembly
cellular protein complex disassembly
cilium assembly
regulation of cell cycle process
regulation of cell cycle process
G1/S transition of mitotic cell cycle
glucocorticoid mediated signaling pathway
regulation by virus of viral protein levels in host cell

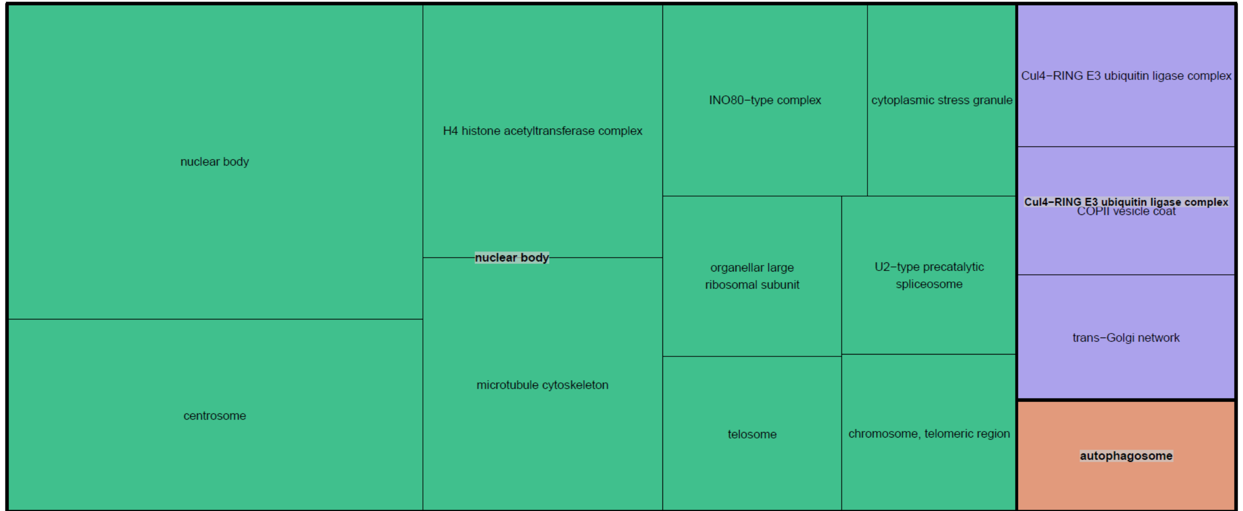
Breast Pseudo−Hypervariable Cellular Component

actin cytoskeleton
actin cytoskeleton
cytosolic large ribosomal subunit
myofibril
actin filament
secretory vesicle
secretory vesicle
secretory granule lumen
endocytic vesicle
calcitonin family receptor complex
cell cortex
cell cortex
filopodium

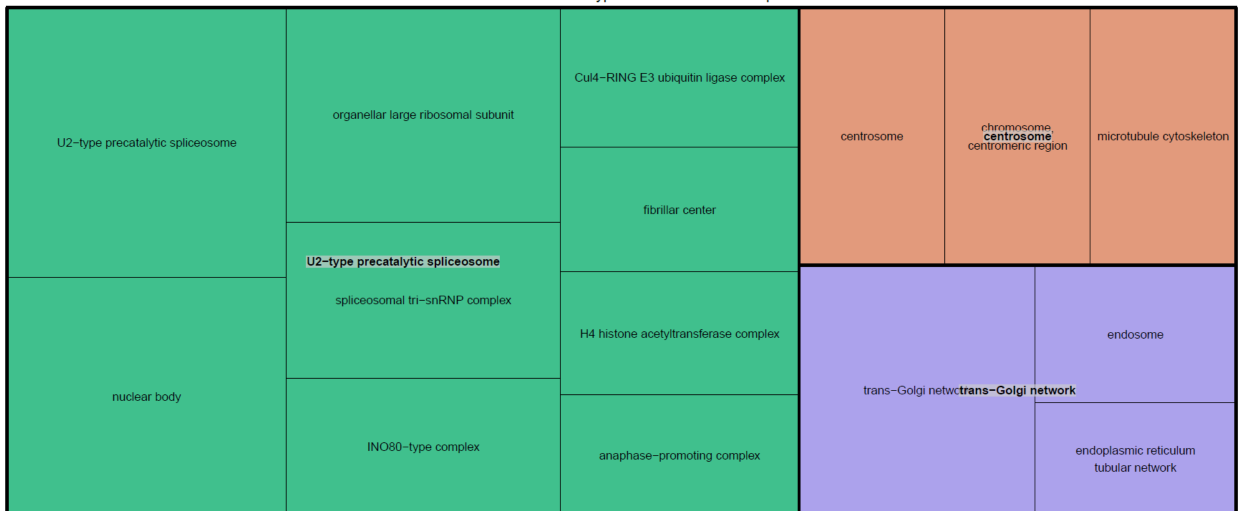

Cerebellum Pseudo−Hypervariable Cellular Component
cation channel complex


Frontal Cortex Pseudo−Hypervariable Cellular Components
cation channel complex
IkappaB kinase complex
cation channel complex
MHC class I protein complex
pigment granule
pigment granule
transport vesicle
microtubule
microtubule
chromatoid body
filopodium


Breast Pseudo−Hypovariable Cellular Component
organellar large ribosomal subunit
chromosome, telomeric region
U2−type precatalytic spliceosome
organellar large ribosomal subunit
P−body
nuclear chromatin
microtubule cytoskeleton
centrosome
U6 snRNP
condensed chromosome
protein kinase complex
Cul3−RING ubiquitin ligase complex
nuclear body
INO80−type complex
nuclear body
H4 histone acetyltransferase complex
fibrillar center
autophagosome
Sin3 complex
lysosomal membrane
MLL1/2 complex
endosome
endosome
trans−Golgi network
pigment granule



Cerebellum Pseudo−Hypovariable Cellular Components
nuclear body
centrosome
H4 histone acetyltransferase complex
nuclear body
microtubule cytoskeleton
INO80−type complex
cytoplasmic stress granule
organellar large ribosomal subunit
U2−type precatalytic spliceosome
telosome
chromosome, telomeric region
Cul4−RING E3 ubiquitin ligase complex
Cul4−RING E3 ubiquitin ligase complex
COPII vesicle coat
trans−Golgi network
autophagosome



Frontal Cortex Pseudo−Hypovariable Cellular Components
U2−type precatalytic spliceosome
nuclear body
organellar large ribosomal subunit
U2−type precatalytic spliceosome
spliceosomal tri−snRNP complex
INO80−type complex
Cul4−RING E3 ubiquitin ligase complex
fibrillar center
H4 histone acetyltransferase complex
anaphase−promoting complex
centrosome
chromosome, centromeric region
centrosome
microtubule cytoskeleton
trans−Golgi network
endosome
endoplasmic reticulum tubular network