Allelic Imbalance Reveals Genetic and Epigenic Underpinnings of metabolic diseases

Sunjin Moon[1,2], So-Young Park[1,2,*]

[1]Senotherapy-based Metabolic Disease Control Research Center

[2]Department of Physiology, College of Medicine, Yeungnam University, Daegu, 42415, Republic of Korea

*To whom correspondence should be addressed: sypark@med.yu.ac.kr

ABSTRACT

Not all alleles express equally. Allelic imbalance (AI) denotes variations in gene expression based on the inherited versions, or alleles, of those genes. While AI has the potential to shed light on many complex diseases, its specific role in obesity and connection with gene regulation remains largely elusive. In this study, we delve into AI's correlation with gene expression in obesity and its potential ramifications for susceptibility to associated health complications. By melding transcriptomic and epigenomic data through a colocalization approach, we discerned significant variations in the length and functional roles of both coding and noncoding differentially expressed genes across AI spectrums. Further, our examination unveiled DNaseI hypersensitive sites in close association with genetic markers identified in large-scale association studies (GWAS-identified SNPs) and polygenic risk variants, emphasizing the importance of length-dependent regulation and allelic interactions. These discoveries not only expand our understanding of the genetic underpinnings of obesity but also highlight new paths for personalized diagnostics and treatments. Our findings robustly advocate for intensified research for integration of rare pathogenic variants into the intricate interplay between genes and the environment.

INTRODUCTION

Not all alleles are expressed equally. Allelic imbalance (AI), a phenomenon where one allele of a heterozygous gene pair is preferentially expressed over the other, provides a nuanced yet essential layer of gene regulation (Li et al. 2023) that can contribute to individual variation in disease susceptibility, including obesity (Yang et al 2003; Pierre et al., 2022). The concept of allelic imbalance has evolved considerably over the last few years (Luquette et al. 2019), AI lies at the epicenter of genetic phenomena, conserved strikingly across organisms (Kravitz et al. 2023; Zwemer et al. 2012; Campbell et al. 2008). Its significance extends through various tissues, developmental stages, and genetic studies, demonstrating the complexity and importance of AI (Crowley et al. 2015; Keane et al. 2011; Goncalves et al. 2012). While this inherent variance has been recognized in the realm of genetics for some time, its implications in complex disorders such as obesity remain underexplored.

Several genetic mechanisms, including parent-of-origin specific gene expression and genomic imprinting, underscore AI's role in developmental potential and stability (Dirks et al. 2019; Swanzey et al. 2020). Moreover, AI provides insights into genomic imprinting, X-link gene dynamics, and the evolutionary trajectory of imprinted genes(Ferguson-Smith and Surani 2001; Carrion, Michal, and Jiang 2023; E. Y. Wu et al. 2023; Zhabotynsky et al. 2022). The realm of metabolic disorders, particularly obesity, captures the profound influence of AI, emphasizing its role in the intricate dance between genetic predispositions and environmental factors (Rosen and Spiegelman, 2006; R. Liu and Nikolajczyk 2019)."

Allelic-specific expression (ASE) remains one of the most direct ways to observe allelic imbalance at the RNA level (Montgomery et al., 2010;), where cis-acting regulatory elements play a pivotal role in chromatin structural alterations and DNA methylation, influencing the evolutionary onset of imprinting (Ferguson-Smith and Surani 2001; Khamlichi and Feil 2018). Regulatory variants control allele-specific TF binding, DNase I hypersensitivity, and histone modifications, further shaping AI dynamics (W. Shi et al. 2016). Distinct expression heterogeneity is observed among X-linked genes, shedding light on potential sex biases affecting various biological functions (Nguyen and Disteche 2006; Carrel and Willard 2005; Xiong et al. 2010; Oliva et al. 2020).

The diagnostic prowess of AI in diseases is undeniable, influencing pathways from tumorigenesis to viral respiratory conditions and beyond (Bielski and Taylor 2021; Yuanxu Zhang et al. 2022; Sinnott-Armstrong et al. 2021). Regulatory variants with chromatin accessibility imbalances play significant roles in a plethora of traits and conditions, such as breast cancer (Q. Li et al. 2013). Nevertheless, challenges like pleiotropy, polygenicity, and "missing heritability" present obstacles in fully deciphering the implications of AI (Gallagher and Chen-Plotkin 2018).

From an evolutionary perspective, AI showcases broader patterns, with data from yeast indicating a dominant role for purifying selection in cis-regulatory evolution (Ronald and Akey 2007). Allele-specific chromatin marks, particularly at imprinted gene loci, bring in added layers of complexity (Ferguson-Smith and Surani 2001). Findings from Neanderthal introgressed alleles further underscore the dynamic nature of AI (McCoy, Wakefield, and Akey 2017).

The advances in technology, such as deep sequencing, paired with the accumulation of large-scale omics data and complemented by refined analytical tools like advanced statistical methods, continue to expand our understanding of AI, particularly in regions like MHC. Notably, the integration of various data sources enables better gene prioritization (Aygün et al. 2023). These advancements have also led to the development and integration of polygenic transcriptomic risk scores, enhancing the accuracy of genetic studies (Liang et

al. 2022; Gensterblum-Miller, Wu, and Sawalha 2018; Mu et al. 2022; Larsson et al. 2019).

Beyond the realm of genetics, epigenetic factors, such as DNA methylation in sperm, interplay with AI. Parent-of-origin effects, influenced by aspects like parental age and BMI, impact a wide range of traits and diseases, deepening AI's footprint (Donkin et al. 2016; Lawson, Cheverud, and Wolf 2013; Macias-Velasco et al. 2022; Potabattula et al. 2018; W. Zhang et al. 2019).

Delving into the realm of metabolic disorders, obesity stands out, highlighting the intricate interactions between genetic predispositions and environmental factors. Not only are lifestyle choices pivotal in the onset of obesity, but variations in genes can also predispose individuals to metabolic diseases, making the interplay between genetics and environment even more profound. As we probe deeper into AI's dynamics, a distinct pattern emerges. Low allelic imbalances might denote a consistent expression across alleles, ensuring stability in expression patterns. Conversely, high allelic imbalances present a spectrum of diverse expression levels, raising questions about the underlying genetic mechanisms and their implications for diseases such as obesity (Jung et al. 2023; Knowles et al. 2017; Kim-Hellmuth et al. 2017).

The genetic determinants of obesity, often located in noncoding genome regions, beckon in-depth exploration (Gallagher and Chen-Plotkin 2018). Many of these determinants have been linked to evolutionary adaptations of adipose tissue and hold implications for various metabolic diseases (Rosen and Spiegelman 2006; Barroso and McCarthy 2019). Techniques like colocalization offer promising insights, but discerning causality from coincidence remains a challenge due to intricate linkage disequilibrium (LD) patterns. Discernment is crucial, reminding us that colocalization doesn't automatically indicate causality and one should be cautious when interpreting these findings (Benaglio et al. 2023). This challenge is further magnified by the effects of LD in methods such as TWAS (Hauberg et al. 2017; Wainberg et al. 2019).

With adipose tissues taking center stage, the expanse of obesity's influence is far-reaching, mediating metabolic diseases and contributing to systemic inflammation (Rosen and Spiegelman, 2006; R. Liu and Nikolajczyk 2019). In conditions like Type 2 Diabetes Mellitus (T2DM), the precision of methods like colocalization unravels the roles of regulatory variants in disease susceptibility, highlighting the changes in immune cells in adipose tissues of obese individuals and their implications in T2DM (Torres et al. 2014; Mathis 2013; SantaCruz-Calvo et al. 2022).

Our study situates itself amidst this vast landscape, seeking to demystify the role of specific classes of allelic imbalances in obesity and associated metabolic diseases. Are genes with low allelic imbalances indicative of stable expression levels across individuals, and what might this mean for metabolic health? Alternatively, what insights can we glean from genes with pronounced allelic imbalances? These are the pivotal questions driving our research. Through a comprehensive integration of transcriptome, epigenome, GWAS, and polygenic risk score analyses, we aim to decipher the intricate interplay of genetic variation, epigenetic regulation, and metabolism. Our findings shed light on the genetic and epigenetic underpinnings of obesity and its associated susceptibility to other complex diseases.

RESULTS

Unraveling the Regulatory Networks of Obesity through Allelic Imbalance and Colocalization Analysis

Understanding the genetic underpinnings of obesity is pivotal for targeted therapeutic interventions and prevention strategies (Figure 1). However, this endeavor is complicated by the intricate interplay between genes, regulatory elements, and environmental factors. To better understand the regulatory landscape, we employed colocalization analyses. This integrative approach helps us to identify common regulatory elements and pathways, revealing not just how genes are regulated but also the mechanisms that underlie their dysregulation in pathological states like obesity. To explore the nuanced roles genes play in obesity and its related complications, it is critical to examine not just the genes themselves but also their varying expression levels from each allele. This phenomenon, known as allelic imbalance (AI), can offer unique insights into gene regulation, and it is particularly relevant in understanding how genes behave differently in specific tissues.

In this study, we use AI as a lens to navigate the complexities of physiological genomics in obesity. Allelic imbalance (AI) emerges as an invaluable parameter in this context, serving as a magnifying lens to scrutinize the nuances of gene expression. Allelic imbalance allows for the interrogation of gene expression at a single nucleotide level. In the context of obesity, this becomes indispensable for revealing differentially regulated genes, particularly within specialized tissues like Adipose tissues. AI not only enables the identification of genes displaying differential expression but also facilitates a deeper understanding of gene regulation. It thereby aids in delineating the intricate relationships between genetic variations, transcriptomic changes, and obesity phenotypes.

To provide an intricate understanding of obesity, we utilized an array of multi-omics data sets. The transcriptomic information is primarily sourced from the Genotype-Tissue Expression (GTEx) project, containing data for (number of samples) samples across (number of tissues) tissues. We also leveraged epigenomic data consisting of DNase I hypersensitive sites (DHSs), with 3.6 million annotated sites across 438 cell and tissue types. Further, phenomic association were derived from Polygenic Score (PGS) Catalog and GWAS catalogs, which included a total of all SNP-trait associations with genome-wide significance ($p \leq 5.0\times10\text{-}8$) from 6545 publications and millions of polygenic score-associated variants in traits of body measurement, lipid or lipoprotein measurement, and metabolic disorder.

Cis-expression quantitative trait loci (cis-eQTLs) offer a more focused and specific avenue for examining gene regulation. Their inclusion in the analysis further refines our understanding of how genetic variants can affect gene expression. In the setting of DHS, cis-eQTLs assume an even more pronounced role. Here, they serve as a dual layer of regulatory control, affecting not only gene expression but potentially influencing disease outcomes as well.

DNase I hypersensitive sites (DHS) serve as vital regulatory landmarks in the genomic landscape. Variants occurring within these sites have a heightened potential to affect gene regulation, either by altering transcription factor binding or by modifying other regulatory elements. Given their importance, these DHS-located variants become crucial for investigating the genetic regulation of obesity.

The analytical power of this study stems from its integrated approach, combining both AI and regulatory variant data in the form of cis-eQTLs within DHS. This allows us to sift through the complexities of the regulatory landscape and pinpoint specific genes and variants that contribute to obesity. More importantly, this integrated approach enables us to identify candidate regulatory variants, which include risk variants as

well as those highlighted in Genome-Wide Association Studies (GWAS). In doing so, we construct a comprehensive view of the myriad factors contributing to obesity, from the genetic to the epigenetic.

In this study, our investigation into the regulatory networks of obesity is heightened through a multifaceted approach that merges allelic imbalance studies with colocalization analysis of regulatory variants. By dissecting the regulatory landscape through these dual lenses, we gain a robust, nuanced view of the genetic mechanisms that underlie obesity.

Characterizing Allelic Imbalance Levels in Adipose Tissues and Their Pleiotropic Impact Across Tissues

We categorized allelic imbalance into five distinct levels: very low, low, moderate, high, and very high, based on the percentiles of allelic fold change (aFC) observed exclusively in adipose tissues. This stratification allowed us to precisely assess the differential impacts of allelic imbalances on gene expression and regulatory mechanisms (Figures 2A and 2B). A total of 2328 PCGs and 671 lncRNAs were identified and stratified into these groups based on their quantile aFC.

In adipose subcutaneous and adipose visceral tissues, which demonstrated the most striking allelic imbalances, we further dissected the genes and regulatory elements involved. In PCGs (Figure 2C), the average absolute value of slopes differed significantly across AI levels, ranging from 0.1566 ±0.0001 in the very low group to 0.4352 ±0.0002 in the very high group. A similar trend was observed in lncRNAs (Figure 2D).

Statistically significant differences were observed in the average number of cis-eQTLs across AI levels. Specifically, for PCGs, the mean number of significant cis-eQTLs varied significantly across the different AI groups (Figures 1B and 1C; $P < 0.05$). lncRNAs of very low AI and low AI groups were the only exceptions, where no significant differences in the number of cis-eQTLs were observed.

Although AI levels were defined based on adipose tissue gene expression, we extended the analysis to a tissue breadth study across 47 tissues. In PCGs (Figure 2E), genes with very low AI had significantly less tissue breadth than those with high (Mann-Whitney U test; p-value=0.032) and moderate (p=0.021) AI levels. For lncRNAs, genes with the highest AI level showed significantly larger numbers of shared genes than those at the lowest AI level (p=0.027; Figure 2F).

Pleiotropic effects of cis-eQTLs appear to be widespread and show similar levels of allelic imbalance across different tissues (Figure 2G and 2H). In PCGs, the average number of shared eGenes were 888.4 ± 77.0 (S.E.M) in the very low group and 1106.3 ± 65 in the high group, showing a significant difference ($P = 0.032$). In lncRNAs, the corresponding numbers were 259.3±19.4 and 310.5±11.4, also demonstrating significant difference (p=0.027).

The observed relationship between higher levels of allelic imbalance in adipose tissues and increased tissue sharing across multiple tissues offers intriguing insights into gene regulation and expression patterns. While allelic imbalance (AI) is captured through allele-specific expression (ASE) in adipose tissues in our study, it's crucial to clarify that a higher degree of AI does not equate to tissue-specific expression. In fact, our data reveal quite the opposite. Genes with high AI levels, as assessed in adipose tissues, were observed to have broader expression profiles across multiple tissues, pointing to a more pleiotropic role. This suggests that genes with higher allelic imbalance could be central players in cellular processes that are fundamental across multiple tissue types, thus making them candidates for investigations into system-wide or multi-

tissue disorders. Conversely, genes with low AI levels displayed more tissue-restricted expression, indicating higher tissue specificity, indicating a more specialized role in adipose tissue functionality.

Association Between Allelic Imbalance, Regulatory Element Density, and Genomic Architecture in Protein-Coding and Non-Coding Genes

Our comprehensive analysis elucidates a clear correlation between the level of allelic imbalance and the density of cis-eQTLs within DNase I hypersensitive sites (DHS), as well as the genomic architecture of the corresponding genes. In both protein-coding genes (PCGs) and long non-coding RNAs (lncRNAs), the number of cis-eQTLs and DHSs increased concomitantly with the level of allelic imbalance (Figure 3A, B3). Specifically, for PCGs and lncRNAs at very low levels of AI, an average of 33.1±0.8 and 38.5±2.4 cis-eQTLs were found within 28.3±0.7 and 31.4±1.6 DHSs, respectively. At very high levels of AI, these figures escalated to 70.1±3.9 and 77.8±7.1 cis-eQTLs within 50.0±1.7 and 55.8±3.4 DHSs for PCGs and lncRNAs, respectively (Figure 3C, 3D).

Remarkably, the difference in the number of cis-eQTLs and DHSs between AI groups outweighed the difference observed between PCGs and lncRNAs, underscoring the significance of allelic imbalance as a determinant of regulatory complexity. Our statistical analyses further indicated a linear relationship between the number of DHSs and the number of cis-eQTLs per gene, substantiating the notion that allelic imbalance may serve as a proxy for assessing the regulatory landscape surrounding a gene.

Intriguingly, while higher levels of AI correlated with greater regulatory complexity, they were also associated with significantly shorter transcript lengths. For instance, genes at lower AI levels had mean transcript lengths of 2741.3±39.1 and 1829.2±153.6 for PCGs and lncRNAs, respectively. In contrast, these lengths considerably reduced to 246.0±39.1 and 1303.4±37.0 for both PCGs and lncRNAs at higher levels of AI (Figure 3E, Mann-Whitney rank-sum test; P-value=6.7e-11 and P-value=1.5e-9). This inverse relationship between AI and transcript length could imply that genes with higher allelic imbalance are perhaps more finely regulated at the transcriptional level.

Diving deeper into the genomic architecture, we discovered differences in untranslated regions (UTRs) across AI levels. The 5' UTR was significantly longer in high AI groups (p-value=4.6e-6), while the 3' UTR was considerably longer in lower AI groups (p-value=2.5e-37). Interestingly, the coding sequence (CDS) lengths showed no such divergence (Figure 3F). Total genic regions in higher AI groups were also significantly larger than those in lower AI groups in PCGs (p-value=2.9e-25), but not in lncRNAs (Figure 3G).

Our analysis uncovers a compelling correlation between allelic imbalance levels and both the density of cis-eQTLs within DNase I hypersensitive sites (DHS) and the structural attributes of the corresponding genes. The data indicate a linear increase in the number of cis-eQTLs and DHSs corresponding to genes with higher allelic imbalance, in both protein-coding genes (PCGs) and long non-coding RNAs (lncRNAs).

The differential between groups of varying allelic imbalance far outweighs the differences observed between PCGs and lncRNAs, accentuating the paramount influence of allelic imbalance on regulatory complexity. Moreover, statistical tests corroborate a linear association between the number of DHSs and cis-eQTLs per gene, underlining the potential of allelic imbalance as an informative metric for scrutinizing a gene's regulatory landscape.

Further inspection of the genomic architecture highlighted disparities in untranslated regions (UTRs) according to the levels of allelic imbalance. In PCGs, the 5' UTR lengths were notably larger in higher AI groups, whereas the 3' UTRs were extended in the lower AI groups. These attributes were accompanied by no significant variation in coding sequence (CDS) lengths, indicating the selective nature of these differences.

Our observations suggest that allelic imbalance serves as more than a mere indicator of differential gene expression; it is a robust gauge of a gene's regulatory complexity and genomic structure. The convergence of higher allelic imbalance with increased regulatory involvement, paired with reduced transcript lengths, presents an intriguing avenue for future investigations. This inverse correlation suggests that genes with high allelic imbalance could be under more intricate transcriptional control.

These findings could be especially relevant for exploring how genes with differing levels of allelic imbalance participate in complex biological networks and may contribute to disease susceptibility or phenotypic diversity. The data highlights the sophisticated interplay between these genomic and epigenomic elements laying the foundation for further investigations into the polygenic architecture of obesity.

Functional Characterization of Allelic Imbalance in Adipose Tissues across Metabolic Conditions

Differential gene expression in adipose tissues is a pivotal factor in understanding metabolic health, particularly in conditions like obesity, insulin resistance, and Type 2 Diabetes (T2D). In our study, we utilized data from eight different studies comparing gene expression profiles in adipose tissues from lean individuals and those who are obese but vary metabolically—from metabolically healthy to those with normal glucose tolerance, insulin resistance, and T2D (Figure 4A, Supplementary Figure S3). Differentially Expressed Genes (DEGs) were initially identified in each study (Supplementary Figure S4) and then aggregated into overarching sets of up- and down-regulated genes. This comprehensive approach facilitated the exclusion of 47 genes that displayed contradictory patterns of expression across different studies. Overall, 2,725 up-regulated and 1,473 down-regulated genes were detected across various types of adipose tissues.

Interestingly, the up-regulation of genes was significantly more prevalent in moderate to very high levels of allelic imbalance (Figure 3B). This trend was predominantly observed in protein-coding genes and was remarkably absent in long non-coding RNAs (lncRNAs; Figure 3C). These findings point toward the potential functional ramifications of allelic imbalance in metabolic health, particularly its potential role in exacerbating conditions like obesity and insulin resistance.

Protein-protein interaction (PPI) networks further substantiated the functional characterization of these DEGs (Figure 4D). Here, a clear dichotomy was observed, wherein one cluster primarily comprised up-regulated genes, while another cluster contained mostly down-regulated genes. Intriguingly, DEGs in the category of very high AI demonstrated significantly lower closeness centrality in these PPI networks (Figure 4E), implying that these genes might play isolated yet potentially crucial roles within biological networks.

The functional significance of these DEGs was also evident when focusing on hub genes—genes that interact with multiple other genes in the network. As per Figures 4F and 4G, the no cis-eQTL group of up-regulated genes displayed significantly more hub genes. Additionally, down-regulated genes also showed a

higher propensity to be hub genes, hinting at their critical role in the cellular network and metabolic pathways.

Finally, the single-cell expression levels of these hub genes were investigated in human white adipose tissues. Notably, up-regulated hub genes like CD4, PTPRC, and ITGAM were predominantly expressed in immune cells such as dendritic cells, macrophages, and monocytes (Figure 5A-C). On the other hand, down-regulated hub genes like PPARG and APOE were more commonly expressed in adipose precursor cells and adipocytes (Figure 5D-H).

The preponderance of up-regulated genes in higher AI levels may indicate that allelic imbalance contributes to the severity of metabolic conditions. Notably, up-regulated genes in higher AI levels, such as CD4 and PTPRC, are known to regulate immune responses, pointing toward their role in systemic inflammation commonly associated with obesity and metabolic disorders like insulin resistance. The distinction in hub gene expression, especially the up-regulation of CD4, PTPRC, and ITGAM in adipose tissues with obesity, provides a crucial insight into the underlying metabolic health. CD4 and PTPRC are integral components of immune response and cellular communication, suggesting their heightened role in inflamed adipose tissue. Conversely, down-regulated hub genes like PPARG and APOE have established roles in lipid metabolism, indicating potential defects in adipocyte function and lipid storage in conditions of high AI.

The distinct clustering and centrality metrics observed in the PPI networks for DEGs across different AI levels suggest that genes with higher AI may serve as isolated but critical nodes within these networks, potentially functioning as key modulatory points in metabolic health. The identification of hub genes, especially those like PPARG and APOE, opens up avenues for therapeutic targeting, given their well-documented roles in metabolic pathways. Their down-regulation in high AI levels particularly emphasizes the need for targeted interventions in metabolic conditions like obesity and T2D.

Single-cell expression data adds another layer of complexity, particularly pointing toward cellular heterogeneity in adipose tissues. Down-regulated genes like PPARG and APOE primarily found in adipose precursor cells and adipocytes may indicate a disrupted adipogenic process, especially in high AI levels.

Differential Functional Enrichment and Ligand-Receptor Dynamics Across AI Levels

Our functional enrichment analysis reveals stark differences in biological processes influenced by both up- and down-regulated genes across various AI levels. Critically, the up-regulated genes in very high AI groups were statistically enriched (p-value=0.0024) for roles in 'response to stimulus.' Conversely, those in the very low AI groups were linked to 'Signal Transduction' (Figure 6A). This functional divergence was further nuanced for genes in the low AI, moderate, and no cis-eQTL groups, which were enriched in 'Immune System Process.'

Down-regulated genes presented another layer of complexity. Except for the moderate AI group, these genes prominently figured in 'Small Molecule Metabolic Process' (Figure 6B). In the very low and low AI groups, as well as the no cis-eQTL group, there was significant involvement in 'cytoplasmic translation' (p-value=0.0007).

Our ligand-receptor signaling analysis offers an intricate view of the directional flow of cellular information, pointing to key regulatory mechanisms. Utilizing an experimentally validated ligand-receptor signaling map of DEGs (Figure 7A; Supplementary Figure S5), we found compelling evidence for the role of ligand-receptor dynamics in the regulation of adipose tissue. Specifically, up-regulated genes in the no cis-eQTL,

high, and moderate AI groups were statistically enriched with a significantly larger number of ligands (p-values: 0.0024, 0.023, 0.039, respectively; Figure 7B). Moreover, genes within the no cis-eQTL and moderate AI groups showed a statistically larger number of receptors (p-values: 0.0126 and 0.0007, respectively; Figure 7C).

Examining the top 5% of hub ligands and receptors (Figures 7D and 7E) based on the number of directed target genes further enriches this narrative. FN1, classified under the moderate AI group, emerged as an up-regulated top hub ligand (Figure 7F). It demonstrated an expansive sphere of influence, interacting with 14 out of 29 DEGs as target receptors. Additionally, FN1 was up-regulated across multiple cell types in obese adipose tissue, including ASPC, LEC, SMC, and adipocytes.

On the contrary, WNT5A from the no cis-eQTL group featured as a down-regulated top hub ligand (Figure 7G). It engaged with a notably smaller pool of target receptors—only 5 out of 24 DEGs. Despite being down-regulated, WNT5A's expression in mesothelium cells remained neutral.

In the receptor landscape, GRM7 stands out as a top hub receptor within the very high AI group (Figure 7H). It interacted with a remarkable 23 out of 51 source ligands that were DEGs and exhibited up-regulation in ASPC cells in obese adipose tissue. Its associations with psychiatric disorders like major depressive disorder (Li et al. 2016), bipolar disorder (Kandaswamy et al. 2014), and attention deficit hyperactive disorder (Park et al. 2013) suggest multifaceted roles.

TSHR from the very high AI group also commands attention as a down-regulated top hub receptor (Figure 7I). It exhibited a more restricted interaction, coupling with only four out of 30 source ligands that were DEGs. Interestingly, TSHR presented divergent expression patterns, being down-regulated in adipocytes while up-regulated in ASPC cells, and its role in adipocyte differentiation and obesity (Lu et al. 2012) cannot be overlooked.

The richness of ligand-receptor signaling data complements and amplifies our broader observations by emphasizing the mechanistic basis underlying cellular regulation in adipose tissues. The marked diversity in the number of ligands and receptors across different AI levels, as exemplified by hub genes like FN1 and GRM7, hints at the presence of complex regulatory networks. These genes don't merely serve as nodes within these systems; they emerge as pivotal points of convergence that integrate multiple signaling pathways. Their expansive set of interactions and distinct expression patterns across varying AI groups and cell types implicate them as crucial components within the pathological landscape of adipose tissues. This complexity offers avenues for understanding how genetic predispositions may interact with environmental factors, such as diet or stress, thereby offering an integrated path to more personalized treatments for metabolic disorders and associated conditions.

Differential Transcript and UTR Lengths Across AI Levels Reveal Varied miRNA Targeting and Functional Implications

Our meticulous examination of transcript lengths in protein-coding genes (PCGs) unearthed intriguing disparities that could be functionally impactful. Specifically, we observed that up-regulated genes within lower AI groups possess significantly larger transcript lengths compared to those in higher AI groups (Figure 8A, p=0.010). Conversely, down-regulated genes in higher AI groups exhibited substantially larger transcript lengths than those in lower AI groups (Figure 8B, p=3.4e-9). These disparate findings not only underscore the complexity of gene regulation but also suggest that transcript length could be a

discriminating feature in the regulation of gene expression across different AI levels.

Beyond transcript lengths, DNase I hypersensitive sites (DHS) density—a marker for chromatin accessibility—also displayed AI-level-dependent variation. For both up- and down-regulated genes, DHS density was significantly higher in the higher AI group than in the lower AI group (p=0.003 and p=0.015, respectively). Moreover, up-regulated long non-coding RNAs (lncRNAs) manifested a higher DHS density in the highest AI group, compared to their counterparts in lower AI groups (Figure S6, p=0.007). These observations offer compelling evidence that chromatin structure and accessibility could serve as vital mechanisms underpinning gene regulation across different AI strata.

As we shifted our gaze to untranslated regions (UTRs), intriguing patterns further emerged. Down-regulated genes in higher AI groups exhibited markedly larger 5'UTR lengths than those in lower AI groups (p=1.4e-6), while 3'UTR lengths of both up- and down-regulated genes were significantly larger in very high AI groups compared to lower AI groups (Figure 8C and 8D, p=0.027 and p=4.4e-5 respectively). These variances in UTR lengths, particularly in 3'UTRs, could have far-reaching implications for miRNA targeting and post-transcriptional regulation of these genes.

To elucidate this point, we turned our attention to microRNAs (miRNAs), the post-transcriptional maestros. An analysis of experimentally validated miRNA targets of differentially expressed genes (DEGs) revealed 29 up-regulated miRNAs that predominantly target down-regulated PCGs (Figure 8E) and 35 down-regulated miRNAs mainly targeting up-regulated PCGs (Figure 8F). The biological processes enriched in the targets of up-regulated miRNAs notably included "Small Molecule Metabolic Process," whereas those enriched in down-regulated miRNAs prominently featured "Regulation of Cellular Processes," including phagocytosis, myotube differentiation, and intrinsic apoptotic signaling pathway in response to DNA damage (Figure 8G).

The distinct patterns we have observed in transcript and UTR lengths across various AI levels are not mere coincidences but could be functionally instrumental. They indicate the potential roles of these structural features in gene regulation, from chromatin accessibility to post-transcriptional regulation by miRNAs. These intricacies in gene regulation could be the scaffolding upon which differential gene expression is built across AI levels, and they offer tantalizing avenues for future exploration.

Genetic Risk Susceptibility Profiles of Differentially Expressed Genes Across Allelic Imbalance Levels

A critical dimension of understanding obesity-related gene expression involves linking allelic imbalance (AI) levels with established genetic risk susceptibilities. To address this, we examined DNase I hypersensitive sites (DHSs) that colocalize with Single Nucleotide Polymorphisms (SNPs) derived from genome-wide association studies (GWAS) for both protein-coding genes (PCGs) and long non-coding RNAs (lncRNAs).

We observed a significant proportion of non-redundant GWAS-SNPs colocalizing with up-regulated genes compared to down-regulated ones, specifically in high AI levels for PCGs (p-value=0.0002; Figure 9A) and very high AI levels for lncRNAs (p-value=5.7e-20; Figure 9B). Intriguingly, down-regulated lncRNAs in the moderate AI level had a significantly larger number of such GWAS-associated SNPs compared to their up-regulated counterparts (p-value=0.0065).

Expanding on these findings, we quantified DHSs directly associated with GWAS data and discovered that

the number of such DHSs was significantly higher in up-regulated PCGs across AI levels (p-value < 0.002). Similarly, up-regulated lncRNAs in both the low and highest AI levels bore more GWAS-associated DHSs (p-value < 0.0012). Conversely, the down-regulated genes in this context had significantly fewer such DHSs (p-value=2.3e-5).

The Standardized Mean Differences (SMDs) of effect weights indicated that variants located within DHSs in high AI groups showed greater impact on complex traits. These included Body Mass Index (BMI), Low-Density Lipoprotein (LDL), High-Density Lipoprotein (HDL), total cholesterol levels, triglycerides, coronary artery disease, recent poor appetite or overeating, and major depressive disorder (Supplementary Figure S7)—all reaching statistical significance (adjusted p-value < 0.05; Figure 10).

For lncRNAs, the SMDs revealed that up-regulated genes in very high AI levels and down-regulated genes in low and very low AI levels were differentially affected by risk variants associated with depressive disorders. Down-regulated lncRNAs in the very low AI level were significantly associated with risks for being overweight and having coronary artery disease.

These results corroborate the multifaceted nature of AI levels and demonstrate their integral role in dictating the genetic risk landscape, especially in the context of obesity. Moreover, the colocalization of GWAS-associated SNPs and DHSs offers intriguing insights into how chromatin accessibility may modulate these complex relationships, thereby providing avenues for targeted interventions in obesity and its associated comorbidities.

DISCUSSION

In a world grappling with rising obesity rates, understanding the genetic factors behind this condition becomes ever more crucial. Our study sheds light on such genetic and epigenetic intricacies, particularly emphasizing the role of allelic imbalance in gene expression patterns associated with complex traits. By delving deep into both coding and non-coding genes, we've unveiled new insights about transcript length variations, miRNA interactions, and genetic predispositions. These findings not only enhance our comprehension of obesity's genetic basis but also point to potential avenues for future research and therapeutic strategies.

Allelic Imbalance and Gene Expression

Allelic imbalance (AI) represents a simple spectrum of the differential expression of genes based on their inherited versions or alleles. AI's concept has long been acknowledged in the context of various diseases, such as cancer, its role in obesity genetics has remained largely untapped. While AI's exploration in the sphere of obesity genetics has been sporadic, our research delves deeper, positioning AI as a cornerstone for decoding the intricate patterns of gene expression in obesity. This fresh perspective not only complements but also amplifies insights from past studies. Our study identifies an intricate pattern of interplay: higher AI levels correlate with up-regulated short genes accompanied by a larger number of DHSs, whereas their lower counterparts align with down-regulated long genes, but with fewer DHSs.

The broader ramifications of these findings are manifold. Understanding this nuanced interplay between AI, gene length, and DHS abundance can open avenues for more individualized therapeutic interventions. By deciphering how AI and associated factors shape gene expression in obesity, medical practitioners can craft therapies tailored to a patient's specific genetic blueprint. This precision might manifest in targeted treatments using pharmacological agents or advanced gene editing technologies. Furthermore, these AI profiles, combined with gene length and DHS data, can act as potential biomarkers. Such markers can be instrumental in identifying individuals predisposed to obesity-linked complications, thereby bolstering preventive measures to be both timely and impactful.

Functional Enrichment and Ligand-Receptor Interactions in the Context of Allelic Imbalance

Past studies on functional enrichment analyses often examined gene functionality in isolation, missing connections to genetic nuances like allelic imbalance[]. By merging insights from allelic imbalance, gene expression data, functional enrichment, and ligand-receptor interactions, our research crafts an integrative perspective not commonly seen in prior works[]. This panoramic view enriches our understanding of the molecular intricacies involved in obesity.

Our findings spotlight crucial ligand-receptor interactions and functionally enriched classes that could be pivotal in unraveling obesity's complex molecular architecture. The prospect of these interactions as potential druggable targets is an enticing one, offering innovative avenues for obesity management[]. Moreover, the functional enrichment analysis accentuates the value of developing multi-target therapies. It offers a blueprint of interlinked pathways and molecular players, suggesting a comprehensive strategy for treatment.

Diverging from traditional differential gene expression explorations, our study delves deep into the functional richness and interactions among the DEGs[]. Of particular note are our findings on the enrichment of cellular signaling pathways among down-regulated genes and immune response mechanisms

in up-regulated genes, reinforcing their significance in disease pathways[].

Shedding Light on Transcript Lengths, UTRs, and non-coding RNAs Targets in Obesity

Historically, the importance of transcript lengths and UTRs has been a focal point in cancer biology[], but their relevance in metabolic diseases like obesity hasn't been adequately emphasized. Our findings not only underscore the importance of these features in obesity but also highlight their connection with allelic imbalance. This offers a holistic view of obesity's transcriptomic framework that was previously fragmented[].

We chart novel territory by establishing the interplay between transcript length, UTR dimensions, and differential gene expression across diverse AI levels in obesity. In our analysis, up-regulated genes in the lower AI bracket boasted extended transcripts, while their down-regulated counterparts in the higher AI realm showcased larger transcripts. Further, we observed notable variations in both 3'-UTR and 5'-UTR lengths—a connection to allelic imbalance not previously drawn in obesity's context[].

Such discrepancies in transcript length can substantially influence post-transcriptional regulation avenues, including the dynamics of miRNA binding. Our insights into UTR alterations pave the way for probing the potential role differential miRNA targeting plays in obesity's onset and progression. We found that up-regulated miRNAs specifically target down-regulated genes engaged in the "Small Molecule Metabolic Process". These discoveries could usher in a new era of RNA-centric therapeutic strategies that cater to individual AI profiles[].

Associating Allelic Imbalance with Genetic Risk Susceptibility Profiles

Treading on previously unexplored territory, our investigation bridges allelic imbalance (AI) with key genetic risk markers such as SNPs and DNase I hypersensitive sites (DHSs). Across diverse AI thresholds, our findings highlighted pronounced variations in the distribution of non-redundant GWAS-SNPs and DHSs, more so among up-regulated genes—encompassing both protein-coding and long non-coding RNAs.

While recent strides have been made in fusing GWAS outcomes with other genetic indicators, our study distinguishes itself by integrating these findings with AI. This innovative approach not only deepens our grasp of obesity's genetic risk landscape but also paves the way for similar explorations into other intricate traits.

By discerning these genetic risk susceptibility sites in AI's context, we're introducing an added dimension of complexity. However, this also unearths new opportunities for tailoring medical interventions. Further experiment is also needed to validate genetic mechanism of how regulatory variants of lncRNAs in obesity are associated with depression disorder. Colocalization of these genetic risk determinants promises to fine-tune the accuracy of existing genetic risk models. Such refined models could substantially guide clinical choices, particularly concerning prevention and management strategies for conditions tied to obesity.

Delving into Gene-Environment Interactions in Obesity

Obesity's multifactorial nature stems from intricate gene-environment interactions. Decades of research affirm the role of environmental factors in obesity, from parental origins influencing disease risk[] to the impact of importing dietary habits from one culture to another[]. Our explorations offer distinct insights, pinpointing instances where AI levels and GWAS-associated SNPs may converge with external factors, including diet and lifestyle.

The potential intersections we identified accentuate the necessity for cross-disciplinary research to robustly address the obesity crisis. A deeper comprehension of these gene-environment synergies can spawn targeted interventions that embrace both genetic inclinations and environmental exposures[].

Although environmental aspects in obesity have been the subject of extensive study, our work contributes uniquely by detecting genetic profiles possibly rendering individuals more vulnerable to environmental provocations.

Limitations

Our study, while providing a fresh perspective on the nexus between allelic imbalance and gene expression in obesity, isn't without its constraints:

The breadth of our investigation, though expansive on the genetic facets of obesity, falls short in robustly addressing the intertwined environmental influences central to this condition. The complexity of obesity arises not only from genetic underpinnings but also from myriad environmental factors such as diet, lifestyle, psychological intricacies, and societal health determinants. A holistic understanding necessitates a closer look at these elements, complementing the genetic findings[].

Drawing from GWAS and PGS datasets, we unearthed invaluable insights into risk-associated SNPs. However, the nuances of GWAS can sometimes tether its findings to particular population specifics, potentially sidelining the richness of genetic diversity inherent in mixed or heterogeneous populations[].

The treasure trove of high-throughput genomic data, while offering a depth of knowledge, is not without pitfalls. The challenges of data sparsity and embedded noise present hurdles that could sway the fidelity of our results[].

Our focus, honed in on specific tissues and cell types, may overlook the broader tapestry of tissues instrumental in obesity's genesis and progression. This selectiveness inadvertently omits the nuanced dynamics of cell type heterogeneity inherent within these tissues. Such an oversight might lead to potential blind spots, obfuscating essential cell-specific influences on obesity[].

Lastly, the computational prowess of our research, though illuminating, isn't the conclusive arbiter of truth. The associations and non-coding genes spotlighted demand rigorous experimental validation to underscore their true functional relevance in the obesity landscape.

Our acknowledgment of these limitations underscores our commitment to clarity and transparency. It's a clarion call for future endeavors in this realm to bridge these gaps, ensuring a more integrated, comprehensive model of obesity that seamlessly melds genetic insights with environmental realities.

Concluding remarks

In this study, we have charted new territory, offering an unparalleled view into the intricate interplay between allelic imbalance, gene expression, and the predispositions related to obesity. Moreover, our findings draw parallels with other diet-influenced complex traits, such as type 2 diabetes and cardiovascular diseases.

Our integration of genomic and epigenomic data has unearthed previously unexplored correlations between gene regulatory mechanisms and obesity-specific risks. These correlations bridge crucial knowledge gaps,

helping clarify the multifaceted nature of obesity's etiology. It's essential, however, to recognize the study's limitations, including challenges tied to cell-type heterogeneity, the applicability of trans-ancestry polygenic risk, and external environmental influences.

Looking forward, the insights gained not only set a robust foundation for subsequent research endeavors but also spotlight potential avenues for improving diagnostic and therapeutic methodologies. Embracing this multidimensional approach, we envisage a future where precise diagnostics, proactive alert systems, and bespoke interventions become the standard for managing obesity and related complexities. It's our hope that this study acts as a beacon, galvanizing interdisciplinary alliances aimed at enhancing individualized care for those at heightened risk of obesity-associated challenges.

ACKNOWLEDGEMENTS

This work is supported by the Medical Research Center Program (2022R1A5A2018865) through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Science and ICT (MSIT).

AUTHOR CONTRIBUTIONS

S.M designed the study with S.-Y.P. S.M run the experiments. All authors contributed to the development of the project. All authors participated in construction of the manuscript.

COMPETING INTERESTS

The authors declare no conflict of interest.

METHODS

Datasets Collection

Patient Samples and Gene Expression Data

We sourced RNA-sequencing data of human adipose tissues from both obese and lean individuals from the Gene Expression Omnibus (GEO). Specific GEO study IDs included are GSE179455, GSE165932, GSE55008, GSE156906, GSE205668, GSE110729, GSE141432, and GSE162653. We identified differentially expressed genes (DEGs) across various adipose tissue types, such as epicardial, omental, and subcutaneous tissues (Supplementary Table 1). Groups were classified based on obesity with/without insulin resistance, normal glucose tolerance (NGT), Type 2 Diabetes (T2D), and metabolically unhealthy conditions compared to lean metabolically healthy controls. A significance level cutoff of 0.05 was employed following Benjamini & Hochberg adjustment for P-values, with a fold change threshold of 1.5 (DESeq2, Love et al., 2014). DEGs that showed contradictory regulation across studies were excluded from further analysis.

cis-eQTL analysis

Cis-expression Quantitative Trait Loci (eQTL) data was acquired from the GTEx project V8, which comprises RNA-seq and whole-genome sequencing data from 49 tissues of 838 healthy individuals (dbGaP: phs000424.v8.p2). We extracted cis-eQTL variants below the predefined significance threshold and calculated the slope of significant eGenes using a q-value < 0.05, derived from beta-approximated permutation p-values.

Allelic Imbalance Classification and Tissue Breadth

cis-eQTL effect sizes were quantified as the allelic fold change in gene expression between reference and alternative alleles. We categorized the magnitude of allelic imbalance into five classes: 'Very Low', 'Low', 'Moderate', 'High', and 'Very High', based on {0, 20, …, 80, 100} percentile distributions of absolute log2-transformed allelic fold change (aFC) values in adipose tissues. For genes with multiple adipose tissues estimates, the largest absolute log2(aFC) was selected. Tissue breadth was assessed by the number of cis-eQTL target genes across the 49 GTEx tissue types.

Regulatory DNA Mapping and Genetic Association

In our approach, we employed a direct colocalization strategy to link SNPs identified in GWAS studies with regulatory DNA regions characterized by DNase I hypersensitive sites (DHSs). This analysis was designed to pinpoint potential regulatory SNPs that may contribute to complex traits. By overlaying GWAS SNPs onto the DHS Index data, we identified SNPs that fall within regulatory regions, potentially influencing gene expression. Additionally, we integrated this data with cis-eQTLs associated with genes present in these regulatory DNA regions. This colocalization enables the direct identification of candidate target genes that are not only implicated in trait variation but are also likely to be directly regulated by these SNPs. This streamlined approach underscores the potential regulatory mechanisms by which genetic variation can influence complex traits and provides a clear path to deciphering gene-trait association

Protein-Protein Interaction (PPI) Network Analysis

PPI networks were generated using high-confidence interactions from the STRING database, considering only interactions with a confidence score above 0.7. Network construction and analysis were performed using Cytoscape, integrating differentially expressed genes (DEGs) with varying levels of allelic imbalance.

Topologically significant nodes, characterized by high connectivity, were identified by using Network Analyzer. These key hub genes were then used to reconstruct the network, highlighting potential pathways and biological processes that might be disrupted in obesity and associated metabolic conditions.

Single cell expression heterogeneity analysis

Single-nucleus RNA-seq data of human white adipose tissue from individuals with obesity were analyzed to assess the cellular distribution of DEGs associated with allelic imbalance. Expression profiles were visualized using 2D Uniform Manifold Approximation and Projection (UMAP) plots.

Ligand-Receptor Interaction mapping

We identified potential ligand-receptor interactions from the upregulated and downregulated DEGs using the CellTalkDB database. Only those pairs found within both datasets were considered for constructing the interaction map in Cytoscape.

Functional Enrichment Analysis

Enrichment analyses for biological pathways were conducted using g:Profiler with DEGs across allelic imbalance levels. Biological processes and pathways with a false discovery rate (FDR) below 0.05 were reported as significantly enriched.

miRNA Target Prediction

For identifying differentially expressed miRNA targets, we utilized miRTarBase to extract experimentally validated miRNA-target interactions, cross-referencing them with our list of DEGs.

Genetic Risk Susceptibility difference

We integrated polygenic risk scores (PRS) for various metabolic and obesity-related traits obtained from the PGS Catalog. Each PRS was used to estimate individual genetic susceptibility. Statistical analysis was conducted to associate allelic imbalance with genetic risk factors, employing standardized mean difference techniques for comparing case and control mean differences, accounting for within-group variance. This analysis provided insight into the genetic predisposition of individuals with obesity, potentially linking differential allelic expression to the polygenic nature of obesity and its metabolic consequences.

Statistical analysis

All statistical tests employed are specified in corresponding text and figure legends. For non-parametric distributions, the Mann-Whitney U test was utilized. The hypergeometric test was applied for evaluating the overlap between gene sets. Analyses were conducted using appropriate Python and R packages, with multiple hypothesis correction applied using the Benjamini-Hochberg procedure where appropriate.

Gene annotation

Gene features such as biotype, UTR lengths, and transcript lengths were extracted from the ENSEMBL BioMart database. Protein-coding genes were defined based on the presence of an associated protein product, excluding transcripts classified as retained introns or those undergoing nonsense-mediated decay. Long non-coding RNAs (lncRNAs) were identified based on the biotype provided in the ENSEMBL annotations.

Data Availability

The datasets and resources used in our study are publicly accessible via the provided URLs. The GTEx V8 cis-eQTL and eGene data are available at the GTEx portal. DHS Index data can be accessed through the dedicated repository. CellTalkDB, miRTarBase, the GWAS Catalog, the PGS Catalog, and ENSEMBL BioMart provide extensive databases for genetic, epigenetic, and gene annotation resources. All data used for analyses are available within these resources.

GTEx Analysis V8 cis-eQTL and eGene: https://storage.cloud.google.com/adult-gtex/bulk-qtl/v8/single-tissue-cis-qtl/GTEx_Analysis_v8_eQTL.tar

DHS Index: https://www.meuleman.org/DHS_Index_and_Vocabulary_hg38_WM20190703.txt.gz

CellTalkDB: http://tcm.zju.edu.cn/celltalkdb

miRTarBase: https://miRTarBase.cuhk.edu.cn/

A single-cell atlas of human white adipose tissue: https://singlecell.broadinstitute.org/single_cell/study/SCP1376/a-single-cell-atlas-of-human-and-mouse-white-adipose-tissue?cluster=Human%20WAT&spatialGroups=--&annotation=cluster--group--study&subsample=all#study-visualize

Imprinted gene database: https://www.geneimprint.com/site/genes-by-species.Homo+sapiens

GWAS Catalog: https://www.ebi.ac.uk/gwas/api/search/downloads/alternative

PRS Catalog: https://ftp.ebi.ac.uk/pub/databases/spot/pgs/scores/

Ensembl BioMart: https://useast.ensembl.org/biomart/martview/

Code Availability

The complete set of scripts and codes used for data analysis and visualization in this study are maintained and documented on GitHub at: https://github.com/moon-s/allelic-imblance

Figure Legends

Figure 1. Study Overview and Main Findings

This schematic illustrates the integrated analysis of allelic imbalance utilizing gene expression data from adipose tissues in obesity, combined with cis-eQTL, regulatory DNA, and GWAS variants. It highlights key outcomes, revealing potential mechanistic insights into gene regulation in obesity.

Figure 2. Stratified Analysis of Pleiotropic Effects in cis-eQTLs

A,B. The mean absolute allelic fold change (aFC) for protein-coding genes (PCGs) and long non-coding RNAs (lncRNAs), respectively, alongside the mean number of cis-eQTLs associated with adipose tissues per gene across different levels of allelic imbalance. **C,D**. the distribution of effect sizes in adipose tissues for PCGs and lncRNAs. **E,F**. the tissue-specific expression breadth of these genes across 49 tissues. **G, H**. the average number of shared genes with significant effect weight between tissues at each level of allelic imbalance for both gene types.

Figure 3. cis-eQTL Distribution within DNase I Hypersensitive Sites (DHS) and Sequence Feature by Allelic Imbalance Levels

A,B. The count of cis-eQTLs located within DHS for both PCGs and lncRNAs, respectively, across varying levels of allelic imbalance (AI). **C,D**. the average number of cis-eQTLs and DHSs linked to cis-eQTLs per gene. The average transcript length (**E**), feature length (**F**), and the span of genic regions (**G**) within each AI category.

Figure 4. Differential Gene Expression in Adipose Tissue Associated with Obesity

A. Heatmap of DEGs across adipose tissue types. **B, C**. the number of differentially expressed PCGs and lncRNAs across AI levels. **D**. a protein-protein interaction network for the differentially expressed PCGs in obese adipose tissues. **E**. Closeness centrality in genes with very high AI levels. **F, G**. list the top 10 hub genes that are up- and down-regulated in obesity, respectively.

Figure 5. Single-Cell Expression Patterns of Hub Genes in High BMI Human adipose tissue

We showed single-cell expression patterns of top differentially expressed (DE) hub genes. Up-regulated genes (**A-C**) and down-regulated genes (**D-H**) are shown with cell type-specific expression heterogeneity, highlighted in red and blue, respectively. Labeled cell types depicts that the gene is differentially expressed in high BMI from the single cell analysis.

Figure 6. Functional Enrichment across Allelic Imbalance Levels

A. genes with high AI levels and up-regulation are significantly enriched for response to stimulus and signaling functions. **B**. down-regulated genes with low AI levels are enriched for metabolic processes and the generation of precursor metabolites and energy.

Figure 7. Ligand-Receptor Interaction Pathways and Signaling in Different AI Levels

A. An overview of ligand-receptor interactions among DEGs at various AI levels. **B,C**. Number of differentially expressed ligands and receptors, respectively. **D, E**. The top 5% of hub ligands and receptors, respectively. **F-I**. single-cell expression heterogeneity for the top hub ligands and receptors, such as FN1, WNT5A, GRM7, and TSHR, with directed edges indicating ligand-to-receptor interactions.

Figure 8. Transcript Length Dynamics Relative to Allelic Imbalance

A.The relationship between transcript length and number of DHSs in each AI level is depicted, with up-regulated PCGs having shorter transcripts and higher AI. **B**. down-regulated PCGs having longer transcripts and higher AI. **C, D**. the differences in 5'-UTR and 3'-UTR lengths across AI levels. **E,F** the miRNA target genes and their enrichment for both down- and up-regulated miRNAs. **G**. Functional enrichment of target genes for up- and down-regulated miRNAs.

Figure 9. GWAS-SNPs within DHS Regions Across Allelic Imbalance Levels

GWAS-associated SNPs present within DHS regions for both PCGs (**A**) and lncRNAs (**B**) categorized by AI level.

Figure 10. Polygenic Risk Scores Linked to DHS in Differentially Expressed Geness

Differential mean effect weights of polygenic risk variants are shown for the first (1Q) and fourth quartiles (4Q) of effect weight distribution across complex traits, for both up-regulated (**A**) and down-regulated (**B**) PCGs, as well as lncRNAs (**C** and **D**).

Supplementary tables

Table S1. RNA-Seq of adipose tissues from individuals with obesity and metabolic disease states

GEO ID	Control group	Test group	Tissue/cell
GSE179455	Lean	Obese/overweight	Epicardial adipose tissue
GSE165932	Lean (IS)	Obese (IR)	Omental adipose tissue
GSE55008	Lean	Obese	Omental adipose tissue
GSE156906	Lean	Obese (MUH)	Subcutaneous adipose tissue
GSE156906	Lean	Obese (MH)	Subcutaneous adipose tissue
GSE205668	Normal weight	Obese	Subcutaneous adipose tissue
GSE110729	Lean	Obese	Subcutaneous white adipose tissue
GSE141432	Lean	Obese (NGT)	White adipose tissue
GSE141432	Lean	Obese (T2D)	White adipose tissue
GSE162653	Lean	Obese	White adipose tissue

*IS: insulin sensitive, IR: insulin resistance, MUH: metabolically unhealthy, MH: metabolically healthy, NGT, Normal glucose tolerance, T2D: type 2 diabetes

Supplementary figures

Figure S1. Tissue-Specific Allelic Fold Change in eGenes

This heat map displays the allelic fold change (aFC) of genes with significant cis-eQTLs across multiple tissue types. The x-axis represents individual genes, and the y-axis represents tissue types. A distinct cluster of eGenes in adipose tissues is highlighted in orange, suggesting tissue-specific regulatory mechanisms.

Figure S2. Allelic Fold Change Distribution in Adipose Tissues

This figure illustrates the distribution of the absolute values of allelic fold change (aFC) in adipose tissues. The visualization facilitates the comparison of the extent of allelic imbalance across the genes examined.

Figure S3. Differential Expression in Adipose Tissues by Obesity Status

Gene expression differences between adipose tissues from obese individuals and those with normal weight or lean mass are shown. Genes were identified as significantly up-regulated (red) or down-regulated (blue) based on an adjusted p-value threshold of <0.05 and a fold change cutoff of 1.5.

Figure S4. Adipose Tissue DEGs and Associated Protein-Protein Interaction Networks

A and B. depict protein-protein interaction (PPI) networks for up- and down-regulated genes associated with cis-eQTLs, respectively. **C and D** show the same for genes without cis-eQTL associations. Each network's top hub gene is encased in a yellow box, indicating central nodes within the PPI networks.

Figure S5. Ligand-Receptor Interaction Map for Adipose Tissue DEGs

The map presents the ligand-receptor interactions among differentially expressed genes (DEGs) in adipose tissues from individuals with obesity. This supplementary figure complements the detailed signaling pathways inferred from DEGs presented in the Figure 7A in main text.

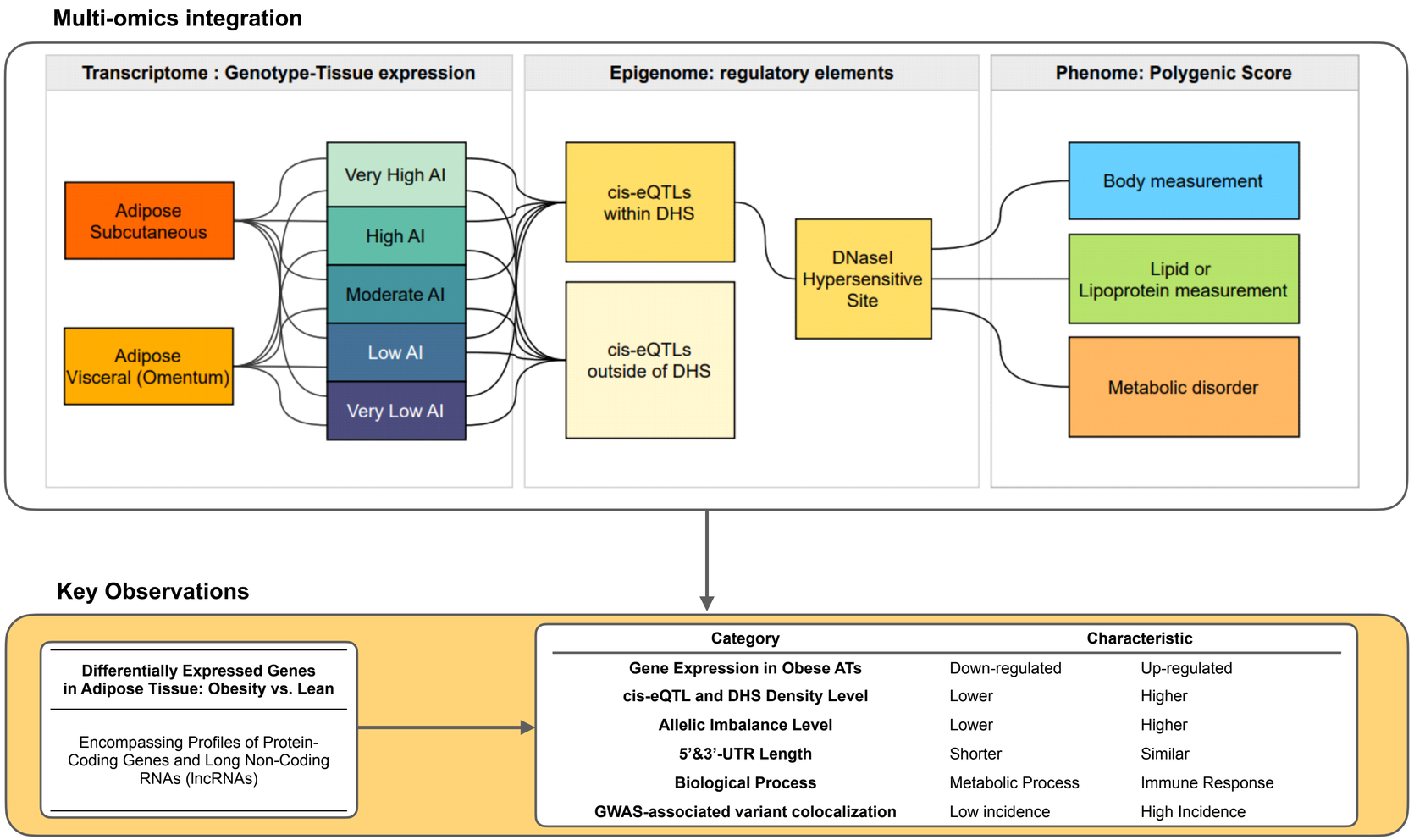
Figure S6. Transcript Length Dynamics Relative to Allelic Imbalance

A.The relationship between transcript length and number of DHSs in each AI level is depicted, with up-regulated lncRNAs having shorter transcripts and higher AI. **B**. down-regulated lncRNA having longer transcripts and higher AI.

Figure S7. Polygenic Score Distribution Across Studies

Each plot shows the distribution of polygenic scores (PGS) with the study identifier from the PGS Catalog as the title. This allows for a cross-study comparison of polygenic score distributions related to different traits or conditions.

Figure 1



Category	Characteristic	
Gene Expression in Obese ATs	Down-regulated	Up-regulated
cis-eQTL and DHS Density Level	Lower	Higher
Allelic Imbalance Level	Lower	Higher
5'&3'-UTR Length	Shorter	Similar
Biological Process	Metabolic Process	Immune Response
GWAS-associated variant colocalization	Low incidence	High Incidence

Figure 2

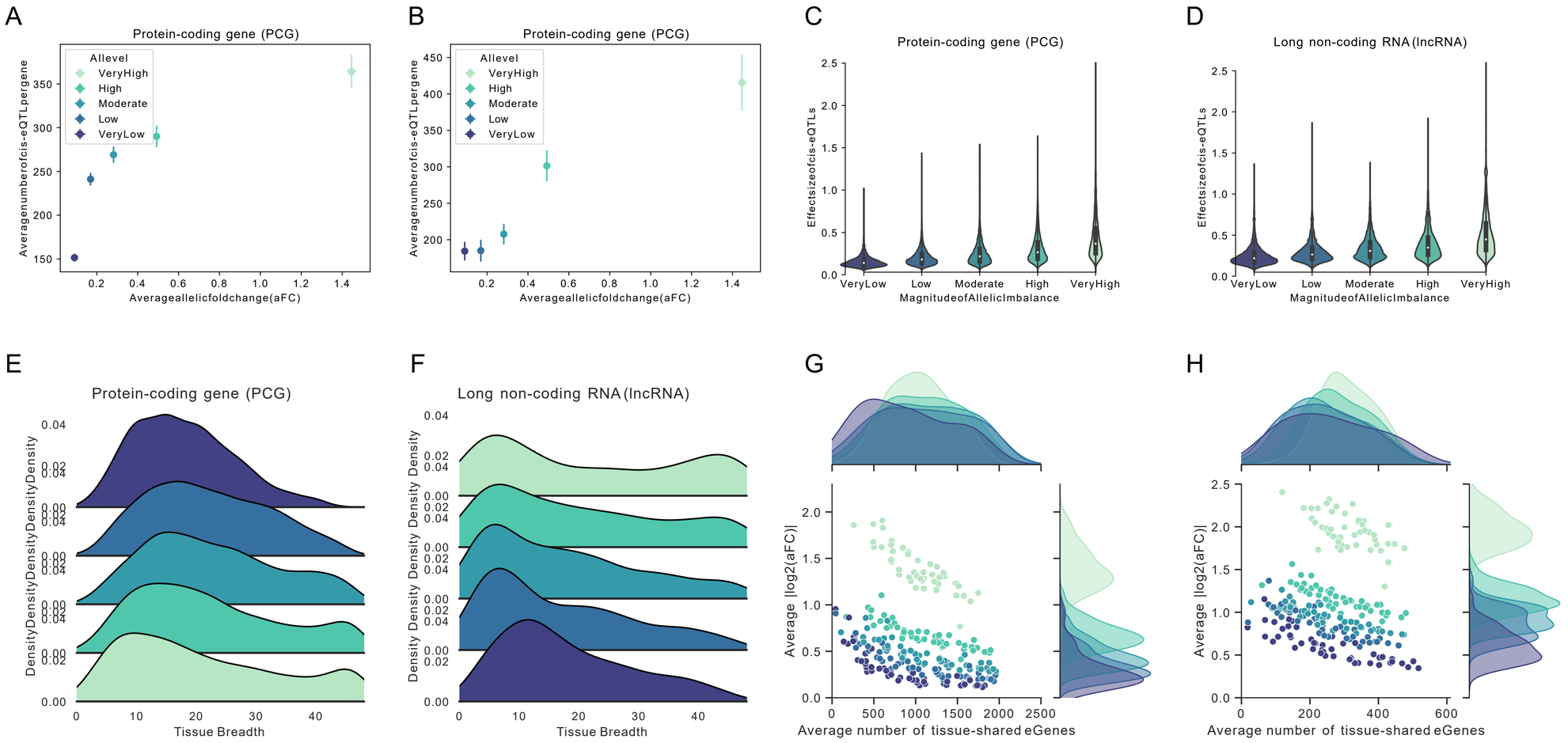

A
Protein-coding gene (PCG)
Allevel
VeryHigh
High
Moderate
Low
VeryLow
Averagenumberofcis-eQTLpergene
Averageallelicfoldchange(aFC)
B
Protein-coding gene (PCG)
Averagenumberofcis-eQTLpergene
Averageallelicfoldchange(aFC)
C
Protein-coding gene (PCG)
Effectsizeofcis-eQTLs
VeryLow
Low
Moderate
High
VeryHigh
MagnitudeofAllelicImbalance
D
Long non-coding RNA (lncRNA)
Effectsizeofcis-eQTLs
VeryLow
Low
Moderate
High
VeryHigh
MagnitudeofAllelicImbalance
E
Protein-coding gene (PCG)
Density
Tissue Breadth
F
Long non-coding RNA (lncRNA)
Density
Tissue Breadth
G
Average |log2(aFC)|
Average number of tissue-shared eGenes
H
Average |log2(aFC)|
Average number of tissue-shared eGenes

Figure 3

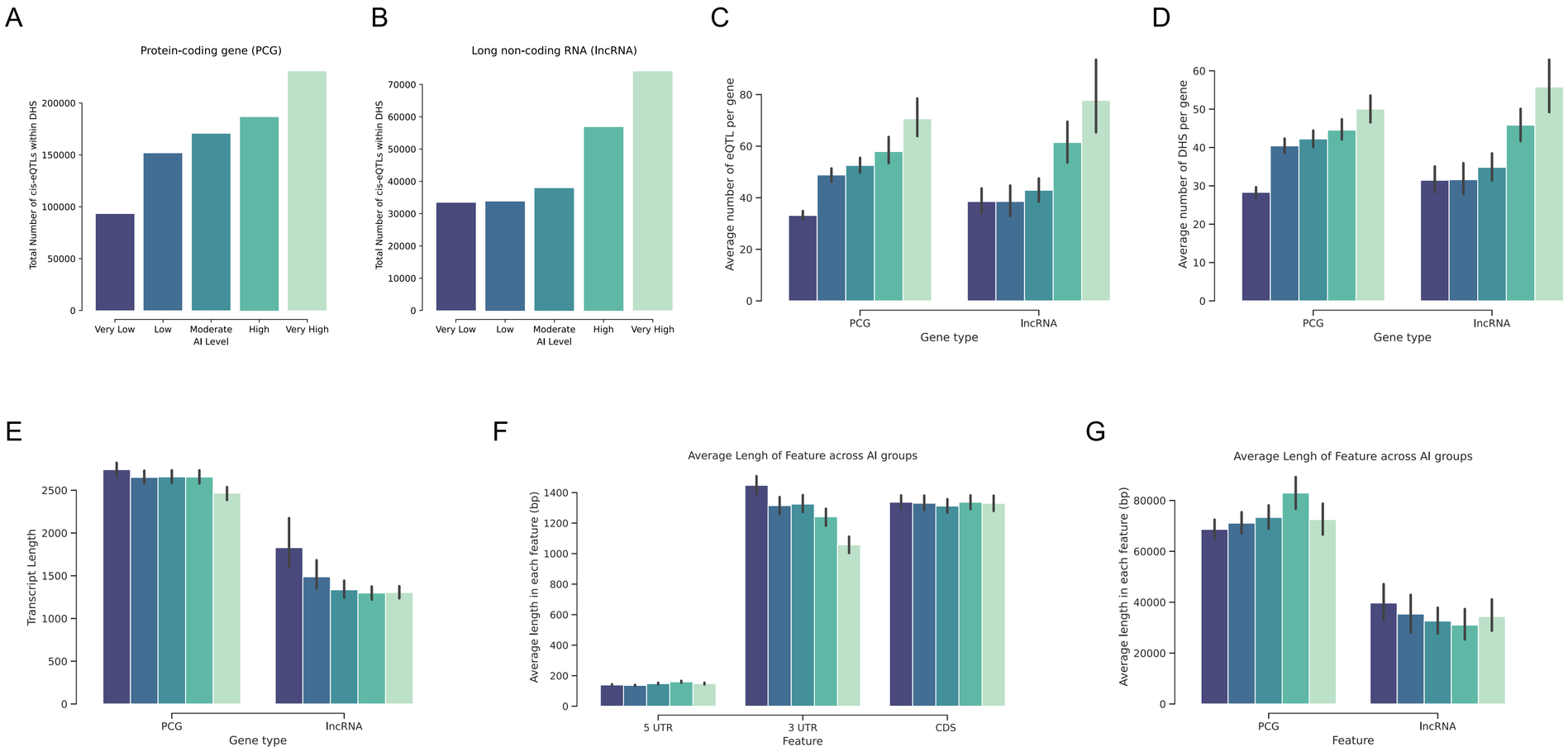

A
Protein-coding gene (PCG)
Total Number of cis-eQTLs within DHS
200000
150000
100000
50000
0
Very Low
Low
Moderate
High
Very High
AI Level
B
Long non-coding RNA (lncRNA)
Total Number of cis-eQTLs within DHS
70000
60000
50000
40000
30000
20000
10000
0
Very Low
Low
Moderate
High
Very High
AI Level
C
Average number of eQTL per gene
80
60
40
20
0
PCG
lncRNA
Gene type
D
Average number of DHS per gene
60
50
40
30
20
10
0
PCG
lncRNA
Gene type
E
Transcript Length
2500
2000
1500
1000
500
0
PCG
lncRNA
Gene type
F
Average Lengh of Feature across AI groups
Average length in each feature (bp)
1400
1200
1000
800
600
400
200
0
5 UTR
3 UTR
CDS
Feature
G
Average Lengh of Feature across AI groups
Average length in each feature (bp)
80000
60000
40000
20000
0
PCG
lncRNA
Feature

Figure 4

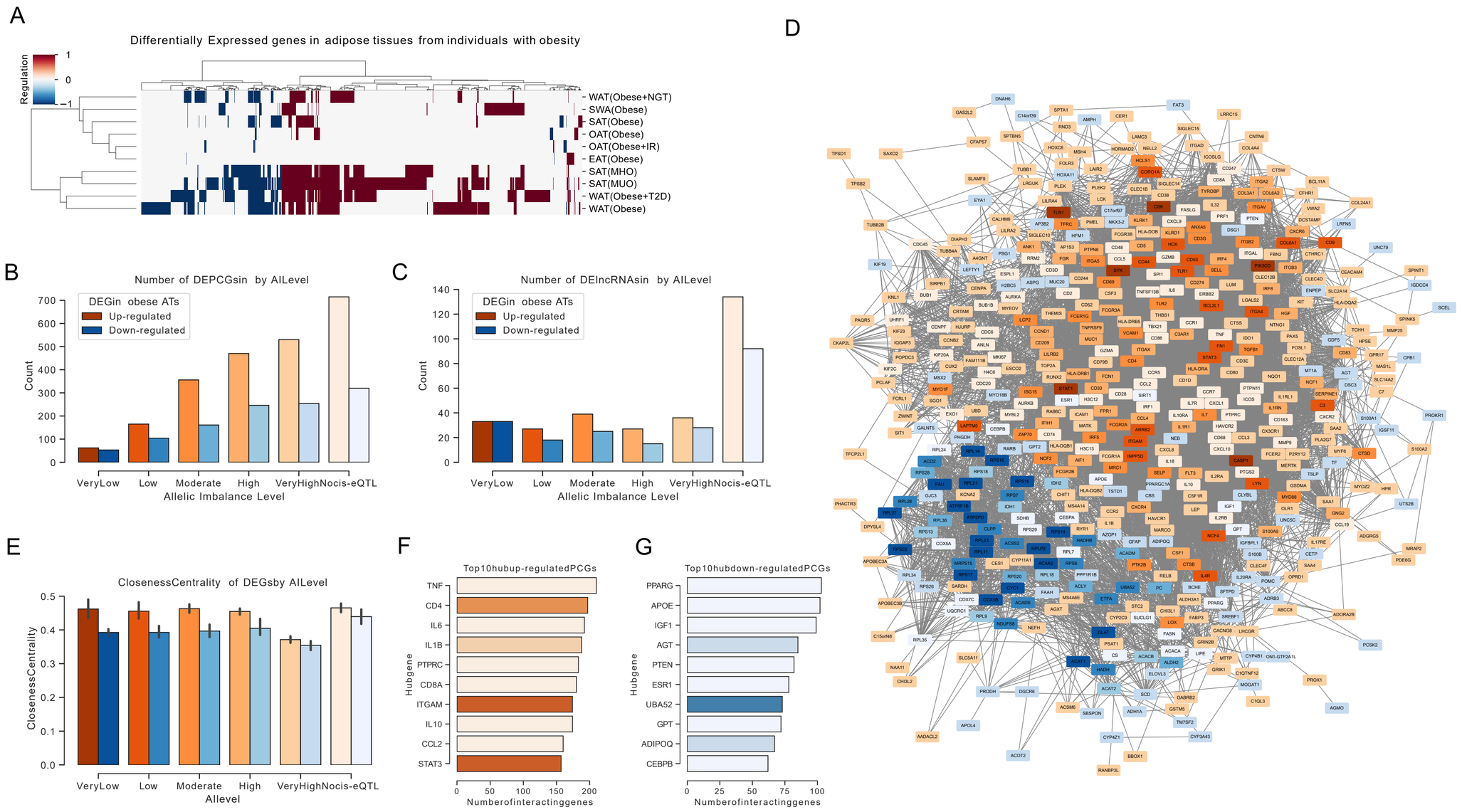

A
Differentially Expressed genes in adipose tissues from individuals with obesity
Regulation
1
0
-1
WAT(Obese+NGT)
SWA(Obese)
SAT(Obese)
OAT(Obese)
OAT(Obese+IR)
EAT(Obese)
SAT(MHO)
SAT(MUO)
WAT(Obese+T2D)
WAT(Obese)
B
Number of DEPCGsin by AILevel
DEGin obese ATs
Up-regulated
Down-regulated
Count
VeryLow
Low
Moderate
High
VeryHigh
Nocis-eQTL
Allelic Imbalance Level
C
Number of DEIncRNAsin by AILevel
DEGin obese ATs
Up-regulated
Down-regulated
Count
VeryLow
Low
Moderate
High
VeryHigh
Nocis-eQTL
Allelic Imbalance Level
D
E
ClosenessCentrality of DEGsby AILevel
ClosenessCentrality
VeryLow
Low
Moderate
High
VeryHigh
Nocis-eQTL
AILevel
F
Top10hubup-regulatedPCGs
Hubgene
TNF
CD4
IL6
IL1B
PTPRC
CD8A
ITGAM
IL10
CCL2
STAT3
Numberofinteractinggenes
G
Top10hubdown-regulatedPCGs
Hubgene
PPARG
APOE
IGF1
AGT
PTEN
ESR1
UBA52
GPT
ADIPOQ
CEBPB
Numberofinteractinggenes

Figure 5

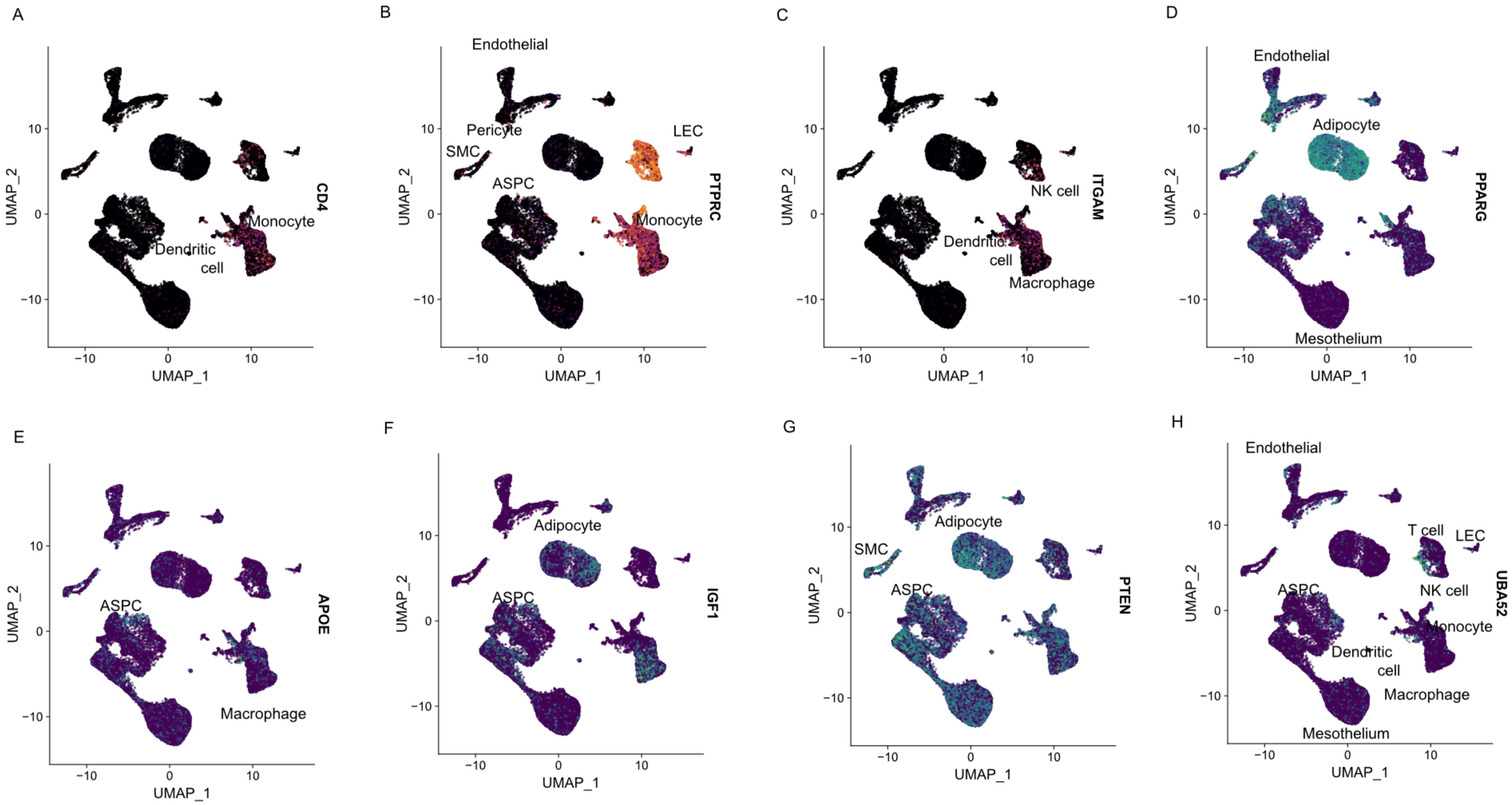

A
CD4
Monocyte
Dendritic cell
B
PTPRC
Endothelial
Pericyte
SMC
ASPC
LEC
Monocyte
C
ITGAM
NK cell
Dendritic cell
Macrophage
D
PPARG
Endothelial
Adipocyte
Mesothelium
E
APOE
ASPC
Macrophage
F
IGF1
Adipocyte
ASPC
G
PTEN
Adipocyte
SMC
ASPC
H
UBA52
Endothelial
T cell
LEC
ASPC
NK cell
Monocyte
Dendritic cell
Macrophage
Mesothelium
UMAP_1
UMAP_2
10
0
−10

Figure 6

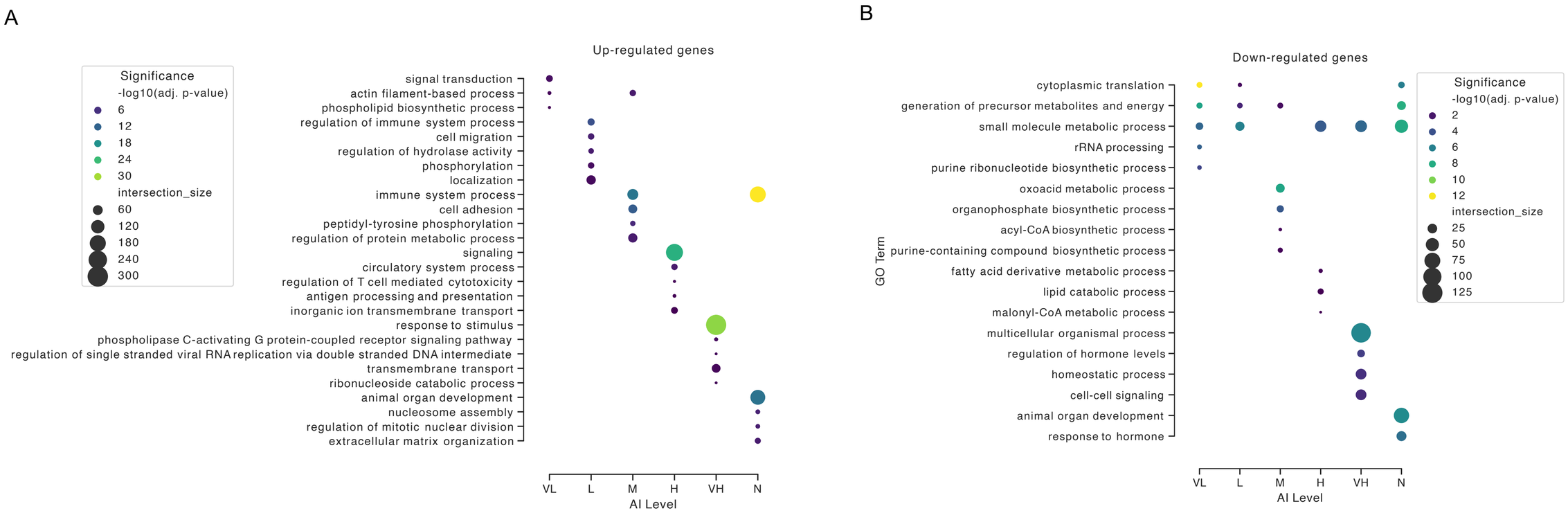

A
Up-regulated genes
Significance
-log10(adj. p-value)
6
12
18
24
30
intersection_size
60
120
180
240
300
signal transduction
actin filament-based process
phospholipid biosynthetic process
regulation of immune system process
cell migration
regulation of hydrolase activity
phosphorylation
localization
immune system process
cell adhesion
peptidyl-tyrosine phosphorylation
regulation of protein metabolic process
signaling
circulatory system process
regulation of T cell mediated cytotoxicity
antigen processing and presentation
inorganic ion transmembrane transport
response to stimulus
phospholipase C-activating G protein-coupled receptor signaling pathway
regulation of single stranded viral RNA replication via double stranded DNA intermediate
transmembrane transport
ribonucleoside catabolic process
animal organ development
nucleosome assembly
regulation of mitotic nuclear division
extracellular matrix organization
VL
L
M
H
VH
N
AI Level
B
Down-regulated genes
GO Term
cytoplasmic translation
generation of precursor metabolites and energy
small molecule metabolic process
rRNA processing
purine ribonucleotide biosynthetic process
oxoacid metabolic process
organophosphate biosynthetic process
acyl-CoA biosynthetic process
purine-containing compound biosynthetic process
fatty acid derivative metabolic process
lipid catabolic process
malonyl-CoA metabolic process
multicellular organismal process
regulation of hormone levels
homeostatic process
cell-cell signaling
animal organ development
response to hormone
Significance
-log10(adj. p-value)
2
4
6
8
10
12
intersection_size
25
50
75
100
125
VL
L
M
H
VH
N
AI Level

Figure 7

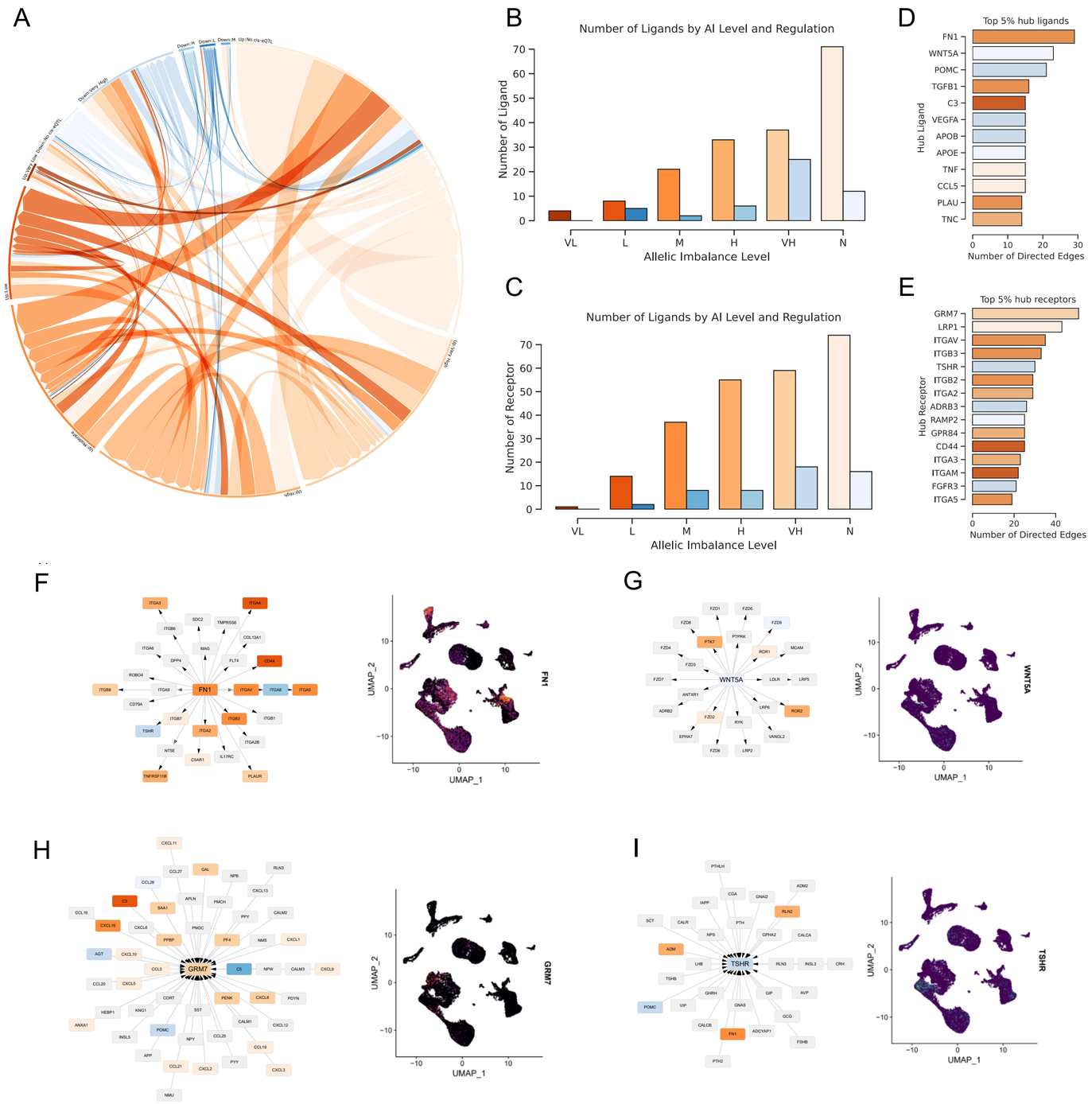

A
B
Number of Ligands by AI Level and Regulation
Number of Ligand
Allelic Imbalance Level
VL
L
M
H
VH
N
C
Number of Ligands by AI Level and Regulation
Number of Receptor
Allelic Imbalance Level
D
Top 5% hub ligands
Hub Ligand
FN1
WNT5A
POMC
TGFB1
C3
VEGFA
APOB
APOE
TNF
CCL5
PLAU
TNC
Number of Directed Edges
E
Top 5% hub receptors
Hub Receptor
GRM7
LRP1
ITGAV
ITGB3
TSHR
ITGB2
ITGA2
ADRB3
RAMP2
GPR84
CD44
ITGA3
ITGAM
FGFR3
ITGA5
Number of Directed Edges
F
FN1
UMAP_1
UMAP_2
G
WNT5A
UMAP_1
UMAP_2
H
GRM7
UMAP_1
UMAP_2
I
TSHR
UMAP_1
UMAP_2

Figure 8

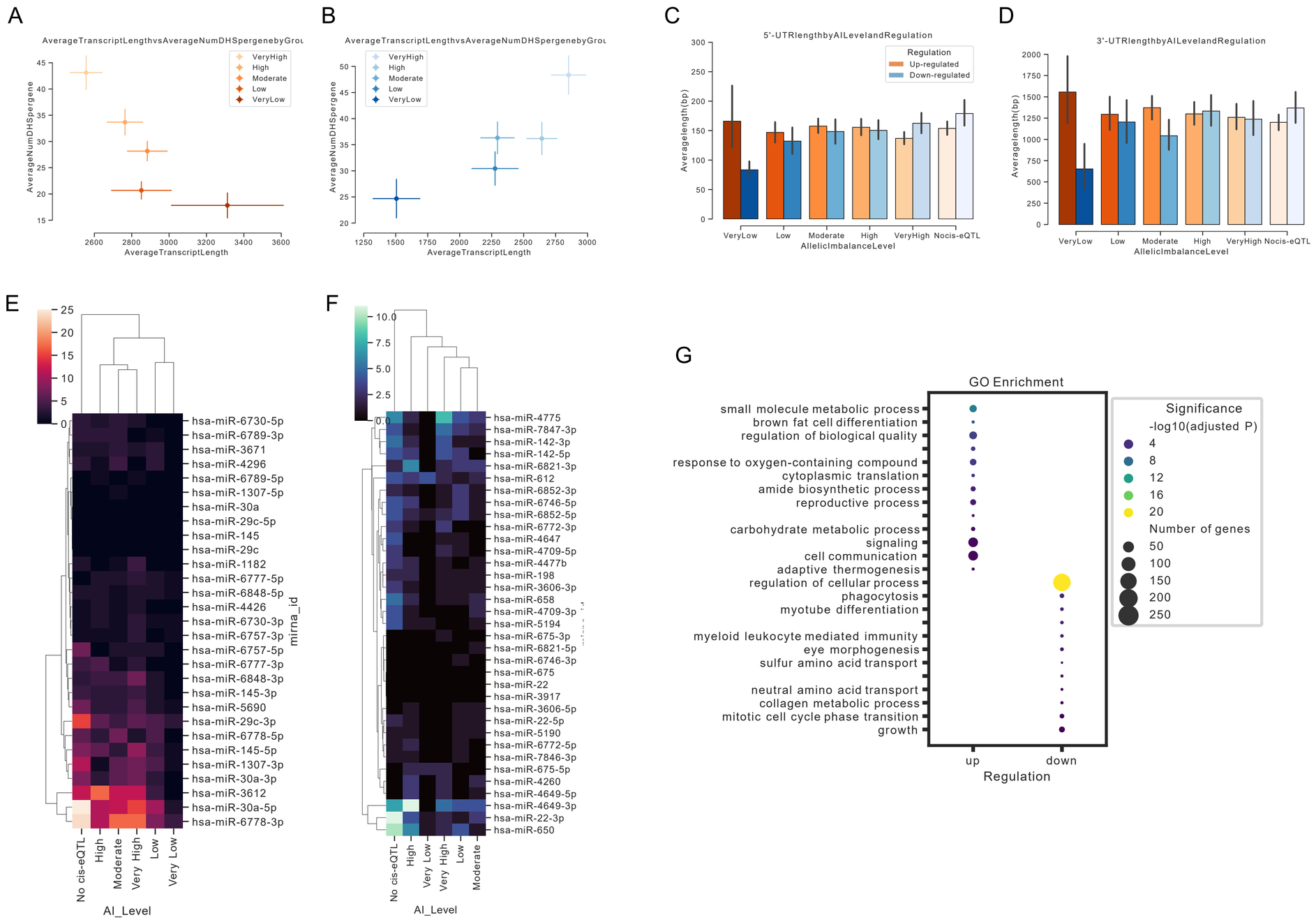

A
AverageTranscriptLengthvsAverageNumDHSpergenebyGrou
VeryHigh
High
Moderate
Low
VeryLow
AverageNumDHSpergene
AverageTranscriptLength
B
AverageTranscriptLengthvsAverageNumDHSpergenebyGrou
VeryHigh
High
Moderate
Low
VeryLow
AverageNumDHSpergene
AverageTranscriptLength
C
5'-UTRlengthbyAILevelandRegulation
Regulation
Up-regulated
Down-regulated
Averagelength(bp)
VeryLow
Low
Moderate
High
VeryHigh
Nocis-eQTL
AllelicImbalanceLevel
D
3'-UTRlengthbyAILevelandRegulation
Averagelength(bp)
VeryLow
Low
Moderate
High
VeryHigh
Nocis-eQTL
AllelicImbalanceLevel
E
hsa-miR-6730-5p
hsa-miR-6789-3p
hsa-miR-3671
hsa-miR-4296
hsa-miR-6789-5p
hsa-miR-1307-5p
hsa-miR-30a
hsa-miR-29c-5p
hsa-miR-145
hsa-miR-29c
hsa-miR-1182
hsa-miR-6777-5p
hsa-miR-6848-5p
hsa-miR-4426
hsa-miR-6730-3p
hsa-miR-6757-3p
hsa-miR-6757-5p
hsa-miR-6777-3p
hsa-miR-6848-3p
hsa-miR-145-3p
hsa-miR-5690
hsa-miR-29c-3p
hsa-miR-6778-5p
hsa-miR-145-5p
hsa-miR-1307-3p
hsa-miR-30a-3p
hsa-miR-3612
hsa-miR-30a-5p
hsa-miR-6778-3p
mirna_id
No cis-eQTL
High
Moderate
Very High
Low
Very Low
AI_Level
F
hsa-miR-4775
hsa-miR-7847-3p
hsa-miR-142-3p
hsa-miR-142-5p
hsa-miR-6821-3p
hsa-miR-612
hsa-miR-6852-3p
hsa-miR-6746-5p
hsa-miR-6852-5p
hsa-miR-6772-3p
hsa-miR-4647
hsa-miR-4709-5p
hsa-miR-4477b
hsa-miR-198
hsa-miR-3606-3p
hsa-miR-658
hsa-miR-4709-3p
hsa-miR-5194
hsa-miR-675-3p
hsa-miR-6821-5p
hsa-miR-6746-3p
hsa-miR-675
hsa-miR-22
hsa-miR-3917
hsa-miR-3606-5p
hsa-miR-22-5p
hsa-miR-5190
hsa-miR-6772-5p
hsa-miR-7846-3p
hsa-miR-675-5p
hsa-miR-4260
hsa-miR-4649-5p
hsa-miR-4649-3p
hsa-miR-22-3p
hsa-miR-650
No cis-eQTL
High
Very Low
Very High
Low
Moderate
AI_Level
G
GO Enrichment
small molecule metabolic process
brown fat cell differentiation
regulation of biological quality
response to oxygen-containing compound
cytoplasmic translation
amide biosynthetic process
reproductive process
carbohydrate metabolic process
signaling
cell communication
adaptive thermogenesis
regulation of cellular process
phagocytosis
myotube differentiation
myeloid leukocyte mediated immunity
eye morphogenesis
sulfur amino acid transport
neutral amino acid transport
collagen metabolic process
mitotic cell cycle phase transition
growth
up
down
Regulation
Significance
-log10(adjusted P)
Number of genes

Figure 9

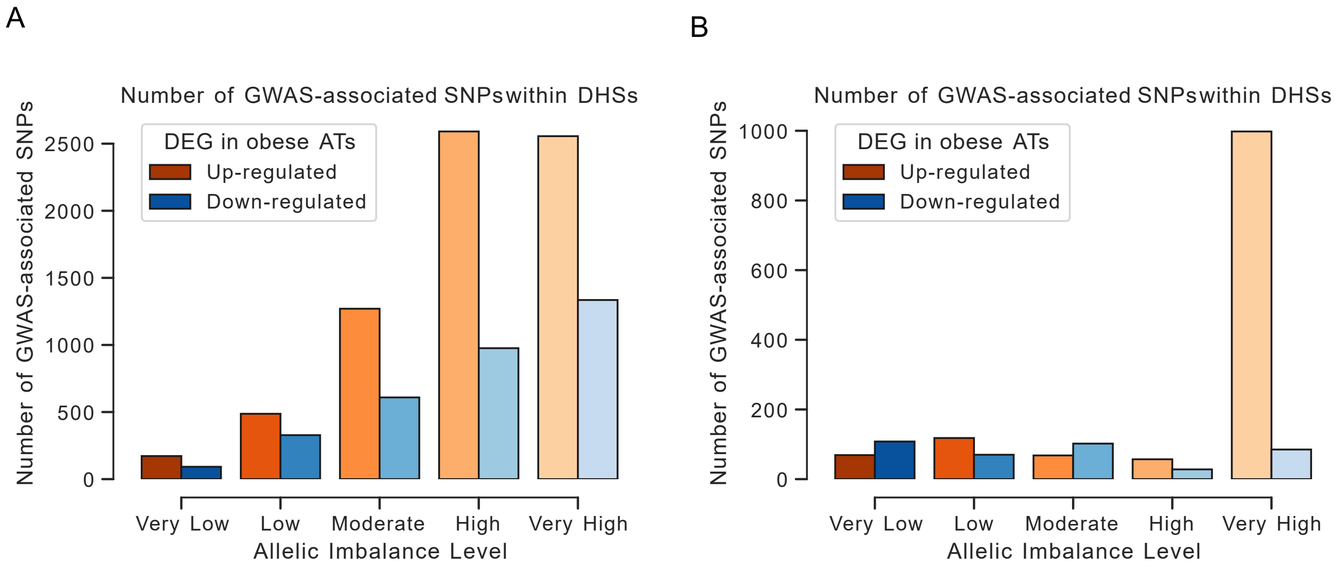

A
Number of GWAS-associated SNPswithin DHSs
Number of GWAS-associated SNPs
DEG in obese ATs
Up-regulated
Down-regulated
0
500
1000
1500
2000
2500
Very Low
Low
Moderate
High
Very High
Allelic Imbalance Level
B
Number of GWAS-associated SNPswithin DHSs
Number of GWAS-associated SNPs
DEG in obese ATs
Up-regulated
Down-regulated
0
200
400
600
800
1000
Very Low
Low
Moderate
High
Very High
Allelic Imbalance Level

Figure 10

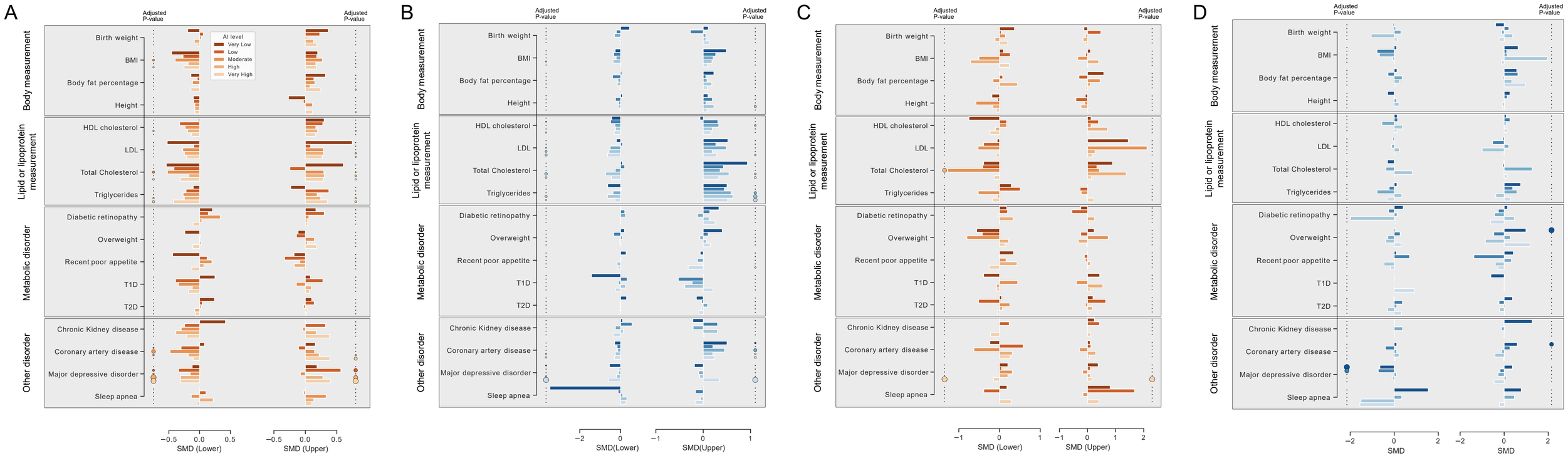
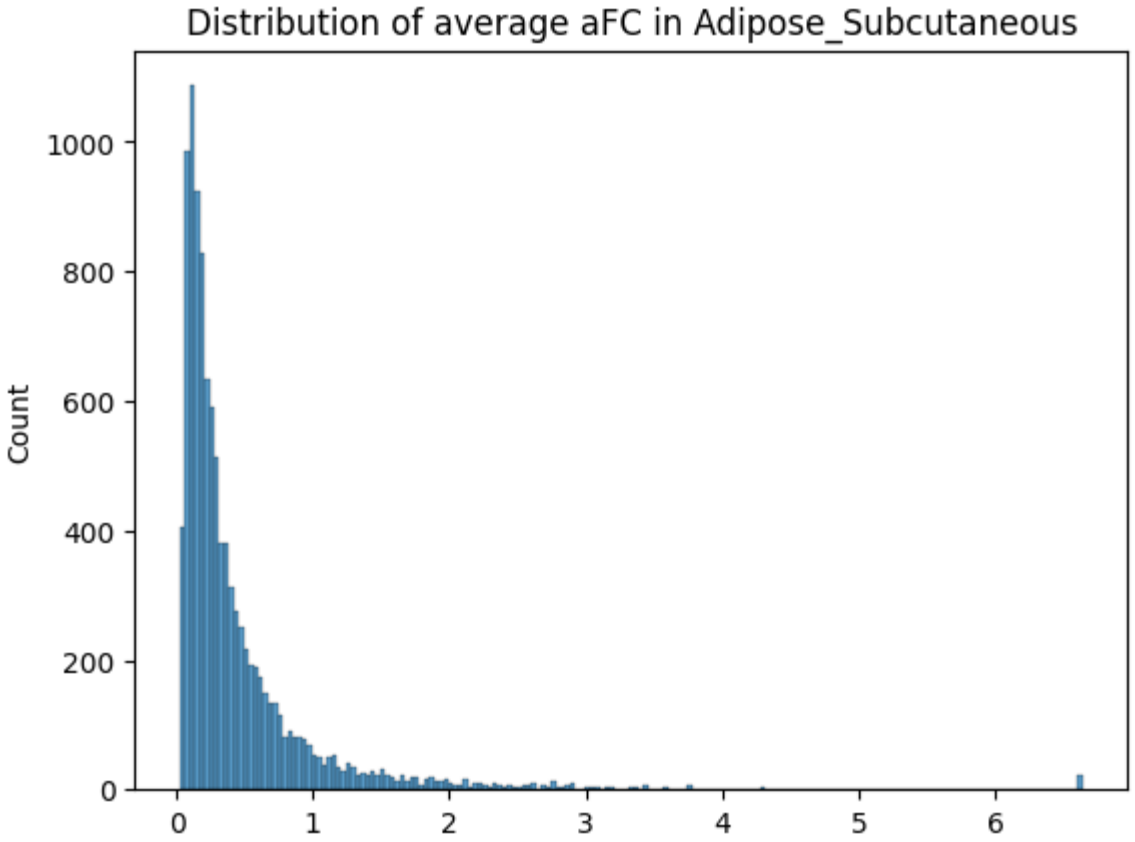

A
B
C
D
Adjusted P-value
Body measurement
Lipid or lipoprotein measurement
Metabolic disorder
Other disorder
Birth weight
BMI
Body fat percentage
Height
HDL cholesterol
LDL
Total Cholesterol
Triglycerides
Diabetic retinopathy
Overweight
Recent poor appetite
T1D
T2D
Chronic Kidney disease
Coronary artery disease
Major depressive disorder
Sleep apnea
AI level
Very Low
Low
Moderate
High
Very High
SMD (Lower)
SMD (Upper)
SMD(Lower)
SMD(Upper)
SMD

Figure S1

aFC
5
0
−5
Testis
Spleen
Brain_Cerebellar_Hemisphere
Brain_Cerebellum
Small_Intestine_Terminal_Ileum
Brain_Nucleus_accumbens_basal_ganglia
Cells_EBV-transformed_lymphocytes
Liver
Vagina
Minor_Salivary_Gland
Brain_Anterior_cingulate_cortex_BA24
Brain_Caudate_basal_ganglia
Brain_Hypothalamus
Brain_Cortex
Brain_Frontal_Cortex_BA9
Brain_Hippocampus
Brain_Spinal_cord_cervical_c-1
Brain_Putamen_basal_ganglia
Brain_Amygdala
Brain_Substantia_nigra
Prostate
Pituitary
Cells_Cultured_fibroblasts
Ovary
Thyroid
Whole_Blood
Adrenal_Gland
Esophagus_Mucosa
Skin_Not_Sun_Exposed_Suprapubic
Skin_Sun_Exposed_Lower_leg
Colon_Sigmoid
Colon_Transverse
Lung
Nerve_Tibial
Artery_Aorta
Pancreas
Stomach
Uterus
Breast_Mammary_Tissue
Adipose_Subcutaneous
Adipose_Visceral_Omentum
Esophagus_Gastroesophageal_Junction
Esophagus_Muscularis
Artery_Coronary
Artery_Tibial
Heart_Atrial_Appendage
Heart_Left_Ventricle
Kidney_Cortex
Muscle_Skeletal

Figure S2

A


Distribution of average aFC in Adipose_Subcutaneous
Count
1000
800
600
400
200
0
0
1
2
3
4
5
6

B

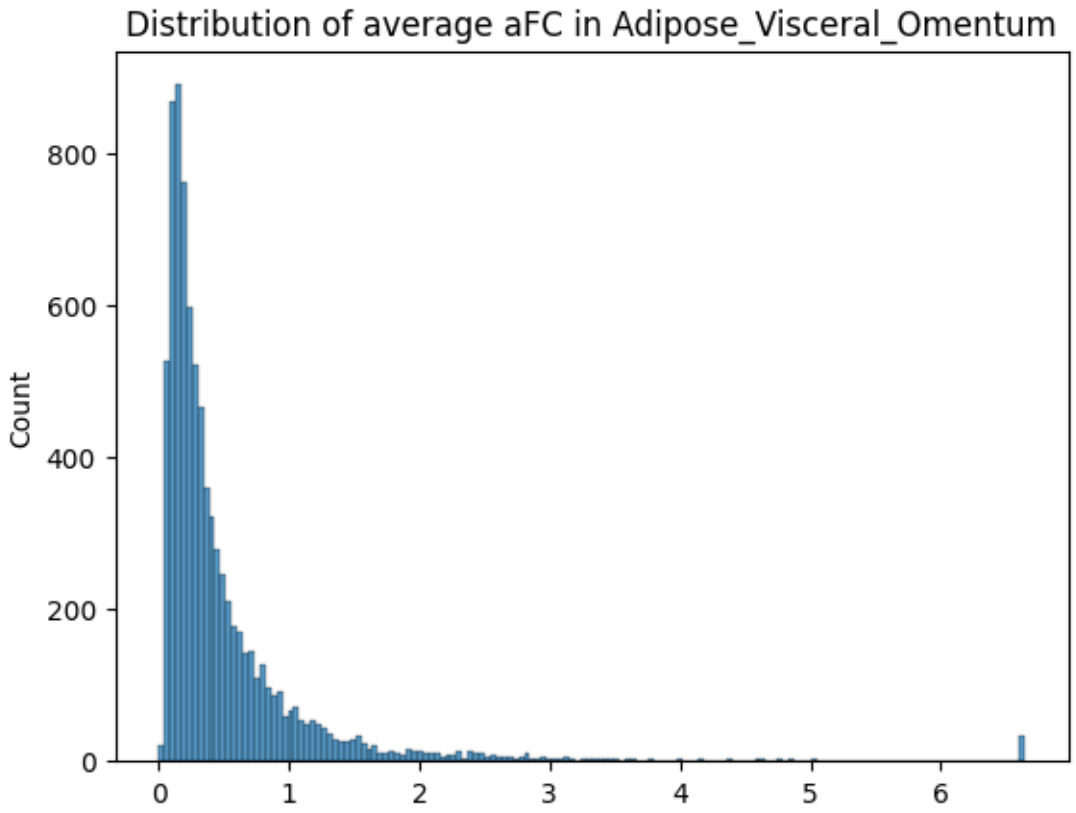
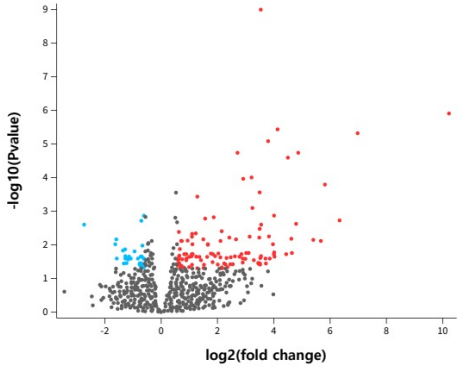
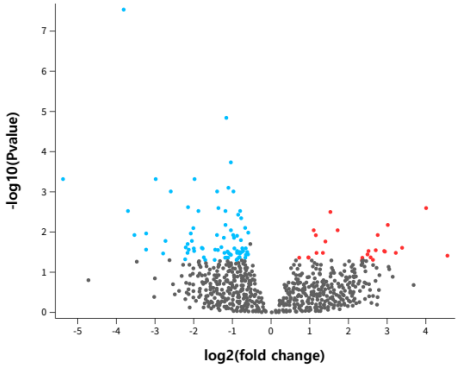

Distribution of average aFC in Adipose_Visceral_Omentum
Count
800
600
400
200
0
0
1
2
3
4
5
6

Figure S3

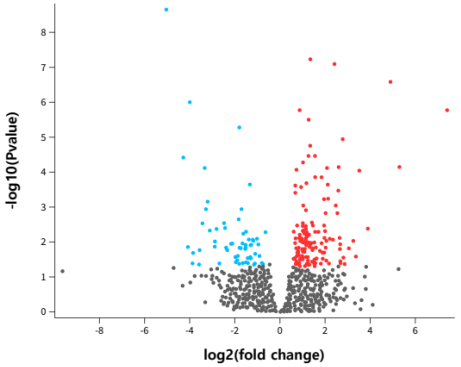
GSE179455: Transcriptome analysis of Epicardial Adipose Tissue from...
obese vs lean, Padj<0.05



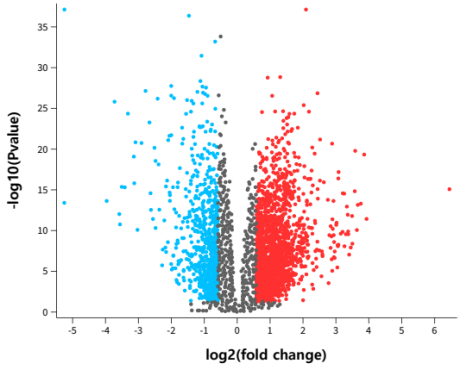
GSE165932: Insulin resistance rewires the metabolic gene program and...
obese vs lean, Padj<0.05



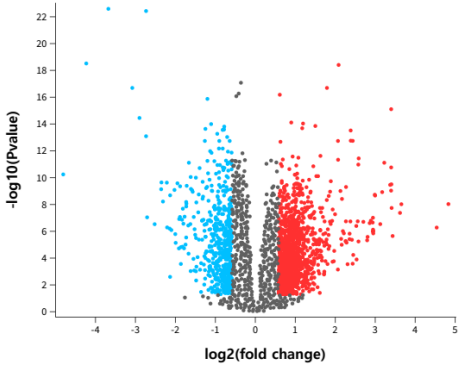
GSE55008: Transcriptome profiling of omental adipose tissues in human...
obese vs lean, Padj<0.05



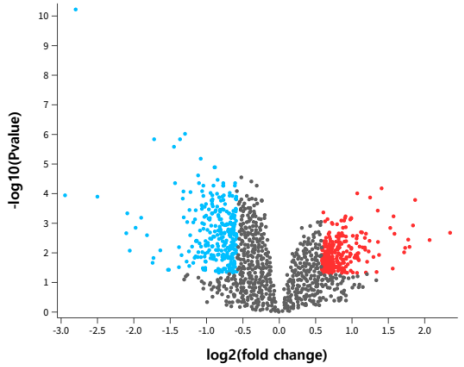
GSE156906: Adipose tissue immunology, inflammation and exosomes in...
obese vs lean, Padj<0.05



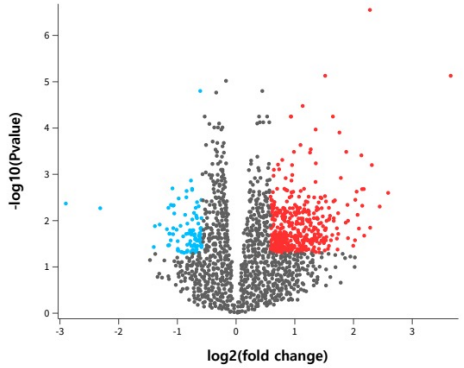
GSE156906: Adipose tissue immunology, inflammation and exosomes in...
obese vs lean, Padj<0.05



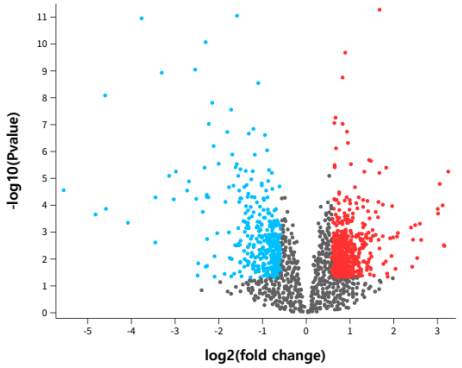
GSE205668: Independent phenotypic plasticity axes define mammalian...
obese vs normal weight, Padj<0.05



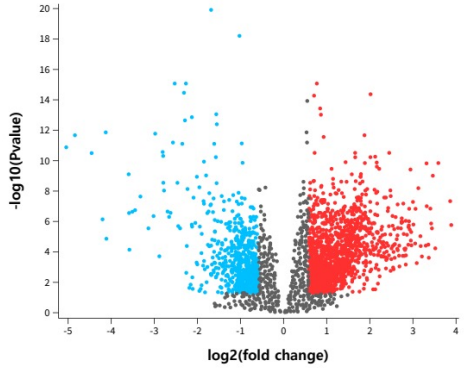
GSE110729: Long non-coding RNAs associated with metabolic traits in...
obese vs lean, Padj<0.05



GSE141432: Obese Insulin Resistant Humans with Compensatory...
obese vs lean, Padj<0.05



GSE141432: Obese Insulin Resistant Humans with Compensatory...
obese T2D vs lean, Padj<0.05



GSE162653: RNA-seq analysis of human white adipose tissue from normal...
obese vs normal weight, Padj<0.05

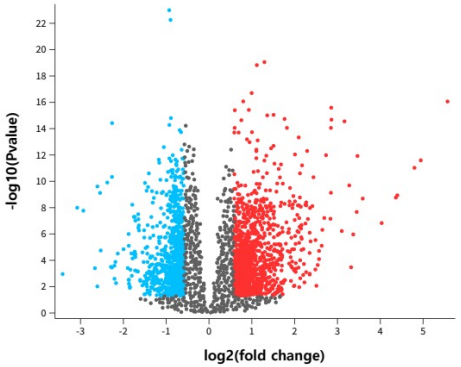
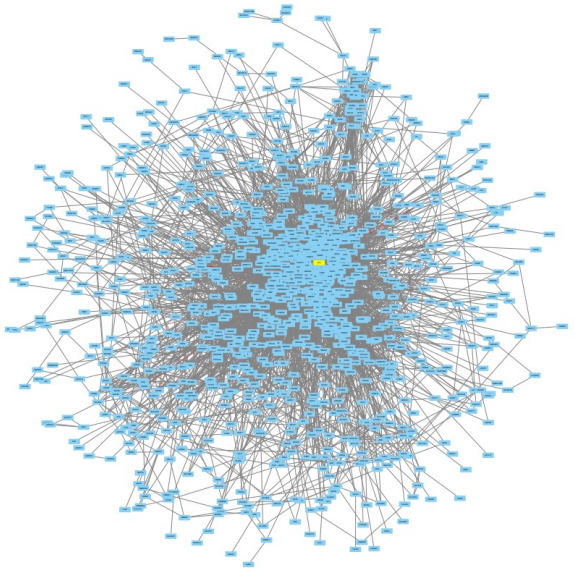


Figure S4

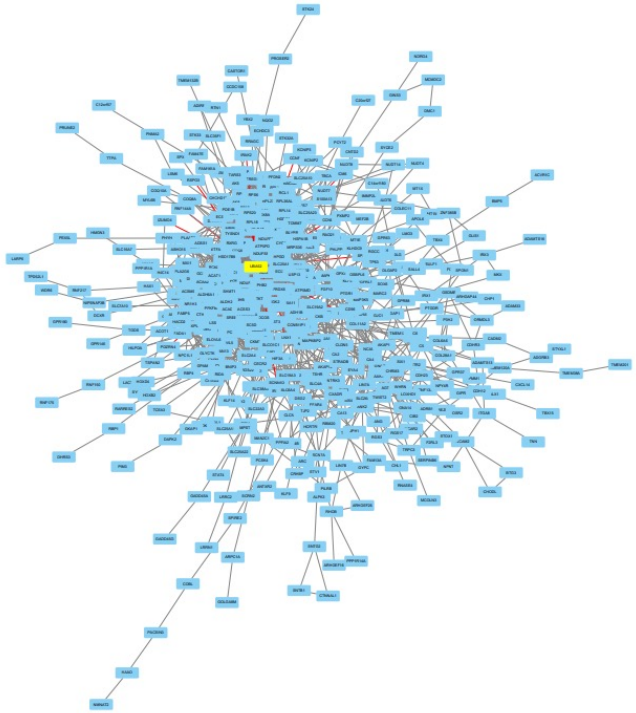
A



Hub genes (up)

Gene symbol	Number Of Undirected Edges
CD4	211
ITGAM	173
FN1	157
CD44	148
ITGB2	148
FCGR3A	146
TYROBP	136
STAT3	133
TGFB1	126
SYK	125
ITGAX	124
TLR2	123
CXCR4	122
CSF1R	122
STAT1	119
CD80	109
MYD88	109
SELL	107
CD274	105
FCER1G	103
VCAM1	100
LYN	100
PLEK	98
CASP1	98
FCGR2B	96
IRF8	95
CTSS	95
KIT	94

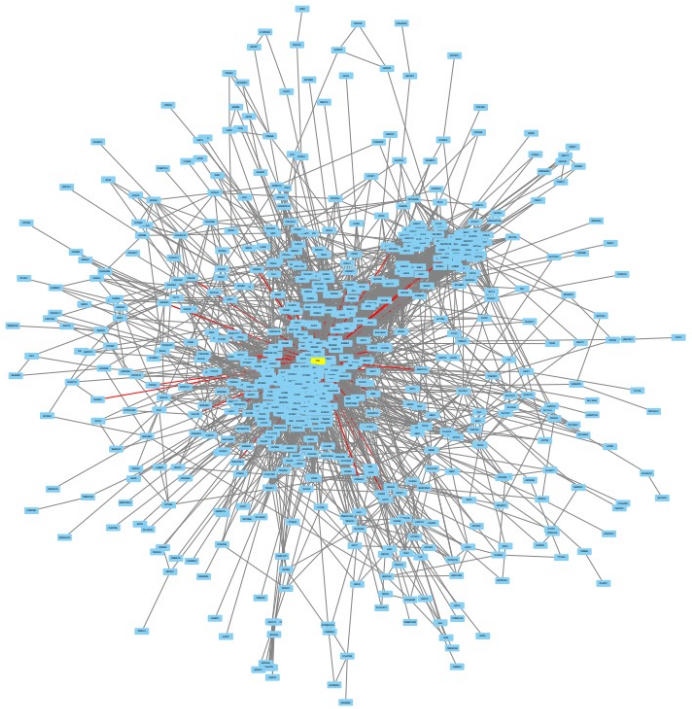
B



Hub genes (down)

Gene symbol	Number Of Undirected Edges
UBA52	70
ACO2	66
ACLY	56
RPL23	56
NDUFS8	55
RPS18	55
HADHB	54
ACADM	53
RPL9	53
RPS20	53
ATP5F1B	52
ACAA2	51
IDH1	50
RPS16	50
RPS14	50
RPL11	50
RPS15	47
ACSS2	45
ALDH2	43
RPS6	43
RPS13	43
SREBF1	42
IDH2	42
MRPS15	42
RPL18	42
RPL27	42
FAU	42
HADH	41
PC	41
PPARGC1A	40

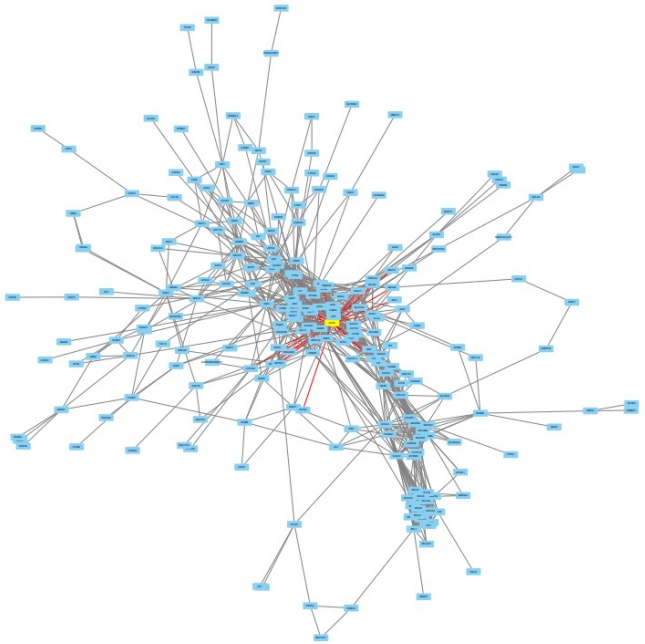
C



Hub genes (up)

Gene symbol	Number Of Undirected Edges
TNF	183
IL6	171
PTPRC	151
CD8A	146
IL10	129
CCL2	122
CD86	104
MMP9	104
CCL5	101
CCR7	97
CCR5	94
CXCL10	94
IL7R	93
GZMB	87
CD163	84
CD68	83
SPI1	82
H3C13	80
CXCL9	79
H3C12	79
CD28	78
CD48	78
CENPA	77
PTGS2	76
PRF1	75
CD247	74
CD2	74
MKI67	74
CXCL1	74

D



Hub genes (down)

Gene symbol	Number Of Undirected Edges
FASN	41
PPARG	41
ESR1	40
IGF1	33
PTEN	32
SDHB	31
CS	30
COX7C	29
SIRT1	28
ACACA	27
CEBPA	27
CEBPB	26
GPT	25
APOE	24
LIPE	24
RPS29	24
COX5A	24
AGT	23
RPL35	23
RPL24	22
RPL7	22
SUCLG1	21
UQCRC1	21
COX6A1	21
RPL23A	21
RPL38	21
ACSL1	20
APOB	20
RPL35A	20

Figure S5

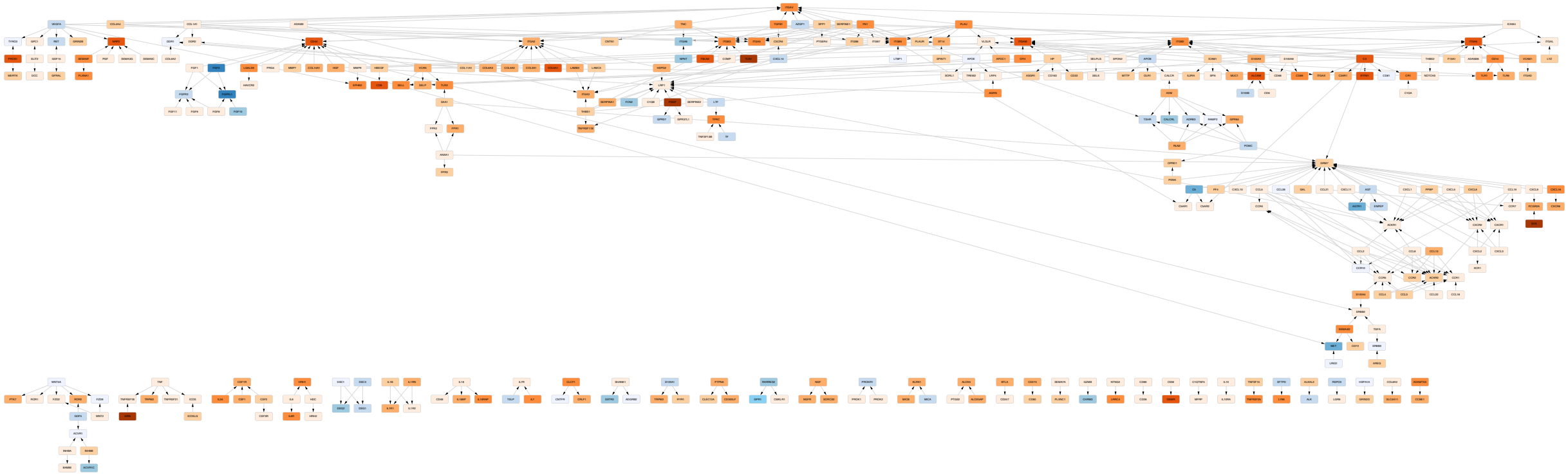
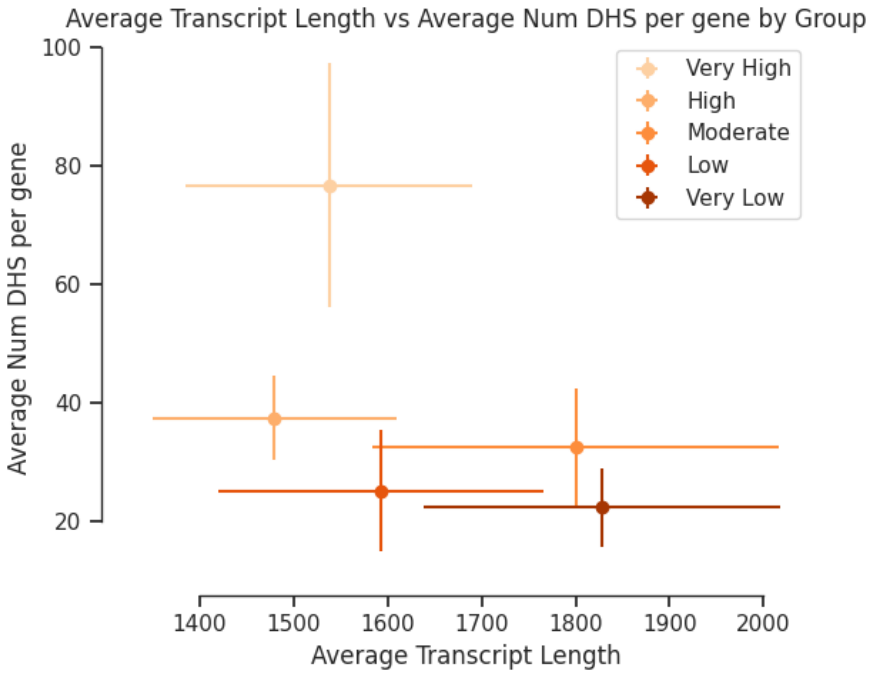


Figure S6

A


Average Transcript Length vs Average Num DHS per gene by Group
Average Num DHS per gene
100
80
60
40
20
1400
1500
1600
1700
1800
1900
2000
Average Transcript Length
Very High
High
Moderate
Low
Very Low

B

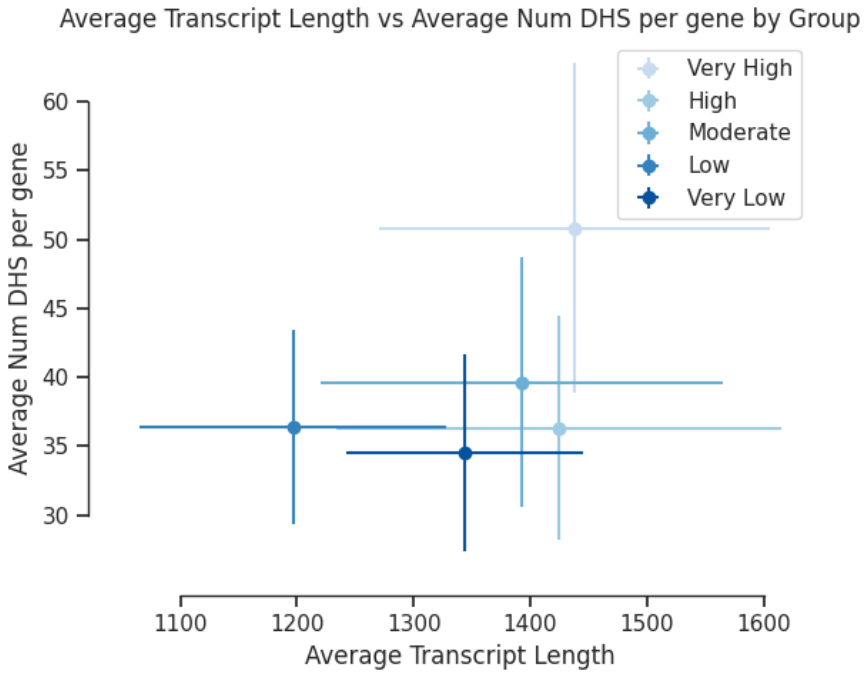
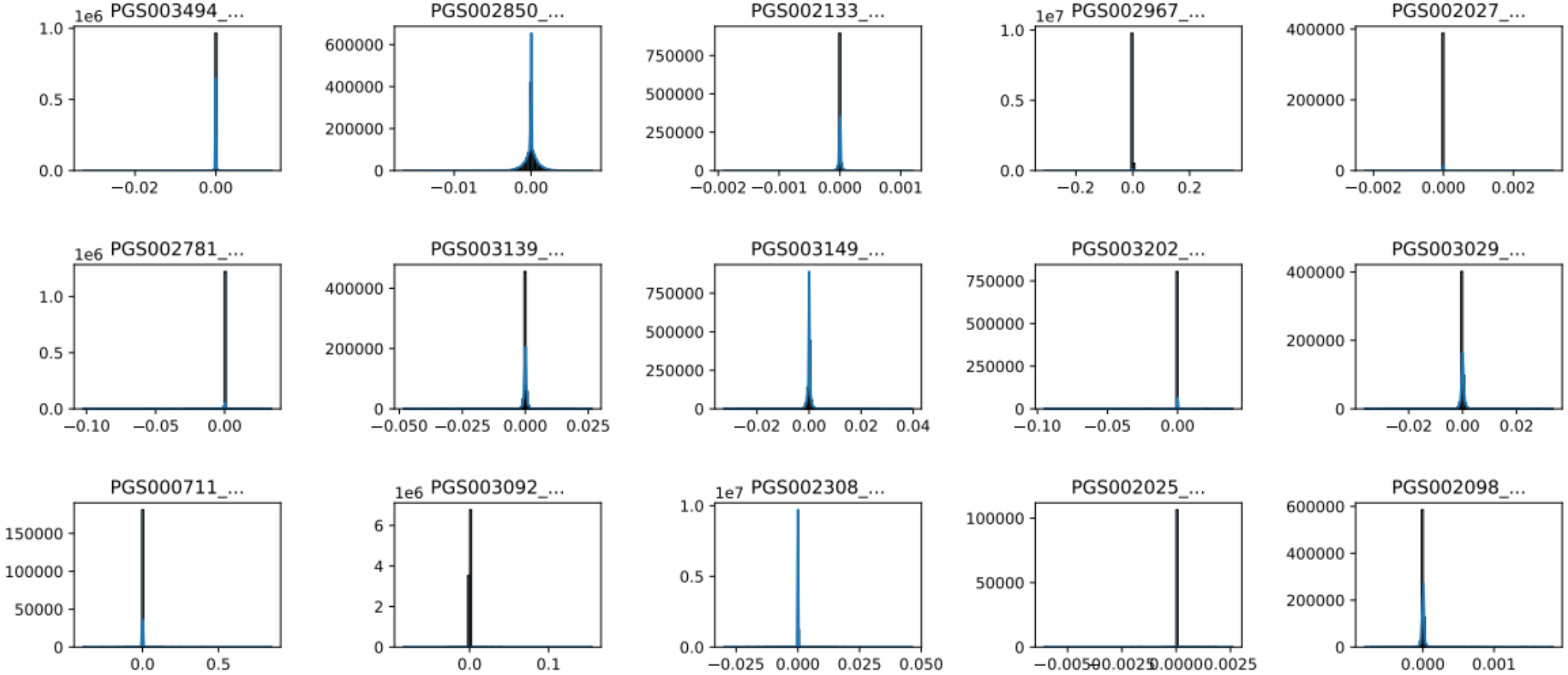

Average Transcript Length vs Average Num DHS per gene by Group
Average Num DHS per gene
60
55
50
45
40
35
30
1100
1200
1300
1400
1500
1600
Average Transcript Length
Very High
High
Moderate
Low
Very Low

Figure S7


PGS003494_...
PGS002850_...
PGS002133_...
PGS002967_...
PGS002027_...
PGS002781_...
PGS003139_...
PGS003149_...
PGS003202_...
PGS003029_...
PGS000711_...
PGS003092_...
PGS002308_...
PGS002025_...
PGS002098_...