The landscape of *SETBP1* gene expression and transcription factor activity across human tissues

Jordan H. Whitlock[1], Elizabeth J. Wilk[1], Timothy C. Howton[1], Amanda D. Clark[1], Brittany N. Lasseigne[1]

[1] Department of Cell, Developmental and Integrative Biology, Heersink School of Medicine The University of Alabama at Birmingham, Birmingham, AL, U.S.A

Corresponding Author:
Brittany N. Lasseigne[1]
1900 University Boulevard, Tinsley Harrison Tower, Room 960, Birmingham, Alabama 35233
Email address: bnp0001@uab.edu

Abstract

Background. The SET binding protein 1 (*SETBP1*) gene encodes a transcription factor (TF) involved in various cellular processes. Distinct *SETBP1* variants have been linked to three different diseases. Germline variants cause the ultra-rare pediatric Schinzel Giedion Syndrome (SGS) and *SETBP1* haploinsufficiency disorder (*SETBP1*-HD), characterized by severe multisystemic abnormalities with neurodegeneration or a less severe brain phenotype accompanied by hypotonia and strabismus, respectively. Somatic variants in *SETBP1* are associated with hematological malignancies and cancer development in other tissues in adults.
Results. To better understand the tissue-specific mechanisms involving *SETBP1*, we analyzed publicly available RNA-sequencing data from the Genotype-Tissue Expression (GTEx) project. We found *SETBP1,* and its known target genes were widely expressed across 31 adult human tissues. K-means clustering identified three distinct expression patterns of SETBP1 targets across tissues. Functional enrichment analysis (FEA) of each cluster revealed gene sets related to transcription regulation, DNA binding, and mitochondrial function. TF activity analysis of SETBP1 and its target TFs revealed tissue-specific TF activity, underscoring the role of tissue context-driven regulation and suggesting its impact in SETBP1-associated disease. In addition to uncovering tissue-specific molecular signatures of *SETBP1* expression and TF activity, we provide a Shiny web application to facilitate exploring TF activity across human tissues for 758 TFs.
Conclusions. This study provides insight into the landscape of *SETBP1* expression and TF activity across 31 non-diseased human tissues and reveals tissue-specific expression and activity of *SETBP1* and its targets. In conjunction with the web application we constructed, our framework enables researchers to generate hypotheses related to the role tissue backgrounds play with respect to gene expression and TF activity in different disease contexts.

Introduction

SETBP1 is a gene located on the long (q) arm of chromosome 18 that encodes the transcription factor (TF) and oncogene SET binding protein 1 (Coccaro et al., 2017). Referred to as a DNA-binding protein, SETBP1 has several motifs, including three nuclear localization signals, a SKI homology region, and a binding region for SET nuclear oncogene. As a protein, it has a role in DNA replication and transcriptional regulation (Piazza et al., 2018). Different pathogenic variants in *SETBP1* can result in three distinct diseases (Kohyanagi & Ohama, 2023). Germline variants cause two unique ultra-rare, de novo pediatric diseases: Schinzel Giedion Syndrome (SGS) (Schinzel & Giedion, 1978) and SETBP1 haploinsufficiency disorder (*SETBP1*-HD) (Morgan et al., 2021). These conditions are differentiated by variant location, phenotypic severity, and accompanying protein gain or loss of function (GoF, LoF), respectively (Schinzel & Giedion, 1978; Morgan et al., 2021). SGS is multisystemic, involving gastrointestinal, cardiorespiratory, neurological, musculoskeletal, and urogenital abnormalities. It has a more severe phenotype than *SETBP1*-HD and affected individuals are characterized by progressive neurodegeneration and shortened life expectancy ("Schinzel Giedion Syndrome - NORD (National Organization for Rare Disorders)," 2015; Liu et al., 2018). However, SGS and *SETBP1*-HD, as disorders of protein dosage, have overlapping phenotypes, including intellectual disability, developmental delay, language impairment, distinctive craniofacial and skeletal features, and hypotonia (Schinzel & Giedion, 1978; Liu et al., 2018; Morgan et al., 2021). In contrast, somatic variants in *SETBP1* are associated with hematological malignancies and exhibit varying evidence for predisposing or promoting cancer in other adult tissue systems (reviewed in Coccaro et al., 2017).

There are multiple hypothesized mechanisms for the tissue-specificity of disease (i.e., clinical manifestations in some tissues but not others) related to intrinsic and extrinsic molecular processes spanning epigenetic, genetic, expression, regulation, and network-based mechanisms (Hekselman & Yeger-Lotem, 2020). Despite genomic advances in variant identification and sequencing technology, gaps remain in translating the role of specific genomic variants to observed phenotypic outcomes. Of these hypothesized mechanisms of tissue-specific disease manifestation, preferential or exclusive gene expression of SETBP1 targets and their altered regulation remain understudied in SETBP1-associated disorders. Potential mechanisms of SETBP1 dysfunction within neurodevelopment involving disrupted cell cycle control, DNA damage mechanisms, phosphatase activity, and chromatin remodeling have been hypothesized (Antonyan & Ernst, 2022). However, how the expression of *SETBP1* and its known TF targets function across additional tissue contexts and non-diseased human tissues requires further study.

Because of this, publicly available non-diseased data, such as from the Genotype-Tissue Expression (GTEx) project, provide an opportunity for investigating and generating hypotheses about the underlying function of disease-associated genes in different contexts, including in non-diseased and diseased contexts. Here, we investigated the gene expression of *SETBP1* and its known targets in RNA-sequencing (RNA-seq) data for each GTEx tissue. Then, we evaluated the functional enrichment of those targets based on how they clustered by expression across tissues. Next, we inferred the TF activity of SETBP1 and other TFs it is known to directly target by leveraging multivariate linear models to calculate enrichment scores representing activity for all TFs across tissues using GTEx expression and CollecTRI, a curated collection of TFs and

their directional regulation on transcriptional targets (**Fig. 1**). Collectively, we have mapped the gene expression and activity of SETBP1 and its targets across human non-diseased tissues, underscoring the potential impact of tissue background. Further, we have developed a Shiny web application (https://lasseignelab.shinyapps.io/gtex_tf_activity/) to facilitate the exploration and hypothesis generation of TF activity across human tissues for 758 TFs **(Table S1)**.

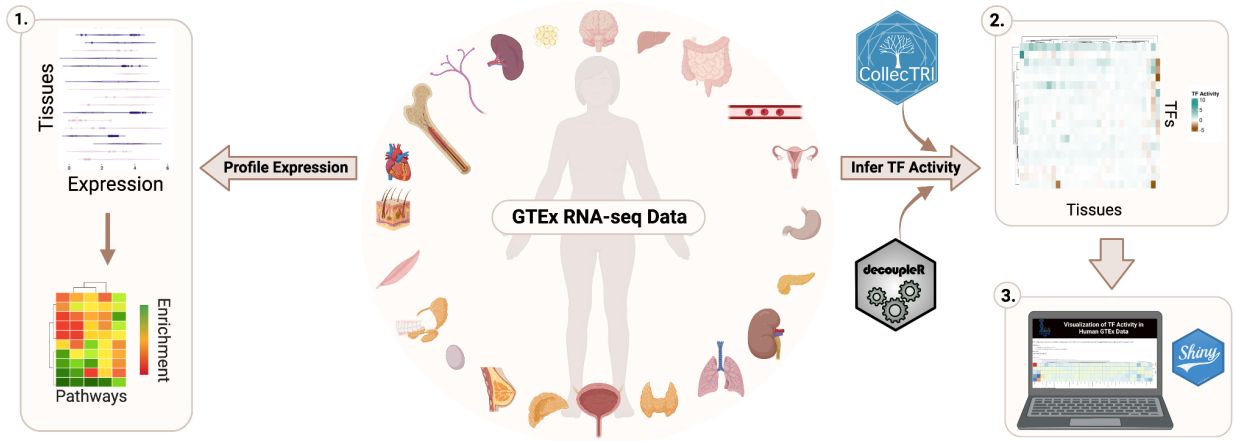

Figure 1 Study Overview. To investigate the tissue-specific expression and transcription factor (TF) activity of *SETBP1*, we analyzed publicly available RNA-sequencing (RNA-seq) data from the Genotype-Tissue Expression (GTEx) project. We profiled the tissue-specific expression of *SETBP1* and its known targets and performed functional enrichment analysis (1). We also inferred TF activity for SETBP1 and TFs it regulates by tissue (2) and developed an interactive web application to enable the exploration of TF activity for 758 TFs across all GTEx tissues (3).

Materials & Methods

Data Availability

Data and code supporting the Shiny app and to reproduce all analyses for this study are available at Zenodo (DOI:10.5281/zenodo.8222799) and https://github.com/lasseignelab/230323_JW_DiseaseNetworks (DOI:10.5281/zenodo.8225613).

The docker image used for these analyses is publicly available on Docker Hub (jordanwhitlock/setbp1_manuscript:1.0.13) and Zenodo (DOI:10.5281/zenodo.8428932). Our interactive web application can be accessed at https://lasseignelab.shinyapps.io/gtex_tf_activity/ (DOI:10.5281/zenodo.8225317).

SETBP1 target gene set construction

We used the SETBP1 target gene set compiled in Whitlock et al. 2023 to obtain a list of known TF targets of SETBP1 (Whitlock et al., 2023). We converted a list of human HGNC to human Ensembl IDs using gprofiler2 (Kolberg et al., 2020) (v.0.2.1), resulting in a final list of 209 genes.

RNA-sequencing expression data

Using recount3 (Wilks et al., 2021) (accessed August 2022), we obtained bulk RNA-seq data represented as transcripts per million (TPM) counts for human tissues (n=31) from the publicly available Genotype-Tissue Expression (GTEx) project. We included all tissues not classified as "study_na".

Classification of disease-associated affected tissues

We compiled a list of affected tissues in Schinzel Giedion Syndrome (SGS), SETBP1 haploinsufficiency disorder (SETBP1-HD), and SETBP1-associated cancer based on the literature, clinical manifestations noted within Online Mendelian Inheritance of Man (OMIM) (MIM #: 616078, 269150), and UpToDate (Schinzel & Giedion, 1978; Piazza et al., 2013; Patnaik et al., 2014; "Schinzel Giedion Syndrome - NORD (National Organization for Rare Disorders)," 2015; Elena et al., 2016; Amberger et al., 2019; Jansen et al., 2021; Morgan et al., 2021). Affected tissues for each SETBP1-associated disease phenotype included the following:

- SETBP1-HD: brain and muscle
- SGS: brain, muscle, heart, kidney, bladder, lung, small intestine, stomach, esophagus
- SETBP1-associated cancer: bone marrow and blood

While individuals with SGS and SETBP1-HD also present with vision problems and distinctive craniofacial and skeletal abnormalities, ocular and additional bone tissues are not present within GTEx v8 so they were not included in our analyses.

SETBP1 and target gene expression and pathway enrichment

We calculated the median TPM of *SETBP1* and its targets across samples in each tissue from GTEx and scaled (log2 + 1 transformed) for visualizations. We performed complete linkage of Euclidean distances for hierarchical clustering of GTEx tissues. For target gene expression, we identified optimal k-means clusters by plotting the total within-cluster sum of squared distances between samples (inertia) for each cluster tested (k = 1-15). We identified an elbow at 3 clusters, indicating a decrease in inertia and a sufficient trade-off between information and the number of clusters captured based on the expression of *SETBP1* and targets and plotted median scaled TPM values (**Fig. S1 A**). We further verified 3 clusters to be sufficient using the Trace(W) method, which uses the trace (sum of the diagonal) of the dispersion matrix (W), finds the second differences, and selects the cluster with the maximum value between indices (**Fig. S1 B**) (Charrad et al. 2014; Milligan and Cooper 1985). To visualize this clustering and the expression of SETBP1 and its targets across GTEx tissues, we used ComplexHeatmap (version 2.10.0) (Gu, Eils & Schlesner, 2016). We next performed functional enrichment analysis (FEA) with gprofiler2 (version 0.2.1) (Reimand et al., 2007; Kolberg et al., 2020) with GO sources (GO:BP, GO:MF, and GO:CC) to identify the expression of pathways from *SETBP1* target genes for each cluster. We applied the Bonferroni procedure for multiple hypothesis correction and used a p-adjusted threshold of 0.05 and, for the background gene list, included *SETBP1* and its targets.

TF activity

We acquired prior knowledge on the direction of TF regulation from CollecTRI (accessed May 2023) (Müller-Dott et al., 2023) and combined it with GTEx expression TPM to infer TF activity using decoupleR (Badia-I-Mompel et al., 2022) (v.2.6.0). We used a multivariate linear model

(run_mlm) with a minimum threshold of 5 targets per TF to calculate activity scores (represented as t-values) for all 758 TFs. We scaled and centered data before summarizing each TF's average regulator activity. Positive and negative scores denote TF activity and inactivity, respectively.

List of Abbreviations

FEA: Functional Enrichment Analysis
GTEx: Genotype-tissue expression
GoF: gain of function
LoF: loss of function
OMIM: Online Mendelian Inheritance of Man
RNA-seq: RNA-sequencing
SETBP1-HD: SETBP1 haploinsufficiency disorder
SGS: Schinzel Giedion Syndrome
TF: Transcription Factor
TPM: Transcripts per Million

Supporting Information

Table_S1.xlsx
Table_S2.xlsx
"Visualization of TF Activity in Human GTEx Data" Shiny app available at https://lasseignelab.shinyapps.io/gtex_tf_activity/

Results

We first examined the gene expression of *SETBP1* across 31 adult human tissues from the GTEx consortium (n=19,081 samples total). We found that while *SETBP1* was expressed ubiquitously (median transcript per million (TPM) range, 0.364-16.719; **Fig. 2A**), it was most highly expressed in the cervix, blood vessel, and uterus (median TPM 16.719, 15.422, and 13.730, respectively) and most lowly expressed in the blood, bone marrow, and the adrenal gland (median TPM 0.135, 0.384, and 0.963, respectively). When we investigated the brain-region-specific expression of *SETBP1*, we found it was similarly expressed across subregions (**Fig. 2B**).

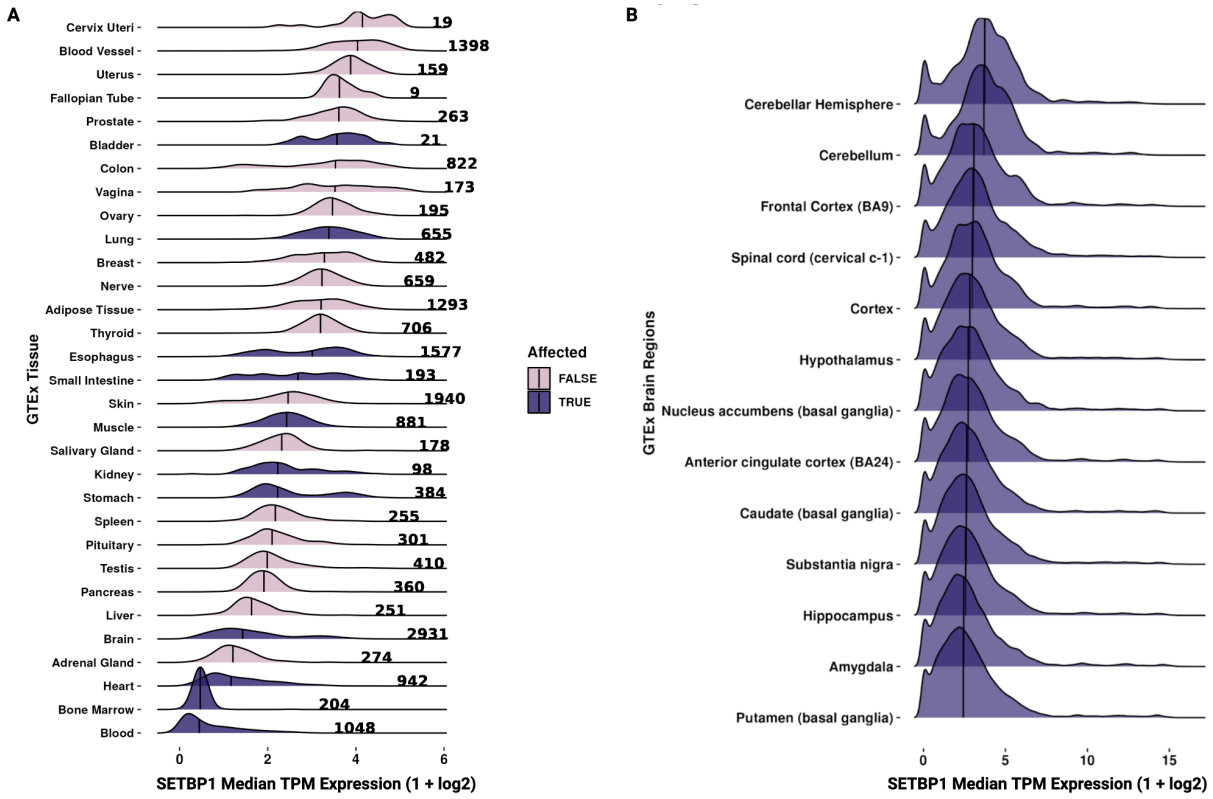

Figure 2 Gene expression of *SETBP1* across GTEx tissues. Ridgeline plot of the scaled median TPM values (x-axis) **(A)** across samples (denoted by bold number) for each GTEx tissue (y-axis) and **(B)** brain subregions (y-axis) for *SETBP1* where affected tissues are colored by dark purple and pink for true and false, respectively. The median is denoted by the vertical black line.

Further, we examined the gene expression of 209 known SETBP1 targets we previously compiled (Whitlock et al., 2023) from SETBP1 ChIP-seq binding sites (Kolmykov et al., 2021), MSigDB (Liberzon et al., 2015), SIGNOR (Lo Surdo et al., 2023), TRRUST (Han et al., 2018), Piazza et al. (Piazza et al., 2018), and Antonyan et al. (Antonyan & Ernst, 2022). We found that most known SETBP1 targets were also broadly expressed across adult human tissues (62.2% of known SETBP1 targets had a TPM>3 in >90% of tissues; **Fig. 3**).

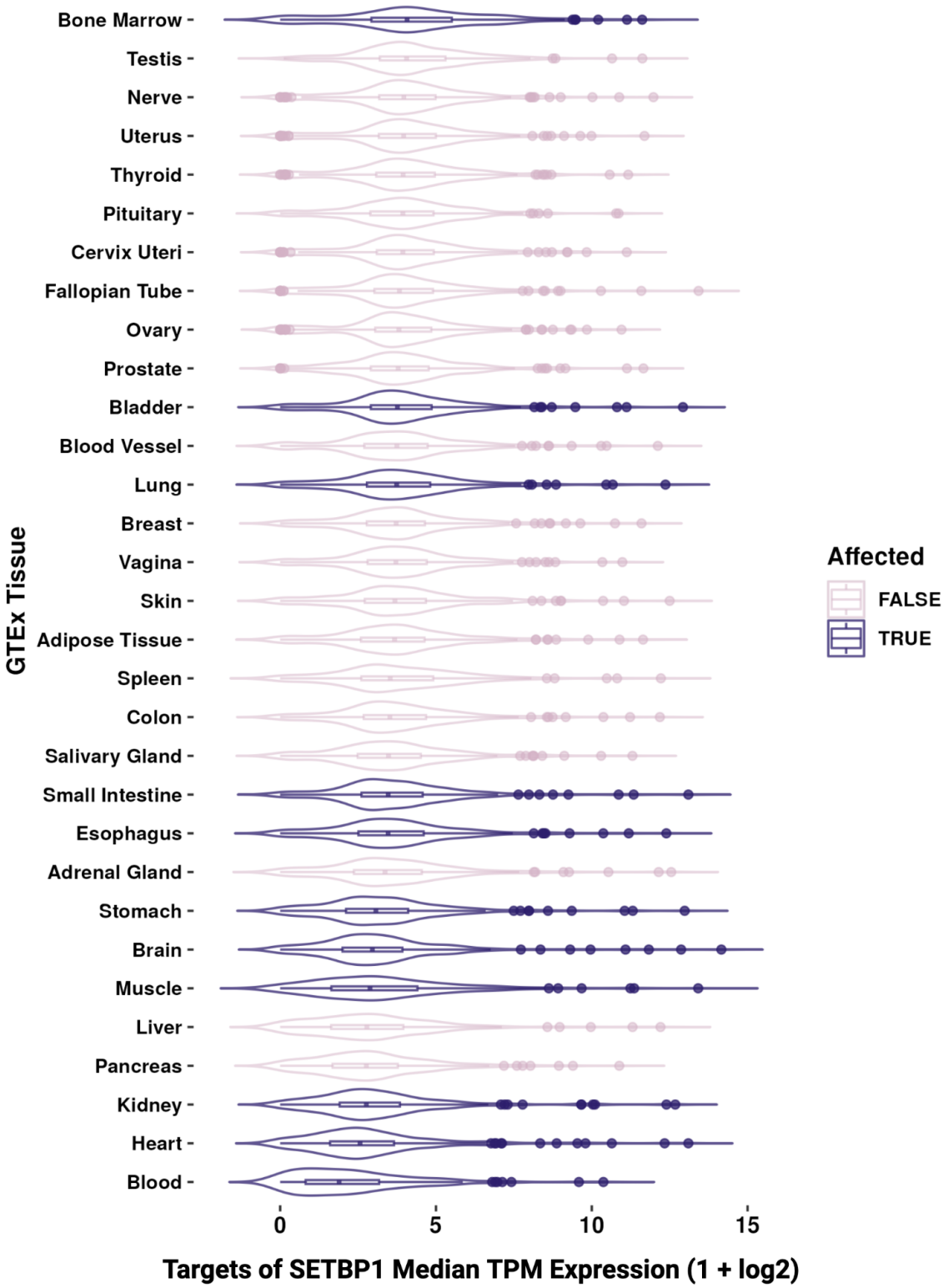

Figure 3 Gene expression of known SETBP1 targets across GTEx tissues. Boxplot of the scaled median TPM values (x-axis) across samples for each GTEx tissue (y-axis) for known

targets of SETBP1 where affected tissues are colored by dark purple and pink for true and false, respectively.

Targets of SETBP1 are widely expressed across tissues and cluster by gene expression into functionally distinct pathways

We next assessed the variability of SETBP1 targets' expression across tissues and if particular functions were enriched based on their co-expression across GTEx tissues. When clustering *SETBP1* targets by gene expression, we selected three k-means clusters by both the elbow and Trace(W) method (**Fig. S1, Table S2**) and found a higher correlation of expression between *SETBP1* and its targets in clusters 1 and 2 than in cluster 3 (rho ranges -0.25 to 0.82, -0.28 to 0.84, and - 0.46 to 0.69, respectively) **(Table S2)**. The targets whose gene expression most correlated to *SETBP1* included *ZEB1* (cluster 2, rho 0.84, p-value = 4.63e-9), *TTC23* (cluster 2, rho 0.82, p-value = 1.14e-8), *RALGAPA1* (cluster 1, rho 0.82, p-value = 1.77e-8), *RAD52* (cluster 2, rho 0.82, p-value = 2.13), *CZIB* (cluster 2, rho 0.81, p-value = 3.07e-8), and *PBRM1* (cluster 2, rho 0.77, p-value = 3.54e-7). The most anti-correlated genes were the mitochondrial-associated genes *MT-RNR1* and *MT-TF* (cluster 3, rho -0.462 and -0.459, p-value = 0.89e-2 and 0.94e-2, respectively). Complete linkage hierarchical clustering of all tissues revealed one clade that included predominantly SETBP1-affected tissues (5 out of 7), including brain, muscle, kidney, heart, and blood (**Fig. 4A**). We further analyzed if the SETBP1 gene targets within these k-means clusters had similar functions using over-representation analysis and found that clusters 1, 2, and 3 included genes enriched for the transcription regulator complex, minor groove AT-rich DNA binding, and mitochondrial structure and function, respectively (**Fig. 4B**).

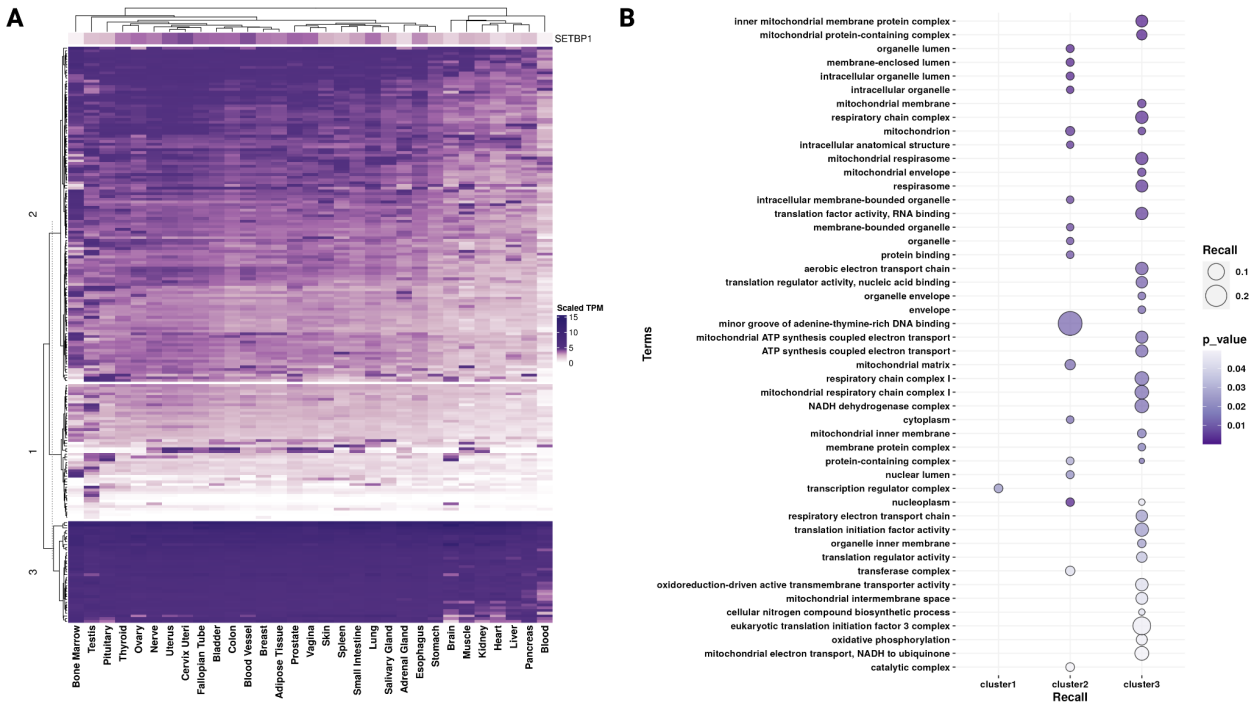

Figure 4 Gene expression of *SETBP1* and its known TF targets across GTEx tissues. (A) Heatmap clustering (y-axis) *SETBP1* and its known transcription factor (TF) targets' scaled TPM gene expression by tissue (x-axis). **(B)** Dot plot representing the functional enrichment analysis (FEA) results for genes from clusters 1, 2, and 3 identified in A. Here dot size indicates recall, the proportion of functionally annotated input genes to each term's full geneset size (y-axis), and color indicates significance (darker purple; more significant).

SETBP1 TF activity is decreased in blood and increased in the pituitary

Due to its role as a TF, we next inferred the TF activity of SETBP1 across GTEx tissues. We did this by extracting the direction of TF regulation for each tissue based on prior knowledge in CollecTRI, a comprehensive, species-specific, curated database of TFs and their transcriptional targets. Using decoupleR, we then built a multivariate linear model taking into account the weight and sign of TF-target interactions and the GTEx gene expression values, where for each tissue, we interpreted positive scores as indicating SETBP1 acting as an activator, negative scores as indicating SETBP1 acting as a repressor, and zero as indicating a lack of coordinated regulation by SETBP1 **(Fig. 5)**. We found that across tissues, SETBP1 TF activity was mostly near zero (**Table S1**; range: -0.739 to 3.760 with median = -0.024 and variance = 0.53), except pituitary (TF activity = 3.760 indicating SETBP1 is functioning as an activator) and blood (TF activity = -0.739 indicating a repressive effect by SETBP1). Except for blood, we did not find that tissues known to be affected by SETBP1-associated diseases had strong evidence of SETBP1 activator or repressor roles in non-diseased adult tissues.

SETBP1 TF targets demonstrate tissue-specific TF activity

In addition to SETBP1, we also inferred TF activity for targets of SETBP1 that act as TFs (**Fig. 5**). We found that these TFs, in addition to SETBP1, had a lack of coordinated TF activity in the

brain, heart, and kidney, all of which are SETBP1-HD or SGS affected tissues (median = -0.004, variance = 0.008 for brain; median = -0.034, variance = 0.014 for heart; and median = -0.026, variance = 0.022 for kidney). However, SETBP1 targets associated with cancer and DNA damage had variable TF activity in the blood. For example, we predicted BCL11A (linked to multiple blood cancers as reviewed in (Bao, Cheng & Sankaran, 2019; Yin et al., 2019)) acts as a repressor in both blood and bone marrow under non-diseased conditions (TF activity -4.124 and -3.287, respectively). MEF2C (often deregulated and associated with recurrence in leukemia (Brown et al., 2018)) and ZEB1 (shown to modulate hematopoietic stem cell fates (Almotiri et al., 2021; Li et al., 2021)) also act as repressors in the blood (TF activity -0.882 and -1.532, respectively).

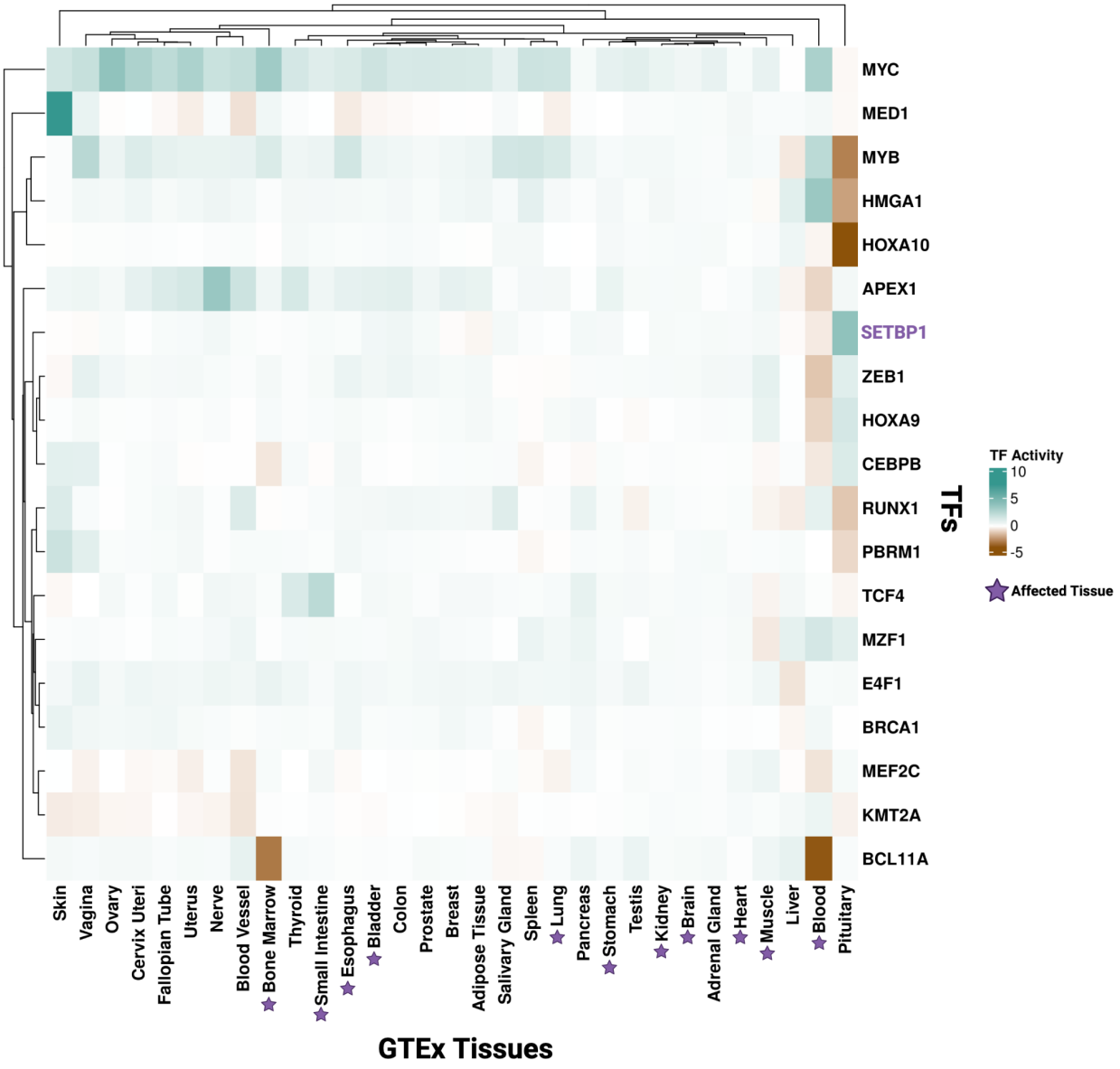

Figure 5 Tissue-specific TF activity of SETBP1 and TFs it targets. Heatmap representing transcription factor (TF) activity scores of SETBP1 (purple) and its known TF targets (black, y-axis) across GTEx tissues (x-axis). Teal and brown represent TF activity as an activator and repressor, respectively.

Furthermore, prior research showed that oncogenic apoptosis resistance and unresolved DNA damage signatures persist due to *SETBP1* variants in SGS (Banfi et al., 2021; Whitlock et al., 2023). Here, we found that in non-diseased adult tissue, TCF4, a known apoptosis regulator (Xie et al., 2012; Forrest et al., 2013), and APEX1, part of the SET complex in DNA damage response (Antonyan & Ernst, 2022), also had variable TF activity across SGS-affected tissues. We predicted TCF4 to function as a repressor in muscle (TF activity = -0.553) and have an activator role in the small intestine (TF activity = 2.181) (**Fig. 5**). Additionally we found APEX1 had predicted repressive activity in the blood (TF activity = -1.810), but was a predicted activator in the bladder (TF activity = 0.694) and esophagus (TF activity = 0.590) (**Fig. 5**). Furthermore, we identified a lack of coordinated activity in kidney, blood, bone marrow, esophagus and muscle for TCF4 and a lack of coordinated activity in blood, muscle, and heart for APEX1 (**Fig. 5**). In summary, we have provided a map of SETBP1 and TF targets of SETBP1 activity across non-diseased adult human tissues.

Discussion

Our study maps tissue-specific molecular signatures associated with SETBP1 across 31 human non-diseased adult tissues. We found that *SETBP1* and its known targets were widely expressed across those tissues, and FEA revealed gene sets related to transcriptional regulation, DNA binding, and mitochondrial function. Further, we uncovered tissue-specific TF activity through TF activity analysis of SETBP1 and its TF targets, underscoring the role of tissue context-driven regulation that may serve to generate hypotheses regarding the importance of TF activity in disease contexts. We also provide a framework for investigating tissue-specific gene expression and TF activity for other genes, particularly those associated with multiple diseases or multi-syndromic diseases with pathophysiological impacts across multiple tissues. As many genes associated with developmental disorders are also associated with a predisposition or increased risk of cancer (reviewed in (Nussinov, Tsai & Jang, 2022; Balachandran & Narendran, 2023)), applying this framework to those genes may be fruitful. To that end, we developed a Shiny web application with pre-computed TF activity scores that we inferred for 758 TFs for each of the 31 GTEx tissues. This application allows users to search for TFs and compare their activity across all 31 GTEx tissues (https://lasseignelab.shinyapps.io/gtex_tf_activity/).

With complete-linkage hierarchical clustering of SETBP1 target gene expression by tissue, we discovered one clade largely consisted of affected tissues (5 out of 7 of the tissues in the clade), including brain, kidney, heart, and blood, the most frequently noted tissues impacted by SETBP1 perturbations (**Fig. 4A**) (Schinzel & Giedion, 1978; Matsumoto et al., 2005; Acuna-Hidalgo et al., 2017; Coccaro et al., 2017). These results suggest that, in non-diseased tissues, there is a similarity in the expression of SETBP1 targets across multiple tissues known to exhibit a

phenotype in SETBP1-associated diseases. We further investigated the trends of expression of SETBP1 targets by k-means clustering and the correlation of each target's expression to *SETBP1* gene expression (**Fig. 4A**, **Table S2**). We compared the total within-cluster sum of squares ("elbow" method) as well as Trace(W) method to select 3 k-means clusters (Charrad et al. 2014; Milligan and Cooper 1985). We highlighted the highest correlated and anti-correlated genes associated with SETBP1's roles as an activator or a repressor (Coccaro et al., 2017). Of note, some of the most significantly correlated and anti-correlated targets included genes with known critical roles in development (*ZEB1,* p-value = 4.63e-9(Drápela et al., 2020) and *RALGAPA1,* p-value = 1.77e-8 (Wagner et al., 2020)), with additional involvement in metastasis and therapy resistance for *ZEB1* (Drápela et al., 2020). Among the non-significant highest correlated and anti-correlated genes were *HMGA1* (p-value = 0.13) and *MYB* (p-value = 0.17). Their non-significance in correlation analyses to SETBP1 across tissues may highlight a potential context dependence or tissue-specific regulatory roles. In addition to having functions in organism development, HMGA1 enhances recovery from double-stranded DNA breaks. When overexpressed, it sensitizes cells to DNA damage and is a driver of malignant tumors (Fujikane et al. 2016). Additionally, SETBP1's consensus binding site largely overlaps with the AT-hook consensus motif of HMGA1 (Piazza et al. 2018). On the other hand, *MYB* expression is known to be critical for myeloid leukemia induced by SETBP1 activation, and its inhibition could be beneficial for treating SETBP1-associated neoplasms (Coccaro et al. 2017; Nguyen et al. 2016). The genes we identified here highlight previously known mechanisms underlying SETBP1-associated disease and provide additional potential targets for future investigation.

To uncover potential functional patterns by expression clusters, we also tested each of our 3 k-means clusters of gene targets for over-representation of GO terms where recall represented the proportion of functionally annotated input genes to each term's full gene set size (**Fig. 4B**). We found cluster 3 genes were ubiquitously and highly expressed across tissues. These genes were enriched for terms regarding mitochondrial structure and function (**Fig. 4B**) and included many mitochondrially-encoded *SETBP1* targets (*MT-ND6*, *MT-RNR1*, *MT-TE*, *MT-TP*, *MTCO3P12*, *MT-TF*) (**Table S2**). Cluster 2 exhibited the highest recall with GO terms enriched for minor groove AT-rich DNA binding (**Fig. 4B**), similar to regulation by SETBP1 as well (Piazza et al., 2018). Cluster 1 genes enrichment was for just one GO term, transcription regulator complex (**Fig. 4B**), and included genes for the cancer-associated transcription factors *RUNX1*, *HOXA9*, and *MYB* (Bhagwat & Vakoc, 2015). These results further highlight the varying roles of SETBP1 targets and their expression patterns across tissues.

Out of all SETBP1 disease-associated tissues, blood was the only tissue exhibiting notable SETBP1 TF repression. Previous studies support that surpassing a higher functional threshold (i.e., more damaging or impacting variants) is required for SETBP1-driven cancer (Acuna-Hidalgo et al., 2017). Our findings here with respect to SETBP1 in non-diseased blood, suggest that GoF variants (already known to drive blood-associated cancers such as myeloid leukemia (Albano et al., 2012; Piazza et al., 2013; Acuna-Hidalgo et al., 2017)) may impact SETBP1's TF repressive role, suggesting future research directions that may shed light on the mechanism behind SETBP1-associated blood cancers. For example, SETBP1 has been shown to activate *MYC (Carratt et al., 2022)*, which we calculated has a TF activity score of 2.77 in blood (activator role) and positive TF activity scores in 27 out of 30 other tissues (**Fig. 5**). Likewise, we and others (Banfi et al., 2021; Antonyan & Ernst, 2022; Whitlock et al., 2023) have hypothesized

that in the presence of pathogenic germline variants, alterations in TCF4 or APEX1 activity lead to apoptosis resistance or increased DNA damage in SGS. Our results suggest both exhibit activator roles across many tissues in non-diseased adult tissues, so disease-associated perturbations may impact SETBP1 TF activity, contributing to SGS.

Our study relies on the assumption that TF protein activity can be inferred by the weighted mode of regulation and transcript levels of its target genes (Badia-I-Mompel et al. 2022). A major limitation of this study is that we conducted analyses in adult bulk expression profiles. For disease-associated genes with developmental phenotypes like *SETBP1*, temporal expression (Cardoso-Moreira et al., 2019, 2020) and TF activity are likely key to linking disrupted genes to molecular and physiological phenotypes. Comprehensive prenatal and postnatal gene expression atlases in active development (e.g., developmental GTEx, dGTEx) will provide an unprecedented opportunity to repeat and expand the analyses in this study across developmental time points. Additionally, as we conducted the analyses here in bulk profiles, we cannot assess the gene expression and TF activity of particular cell types. However, as we recently reported, SETBP1's role as an epigenetic hub leads to cell-type-specific differences in TF activity, gene targeting, and regulatory rewiring in the mouse cerebral cortex and kidney (Whitlock et al., 2023). This underscores the importance of future studies that generate and analyze the necessary data to understand cell-type-specific gene expression and TF activity across human tissues. Furthermore, some affected tissues related to the skeleton and eyes are not included within GTEx v8. If this data becomes available, future studies could investigate tissue-specific expression and activity for vision loss, craniofacial, and skeletal abnormalities within SGS and SETBP1-HD. Finally, the GTEx tissues may have subclinical pathologies that have not been previously reported, so care must be taken with the interpretation of non-affected tissues.

Conclusions

In summary, our study highlights the importance of considering tissue-specific expression and regulatory properties in investigating disease-related genes. It provides a basis for future investigations of TFs involved in processes across many tissues, including developmental and cancer contexts.

Acknowledgements

The authors thank the Lasseigne Lab members Vishal Oza, Tabea Soelter, Emma Jones, and Victoria Flanary for their feedback throughout this study. We also thank the UAB Biological Data Science group (RRID:SCR_021766) for providing a script for helping to run containers on the UAB high-performance cluster (https://github.com/U-BDS/training_guides/blob/main/run_rstudio_singularity.sh)

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported in part by the UAB Lasseigne Lab funds, UAB Pilot Center for Precision Animal Modeling (C-PAM)(1U54OD030167)(to BNL), and the UAB Predoctoral Training Grant in Cell, Molecular, and Developmental Biology (CMDB T32)(5T32GM008111-35)(to JHW).

Authors' contributions

Conceptualization: JHW, EJW, TCH, ADC, and BNL. Methodology: JHW and BNL. Software: JHW (including the Shiny app) and EJW. Validation: JHW, EJW, TCH, and ADC. Formal Analysis: JHW and EJW. Investigation: JHW, EJW. Resources: BNL. Data Curation: JHW. Writing - Original Draft: JHW. Writing - Review & Editing: JHW, EJW, TCH, ADC, and BNL. Visualization: JHW and EJW. Supervision: BNL. Project Administration: BNL. Funding Acquisition: BNL.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing Interests

The authors declare that they have no competing interests

References

Acuna-Hidalgo R, Deriziotis P, Steehouwer M, Gilissen C, Graham SA, van Dam S, Hoover-Fong J, Telegrafi AB, Destree A, Smigiel R, Lambie LA, Kayserili H, Altunoglu U, Lapi E, Uzielli ML, Aracena M, Nur BG, Mihci E, Moreira LMA, Borges Ferreira V, Horovitz DDG, da Rocha KM, Jezela-Stanek A, Brooks AS, Reutter H, Cohen JS, Fatemi A, Smitka M, Grebe TA, Di Donato N, Deshpande C, Vandersteen A, Marques Lourenço C, Dufke A, Rossier E, Andre G, Baumer A, Spencer C, McGaughran J, Franke L, Veltman JA, De Vries BBA, Schinzel A, Fisher SE, Hoischen A, van Bon BW. 2017. Overlapping SETBP1 gain-of-function mutations in Schinzel-Giedion syndrome and hematologic malignancies. *PLoS genetics* 13:e1006683. DOI: 10.1371/journal.pgen.1006683.

Albano F, Anelli L, Zagaria A, Coccaro N, Casieri P, Minervini A, Specchia G. 2012. SETBP1 and miR_4319 dysregulation in primary myelofibrosis progression to acute myeloid leukemia. *Journal of hematology & oncology* 5:48. DOI: 10.1186/1756-8722-5-48.

Almotiri A, Alzahrani H, Menendez-Gonzalez JB, Abdelfattah A, Alotaibi B, Saleh L, Greene A, Georgiou M, Gibbs A, Alsayari A, Taha S, Thomas L-A, Shah D, Edkins S, Giles P, Stemmler MP, Brabletz S, Brabletz T, Boyd AS, Siebzehnrubl FA, Rodrigues NP. 2021. Zeb1 modulates hematopoietic stem cell fates required for suppressing acute myeloid leukemia. *The Journal of clinical investigation* 131. DOI: 10.1172/JCI129115.

Amberger JS, Bocchini CA, Scott AF, Hamosh A. 2019. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic acids research* 47:D1038–D1043. DOI: 10.1093/nar/gky1151.

Antonyan L, Ernst C. 2022. Putative Roles of SETBP1 Dosage on the SET Oncogene to Affect Brain Development. *Frontiers in neuroscience* 16:813430. DOI: 10.3389/fnins.2022.813430.

Badia-I-Mompel P, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, Taus P, Dugourd A, Holland CH, Ramirez Flores RO, Saez-Rodriguez J. 2022. decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics advances* 2:vbac016. DOI: 10.1093/bioadv/vbac016.

Balachandran S, Narendran A. 2023. The Developmental Origins of Cancer: A Review of the Genes Expressed in Embryonic Cells with Implications for Tumorigenesis. *Genes* 14. DOI: 10.3390/genes14030604.

Banfi F, Rubio A, Zaghi M, Massimino L, Fagnocchi G, Bellini E, Luoni M, Cancellieri C, Bagliani A, Di Resta C, Maffezzini C, Ianielli A, Ferrari M, Piazza R, Mologni L, Broccoli V, Sessa A. 2021. SETBP1 accumulation induces P53 inhibition and genotoxic stress in neural progenitors underlying neurodegeneration in Schinzel-Giedion syndrome. *Nature communications* 12:4050. DOI: 10.1038/s41467-021-24391-3.

Bao EL, Cheng AN, Sankaran VG. 2019. The genetics of human hematopoiesis and its disruption in disease. *EMBO molecular medicine* 11:e10316. DOI: 10.15252/emmm.201910316.

Bhagwat AS, Vakoc CR. 2015. Targeting Transcription Factors in Cancer. *Trends in cancer research* 1:53–65. DOI: 10.1016/j.trecan.2015.07.001.

Brown FC, Still E, Koche RP, Yim CY, Takao S, Cifani P, Reed C, Gunasekera S, Ficarro SB, Romanienko P, Mark W, McCarthy C, de Stanchina E, Gonen M, Seshan V, Bhola P, O'Donnell C, Spitzer B, Stutzke C, Lavallée V-P, Hébert J, Krivtsov AV, Melnick A, Paietta EM, Tallman MS, Letai A, Sauvageau G, Pouliot G, Levine R, Marto JA, Armstrong SA, Kentsis A. 2018. MEF2C Phosphorylation Is Required for Chemotherapy Resistance in

Acute Myeloid Leukemia. *Cancer discovery* 8:478–497. DOI: 10.1158/2159-8290.CD-17-1271.

Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, Liechti A, Ascenção K, Rummel C, Ovchinnikova S, Mazin PV, Xenarios I, Harshman K, Mort M, Cooper DN, Sandi C, Soares MJ, Ferreira PG, Afonso S, Carneiro M, Turner JMA, VandeBerg JL, Fallahshahroudi A, Jensen P, Behr R, Lisgo S, Lindsay S, Khaitovich P, Huber W, Baker J, Anders S, Zhang YE, Kaessmann H. 2019. Gene expression across mammalian organ development. *Nature* 571:505–509. DOI: 10.1038/s41586-019-1338-5.

Cardoso-Moreira M, Sarropoulos I, Velten B, Mort M, Cooper DN, Huber W, Kaessmann H. 2020. Developmental Gene Expression Differences between Humans and Mammalian Models. *Cell reports* 33:108308. DOI: 10.1016/j.celrep.2020.108308.

Carratt SA, Kong GL, Curtiss BM, Schonrock Z, Maloney L, Maniaci BN, Blaylock HZ, Baris A, Druker BJ, Braun TP, Maxson JE. 2022. Mutated SETBP1 activates transcription of Myc programs to accelerate CSF3R-driven myeloproliferative neoplasms. *Blood* 140:644–658. DOI: 10.1182/blood.2021014777.

Chang X, Liu J, Yang Q, Gao Y, Ding X, Zhao J, Li Y, Liu Z, Li Z, Wu Y, Zuo D. 2023. Targeting HMGA1 contributes to immunotherapy in aggressive breast cancer while suppressing EMT. *Biochemical pharmacology* 212:115582. DOI: 10.1016/j.bcp.2023.115582.

Cicirò Y, Sala A. 2021. MYB oncoproteins: emerging players and potential therapeutic targets in human cancer. *Oncogenesis* 10:19. DOI: 10.1038/s41389-021-00309-y.

Coccaro N, Tota G, Zagaria A, Anelli L, Specchia G, Albano F. 2017. SETBP1 dysregulation in congenital disorders and myeloid neoplasms. *Oncotarget* 8:51920–51935. DOI: 10.18632/oncotarget.17231.

Drápela S, Bouchal J, Jolly MK, Culig Z, Souček K. 2020. ZEB1: A Critical Regulator of Cell Plasticity, DNA Damage Response, and Therapy Resistance. *Frontiers in molecular biosciences* 7:36. DOI: 10.3389/fmolb.2020.00036.

Elena C, Gallì A, Such E, Meggendorfer M, Germing U, Rizzo E, Cervera J, Molteni E, Fasan A, Schuler E, Ambaglio I, Lopez-Pavia M, Zibellini S, Kuendgen A, Travaglino E, Sancho-Tello R, Catricalà S, Vicente AI, Haferlach T, Haferlach C, Sanz GF, Malcovati L, Cazzola M. 2016. Integrating clinical features and genetic lesions in the risk assessment of patients with chronic myelomonocytic leukemia. *Blood* 128:1408–1417. DOI: 10.1182/blood-2016-05-714030.

Forrest MP, Waite AJ, Martin-Rendon E, Blake DJ. 2013. Knockdown of human TCF4 affects multiple signaling pathways involved in cell survival, epithelial to mesenchymal transition and neuronal differentiation. *PloS one* 8:e73169. DOI: 10.1371/journal.pone.0073169.

Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32:2847–2849. DOI: 10.1093/bioinformatics/btw313.

Han H, Cho J-W, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E, Lee S, Kang B, Jeong D, Kim Y, Jeon H-N, Jung H, Nam S, Chung M, Kim J-H, Lee I. 2018. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic acids research* 46:D380–D386. DOI: 10.1093/nar/gkx1013.

Hekselman I, Yeger-Lotem E. 2020. Mechanisms of tissue and cell-type specificity in heritable traits and diseases. *Nature reviews. Genetics* 21:137–150. DOI: 10.1038/s41576-019-0200-9.

Jansen NA, Braden RO, Srivastava S, Otness EF, Lesca G, Rossi M, Nizon M, Bernier RA, Quelin C, van Haeringen A, Kleefstra T, Wong MMK, Whalen S, Fisher SE, Morgan AT, van Bon BW. 2021. Clinical delineation of SETBP1 haploinsufficiency disorder. *European journal of human genetics: EJHG* 29:1198–1205. DOI: 10.1038/s41431-021-00888-9.

Kohyanagi N, Ohama T. 2023. The impact of SETBP1 mutations in neurological diseases and cancer. *Genes to cells: devoted to molecular & cellular mechanisms*. DOI: 10.1111/gtc.13057.

Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. 2020. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research* 9. DOI: 10.12688/f1000research.24956.2.

Kolmykov S, Yevshin I, Kulyashov M, Sharipov R, Kondrakhin Y, Makeev VJ, Kulakovskiy IV, Kel A, Kolpakov F. 2021. GTRD: an integrated view of transcription regulation. *Nucleic acids research* 49:D104–D111. DOI: 10.1093/nar/gkaa1057.

Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems* 1:417–425. DOI: 10.1016/j.cels.2015.12.004.

Li L, Feng Y, Hu S, Du Y, Xu X, Zhang M, Peng X, Chen F. 2021. ZEB1 serves as an oncogene in acute myeloid leukaemia via regulating the PTEN/PI3K/AKT signalling pathway by combining with P53. *Journal of cellular and molecular medicine* 25:5295–5304. DOI: 10.1111/jcmm.16539.

Liu W-L, He Z-X, Li F, Ai R, Ma H-W. 2018. Schinzel–Giedion syndrome: a novel case, review and revised diagnostic criteria. *Journal of genetics* 97:35–46. DOI: 10.1007/s12041-017-0877-5.

Lo Surdo P, Iannuccelli M, Contino S, Castagnoli L, Licata L, Cesareni G, Perfetto L. 2023. SIGNOR 3.0, the SIGnaling network open resource 3.0: 2022 update. *Nucleic acids research* 51:D631–D637. DOI: 10.1093/nar/gkac883.

Matsumoto F, Tohda A, Shimada K, Okamoto N. 2005. Malignant retroperitoneal tumor arising in a multicystic dysplastic kidney of a girl with Schinzel-Giedion syndrome. *International journal of urology: official journal of the Japanese Urological Association* 12:1061–1062. DOI: 10.1111/j.1442-2042.2005.01213.x.

Morgan A, Srivastava S, Duis J, van Bon B. 2021. SETBP1 Haploinsufficiency Disorder. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Gripp KW, Mirzaa GM, Amemiya A eds. *GeneReviews®*. Seattle (WA): University of Washington, Seattle,.

Müller-Dott S, Tsirvouli E, Vázquez M, Ramirez Flores RO, Badia-i-Mompel P, Fallegger R, Lægreid A, Saez-Rodriguez J. 2023. Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. *bioRxiv*:2023.03.30.534849. DOI: 10.1101/2023.03.30.534849.

Nussinov R, Tsai C-J, Jang H. 2022. How can same-gene mutations promote both cancer and developmental disorders? *Science advances* 8:eabm2059. DOI: 10.1126/sciadv.abm2059.

Patnaik MM, Itzykson R, Lasho TL, Kosmider O, Finke CM, Hanson CA, Knudson RA, Ketterling RP, Tefferi A, Solary E. 2014. ASXL1 and SETBP1 mutations and their prognostic contribution in chronic myelomonocytic leukemia: a two-center study of 466 patients. *Leukemia* 28:2206–2212. DOI: 10.1038/leu.2014.125.

Piazza R, Magistroni V, Redaelli S, Mauri M, Massimino L, Sessa A, Peronaci M, Lalowski M, Soliymani R, Mezzatesta C, Pirola A, Banfi F, Rubio A, Rea D, Stagno F, Usala E, Martino B, Campiotti L, Merli M, Passamonti F, Onida F, Morotti A, Pavesi F, Bregni M, Broccoli V, Baumann M, Gambacorti-Passerini C. 2018. SETBP1 induces transcription of a network of development genes by acting as an epigenetic hub. *Nature communications* 9:2192. DOI: 10.1038/s41467-018-04462-8.

Piazza R, Valletta S, Winkelmann N, Redaelli S, Spinelli R, Pirola A, Antolini L, Mologni L, Donadoni C, Papaemmanuil E, Schnittger S, Kim D-W, Boultwood J, Rossi F, Gaipa G, De Martini GP, di Celle PF, Jang HG, Fantin V, Bignell GR, Magistroni V, Haferlach T, Pogliani EM, Campbell PJ, Chase AJ, Tapper WJ, Cross NCP, Gambacorti-Passerini C. 2013. Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. *Nature genetics* 45:18–24. DOI: 10.1038/ng.2495.

Reimand J, Kull M, Peterson H, Hansen J, Vilo J. 2007. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic acids research* 35:W193–200. DOI: 10.1093/nar/gkm226.

Schinzel A, Giedion A. 1978. A syndrome of severe midface retraction, multiple skull anomalies, clubfeet, and cardiac and renal malformations in sibs. *American journal of medical genetics* 1:361–375. DOI: 10.1002/ajmg.1320010402.

Schinzel Giedion Syndrome - NORD (National Organization for Rare Disorders). 2015. *Available at https://rarediseases.org/rare-diseases/schinzel-giedion-syndrome/* (accessed November 15, 2022).

Wagner M, Skorobogatko Y, Pode-Shakked B, Powell CM, Alhaddad B, Seibt A, Barel O, Heimer G, Hoffmann C, Demmer LA, Perilla-Young Y, Remke M, Wieczorek D,

Navaratnarajah T, Lichtner P, Klee D, Shamseldin HE, Al Mutairi F, Mayatepek E, Strom T, Meitinger T, Alkuraya FS, Anikster Y, Saltiel AR, Distelmaier F. 2020. Bi-allelic Variants in RALGAPA1 Cause Profound Neurodevelopmental Disability, Muscular Hypotonia, Infantile Spasms, and Feeding Abnormalities. *American journal of human genetics* 106:246–255. DOI: 10.1016/j.ajhg.2020.01.002.

Whitlock JH, Soelter TM, Howton TC, Wilk EJ, Oza VH, Lasseigne BN. 2023. Cell-type-specific gene expression and regulation in the cerebral cortex and kidney of atypical Setbp1S858R Schinzel Giedion Syndrome mice. *bioRxiv*:2023.07.31.551338. DOI: 10.1101/2023.07.31.551338.

Wilks C, Zheng SC, Chen FY, Charles R, Solomon B, Ling JP, Imada EL, Zhang D, Joseph L, Leek JT, Jaffe AE, Nellore A, Collado-Torres L, Hansen KD, Langmead B. 2021. recount3: summaries and queries for large-scale RNA-seq expression and splicing. *Genome biology* 22:323. DOI: 10.1186/s13059-021-02533-6.

Xie J, Xiang D-B, Wang H, Zhao C, Chen J, Xiong F, Li T-Y, Wang X-L. 2012. Inhibition of Tcf-4 induces apoptosis and enhances chemosensitivity of colon cancer cells. *PloS one* 7:e45617. DOI: 10.1371/journal.pone.0045617.

Yin J, Xie X, Ye Y, Wang L, Che F. 2019. BCL11A: a potential diagnostic biomarker and therapeutic target in human diseases. *Bioscience reports* 39. DOI: 10.1042/BSR20190604.