



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

CENTRO DE CIENCIAS GENÓMICAS
INSTITUTO DE BIOTECNOLOGÍA

DIFFERENTIAL ACTIVITY OF TRANSCRIPTIONAL
REGULATORY ELEMENTS DURING GLUCOSE STARVATION

REPORTE DE INVESTIGACIÓN

QUE PARA OBTENER EL TÍTULO DE:

LICENCIADO EN CIENCIAS GENÓMICAS

PRESENTA:

ROBERTO OLAYO ALARCÓN

TUTORES:

DR. SAULIUS LUKAUSKAS
DR. ROBERT SCHNEIDER



Cuernavaca, Morelos, 2020



Instituto de Biotecnología
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Contents

1	Introduction	2
1.1	Transcription and Transcriptional Regulatory Elements	2
1.2	Chromatin Accessibility	3
1.3	Histone Modifications	4
1.4	Automatic enhancer annotation and linkage to target genes	5
1.5	Regulatory response to glucose starvation	7
2	Justification	8
3	Aims	9
4	Hypothesis	9
5	Results	10
5.1	Transcriptional response to glucose starvation	10
5.2	Differential Accessibilty detected in ATAC-seq	10
5.2.1	Normalization of ATAC-seq datasets	12
5.2.2	Differentially Accessibility analysis	16
5.3	Correlation between H3K27ac ChIP-seq, ATAC-seq and Gene Expression	19
5.4	Adaption of Activity by Contact for condition comparisons	23
5.5	Additional filters to infer gene-enhancer relationships	24
5.6	Distinct enhancer responses to glucose starvation	32
5.7	Differentially Accessible Enhancers	35
5.8	Discussion	39
6	Methods	41
6.1	Cell Culture	41
6.2	Glucose Starvatino	41
6.3	RNA-seq	42
6.4	Assay for Transposase-Accessible Chromatin using sequencing	42
6.5	Chromatin Immunoprecipitation	44
6.6	Activity by Contact	46
6.7	Term Enrichment Analysis	47
6.8	Software	47
References		47

1 Introduction

In order to survive, organisms must be able to overcome acute environmental stress. At the cellular level, transcriptional programs must be adapted to favor gene products that can satisfy the physiological needs that the cell's circumstances demand. The study of these stress-responsive expression programs has revealed that they are fine-tuned (De Nadal, Ammerer, & Posas, 2011). Eukaryotic cells in particular, possess several mechanisms that can give rise to complex transcriptional behaviors. Individual examples have revealed that transcriptional regulatory elements, such as promoters and enhancers, play a key role in the coordination of the transcriptional response to diverse environmental stressors such as temperature changes, osmolarity imbalances, and the variable availability of nutrients, etc. (De Nadal et al., 2011).

1.1 Transcription and Transcriptional Regulatory Elements

Transcription is a highly regulated process by which a DNA sequence is used as a template to create a new RNA molecule. In the case of protein coding genes, the DNA sequence is transcribed into a messenger RNA molecule (mRNA) which is then recognized by ribosomal complexes as instructions to assemble an amino acid chain. In eukaryotes, the transcription of protein coding genes is carried out by the enzyme RNA polymerase II (RNAP II).

Through several years of research, many experimental protocols to measure gene expression have been developed. Of these, RNA-seq has become a popular experiment to detect the presence of, and quantify, RNA molecules at genome-wide scale (Stark, Grzelak, & Hadfield, 2019). To achieve this, RNA is first purified from a biological sample. After RNA isolation, several variations of this technique can take place; if one is interested in detecting and quantifying protein coding transcripts, then mRNA molecules can be further enriched by selecting RNA with a polyadenylated sequence at their 3' end. Once the desired RNA molecules have been selected, they are reverse transcribed into complementary DNA molecules (cDNA), which can then be sequenced. By quantifying the frequency with which a nucleotide sequence is present in a biological sample, one can obtain an estimate of abundance of said transcript. Comparing transcript abundances between samples from different biological contexts (different growth conditions, cell-types, diseased tissues, etc.), is a common use for these types of data, as it reveals the cellular processes that are active or inactive in the compared contexts.

In order for transcription to take place, RNAP II must first be recruited to a gene's transcription start site (TSS). This is often carried out by a group of proteins collectively known as transcription factors (TFs). These proteins typically bind directly to the DNA molecule, upstream from TSS, in a genomic region known as the promoter. TFs recognize specific nucleotide sequences (sequence motifs) present in the promoter region and bind to them. Once bound, TFs recruit co-activator molecules, as well as RNAP II, to the gene promoter to a structure known as the pre-initiation complex (PIC). Once this complex is formed, transcription can be initiated.

While promoters are sufficient to initiate transcription, the rate of initiation is often aided by distal regulatory elements known as enhancers (Haberle & Stark, 2018; Bell, Tiwari, Thomä, & Schübeler, 2011). Enhancers also contain sequence motifs that can be bound by TFs. The relative location of enhancers, with respect to their target promoter, is variable; they can be found both upstream and downstream of the gene they regulate. While the position and distance in the genomic sequence between enhancers and their target gene varies, they are

consistently found in close spatial proximity within the nucleus (Shlyueva, Stampfel, & Stark, 2014). This is achieved by looping the three-dimensional structure of the DNA molecule. In this way, the TFs that bind to the enhancer sequence can further recruit other proteins to assemble the PIC at their target TSS. Importantly, enhancers are modular and make individual, additive contributions towards the expression of their target gene (Shlyueva et al., 2014).

The spatial proximity of enhancers and promoters can be measured by experimental techniques such as Chromosome Conformation Capture (3C) (Kempfer & Pombo, 2019). The procedure for this assay consists of crosslinking interacting regions of the genome. Once these spatial contacts have been fixed, restriction enzymes are used to fragment the genome, and crosslinked fragments are then ligated. In a variant of the 3C procedure known as Hi-C, the DNA sequence of these ligated fragments is determined using high-throughput sequencing technology. In this way, all ligation events that take place in the genome are captured. By mapping ligated sequences to the genome being studied, the frequency with which any two regions of genome interact can be determined.

As was noted previously, promoters and enhancers play pivotal roles in the regulation of gene transcription. Together, these regulatory elements can have a collective effect of either upregulating, downregulating, or maintaining gene expression levels. The activity of these regions of the genome has often been found to be cell-type and condition specific (Andersson & Sandelin, 2019; Gasperini, Tome, & Shendure, 2020). At the same time, alterations to enhancers and promoters have been linked to developmental defects, congenital diseases and cancerous phenotypes (Shlyueva et al., 2014; Andersson & Sandelin, 2019). Therefore, tight control over these regions of the genome is necessary in order to create and maintain complex and beneficial transcriptional programs within a cell.

1.2 Chromatin Accessibility

The organization of the genome has a direct impact on transcriptional dynamics (Bell et al., 2011). The nuclear genome of eukaryotic cells is organized into basic units called nucleosomes; these units consist of DNA wrapped around an octamer of histone proteins (H3, H4, H2A, and H2B). Collectively, nucleosomes, and other associated proteins, are referred to as chromatin. This form of organization compacts the genetic material and also allows for the regulation of gene expression by controlling access to the DNA sequence of transcriptional regulatory elements. This regulation is achieved because nucleosome occupancy has a direct impact on the binding activities of TFs (Barisic, Stadler, Iurlaro, & Schübeler, 2019). The tight packaging of nucleosomes often acts as a barrier that impedes the binding of TFs to the DNA sequence. Conversely, regions of the genome where nucleosome occupancy is low (nucleosome depleted regions (NDRs)) have an exposed DNA sequence that can be recognized by different regulatory proteins.

Consequently, active promoters and enhancers are characteristically NDRs themselves. The accessible DNA sequence allows TFs to bind to these regulatory elements, thereby enabling transcription to take place. The promoters of transcriptionally inactive genes are often found to be wrapped in nucleosomes (Bell et al., 2011; Andersson & Sandelin, 2019). Given the importance of chromatin accessibility in the regulation of expression, it is not surprising to find that accessible regions of the genome tend to be cell-type specific. Additionally, cells have the ability to dynamically modulate chromatin accessibility in response to stimuli (De Nadal et al., 2011).

Nucleosome occupancy is, in part, regulated by the actions of protein complexes with the ability to move, disassemble or reassemble nucleosomes. When cells are faced with environmental stress, these chromatin remodelers can be recruited to promoter regions and thereby contribute toward the establishment of a stress inducible transcriptional program (De Nadal et al., 2011). As an example, when yeast cells face heat stress chromatin remodeler complexes are recruited to the promoters of genes whose transcription is induced by this environmental stimuli, in order to remove or displace nucleosomes in these regions. These protein complexes are also required to reassemble nucleosomes at promoter regions, once the expression of stress-responsive genes is no longer required (De Nadal et al., 2011).

Advances in DNA sequencing technology have prompted the development of several experimental techniques that can assess genome-wide chromatin accessibility. One such technique, is known as Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq). (Buenrostro, Wu, Chang, & Greenleaf, 2015). Briefly, chromatin from a population of cells is purified and treated with a genetically engineered hyperactive Tn5 transposase. This enzyme has the ability to cut chromatin, while at the same time inserting another DNA sequence to which it is associated. Due to steric hinderance by nucleosomes, Tn5 preferentially carries out its activity at sites of accessible chromatin. This behavior is leveraged in order to introduce sequencing adapters in open chromatin. These tagged DNA fragments of accessible chromatin can then be purified, amplified and sequenced. The sequencing reads produced are finally mapped to the genome sequence of the organism being studied, and a genome-wide profile of chromatin accessibility is determined.

1.3 Histone Modifications

Chemical post-translational modifications of histone proteins (PTMs) also play a role in determining nucleosome positioning and stability. Histones can be reversibly modified by the addition or removal of various chemical groups to many residues of their amino acid side chains. The consequences of these PTMs greatly depends on the exact chemical modification that takes place, as well as the residue where it occurs. The position within the genome of chemically modified histones can be used to identify regulatory elements, and can also indicate the level of activity of said regions.

Promoters whose genes have undergone multiple rounds of transcription can be identified by the presence of H3 histones whose lysine 4 residue are tri-methylated (H3K4me3) in close proximity to the TSS. The continuous transcription of these genes enables the appearance of the H3K4me3 which, in turn, is further recognized by additional activating regulatory proteins (Soares et al., 2017). Actively expressed genes are also characterized by the presence of H3K27 acetylated histones (H3K27ac) at their promoters. H3K27ac also marks active enhancers. While the precise role that this PTM plays in transcriptional regulation is still an active area of research, it has been suggested that it may facilitate chromatin remodelling (Andersson & Sandelin, 2019). H3K27ac has frequently found to be associated with NDRs of active regulatory elements (Andersson & Sandelin, 2019), which in turn facilitates TF binding at enhancers and promoters. Enhancers can also be recognized by the presence of H3K4 monethylated histones (H3K4me1). Both H3K27ac and H3K4me1 are often used to identify enhancers, and are also used to gather a measurement of their activity. As with chromatin accessibility, the addition and removal of PTMs at the site of transcriptional regulatory elements, can also modulated in response to environmental stress.

The gain or loss of certain histone PTMs can mark the differential activity of regulatory elements in response to stress. Enhancers that become active in when the cell is presented with inflammatory agents, for example, have been shown to acquire H3K27ac and H3K4me1 modifications at histones that flank the NDR; furthermore, these activated regions become bound by the Pu.1 TF (Ostuni et al., 2013). In another example, when continuously exposed to drug treatment, an enhancer that regulates the expression of the gene ABCB1, acquires H3K27ac thereby enabling it to up-regulate the expression of said gene. The higher expression of the ABCB1 efflux pump allows acute myloid leukemia cells to become resistant to the same drug treatments that activated its stress-responsive enhancer. The misregulation of PTMs in response to certain environmental stressors leads to aberrant regulatory element activity and can promote the appearance of cancerous phenotypes (Maiques-Diaz et al., 2018).

Much of the current knowledge on histone PTMs has been gathered thanks to the development experimental techniques such as Chromatin Immunoprecipitation (ChIP). In this experiment, DNA associated proteins are crosslinked with the DNA molecule itself, thereby effectively fixing any protein-DNA interactions. This crosslinked chromatin is then sonicated to form smaller fragments. Once these fragments have the desired length, specific protein-DNA interactions are purified by using an antibody that recognizes the protein of interest. The selection of antibodies enables the purification of histones with specific modifications, as well as other proteins such as TFs. Following this purification, the DNA and associated proteins are separated and the recovered DNA fragments are then analyzed; when these are sequenced, the technique is then referred to as Chromatin Immunoprecipitation followed by Sequencing (ChIP-seq). As with ATAC-seq, the resulting sequencing reads are then mapped to the genome, thereby revealing a genome-wide profile of protein location.

Large-scale efforts such as the Encyclopedia of DNA Elements (ENCODE) (Consortium et al., 2012) and the NIH Roadmap Epigenomics Mapping Consortium (Bernstein et al., 2010) aim to establish a set of guidelines to effectively carry out experiments that evaluate the state of chromatin, including ChIP-seq and ATAC-seq. Experiments that have been performed on various cell lines are collected and made available to the public, thereby eliminating the need of repeating costly and time-consuming experiments. The ENCODE consortium has the additional aim of leveraging the collected datasets to create a catalogue of regulatory elements, including enhancers and promoters.

1.4 Automatic enhancer annotation and linkage to target genes

Because of the importance that the interactions between enhancers and promoters have on creating and maintaining transcriptional programs, great efforts have been made towards identifying enhancers and their regulatory targets at a genome-wide scale. The genome-wide identification of putative enhancer regions is often carried out bioinformatically. These automatic algorithms are faced with the challenge that enhancers cannot be recognized by their DNA sequence alone, and that their relative location and distance from their target gene varies. Enhancer annotation pipelines, therefore, make use of a combination of experimental datasets. For example, the ENCODE consortium leverages chromatin accessibility data, H3K4me3, H3K27ac and CTCF (a DNA binding protein) ChIP-seq assays, in order to locate candidate enhancer regions. The annotation pipeline used by ENSEMBL requires additional datasets, such as H3K4me1, and H3K27me3 ChIP-seq experiments, among others. These automatic approaches often yield a catalogue of putative enhancer regions that vastly outnumbers the number of

genes present in the genome that is studied. In some cases, these predicted enhancers fail to show enhancer activity upon experimental validation, leading to some scepticism about the reliability of these methods to accurately recognize these regulatory elements. Experimental validation remains the gold standard to show true enhancer activity (Gasperini et al., 2020).

A further challenge arises when attempting to identify the target gene(s) of a specific enhancer (Moore, Pratt, Purcaro, & Weng, 2020). While regulatory relationships can be confirmed through various experimental techniques, these procedures are often time-consuming and are complicated to scale up to a genome-wide assay. As a consequence, several predictive tools have been developed, with the goal of inferring a complete set of regulatory relationships. The simplest automatic approach consists of linking candidate enhancers to the TSS that is closest to it in the linear genome sequence. This distance-based criteria has major limitations, given that there are several examples of enhancers that have been shown to skip the nearest gene and maintain long-range interactions with its target (Lettice et al., 2003). Furthermore, it has been shown that some enhancers are able to influence the expression of more than one gene (Pennacchio, Bickmore, Dean, Nobrega, & Bejerano, 2013). This obscure nature of gene-enhancer relationships has prompted the further development of predictive algorithms that incorporate additional criteria. One major approach consists on correlating the activity signals of candidate enhancers and promoters across different cell-lines, thereby linking regulatory regions with similar activity profiles across all samples. One of the earliest implementations of this strategy correlated the signal of various histone modification ChIP-seq experiments (including H3K27ac and H3K4me1, among others) at possible enhancer sites, with gene expression (Ernst et al., 2011). More recent methods have relied on training machine learning models with sets of validated enhancer-promoter interactions. These methods also require the combined input of various experimental datasets. A tool known as Join Effect of Multiple Enhancers (JEME), for example, relies on a combination of linear regression and a Random Forest model, in order to classify possible regulatory interactions. JEME determines the validity of candidate enhancer-promoter pairs by considering H3K27ac, H3K4me1, and H3K27me3 ChIP-seq experiments, as well as chromatin accessibility information and gene expression measurements (Cao et al., 2017). Another popular method, named TargetFinder, also relies on an ensemble of Decision Tree classifiers, known as Gradient Boosted Decision Trees. This tool, in addition to the ChIP-seq experiments used by JEME, also requires further information about the DNA binding activities of proteins such as CTCF, as well as measurements of chromatin spatial conformation (Whalen, Truty, & Pollard, 2016).

One of the more recent algorithms developed to link enhancers to their target gene(s) is known as Activity by Contact (ABC) (Fulco et al., 2019). Interestingly, ABC does not require as much experimental data as other methods developed with the same purpose; requiring only H3K27ac ChIP-seq, a chromatin accessibility assay (in this work we utilize ATAC-seq), and a Hi-C dataset. Using this limited input, the ABC algorithm showed great performance when predicting genes whose expression was affected when an enhancer sequence was altered. This method was able to successfully recapitulate a set of experimentally validated enhancer-gene pairs, with a precision of 70% and a recall of 59%. It outperformed previously established tools such as JEME, TargetFinder, and the linear-distance criteria. In addition, the algorithm itself is relatively simple. Briefly, all NDRs within 5 Mbs of a gene's TSS are considered to be potential enhancers that regulate said gene. The activity of each candidate is estimated with the following formula:

$$A_E = \sqrt{ATAC * H3K27ac} \quad (1)$$

Where $ATAC$ represents the number of reads from ATAC-seq that map to the candidate region and $H3K27ac$ represents the number of ChIP-seq reads that overlap that same region. The geometric mean of these two values yields that activity of the enhancer (A_E). Then, the activity of all candidates is linked to the evaluated gene by:

$$WA_{E,G} = A_E * C_{E,G} \quad (2)$$

The Weighted Activity of enhancer E for gene G ($WA_{E,G}$) is estimated as the product of A_E and the contact frequency, reported by Hi-C, between the regions of the genome that contain E and G 's TSS ($C_{E,G}$). Finally, a score is obtained by comparing the relative influence of enhancer E 's weighted activity to the sum of the weighted activities of all other candidates ($WA_{e,G}$) in the following way:

$$ABC\text{Score}_{E,P} = \frac{WA_{E,G}}{\sum WA_{e,G}} \quad (3)$$

The activity of the gene's promoter is also included in the denominator of this equation.

In this way, the ABC score represents the relative contribution that a given candidate enhancer makes toward the total regulatory input received by the target gene. By selecting an appropriate ABC score threshold, true enhancer-promoter pairs can be determined. The combined consideration of activity and Hi-C contacts yields better predictions than either value on their own. The A_E value on its own does not change when evaluating different target genes; the biochemical signals from ChIP-seq and ATAC-seq cannot determine specificity. To link enhancers to genes, ABC relies on Hi-C contacts. However, close spatial proximity between enhancers and genes does not always mean that a regulatory interaction between these two elements is taking place. The activity of enhancers tends to be cell-type and condition specific, therefore a biochemical signature is needed to assess whether an enhancer is active.

Strikingly, the authors of this method demonstrate that cell-type specific Hi-C experiments are not necessary in order to achieve meaningful predictions. The use of an average Hi-C dataset, made from the contact frequencies of several cell-types, can yield a similar predictive performance as when using a cell-type specific dataset. This observation is of great value, given that three-dimensional contact frequency maps are only available for a limited amount of cell lines due to the high cost and technical complexity of these experiments (Fulco et al., 2019). The H3K27ac ChIP-seq and ATAC-seq datasets, however, must be obtained from the cell-line that is analyzed.

1.5 Regulatory response to glucose starvation

For many organisms, glucose represents an important source of carbon that is necessary to carry out certain metabolic processes. Human cells mainly use glucose as their primary source of energy; by breaking down the glucose molecule, cells obtain energy in the form of adenosine tri-phosphate (ATP). ATP is vital for several chemical reactions that are necessary to sustain life, such as the activation of ion pumps that enable membrane polarization in neuronal cells. Glucose is mainly obtained by the ingestion of food, where the absorption of nutrients takes place during digestion. Given its ubiquitous use by many organisms, as well as the important

role it plays in the production of energy, cells must be able to contend with constant changes in the availability of glucose.

It has been shown that mammal liver cells adapt their transcriptional program in order to be able to grow in glucose depleted conditions (Rui, 2011). This transcriptional response favors the expression of enzymes that participate in metabolic processes that allows cells to produce energy, without further intake of glucose molecules. During periods of short term-fasting, liver cells can make use of stored glycogen molecules to produce glucose; however, the amount of glycogen that cells can store is limited, and when faced with longer periods of glucose depravation, they must rely on alternative processes (Rui, 2011). Once glycogen molecules have been depleted, hepatocytes are capable of synthesizing new glucose molecules from lactate, pyruvate, glycerol, and amino acids; this procedure requieres the breakdown of proteins in order to make amino acids available, which can also have negative consequences if continued for long periods of time (Rui, 2011). Furthermore, the depravation of glucose has also been shown to induce the expression of proteins that limit the progression of the cell cycle, such as p53 (Okoshi et al., 2009).

Changes at transcriptional regulatory elements, such as increased chromatin accessibility at promoter regions or the acquirement of activating biochemical signatures at enhancers (examples given above), are clearly linked to the up-regulation of gene products necessary for survival. In the context of glucose availability, studies performed on plants, mice, and yeast have provided individual examples of regulatory elements, particularly promoters, that become active during growth in a glucose deprived environment (Bheda et al., 2020; Senmatsu et al., 2019; Lu, Lim, & Yu, 1998). In some cases, this increased activity is due to the binding of condition specific TFs. However, increases in chromatin accessibility and the addition of histone modifications at promoters have also been observed to respond to glucose availability (Bheda, 2020; Nicolaï et al., 2006). Even so, to the best of our knowledge, a comprehensive genome-wide effort to describe and identify regulatory elements that are induced by glucose starvation has still not been carried out.

2 Justification

The stress-induced activity of transcriptional regulatory elements, such as promoters and enhancers, is linked to cellular processes that enable cell survival. While it has been established that glucose-starvation can provoke a transcriptional response in liver cells, a genome-wide identification of the responsive transcriptional regulatory elements has not been carried out. The study of glucose-responsive regulatory elements would reveal the regulatory mechanisms and dynamics responsible for the correct use of alternative sources of energy during periods of glucose starvation. Identifying these responsive regions can have important implications for the study of aberrant gene expression patterns that give way to conditions such as hyperglycemia and diabetes.

3 Aims

General:

- Identify transcriptional regulatory elements with differential regulatory activity in response to glucose-starvation.

Particular:

- Contrast ATAC-seq datasets to identify genomic regions that exhibit differential accessibility in response to glucose-starvation.
- Relate changes in chromatin accessibility and presence of histones with activating modifications at gene promoters, with the observed gene transcriptional response.
- Identify non-promoter nucleosome deprived regions with enhancer-like biochemical profiles, and link them to their target gene.
- Relate changes in chromatin accessibility and presence of histones with activating modifications at enhancers, with their target gene's transcriptional response.
- Identify metabolic processes regulated by differentially active transcriptional regulatory elements.

4 Hypothesis

We will be able to identify genome-wide significant changes in chromatin accessibility and presence of activating histone PTMs in response to glucose-starvation by contrasting these biochemical profiles when they are gathered during periods of growth in nutrient-rich media with profiles obtained during glucose-deprived cultivation. These changes will largely occur at the sites of transcriptional regulatory elements, such as enhancers and promoters. Additionally, changes in the biochemical activating signatures at transcriptional regulatory elements will be able to explain the observed transcriptional response of their regulated genes.

5 Results

In order to identify glucose-responsive regulatory elements, we made use of datasets that had been previously been generated by an experimental collaborator. A population of a human liver cancer line (Huh7) was studied, following the methodology laid out in Figure 1. Briefly, Huh7 cells that had been cultivated in nutrient-rich media were transferred to a medium that did not contain glucose. The cells were allowed to grow in these conditions for 24 hours. Afterwards, the same population of cells was then transferred to a glucose-rich medium for another 24 hours before being transferred back to glucose-depleted medium. In total, cells were subjected to three, 24 hour periods of growth in medium without glucose, and three periods of growth in nutrient-rich media (including the initial growth state). In order to assess the effect of glucose availability on gene expression, RNA-seq was carried out after each 24 hour period of growth, before media transfer took place. In the same manner, changes in chromatin accessibility were determined via ATAC-seq, and the presence of activating PTMs was measured by H3K27ac, H3K4me1 and H3K4me3 ChIP-seq.

5.1 Transcriptional response to glucose starvation

Past studies have shown that a glucose depleted medium can elicit a transcriptional response, that up-regulates the expression of genes involved in various pathways such as: the unfolded protein response (Park et al., 2004), the reorganization of the cytoskeleton (Joyner et al., 2016), cellular adhesion and migration (Fu, Tang, Xiang, Liu, & Xu, 2019), etc. At the same time, genes involved in cell growth and the progression of the cell-cycle are down-regulated (Okoshi et al., 2009). To characterize the transcriptional response of our studied cell-line in response to glucose starvation, we performed differential gene expression analysis using the RNA-seq experiments described previously.

We identify 1,685 genes that exhibit increased expression during starved periods of growth, while 1,602 genes show decreased expression in these same growth condition (Figure 2A). Genes whose expression becomes up-regulated are present in pathways that have previously been associated with glucose starvation, such as the actin cytoskeleton organization, regulated exocytosis, cell adhesion etc (Figure 2B). Similarly, differentially down-regulated genes are present in familiar pathways related to the progression of the cell-cycle and cell growth (Figure 2C).

In summary, glucose-starvation elicits a transcriptional response in our studied Huh7 cell-line. This adapted expression program affects pathways that have been previously shown to be part of the cellular response to glucose availability. This led us to explore the role of that differentially active transcriptional regulatory elements might play in the establishment of this transcriptional program.

5.2 Differential Accessibilty detected in ATAC-seq

An accessible DNA sequence is a common feature of active transcriptional regulatory elements, as this allows binding of TFs that can influence transcription (Bell et al., 2011). Additionally, several examples stimuli-responsive promoters have shown that these regions have an increased level of chromatin accessibility when the cell faces environmental stress (De Nadal et al., 2011). These observations have created great interest in the genome-wide identification of genomic loci that present increased or decreased chromatin accessibility (differential

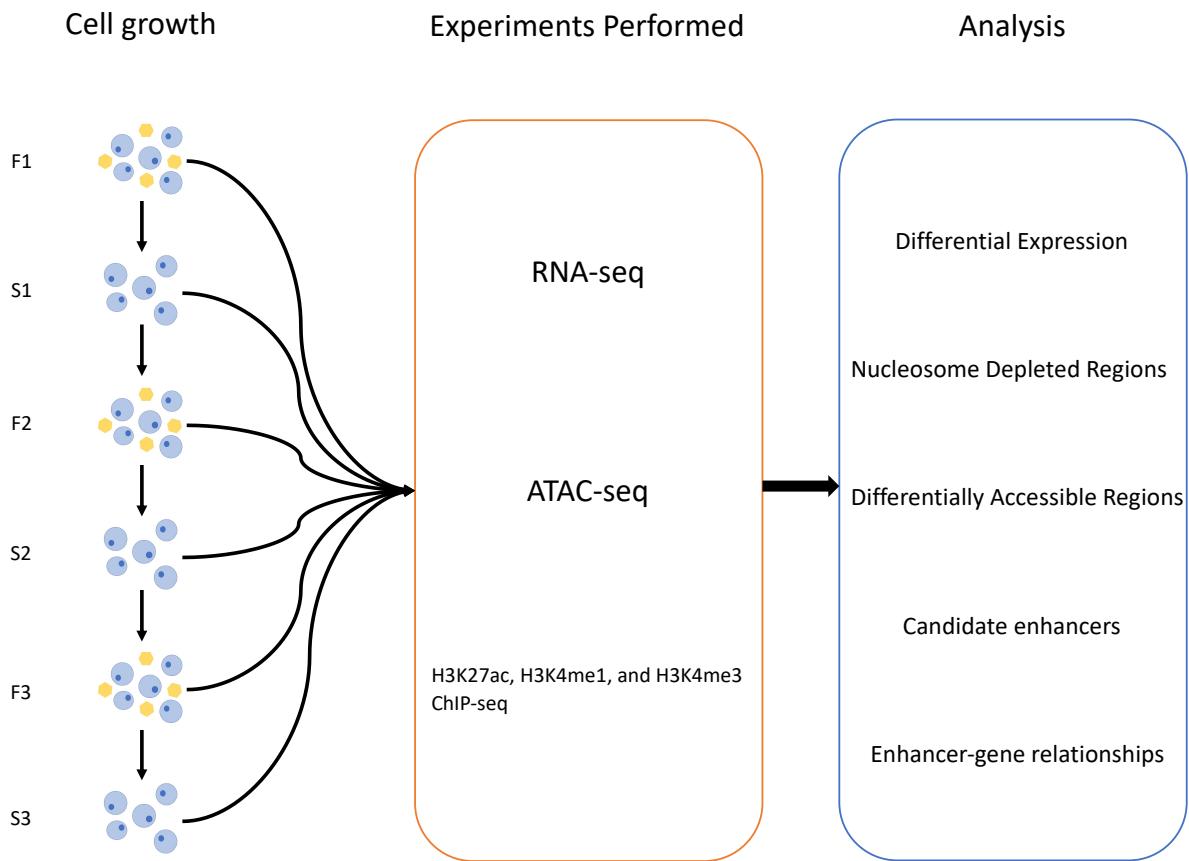


Figure 1: Procedure for the identification of glucose-responsive transcriptional regulatory elements. The procedure can be divided into 3 general steps. Cell Growth: A population of Huh7 cells (blue circles) were grown in glucose-rich media (yellow hexagons) and then transferred to grow glucose-deprived media, in an alternating pattern. Each growth period lasted 24 hours. This yielded three, 24 hour, periods of growth in glucose-rich media (labelled F1, F2 and F3), and three growth periods in glucose-deprived media (labelled S1, S2, and S3). Experiments Performed: After each 24 hour period of growth, and before media transfer took place, the effect of glucose availability was measured by RNA-seq, ATAC-seq, and ChIP-seq for various histone PTMs (H3K27ac, H3K4me1, and H3K4me3). Analysis: Once all experiments were performed for each growth period, the data was analyzed with the goal of identifying glucose-responsive transcriptional elements. ATAC-seq was used to locate NDRs and determine if their level of accessibility changed in response to glucose presence or absence. Potential regulatory activity at NDRs was determined via ChIP-seq analysis. H3K27ac was used, together with ATAC-seq, in order to assess the regulatory activity of NDRs and link potential enhancers to genes. Regulatory activity levels were related to gene expression via differential expression analysis of RNA-seq

accessibility (DA)) in response to various external stressors (Reske, Wilson, & Chandler, 2020). The observed transcriptional response in our studied cell-line, led us to believe that changes

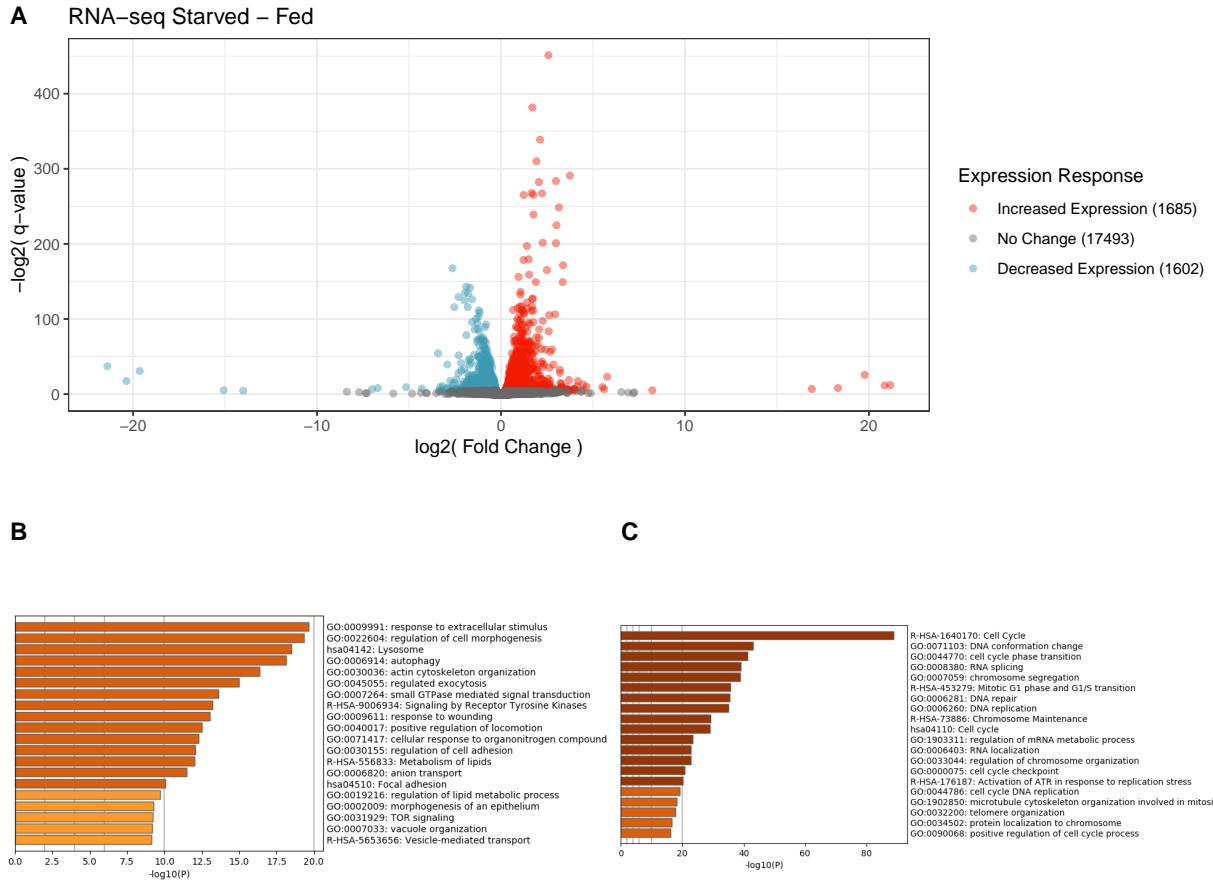


Figure 2: Transcriptional response to glucose starvation. **A.** log Fold Change (FC) in gene expression is plotted against $-\log_{10}(q\text{-value})$. 1,685 genes are identified as having up-regulated expression ($q\text{-value} < 0.05$, $\log_2(\text{expression FC}) > 0$) during glucose starvation (shown in red), while 1602 genes exhibit significant down-regulated expression ($q\text{-value} < 0.05$, $\log_2(\text{expression FC}) < 0$) during the same growth period (shown in blue). **B.** Enrichment analysis, as performed by Metascape, for genes with increased expression during glucose- tarnation. **C.** Metascape enrichment analysis for genes that decrease in expression during glucose starvation. The transcriptional response is in line with previous reports of cell adaptation to growth in a glucose-depleted medium

in chromatin accessibility might also take place in response to glucose availability. Identifying these DA sites, would reveal regulatory elements whose function responds to glucose starvation. In the experimental procedure, ATAC-seq experiments with two replicates at the end of each growth period were performed, yielding a total of 12 samples.

5.2.1 Normalization of ATAC-seq datasets

Firstly, in order to locate candidate regulatory regions that are active in Huh7 cells, we used the information gathered in ATAC-seq experiments to determine the position of NDRs. This

was achieved by using the MACS2 software (Y. Zhang et al., 2008), which identifies regions of the genome that have an enriched number of ATAC-reads mapped to it. These enriched regions are commonly referred to as *ATAC-seq peaks*, and they are meant to represent NDRs. After applying the MACS2 procedure to all samples from every growth period, a total of 116,763 ATAC-seq peaks were identified. These regions were mostly located at gene promoters, gene introns and intergenic regions (Figure 7 All peaks).

Next, with the intention of identifying DA regions, we compared the ATAC-seq measurements gathered during periods of glucose-rich growth, to measurements gathered during glucose-starved growth. An initial impression of the variation present in the samples, can be seen in Figure 3. We notice obvious differences in the number of reads contained in ATAC-seq peaks when comparing the different samples. Noteably, sample F1-1 contains less counts per million (CPM) in many ATAC-seq peaks when compared to all other samples (Figure 3 F1-1). Additionally, there is also great variability between replicates, which is the case for samples F1-1 and F1-2. While the variability between ATAC-seq samples is likely due to a combination of biological and technical noise, the notable differences between replicates indicates that technical noise may account for an important part of the observed variance. To correct this, we evaluate several normalization procedures.

Special care was taken to correctly normalize these experiments in order to identify meaningful changes in chromatin accessibility. The choice of normalization strategy for ATAC-seq datasets greatly influences the final results and their interpretation (Reske et al., 2020). The procedures evaluated in this work, mainly vary in their criteria for deciding which genomic regions will be included to estimate normalization factors. Currently, there is no standard ATAC-seq normalization procedure that is widely used to carry out DA analysis. With this goal in mind, Reske, et al. evaluated the following normalization procedures:

- **binned | TMM (B-T)**: The number of reads that map to continuous 10kb genomic bins is used to estimate normalization factors using the trimmed mean of M values (TMM) (Robinson & Oshlack, 2010). By taking into account the genome-wide ATAC-seq signal, detected DA sites can be said to exhibit changes that are greater than the background signal.
- **Local Enrich | TMM (LE-T)**: This strategy follows the same steps laid out in **B-T**. An additional step is added, which selects ATAC-seq peaks that show a 3-fold enrichment over the ATAC-seq signal present in the immediate 2kb vicinity. In this way, weak ATAC peaks that might show increased accessibility due to small changes in the number of mapped reads, are removed.

We have re-evaluated these methods, together with the following normalization strategies inspired by procedures mentioned in the CSAW (Lun & Smyth, 2016) user manual:

- **Filter Peak | TMM (FP-T)**: In this procedure, we remove ATAC-seq peaks that contain less than 0.75 counts per million (CPM) in at least 6 samples. The number of reads overlapping the remaining peaks is used to estimate normalization factors with the TMM method. By limiting the estimation of normalization factors to only include reads at ATAC-seq peaks, we limit the influence of uneven background signals, as well as the influence of weak ATAC peaks.

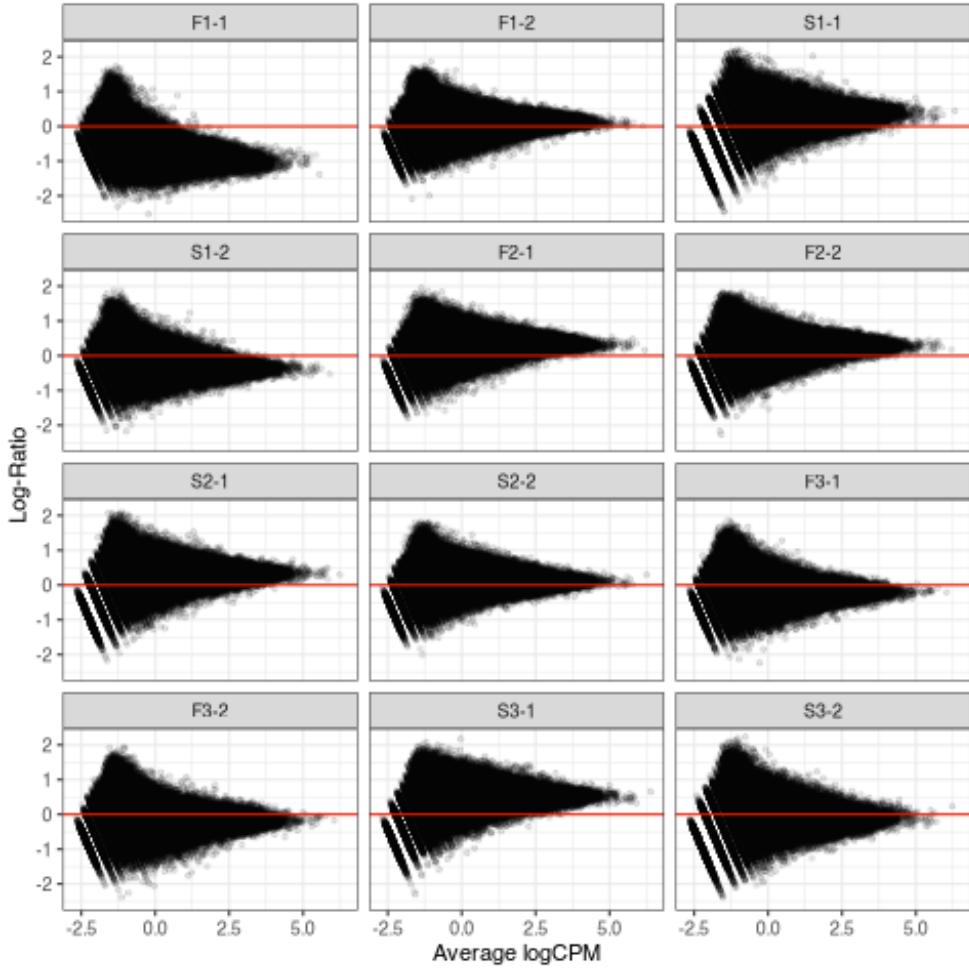


Figure 3: **Mean-Average plot of ATAC-seq samples prior to normalization.** The log-ratio between the number of CPMs found in each ATAC-seq peak in a given sample (y-axis) is plotted against the average CPM value of each peak in all other samples (x-axis). Sample F1-1 shows a lower number of reads in several ATAC-seq peaks in comparison to the reads present in all other samples, including with respect to its replicate, F1-2. This indicates the presence of important technical variation between this sample and all others.

- **Filter Peak | UPQ (FP-U):** Similar to FP-T, the only difference being that normalization factors are estimated using the Upperquartile (UPQ) method, instead of TMM.

We normalized the ATAC-seq samples using each the previously mentioned strategies, and focused our observations on their effect on sample F1-1 (Figure 4). It can be observed that the procedures that estimate normalization factors from reads from the binned genome, do not adequately correct sample biases; in fact, these biases seem to be exacerbated (Figure 4 F1-1: B-T and F1-1: LE-T). The number of adjusted CPMs contained in the majority of ATAC-seq peaks is still below the average of other samples. This problem does seem to be corrected when applying strategies that remove low-count peaks before the estimating normalization factors. When using these criteria, the average scaled number of reads in sample F1-1 is similar to the average of all the samples (Figure 4 F1-1: FP-T and F1-1: FP-U). Empirically, we find that

FP-U better adjusts the values for peaks with a high number of read counts; therefore, we use this normalization procedure in downstream analysis.

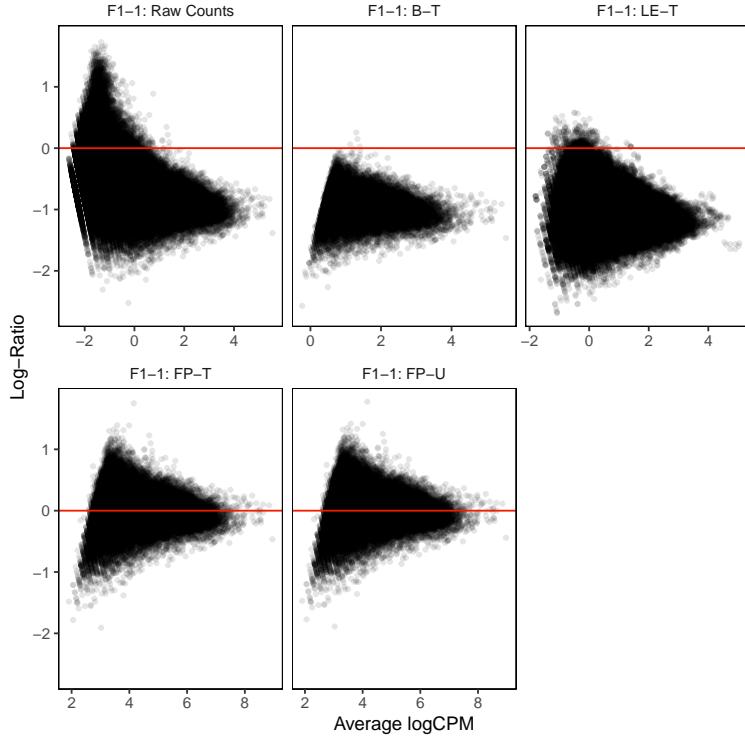


Figure 4: Application of normalization strategies on ATAC-seq samples. After applying normalization procedure on all ATAC-seq samples, we observe their effect on sample F1-1. A successful normalization will adjust the number of reads in each ATAC-seq peak, such that most peaks will have a comparable number of reads to the average of all other samples, and therefore present a log ratio of, or close to, 0. After applying B-T and LE-T procedures, most peaks exhibit a log ratio lower than 0. In contrast, FP-T and FP-U adjust the CPM values such that most peaks have a log ratio close to 0. Additionally, FP-U better adjusts the log ratios of peaks with an average high number of reads.

The observation that the procedures that estimate normalization factors from genome-wide counts do not correct biases present in the samples, indicates important differences in the background genomic ATAC-seq signal between samples. This is possibly due to variations in the experimental protocol, such as varied Tn5 activity efficiencies. Focusing on regions that are rich in ATAC-seq reads, namely ATAC-seq peaks, allows for comparisons with less background variation. Also, the removal of peaks with a low number of mapped reads will limit the detection of changes in chromatin accessibility that are due to small increases or decreases of low values. The improved performance of FP-T and FP-U does not indicate that the B-T and LE-T normalization procedures are incorrect; rather we suggest that these last procedures are not the most adequate for this dataset.

To further assess the effect of normalization on our ATAC-seq samples, we performed Principal Components Analysis (PCA) before and after the application of FP-U (Figure 5). Prior to normalization (Figure 5A) there is an overlap between experiments performed after growth periods in glucose-rich media and samples gathered during glucose starvation. However, once

samples are normalized, we can observe a clear separation between the two growth conditions along the PC1 (Figure 5B). Therefore, the normalization of ATAC-seq samples reveals differences between growth conditions, further suggesting the presence of DA sites.

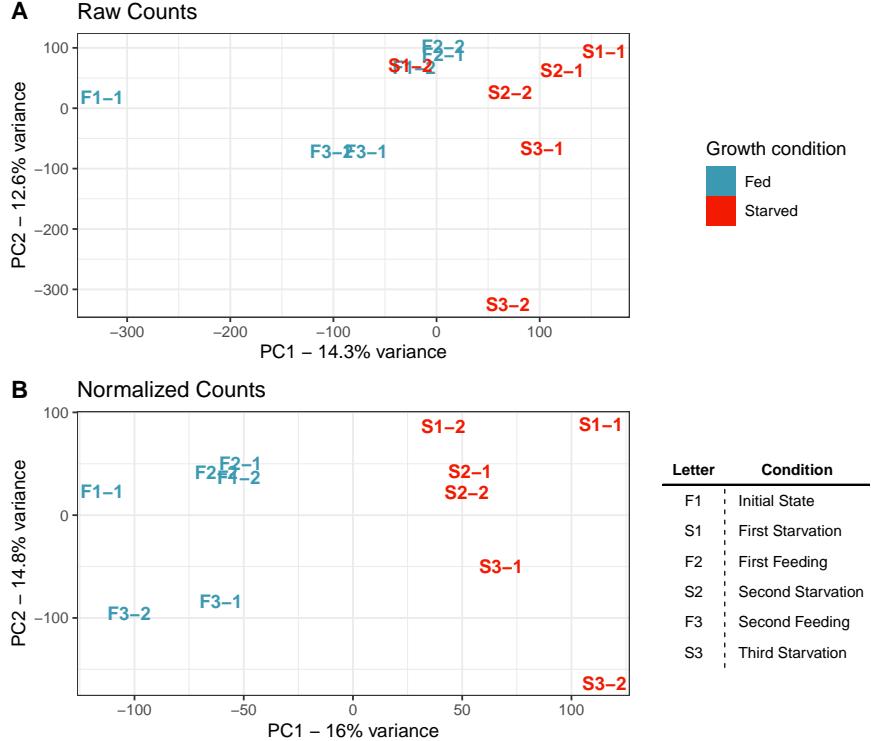


Figure 5: Principal Components Analysis of ATAC-seq samples. **A.** PCA performed on ATAC-seq read counts on all samples prior to normalization shows that there is an overlap between samples gathered during periods of growth in glucose-rich media (Fed) and samples gathered from periods of growth in glucose-depleted media (Starved). **B.** After applying the FP-U normalization procedure, PCA of read counts reveals a clear separation between the samples from both growth conditions along Principal Component 1 (PC1).

5.2.2 Differentially Accessibility analysis

As was previously mentioned, PCA of normalized ATAC-seq samples revealed differences in the adjusted number of reads that are able to distinguish samples gathered during glucose starvation from those gathered during glucose-rich growth (Figure 5B). These key differences in read counts suggest the presence of peaks whose chromatin accessibility responds to the availability of glucose. Therefore, we performed differential accessibility analysis in order to identify these regions (Figure 6A).

This was done by applying a previously established methodology, originally used for the detection of differentially expressed genes by tools such as edgeR (Robinson, McCarthy, & Smyth, 2010). However, this procedure has also been applied assays such as ATAC-seq and ChIP-seq, leading to the development of tools such as CSAW (Lun & Smyth, 2016). Briefly, the number of reads overlapping ATAC-seq peaks is modeled using a negative binomial distribution.

Then, a quasi-likelihood F-test is applied to all peaks in order to detect significant changes in the number of mapped ATAC-seq reads between the two growth conditions.

In total, we identified 771 ATAC-seq peaks (quasi-likelihood F-test, FDR < 0.05) that presented a significant increase in the number of reads when cells are grown in glucose-depleted media (DA-Up peaks) (Figure 6B). At the same time, 356 peaks exhibit a significant decrease in the number of reads upon glucose starvation (DA-Down peaks)(Figure 6C). DA-Up peaks, therefore, represent NDRs that have an increased chromatin accessibility in response to glucose starvation, while DA-Down peaks correspond to NDRs that exhibit the opposite behavior.

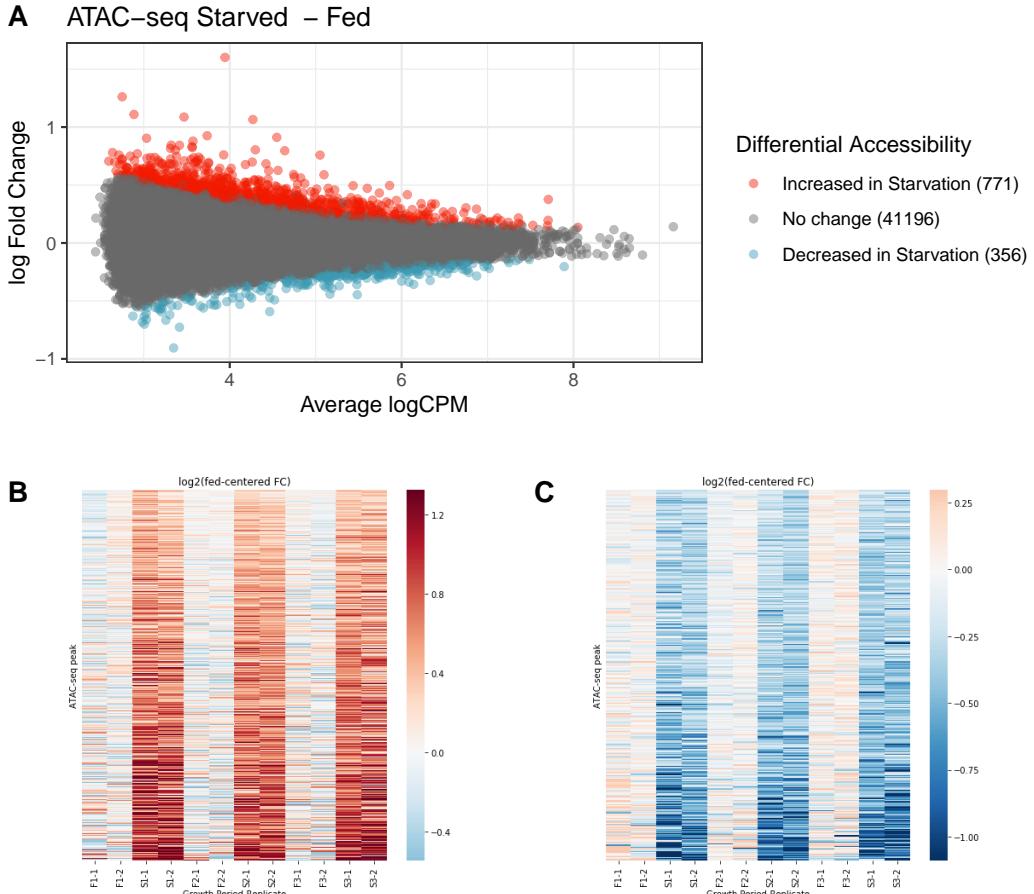


Figure 6: Detection of Differentially Accessible ATAC-seq peaks. **A.** We were able to identify ATAC-seq peaks that show increased and decreased chromatin accessibility in reponse to glucose starvation. **B.** Using the quasi-likelihood F-test and requiring a False Discovery Rate (FDR) less than 0.05, we detect 771 ATAC-peaks with a significant increase in the normalized number of reads mapped to them, and **C.** 356 peaks the exhibit a significant decrease in the adjusted number of mapped reads. B and C show log fold change in ATAC-seq signal with respect to average signal during glucose-rich growth

Interestingly, we note that most DA-Down peaks are located in gene promoters (Figure 7 Decreased in starve). DA-Up peaks, on the other hand, are more commonly found within gene introns and intergenic regions (Figure 7 Increased in starve). In fact, the proportion of the different annotation categories for both DA groups does not follow the observed annotation proportions obtained from all of the ATAC-seq peaks. In the context of DA-Up regions, this

observation is quite interesting given that introns and intergenic regions have been previously reported as possible locations for enhancers (Pennacchio et al., 2013).

	Promoter	Other
All peaks	13306	29017
Increased in starve	95	676
Decreased in starve	250	106

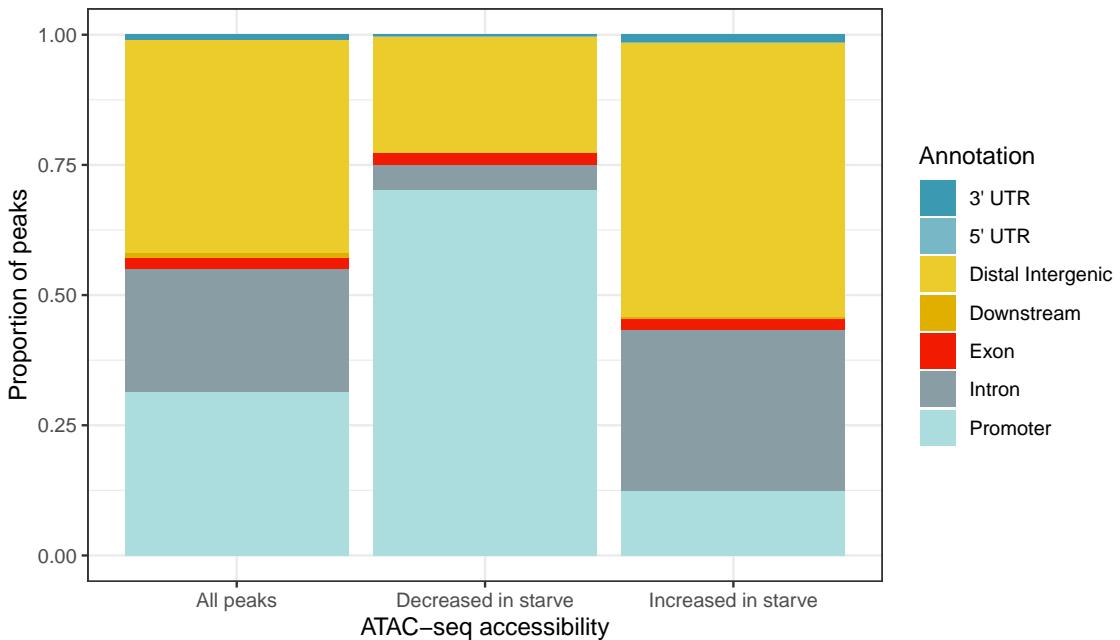


Figure 7: Annotation of ATAC-seq peaks. Using the GENCODE v33 primary assembly annotations as reference, we were able to annotate ATAC-seq peaks based on their genomic coordinates. Most of all ATAC-peaks are found within promoters (2kb upstream or downstream of a gene’s TSS), gene introns and in intergenic regions. DA-Down peaks are mainly found within promoter regions, with few intronic or intergenic peaks. Meanwhile most the most common annotations for DA-Up peaks are introns and intergenic regions, followed by promoters. The fact that most DA-Up peaks are located in introns and intergenic regions is interesting, given that these are locations where enhancers are known to reside.

The identified DA peaks could be indicating the existence of regulatory regions whose activity is modulated by the addition or removal of nucleosomes in response to glucose availability. The regulatory potential of these DA regions is supported by the fact that most are found within promoters (which is especially true for DA-Down peaks), and in previously established locations for enhancers (in the case of DA-Up peaks). This led us to further assess the regulatory potential of these regions, by incorporating information from H3K27ac ChIP-seq.

5.3 Correlation between H3K27ac ChIP-seq, ATAC-seq and Gene Expression

In addition to having an accessible DNA sequence, active promoters and enhancers are also characterized by the presence of H3K27ac modified histones that flank the NDR (Andersson & Sandelin, 2019). Integrating the accessibility and the presence of H3K27ac at candidate regulatory elements can give a measurement of these region's regulatory activity (Andersson & Sandelin, 2019; Creyghton et al., 2010). Additionally, previous studies have shown that, when facing environmental stress, enhancers can acquire increased levels of H3K27ac, leading them to further up-regulate the transcription of their target genes (Creyghton et al., 2010; Simeonov et al., 2017; Williams, Amaral, Simeoni, Somervaille, et al., 2020). Therefore, we evaluate the regulatory potential of our previously identified NDRs, by measuring the presence of H3K27ac at these sites. We also explore whether changes in the chromatin accessibility and H3K27ac signal at gene promoters is correlated with changes in gene expression.

Given that H3K27ac modified histones are associated with the NDRs of transcriptional regulatory elements (Andersson & Sandelin, 2019), we reasoned that the FP-U normalization procedure would be able to successfully normalize H3K27ac ChIP-seq datasets. We extended all ATAC-seq peaks to a minimum length of 500 base pairs, in order to capture the signal of this modification at flanking histones. We then used the number of ChIP-seq reads overlapping these extended regions, as well as the number of reads overlapping promoter regions, in order to estimate normalization factors with the Upperquartile procedure. The result of this procedure can be observed in Figure 8. This normalization strategy can adequately adjust read counts in these samples, as the average adjusted count is similar across all samples.

To assess the regulatory potential of detected NDRs, we studied the relationship between ATAC-seq signal and H3K27ac ChIP-seq signal at extended ATAC-seq peaks and gene promoters (Figure 9A and B). We observe that the signal intensity of these two assays is highly correlated at both ATAC-seq peaks (Figure 9A, spearman $\rho = 0.534$) and gene promoters (Figure 9B, spearman $\rho = 0.864$). This observation further supports the regulatory potential of many of the NDRs present in our dataset, as they exhibit both chromatin accessibility and the presence of H3K27ac modified histones.

Given that both, chromatin accessibility and presence of H3K27ac modified histones, have been previously associated to gene expression (Andersson & Sandelin, 2019), we sought to evaluate these relationships in our datasets. To do this, we compared the signal intensity of ATAC-seq and H3K27ac ChIP-seq at promoters, with the intensity of gene expression as measured by RNA-seq. We notice a correlation between ATAC-seq signal at promoters and the RNA-seq signal intensity of the corresponding gene (Figure 9C, $\rho = 0.389$). This positive correlation value indicates that promoters with a higher degree of accessibility tend to regulate highly expressed genes. There is also a positive correlation between H3K27ac ChIP-seq signal and RNA-seq signal (Figure 9D, $\rho = 0.491$), that is slightly higher than the correlation of ATAC-seq vs RNA-seq. Greater presence of H3K27ac histones at the promoter is related to a higher level of expression. Therefore, our datasets exhibit similar behavior to what has been reported by previous studies (Andersson & Sandelin, 2019; Creyghton et al., 2010; De Nadal et al., 2011).

Changes in chromatin accessibility and presence of H3K27ac modified histones have been shown to occur in response to environmental stress (De Nadal et al., 2011; Williams et al., 2020; Ostuni et al., 2013). We explored whether these changes also take place in response to

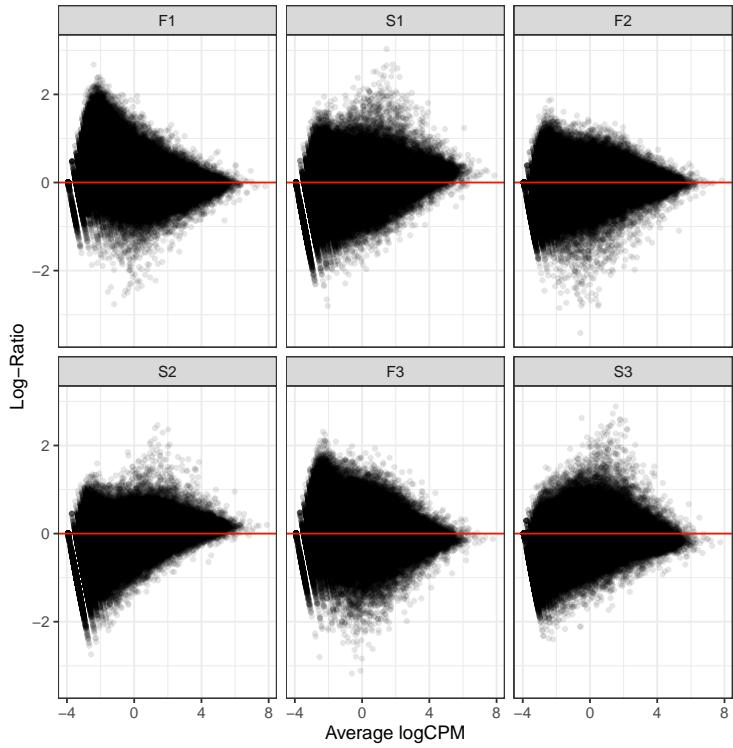


Figure 8: Normalization of H3K27ac ChIP experiments. In order to capture the signal of H3K27ac ChIP-seq at histones flanking NDRs, all high count ATAC-seq peaks were extended to minimum length of 500 base pairs. Then, the number of H3K27ac ChIP-seq reads overlapping these extended regions was used, together with the number of reads in promoter regions (500bp +/- TSS), to estimate normalization factors with the Upperquartile procedure. For each gene, only the promoter of the most expressed isoform was considered. This normalization strategy can adequately adjust biases in these samples, as the average adjusted count is similar across all samples.

glucose starvation, and whether local modulation of chromatin accessibility is also related to changes in the presence H3K27ac modified histones (Figure 10). To do this, we compared the fold change (FC) of ATAC-seq signal to the observed FC in H3K27ac ChIP-seq signal, at both promoter regions and NDRs. When assessing signal changes in ATAC-seq peaks, we observe the ATAC-seq signal FC and H3K27ac ChIP-seq signal FC are correlated (Figure 10A. spearman $\rho = 0.262$). This relationship is slightly stronger at promoter regions (Figure 10B. spearman $\rho = 0.307$). These observations indicate that both chromatin accessibility and the presence of H3K27ac modified histones react to the availability of glucose availability in a similar manner, although this relationship is not true at all regions.

Given the previously established link between ATAC-seq signal, H3K27ac ChIP-seq signal and RNA-seq signal (Figure 9), we next assesed whether the observed changes in chromatin accessibility and presence of H3K27ac modified histones could also be closely related to changes in gene expression in response to glucose starvation. This was done by comparing log FC of these assays. For ATAC-seq and H3K27ac ChIP-seq, FC was estimated at promoter regions, while RNA-seq FC was measured for the corresponding gene (Figure 10). Interestingly, it can be observed that changes in ATAC-seq signal at gene promoters, are not correlated with changes

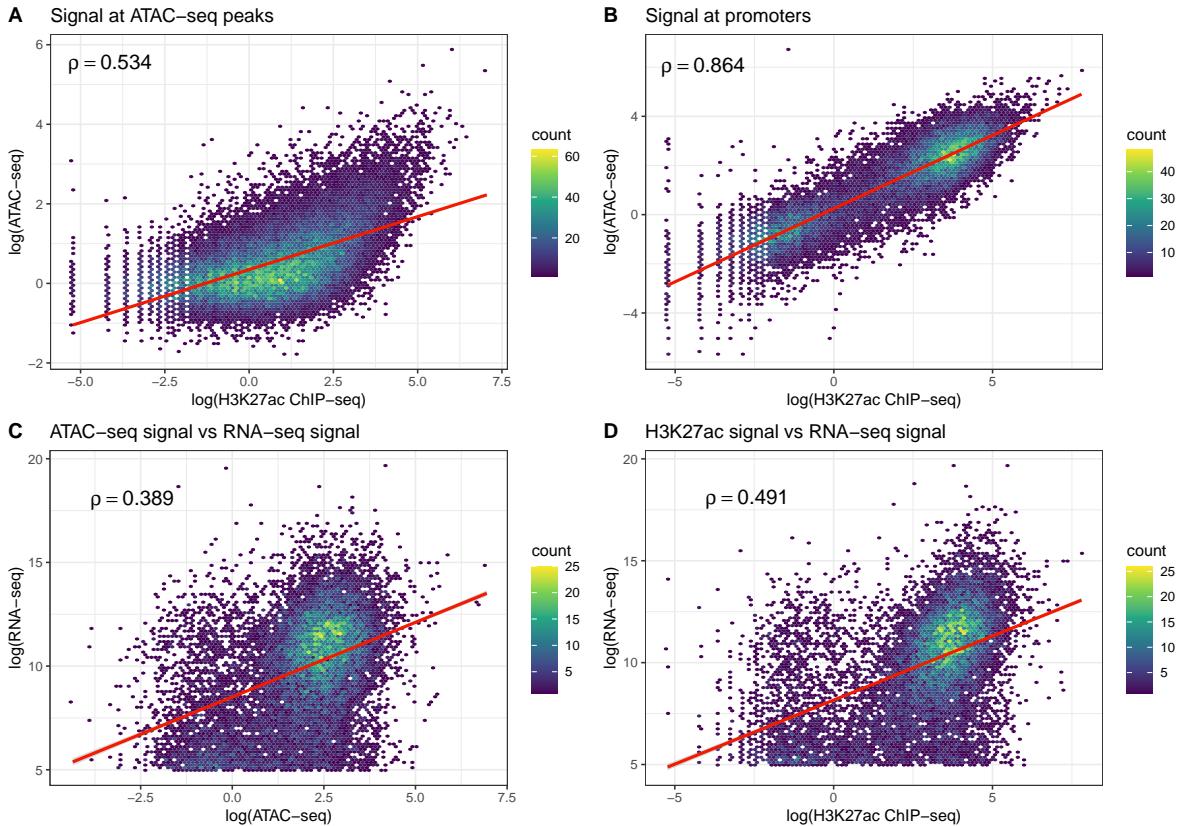


Figure 9: Relationship between ATAC-seq, H3K27ac ChIP-seq and Gene Expression. The normalized CPMs of ATAC-seq and H3K27ac ChIP-seq from growth period F1 was measured at promoter regions (500bp +/- TSS) and at ATAC-seq peaks. Normalized RNA-seq CPMs where also gathered from growth period F1. Individual points are aggregated into hexagonal bins along both axis (*count* legend). A linear model was fit (red line) to the resulting spread. **A.** Spearman correlation of ATAC-seq signal and H3K27ac signal at ATAC-seq peaks (spearman $\rho = 0.534$, p-value $< 2.2e-16$). **B.** Spearman correlation of ATAC-seq and H3K27ac ChIP-seq signal intensity at promoter regions (spearman $\rho = 0.864$, p-value $< 2.2e-16$). In both cases, the signal of these assays is correlated, thereby indicating a high regulatory potential for determined NDRs. **C.** Spearman correlation of ATAC-seq signal and RNA-seq signal (spearman $\rho = 0.389$, p-value $< 2.2e-16$) **D.** Spearman correlation of H3K27ac ChIP-seq signal vs RNA-seq signal (spearman $\rho = 0.491$, p-value $< 2.2e-16$). Both results indicate that the degree of chromatin accessibility and presence of H3K27ac modified histones at promoter regions is related to the expression level of the corresponding gene.

in gene expression (Figure 10C spearman $\rho = 0.026$). However, there is a stronger relationship between gene expression and H3K27ac ChIP-seq signal at promoters (Figure 10D spearman $\rho = 0.294$); particularly for genes whose promoters show a positive FC in H3K27ac presence. A loss of H3K27ac modified histones, however, does seem to indicate an equal reduction in gene transcription.

These results indicate that several of the NDRs that were identified, are potential regulatory elements as they show a degree of chromatin accessibility that is also associated with the presence of H3K27ac modified histones (Figure 9A and B). Both of these characteristics are

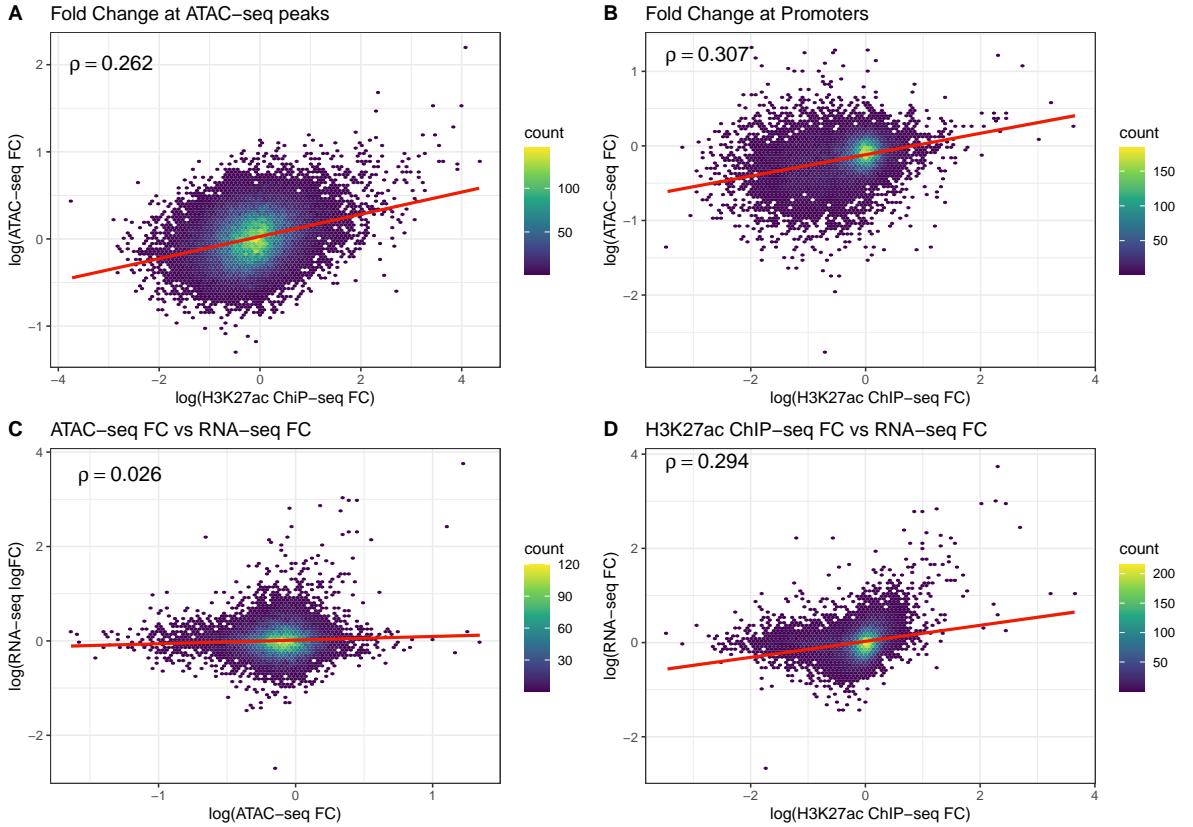


Figure 10: ATAC-seq signal, H3K27ac ChIP-seq singal and RNA-seq signal response to glucose starvation. The log FC for ATAC-seq and H3K27ac ChIP-seq signals was estimated by subtracting the average log CPM of samples from glucose-rich growth (F1, F2, and F3) from the log CPM of all samples. Individual points are aggregated into hexagonal bins along both axis (*count* legend). A linear model was fit (red line) to the resulting spread. **A.** Spearman correlation of ATAC-seq logFC and H3K27ac ChIP-seq logFC at ATAC-seq peaks (spearman $\rho = 0.262$, p-value $< 2.2e-16$). **B.** Spearman correlation of ATAC-seq logFC and H3K27ac ChIP-seq logFC at gene promoters (500bp +/- TSS) (spearman $\rho = 0.307$, p-value $< 2.2e-16$). Both chromatin accessibility and the presence of H3K27ac modified histones react to the availability of glucose availability in a similar manner. **C.** Spearman correlation between ATAC-seq logFC and RNA-seq logFC (spearman $\rho = 0.026$, p-value = 0.0019). **D.** Spearman correlation between H3K27ac ChIP-seq logFC and RNA-seq logFC (spearman $\rho = 0.294$, p-value $< 2.2e-16$).

associated with active transcriptional regulatory elements. This relationship to transcription was observed at promoter regions, where the degree of chromatin accessibility and presence of H3K27ac modified histones was correlated with a corresponding level of gene expression (Figure 9C and D). Interestingly, the observed changes of these two measurements in response to glucose starvation are also similar (Figure 10A and B). However, the modulation of promoter chromatin accessibility did not reflect the observed differential expression of the corresponding gene (Figure 10C). Changes in the presence of H3K27ac did have a closer association to changes in gene expression (Figure 10D); however it can also be observed that a loss of H3K27ac at histones does not indicate a decrease in gene expression.

Given that changes in promoter regions do not fully explain the observed gene transcriptional response, we decided to further assess the regulatory role of non-promoter NDRs with characteristics of active transcriptional regulatory elements, by evaluating their possible function as transcriptional enhancers.

5.4 Adaption of Activity by Contact for condition comparisons

In the previous section, we found that while changes in chromatin accessibility and presence of H3K27ac modified histones in response to glucose availability take place at promoters, they do not fully explain the observed changes in gene expression (Figure 10C and D). At the same time, several non-promoter NDRs in our dataset exhibit characteristics of active regulatory regions (Figure 9A and B), and also present changes in these signals when cells face glucose starvation (Figure 6, Figure 7, and Figure 10A). This led us to investigate their possible role as enhancers in the transcriptional response to glucose starvation. To do this, we employed the Activity By Contact (ABC) model (Fulco et al., 2019). Briefly, this model aims to identify regulatory interactions between genes and enhancers, based on a candidate enhancer's activity (see Equation 1), and their spatial proximity to a given gene (weighted-activity see Equation 2). By comparing the contact weighted-activities of all NDRs within 5 Mb of a gene's TSS, each candidate can be assigned a score (ABC-score see Equation 3) that reflects its relative influence on the gene's expression.

While the ABC model has shown promising results, it was not originally designed to compare the activity of enhancers in different conditions. Therefore, we first developed a strategy that would allow a valid comparison of enhancer activities between samples that were gathered from different growth conditions. Our approach consisted of using read counts that had been previously adjusted with the FP-U procedure. To determine the effect that this normalization has on the estimation of enhancer activity, we applied this normalization procedure to the dataset from the original ABC publication (Fulco et al., 2019), by normalizing assay replicates.

As can be seen in Figure 11, the use of normalized adjusted read counts does not present significant changes in the estimation of enhancer activity. In fact, the activity reported by the authors and the activity estimated from adjusted read counts are highly similar (spearman $\rho = 0.99$). The activity values reported by the authors are slightly higher, but the relative behavior of enhancer candidate elements is maintained.

While the use of normalized datasets does not change the general estimation of enhancer activity, we note that it does reveal differences between samples that come from the same growth condition (Figure 12). The use of unnormalized read counts yields a PCA landscape where there is a slight separation between activity estimations along PC1 (Figure 12A). Meanwhile, the same analysis done with normalized activity estimates, also reveals a clear separation between growth conditions along PC1, and also reveals a greater separation between samples gathered during the same growth condition along PC2 (Figure 12B). In fact, the order in which experiments are performed seems to be recapitulated along this component, possibly indicating activity patterns dependent on repeated exposure to glucose starvation.

The use of normalized samples does not greatly influence the estimation of activity in a single sample (Figure 11); but it does have important implications for the comparison of activity values from different growth conditions. In the following analysis, we used normalized datasets to estimate candidate enhancer activity.

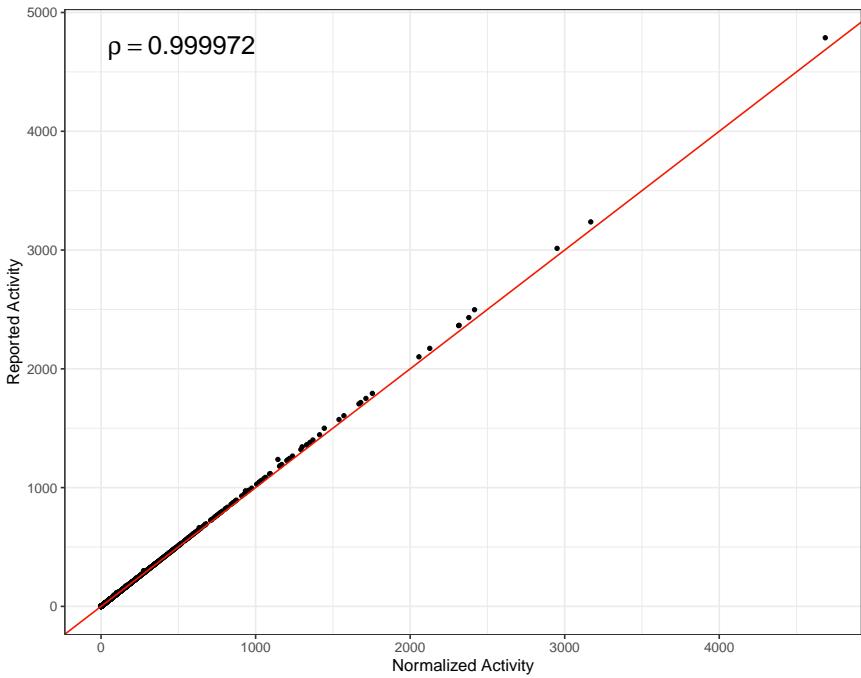


Figure 11: Comparison of Reported Activity and Normalized Activity. In the original ABC publication, H3K27ac ChIP-seq and DNase-seq assays are performed on the K562 cell line. Data from chromosome 22 was used to estimate activity values. The activity obtained by the authors (Reported Activity) is plotted against the activity estimated with FP-U normalized counts (Normalized Activity), according to Equation 1. Each point represents a candidate enhancer. A straight line with slope=1 is plotted to compare value magnitude (red line). Spearman correlation $\rho = 0.99$, $p\text{-value} < 2.2\text{e-}16$.

5.5 Additional filters to infer gene-enhancer relationships

After establishing a strategy that would allow for the comparison of enhancer activity estimates, we applied the ABC model to our FP-U normalized datasets. By requiring a minimum ABC-score of 0.02, the authors of the ABC method were able to predict gene expression changes in response to enhancer perturbation with a precision of 79% and a recall value of 59% (Fulco et al., 2019). Therefore, this is the threshold that we use in our work. Additionally, we leverage the average Hi-C dataset produced by the authors of the ABC method.

We considered two main strategies in order to obtain a set of gene-enhancer pairs for subsequent analysis. One criteria consisted of using the gene-enhancer pairs predicted using the data gathered from a single growth period for further analysis. The second option (referred to as Max Values), consisted of pooling the estimated ABC-scores from all growth periods, and consider whether the highest score for each gene-enhancer pair passed the established threshold. The Max Values criteria was considered with the intention of capturing regulatory relationships that only occur after repeated stress. By considering the highest value, we predict regulatory pairs independently of the growth period in which they occur.

We compared both criteria by observing the distribution of the predicted number of enhancers that regulate one gene (Figure 13A) and the distribution of the number of genes one enhancer is predicted to regulate (Figure 13B). While all growth periods present a similar dis-

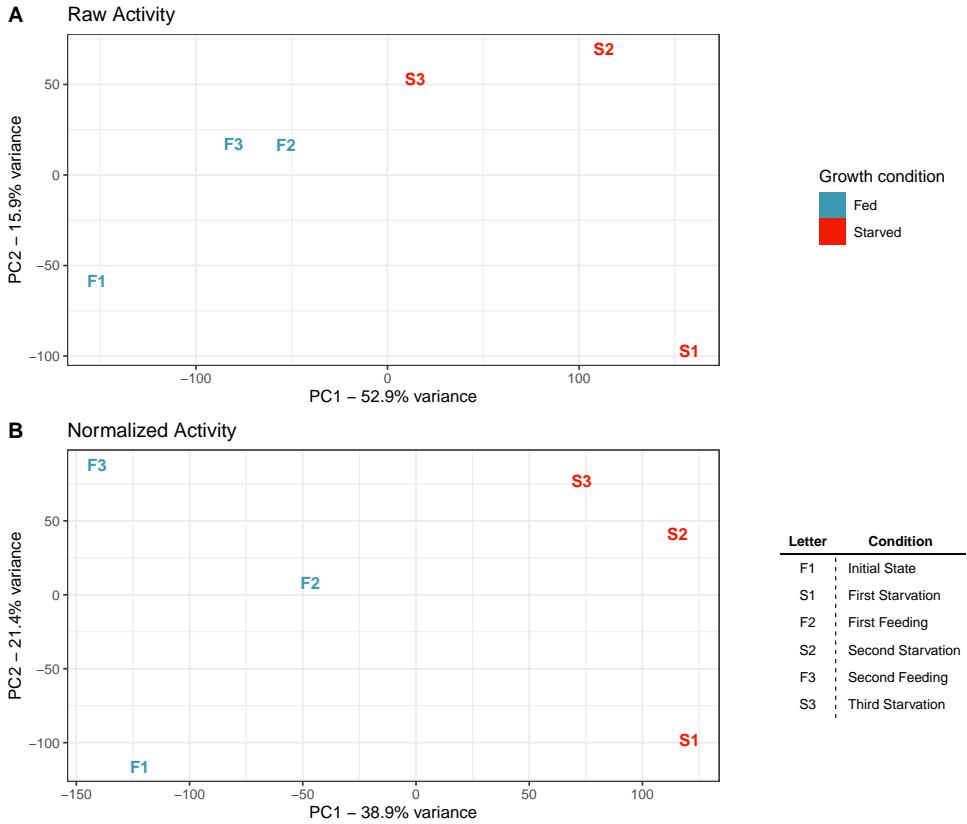


Figure 12: Principal Component Analysis of Activity estimates. Activity was estimated in our datasets for non-promoter extended ATAC-seq peaks, according to Equation 1. **A** PCA performed on unnormalized datasets. **B** PCA performed on FP-U normalized datasets. In both instances, a separation between samples that were gathered in different growth conditions along PC1. However, PCA of normalized activity also reveals separation between samples from the same growth condition, as well as recapitulating the order in which the experiments were performed along PC2)

tribution, we observe that the maximum number of enhancers per gene tends to be higher in samples gathered during glucose-starvation (S1, S2, and S3) and that the Max Values criteria shows an increased number of enhancers per gene predicted. A similar pattern can be observed in the distribution of the number of genes an enhancer regulates; glucose-starved samples show a higher number of genes regulated per enhancer, with Max Values showing the highest value.

The reason behind these different distributions can be observed in Figure 14. There, we note that the distribution of ABC-scores varies in each growth period. Estimations made with datasets gathered during glucose-starved periods of growth, exhibit higher ABC-score values, with more gene-enhancer pairs passing the established threshold. The Max Values strategy exhibits behavior similar to predictions made with glucose-starved samples, suggesting that many of the highest ABC-scores come from glucose-starved samples

In all strategies, it can be observed that many enhancers are predicted to regulate multiple genes; some enhancers are predicted to regulate upward of 10 genes (Figure 13B). This led us explore the properties of candidate enhancer regions and their relationship to the assignment of multiple genes (prolific enhancers) (Figure 15). We find that as the activity measurement

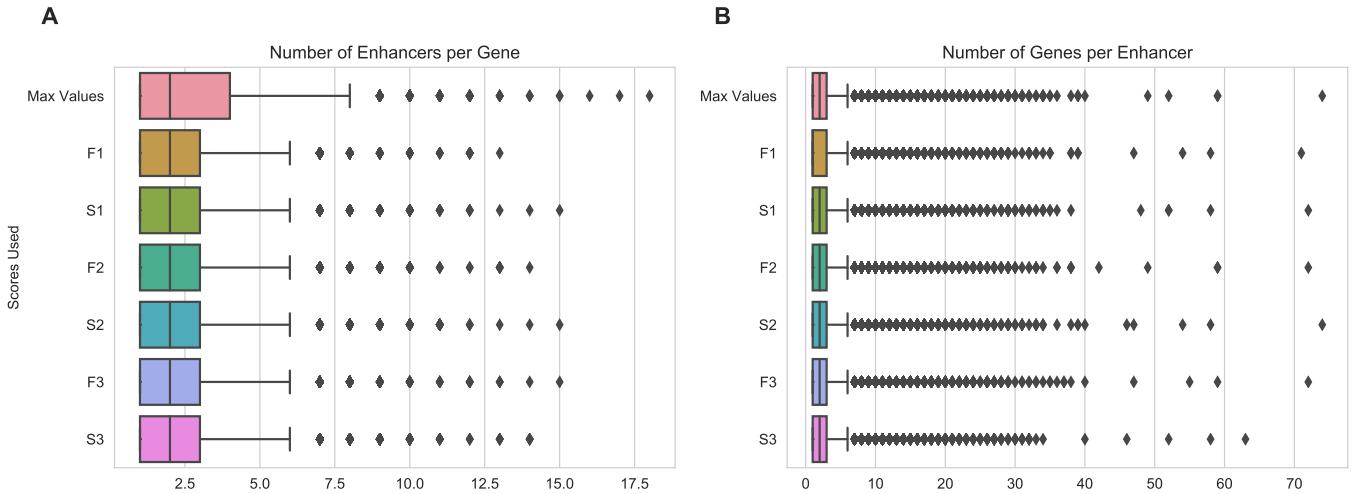


Figure 13: **Distribution of gene-enhancer relationships** ABC-scores for enhancer-gene pairs were estimated using data from each growth period (see Equation 3). We consider Gene-enhancer predictions in each growth period, as well as predictions made with the highest ABC-score for each gene-enhancer pair in all growth conditions. **A.** Distribution of the number of enhancers that regulate a gene. **B** Distribution of the number of genes that an enhancer regulates. In both cases, values tend to be higher during periods of glucose-starved growth

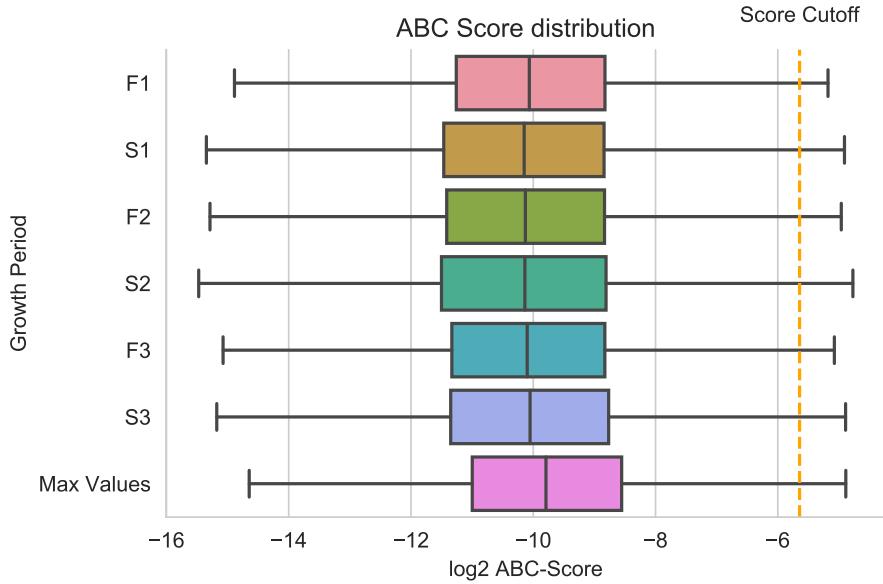


Figure 14: **Distribution of ABC scores.** ABC-score was estimated using datasets gathered during each growth-period. The highest value for each gene-enhancer pair was pooled together to evaluate the Max Values criteria. The distribution of ABC-scores estimated in each growth period and Max Values is shown, along with the established ABC-score cutoff (0.02, orange line). ABC-scores estimated from glucose-starved periods of growth tend to be higher than those gathered during non-starved periods of growth.

of an enhancer increases, so does the number of genes it is predicted to regulate (Figure 15A). Genes that are regulated by these prolific enhancers do not have weak promoters (15B). Rather, the ratio between an enhancer's weighted-activity and the promoter's activity increases as the number of regulated genes is higher (15C). These observations mean that some enhancers are predicted to regulate several genes because of their high activity value in comparison to the regulated promoter's activity. Highly active enhancers are predicted to regulate more genes.

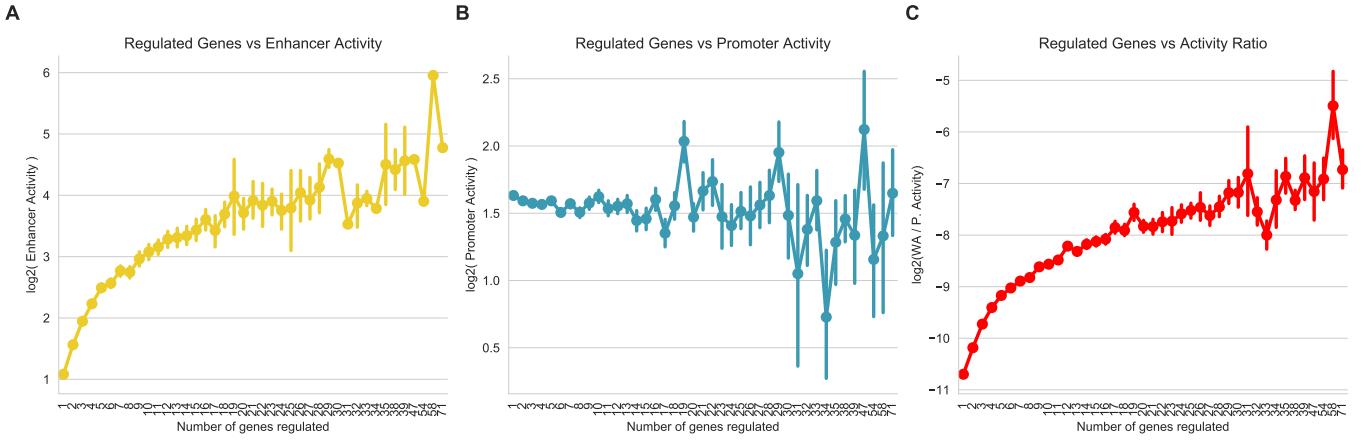


Figure 15: Relationship between enhancer activity and number of genes regulated. The average Activity and Weighted-Activity (WA) values for sample F1 are shown. In all cases error bars represent *average +/- standard error*. **A.** Average log enhancer activity is plotted against the number of genes regulated. As the activity of the enhancer increases, so does the number of genes it is predicted to regulate. **B.** Average log promoter activity is plotted against the number of genes the enhancer that regulates it is predicted to regulate. The activity of the promoter (500bp +/- TSS) does not vary in response to the activity of the enhancer that regulates it. **C.** Enhancer WA is divided by regulated promoter activity. These values are plotted in log scale against the number of genes the enhancer is predicted to regulate. The ratio between weighted-activity of enhancers and promoter activity increases as the number of genes an enhancer regulate increases.

We had previously observed that several ATAC peaks increase in accessibility in response to glucose starvation (figure 6). This increase in the level of accessibility would contribute to higher activity estimations for the same enhancer during glucose starvation, making it a viable candidate for a greater number of genes. This would satisfactorily explain the observations from Figure 14, where glucose-starved growth periods show higher ABC scores. It would also explain why these growth periods, along with the Max Values criteria, result in greater amounts regulated genes per enhancer (Figure 13B).

In order to assess the validity of multiple predicted interactions for a given enhancer, we explored the spatial proximity that such enhancers share with the genes they are predicted to regulate. To do this, we calculate the percentage of total contacts (as reported by the average Hi-C dataset) between a given enhancer region and each of the genes for which it is considered a candidate regulator. As can be seen for *peak 115744* (Figure 16), several of the genes for which it passes the ABC-score threshold (shown in blue) do not share a greater proportion of the total contacts of this enhancer than other genes for which it is not a predicted regulator (shown in red). In the case of this enhancer, the gene PRICKLE3 stands out from other genes, as it

shares a greater proportion of the enhancer's total contacts than the following most contacted genes. This indicates that, while there are several genes for which *peak 115744* represents a viable regulatory candidate, many of them are not in close spatial proximity to this enhancer region.

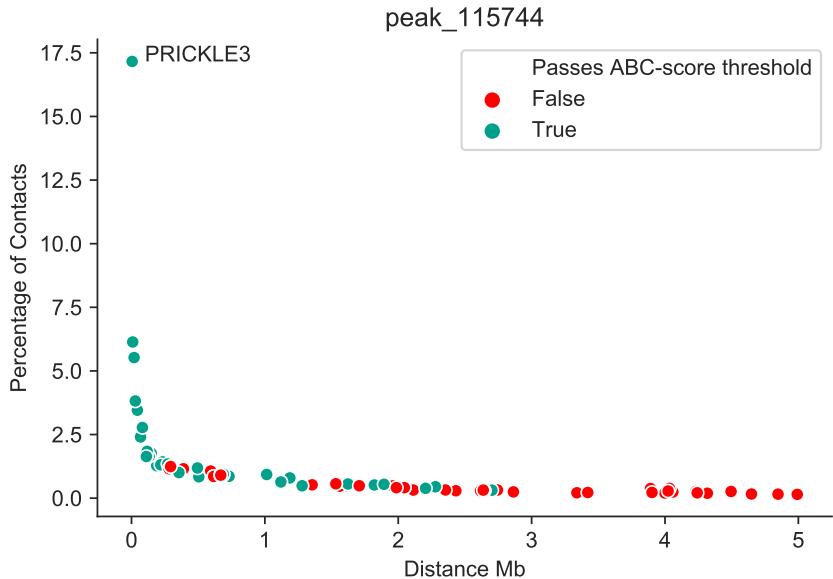


Figure 16: Percentage of total contacts shared for *peak 115744*. After estimating ABC-scores in sample F1, *peak 115744* passes the 0.02 threshold for 35 genes (shown in blue). Genes for which it does not pass the established threshold are shown in red. The normalized Hi-C contacts that *peak 115744* shares with its candidate genes are gathered, and the percentage of this total of contacts that the enhancer shares with a given gene is plotted on y-axis. The linear distance (given in Mb) between the enhancer and a gene is shown on the x-axis. The gene PRICKLE3 has the greatest share of *peak 115744*'s contacts.

The observation that many of the genes that *peak 115744* is predicted to regulate do not share a greater proportion of its contacts than non-regulated genes, reveals a weakness of the ABC method. While the ABC-score can successfully rank the input of several candidate enhancer elements for a given gene, it does not closely consider ranking the potential regulatory interactions that an enhancer can have with several genes. In other words, the ABC-score can select the most viable candidate enhancer for a given gene, but it does not select the most viable gene to be regulated by each enhancer.

It is because of this behavior, that we decided to evaluate an additional filter when selecting positive regulatory interactions using the ABC method. In an effort to select the most meaningful gene candidates for a given enhancer, we explored the addition of a contact percentage filter. In this way, we select for regulatory interactions where the enhancer is a viable candidate for a given gene ($\text{ABC-score} > 0.02$), and said gene has a high potential to be regulated by the evaluated enhancer.

In order to determine a minimum percentage of total enhancer contacts to that a gene must share in order to be considered a viable regulatory target, we observed the distribution of contact percentages for enhancers that regulate a single gene (Figure 17). We note that the

percentage of shared contacts between an enhancer and its regulated genes tends to higher in comparison to the contacts shared with genes it does not regulate. We select an 8.5% contact threshold, as it close to the lower quartile value of contacts shared with regulated genes.

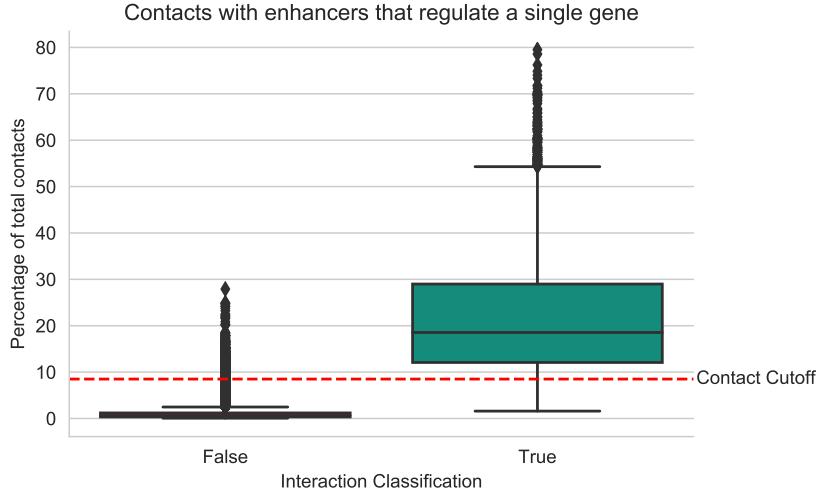


Figure 17: Distribution of the percentage of total enhancer contacts. Enhancers that are predicted to regulate one gene (in sample F1) are analyzed to determine the share of total contacts they share with genes that they do not regulate (False), and the percentage of total contacts they share with genes they are predicted to regulate (True). The distribution of contacts shared with regulated genes has higher values than the percentage shared with genes they do not regulate. An 8.5% contact filter is selected in order to consider a gene as a viable regulatory target for a given enhancer (red line).

A consequence of this additional filter, is a reduction in the number of genes an enhancer is predicted to regulate. When re-evaluating *peak 115744*, only PRICKLE3 shares a distinct number of contacts with this enhancer, thereby discarding all other genes for which this enhancer achieved a high ABC-score. We can observe this pattern replicated for several other enhancers in Figure 18. An 8.5% minimum percentage of contacts seems to identify genes with a distinct number of contacts from other candidate genes with a similar number of contacts. In many cases only one candidate gene would be predicted to be regulated; in other cases all predicted interactions are discarded (such is the case of *peak 106506*).

After contact filtering, enhancers that regulate more than three genes are now considered to be outliers (Figure 19B), while we still observe genes that are regulated by more than one enhancer (Figure 19A). Currently, there is no clear consensus on the number of genes an average enhancer can regulate in the human genome; however, previous studies have suggested an average range of 1 to 3 regulated TSSs for an enhancer (Jin et al., 2013). The application of the contact percentage filter maintains this relationship.

The impact of this additional filter for genes such as GABRG1, is that not all enhancers that pass the ABC score threshold are considered to be regulators of this gene. This would be because some of these enhancers do not pass the contact percentage threshold, and are therefore more likely to regulate other genes (Figure 20).

To the best of our knowledge, no Hi-C dataset has been generated for in our studied cell-line, neither during standard growth conditions or during glucose starvation. However, the use

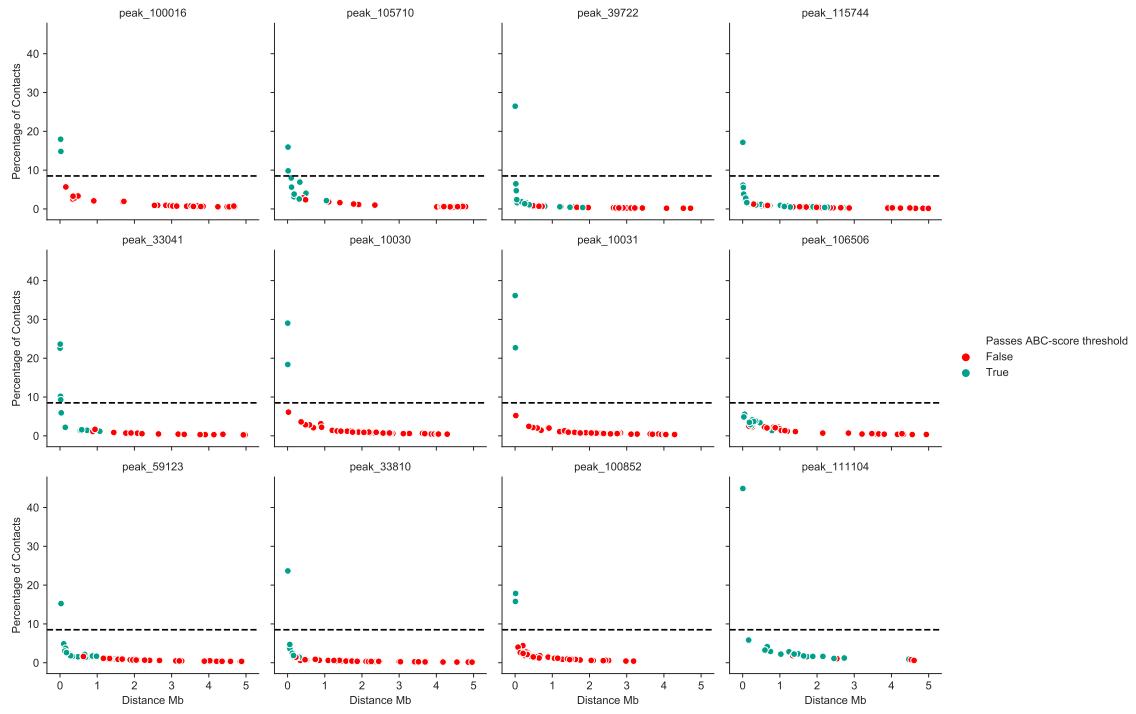


Figure 18: Distribution of percentage of contacts with candidate genes. For the presented set of enhancers, Figure 16 is re-created. The number of shared contacts with candidate genes is shown on y-axis, while the linear distance between these candidate genes and the enhancer is shown on x-axis. Genes for which the enhancer passes 0.02 ABC-score threshold are presented in blue, predicted negative interactions are shown in red. The 8.5% contact threshold is shown as a black line. ABC-scores estimated with data from growth period F1.

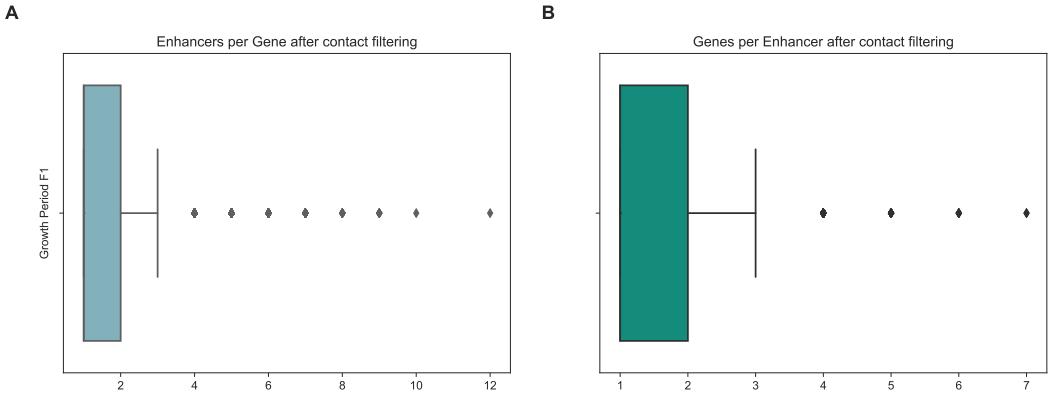


Figure 19: Gene-enhancer relationships after contact filtering. Gene-enhancer interactions are predicted using data from growth period F1. **A.** Distribution of the number of enhancers that regulate a given gene. **B.** Distribution of the number of genes that are regulated by a given enhancer. The application of the contact percentage filter limits the number of genes a given enhancer regulates, bringing the values more in line to previously suggested distributions

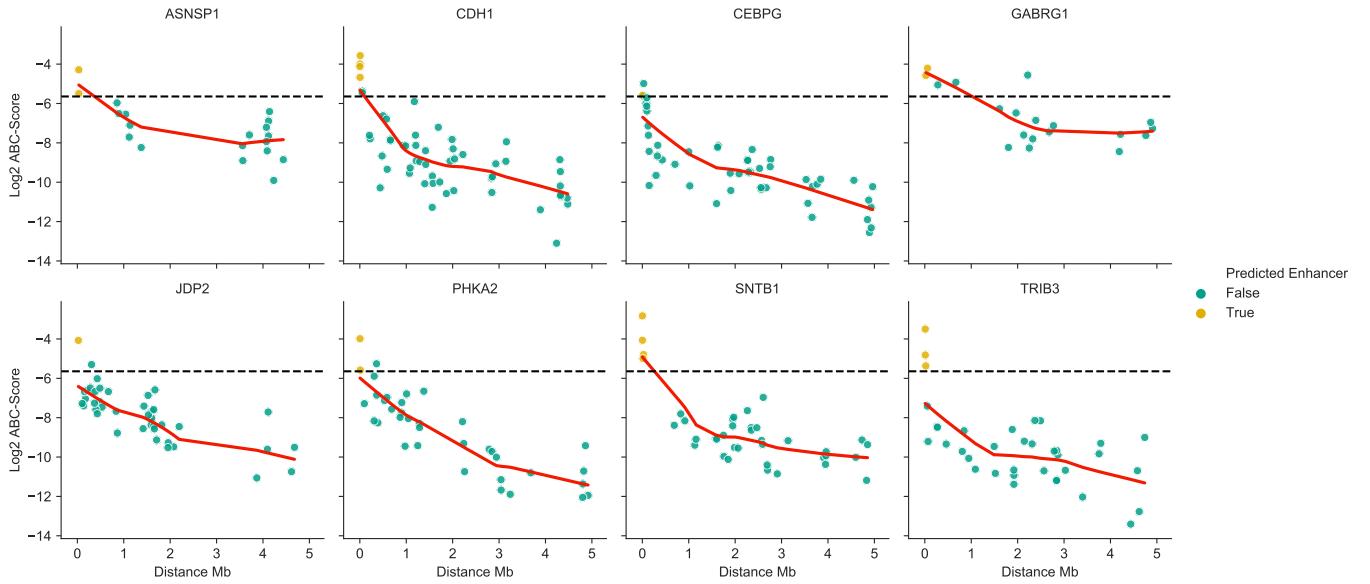


Figure 20: Enhancer prediction for genes after applying a contact filter. ABC-scores are estimated using data from sample F1. For the presented set of genes, the ABC-score for candidate regulatory elements are shown along the y-axis, and the distance to the candidate enhancer is shown along the x-axis. Enhancers that pass both the ABC-score threshold (0.02, shown as black line) and the contact percentage filter (8.5%) are shown in yellow. All other negative interactions are shown in blue. Loess regression shown as red line. The additions of the contact filter means that not all enhancers that overcome the ABC-score threshold are considered true regulators.

of a cell-type averaged Hi-C dataset should be able to provide meaningful predictions (Fulco et al., 2019). Previous studies have shown that stimuli-responsive enhancers are already in spatial proximity to their regulated gene, even before the stimulating event occurs (Gasperini et al., 2020; Jin et al., 2013; Ray et al., 2019). Therefore, in the context of glucose-starvation, the addition of the contact percentage filter, using the averaged Hi-C dataset provided by the authors of ABC, should not limit the study of differentially active enhancers. Given that the proposed contact percentage filter reduces the number of genes an enhancer is predicted to regulate, it is possible that it will improve the precision of the ABC model. It will also clarify the global relationship between enhancers and genes, as can be seen in Figure 21.

By considering a minimum number of shared contacts, the predicted regulatory interactions between enhancers and their target genes can reveal a clearer picture of chromatin topology. An example of this are the predictions made for the two enhancers found downstream of the CPEB4-204 transcript (Figure 21). When only considering an ABC-score threshold, these two enhancers are predicted to regulate multiple genes that are found at various distances along the linear genome, including the closest gene, CPEB4. Once the contact percentage filter is taken into account, the same enhancers are only predicted to regulate BOD1. These interactions become examples of enhancers skipping the closest gene in the linear genome (CPEB4) to maintain long-range regulatory relationships with their target gene. These enhancers and the BOD1 promoter are likely brought in close spatial proximity by chromatin looping; a claim that is further supported by the prediction of short-range gene-enhancer interactions between

the BOD1 gene and its regulatory elements. As stated before, the inclusion of the contact percentage filter reduces the number of predicted regulatory interactions; however, Figure 21 reveals that these interactions might recapitulate the spatial conformation of chromatin more accurately.

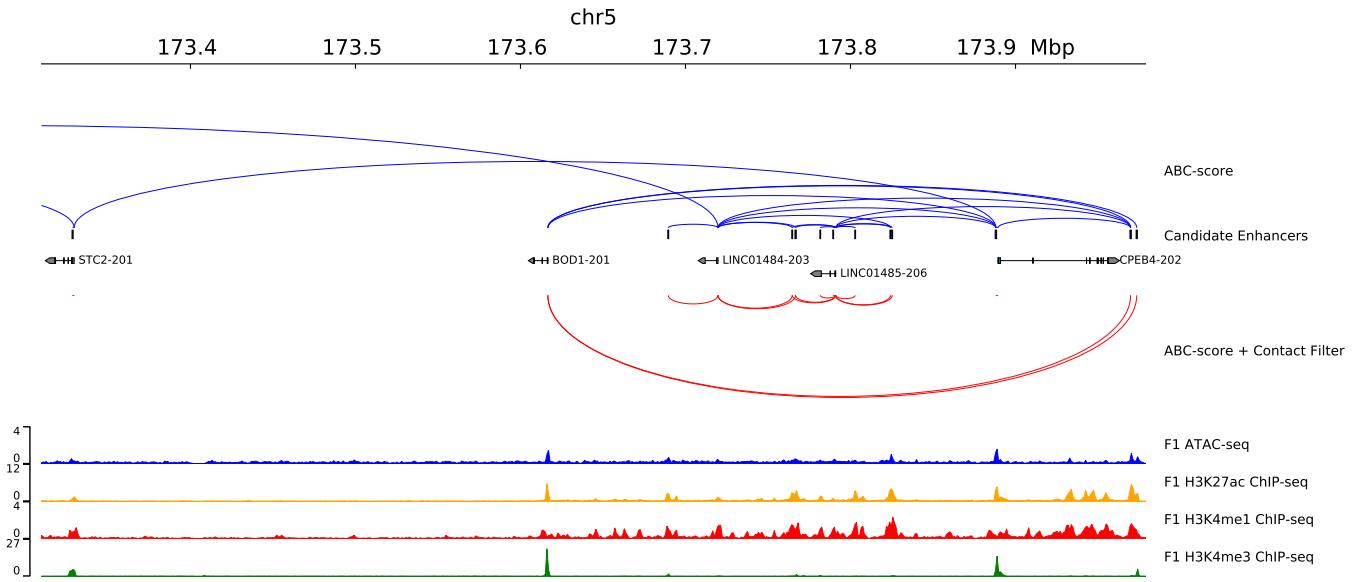


Figure 21: Comparison of predicted gene-enhancer regulatory interactions. Here we show the predicted regulatory interactions, together with ATAC-seq and ChIP-seq profiles for genomic coordinates in chromosome 5 173309713 to 173978817. ATAC-seq and ChIP-seq profiles from sample F1 are shown as FP-U normalized CPMs. Regulatory predictions made by only considering the ABC-score threshold are shown as blue arcs (labeled ABC-score). Gene-enhancer pairs predicted by including the contact percentage filter are shown as red arcs (labeled ABC-score + Contact Filter). The use of the additional contact filter reduces the number of interactions per enhancer and provides a clearer picture of chromatin topology.

As was previously observed in Figure 14, ABC-scores for candidate enhancer elements tend to be higher when estimated using datasets gathered during glucose starvation. This would lead to many enhancers passing the ABC-score threshold for a greater number of genes. Therefore, to further limit the effect of false gene-enhancer pairs in subsequent analysis, we use data from growth period F1, as the activity of candidate enhancers has not been perturbed by glucose starvation or repeated stress. We use these criteria (F1 ABC-score > 0.02 and contact percentage $> 8.5\%$) in order to infer a set of gene-enhancer regulatory interactions that are analyzed for their response to glucose starvation.

5.6 Distinct enhancer responses to glucose starvation

As was observed in Figure 10, changes in chromatin accessibility and presence of H3K27ac modified histones occur at ATAC-seq peaks. Furthermore, some of these peaks are predicted to be enhancers, according to the criteria described in the previous section. Therefore, we were interested in exploring the different types of activity responses these predicted enhancers exhibit

in when faced with glucose starvation, and how they might influence the expression of the genes they regulate. To do this, we clustered each enhancer's fold change in weighted-activity with respect to its values during glucose-fed periods of growth (Figure 22A). The average profile of weighted-activity fold-change can be observed in Figure 22B. This allowed us to observe clear increases and decreases in activity during starved periods of growth (Cluster 1 and Cluster 2). Noteably, enhancers that had been previously identified as DA-Up peaks can be found together with other enhancers that increase their activity during glucose-starvation (Figure 22A). Additionally, we can also observe patterns of activity response that seem to depend on continuous stress (Clusters 3 and 4).

To determine the influence that these changes in enhancer activity might have on the expression regulated genes, we explored the average expression fold-change profile of genes grouped by the cluster of enhancers that they are regulated by (Figure 22C). Interestingly, changes in gene expression to closely mimick the weighted-activity response of the enhancers they are regulated by (comparing Figure 22B and C). This is especially true for genes regulated by enhancers in Cluster 2, as their increases in expression during glucose starvation correspond to increases in activity of their enhancer regulators. A weaker relationship can be observed between the weighted-activity profile of Cluster 1 and the genes they regulate. While genes regulated by this cluster of enhancers does seem to decrease in expression in the same growth periods where enhancer activity decreases, the overall trend points to a growth in transcriptional activity. Closer associations can be observed between Clusters 3 and 4 and the genes they are predicted to regulate. Interestingly, both gene expression and enhancer activity exhibit a response to glucose starvation that seems to be related to repeated stress.

While clustering weighted-activity profiles can give a general impression of enhancer behaviour, it is not a robust method of determinining significant increases, or decreases, in activity in response to starvation. In the case of Cluster 1 this could explain why the average expression profile of the genes they are predicted to regulate, does not mimick enhancer behaviour so closely. Alternatively, a weaker correspondence might be due to the fact that their genes might be regulated by multiple enhancers (Figure 19A). The combined regulatory effect of many enhancers might lead to an expression profile that cannot be easily explained by one single enhancer behaviour. This possibility led us to explore enhancer behaviour, grouped by the transcriptional response of the genes they regulate (Figure 23A).

Strikingly, when grouped by transcriptional response, the average enhancer weighted-activity profile closely mimicks the behaviour of their regulated genes. Meaning, for example, that genes that present increased transcription during glucose-starvation, are regulated by enhancers that show a net increase in their weighted-activity during the same growth periods. This correspondence is also observed for genes with down-regulated expression and genes whose transcriptional activity does not change during glucose-deprived growth.

Interestingly, gene promoter activity does not exhibit this same level of similarity (Figure 23B). When observing genes with up-regulated expression during glucose starvation, we notice that the promoters of these genes do not show a net increase in activity in the same growth periods. However, promoters of genes that decrease in expression during Starved growth periods, do seem to show a profile that closely mimicks the change in expression.

These behaviours suggest the existence of potential regulatory dynamics between promoters and enhancers. The combined net change in enhancer activity closely resembles the transcriptional response of the genes they regulate. Additionally, the transcriptional response of

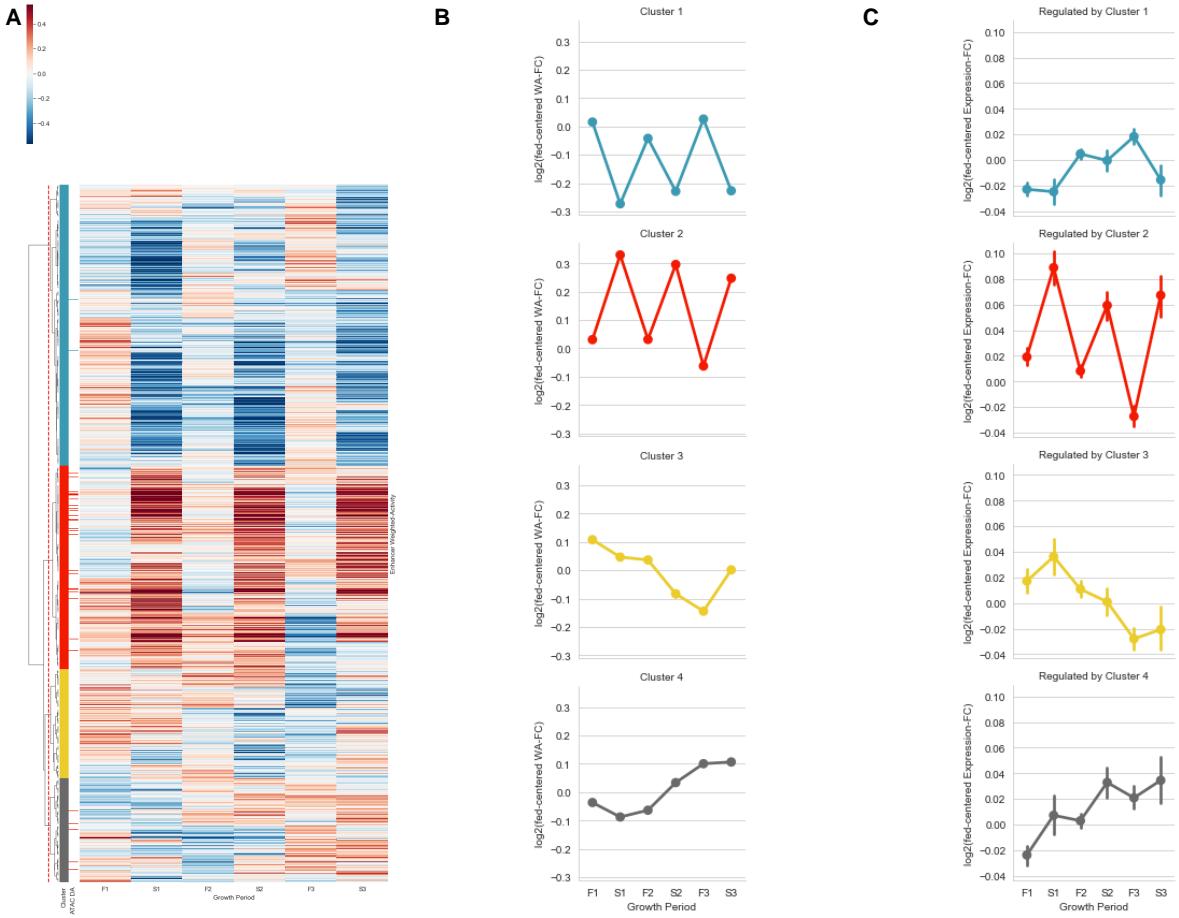


Figure 22: Clustering of enhancer weighted-activity response profiles. The logFC in weighted-activity (WA) was estimated by subtracting the average log WA values of samples F1, F2 and F3 from the log WA values in all samples. Gene expression logFC was estimated in the same way. **A.** Hierarchical clustering of enhancer WA-FC values in each growth period (ward clustering on cosine distance). Cluster dendrogram is cut at a height of 30. *Cluster* column indicates the cluster a row belongs to, after dendrogram partition. *ATAC-DA* column indicates whether an enhancer was determined to be an DA-Up peak (red), a DA-Down peak (blue), or neither (white). **B** Average enhancer WA-FC values (+/- standard error) for each cluster formed in **A**. **C**. Average expression FC of genes (+/- standard error) in each growth period, grouped by the cluster of enhancers they are regulated by. Enhancer activity profile mimicks the transcriptional response of the genes they regulate; there are some exceptions.

up-regulated genes is more closely correlated to the net changes in activity observed in the enhancers that regulate them, rather than the changes occurring at their promoters. This could indicate that enhancers might play a more active role in the achieving an increased expression of the genes they regulate, in comparison to their promoters. For genes that become down-regulated, both promoters and enhancers show a decrease in activity, meaning that a combined effect of these regulatory elements achieves the observed decrease in expression in the genes they regulate.

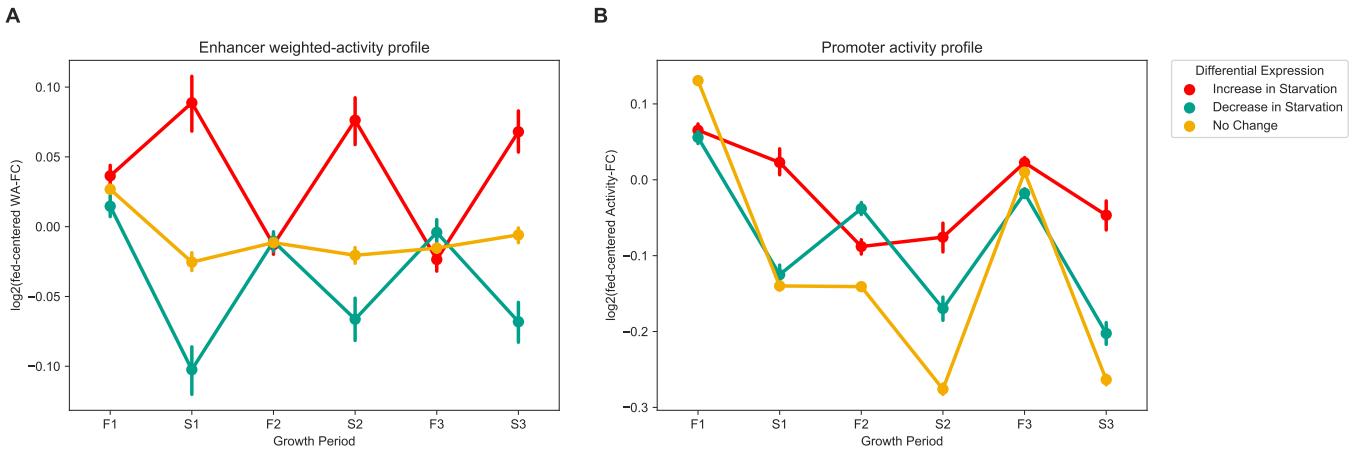


Figure 23: Enhancer and promoter activity grouped by gene transcriptional response. Genes are classified as either having increased expression during starvation, decreased expression during starvation or non-changing according to the results of differential expression analysis (Figure 2). The average log enhancer WA-FC and promoter activity FC with respect to the average nutrient-rich values is shown (+/- standard error). **A.** Average WA profile of enhancers, according to the transcriptional response of the genes they regulate. **B.** Average activity profile of promoters, according to the transcriptional response of the genes they belong to. Net enhancer activity profiles closely follows the transcriptional response of the genes they regulate; this phenomenon is not observed for the genes promoters.

5.7 Differentially Accessible Enhancers

As was previously mentioned in Figure 22A, some enhancers that present increases in their weighted-activity during glucose starvation, were also previously determined to be DA-Up ATAC-seq peaks. Given that this group of regulators has shown increased accessibility in response to glucose starvation, we looked more closely into the possible effect they might have on the expression of the genes they are predicted to regulate, as well as the pathways they participate in. Of the 676 peaks that increased in accessibility during starvation, 207 were determined to be enhancers.

When exploring the expression profile of the genes regulated by DA-Up enhancers, we observe a variety of transcriptional responses to glucose-starvation. On average, genes regulated by these differentially accessible enhancers do increase in expression during starvation (Figure 24A). Indeed, several of the genes that are regulated by DA-Up enhancers are identified to have significantly increased differential expression (DA-Up Regulated Gene Cluster 2). However, we also notice other gene expression patterns that do not necessarily correspond with the increased activity of these enhancers. In DA-Up enhancer regulated gene Clusters 3 and 5, we observe genes whose expression decreases in response to starvation, or decreases in the face of repeated stress. The genes in Regulated Cluster 4 do not exhibit obvious changes in gene expression. Interestingly, Cluster 1 exhibits continuous expression growth (Figure 24B).

While there is heterogeneity in the expression profiles of genes regulated by DA-Up enhancers, this is not entirely unexpected. As was mentioned previously, the many-to-many relationship between enhancers and genes could lead to complex expression dynamics, not easily explained by one regulatory behaviour. The combined regulatory effect of enhancers

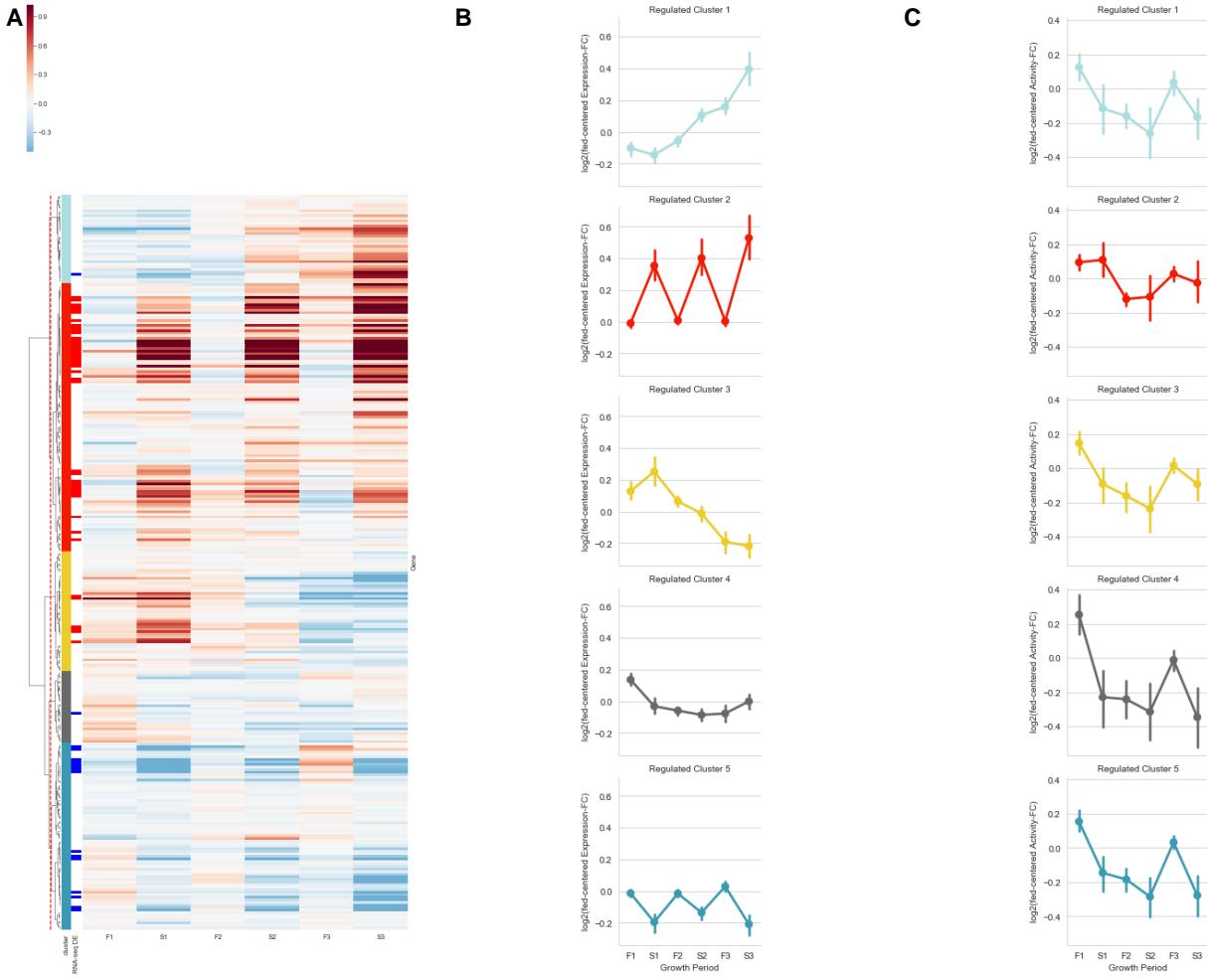


Figure 24: Gene expression regulated by DA-Up enhancers. The logFC in gene expression was estimated by subtracting the average log expression values of samples F1, F2 and F3 from the log expression values in all samples. The logFC in promoter activity was estimated in the same way. **A.** Hierarchical clustering of expression-FC values in each growth period (ward clustering on cosine distance). Cluster dendrogram is cut at a height of 30. *Cluster* column indicates the cluster a row belongs to, after dendrogram partition. *RNA-seq DE* column indicates whether a gene was determined to be differentially up-regulated (red), a down-regulated (blue), or neither (white). **A.** Average gene expression-FC profile according to the transcriptional response of the genes they regulate. **B.** Average gene expression-FC values (+/- standard error) for each cluster formed in **A**. **C.** Average activity FC of promoters (+/- standard error) in each growth period, grouped by the cluster of genes they regulate. A variety of transcriptional responses appear, some of which show increased expression during enhancer accessibility.

regulating genes Regulated Clusters 3 and 5 might present a net decrease in activity during glucose starvation. Additionally, some DA-Up enhancers are predicted to regulate more than one gene (207 DA-Up enhancers regulate 306 genes); these types of relationships could represent a distribution of the effect of these enhancers. Finally, even though the use of the contact percentage filter should greatly improve the precision of the relationship predictions, there is still room for misclassifications.

When observing the average activity profiles of gene promoters, we can notice that for all Regulated Clusters, except cluster 2, there seems to be a marked decrease in activity in response to glucose starvation (Figure 24C). The promoters of genes in DA-Up Regulated Cluster 2, do present slight changes in activity during the different growth periods, but for the most part their activity remains stable in spite of repeated stress. This observation matches the result presented in Figure 23B, where the promoters of up-regulated genes do not present large decreases in activity in response to glucose availability. Therefore, the increased expression during starved growth periods observed in genes belonging to DA-Up Regulated Cluster 2, cannot be solely explained by changes in activity at their promoters; rather, this up-regulated expression can be explained by increases in activity at the enhancers that regulate them; specifically, by a significant increase of chromatin accessibility during glucose starvation.

An example of these behaviours can be noted for the gene MKNK2 in Figure 25. Here, we observe that MKNK2 is receives the regulatory input of three enhancers; two are found upstream of the TSS, and the third is located within MKNK2's first intron. The classification of these three ATAC-seq peaks as enhancer regulators, is further supported by the presence of H3K4me1 modified histones and little to no H3K4me3 signal. One of these enhancers (colored in red), is a DA-Up peak whose differential accessibility during glucose starvation can be noted in the ATAC-seq signal profiles from the different growth periods. We also notice an increased signal of H3K27ac modified histones during periods of glucose-starved growth. The combined increase in chromatin accessibility and presence of H3K27ac modified histones results in an elevated activity estimate when cells are grown in a glucose-depleted medium. The increased presence of H3K4me3 modified histones at MKNK2's TSS confirms that this gene is being transcribed. However, the ATAC-seq signal profiles at this location do not reveal an obvious increase or decrease in chromatin accessibility in this region. Upon close inspection of the H3K27ac signal, we do notice slight increases in response to glucose starvation, which is in line to the observation made in Figure 10D. However, the combined input of ATAC-seq and H3k27ac ChIP-seq does not yield an noticeable increased promoter activity during glucose starvation. The other two enhancers that regulate MKNK2 maintain a constant ATAC-seq and H3K27ac signal. The combined effect of the three enhancers regulating this gene results in a net increase of enhancer activity during starvation. These observations suggest that MKNK2's increased transcription might be mainly due to the increased activity of the upstream enhancer that regulates it.

Given that the increased activity of DA-Up enhancers can explain the up-regulated expression of genes in Regulated Cluster 2 during glucose starvation, we explored the metabolic pathways that the products of these genes participate in (Figure 26). Enrichment analysis reveals that they play a role metabolic pathways dealing with the use of nutrients, such as the import of amino acids, carbohydrate metabolism, and the metabolism of lipids. Additionally, these genes are also associated with health conditions such as insulin resistance.

The results presented in this section indicate the existence of ATAC-seq peaks with an increased accessibility during glucose-starvation, that exhibit the characteristics of active enhancers. Furthermore, these DA enhancers are predicted to regulate genes with various transcriptional responses to starvation, a group of which presents an increased expression during periods of growth where the enhancer is more accessible. This increase in activity is not observed in the promoters of this group of regulated genes, suggesting that the enhancer plays a central role to establish the increased transcriptional response. Finally, this group of genes is

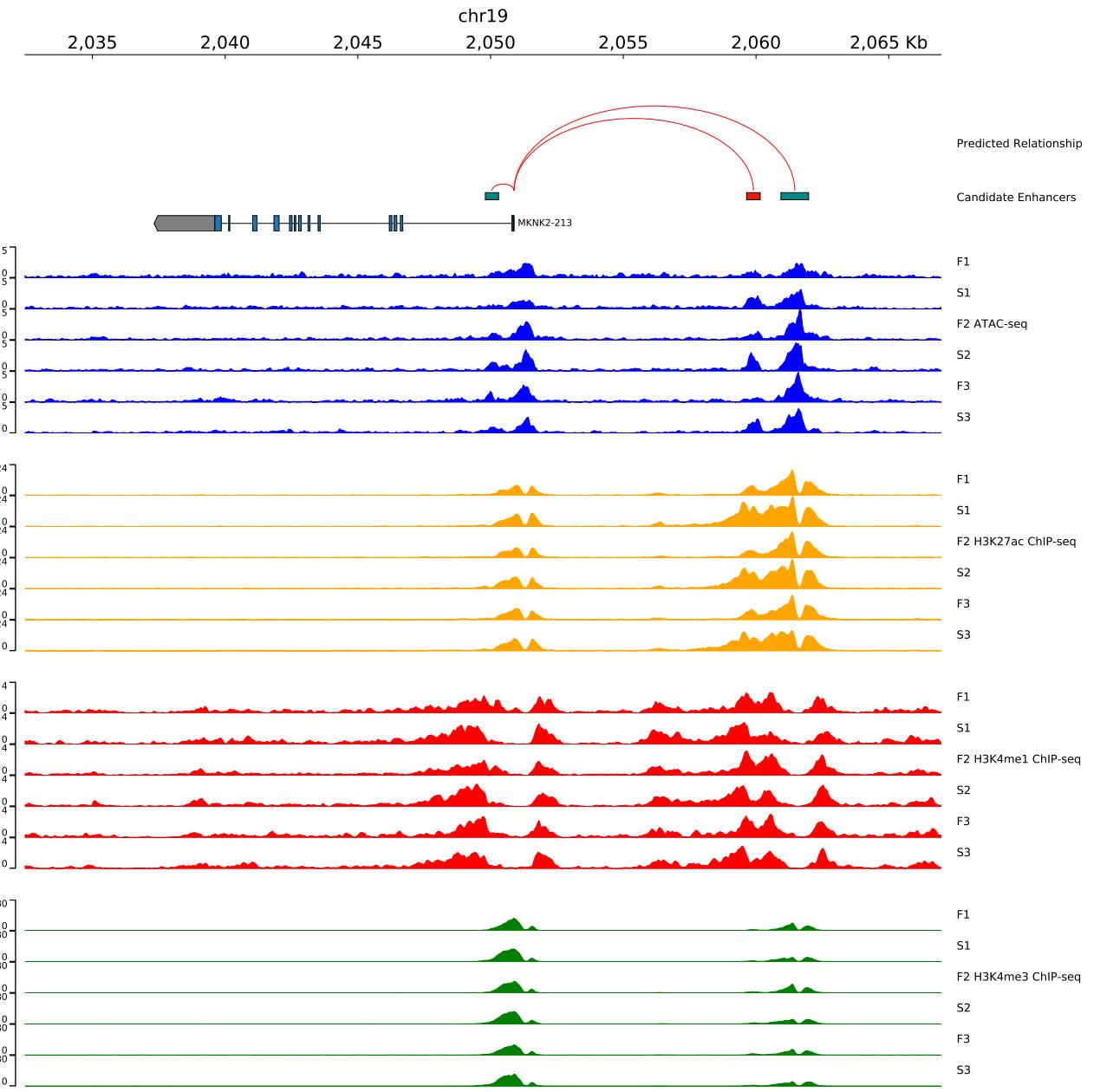


Figure 25: Enhancer regulation of MKNK2. Here we show the predicted regulatory interactions, together with ATAC-seq and ChIP-seq profiles for genomic coordinates in chr19:2032465-2066981 (hg38). ATAC-seq and ChIP-seq profiles are shown as FP-U normalized CPMs. ATAC-seq peaks are colored according to their DA classification: DA-Up peaks are shown in red, DA-Down peaks are colored in purple (none appear here), and all other peaks are colored in green. Regulatory relationships between genes and enhancers predicted by ABC are shown as red arcs. The combined regulatory input of the enhancers regulating MKNK2 yields a net increase in enhancer activity the can explain the up-regulated expression of this gene during glucose starvation.

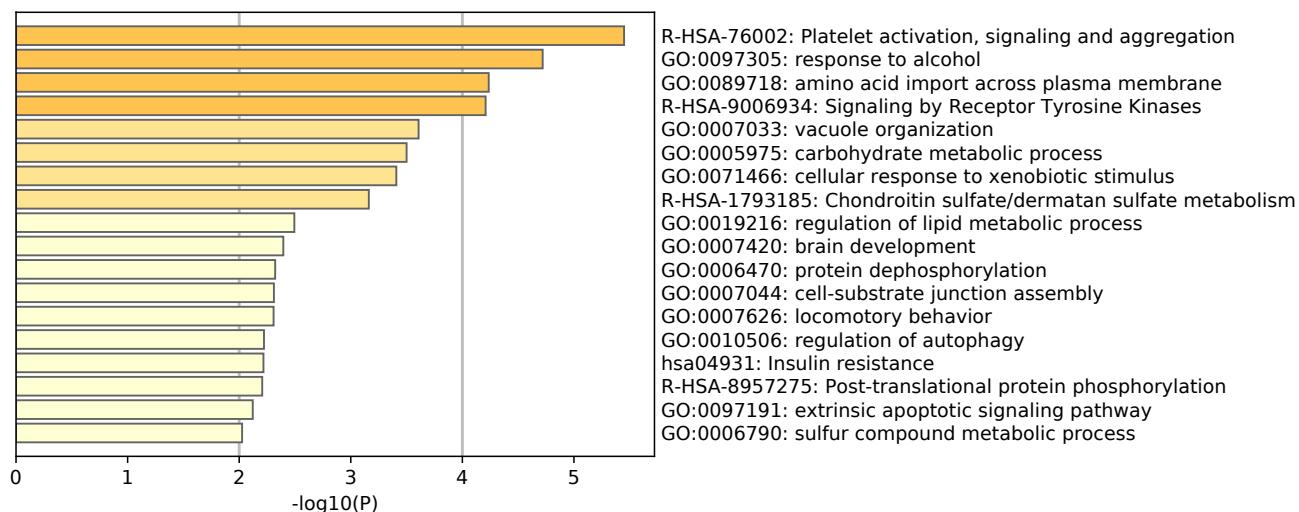


Figure 26: Metascape enrichment analysis of DA-Up Regulated Cluster 2. Metascape enrichment analysis of genes present in Regulated Cluster 2 24A and B. Genes in this group are involved in nutrient metabolism and the appearance of insulin resistance.

present in pathways associated with the metabolic processing of nutrients, and their misregulation might play a role in the appearance of common diseases, such as diabetes. In summary, enhancers that exhibit increased accessibility in response to glucose starvation are suggested to play a central role in the up-regulation of gene products that participate in the adapted used of nutrients and whose mis-regulation is linked to the appearance of metabolic diseases.

5.8 Discussion

Throughout the course of the present work, we encountered several difficulties when trying to compare epigenetic datasets (ATAC-seq and ChIP-seq) gathered from different growth conditions. Dealing with these difficulties led to many interesting observations and has revealed interesting points about the regulation of differential gene expression.

Among the first difficulties we faced, was the correct normalization of ATAC-seq datasets. While normalization strategies have been established for the comparison of other assays, such as RNA-seq, the procedures for the normalization of ATAC-seq and ChIP-seq still vary from author to author (Reske et al., 2020). Many such methodologies exist and can be found in various sources of information (Reske et al., 2020; Lun & Smyth, 2016). We find that the choice of normalization procedure should depend on the biases present in the data that one is analyzing. Several strategies must be applied in order to determine which one is the most adequate. However, this variation in applied normalization techniques would further complicate the comparison against similar datasets generated by other groups (Reske et al., 2020). A careful and reproducible technique during the performance of experiments, must be a high priority when performing ATAC-seq (or ChIP-seq) with goal of future differential analysis.

In our attempt to link the differential activity of transcriptional regulatory elements with the observed gene expression response, we found that neither fold change in chromatin accessibility or fold change in the presence H3K27ac modified histones at gene promoters seemed to fully explain the corresponding gene's transcriptional output (Figure 10C and D). However, as can be

observed in Figure 10D, for genes exhibiting an elevated increase in gene expression, there is a correlation with the acquisition of H3K27ac modifications at their promoter regions; conversely this relationship is not maintained for genes that decrease in expression. An explanation for this behaviour might be that H3K27ac modified histones, while being a characteristic of active regulatory elements (Creyghton et al., 2010), are not essential for these elements to carry out their function. A recent study give support to this claim, by observing that the loss of H3K27ac modifications at active enhancers does not inhibit the regulatory activity of these regions (T. Zhang, Zhang, Dong, Xiong, & Zhu, 2020). Therefore, it could be expected that the loss of H3K27ac at gene promoters does not necessarily lead a decrease in expression. Even so, when observing the average activity response (according to Equation 1) to glucose-starvation in gene promoters, we note that it does not correspond to the increased expression of up-regulated genes.

While promoter activity does not fully explain the all the types of transcriptional response to glucose, we do find a more clear relationship between the average enhancer activity profile and gene expression response (Figure 22). In most cases the general behaviour of enhancer activity in response to glucose starvation is reflected in the expression pattern of the genes that they are predicted to regulate. However, we also find that some expression patterns, particularly those associated with down-regulated transcription, is not so easily explained by a single type of enhancer behavior. This can be explained by previous observations of enhancer activity during the development of *Drosophila melanogaster*. There, it has been observed that the regulatory input of several enhancers on a single gene can be complementary and additive (Bothma et al., 2015; Dunipace, Ákos, & Stathopoulos, 2019). Our results suggest that these enhancer properties might also play a role in the cellular response to glucose availability.

We notice that the average enhancer activity profile resembles the differential transcriptional response of the genes they regulate, while the same is not always true for promoters 23. A lack of change in promoter activity, combined with a net increase in enhancer activity, might explain the observed increase in transcription for up-regulated genes. A combined decrease in enhancer and promoter activity during glucose-starvation also correlates with the decreased expression of the genes they regulate in the same growth periods. This points to some possible interesting dynamics between promoters and enhancers. Future efforts might attempt to model a gene's transcriptional response to glucose starvation by considering the net gain or loss of activity in its corresponding enhancers and promoter.

It must be recognized that, while we make use of measurements of chromatin accessibility and presence of H3K27ac modified histones to compare the activities of enhancers and promoters, these data might not represent the most accurate portrayal of regulatory potential. Promoters, for example, can also be described by the presence of H3K4me3 modified histones (Soares et al., 2017), while enhancers are also characterized by the presence of H3K4me1 histones (Shlyueva et al., 2014). It is possible that by incorporating these signals, new patterns will be revealed. This highlights the need for a more basic understanding of the role of these different PTMs during regulatory activity, as well as the need to dissect the effects of their combined presence.

We note that the large-scale identification and target-linking of enhancers remains a challenge. While the Activity by Contact model represents a great advancement, it is still prone to errors, specifically by considering only potential regulators from the point of view of a gene, while failing to consider a ranking system that determines which gene an enhancer is likely

to regulate. We propose an additional criteria, that consists of requiring a minimal spatial proximity between candidate enhancers and promoters, in order for a regulatory interaction to be considered as true. This is based on the fact that previous studies have found that stress-induced enhancers are found in spatial proximity to their target gene, even before the inducing event takes place (Gasperini et al., 2020; Jin et al., 2013; Ray et al., 2019). It still remains to be evaluated whether this consideration truly increases the precision of the ABC algorithm, without sacrificing most of the recall value. If so, more development would be required in order to properly incorporate this measurement into a single predictive score.

Even so, the predictions made with the currently presented algorithm reveal interesting relationships between genes and their enhancers. Additionally, we were able to link enhancers with increased accessibility during glucose-starved periods of growth, to genes whose transcription is also increases in the same growth conditions. The fact that the increased accessibility of these enhancers corresponds to increased expression, points to a possible mechanism where these exposed DNA segments allow for the binding of transcription factors that enhance gene transcription. Further analysis is necessary to reveal the regulatory proteins that participate in the up-regulation of genes by binding to their differentially accessible enhancers. This set of genes is involved in the metabolism of nutrients, and are associated with the appearance of pervasive health conditions. The relationship between these enhancers and their regulated genes could be further assessed via experimental techniques such as CRISPRi. By altering the DNA sequence of DA-Up enhancers, we expect that the genes they are predicted to regulate may no longer exhibit the increases in expression in response to glucose starvation. The aberrant activity of these enhancers might play a role in establishing a transcriptional program that leads to insulin resistance, making them interesting candidates for the future development of gene therapies.

Finally, here we have described the relationships between changes in gene transcription and the modulation of the activity of their regulatory elements in response to glucose-starvation, in the Huh7 cell-line. By comparing these behaviours, to those that present themselves in a non-cancerous cell-line, we might be able to determine which differentially active regions are responsible for promoting cancer-cell survival in a nutrient-depleted medium.

6 Methods

6.1 Cell Culture

Experiments were performed using the Huh7, human hepatocarcinoma cell line. Cells were cultivated in Dulbecco's Modified Eagle's Media (DMEM) - high glucose with 4500 mg/L (25mM) glucose and sodium bicarbonate, supplemented with L-glutamine, Penicillin-streptomycin, Non-essential amino acids, and sodium pyruvate.

6.2 Glucose Starvation

Cells were counted and plated, then cultured in DMEM with 25mM of glucose for 24 hours. To expose Huh7 cells to glucose-starved conditions, they were first washed 2 times with PBS and then DMEM media that contained 0mM of glucose was added. For refeeding, 0 mM glucose DMEM media is removed and 25mM glucose DMEM is added to the plate.

6.3 RNA-seq

Experimental Procedure

RNA extraction: Between 1e5 to 2e6 cells were either trypsinized and pelleted, or lysed in the cell culture dish and processed according to manufacturer's directions with the Quick RNA Miniprep kit, including a 15 minute DNase treatment. They were then eluted in 3 μ l RNase, DNase-free water. RNA is measured using Multiskan plate reader, uDrop microplate.

Sequencing: RNA samples were submitted to the Institut de genetique et de biologie moleculaire et cellulaire (IGBMC) Sequencing platform. Libraries were prepared by the platform, using TruSeq Sample Prep Kit for stranded mRNA-seq. Single ended 50 base pair sequencing was performed. Reads were mapped onto the hg38 assembly of human genome using Bowtie2 v2.1.0 aligner (Langmead & Salzberg, 2012).

Differential Expression Analysis

The GENCODE v33 GRCh38, primary genome assembly sequence and annotations were used to generate a decoy transcriptome using SalmonTools. A Salmon (Patro, Duggal, Love, Irizarry, & Kingsford, 2017) index was created for GENCODE v33 transcripts, against the aforementioned decoy transcriptome. Index k-mer length was set to 21.

Adapter sequences were trimmed from fastq reads by Trim Galore!. Trimmed reads were used to quantify transcript abundances by Salmon. Salmon was run with the –seqBias –gcBias –validateMappings –numGibbsSamples 200 parameters; as well as –fldMean and –fldSD parameters specifying the fragment length distributions of single-end reads.

The resulting transcript quantification was analyzed using DESeq2 (Love, Huber, & Anders, 2014). Transcripts were aggregated to the gene level, and genes which did not have at least 10 reads in at least 3 samples were not considered for further analysis. Additionally, the most expressed isoform on average across all growth periods of each gene was annotated.

Transcript abundances were modelled as *pair* * *starvation*, where *pair* tracks related sample pairs (F1-S1, F2-S2 and S3-F3) and *starvation* refers to whether the same was gathered during glucose starvation or glucose-fed conditions. Statistical testing was done on the estimates of the *starvation* parameter and *starvation:pair* interaction terms. Null hypothesis assumed $|\log FC| < 0.1$, significance was assumed at adjusted p-value of 0.05. In our analysis, genes are considered to be differentially expressed when their $|\log FC| > 0.1$ and adjusted the adjusted p-value for the *starvation* coefficient was lower than 0.05.

The Metascape (Zhou et al., 2019) Express Analysis in the online-portal was used to carry out functional enrichment of gene sets.

6.4 Assay for Transposase-Accessible Chromatin using sequencing

Experimental Procedure

Huh7 cells were treated with trypsin. 50,000 cells were counted and placed in 50 μ l of TE in a 1.5ml tube. These tubes were centrifuged at 1200 rpm and the resulting supernatant was discarded. Lysis buffer (10 μ l of 1M Tris·Cl, pH 7.4 (final 10 mM), 2 μ l of 5M NaCl (final 10 mM), 3 μ l of 1M MgCl₂ (final 3 mM), 10 μ l of 10% NP-40 (final 0.1% v/v), and 975 μ l nuclease-free H₂O) was freshly prepared on ice. 50 μ l of this buffer was added to the cell pellet,

which was resuspended by repeated pipetting. Lysed cells were centrifuged at 500 x g for 10 minutes at 4°C and the supernatant cytoplasm was discarded.

Transposition reaction mix was prepared using the Nextera DNA Library Prep. For samples with a volume of 50,000 cells, 50 μ l transposition reaction mix was prepared with 25 μ l of 2X TD Buffer, 2.5 μ l of Tn5 Transposase and 22.5 μ l of nuclease-free H₂O. This mix was added to the previously formed pellet and the nuclei were resuspended. This mixture was incubated at 37°C for 30 minutes. Next, the DNA was purified using Qiagen MiniElute Reaction Cleanup Kit with the final elution of eDNA in 10 μ l Elution Buffer (EB).

Library preparation was done by amplifying DNA with indexed primers. A volume of 10 μ l of purified transposed DNA was mixed with 10 μ l of nuclease-free H₂O, and 2.5 μ l of Ad1-noMX primer (25 μ M), 2.5 μ l of Ad2.* (*depending on index number) indexing primer (25 μ M), and 25 μ l of NEBNext High-Fidelity 2X PCR Master Mix, in a PCR tube and amplified. The following cycle was used:

Phase	Temperature	Duration	Cycles
	72°C	5 min	
Initial Denaturation	98°C	30 sec	1
Denaturation Elongation	98°C	10 sec	5
	63°C	30 sec	
Hold	72°C	1 min	1

To determine how many more cycles of amplification were necessary, 5 μ l of the previously partially amplified library was removed and placed in a new tube. 4.41 μ l of nuclease-free H₂O, 0.25 μ l of Ad1-noMX primer, 0.25 μ l of Ad2.* indexing primer, 0.09 μ l of 00X SYBR Green I and 5 μ l of NEBNext High-Fidelity 2X PCR Master Mix were added to this 5 μ l of partially-amplified library. This mixture was placed in a 96 well plate, mixed, and subjected to the following program:

Phase	Temperature	Duration	Cycles
Initial Denaturation	98°C	30 sec	1
Denaturation Elongation	98°C	10 sec	20
	63°C	30 sec	
Hold	72°C	1 min	1

In order to calculate the number of additional PCR cycles needed for each sample, we examined the plot of R vs Cycle Number and determined the number of cycles needed to reach 1/3 of the maximum R. Using the remaining 45 μ l of the partially-amplified library run the PCR for the appropriate number of cycles (in our experiment it was 7 additional cycles). To remove primer dimers, single left-sided bead purification was performed by adding 1.8X volume (81 μ l) of AMPure XP beads and mixed by pipetting up and down. This mixture was incubated at RT for 10 minutes. These tubes were then placed on a magnetic stand and the cleared supernatant was discarded. While on the magnetic stand, the beads were washed twice with 80% freshly prepared ethanol. After the second wash, the residual liquid was removed and beads were dried for 5 minutes at RT. Once tubes were removed from the magnetic stand, 20 μ l of nuclease-free H₂O was used to elute the DNA from the beads, by incubating for 2 minutes at RT. The tube was then placed back on the magnetic rack; 20 μ l of the cleared supernatant was transferred to a new tube and stored at -20°C.

The library quality was assessed by testing a 1:3 dilution of the prepared library in water, on a Bioanalyzer high sensitivity chip and with the Qubit high sensitivity DNA kit. ATAC libraries were submitted to Helmholtz Next Generation sequencing core facility for 50 bp paired-end sequencing on the Illumina 4500 platform. Sequence reads were mapped to reference genome hg38 using Bowtie v2.1.0.

Differential Accessibility Analysis

ATAC-seq bamfiles were used to detect regions with enriched ATAC-seq signal, using the MACS2 software. The callpeak command was used with the –keep-dup all and -g hs parameters. We then created a concatenated list containing all of the ATAC-seq peaks contained in each sample from each growth period; peaks with overlapping coordinates were merged into a single peak.

The GENCODE v33 GRCh38, primary genome assembly annotations were used to annotate ATAC-seq peaks, based on their location. Annotation was carried out by the ChIPseeker v.1.8.6 R package. Accordingly, peaks were classified as promoters, introns, exons, 5' UTR, 3' UTR or distal intergenic. Regions within 2000 bp of a gene most expressed transcript's TSS were considered to be within the promoter of said gene.

For normalization, many strategies were attempted several normalization strategies: The number of reads that overlap with a given peak in each sample was determined using the GenomicAlignments (v1.24.0) R package. We counted reads that overlapped any portion of exactly one peak, and had a mapping quality above or equal to 20. Reads that overlap multiple peaks were discarded. We discarded peaks that had less than 0.75 CPM but maintained the original library size. The remaining ATAC peaks' counts were normalized using the 'tmm' (FP-T) or 'upperquartile' (FP-U) method provided in the edgeR (v3.28.1) calcNormFactors function.

Alternatively, the CSAW (v1.22.1) package was used to count the number of reads within ATAC-seq peaks. Only reads with a minimum mapping quality of 20 were considered. At the same time, for the estimation of normalization factors, the number of reads found within overlapping 300 bp genomic sliding windows was used (B-T). In an additional step, we conserve peaks whose signal is at least $\log_2(3)$ times greater than its surrounding 2000 bp surroundings (LE-T).

Principal Components Analysis was carried out using the prcomp function from the R programming language. Analysis was performed on the number of reads within each filtered ATAC-seq peak, before and after normalization.

To identify ATAC-seq peaks with differential accessibility, we consider all samples where cells are starved and fed as their respective replicates. Significant changes were determined using the Quasi Likelihood F-test (using edgeR) and a False Discovery Rate of 0.05.

6.5 Chromatin Immunoprecipitation

Experimental Procedure

Huh7 cells were grown on 15cm^2 plate (around $1\text{e}7$ cells per 15cm^2 plate) and crosslinked at room temperature (RT), under a fume hood with 1% formaldehyde (0.583ml of 37% formaldehyde solution in 21 ml cell culture media) for 10 minutes with intermittent agitation. These

plates were quenched with 1.5 ml of 2.5M glycine for 5 minutes with intermittent agitation. Afterwards, they were aspirated and washed 2 times with 20 ml of cold PBS. Cells were then harvested by using a cell filter in 5ml of Scraping buffer (PBS with NaBu and Complete protease inhibitor). They were then centrifuged at 1260g (2500rpm) in 4°C for 10 minutes; the supernatant was aspirated and the sample was frozen at -80°C until samples from all growth periods were collected.

Once samples from all growth periods were collected, cell pellets were resuspended in 1ml of Lysis Buffer 1 (L1) (plus NaBu and Complete) and incubated on ice for 5 minutes; after which they were spun down to pellet nuclei at 800 x g, for 5 minutes at 4°C. Nuclei were resuspended in Lysis Buffer 2 (L2) (plus NaBu and Complete) and divided into equal volumes of 300-600 μ l per tube. Each tube was then sonicated at 80% amplitude for 20 seconds ON, 40 seconds OFF, for a total of 25 min, using the qSonica sonicator. The sonicated samples were centrifuged for 10 minutes at 14 000 x g, at 4°C to remove cell debris. The supernatant was transferred to a 1.5 ml tube and chromatin concentration was measured; Sonication quality was tested by decrosslinking 15 μ g of sonicated chromatin into a final volume of 500 μ l DB buffer with 20 μ l NaCl 5M (0.2M) at 65°C overnight, with shaking; then, the sample was purified using the QIAquick PCR Purification kit. 10 μ l of sample was run on a 1.5% agarose gel to confirm sonicated DNA sizes between 100 and 600 base pairs.

To Pre-block protein A/G – sepharose beads, 500 ul of Protein A – sepharose beads was mixed with 500ul of Protein G sepharose in a 2 ml centrifuge tube then centrifuged for 2 minutes at 4,000 rpm; the resultsng supernatant was discarded. The A/G bead mix was washed twice with 1 ml of Tris-EDTA (TE), and then resuspended in the same buffer. For blocking, 100 μ l of Bovine Serum Albumin (BSA) (10mg/ml) and 200 μ l of denatured tRNA (95°C for 5 minutes) (sigma) were added to the previously described 1 ml A/G bead mixture; the resulting mixture was incubated for 2 hours, with rotation, at 4°C. Blocked beads were washed with 1ml of TE, as described previously, and resuspended in 500 μ l of TE. 60 μ g of chromatin was diluted in dilution buffer (DB) to a final volume of 500 μ l in low-DNA-binding tubes. These tubes were then pre-cleared with 30 μ l of blocked A/G beads for 1 hour at 4°C with overhead rotation. Afterwards samples were centriufuged for 1 minute at 4000 rpm; the supernatant was transferred to a new 1.5ml tube.

For immunoprecipitation (IP), a pre-determined amount of antibody was added to the 60 μ g of pre-cleared chromatin and incubated overnight at 4°C with overhead rotation. The next day, 50 μ l pre-blocked Protein A/G - Sepharose beads were added and this was incubated for two hours at 4°C with overhead shaking. The immunoprecipitaed chromatin was centrifuged for 1 minute at 5000rpm, at RT; the resulting supernatant was discarded; the beads were resuspended in 1 ml of washing buffer and incubated at RT with overhead shaking. Finally, the beads were centrifuged for 5 minutes at 5000rpm at RT and the supernatant was discarded. The wash was repeated twice more; once with wash buffer and one last time with final wash buffer. The immunoprecipitated chromatin was eluted from the beads in 150 μ l elution buffer (EB), with a 10 minute incubation at RT with overhead shaking. Then, tubes were centrifuged for 5 minutes at 5000 rpm at RT and the resulting supernatant was transferred to a new tube; this process was repeated for a final elution volume of 300 μ l. Next, samples were de-crosslinked with 12 μ l of 5M NaCl and 1.56 μ l of 10mg/ml RNase A 50 μ g/ml and incubated at 37°C for 30 minutes then 1 μ l of 20mg/ml Proteinase K was added and incubated for at least 5 hours at 65°C with shaking. Sample were then purified using the QIAquick PCR Purification kit by diluteing in

5 volumes ($1500\mu\text{l}$) of binding buffer and ultimately eluted in $50\mu\text{l}$ of Qiagen elution buffer. Samples are stored at -20°C or used for ChIP-seq library preparation.

Prepared ChIP libraries were submitted to Helmholtz Next Generation sequencing core facility for 50 bp paired-end sequencing on the Illumina 4500 platform. Sequencing reads were mapped to reference genome hg38 using Bowtie2 v2.1.0.

Signal normalization

The length of the previously merged ATAC-seq peaks was extended such that each peak had a minimum length of 500 bp; ATAC-seq peaks whose length was already greater than 500 bp was left the same. Also, gene promoters were taken into account, by considering $500 \pm$ the TSS of each gene's most expressed transcript. As with ATAC-seq, ChIP-seq reads overlapping the extended ATAC-seq peaks and gene promoters was counted such that only reads that overlapped any portion of exactly one peak, and had a mapping quality above or equal to 20 were taken into consideration. Reads that overlap multiple peaks were discarded. Then, the number of reads was normalized with the 'upperquartile' method, keeping original library sizes.

Correlation

The number of FP-U normalized CPMs found within extended ATAC-seq peaks and gene promoters was determined for ATAC-seq and H3K27ac ChIP-seq. Correlation between CPM values was done with spearman correlation. Response to glucose starvation was determined by subtracting the average log CPM values collected in fed samples (F1, F2, and F3) from the average log CPM values from starved samples (S1, S2, and S3). The same was done for RNA-seq samples. Correlation of these logFC values was also done with the spearman correlation.

6.6 Activity by Contact

Replicating Activity Estimates

Data for recreating activity estimates for enhancers on chromosome 22 of the K562 cell-line was taken from the Activity by Contact repository (<https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction>). DNase-hypersensitivity assays and H3K27ac ChIP-seq replicate samples were normalized with the FP-U procedure. The normalized CPM values at DHS peaks were used to estimate activity according to equation 1. Spearman correlation was used to evaluate the similarity between activity estimated with FP-U normalized values and the activity reported by the authors. A p-value < 0.05 was considered as significant.

The effect of activity estimation was analyzed using PCA, which was carried out using the prcomp function from the R programming language.

Inferring gene-enhancer pairs

In-house developed scripts were developed in order to normalize ATAC-seq and H3K27ac samples from the different growth periods. The number of adjusted CPMs found within

ATAC-seq peaks and gene promoters was also calculated with in-house scripts. The activity (according to equation 1, was estimated for all ATAC-seq peaks and gene promoters using the FP-U normalized CPMs. Finally, estimation of the number of contacts between promoters and enhancers, as well as the esimation of the ABC-score, was done using a modified version of ‘predict.py’ and ‘predictor.py’ scripts provided by the ABC developers (<https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction/tree/master/src>). These scripts were modified in order to convert coordinates from hg38 assembly to hg19 assembly, using the pyliftover (0.4) python package. This was done since the averaged Hi-C dataset provided by the authors was happed to the hg19 assembly. ATAC-seq peaks found within 2000 bp of a transcripts TSS were not considered as candidate enhancers for any gene.

Contact percentage was estimated by considering the total number of powerlaw-scale adjusted contacts between a candidate enhancer and all of the genes for which it is considered a candidate regulator.

Positive gene-enhancer interactions were considered as those with an ABC-score equal to, or greater than, 0.02 and the contact percentage is equal to, or greater than, 8.5%. Changes in activity were presented by subtracting the average log(activity) value during fed growth periods (F1, F2, and F3) from the log(activity) values from all samples.

6.7 Term Enrichment Analysis

During differential expression analysis and DA-Up regulated cluster analysis, we performed gene term enrichment analysis with the Metascape tool (Zhou et al., 2019) using the online portal (<https://metascape.org/gp/index.html#/main/step1>). Ensembl gene ids were used to carry out an ‘Express Analysis’. The figures presented in this work only present the top 20 most enriched terms for each gene group.

6.8 Software

R 3.6.2 Python 3.7.6

References

- Andersson, R., & Sandelin, A. (2019). Determinants of enhancer and promoter activities of regulatory elements. *Nature Reviews Genetics*, 1–17.
- Barisic, D., Stadler, M. B., Iurlaro, M., & Schübeler, D. (2019). Mammalian iswi and swi/swf selectively mediate binding of distinct transcription factors. *Nature*, 569(7754), 136–140.
- Bell, O., Tiwari, V. K., Thomä, N. H., & Schübeler, D. (2011). Determinants and dynamics of genome accessibility. *Nature Reviews Genetics*, 12(8), 554–564.
- Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., ... others (2010). The nih roadmap epigenomics mapping consortium. *Nature biotechnology*, 28(10), 1045–1048.
- Bheda, P. (2020). Metabolic transcriptional memory. *Molecular Metabolism*, 38, 100955.
- Bheda, P., Aguilar-Gómez, D., Becker, N. B., Becker, J., Stavrou, E., Kukhlevich, I., ... others (2020). Single-cell tracing dissects regulation of maintenance and inheritance of transcriptional reinduction memory. *Molecular Cell*.

- Bothma, J. P., Garcia, H. G., Ng, S., Perry, M. W., Gregor, T., & Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the drosophila embryo. *Elife*, 4, e07956.
- Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). Atac-seq: a method for assaying chromatin accessibility genome-wide. *Current protocols in molecular biology*, 109(1), 21–29.
- Cao, Q., Anyansi, C., Hu, X., Xu, L., Xiong, L., Tang, W., ... others (2017). Reconstruction of enhancer–target networks in 935 samples of human primary cells, tissues and cell lines. *Nature genetics*, 49(10), 1428.
- Consortium, E. P., et al. (2012). An integrated encyclopedia of dna elements in the human genome. *Nature*, 489(7414), 57–74.
- Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., ... Jaenisch, R. (2010). Histone h3k27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, 107(50), 21931–21936. Retrieved from <https://www.pnas.org/content/107/50/21931> doi: 10.1073/pnas.1016071107
- De Nadal, E., Ammerer, G., & Posas, F. (2011). Controlling gene expression in response to stress. *Nature Reviews Genetics*, 12(12), 833–845.
- Dunipace, L., Ákos, Z., & Stathopoulos, A. (2019). Coacting enhancers can have complementary functions within gene regulatory networks and promote canalization. *PLoS Genetics*, 15(12), e1008525.
- Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., ... others (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345), 43–49.
- Fu, Y., Tang, M., Xiang, X., Liu, K., & Xu, X. (2019). Glucose affects cell viability, migration, angiogenesis and cellular adhesion of human retinal capillary endothelial cells via sparc. *Experimental and therapeutic medicine*, 17(1), 273–283.
- Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., ... others (2019). Activity-by-contact model of enhancer–promoter regulation from thousands of crispr perturbations. *Nature genetics*, 51(12), 1664–1669.
- Gasperini, M., Tome, J. M., & Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nature Reviews Genetics*, 1–19.
- Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature reviews Molecular cell biology*, 19(10), 621–637.
- Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., ... Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*, 503(7475), 290–294.
- Joyner, R. P., Tang, J. H., Helenius, J., Dultz, E., Brune, C., Holt, L. J., ... Weis, K. (2016). A glucose-starvation response regulates the diffusion of macromolecules. *Elife*, 5, e09376.
- Kempfer, R., & Pombo, A. (2019). Methods for mapping 3d chromosome architecture. *Nature Reviews Genetics*, 1–20.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature methods*, 9(4), 357.
- Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., ... de Graaff, E. (2003). A long-range shh enhancer regulates expression in the developing limb and fin and

- is associated with preaxial polydactyly. *Human molecular genetics*, 12(14), 1725–1735.
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, 15(12), 550.
- Lu, C.-A., Lim, E.-K., & Yu, S.-M. (1998). Sugar response sequence in the promoter of a rice α -amylase gene serves as a transcriptional enhancer. *Journal of Biological Chemistry*, 273(17), 10120–10131.
- Lun, A. T., & Smyth, G. K. (2016). csaaw: a bioconductor package for differential binding analysis of chip-seq data using sliding windows. *Nucleic acids research*, 44(5), e45–e45.
- Maiques-Diaz, A., Spencer, G. J., Lynch, J. T., Ciceri, F., Williams, E. L., Amaral, F. M., ... others (2018). Enhancer activation by pharmacologic displacement of lsd1 from gfi1 induces differentiation in acute myeloid leukemia. *Cell reports*, 22(13), 3641–3659.
- Moore, J. E., Pratt, H. E., Purcaro, M. J., & Weng, Z. (2020). A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods. *Genome biology*, 21(1), 17.
- Nicolaï, M., Roncato, M., Canoy, A., Rouquie, D., Sarda, X., Freyssinet, G., & Robaglia, C. (2006). Large-scale analysis of mrna translation states during sucrose starvation in arabidopsis cells identifies cell proliferation and chromatin structure as targets of translational control. *Plant Physiology*, 141(2), 663–673.
- Okoshi, R., Ando, K., Suenaga, Y., Sang, M., Kubo, N., Kizaki, H., ... Ozaki, T. (2009). Transcriptional regulation of tumor suppressor p53 by camp-responsive element-binding protein/amp-activated protein kinase complex in response to glucose deprivation. *Genes to Cells*, 14(12), 1429–1440.
- Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., ... Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. *Cell*, 152(1-2), 157–171.
- Park, H.-R., Tomida, A., Sato, S., Tsukumo, Y., Yun, J., Yamori, T., ... Shin-Ya, K. (2004). Effect on tumor cells of blocking survival response to glucose deprivation. *Journal of the National Cancer Institute*, 96(17), 1300–1310.
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods*, 14(4), 417–419.
- Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. (2013). Enhancers: five essential questions. *Nature Reviews Genetics*, 14(4), 288–295.
- Ray, J., Munn, P. R., Vihervaara, A., Lewis, J. J., Ozer, A., Danko, C. G., & Lis, J. T. (2019). Chromatin conformation remains stable upon extensive transcriptional changes driven by heat shock. *Proceedings of the National Academy of Sciences*, 116(39), 19431–19439.
- Reske, J. J., Wilson, M. R., & Chandler, R. L. (2020, April). ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. *Epigenetics & Chromatin*, 13(1). Retrieved from <https://doi.org/10.1186/s13072-020-00342-y> doi: 10.1186/s13072-020-00342-y
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140.
- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of rna-seq data. *Genome biology*, 11(3), 1–9.
- Rui, L. (2011). Energy metabolism in the liver. *Comprehensive physiology*, 4(1), 177–197.

- Senmatsu, S., Asada, R., Abe, T., Hoffman, C. S., Ohta, K., & Hirota, K. (2019). lncrna transcriptional initiation induces chromatin remodeling within a limited range in the fission yeast *fbp1* promoter. *Scientific reports*, 9(1), 1–8.
- Shlyueva, D., Stampfel, G., & Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews Genetics*, 15(4), 272–286.
- Simeonov, D. R., Gowen, B. G., Boontanrart, M., Roth, T. L., Gagnon, J. D., Mumbach, M. R., ... others (2017). Discovery of stimulation-responsive immune enhancers with crispr activation. *Nature*, 549(7670), 111–115.
- Soares, L. M., He, P. C., Chun, Y., Suh, H., Kim, T., & Buratowski, S. (2017). Determinants of histone h3k4 methylation patterns. *Molecular cell*, 68(4), 773–785.
- Stark, R., Grzelak, M., & Hadfield, J. (2019). Rna sequencing: the teenage years. *Nature Reviews Genetics*, 20(11), 631–656.
- Whalen, S., Truty, R. M., & Pollard, K. S. (2016). Enhancer–promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nature genetics*, 48(5), 488–496.
- Williams, M. S., Amaral, F. M., Simeoni, F., Somervaille, T. C., et al. (2020). A stress-responsive enhancer induces dynamic drug resistance in acute myeloid leukemia. *The Journal of Clinical Investigation*, 130(3).
- Zhang, T., Zhang, Z., Dong, Q., Xiong, J., & Zhu, B. (2020). Histone h3k27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. *Genome biology*, 21(1), 1–7.
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., ... others (2008). Model-based analysis of chip-seq (macs). *Genome biology*, 9(9), 1–9.
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., ... Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature communications*, 10(1), 1–10.