

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by

Qi Wang, M.Sc.
Born in Liaoning, China
Oral-examination: 8 Oct. 2018

Integrative methods for epigenetic profiling in cancer and development

Referees: Prof. Dr. Roland Eils
Prof. Dr. Benedikt Brors

Chapter 3

Integrative analysis of differential epigenetic alterations

3.1 Introduction

Integrating multi-omics data is challenging due to high variability and noises across different data types, yet it is essential in cancer research since cancer usually harbors all type of alterations, either in DNA sequences or epigenetic modifications. Although tools such as ChromHMM have proven their usefulness in studying the combinatorial patterns of multiple epigenetic marks, it is limited to binary measurement such as presence or absence of the peaks. In the case where enrichment peaks are present in all biological conditions, the intensity of epigenetic modification level may dramatically affect the binding probabilities. Therefore a quantitative comparison between epigenetic data is necessary for understanding the impact of a binding event. However, there are very limited ways in doing this. Although meta-analysis [298] and time course analysis [299] approaches are common in RNA-Seq studies, they are generally not applicable in studying multiple epigenetic marks. In this chapter, I present my approach to integrating multiple epigenetic datasets, which leads to more insightful results in understanding the underlying biological relationship of epigenetic alterations.

In this approach, I will deliberately treat gene expression independently of epigenetic modifications. One of the justifications for this choice comes from the concept of epigenetic priming, an event in which epigenetic modifications initiate before gene expression. Hence, epigenetic alterations and gene expression appear to be decoupled from this point of view. The notion

of epigenetic priming is important in cancer studies because it might allow one to identify potential oncogenic processes through epigenetic alterations before they become detectable from gene expression or at the protein level. To investigate the concept of epigenetic priming, I considered datasets containing different time points during developmental progression. For example, I studied four stages in neural progenitor cells (NPC) development into neuroepithelial (NE, day 12), early radial glial (ERG, day 12), mid radial glial (MRG, day 35) and late radial glial (LRG, day 80), in which Ziller *et al.* observed that gain of H3K4me1 and loss of DNA methylation appeared in the early stages of the differentiation from ESC to NPC [300]. By comparing the epigenetic patterns in promoter regions of NE stage to gene expression of all five stages, the result shows that the correlations of epigenetic marks with expression at later stages, as opposed to the NE stage, reaches the highest level (Fig. 3.1), indicating the expression level is more related to the epigenetic modifications in earlier stages. Comparing to histone modifications, DNA methylation have longer-term effect, which confers later expression in development [301]. Epigenetic priming is the reason why my approach focuses mostly on epigenetic alterations rather than gene expression. The delay in gene expression in this case is understandable since time is needed for mRNA accumulation after the epigenetic regulations take place.

To address such concepts in differential epigenetic analyses, I developed a new method termed "cancer regulatory landscapes" (*crl*) which integrates the quantitative information from multiple epigenetic marks, on genome-wide non-coding regulatory elements, allowing one to discover significantly epigenetically altered genomic regions and pathways. In the benchmarking, I proved that the genes of interests found using this method are highly relevant to cancer and developmental test cases.

3.2 Data sources

3.2.1 Epigenetic datasets

As illustrated in Fig. 3.1, both histone modifications and DNA methylation can exhibit epigenetic priming. Besides epigenetic modifications in promoters, I also included genome-wide epigenetic alterations in non-coding regulatory elements, which covers both alternative promoters and enhancers. With WGBS data I am able to evaluate the effects of lowly methylated regions (LMRs) inside enhancers, which have been shown to contribute to its activity [302]. Genome-wide epigenome cohorts are publicly available from many consortia, such as NIH Roadmap Epigenomics [303], ENCODE

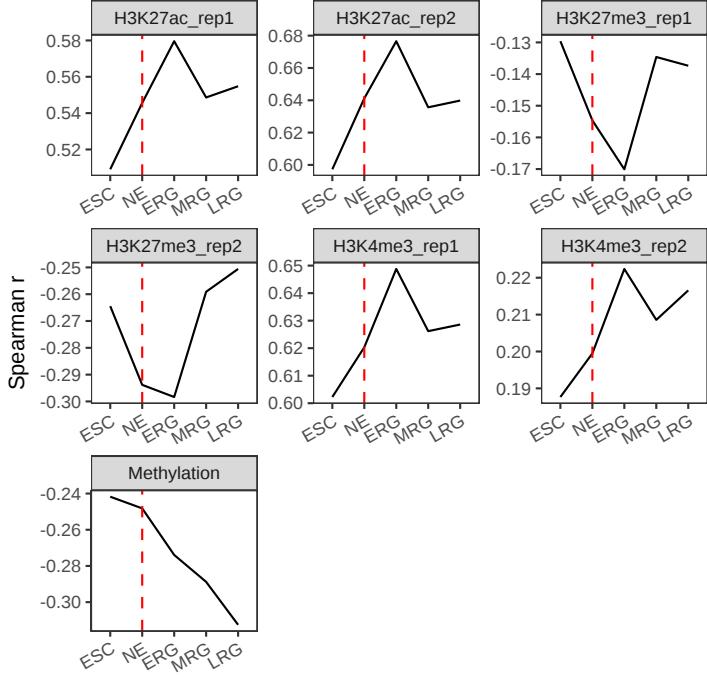


Figure 3.1: Epigenetic priming in neural progenitor development stages. Each histone mark has two replicates, and Spearman correlation between epigenetic patterns of the NE stage and gene expression of all stages are shown. Histone modifications present short-term effects, while DNA methylation present long-term effects.

[187], Blueprint [304], and the International Human Epigenome Consortium (IHEC) [305]. These resources allow me to investigate the epigenetic relationships between embryonic stem cells and differentiated cells, or between tumor and normal tissues (Table 3.1).

Due to the public policies of some data providers, their data are generally provided for visualization purpose. Mostly only Wig and BigWig format files [306] are accessible, rather than raw sequences (Fastq) or alignment files (BAM/BED). Therefore statistical methods specific to raw counts [307, 308] are not applicable in my study. The peak calls were done by these data providers, and these peak regions are used in the validation of cell-type specific enhancers in later sections.

For restricting the epigenetic comparisons to the genomic loci of interests (promoters and enhancers), I downloaded genomic coordinates of promoters from the eukaryotic promoter database (EPD) [309], and enhancers from the GeneHancer database [310]. There are $\sim 285,000$ enhancers in GeneHancer

Table 3.1: Test cases for phenotypic studies.

Tests	Controls		Num.*	Data**	Accession codes/src
Neural Progenitor Cells (NPC)	Embryonic stem cells		8	61	GSE16256
Neuroepithelial (NE)	Embryonic stem cells		5	10	GSE62193
Early radial glial (ERG)	Embryonic stem cells		5	10	GSE62193
Mid radial glial (MRG)	Embryonic stem cells		5	10	GSE62193
Mesenchymal stem cells (MSC)	Embryonic stem cells		8	51	GSE16256
Trophoblast stem cells (TSC)	Embryonic stem cells		8	64	GSE16256
Chronic lymphocytic leukemia (CLL)	B cells from healthy cases		4	121	CEEHRC
Lower grade glioma (LGG)	Hippocampus middle, Inferior temporal lobe, Mid frontal lobe		7	112	CEEHRC, GSE17312
Colorectal cancer (CRC)	Sigmoid colon from healthy cases		7	154	CEEHRC
Papillary thyroid cancer (PTC)	Thyroid from healthy cases		7	54	CEEHRC

* Number of epigenetic marks.

** Number of epigenetic datasets.

database, incorporated from four different sources: the Encyclopedia of DNA Elements (ENCODE) [311], the Ensembl regulatory build [312], the VISTA Enhancer Browser [313], and the functional annotation of the mammalian genome (FANTOM) project [234]. This database contains enhancers for a large number of cell-lines and tissues. Both datasets were converted from GRCh38 to GRCh37 using the LiftOver tool [314].

The promoter coordinates were extended to ± 1000 base pairs around the original coordinates. A BigWig file consists of a number of blocks, each containing a declaration of a fixed or variable genomic region. The numerical signals from BigWigs for region $i \in [m, n]$ were calculated as $S_i = \sum_m^n s_i$, where the promoter/enhancer ranges from m to n , and s_i is the signal for each genomic window in the region. The enhancers from GeneHancer database are

collected from all tissues. To avoid unspecific enhancers, I require enhancers to overlap with H3K4me1 peaks of at least two samples in the tissue I am studying. Fig. 3.2 shows the number of common and specific enhancers among four cancer types.

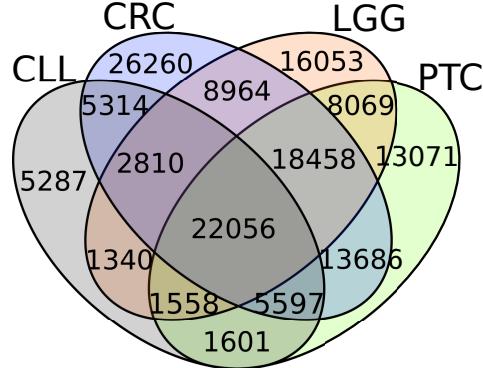


Figure 3.2: Number of common and specific cancer enhancers among four cancer test cases.

3.2.2 Data processing procedure

The epigenetic data between samples have large heterogeneity and must be normalized before statistical comparisons can be made. The promoters and enhancers are of different genomic length, so I divided the intensities by their length before the normalization. The data are heavily right skewed, and I used Box-Cox transformation (3.1) to transform the data into normal distributions.

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log y, & \text{if } \lambda = 0 \end{cases} \quad (3.1)$$

By finding the most likely λ that minimizes the variation, a universal λ of ~ 0.182 was proposed to apply to all data. After power-transformation, I checked the global variability across and within biological groups using *quantro* [315]. It estimated the variabilities were caused by technical variation (e.g. batch effects), and a global normalization is applicable.

I used quantile normalization which was firstly implemented in microarray analysis by Bolstad *et al.* [316], and quickly adapted to a variety of data types such as RNA-Seq analysis [254, 317–319], DNA methylation [320], ChIP-Seq [321, 322]. This method makes the quantiles of each distribution equal. After normalization, except a number of genomic regions showing no signal

of epigenetic modifications, the rest of the data follows normal distributions (Fig. 3.3), which is the prerequisite for using differential principal component analysis in the next step.

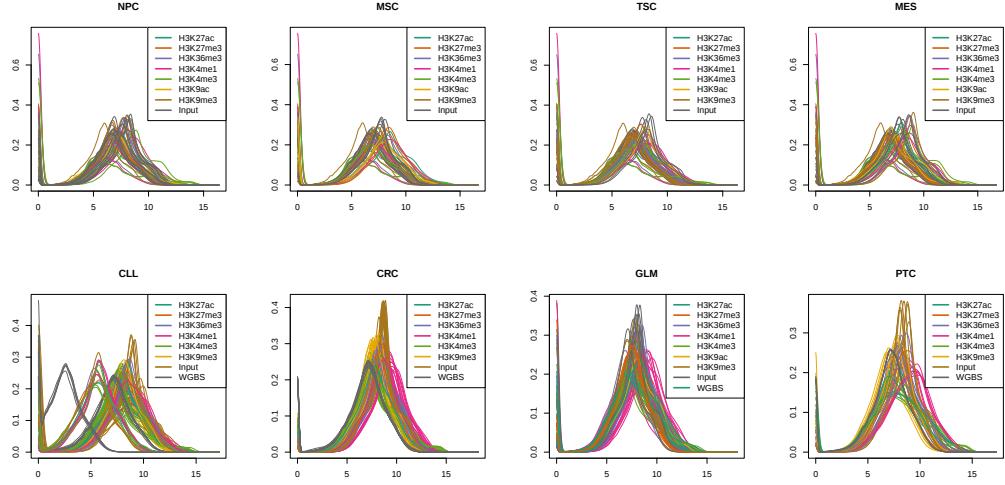


Figure 3.3: Data distributions of epigenetic marks after Box-Cox transformation.

Correlation structures of data

I first compared the aforementioned processed data with gene expression to verify that the normalization maintains the correlation structure. The result shows that for each sample, the gene expression positively correlates with the epigenetic signal in the promoter regions of H3K4me3 and H3K27ac, and negatively correlates with the epigenetic signal of H3K9me3 and H3K27me3 (Fig. 3.4. a).

Although some epigenetic marks correlate very well with the gene expression of the corresponding dataset, I wanted to perform a differential analysis between conditions. Therefore, I tried to resolve the relationships of expression differences and the differences of epigenetic modifications between two biological conditions, e.g. tumor cells and normal cells. Again, I normalized the data using quantile normalization, making the two groups have same standard deviations. Afterwards, I took the average differences of epigenetic modification levels and gene expression levels. The correlations of differential histone modification with expression in the promoter regions are considerably lower comparing to the previous test, but still correlate positively with activation marks and negatively with repression marks (Fig. 3.4. b). The

weaker correlations may be attributed to the fact that I have not included the epigenetic modification outside of the promoter.

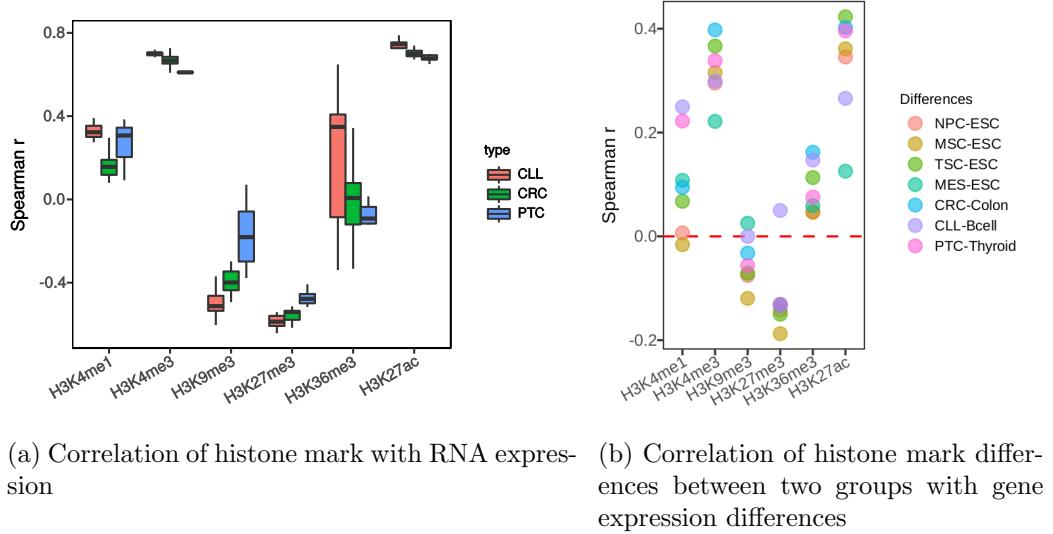


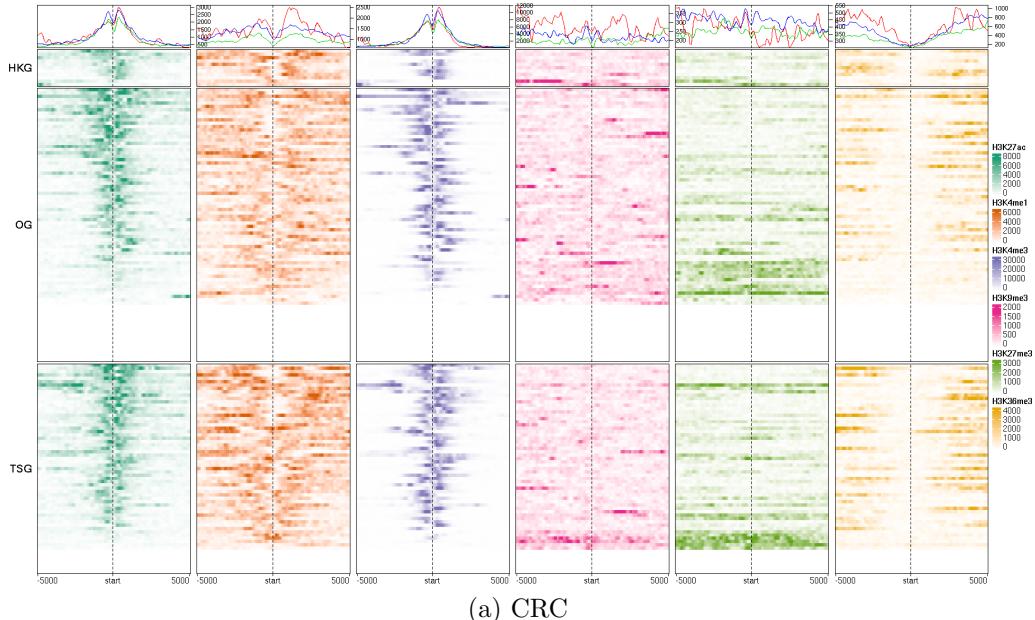
Figure 3.4: Correlation of epigenetic marks with gene expression.

Modifications at oncogenes and tumor suppressor genes

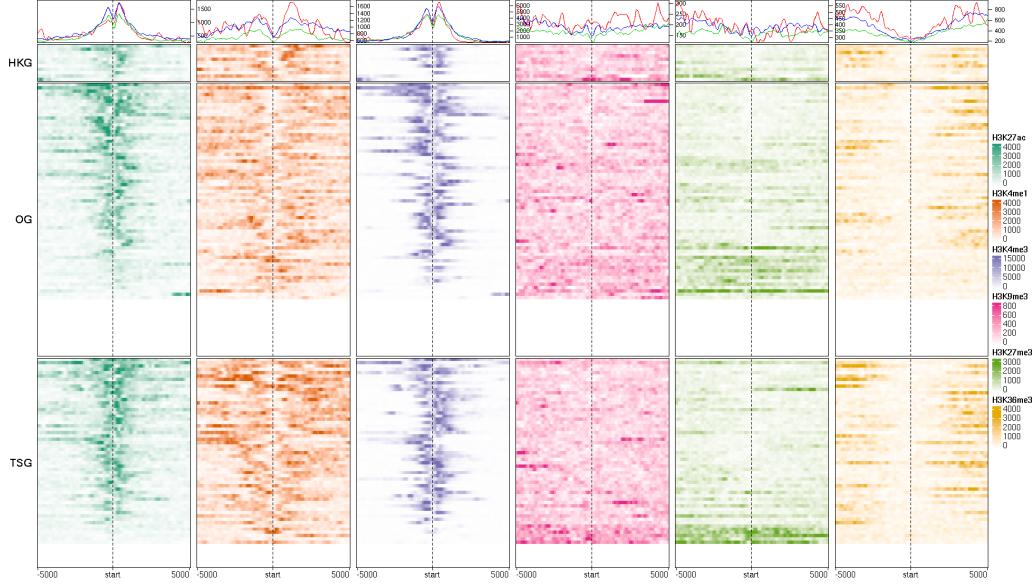
With respect to the role in cancer development, it has been known that hyperacetylation of oncogenes (OG) results in an increase of their gene expression, whereas hypoacetylation of tumor suppressor genes (TSG) reduce their expression levels [323]. OG and TSG are both retrieved from a compiled list by Walker *et al.* [324] (Supplementary table S1). Here I used 11 housekeeping genes (HKG) with constant expression level from RNA-Seq profiles as control [325]. The epigenetic modification levels around the TSS of HKG, OG and TSG are not significantly altered compared to each other (Fig. 3.5 and Fig. S1), suggesting that differential epigenetic modifications mainly occur at distal regulatory regions.

3.2.3 Multivariate data analysis

In order to represent the overall differences from multiple epigenetic modifications, a single measure is needed to represent the variances between different datasets. After subtracting each epigenetic mark with the control, the Pearson's correlations between the average values of each epigenetic mark at promoters and enhancers across the samples indicate that there are strong positive correlations between the repressive marks (H3K27me and H3K9me3) and



(a) CRC



(b) Colon

Figure 3.5: Histone mark signals around OG/TSG/HKG in colorectal cancer and normal sigmoid colon. Other test cases are in Supplementary Fig. S1

active marks (H3K4me1, H3K4me3, and H3K27ac), as well as negative correlation between active marks and repressive marks (Fig. 3.6). Co-occurrence of epigenetic marks is common for multiple activation marks [33]. In addition, DNA methylation and repressive marks are also negatively correlated,

as they are often replaced with each other during gene silencing [326, 327].

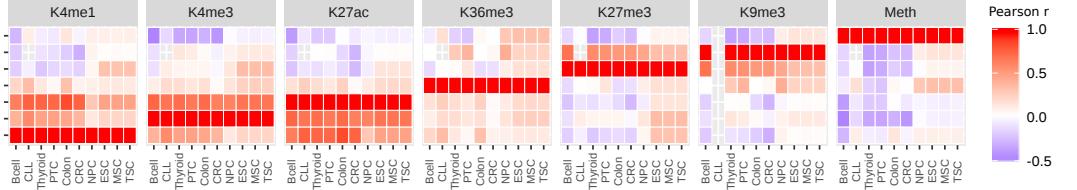


Figure 3.6: Epigenetic marks in promoter and enhancer regions show strong correlations with each other.

To statistically test the differences between two biological conditions with each epigenetic mark, I used the epigenetic datasets of embryonic stem cells (ESC) and compared with their differentiated forms (NPC, MSC, TSC, MES). The datasets are available in BAM formats. The p-values of the differential epigenetic signals in both promoter and enhancer regions were computed with ChIPComp for ChIP-Seq and BiSeq for WGBS. The combined p-values using Fisher's method (3.2) are very close to the smallest p-value in the every comparison (Fig. 3.7). In this situation, combining p-values from multiple hypothesis testing is not applicable as it may lead to severe inflation of false positive rates when applied to highly correlated datasets [328]. Indeed, Fisher's method makes the assumption of independence of the tests, which is not fulfilled here.

$$X_{2k}^2 \sim -2 \sum_{i=1}^k \ln(p_i) \quad (3.2)$$

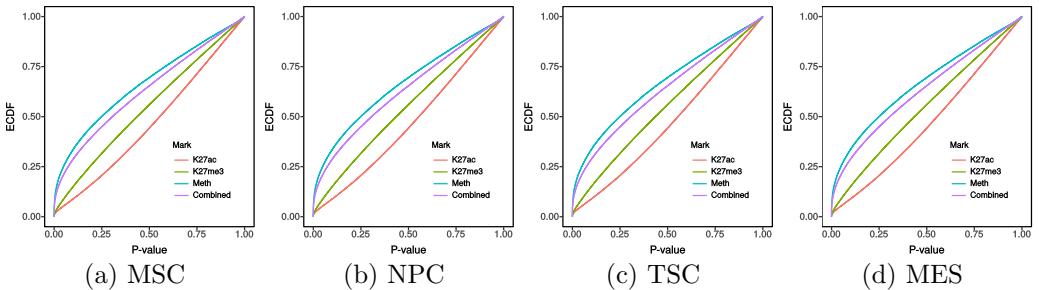


Figure 3.7: Combined p-values from histone modifications and DNA methylation.

Alternatively, I used a method based on the principal component analysis (PCA) applied to differential signals to represent the overall epigenomic

differences. Differential principal component analysis (dPCA) is one of the methods built on singular value decomposition (SVD) that compares differential epigenetic signals across multiple histone marks and replicates between two biological groups [329]. It takes the arithmetic means of each epigenetic datasets across replicates, and summarizes the observed differences between the two groups into a matrix D with genomic loci as rows and datasets as columns. The primary difference between the dPCA and conventional principal component analysis (PCA) is that it analyses the underlying true differences by decomposing D into two matrices: $D = \Delta + E$, where Δ is the underlying true differences which I am interested in, and E is the random sampling noise. E is calculated as $E = \sigma^2\Omega$, where Ω represents the diagonal matrix of eigenvalues, and σ^2 is estimated from a normal distribution over all loci. Since the number of genomic loci is typically much larger than the number of epigenetic datasets, SVD can be used to decompose the matrix $\Delta = B \times V'$, where V' is a transposed diagonal matrix, and $B = (\beta_1, \beta_2, \dots, \beta_n)$ in which β_j characterizes the variation in Δ contributed by pattern v_j .

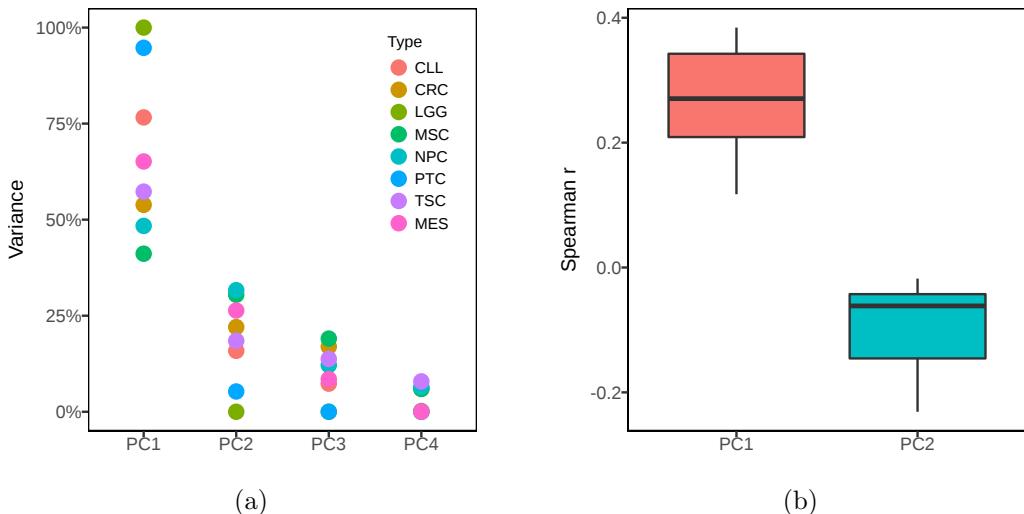


Figure 3.8: Characteristics of dPCs. (a) Variances explained by each dPCs. (b) Correlation of dPCs with gene expression differences in all test cases, in which dPC1 is positively correlated with gene expression, and dPC2 is slightly negatively correlated with gene expression.

In these test cases, like the correlation of histone mark differences with gene expression differences (Fig. 3.4 b), the dPCs still have correlations with the gene expression differences (Fig. 3.8 b). Among the dPCs, dPC1 explained $\sim 40\% - 100\%$ variances (Fig. 3.8 a), and usually these variances

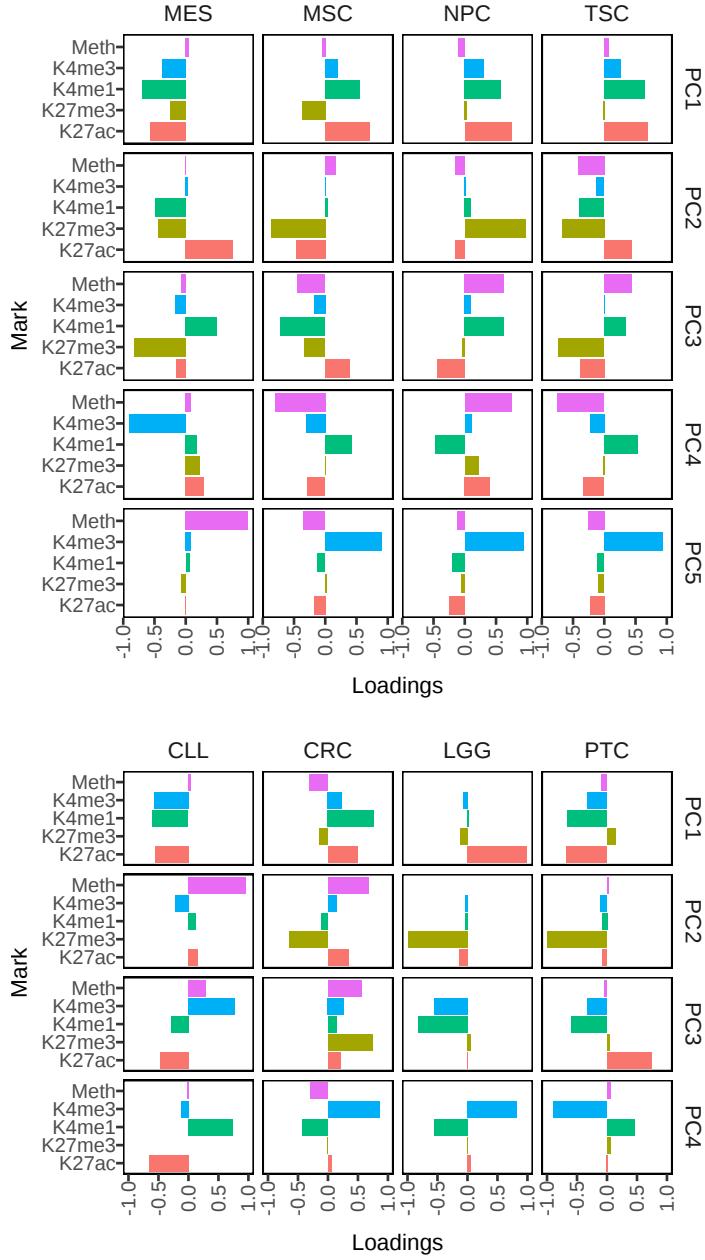


Figure 3.9: Histone mark contributions to each PC.

are mainly contributed by one or two epigenetic marks. Coinciding with the results from Ji *et al.*, dPC1 appears to be mainly driven by active epigenetic marks (Fig. 3.9).

Plotting the computed dPC1 values in both promoter and enhancer re-

gions in the context of the three of the activation marks (H3K27ac, H3K4me1 ,and H3K4me3) against each biological condition, indicates that dPC1 alone is able to represent all three marks (Fig. S4 a, b). Other dPCs, taking dPC2 for example, are not representative for the above three marks (Fig. S4 c, d).

3.3 Network representation of promoter-enhancer relationships

3.3.1 Enhancer-promoter interactions

Both *in vivo* [330] and *in vitro* [50], one enhancer can regulate multiple promoters, and one promoter can be under the control of multiple enhancers, too. Therefore, enhancer-promoter interactions can be presented as a bipartite graph, in which both enhancers and promoters are represented as vertices, and directed edges link enhancers to their target promoters. In this oriented graph, the vertices are weighted according to the magnitude of epigenetic alterations at enhancers and promoters (as measured by the dPCs), and the edges are weighted according to the probability of such promoter-enhancer interactions, as will be described in the next section.

3.3.2 Estimating interaction frequencies between enhancers and promoters

Enhancers have been found to interact with promoters on the same chromosome (*cis*) or different chromosomes (*trans*). It is estimated that most often, the enhancers are within a limited distance from the target promoter of the same chromosome, a fact which is supported either by polymer physics [331, 332] or looking at the interaction densities from both experimental data (CD34 and GM12878 [333])(17 blood cells [334]) and *in silico* predictions [235, 242]. Taking the promoter-enhancer (P-E) interactions from these observations and predictions, most of the interactions occur within ± 1 Mb of the TSS (Fig. 3.10).

Chromatin interaction data in cancer are generally not available to us. Moreover, besides the dysregulation of histone modifications and DNA methylation, the chromosome loops in cancer cells are believed to be altered [335– 337] comparing to their normal counterparts. I want to develop a method which can be applied to many different biological contexts, for which, in general, no experimental interaction data is available. Therefore, I chose to implement a universal model for interaction probabilities, fitted on a large set of experimental datasets. The interaction probabilities can be modeled

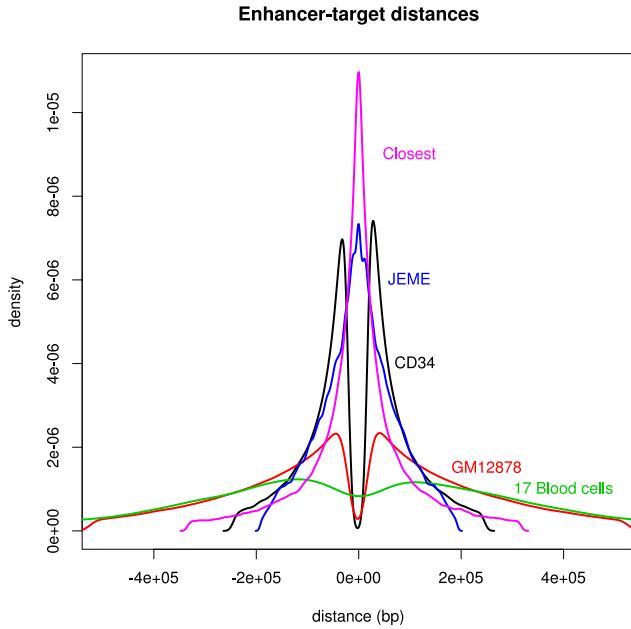


Figure 3.10: Interaction density from capture Hi-C and imputation data. Most of the inferred chromatin interactions occur within $\pm 1\text{Mb}$ range from the target.

using an exponential function based on P-E distances [235]. For each genomic region window $i \in (m, m + 200\text{bp}]$, $j \in (n, n + 200\text{bp}]$ for a given promoter position m and enhancer position n , I counted the number of interactions f_{ij} for each interaction (i, j). Afterwards, the probability y was fitted to an exponential function (3.3).

$$y \sim \exp(f_{ij}, d_{ij}) \quad (3.3)$$

Although physical interactions are generally organized along the whole chromosome and even trans-chromosomes, functional interactions are more likely to be limited within topologically associating domains (TADs) [236]. Therefore, I also restricted the promoter and its interacting enhancers to stay within the same topologically associating domains (TAD). Although TADs are generally believed to be tissue-specific [338], a test using TADs from five human cell lines (embryonic stem cell, mesendoderm cell, mesenchymal stem cell, neural progenitor cell, trophoblast-like cell) provided by Schmitt *et al.* [338], has shown that 57% of the promoters have the same contacting profile in at least 80% of the cell lines.

The likelihoods of enhancer-promoter interaction can be mapped to enhancer-

promoter distances with a power-law decay function [339]. I estimated consistency of contact frequency profiles from several publicly available capture Hi-C datasets, including GM12878 (E-MTAB-2323 [333]), 17 blood cells (EGAS00001001911 [334]), breast cancer (PRJEB23968 [340]), stem cells (GSE84660 [341]), and colorectal cancer (EGAS00001001085 [342]). During processing the capture Hi-C data, I required ≥ 10 reads mapped to the other end of the fragment to infer a reliable interaction.

Discrete binning is used to estimate the parameters in the distance-decay function. Afterwards, an interpolation method, as implemented by Lajoie *et al.* [339] is used. By fitting the average number of interactions falling into each bin against the distances to the promoter to an exponential function, I obtained exponent values ranged from -8.17 to -1.74 (Fig. 3.11). As exponent coefficients ranging from -1 to -30 display similar performances in later benchmarking (Fig. 3.14 b), I chose -20 in the following tests.

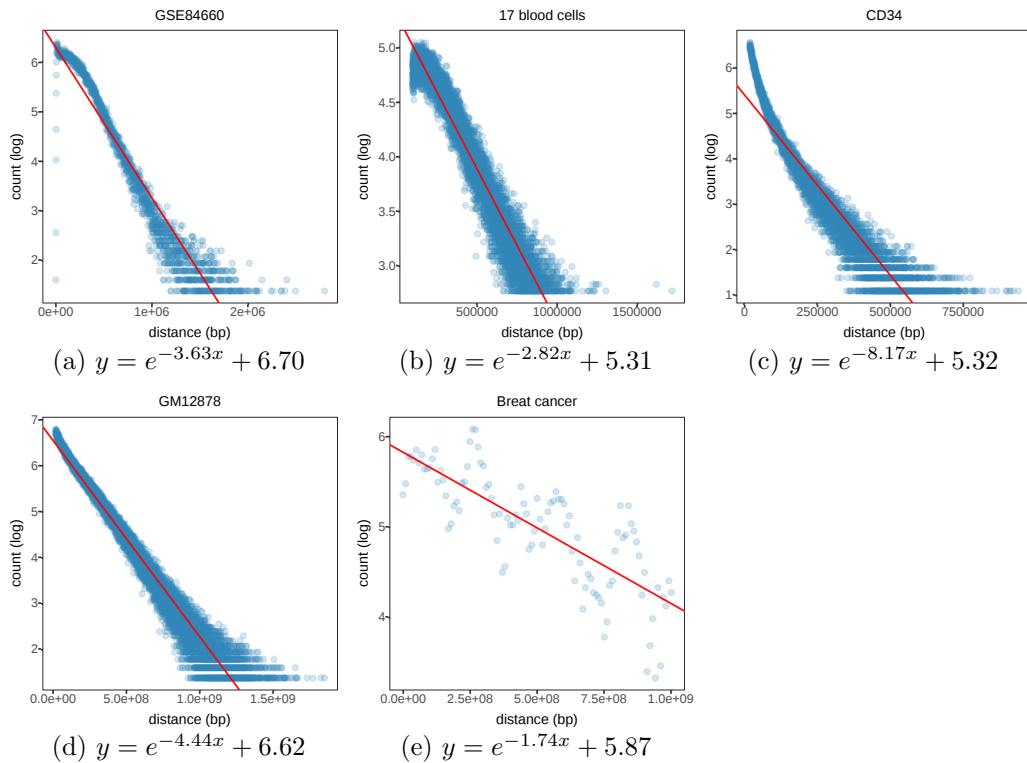


Figure 3.11: Probability density functions of several capture Hi-C contact frequencies. y is the probability of the contacts, while x represents the distance (in Mb) of the interacting region to the promoter.

3.3.3 Ranking genes using personalized PageRank

I defined a meta-gene as the union of all promoters of the gene and its targeting enhancers. I adopted PageRank to summarize the weights of promoters and connected enhancers into a unique meta-gene score. PageRank is originally designed in the valuation the importance of web pages [343]. It was also adopted in bioinformatics in ranking the impact of nodes in metabolic network [344] or gene ontology network [345]. PageRank yields an importance score computed through a random walk process, in which a walker starts from a random vertex, and walks to another connected vertex randomly. This process can be repeated many times. In the end, a rank is used to represent the frequency of visit of each vertex, which is calculated as (3.4),

$$PR(u) = \sum_{v \in B(u)} \frac{PR(v)}{N_v} \quad (3.4)$$

where u represents a vertex, and $B(u)$ are the incoming vertices linked to u , $PR(u)$ and $PR(v)$ represent ranks of vertices u and v , respectively. N_v denotes the number of outgoing links from node v .

Depending on the network structure, in a scenario where there are only incoming links to a vertex, but no outgoing links from that vertex, the walker will stop at the vertex and the process terminates. To solve this problem, a reset parameter α is added to allow the walker to restart at any other random vertex, therefore the final rank becomes (3.5).

$$PR(u) = (1 - \alpha) + \alpha \sum_{v \in B(u)} \frac{PR(v)}{N_v} \quad (3.5)$$

In practice, α is usually set to 0.85, which means the walker has 85% probability to follow an outgoing link from current vertex, and 15% probability to hop to a random vertex.

In another scenario specific to my application, the walker has a preference for some vertices or links over the other ones, and therefore weights of vertices and edges are introduced. In a "personalized" PageRank, the rank is also dependent on the weights of the incoming and outgoing links (3.6),

$$PR(u) = (1 - \alpha) + \alpha \sum_{v \in B(u)} PR(v) W_{(v,u)}^{in} W_{(v,u)}^{out} \quad (3.6)$$

where $W_{(v,u)}^{in}$ and $W_{(v,u)}^{out}$ are calculated based on the number and weight of incoming links and outgoing links of vertex v and u .

I used personalized PageRank implemented in *igraph* [189] to uncover important epigenetic alterations for a gene by taking into account of the

regulatory contribution of enhancers. Upon setting the weights of vertices in the random walk, the promoters or enhancers with more significant epigenetic alterations have higher weights, which is represented by the dPCs previously computed. The dPCs are sorted in decreasing order regardless of their direction of alteration, as several studies suggest that both up and down-regulation of histone acetylation can contribute to gene activation [346], as well as hypomethylation and hypermethylation both affect transcription factor binding, depending on the preferences of transcription factors [176].

I also set edge weights in accordance with the probabilities of chromatin contacts to ensure that only highly confident enhancers are contributing. Since the enhancer-promoter network is a directed graph, all the enhancer scores will eventually be attributed to their interacting promoter. This ensures that, even in a case where the promoter shows little epigenetic alterations, the corresponding meta-gene might have a high score due to the contribution of enhancers. In a next section, I will specifically discuss such cases. In the end, PageRank returns a vector with the rank scores for all meta-genes, in which all the genes are ordered according to the cumulative score of their promoters and associated enhancers.

3.3.4 Benchmarks

In order to validate the outlined procedure, I designed several benchmarking strategies which I will discuss in the next sections.

Benchmarking using rank lists

The PageRank algorithm sorts the genes into descending order which is in accordance with the significance of alterations from both promoter and enhancers (abbreviated as "PromEnh" rank list). First, I found that the PromEnh rank list is more relevant to the biological conditions than the rank list derived from the dPC1 order of only promoters (abbreviated as "PromOnly" rank list). In order to show how much improvements I have using the PromEnh rank list, I compiled a list of 14 ~ 36 marker genes for each biological test case (Supplementary table S2 and S3), either selected from comprehensive literature reviews [347–351], or cancer signature databases, including COSMIC [352], Intogen [353], MalaCards [354], etc.

Afterwards, the receiver-operating characteristics (ROC) of cumulative fraction of markers genes through the ranked list of genes is determined for assessing the sensitivity and specificity of the ranking. The rank list can be generated from other dPCs as well. Taking the area under curve (AUC) in the CLL case as an example, as 70%-80% of the CLL specific marker genes

are enriched in the top 20% genes of the PromEnh rank list, while for the PromOnly rank list this value is below 60% (Fig. 3.12 a). This conclusion is also true for other test cases as well as most of the other dPCs (Fig. 3.12 b).

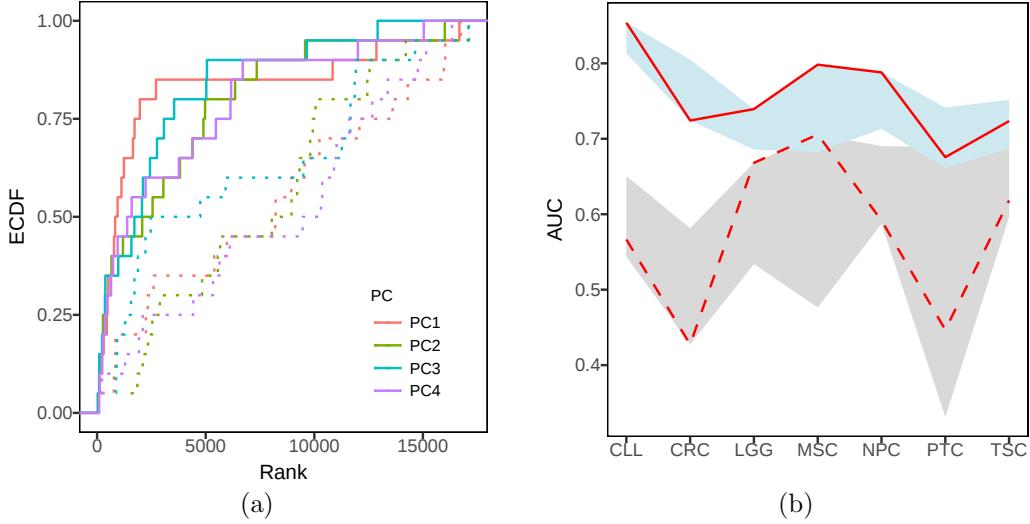


Figure 3.12: ROC of CLL and AUC of all test cases. (a) CLL ROC of PC1-4 using PromEnh rank list (solid lines) and PromOnly rank list (dashed lines). (b) AUC of dPC1 are plotted as red lines (solid: PromEnh, dashed: PromOnly), and AUCs from other dPCs are plotted as shaded bands between the minimum and maximum scores (light blue: PromEnh, grey: PromOnly).

Benchmarking with transformed vertex weights

So far it is still questionable whether dPC scores are directly related to the importance of the vertices. Indeed, it could be that, below a certain threshold, small differences in epigenetic signals (and hence small dPC scores) have no impact on the state of the gene, and should have a zero contribution to the overall score. Hence, I tested a number of transformations of the raw dPC scores using functions frequently used in data transformation in artificial neural networks (Table 3.2) to introduce non-linearity into the vertices weights (Fig. 3.13 a). Benchmarking with dPC1 using different transformation functions ended up with similar AUCs (Fig. 3.13 b). Therefore, I directly used dPCs as the vertex weights to avoid introducing any undesired side effect.

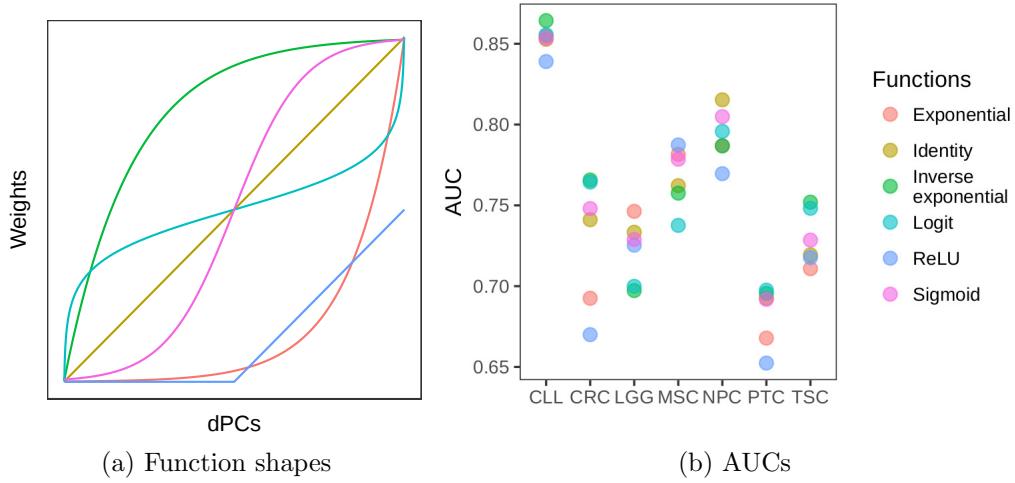


Figure 3.13: Transforming functions and AUCs of each function.

Table 3.2: Weights transforming functions

Function	Equation
Sigmoid	$f(x) = \frac{1}{1+e^{-x}}$
Logit	$f(x) = \ln(\frac{x}{1-x})$
Exponential	$f(x) = e^x$
Inverse exponential	$f(x) = \ln(x)$
Identity	$f(x) = x$
Rectified linear unit (ReLU)	$f(x) = \begin{cases} 0 & \text{if } i < \frac{n}{2} \\ x & \text{if } i \geq \frac{n}{2} \end{cases}$

Benchmarking with biological evidence

The PromEnh rank lists not only show better coincidence with selected marker genes comparing to the PromOnly rank lists, but also show better tissue specificity and are enriched with oncogenes in cancer samples. To show this, tissue specific genes were retrieved from ARCHS4_Tissues [355] which is provided with the EnrichR tool [356]. Adjusted p-values from EnrichR indicate tissue specific enrichment in corresponding test cases. The most significant term from the enrichment of the top 1000 genes in each PromEnh rank list shows a remarkable enrichment of corresponding tissues, while the most significant terms from the top 1000 genes in PromOnly rank list are not relevant to the corresponding tissues (Table 3.3). Given the fact that super-enhancers near cell-type specific genes often accumulate disease associated non-coding variants [58, 357], this phenomenon is not a coincidence as it is

applicable to all the cancer test cases, implying enhancers play a dominant role in cell-type specific regulations.

Table 3.3: Most enriched tissues corresponding to the PromEnh and PromOnly rank list of test cases

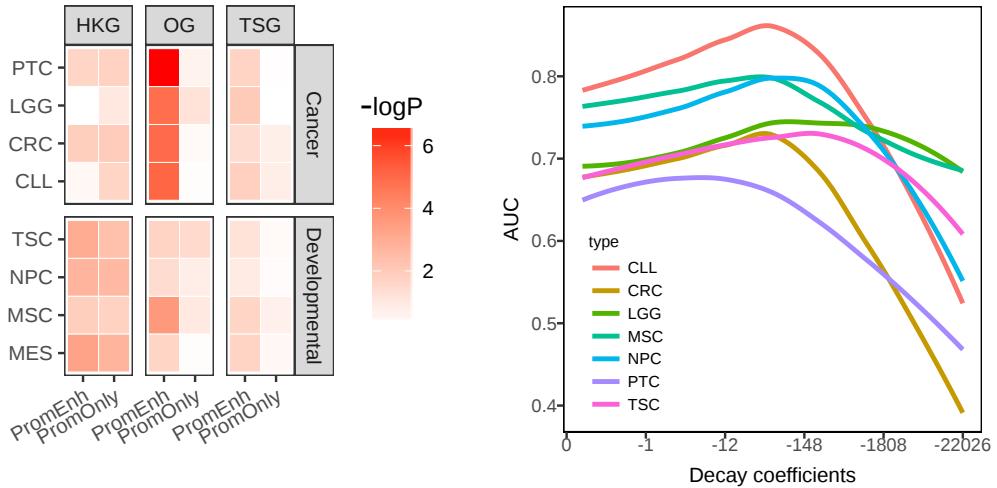
Test case	PromEnh tissue	Adj. P*	PromOnly tissue	Adj. P*
CLL	CD19+ B cells	6.1e-10	Breast (bulk)	2.3e-23
CRC	Small intestine (bulk)	1.5e-16	Spinal cord	2.5e-108
LGG	Prefrontal cortex	3.9e-24	Testis (bulk)	0.07
PTC	Thyroid (bulk)	2.1e-15	Brain (bulk)	7.0e-33
NPC	Spinal cord	3.2e-27	Renal cortex	1.6e-4
MSC	Astrocyte	5.7e-37	Fibroblast	3.9e-24
TSC	Fibroblast	3.9e-24	Lung (bulk)	6.8e-12

* Adjusted p-value (Benjamini-Hochberg method)

In addition, I found that oncogenes are also highly ranked in the PromEnh list. I performed Wilcoxon-Mann-Whitney tests on the positions of oncogenes (OG), tumor suppressor genes (TSG), and housekeeping genes (HKG) in the PromEnh rank list against a uniformly distributed rank list of the same length. Comparing to TSG, HKG, and random gene sets, OGs show a higher ranking which might be interpreted as a tendency to be under strong epigenetic regulation in cancer cells. This phenomenon is not observed in the test cases with normal cells (NPC, MSC, TSC, MES) (Fig. 3.14 a). On the contrary, the test cases related to developmental processes show that HKGs are significantly ranked higher when epigenetic regulation is taken into account. However, all p-values are insignificant when performed using PromOnly rank list, suggesting that the ranks of OGs and HKGs are mainly explained by enhancer contribution. Looking at individual genes, BRAF, KRAS are associated with stronger enhancer regulation. On the other hand, TP53 is often associated with weaker enhancer regulation (Fig. 3.15).

Benchmarking with distance based functions

As the promoter-enhancer contact frequencies discussed above, the distances from the enhancer to the TSS of the promoter determine the likelihood of promoter-enhancer interaction, and the effect can be modeled with power-law decay. I tested a series of exponent values in the decay function from which the edge weights were calculated as (3.7),



(a) Wilcoxon tests of OG, TSG, and HKG ranks comparing to uniformly distributed rank. OGs in cancer are significantly ranked higher in PromEnh rank list, whereas they are not significant in developmental test cases

(b) AUCs of PromEnh list from a range of exponential decay coefficients applied to promoter-enhancer distances. From 0 (no distance decay) to -22026 (edge weight reduced by 99% when P-E distance is over 200bp)

Figure 3.14: Rank list benchmarking.

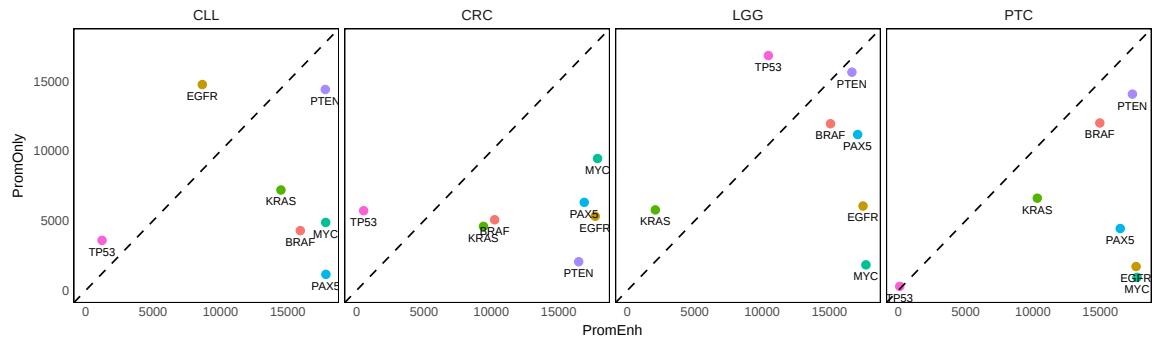


Figure 3.15: Ranking positions of well-known genes in PromEnh and PromOnly list. Genes on the bottom-left corner are ranked worst in both lists, while genes on the top-right corner are ranked best in both lists. Genes on the top-left corner are only ranked best in PromOnly list, and genes on bottom-right corner are only ranked best in PromEnh list.

$$w_i = \exp(m \times d_i + \beta) \quad (3.7)$$

where m is the coefficient I am estimating, and d_i indicates the distance

(in Mb) between the enhancer to the promoter, and β is the intercept from the fitted function. Therefore, setting $m = 0$ implies that all enhancers have the same weights regardless of their distances to the promoter. Also calculated from (3.7), when $m \approx -3912$, the weight of an enhancer with a distance of 1kb to the promoter drops by 98% comparing to an enhancer at the same genomic location of the promoter, which essentially means that the contributions of enhancers in the PageRank are negligible in this scenario, and the AUCs are very close to ranking promoters only. Therefore, a coefficient $m \in (-10, -30)$ seems suitable in maximizing the AUC and matching with the experimental estimation (Fig. 3.14 b), without exaggerating the contribution from enhancers.

Robustness of the gene ranking under random perturbations

To validate that the contributions from enhancers are not an artifact, I used degree-preserving random perturbations, which rewrites the endpoints of the edges with a 50% probability randomly to another vertex in a graph. The randomization can be realized using a rewired promoter-enhancer network in PageRank. During the test with 100 permutations of different rewired network structures, I used the same marker genes as in the benchmarking, and the AUCs with randomly assigned enhancers dropped 10%-20% for most of the test cases (Fig. 3.16). Nevertheless, the AUCs from the rank list including random enhancers still outperformed the rank list using promoter only. I assume that the reason is because that the rewiring did not change the number of enhancers that a gene might have. Therefore, the ranking of cancer and development marker genes will benefit from enhancers no matter which enhancers were linked to them, but the ranks of the other genes are disordered and hence the AUCs are lower.

3.4 Network representation of gene relationships

3.4.1 Network construction

Considering the top-ranked genes could be artifacts, the relatively high ranked genes with known biological functions are more appealing to me, even though they may not be the highest in the rank list. The most common way to perform functional annotation of a rank list of genes is by doing a gene-set enrichment analysis (GSEA). But applying GSEA on high-throughput epigenetic data is still debatable, as severe biases towards genes related to development

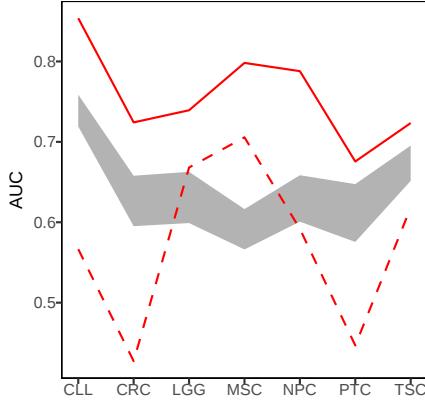


Figure 3.16: AUCs of PromEnh rank list from randomized promoter-enhancer interactions. The grey band regions indicate the quantile ranges from benchmarking each with 100 different rewired promoter-enhancer networks, whereas the red lines show the AUCs with the original promoter-enhancer interactions from PromEnh (solid line) and PromOnly (dashed line) rank lists.

and differentiation have been mistakenly reported in a wide range of DNA methylation studies [358]. Therefore, I performed a network analysis over the rank lists. The rank lists are examined in a context of biological networks such as protein-protein interaction networks, and significantly altered genes might appear clustered together in accordance with their biological functions. The context can be known signaling pathways, co-expression, or protein-protein interactions etc. Here I chose Human Protein Reference Database (HPRD) as the reference network [359]. HPRD contains manually curated scientific information of most human proteins related to their biological functions, including protein-protein interactions, post-translational modifications, enzyme-substrate relationships, and disease associations.

I use *igraph* to find communities from HPRD via short random walks. Multiple edges and self-loops are removed. I limited the genes of interests to a percentage of the top-ranked genes, and the edge weights are computed as the average rank of the two connected genes. The edge directions are ignored in HPRD in a random walk. Using the highly ranked genes as "seeds", the random walk clustering returns several dense subnetworks. The genes in the subnetworks are non-overlapping. This method recovered more cancer-related genes due to the fact they have more interacting partners than non-cancer genes [360–362]. In a test, I took out the top fifteen largest subnetworks, generated with thresholds of the top 10%-50% ranked genes.

The genes enriched in the subnetworks exhibit higher oncogene frequencies compared to the frequencies in the PromEnh rank list. As the frequencies of oncogenes dropped along with the PromEnh rank list, the oncogene frequencies stayed stable in the extracted subnetworks, implying the network clustering succeeded in selecting the biological meaningful genes in cancer (Fig. 3.17).

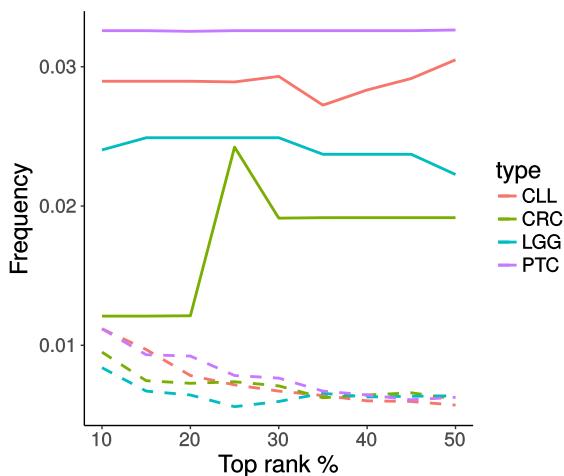


Figure 3.17: Frequencies of oncogenes in the network clustering. The solid lines represent frequencies of oncogenes in the enriched subnetworks built from top 10%-50% genes of PromEnh rank list, while the dashed lines indicate the frequencies of oncogenes in the top 10%-50% of the PromEnh rank list.

Identifying modules and pathways in diseases

Functional analysis of the subnetworks obtained as described previously uncovers a few interesting pathways for my test cases by analyzing the genes in these subnetworks with EnrichR. Many of them are general biological processes and signaling pathways, yet some of them show specific functions related to the cancer types. In table 3.4 I listed a few of the subnetworks ranked in descending order of the sum of vertex weights.

In the center of the subnetworks reside the hub regulators, which are linked by many genes with epigenetic alterations. For example, SMAD2 and SMAD3 expression have been shown to be regulated by histone modifications of their promoters [382]. In the following part, I will highlight two cancer specific pathways.

Table 3.4: Top ranked specific functions from network clustering

Case	Biological functions	Hub regulators*	Adj. P**
CLL	Toll-like receptor signaling pathway	TLR1, TLR4 [363]	1.0e-7
CRC	HIF-1 signaling pathway [364, 365]	EPAS1 [366]	7.6e-4
LGG	TGF- β signaling pathway [367]	SMADs	1.8e-6
LGG	Notch signaling pathway [349]	NOTCH2 [368]	7.5e-14
PTC	PI3K-Akt signaling pathway [369]	FGFs [370, 371]	4.0e-9
MSC	Osteoblast signaling pathway	PTH [372]	1.0e-3
NPC	Hedgehog signaling pathway [373]	ZIC3 [374]	7.6e-9
NPC	Axon guidance [375]	Ephrins [376, 377]	2.8e-18
TSC	MAPK cascade [378, 379]	MAPK1 [380]	6.5e-16
TSC	EPO receptor signaling pathway [381]	PTPRC	0.015

* Hub regulator indicates a gene surrounded by several significantly altered genes.

** Adjusted P-value (Benjamini-Hochberg method)

Chronic lymphocytic leukemia

Toll-like receptors (TLR) are iconic markers in both normal and malignant B-cells. They mediate innate immune response via pattern recognition of antigens. TLR4 and TLR9 gene expression were lower in CLL than in healthy individuals [363, 383], while TLR2 was highly expressed in both CLL [383] and acute myeloid leukemia (AML) [384]. Accordingly, I observed both hyperacetylation and hypoacetylation of the TLR genes and their neighboring enhancers in my test case (Fig. 3.18), which might lead to their differential expression in the end. The toll-like receptor signaling pathway is only observed from network analysis, and GSEA did not identify this pathway, which indicates network analysis is powerful in enriching pathways with fewer members.

Colorectal cancer

Colorectal cancer (CRC) is a malignant cancer affecting colon or rectum, and accounts for $\sim 9\%$ of all cancer deaths [385]. Network clustering of the epigenetic alterations of promoters and enhancers suggests that HIF-1 pathway is the top candidate. HIF-2 α overexpression is frequent in multiple cancers, and is associated with poor prognosis [364, 365]. I further identified EGLN3 and HIF-2 α (EPAS1), whose scores are mainly contributed from

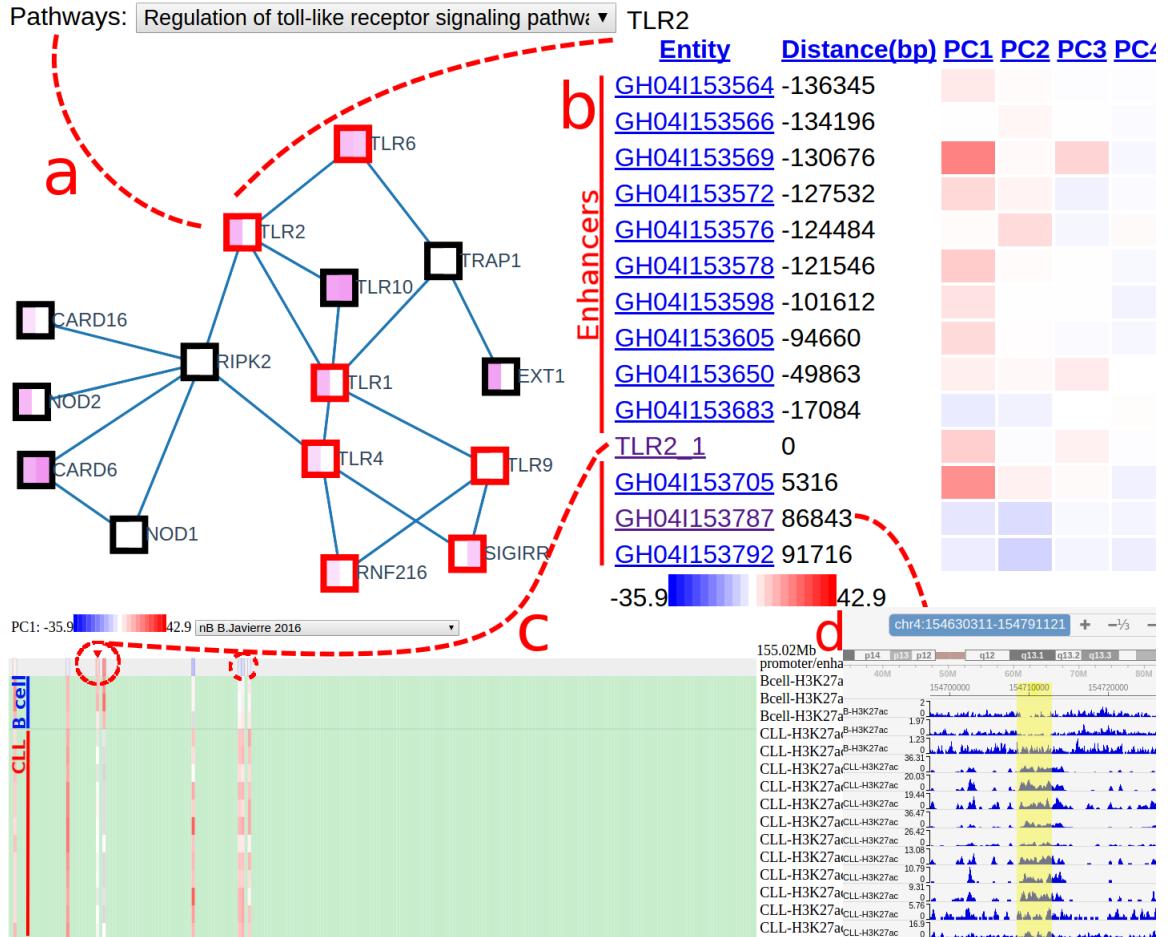


Figure 3.18: Toll-like receptor signaling pathway is associated with strong epigenetic alterations in CLL. Here I attached a screenshot from *crl* program to illustrate the analysis procedure: a. Network browser showing enriched pathways, corresponding genes are highlighted in red on selecting from the drop list; b. Clicking on a gene of interest will direct to a web page showing neighboring enhancers, as well as their dPCs in colored ranges; c. By clicking the promoter ID of the gene, a heatmap-like epigenome browser appears, showing the intensities from each epigenetic mark between tests and controls; d. Clicking the enhancer IDs will allow users to read additional information from GeneCards [383], as well as their predicted targets; Clicking a track in the epigenome browser will redirect users to WashU Epigenome Browser [259], showing a more detailed view of epigenetic signals for that point of genomic region.

epigenetic alterations of nearby enhancers (GH14I033835 and GH02I046346, respectively). These genes are also differentially expressed between CRC and normal cells. Baba *et al.* also confirmed EPAS1 overexpression in a cohort of 731 colorectal cancers [386], and Yoshimura *et al.* proved EPAS1 is associated with high-grade colorectal cancer in 88 patients [366]. The enriched network not only revealed EGLN3 and EPAS1, but also included HIF-1 α , MYC, and MAX as potential regulatory factors.

3.4.2 Discovery of novel cancer-related enhancers

Besides studying the pathways enriched for epigenetic alterations, I also analyzed the highly altered enhancers associated with certain genes of interests. PAX5 and MYC are two candidate genes discovered from their consistent high ranking in PromEnh list from all the test cases (Fig. 3.15). As their ranking in the PromOnly is not high, it suggests PAX5 and MYC are extensively regulated by enhancers. PAX5 is a key TF involved in B-cell development, and its promoters have no significant epigenetic alterations in CLL. However, taking the nearby enhancers into consideration, this gene is associated with several hyperacetylated and hypomethylated enhancers, one of which located 330 kilobases (kb) upstream of the PAX5 TSS has been found as extensively mutated in CLL [387] (Fig. 3.19). Deletion of this enhancer resulted in a 40% reduction in the expression of PAX5 and chromatin interaction of this enhancer and PAX5 have been proven from chromosome conformation capture sequencing (4C-Seq) analysis [387].

By analyzing the significantly altered oncogenes across cancer cohorts, I identified a large enhancer region, also referred to as super-enhancer, regulating MYC [58]. The enhancers are located around 20 ~ 200kb downstream of MYC, overlapping with the long non-coding (lncRNA) PVT1. PVT1 has been found to positively influence MYC expression [388–390], and is also considered as an oncogene. Cho *et al.* found that PVT1 competes with MYC for access to this super enhancer, and disruption of the PVT1 promoter leads to greater enhancer activity of MYC as well as an increase in expression [391]. The enrichments in H3K27ac and H3K4me1 signals of this enhancer have been observed in all cancer test cases (Fig. 3.20). Also, *cis*-interactions were detected using an orthogonal conformation capture technique [392]. The Hi-C profiles from the most similar available tissues or cell lines of each cancer type (K562, Huvec, GM12878 [231], Hippocampus [338] were used in upholding the chromatin interactions in CLL, CRC, PTC, LGG, respectively) are visualized using Hi-C data Browser [393] and placed above the genomic view. Another enhancer in the vicinity of MYC, known as the "Blood ENhancer Cluster" (BENC) [394], has also been recognized from increased H3K27ac

levels in CLL (Fig. 3.20 c) and LGG (Fig. 3.20 g).

3.4.3 Development of an R package for integrative epigenetics analysis

I have developed an R package named "*crl*" incorporating the differential analysis steps described in this chapter. *Crl* features a variety of tools for visualizing network clusters and epigenetic alterations, which is crucial for both hypothesis generation and detection of potential artifacts. Comparing with other network presentation tools, such as Cytoscape [395], the display of networks in *crl* is fully adapted to highlight differential epigenetic alterations from both the PromEnh and PromOnly rank list (Fig. 3.18 a), allowing one to inspect principal components of epigenetic marks for the gene of interests, and further navigate to a snapshot of the genomic regions of epigenetic tracks from each biological replicates, while existing tools, such as the WashU Epigenome Browser [259], load very slowly when visualizing a large amount of epigenetic tracks (Fig. 3.18 c).

An alpha version of this tools is available in GitHub for inspecting the enriched pathways for cancer and development test cases. It can be accessed from <http://qwang-big.github.io/crl-web/>.

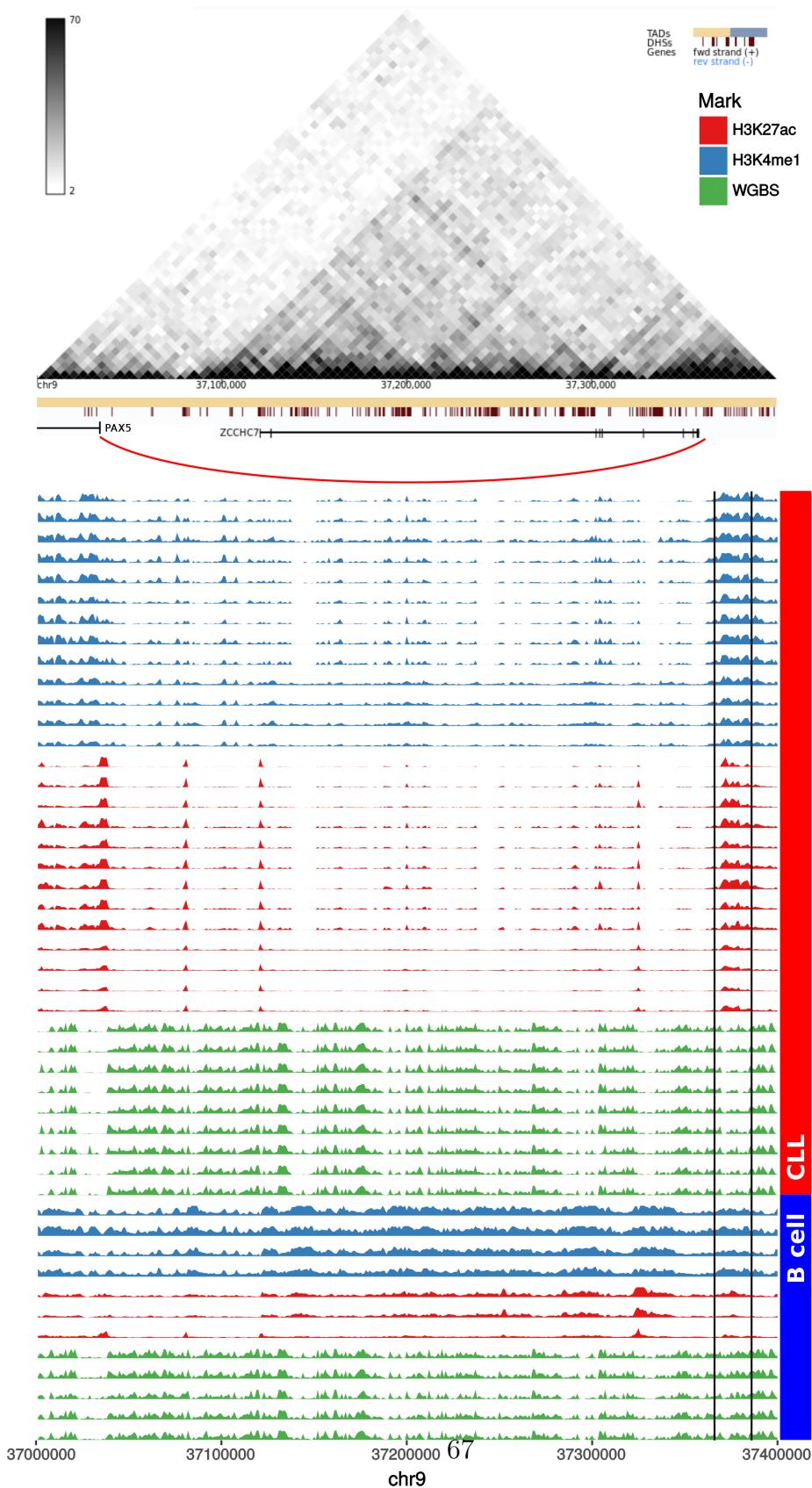


Figure 3.19: A known PAX5 enhancer in CLL exhibits hyperacetylation and hypomethylation.

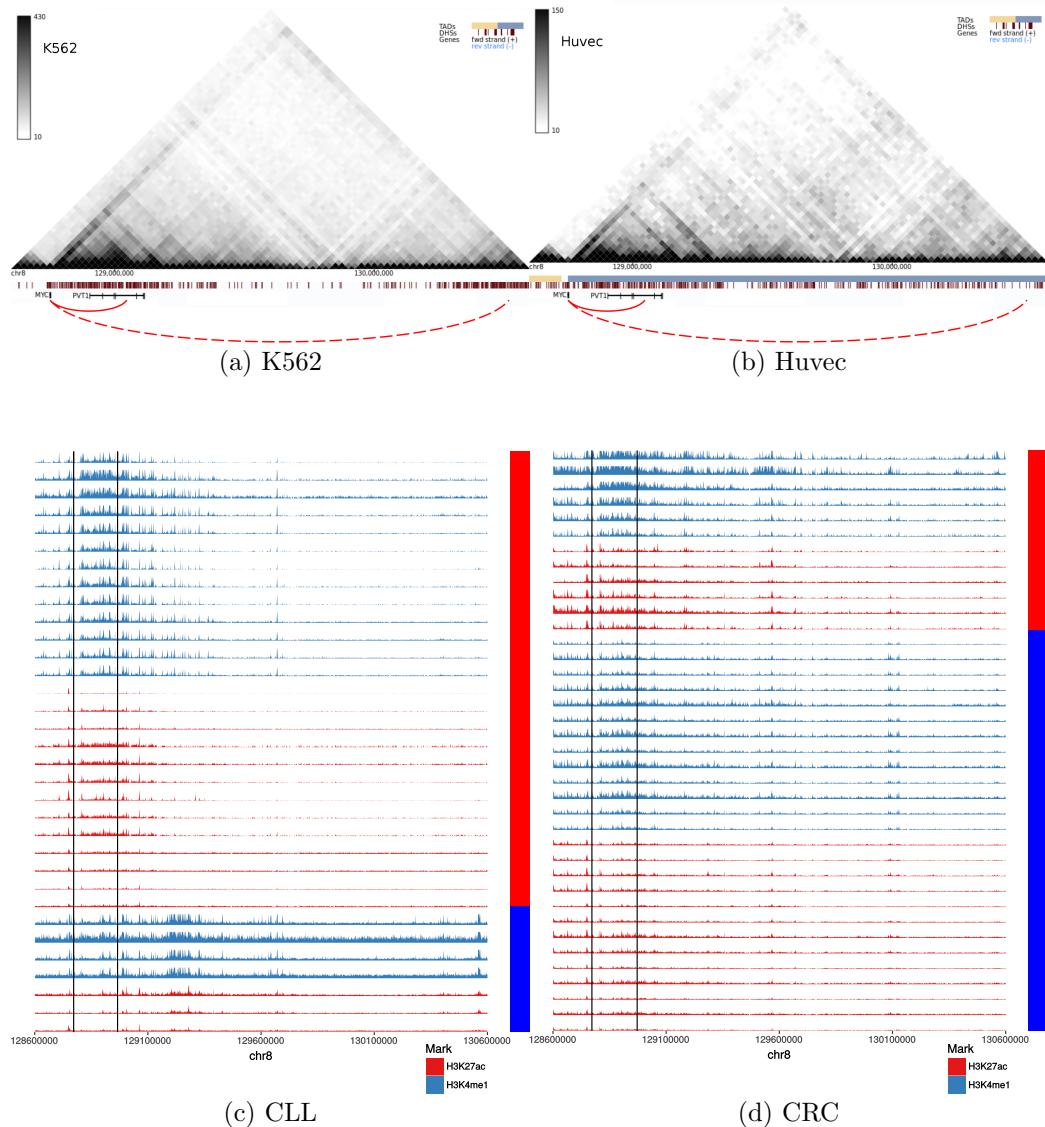


Figure 3.20: Continued on next page.

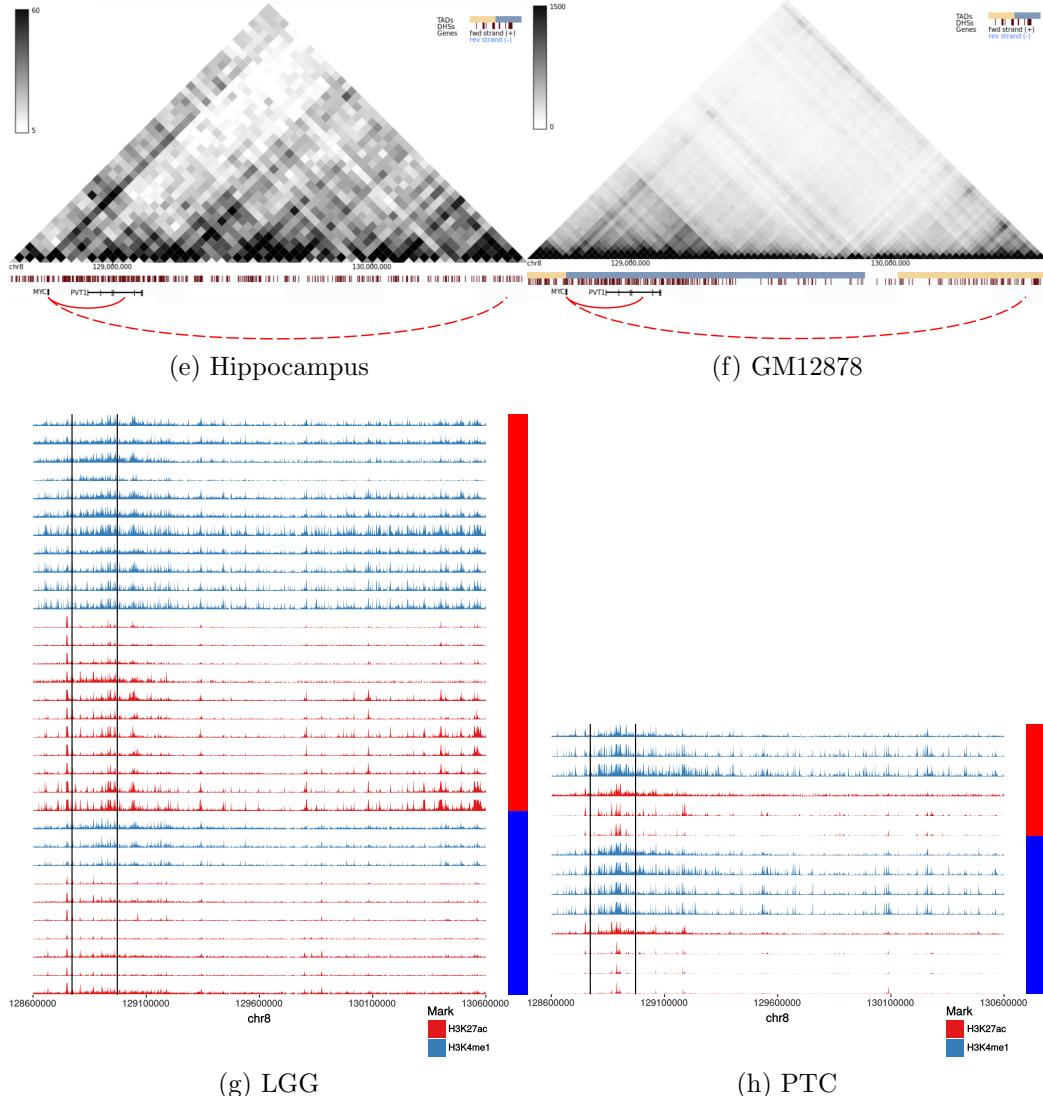


Figure 3.20: A super enhancer near MYC exhibits hyperacetylation in cancers. The proposed super enhancer regions are bordered with black solid lines on the genome tracks, wherein the cancer samples are marked with red strips, and normal samples are marked with blue strips. Solid red lines indicate the chromatin interaction between MYC and this enhancer in Hi-C, and dashed lines indicate the interactions between MYC and another superenhancer "BENC".

Bibliography

- [1] Willbanks A, Leary M, Greenshields M, et al. The evolution of epigenetics: From prokaryotes to humans and its biological consequences. *Genetics and Epigenetics*. 2016;1(8):25–36.
- [2] Jones PL, Veenstra GJC, Wade PA, et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genetics*. 1998;19(2):187–191.
- [3] Ehrlich M. DNA hypomethylation in cancer cells. *Epigenomics*. 2009;1(2):239–259.
- [4] Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*. 1983;301(5895):89–92.
- [5] Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nature Reviews Genetics*. 2002;3(6):415–428.
- [6] Gonzalo S. Epigenetic alterations in aging. *Journal of Applied Physiology*. 2010;109(2):586–597.
- [7] Esteller M. CpG island hypermethylation and tumor suppressor genes: A booming present, a brighter future. *Oncogene*. 2002;21(35 REV. ISS. 3):5427–5440.
- [8] Shen R, Tao L, Xu Y, et al. Reversibility of aberrant global DNA and estrogen receptor-alpha gene methylation distinguishes colorectal precancer from cancer. *International journal of clinical and experimental pathology*. 2009;2(1):21–33.
- [9] Zemach A, McDaniel IE, Silva P, et al. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science (New York, NY)*. 2010;328(5980):916–9.
- [10] Feng S, Cokus SJ, Zhang X, et al. Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107(19):8689–94.
- [11] Kim MY, Zilberman D. DNA methylation as a system of plant genomic immunity. *Trends in Plant Science*. 2014;19(5):320–6.
- [12] Lev Maor G, Yearim A, Ast G. The alternative role of DNA methylation in splicing regulation. *Trends in Genetics*. 2015;31(5):274–280.
- [13] Maunakea AK, Nagarajan RP, Bilenky M, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*. 2010;466(7303):253–257.
- [14] Wang X, Hu L, Wang X, et al. DNA Methylation Affects Gene Alternative Splicing in Plants: An Example from Rice. *Molecular Plant*. 2016;9(2):305–307.
- [15] ALLFREY VG, FAULKNER R, MIRSKY AE. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of RNA Synthesis. *Proceedings of the National Academy of Sciences of the United States of America*. 1964;51(1938):786–94.
- [16] Wu G, Broniscer A, McEachron TA, et al. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nature genetics*. 2012;44(3):251–3.
- [17] Behjati S, Tarpey PS, Presneau N, et al. Distinct H3F3A and H3F3B driver mu-

- tations define chondroblastoma and giant cell tumor of bone. *Nature Genetics*. 2013;45(12):1479–1482.
- [18] Tollervey JR, Lunyak VV. Epigenetics: judge, jury and executioner of stem cell fate. *Epigenetics*. 2012;7(8):823–40.
 - [19] Klocker H, Eder IE, Comuzzi B, et al. Androgen Receptor Function in Prostate Cancer Progression. In: *Prostate Cancer*. Springer, Boston, MA; 2007. p. 87–105. Available from: http://dx.doi.org/10.1007/978-1-59745-224-3{_}6.
 - [20] Hu J, Zhang Y, Zhao L, et al. Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. *Cell*. 2015;163(4):947–959.
 - [21] Vu TH, Nguyen AH, Hoffman AR. Loss of IGF2 imprinting is associated with abrogation of long-range intrachromosomal interactions in human cancer cells. *Human Molecular Genetics*. 2010;19(5):901–19.
 - [22] Hnisz D, Weintraub AS, Day DS, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*. 2016;351(6280):1454–1458.
 - [23] Rickman DS, Soong TD, Moss B, et al. Oncogene-mediated alterations in chromatin conformation. *Proceedings of the National Academy of Sciences*. 2012;109(23):9083–9088.
 - [24] Mourad R, Hsu PY, Juan L, et al. Estrogen Induces Global Reorganization of Chromatin Structure in Human Breast Cancer Cells. *PLoS ONE*. 2014;9(12):e113354.
 - [25] Schwarzer W, Abdennur N, Goloborodko A, et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*. 2017;551(7678):51–56.
 - [26] Chalkley GE, Verrijzer CP. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *The EMBO journal*. 1999;18(17):4835–45.
 - [27] Smale ST. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta - Gene Structure and Expression*. 1997;1351(1-2):73–88.
 - [28] Koudritsky M, Domany E. Positional distribution of human transcription factor binding sites. *Nucleic Acids Research*. 2008;36(21):6795–805.
 - [29] Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences*. 2006;103(5):1412–1417.
 - [30] Mirabella AC, Foster BM, Bartke T. Chromatin deregulation in disease. *Chromosoma*. 2016;125(1):75–93.
 - [31] Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes & Development*. 2011;25(10):1010–1022.
 - [32] Vermeulen M, Mulder KW, Denissov S, et al. Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell*. 2007;131(1):58–69.
 - [33] Karmadiya K, Krebs AR, Oulad-Abdelghani M, et al. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC genomics*. 2012;13:424.
 - [34] Wysocka J, Swigut T, Xiao H, et al. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*. 2006;442(7098):86–90.
 - [35] Sims RJ, Millhouse S, Chen CF, et al. Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. *Molecular Cell*. 2007;28(4):665–676.
 - [36] Chen K, Chen Z, Wu D, et al. Broad H3K4me3 is associated with increased transcrip-

- tion elongation and enhancer activity at tumor-suppressor genes. *Nature Genetics*. 2015;47(10):1149–1157.
- [37] Wang Z, Zang C, Rosenfeld JA, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics*. 2008;40(7):897–903.
 - [38] Azuara V, Perry P, Sauer S, et al. Chromatin signatures of pluripotent cell lines. *Nature Cell Biology*. 2006;8(5):532–538.
 - [39] Bernstein BE, Mikkelsen TS, Xie X, et al. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*. 2006;125(2):315–326.
 - [40] Mikkelsen TS, Ku M, Jaffe DB, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. 2007;448(7153):553–560.
 - [41] Blackwood EM, Kadonaga JT. Going the distance: A current view of enhancer action. *Science*. 1998;281(5373):60–63.
 - [42] Pennacchio LA, Bickmore W, Dean A, et al. Enhancers: Five essential questions. *Nature Reviews Genetics*. 2013;14(4):288–295.
 - [43] Maston GA, Evans SK, Green MR. Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics*. 2006;7(1):29–59.
 - [44] Kowalczyk MS, Hughes JR, Garrick D, et al. Intragenic Enhancers Act as Alternative Promoters. *Molecular Cell*. 2012;45(4):447–58.
 - [45] Fuda NJ, Ardehali MB, Lis JT. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*. 2009;461(7261):186–192.
 - [46] Koch F, Fenouil R, Gut M, et al. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Structural and Molecular Biology*. 2011;18(8):956–963.
 - [47] Schmidt D, Wilson MD, Ballester B, et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science*. 2010;328(5981):1036–40.
 - [48] May D, Blow MJ, Kaplan T, et al. Large-scale discovery of enhancers from human heart tissue. *Nature Genetics*. 2011;44(1):89–93.
 - [49] Creyghton MP, Cheng AW, Welstead GG, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*. 2010;107(50):21931–21936.
 - [50] Heintzman ND, Stuart RK, Hon G, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*. 2007;39(3):311–8.
 - [51] Calo E, Wysocka J. Modification of Enhancer Chromatin: What, How, and Why? *Molecular Cell*. 2013;49(5):825–837.
 - [52] Dorschner MO, Hawrylycz M, Humbert R, et al. High-throughput localization of functional elements by quantitative chromatin profiling. *Nature Methods*. 2004;1(3):219–25.
 - [53] Visel A, Blow MJ, Li Z, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*. 2009;457(7231):854–8.
 - [54] Stasevich TJ, Hayashi-Takanaka Y, Sato Y, et al. Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature*. 2014;516(7530):272–275.
 - [55] Charlet J, Duymich CE, Lay FD, et al. Bivalent Regions of Cytosine Methylation and H3K27 Acetylation Suggest an Active Role for DNA Methylation at Enhancers. *Molecular Cell*. 2016;62(3):422–431.
 - [56] Li W, Notani D, Rosenfeld MG. Enhancers as non-coding RNA transcription units: Recent insights and future perspectives. *Nature Reviews Genetics*. 2016;17(4):207–

223.

- [57] Hah N, Benner C, Chong LW, et al. Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs. *Proceedings of the National Academy of Sciences*. 2015;112(3):E297–302.
- [58] Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155(4):934–47.
- [59] Whyte WA, Orlando DA, Hnisz D, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153(2):307–319.
- [60] Ong CT, Corces VG. CTCF: An architectural protein bridging genome topology and function. *Nature Reviews Genetics*. 2014;15(4):234–46.
- [61] Kim TH, Abdullaev ZK, Smith AD, et al. Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. *Cell*. 2007;128(6):1231–45.
- [62] Huang S, Li X, Yusufzai TM, et al. USF1 Recruits Histone Modification Complexes and Is Critical for Maintenance of a Chromatin Barrier. *Molecular and Cellular Biology*. 2007;27(22):7991–8002.
- [63] Wood AM, Van Bortle K, Ramos E, et al. Regulation of Chromatin Organization and Inducible Gene Expression by a Drosophila Insulator. *Molecular Cell*. 2011;44(1):29–38.
- [64] Liu Z, Scannell DR, Eisen MB, et al. Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell*. 2011;146(5):720–31.
- [65] Casey SC, Vaccari M, Al-Mulla F, et al. The effect of environmental chemicals on the tumor microenvironment. *Carcinogenesis*. 2015;36:S160–S183.
- [66] Chai H, Brown RE. Field effect in cancer—an update. *Annals of clinical and laboratory science*. 2009;39(4):331–7.
- [67] Ge R, Wang W, Kramer PM, et al. Wy-14,643-induced hypomethylation of the c-myc gene in mouse liver. *Toxicological Sciences*. 2001;62(1):28–35.
- [68] Tao L, Yang S, Xie M, et al. Effect of trichloroethylene and its metabolites, dichloroacetic acid and trichloroacetic acid, on the methylation and expression of c-Jun and c-Myc protooncogenes in mouse liver: prevention by methionine. *Toxicological sciences : an official journal of the Society of Toxicology*. 2000;54(2):399–407.
- [69] Sen P, Costa M. Induction of Chromosomal Damage in Chinese Hamster Ovary Cells by Soluble and Particulate Nickel Compounds: Preferential Fragmentation of the Heterochromatic Long Arm of the X-Chromosome by Carcinogenic Crystalline NiS Particles. *Cancer Research*. 1985;45(5):2320–2325.
- [70] Lee YW, Klein CB, Kargacin B, et al. Carcinogenic nickel silences gene expression by chromatin condensation and DNA methylation: a new model for epigenetic carcinogens. *Molecular and Cellular Biology*. 1995;15(5):2547–2557.
- [71] Zhao CQ, Young MR, Diwan Ba, et al. Association of arsenic-induced malignant transformation with DNA hypomethylation and aberrant gene expression. *Proceedings of the National Academy of Sciences of the United States of America*. 1997;94(20):10907–10912.
- [72] Senut MC, Cingolani P, Sen A, et al. Epigenetics of early-life lead exposure and effects on brain development. *Epigenomics*. 2012;4(6):665–674.
- [73] Devóz PP, Gomes WR, De Araújo ML, et al. Lead (Pb) exposure induces disturbances in epigenetic status in workers exposed to this metal. *Journal of Toxicology and Environmental Health - Part A: Current Issues*. 2017;80(19-21):1098–1105.
- [74] Kondo K, Takahashi Y, Hirose Y, et al. The reduced expression and aberrant methylation of p16INK4a in chromate workers with lung cancer. *Lung Cancer*. 2006;53(3):295–302.

- [75] Takahashi Y, Kondo K, Hirose T, et al. Microsatellite instability and protein expression of the DNA mismatch repair gene, hMLH1, of lung cancer in chromate-exposed workers. *Molecular Carcinogenesis*. 2005;42(3):150–158.
- [76] Takiguchi M, Achanzar WE, Qu W, et al. Effects of cadmium on DNA-(Cytosine-5) methyltransferase activity and DNA methylation status during cadmium-induced cellular transformation. *Experimental Cell Research*. 2003;286(2):355–365.
- [77] Huang D, Zhang Y, Qi Y, et al. Global DNA hypomethylation, rather than reactive oxygen species (ROS), a potential facilitator of cadmium-stimulated K562 cell proliferation. *Toxicology Letters*. 2008;179(1):43–47.
- [78] Karaczyn AA, Golebiowski F, Kasprzak KS. Truncation, deamidation, and oxidation of histone H2B in cells cultured with nickel(II). *Chemical Research in Toxicology*. 2005;18(12):1934–1942.
- [79] Broday L, Peng W, Kuo MH, et al. Nickel compounds are novel inhibitors of histone H4 acetylation. *Cancer Research*. 2000;60(2):238–241.
- [80] Chen H, Ke Q, Kluz T, et al. Nickel Ions Increase Histone H3 Lysine 9 Dimethylation and Induce Transgene Silencing. *Molecular and Cellular Biology*. 2006;26(10):3728–3737.
- [81] Ke Q, Davidson T, Chen H, et al. Alterations of histone modifications and transgene silencing by nickel chloride. *Carcinogenesis*. 2006;27(7):1481–1488.
- [82] Golebiowski F, Kasprzak KS. Inhibition of core histones acetylation by carcinogenic nickel(II). *Molecular and Cellular Biochemistry*. 2005;279(1-2):133–139.
- [83] Zhou X, Sun H, Ellen TP, et al. Arsenite alters global histone H3 methylation. *Carcinogenesis*. 2008;29(9):1831–1836.
- [84] Zhong CX, Mass MJ. Both hypomethylation and hypermethylation of DNA associated with arsenite exposure in cultures of human cells identified by methylation-sensitive arbitrarily-primed PCR. *Toxicology Letters*. 2001;122(3):223–234.
- [85] Chai CY, Huang YC, Hung WC, et al. Arsenic salts induced autophagic cell death and hypermethylation of DAPK promoter in SV-40 immortalized human uroepithelial cells. *Toxicology Letters*. 2007;173(1):48–56.
- [86] Mass MJ, Wang L. Arsenic alters cytosine methylation patterns of the promoter of the tumor suppressor gene p53 in human lung cells: A model for a mechanism of carcinogenesis. *Mutation Research - Reviews in Mutation Research*. 1997;386(3):263–277.
- [87] Chanda S, Dasgupta UB, GuhaMazumder D, et al. DNA hypermethylation of promoter of gene p53 and p16 in arsenic-exposed people with and without malignancy. *Toxicological Sciences*. 2006;89(2):431–437.
- [88] Schnekenburger M, Talaska G, Puga A. Chromium Cross-Links Histone Deacetylase 1-DNA Methyltransferase 1 Complexes to Chromatin, Inhibiting Histone-Remodeling Marks Critical for Transcriptional Activation. *Molecular and Cellular Biology*. 2007;27(20):7089–7101.
- [89] Sun H, Zhou X, Chen H, et al. Modulation of histone methylation and MLH1 gene silencing by hexavalent chromium. *Toxicology and Applied Pharmacology*. 2009;237(3):258–266.
- [90] Natelson EA. Benzene-induced acute myeloid leukemia: a clinician's perspective. *American journal of hematology*. 2007;82(9):826–30.
- [91] McMichael AJ, Spirtas R, Kupper LL, et al. Solvent exposure and leukemia among rubber workers: an epidemiologic study. *Journal of occupational medicine : official publication of the Industrial Medical Association*. 1975;17(4):234–9.
- [92] Bollati V, Baccarelli A, Hou L, et al. Changes in DNA methylation patterns in

- subjects exposed to low-dose benzene. *Cancer Research*. 2007;67(3):876–880.
- [93] Xie Y, Zhou JJ, Zhao Y, et al. *H. pylori* modifies methylation of global genomic DNA and the gastrin gene promoter in gastric mucosal cells and gastric cancer cells. *Microbial Pathogenesis*. 2017;108:129–136.
- [94] Chan AOO, Lam SK, Wong BCY, et al. Promoter methylation of E-cadherin gene in gastric mucosa associated with *Helicobacter pylori* infection and in gastric cancer. *Gut*. 2003;52(4):502–6.
- [95] Sugiyama A, Maruta F, Ikeno T, et al. *Helicobacter pylori* infection enhances N-methyl-N-nitrosourea-induced stomach carcinogenesis in the Mongolian gerbil. *Cancer Research*. 1998;58(10):2067–2069.
- [96] Paschos K, Smith P, Anderton E, et al. Epstein-Barr virus latency in B cells leads to epigenetic repression and CpG methylation of the tumour suppressor gene Bim. *PLoS Pathogens*. 2009;5(6).
- [97] Zhang T, Ma J, Nie K, et al. Hypermethylation of the tumor suppressor gene PRDM1/Blimp-1 supports a pathogenetic role in EBV-positive Burkitt lymphom. *Blood Cancer Journal*. 2014;4(11).
- [98] Lambert MP, Paliwal A, Vaissière T, et al. Aberrant DNA methylation distinguishes hepatocellular carcinoma associated with HBV and HCV infection and alcohol intake. *Journal of Hepatology*. 2011;54(4):705–715.
- [99] Sanchez-Cespedes M, Esteller M, Wu L, et al. Gene promoter hypermethylation in tumors and serum of head and neck cancer patients. *Cancer research*. 2000;60(4):892–895.
- [100] Kato K, Hara A, Kuno T, et al. Aberrant promoter hypermethylation of p16 and MGMT genes in oral squamous cell carcinomas and the surrounding normal mucosa. *Journal of Cancer Research and Clinical Oncology*. 2006;132(11):735–743.
- [101] Shaw RJ, Liloglou T, Rogers SN, et al. Promoter methylation of P16, RAR β , E-cadherin, cyclin A1 and cytoglobin in oral cancer: Quantitative evaluation using pyrosequencing. *British Journal of Cancer*. 2006;94(4):561–568.
- [102] Hattori N, Ushijima T. Epigenetic impact of infection on carcinogenesis: mechanisms and applications. *Genome medicine*. 2016;8(1):10.
- [103] Keleher MR, Zaidi R, Shah S, et al. Maternal high-fat diet associated with altered gene expression, DNA methylation, and obesity risk in mouse offspring. *PLoS ONE*. 2018;13(2).
- [104] Dolinoy DC, Huang D, Jirtle RL. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proceedings of the National Academy of Sciences of the United States of America*. 2007;104(32):13056–61.
- [105] Dolinoy DC, Weidman JR, Waterland RA, et al. Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environmental Health Perspectives*. 2006;114(4):567–572.
- [106] Gkountela S, Zhang KX, Shafiq TA, et al. DNA demethylation dynamics in the human prenatal germline. *Cell*. 2015;161(6):1425–1436.
- [107] Morgan HD, Sutherland HGE, Martin DIK, et al. Epigenetic inheritance at the agouti locus in the mouse. *Nature Genetics*. 1999;23(3):314–318.
- [108] Smith ZD, Chan MM, Humm KC, et al. DNA methylation dynamics of the human preimplantation embryo. *Nature*. 2014;511(7511):611–5.
- [109] Cedar H, Bergman Y. Programming of DNA Methylation Patterns. *Annual Review of Biochemistry*. 2012;81(1):97–117.
- [110] Weber RG, Hoischen a, Ehrler M, et al. Frequent loss of chromosome 9, homozy-

- gous CDKN2A/p14(ARF)/CDKN2B deletion and low TSC1 mRNA expression in pleomorphic xanthoastrocytomas. *Oncogene*. 2007;26(7):1088–97.
- [111] Meissner A, Mikkelsen TS, Gu H, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*. 2008;454(7205):766–770.
- [112] Yasukochi Y, Maruyama O, Mahajan MC, et al. X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils. *Proceedings of the National Academy of Sciences*. 2010;107(8):3704–3709.
- [113] Marahrens Y. X-inactivation by chromosomal pairing events. *Genes and Development*. 1999;13(20):2624–2632.
- [114] Krusche CA, Vloet AJ, Classen-Linke I, et al. Class I histone deacetylase expression in the human cyclic endometrium and endometrial adenocarcinomas. *Human Reproduction*. 2007;22(11):2956–2966.
- [115] Guo JZ, Gorski J. Estrogen effects on modifications of chromatin proteins in the rat uterus. *Journal of Steroid Biochemistry*. 1989;32(1A):13–20.
- [116] Heryanto B, Lipson KE, Rogers PAW. Effect of angiogenesis inhibitors on oestrogen-mediated endometrial endothelial cell proliferation in the ovariectomized mouse. *Reproduction*. 2003;125(3):337–346.
- [117] Munro SK, Farquhar CM, Mitchell MD, et al. Epigenetic regulation of endometrium during the menstrual cycle. *Molecular Human Reproduction*. 2010;16(5):297–310.
- [118] Langton AK, Herrick SE, Headon DJ. An extended epidermal response heals cutaneous wounds in the absence of a hair follicle stem cell contribution. *Journal of Investigative Dermatology*. 2008;128(5):1311–1318.
- [119] Yan M, Zhang Z, Brady JR, et al. Identification of a novel death domain-containing adaptor molecule for ectodysplasin-A receptor that is mutated in crinkled mice. *Current Biology*. 2002;12(5):409–413.
- [120] Wullner U, Kaut O, DeBoni L, et al. DNA methylation in Parkinson’s disease. *Journal of Neurochemistry*. 2016;139 Suppl:108–120.
- [121] Irier HA, Jin P. Dynamics of DNA Methylation in Aging and Alzheimer’s Disease. *DNA and Cell Biology*. 2012;31(S1):S42–8.
- [122] De Souza RAG, Islam SA, McEwen LM, et al. DNA methylation profiling in human Huntington’s disease brain. *Human Molecular Genetics*. 2016;25(10):2013–2030.
- [123] Vogelstein B, Papadopoulos N, Velculescu VE, et al. Cancer genome landscapes. *Science*. 2013;340(6127):1546–1558.
- [124] Berman BP, Weisenberger DJ, Aman JF, et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear laminag-associated domains. *Nature Genetics*. 2012;44(1):40–46.
- [125] Eden A, Gaudet F, Waghmare A, et al. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science*. 2003;300(5618):455.
- [126] International Human Genome Sequencing Consortium, Human I, Sequencing G. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860–921.
- [127] Wolff EM, Byun HM, Han HF, et al. Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. *PLoS Genetics*. 2010;6(4).
- [128] Hur K, Cejas P, Feliu J, et al. Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of protooncogenes in human colorectal cancer metastasis. *Gut*. 2014;63(4):635–646.
- [129] Antelo M, Balaguer F, Shia J, et al. A High Degree of LINE-1 Hypomethylation Is a Unique Feature of Early-Onset Colorectal Cancer. *PLoS ONE*. 2012;7(9).

- [130] Shigaki H, Baba Y, Watanabe M, et al. LINE-1 hypomethylation in gastric cancer, detected by bisulfite pyrosequencing, is associated with poor prognosis. *Gastric Cancer*. 2013;16(4):480–487.
- [131] Van Hoesel AQ, Van De Velde CJH, Kuppen PJK, et al. Hypomethylation of LINE-1 in primary tumor has poor prognosis in young breast cancer patients: A retrospective cohort study. *Breast Cancer Research and Treatment*. 2012;134(3):1103–1114.
- [132] Aoki Y, Nojima M, Suzuki H, et al. Genomic vulnerability to LINE-1 hypomethylation is a potential determinant of the clinicogenetic features of multiple myeloma. *Genome Medicine*. 2012;4(12).
- [133] Zhu C, Utsunomiya T, Ikemoto T, et al. Hypomethylation of Long Interspersed Nuclear Element-1 (LINE-1) is Associated with Poor Prognosis via Activation of c-MET in Hepatocellular Carcinoma. *Annals of Surgical Oncology*. 2014;21(4):729–735.
- [134] Kreimer U, Schulz WA, Koch A, et al. HERV-K and LINE-1 DNA Methylation and Reexpression in Urothelial Carcinoma. *Frontiers in Oncology*. 2013;3:255.
- [135] Howard G, Eiges R, Gaudet F, et al. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene*. 2008;27(3):404–408.
- [136] Sunami E, de Maat M, Vu A, et al. LINE-1 hypomethylation during primary colon cancer progression. *PLoS ONE*. 2011;6(4).
- [137] Schichman SA, Caligiuri MA, Strout MP, et al. ALL-1 Tandem Duplication in Acute Myeloid Leukemia with a Normal Karyotype Involves Homologous Recombination between Alu Elements1. *Cancer Research*. 1994;54(16):4277–4280.
- [138] O’Neil J, Tchinda J, Gutierrez A, et al. Alu elements mediate MYB gene tandem duplication in human T-ALL. *The Journal of Experimental Medicine*. 2007;204(13):3059–3066.
- [139] Jeffs AR, Benjes SM, Smith TL, et al. The BCR gene recombines preferentially with Alu elements in complex BCR-ABL translocations of chronic myeloid leukaemia. *Human Molecular Genetics*. 1998;7(5):767–776.
- [140] Morse B, Rothenberg PG, South VJ, et al. Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature*. 1988;333(6168):87–90.
- [141] Teugels E, De Brakeleer S, Goelen G, et al. De novo Alu element insertions targeted to a sequence common to the BRCA1 and BRCA2 genes. *Human mutation*. 2005;26(3):284.
- [142] Oliveira C, Senz J, Kaurah P, et al. Germline CDH1 deletions in hereditary diffuse gastric cancer families. *Human Molecular Genetics*. 2009;18(9):1545–1555.
- [143] Tang Z, Steranka JP, Ma S, et al. Human transposon insertion profiling: Analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer. *Proceedings of the National Academy of Sciences*. 2017;114(5):E733–E740.
- [144] Cui H, Cruz-Correia M, Giardiello FM, et al. Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science*. 2003;299(5613):1753–5.
- [145] Roglerss CE, Yangs D, Rosettin L, et al. Altered body composition and increased frequency of diverse malignancies in insulin-like growth factor-II transgenic mice. *The Journal of biological chemistry*. 1994;269(19):13779–13784.
- [146] Shi W, Dirim F, Wolf E, et al. Methylation Reprogramming and Chromosomal Aneuploidy in In Vivo Fertilized and Cloned Rabbit Preimplantation Embryos1. *Biology of Reproduction*. 2004;71(1):340–347.
- [147] Esteller M. Aberrant DNA methylation as a cancer-inducing mechanism. *Annual Review of Pharmacology and Toxicology*. 2005;45(1):629–656.

- [148] Clouaire T, Stancheva I. Methyl-CpG binding proteins: Specialized transcriptional repressors or structural components of chromatin? *Cellular and Molecular Life Sciences*. 2008;65(10):1509–1522.
- [149] Singh S, Li SSL. Epigenetic effects of environmental chemicals bisphenol A and phthalates. *International Journal of Molecular Sciences*. 2012;13(8):10143–10153.
- [150] Mizuno SI, Chijiwa T, Okamura T, et al. Expression of DNA methyltransferases DNMT1, 3A, and 3B in normal hematopoiesis and in acute and chronic myelogenous leukemia. *Blood*. 2001;97(5):1172–9.
- [151] Lu R, Wang P, Parton T, et al. Epigenetic Perturbations by Arg882-Mutated DNMT3A Potentiate Aberrant Stem Cell Gene-Expression Program and Acute Leukemia Development. *Cancer Cell*. 2016;30(1):92–107.
- [152] Agoston AT, Argani P, Yegnasubramanian S, et al. Increased protein stability causes DNA methyltransferase 1 dysregulation in breast cancer. *The Journal of biological chemistry*. 2005;280(18):18302–10.
- [153] Butcher DT, Rodenhiser DI. Epigenetic inactivation of BRCA1 is associated with aberrant expression of CTCF and DNA methyltransferase (DNMT3B) in some sporadic breast tumours. *European Journal of Cancer*. 2007;43(1):210–9.
- [154] Ibrahim AEK, Arends MJ, Silva AL, et al. Sequential DNA methylation changes are associated with DNMT3B overexpression in colorectal neoplastic progression. *Gut*. 2011;60(4):499–508.
- [155] Noshio K, Shima K, Irahara N, et al. DNMT3B expression might contribute to CpG island methylator phenotype in colorectal cancer. *Clinical Cancer Research*. 2009;15(11):3663–71.
- [156] Saito Y, Kanai Y, Nakagawa T, et al. Increased protein expression of DNA methyltransferase (DNMT) 1 is significantly correlated with the malignant potential and poor prognosis of human hepatocellular carcinomas. *International Journal of Cancer*. 2003;105(4):527–32.
- [157] Zhao Z, Wu Q, Cheng J, et al. Depletion of DNMT3A suppressed cell proliferation and restored PTEN in hepatocellular carcinoma cell. *Journal of Biomedicine and Biotechnology*. 2010;27(6):1160–8.
- [158] Peng DF, Kanai Y, Sawada M, et al. DNA methylation of multiple tumor-related genes in association with overexpression of DNA methyltransferase 1 (DNMT1) during multistage carcinogenesis of the pancreas. *Carcinogenesis*. 2006;27(6):1160–8.
- [159] Kobayashi Y, Absher DM, Gulzar ZG, et al. DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. *Genome Research*. 2011;21(7):1017–27.
- [160] Zhao SL, Zhu ST, Hao X, et al. Effects of DNA methyltransferase 1 inhibition on esophageal squamous cell carcinoma. *Diseases of the Esophagus*. 2011;24(8):601–10.
- [161] Pathania R, Ramachandran S, Mariappan G, et al. Combined inhibition of DNMT and HDAC blocks the tumorigenicity of cancer stem-like cells and attenuates mammary tumor growth. *Cancer Research*. 2016;76(11):3224–3235.
- [162] Soengas MS, Capodieci P, Polsky D, et al. Inactivation of the apoptosis effector Apaf-1 in malignant melanoma. *Nature*. 2001;409(6817):207–211.
- [163] Wargo JA, Robbins PF, Li Y, et al. Recognition of NY-ESO-1+ tumor cells by engineered lymphocytes is enhanced by improved vector design and epigenetic modulation of tumor antigen expression. *Cancer Immunology, Immunotherapy*. 2009;58(3):383–394.
- [164] Li Y, Seto E. HDACs and HDAC Inhibitors in Cancer Development and Therapy.

- Cold Spring Harbor perspectives in medicine. 2016;6(10).
- [165] Ellis L, Hammers H, Pili R. Targeting tumor angiogenesis with histone deacetylase inhibitors. *Cancer Letters*. 2009;280(2):145–153.
 - [166] Marquard L, Poulsen CB, Gjerdrum LM, et al. Histone deacetylase 1, 2, 6 and acetylated histone H4 in B- and T-cell lymphomas. *Histopathology*. 2009;54(6):688–98.
 - [167] Adams H, Fritzsche FR, Dirnhofer S, et al. Class I histone deacetylases 1, 2 and 3 are highly expressed in classical Hodgkin's lymphoma. *Expert opinion on therapeutic targets*. 2010;14(6):577–84.
 - [168] Svechnikova I, Almqvist PM, Ekström TJ. HDAC inhibitors effectively induce cell type-specific differentiation in human glioblastoma cell lines of different origin. *International Journal of Oncology*. 2008;32(4):821–827.
 - [169] Adamopoulou E, Naumann U. HDAC inhibitors and their potential applications to glioblastoma therapy. *Oncogen Immunology*. 2013;2(8).
 - [170] Wallace IV GC, Haar CP, Vandergrift WA, et al. Multi-targeted DATS prevents tumor progression and promotes apoptosis in ectopic glioblastoma xenografts in SCID mice via HDAC inhibition. *Journal of Neuro-Oncology*. 2013;114(1):43–50.
 - [171] Holm K, Grabau D, Lövgren K, et al. Global H3K27 trimethylation and EZH2 abundance in breast tumor subtypes. *Molecular Oncology*. 2012;6(5):494–506.
 - [172] Yamagishi M, Uchimaru K. Targeting EZH2 in cancer therapy. *Current Opinion in Oncology*. 2017;29(5):375–381.
 - [173] Tjian R. The binding site on SV40 DNA for a T antigen-related protein. *Cell*. 1978;13(1):165–79.
 - [174] Roeder RG. The role of general initiation factors in transcription by RNA polymerase II. *Trends in biochemical sciences*. 1996;21(9):327–35.
 - [175] Nikolov DB, Burley SK. RNA polymerase II transcription initiation: a structural view. *Proceedings of the National Academy of Sciences of the United States of America*. 1997;94(1):15–22.
 - [176] Yin Y, Morganova E, Jolma A, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*. 2017;356(6337).
 - [177] Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews Molecular Cell Biology*. 2015;16(3):178–89.
 - [178] Xie H, Hoffmann HM, Iyer AK, et al. Chromatin status and transcription factor binding to gonadotropin promoters in gonadotrope cell lines. *Reproductive biology and endocrinology*. 2017;15(1):86.
 - [179] Ng HH, Surani MA. The transcriptional and signalling networks of pluripotency. *Nature Cell Biology*. 2011;13(5):490–6.
 - [180] Orkin SH, Hochedlinger K. Chromatin connections to pluripotency and cellular reprogramming. *Cell*. 2011;145(6):835–50.
 - [181] Young RA. Control of the embryonic stem cell state. *Cell*. 2011;144(6):940–54.
 - [182] Boyer LA, Lee TI, Cole MF, et al. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell*. 2005;122(6):947–56.
 - [183] Loh YH, Wu Q, Chew JL, et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genetics*. 2006;38(4):431–40.
 - [184] Dohm JC, Lottaz C, Borodina T, et al. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*. 2008;36(16):e105.
 - [185] Kuan PF, Chung D, Pan G, et al. A statistical framework for the analysis of ChIP-Seq data. *Journal of the American Statistical Association*. 2011;106(495):891–903.

- [186] Vega VB, Cheung E, Palanisamy N, et al. Inherent signals in sequencing-based Chromatin-ImmunoPrecipitation control libraries. *PLoS ONE*. 2009;4(4):e5241.
- [187] ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* (New York, NY). 2004;306(5696):636–40.
- [188] Lun ATL, Smyth GK. CsaW: A Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Research*. 2015;44(5):1–10.
- [189] Rye MB, Sætrom P, Drabløs F. A manually curated ChIP-seq benchmark demonstrates room for improvement in current peak-finder programs. *Nucleic Acids Research*. 2011;39(4).
- [190] Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biology*. 2008;9(9):R137.
- [191] Zang C, Schones DE, Zeng C, et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics*. 2009;25(15):1952–1958.
- [192] Chen L, Wang C, Qin ZS, et al. A novel statistical method for quantitative comparison of multiple ChIP-seq datasets. *Bioinformatics*. 2015;31(12):1889–96.
- [193] Xu H, Wei CL, Lin F, et al. An HMM approach to genome-wide identification of differential histone modification sites from ChIP-seq data. *Bioinformatics*. 2008;24(20):2344–9.
- [194] Nair NU, Das Sahu A, Bucher P, et al. Chipnorm: A statistical method for normalizing and identifying differential regions in histone modification chip-seq libraries. *PLoS ONE*. 2012;7(8):e39573.
- [195] Liang K, Keles S. Detecting differential binding of transcription factors with ChIP-seq. *Bioinformatics*. 2012;28(1):121–2.
- [196] Stark R, Brown G. DiffBind: differential binding analysis of ChIP-Seq peak data; 2011. Available from: <http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf>.
- [197] Shao Z, Zhang Y, Yuan GC, et al. MANorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biology*. 2012;13(3):R16.
- [198] Song Q, Smith AD. Identifying dispersed epigenomic domains from ChIP-Seq data. *Bioinformatics*. 2011;27(6):870–1.
- [199] Steinhäuser S, Kurzawa N, Eils R, et al. A comprehensive comparison of tools for differential ChIP-seq analysis. *Briefings in Bioinformatics*. 2016;17(6):953–966.
- [200] Ross-Innes CS, Stark R, Teschendorff AE, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012;481(7381):389–393.
- [201] Du P, Zhang X, Huang CC, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010;11:587.
- [202] Bhasin JM, Hu B, Ting AH. MethylAction: Detecting differentially methylated regions that distinguish biological subtypes. *Nucleic Acids Research*. 2016;44(1):106–16.
- [203] Assenov Y, Müller F, Lutsik P, et al. Comprehensive analysis of DNA methylation data with RnBeads. *Nature Methods*. 2014;11(11):1138–1140.
- [204] Hebestreit K, Dugas M, Klein HU. Detection of significantly differentially methylated regions in targeted bisulfite sequencing data. *Bioinformatics*. 2013;29(13):1647–53.
- [205] Stalder J, Larsen A, Engel JD, et al. Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNase I. *Cell*. 1980;20(2):451–460.

- [206] Gross D. Nuclease Hypersensitive Sites In Chromatin. *Annual Review of Biochemistry*. 1988;57:159–97.
- [207] Boyle AP, Davis S, Shulha HP, et al. High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell*. 2008;132(2):311–322.
- [208] Buenrostro JD, Giresi PG, Zaba LC, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*. 2013;10(12):1213–8.
- [209] Giresi PG, Kim J, McDaniell RM, et al. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Research*. 2007;17(6):877–85.
- [210] Hesselberth JR, Chen X, Zhang Z, et al. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nature Methods*. 2009;6(4):283–9.
- [211] Giresi PG, Lieb JD. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods*. 2009;48(3):233–9.
- [212] Kumar V, Muratani M, Rayan NA, et al. Uniform, optimal signal processing of mapped deep-sequencing data. *Nature Biotechnology*. 2013;31(7):615–622.
- [213] Yan H, Tian S, Slager SL, et al. Genome-Wide Epigenetic Studies in Human Disease: A Primer on -Omic Technologies. *American journal of epidemiology*. 2016;183(2):96–109.
- [214] Daugherty AC, Yeo RW, Buenrostro JD, et al. Chromatin accessibility dynamics reveal novel functional enhancers in *C. elegans*. *Genome Research*. 2017;27(12):2096–2107.
- [215] Quillien A, Abdalla M, Yu J, et al. Robust Identification of Developmentally Active Endothelial Enhancers in Zebrafish Using FANS-Assisted ATAC-Seq. *Cell Reports*. 2017;20(3):709–720.
- [216] Bowman SK. Discovering enhancers by mapping chromatin features in primary tissue. *Genomics*. 2015;106(3):140–144.
- [217] Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods*. 2012;9(3):215–6.
- [218] Mammana A, Chung HR. Chromatin segmentation based on a probabilistic model for read counts explains a large portion of the epigenome. *Genome Biology*. 2015;16(151).
- [219] Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317–30.
- [220] Thurner M, van de Bunt M, Torres JM, et al. Integration of human pancreatic islet genomic data refines regulatory mechanisms at type 2 diabetes susceptibility loci. *eLife*. 2018;7(7):e31977.
- [221] Taberlay PC, Statham AL, Kelly TK, et al. Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. *Genome Research*. 2014;24(9):1421–32.
- [222] Hall AW, Battenhouse AM, Shivram H, et al. Bivalent chromatin domains in glioblastoma reveal a subtype-specific signature of glioma stem cells. *Cancer Research*. 2018;78(10):2463–2474.
- [223] Lin IH, Chen DT, Chang YF, et al. Hierarchical clustering of breast cancer methylomes revealed differentially methylated and expressed breast cancer genes. *PLoS ONE*. 2015;10(2):e0118453.
- [224] Dekker J, Rippe K, Dekker M, et al. Capturing chromosome conformation. *Science*. 2002;295(5558):1306–11.

- [225] Simonis M, Klous P, Splinter E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genetics*. 2006;38(11):1348–1354.
- [226] Dostie J, Richmond TA, Arnaout RA, et al. Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Research*. 2006;16(10):1299–1309.
- [227] Fullwood MJ, Liu MH, Pan YF, et al. An oestrogen-receptor- α -bound human chromatin interactome. *Nature*. 2009;462(7269):58–64.
- [228] Belton JM, McCord RP, Gibcus JH, et al. Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods*. 2012;58(3):268–276.
- [229] Lieberman-Aiden E, Van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289–293.
- [230] Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485(7398):376–80.
- [231] Rao SSP, Huntley MH, Durand NC, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159(7):1665–1680.
- [232] Bourgo RJ, Singhal H, Greene GL. Capture of associated targets on chromatin links long-distance chromatin looping to transcriptional coordination. *Nature Communications*. 2016;7:12893.
- [233] Hughes JR, Roberts N, McGowan S, et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature genetics*. 2014;46(2):205–12.
- [234] Andersson R, Gebhard C, Miguel-Escalada I, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507(7493):455–461.
- [235] He B, Chen C, Teng L, et al. Global view of enhancer-promoter interactome in human cells. *Proceedings of the National Academy of Sciences*. 2014;111(21):E2191–E2199.
- [236] Nora EP, Lajoie BR, Schulz EG, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012;485(7398):381–385.
- [237] Kikuta H, Laplante M, Navratilova P, et al. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Research*. 2007;17(5):545–555.
- [238] Larkin DM, Pape G, Donthu R, et al. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Research*. 2009;19(5):770–777.
- [239] Thurman RE, Rynes E, Humbert R, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012;489(7414):75–82.
- [240] Ernst J, Kheradpour P, Mikkelsen TS, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011;473(7345):43–49.
- [241] Roy S, Siahpirani AF, Chasman D, et al. A predictive modeling approach for cell line-specific long-range regulatory interactions. *Nucleic Acids Research*. 2016;44(4):1977–8.
- [242] Cao Q, Anyansi C, Hu X, et al. Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines. *Nature Genetics*. 2017;49(10):1428–1436.
- [243] Dobes M, Khurana VG, Shadbolt B, et al. Increasing incidence of glioblastoma multi-forme and meningioma, and decreasing incidence of Schwannoma (2000-2008): Find-

- ings of a multicenter Australian study. *Surgical neurology international*. 2011;2:176.
- [244] Johnson DR, O'Neill BP. Glioblastoma survival in the United States before and during the temozolomide era. *Journal of neuro-oncology*. 2012;107(2):359–64.
- [245] Verhaak RGW, Hoadley KA, Purdom E, et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010;17(1):98–110.
- [246] Sturm D, Witt H, Hovestadt V, et al. Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. *Cancer Cell*. 2012;22(4):425–437.
- [247] Xu W, Yang H, Liu Y, et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell*. 2011;19(1):17–30.
- [248] Yang Z, Jiang B, Wang Y, et al. 2-HG Inhibits Necroptosis by Stimulating DNMT1-Dependent Hypermethylation of the RIP3 Promoter. *Cell Reports*. 2017;19(9):1846–1857.
- [249] Zhang W, Xu J. DNA methyltransferases and their roles in tumorigenesis. *Biomarker Research*. 2017;5(1).
- [250] Lu C, Thompson CB. Metabolic regulation of epigenetics. *Cell Metabolism*. 2012;16(1):9–17.
- [251] Q W, B H, X H, et al. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell*. 2017;32(1):42–56.e6.
- [252] Noushmehr H, Weisenberger DJ, Diefes K, et al. Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma. *Cancer Cell*. 2010;17(5):510–522.
- [253] Turcan S, Rohle D, Goenka A, et al. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature*. 2012;483(7390):479–483.
- [254] Dillies MA, Rau A, Aubert J, et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*. 2013;14(6):671–683.
- [255] Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nature biotechnology*. 2011;29(1):24–6.
- [256] Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: From properties to genome-wide predictions. *Nature Reviews Genetics*. 2014;15(4):272–86.
- [257] Heinz S, Romanoski CE, Benner C, et al. The selection and function of cell type-specific enhancers. *Nature Reviews Molecular Cell Biology*. 2015;16(3):144–54.
- [258] Lin CY, Erkek S, Tong Y, et al. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature*. 2016;530(7588):57–62.
- [259] Zhou X, Lowdon RF, Li D, et al. Exploring long-range genome interactions using the WashU Epigenome Browser. *Nature Methods*. 2013;10(5):375–376.
- [260] McLean CY, Bristor D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*. 2010;28(5):495–501.
- [261] Lewis CA, Brault C, Peck B, et al. SREBP maintains lipid biosynthesis and viability of cancer cells under lipid- and oxygen-deprived conditions and defines a gene signature associated with poor survival in glioblastoma multiforme. *Oncogene*. 2015;34(40):5128–40.
- [262] Cherry AE, Stella N. G protein-coupled receptors as oncogenic signals in glioma: emerging therapeutic avenues. *Neuroscience*. 2014;278:222–36.
- [263] Berezowska S, Schlegel J. Targeting ErbB Receptors in High-Grade Glioma. *Current Pharmaceutical Design*. 2011;17(23):2468–87.

- [264] Haley EM, Kim Y. The role of basic fibroblast growth factor in glioblastoma multiforme and glioblastoma stem cells and in their in vitro culture. *Cancer Letters*. 2014;346(1):1–5.
- [265] Johnston ALM, Lun X, Rahn JJ, et al. The p75 neurotrophin receptor is a central regulator of glioma invasion. *PLoS Biology*. 2007;5(8):e212.
- [266] Bernstein BE, Kamal M, Lindblad-Toh K, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*. 2005;120(2):169–181.
- [267] Kim TH, Barrera LO, Zheng M, et al. A high-resolution map of active promoters in the human genome. *Nature*. 2005;436(7052):876–880.
- [268] Sangpairoj K, Vivithanaporn P, Apisawetakan S, et al. RUNX1 Regulates Migration, Invasion, and Angiogenesis via p38 MAPK Pathway in Human Glioblastoma. *Cellular and Molecular Neurobiology*. 2017;37(7):1243–1255.
- [269] Herms JW, Von Loewenich FD, Behnke J, et al. C-MYC oncogene family expression in glioblastoma and survival. *Surgical Neurology*. 1999;51(5):536–42.
- [270] Bjerke L, Mackay A, Nandhabalan M, et al. Histone H3.3 mutations drive pediatric glioblastoma through upregulation of MYCN. *Cancer Discovery*. 2013;3(5):512–9.
- [271] Rheinbay E, Suvà ML, Gillespie SM, et al. An Aberrant Transcription Factor Network Essential for Wnt Signaling and Stem Cell Maintenance in Glioblastoma. *Cell Reports*. 2013;3(5):1567–79.
- [272] Park NI, Guilhamon P, Desai K, et al. ASCL1 Reorganizes Chromatin to Direct Neuronal Fate and Suppress Tumorigenicity of Glioblastoma Stem Cells. *Cell stem cell*. 2017;21(2):209–224.e7.
- [273] Saunders LR, Bankovich AJ, Anderson WC, et al. A DLL3-targeted antibody-drug conjugate eradicates high-grade pulmonary neuroendocrine tumor-initiating cells in vivo. *Science Translational Medicine*. 2015;7(302):302ra136.
- [274] Muraguchi T, Tanaka S, Yamada D, et al. NKX2.2 suppresses self-renewal of glioma-initiating cells. *Cancer Research*. 2011;71(3):1135–1145.
- [275] Trépan AL, Bouchart C, Rorive S, et al. Identification of OLIG2 as the most specific glioblastoma stem cell marker starting from comparative analysis of data from similar DNA chip microarray platforms. *Tumor Biology*. 2015;36(3):1943–53.
- [276] Leelatian N, Ihrie RA. Head of the Class: OLIG2 and Glioblastoma Phenotype. *Cancer Cell*. 2016;29(5):613–615.
- [277] Weigle B, Ebner R, Temme A, et al. Highly specific overexpression of the transcription factor SOX11 in human malignant gliomas. *Oncology Reports*. 2005;13(1):139–44.
- [278] Chen X, Hu H, He L, et al. A novel subtype classification and risk of breast cancer by histone modification profiling. *Breast Cancer Research and Treatment*. 2016;157(2):267–279.
- [279] Bogdanović O, Long SW, Van Heeringen SJ, et al. Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis. *Genome Research*. 2011;21(8):1313–1327.
- [280] Kelley DZ, Flam EL, Izumchenko E, et al. Integrated analysis of whole-genome ChIP-Seq and RNA-Seq data of primary head and neck tumor samples associates HPV integration sites with open chromatin marks. *Cancer Research*. 2017;77(23):6538–6550.
- [281] Gal-Yam EN, Egger G, Iniguez L, et al. Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(35):12979–84.

- [282] Hinoue T, Weisenberger DJ, Lange CPE, et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Research*. 2012;22(2):271–282.
- [283] Hahn MA, Li AX, Wu X, et al. Loss of the polycomb mark from bivalent promoters leads to activation of cancer-promoting genes in colorectal tumors. *Cancer Research*. 2014;74(13):3617–3629.
- [284] Ohm JE, McGarvey KM, Yu X, et al. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nature Genetics*. 2007;39(2):237–242.
- [285] Kretzmer H, Bernhart SH, Wang W, et al. DNA methylome analysis in Burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control. *Nature Genetics*. 2015;47(11):1316–1325.
- [286] Bernhart SH, Kretzmer H, Holdt LM, et al. Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. *Scientific Reports*. 2016;6(1):37393.
- [287] Liang K, Keles S. Detecting differential binding of transcription factors with ChIP-seq. *Bioinformatics*. 2012;28(1):121–2.
- [288] Baldi P, Sadowski P, Whiteson D. Searching for exotic particles in high-energy physics with deep learning. *Nature Communications*. 2014;5:4308.
- [289] Wang HW, Sun HJ, Chang TY, et al. Discovering monotonic stemness marker genes from time-series stem cell microarray data. *BMC Genomics*. 2015;16.
- [290] Hartigan JA, Wong MA. A K-Means Clustering Algorithm. *Applied Statistics*. 1979;28(1):100–108.
- [291] Cao B, Yang Y, Pan Y, et al. Epigenetic silencing of CXCL14 induced colorectal cancer migration and invasion. *Discovery medicine*. 2013;16(88):137–47.
- [292] Meshcheryakova A, Svoboda M, Tahir A, et al. Exploring the role of sphingolipid machinery during the epithelial to mesenchymal transition program using an integrative approach. *Oncotarget*. 2016;7(16):22295–323.
- [293] Thiery JP. Epithelial–mesenchymal transitions in tumour progression. *Nature Reviews Cancer*. 2002;2(6):442–54.
- [294] Tönjes M, Barbus S, Park YJ, et al. BCAT1 promotes cell proliferation through amino acid catabolism in gliomas carrying wild-type IDH1. *Nature medicine*. 2013;19(7):901–908.
- [295] Joshi AD, Parsons DW, Velculescu VE, et al. Sodium ion channel mutations in glioblastoma patients correlate with shorter survival. *Molecular Cancer*. 2011;10:17.
- [296] Puget S, Philippe C, Bax DA, et al. Mesenchymal transition and PDGFRA amplification/mutation are key distinct oncogenic events in pediatric diffuse intrinsic pontine gliomas. *PloS one*. 2012;7(2):e30313.
- [297] Kupp R, Shtayer L, Tien AC, et al. Lineage-Restricted OLIG2-RTK Signaling Governs the Molecular Subtype of Glioma Stem-like Cells. *Cell reports*. 2016;16(11):2838–2845.
- [298] Sudmant PH, Alexis MS, Burge CB. Meta-analysis of RNA-seq expression data across species, tissues and studies. *Genome Biology*. 2015;16(1).
- [299] Sun X, Dalpiaz D, Wu D, et al. Statistical inference for time course RNA-Seq data using a negative binomial mixed-effect model. *BMC Bioinformatics*. 2016;17(1).
- [300] Ziller MJ, Edri R, Yaffe Y, et al. Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature*. 2014;518(7539):355–359.
- [301] Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*. 2007;447(7143):425–432.
- [302] Stadler MB, Murr R, Burger L, et al. DNA-binding factors shape the mouse methy-

- lome at distal regulatory regions. *Nature*. 2011;480(7378):490–495.
- [303] Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH roadmap epigenomics mapping consortium. *Nature Biotechnology*. 2010;28(10):1045–1048.
 - [304] Adams D, Altucci L, Antonarakis SE, et al. BLUEPRINT to decode the epigenetic signature written in blood. *Nature Biotechnology*. 2012;30(3):224–226.
 - [305] Stunnenberg HG, Consortium TIHE, Hirst M, et al. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell*. 2016;167(5):1145–1149.
 - [306] Kent WJ, Zweig AS, Barber G, et al. BigWig and BigBed: Enabling browsing of large distributed datasets. *Bioinformatics*. 2010;26(17):2204–2207.
 - [307] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40.
 - [308] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014;15(12):550.
 - [309] Périer RC, Praz V, Junier T, et al. The eukaryotic promoter database (EPD). *Nucleic acids research*. 2000;28(1):302–3.
 - [310] Fishlevich S, Nudel R, Rappaport N, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*. 2017; p. bax028.
 - [311] ENCODE Project Consortium AIEoDEith. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74.
 - [312] Zerbino DR, Wilder SP, Johnson N, et al. The Ensembl Regulatory Build. *Genome Biology*. 2015;16(1).
 - [313] Visel A, Minovitsky S, Dubchak I, et al. VISTA Enhancer Browser - A database of tissue-specific human enhancers. *Nucleic Acids Research*. 2007;35(Database issue):D88–92.
 - [314] Hinrichs AS. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Research*. 2006;34(Database issue):D590–8.
 - [315] Hicks SC, Irizarry RA. quantro: A data-driven approach to guide the choice of an appropriate normalization method. *Genome Biology*. 2015;16:117.
 - [316] Bolstad BM, Irizarry Ra, Astrand M, et al. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185–93.
 - [317] Cloonan N, Forrest ARR, Kolle G, et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods*. 2008;5(7):613–619.
 - [318] Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*. 2010;11(3):R25.
 - [319] Bullard JH, Purdom E, Hansen KD, et al. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*. 2010;11.
 - [320] Yousefi P, Huen K, Schall RA, et al. Considerations for normalization of DNA methylation data by Illumina 450K BeadChip assay in population studies. *Epigenetics*. 2013;8(11):1141–1152.
 - [321] Bilodeau S, Kagey MH, Frampton GM, et al. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes and Development*. 2009;23(21):2484–2489.
 - [322] Kasowski M, Grubert F, Heffelfinger C, et al. Variation in transcription factor binding among humans. *Science*. 2010;328(5975):232–235.
 - [323] Di Cerbo V, Schneider R. Cancers with wrong HATs: The impact of acetylation.

Briefings in Functional Genomics. 2013;12(3):231–243.

- [324] Walker EJ, Zhang C, Castelo-Branco P, et al. Monoallelic expression determines oncogenic progression and outcome in benign and malignant brain tumors. *Cancer research*. 2012;72(3):636–44.
- [325] Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends in Genetics*. 2013;29(10):569–574.
- [326] Hagarman JA, Motley MP, Kristjansdottir K, et al. Coordinate Regulation of DNA Methylation and H3K27me3 in Mouse Embryonic Stem Cells. *PLoS ONE*. 2013;8(1).
- [327] Yang X, Hu B, Hou Y, et al. Silencing of developmental genes by H3K27me3 and DNA methylation reflects the discrepant plasticity of embryonic and extraembryonic lineages. *Cell Research*. 2018;28(5):593–596.
- [328] Dai H, Leeder JS, Cui Y. A modified generalized fisher method for combining probabilities from dependent tests. *Frontiers in Genetics*. 2014;5:32.
- [329] Ji H, Li X, Wang Qf, et al. Differential principal component analysis of ChIP-seq. *Proceedings of the National Academy of Sciences*. 2013;110(17):6789–6794.
- [330] Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. *Nature*. 2009;461(7261):199–205.
- [331] Fudenberg G, Mirny LA. Higher-order chromatin structure: Bridging physics and biology. *Current Opinion in Genetics and Development*. 2012;22(2):115–124.
- [332] Pombo A, Nicodemi M. Physical mechanisms behind the large scale features of chromatin organization. *Transcription*. 2014;5(2):e28447.
- [333] Mifsud B, Tavares-Cadete F, Young AN, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics*. 2015;47(6):598–606.
- [334] Javierre BM, Sewitz S, Cairns J, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*. 2016;167(5):1369–1384.e19.
- [335] Zeitz MJ, Ay F, Heidmann JD, et al. Genomic Interaction Profiles in Breast Cancer Reveal Altered Chromatin Architecture. *PLoS ONE*. 2013;8(9).
- [336] Elemento O, Rubin MA, Rickman DS. Oncogenic transcription factors as master regulators of chromatin topology: A new role for ERG in prostate cancer. *Cell Cycle*. 2012;11(18):3380–3383.
- [337] Ferraro A. Altered primary chromatin structures and their implications in cancer development. *Cellular Oncology*. 2016;39(3):195–210.
- [338] Schmitt AD, Hu M, Jung I, et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Reports*. 2016;17(8):2042–2059.
- [339] Lajoie BR, Dekker J, Kaplan N. The Hitchhiker’s guide to Hi-C analysis: Practical guidelines. *Methods*. 2015;72(C):65–75.
- [340] Baxter JS, Leavy OC, Dryden NH, et al. Capture Hi-C identifies putative target genes at 33 breast cancer risk loci. *Nature Communications*. 2018;9(1).
- [341] Rubin AJ, Barajas BC, Furlan-Magaril M, et al. Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nature genetics*. 2017;49(10):1522–1528.
- [342] Jäger R, Migliorini G, Henrion M, et al. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nature Communications*. 2015;6:6178.
- [343] Brin S, Page L. The anatomy of a large scale hypertextual Web search engine. *Computer Networks and ISDN Systems*. 1998;30(1-7):107–17.
- [344] Iván G, Grolmusz V. When the web meets the cell: Using personalized PageRank

- for analyzing protein interaction networks. *Bioinformatics*. 2011;27(3):405–407.
- [345] Morrison JL, Breitling R, Higham DJ, et al. GeneRank: Using search engine technology for the analysis of microarray experiments. *BMC Bioinformatics*. 2005;6.
 - [346] Kurdistani SK, Tavazoie S, Grunstein M. Mapping global histone acetylation patterns to gene expression. *Cell*. 2004;117(6):721–733.
 - [347] Lin CS, Xin ZC, Dai J, et al. Commonly used mesenchymal stem cell markers and tracking labels: Limitations and challenges. *Histology and Histopathology*. 2013;28(9):1109–1116.
 - [348] Luo Y, Cai J, Liu Y, et al. Microarray analysis of selected genes in neural stem and progenitor cells. *Journal of Neurochemistry*. 2002;83(6):1481–1497.
 - [349] Kotliarova S, Fine HA. SnapShot: glioblastoma multiforme. *Cancer cell*. 2012;21(5):710–710.e1.
 - [350] Ten Hacken E, Guièze R, Wu CJ. SnapShot: Chronic Lymphocytic Leukemia. *Cancer cell*. 2017;32(5):716–716.e1.
 - [351] Ciccone M, Ferrajoli A, Keating MJ, et al. SnapShot: chronic lymphocytic leukemia. *Cancer cell*. 2014;26(5):770–770.e1.
 - [352] Forbes SA, Beare D, Boutselakis H, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic acids research*. 2017;45(D1):D777–D783.
 - [353] Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, et al. IntOGen-mutations identifies cancer drivers across tumor types. *Nature Methods*. 2013;10(11):1081–1084.
 - [354] Rappaport N, Nativ N, Stelzer G, et al. MalaCards: an integrated compendium for diseases and their annotation. *Database*. 2013; p. bat018.
 - [355] Lachmann A, Torre D, Keenan AB, et al. Massive mining of publicly available RNA-seq data from human and mouse. *Nat Commun*. 2018;9(1):1366.
 - [356] Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*. 2016;44(W1):W90–7.
 - [357] Khan A, Mathelier A, Zhang X. Super-enhancers are transcriptionally more active and cell-type-specific than stretch enhancers. *Epigenetics*. 2018; p. 1–13.
 - [358] Geeleher P, Hartnett L, Egan LJ, et al. Gene-set analysis is severely biased when applied to genome-wide methylation data. *Bioinformatics*. 2013;29(15):1851–7.
 - [359] Keshava Prasad TS, Goel R, Kandasamy K, et al. Human Protein Reference Database—2009 update. *Nucleic acids research*. 2009;37(Database issue):D767–72.
 - [360] Lin J, Gan CM, Zhang X, et al. A multidimensional analysis of genes mutated in breast and colorectal cancers. *Genome Research*. 2007;17(9):1304–1318.
 - [361] Jonsson PF, Bates PA. Global topological features of cancer proteins in the human interactome. *Bioinformatics*. 2006;22(18):2291–2297.
 - [362] Davoli T, Xu AW, Mengwasser KE, et al. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell*. 2013;155(4):948–62.
 - [363] Arvaniti E, Ntoufa S, Papakonstantinou N, et al. Toll-like receptor signaling pathway in chronic lymphocytic leukemia: Distinct gene expression profiles of potential pathogenic significance in specific subsets of patients. *Haematologica*. 2011;96(11):1644–1652.
 - [364] Keith B, Johnson RS, Simon MC. HIF1 α and HIF2 α : sibling rivalry in hypoxic tumour growth and progression. *Nature reviews Cancer*. 2011;12(1):9–22.
 - [365] Qing G, Simon MC. Hypoxia inducible factor-2 α : a critical mediator of aggressive tumor phenotypes. *Current Opinion in Genetics and Development*. 2009;19(1):60–66.

- [366] Yoshimura H, Dhar DK, Kohno H, et al. Prognostic Impact of Hypoxia-Inducible Factors 1 α and 2 α in Colorectal Cancer Patients. *Clinical Cancer Research*. 2004;10(24):8554–8560.
- [367] Joseph JV, Balasubramaniyan V, Walenkamp A, et al. TGF- β as a therapeutic target in high grade gliomas - promises and challenges. *Biochemical pharmacology*. 2013;85(4):478–85.
- [368] Tchorz JS, Tome M, Cloëtta D, et al. Constitutive Notch2 signaling in neural stem cells promotes tumorigenic features and astroglial lineage entry. *Cell death & disease*. 2012;3:e325.
- [369] Nozhat Z, Hedayati M. PI3K/AKT Pathway and Its Mediators in Thyroid Carcinomas. *Molecular diagnosis & therapy*. 2016;20(1):13–26.
- [370] Eggo MC, Hopkins JM, Franklyn JA, et al. Expression of fibroblast growth factors in thyroid cancer. *The Journal of clinical endocrinology and metabolism*. 1995;80(3):1006–11.
- [371] St Bernard R, Zheng L, Liu W, et al. Fibroblast growth factor receptors as molecular targets in thyroid carcinoma. *Endocrinology*. 2005;146(3):1145–53.
- [372] Yu B, Zhao X, Yang C, et al. PTH Induces Differentiation of Mesenchymal Stem Cells by Enhancing BMP Signaling. *J Bone Miner Res J Bone Miner Res*. 2012;27(9):2001–2014.
- [373] Wu SM, Choo ABH, Yap MGS, et al. Role of Sonic hedgehog signaling and the expression of its components in human embryonic stem cells. *Stem cell research*. 2010;4(1):38–49.
- [374] Kumar A, Declercq J, Eggermont K, et al. Zic3 induces conversion of human fibroblasts to stable neural progenitor-like cells. *Journal of molecular cell biology*. 2012;4(4):252–5.
- [375] Hinck L. The versatile roles of "axon guidance" cues in tissue morphogenesis. *Developmental Cell*. 2004;7(6):783–793.
- [376] Qiu R, Wang X, Davy A, et al. Regulation of neural progenitor cell state by ephrin-B. *The Journal of cell biology*. 2008;181(6):973–83.
- [377] Jiao Jw, Feldheim DA, Chen DF. Ephrins as negative regulators of adult neurogenesis in diverse regions of the central nervous system. *Proceedings of the National Academy of Sciences*. 2008;105(25):8778–8783.
- [378] Adams RH, Porras A, Alonso G, et al. Essential role of p38 α MAP kinase in placental but not embryonic cardiovascular development. *Molecular Cell*. 2000;6(1):109–116.
- [379] Vaillancourt C, Lanoix D, Le Bellego F, et al. Involvement of MAPK signalling in human villous trophoblast differentiation. *Mini reviews in medicinal chemistry*. 2009;9(8):962–73.
- [380] Daoud G, Amyot M, Rassart E, et al. ERK1/2 and p38 regulate trophoblasts differentiation in human term placenta. *The Journal of physiology*. 2005;566(Pt 2):409–23.
- [381] Fairchild Benyo D, Conrad KP. Expression of the Erythropoietin Receptor by Trophoblast Cells in the Human Placenta. *Biology of Reproduction*. 1999;60(4):861–870.
- [382] Tang YN, Ding WQ, Guo XJ, et al. Epigenetic regulation of Smad2 and Smad3 by profilin-2 promotes lung cancer growth and metastasis. *Nature Communications*. 2015;6:8230.
- [383] Rybka J, Butrym A, Wróbel T, et al. The Expression of Toll-Like Receptors in Patients with B-Cell Chronic Lymphocytic Leukemia. *Archivum Immunologiae et Therapiae Experimentalis*. 2016;64:147–150.
- [384] Rybka J, Butrym A, Wróbel T, et al. The expression of Toll-like receptors in pa-

tients with acute myeloid leukemia treated with induction chemotherapy. *Leukemia research.* 2015;39(3):318–22.

- [385] Wiseman M. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. *The Proceedings of the Nutrition Society.* 2008;67(3):253–6.
- [386] Baba Y, Noshio K, Shima K, et al. HIF1A overexpression is associated with poor prognosis in a cohort of 731 colorectal cancers. *American Journal of Pathology.* 2010;176(5):2292–2301.
- [387] Puente XS, Beà S, Valdés-Mas R, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature.* 2015;526(7574):519–524.
- [388] Tseng YY, Moriarity BS, Gong W, et al. PVT1 dependence in cancer with MYC copy-number increase. *Nature.* 2014;512(1):82–86.
- [389] Kim T, Cui R, Jeon YJ, et al. Long-range interaction and correlation between MYC enhancer and oncogenic long noncoding RNA CARLo-5. *Proceedings of the National Academy of Sciences.* 2014;111(11):4173–8.
- [390] Shi J, Whyte WA, Zepeda-Mendoza CJ, et al. Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes and Development.* 2013;27(24):2648–62.
- [391] Cho SW, Xu J, Sun R, et al. Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell.* 2018;173(6):1398–1412.e22.
- [392] Schwartzman O, Mukamel Z, Oded-Elkayam N, et al. UMI-4C for quantitative and targeted chromosomal contact profiling. *Nature Methods.* 2016;13(8):685–691.
- [393] Wang Y, Zhang B, Zhang L, et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biology.* 2018;19(1):151.
- [394] Bahr C, Von Paleske L, Uslu VV, et al. A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature.* 2018;553(7689):515–520.
- [395] Shannon P, Markiel A, Ozier O, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research.* 2003;13(11):2498–2504.
- [396] Predeus AV, Gopalakrishnan S, Huang Y, et al. Targeted Chromatin Profiling Reveals Novel Enhancers in Ig H and Ig L Chain Loci. *The Journal of Immunology.* 2014;192(3):1064–1070.
- [397] Rozowsky J, Euskirchen G, Auerbach RK, et al. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nature Biotechnology.* 2009;27(1):66–75.
- [398] Heintzman ND, Hon GC, Hawkins RD, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature.* 2009;459(7243):108–112.
- [399] Mohrs M, Blankenspoor CM, Wang ZE, et al. Deletion of a coordinate regulator of type 2 cytokine expression in mice. *Nature Immunology.* 2001;2(9):842–7.
- [400] Jadamba E, Shin M. A novel approach to significant pathway identification using pathway interaction network from PPI data. *Biochip Journal.* 2014;8(1):22–27.
- [401] Paulsen J, Rødland EA, Holden L, et al. A statistical model of ChIA-PET data for accurate detection of chromatin 3D interactions. *Nucleic Acids Research.* 2014;42(18).
- [402] von Mering C, Huynen M, Jaeggi D, et al. STRING: A database of predicted functional associations between proteins. *Nucleic Acids Research.* 2003;31(1):258–261.

- [403] Islam MO, Kanemura Y, Tajria J, et al. Functional expression of ABCG2 transporter in human neural stem/progenitor cells. *Neuroscience Research*. 2005;52(1):75–82.
- [404] Castro DS, Martynoga B, Parras C, et al. A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. *Genes and Development*. 2011;25(9):930–945.
- [405] Molofsky AV, Pardal R, Iwashita T, et al. Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature*. 2003;425(6961):962–967.
- [406] Peh GSL, Lang RJ, Pera MF, et al. CD133 Expression by Neural Progenitors Derived from Human Embryonic Stem Cells and Its Use for Their Prospective Isolation. *Stem Cells and Development*. 2009;18(2):269–282.
- [407] Bang SY, Kwon SH, Yi SH, et al. Epigenetic activation of the Foxa2 gene is required for maintaining the potential of neural precursor cells to differentiate into dopaminergic neurons after expansion. *Stem Cells and Development*. 2015;24(4):520–533.
- [408] Kirkeby A, Greathouse S, Wolf DA, et al. Generation of Regionally Specified Neural Progenitors and Functional Neurons from Human Embryonic Stem Cells under Defined Conditions. *Cell Reports*. 2012;1(6):703–714.
- [409] Kim DY, Hwang I, Muller FL, et al. Functional regulation of FoxO1 in neural stem cell differentiation. *Cell Death and Differentiation*. 2015;22(12):2034–2045.
- [410] Fathi A, Hatami M, Hajihosseini V, et al. Comprehensive gene expression analysis of human embryonic stem cells during differentiation into neural cells. *PLoS ONE*. 2011;6(7).
- [411] Livesey FJ, Young TL, Cepko CL. An analysis of the gene expression program of mammalian neural progenitor cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(5):1374–1379.
- [412] Zhao J, Yao Y, Xu C, et al. Expression of GAP-43 in fibroblast cell lines influences the orientation of cell division. *International journal of developmental neuroscience : the official journal of the International Society for Developmental Neuroscience*. 2011;29(4):469–74.
- [413] Doetsch F, Caillé I, Lim DA, et al. Subventricular Zone Astrocytes Are Neural Stem Cells in the Adult Mammalian Brain. *Cell*. 1999;97(6):703–716.
- [414] Middeldorp, Boer, Sluijs, et al. GFAPdelta in radial glia and subventricular zone progenitors in the developing human cortex. *Development*. 2010;137:313–321.
- [415] Maurer MH, Geomor HK, Bürgers HF, et al. Adult neural stem cells express glucose transporters GLUT1 and GLUT3 and regulate GLUT3 expression. *FEBS Letters*. 2006;580(18):4430–4434.
- [416] Kobayashi T, Kageyama R. Hes1 regulates embryonic stem cell differentiation by suppressing Notch signaling. *Genes to Cells*. 2010;15(7):689–698.
- [417] Shimojo H, Ohtsuka T, Kageyama R. Dynamic Expression of Notch Signaling Genes in Neural Stem/Progenitor Cells. *Frontiers in Neuroscience*. 2011;5:78.
- [418] Kaneko J, Chiba C. Immunohistochemical analysis of Musashi-1 expression during retinal regeneration of adult newt. *Neuroscience Letters*. 2009;450(3):252–257.
- [419] Murdoch B, Roskams AJ. A Novel Embryonic Nestin-Expressing Radial Glia-Like Progenitor Gives Rise to Zonally Restricted Olfactory and Vomeronasal Neurons. *Journal of Neuroscience*. 2008;28(16):4271–4282.
- [420] Steiner B, Zurborg S, Hörster H, et al. Differential 24 h responsiveness of Prox1-expressing precursor cells in adult hippocampal neurogenesis to physical activity, environmental enrichment, and kainic acid-induced seizures. *Neuroscience*. 2008;154(2):521–529.
- [421] Heng YHE, McLeay RC, Harvey TJ, et al. NFIX Regulates Neural Progen-

- itor Cell Differentiation During Hippocampal Morphogenesis. *Cerebral cortex*. 2014;24(1):261–79.
- [422] Yang X, Klein R, Tian X, et al. Notch activation induces apoptosis in neural progenitor cells through a p53-dependent pathway. *Developmental Biology*. 2004;269(1):81–94.
- [423] Cui XY, Hu QD, Tekaya M, et al. NB-3/Notch1 pathway via Deltex1 promotes neural progenitor cell differentiation into oligodendrocytes. *Journal of Biological Chemistry*. 2004;279(24):25858–25865.
- [424] Dominici C, Moreno-Bravo JA, Puiggros SR, et al. Floor-plate-derived netrin-1 is dispensable for commissural axon guidance. *Nature*. 2017;545(7654):350–354.
- [425] Li XJ, Du ZW, Zarnowska ED, et al. Specification of motoneurons from human embryonic stem cells. *Nat Biotechnol*. 2005;23(2):2121–2152.
- [426] Basch ML, Bronner-Fraser M, García-Castro MI. Specification of the neural crest occurs during gastrulation and requires Pax7. *Nature*. 2006;441(7090):218–222.
- [427] Blake JA, Ziman MR. Pax genes: regulators of lineage specification and progenitor cell maintenance. *Development*. 2014;141(4):737–751.
- [428] Pankratz MT, Li XJ, LaVaute TM, et al. Directed Neural Differentiation of Human Embryonic Stem Cells via an Obligated Primitive Anterior Stage. *Stem Cells*. 2007;25(6):1511–1520.
- [429] Vives V, Alonso G, Solal AC, et al. Visualization of S100B-positive neurons and glia in the central nervous system of EGFP transgenic mice. *Journal of Comparative Neurology*. 2003;457(4):404–419.
- [430] Matsumoto S, Banine F, Struve J, et al. Brg1 is required for murine neural stem cell maintenance and gliogenesis. *Developmental Biology*. 2006;289(2):372–383.
- [431] Venere M, Han YG, Bell R, et al. Sox1 marks an activated neural stem/progenitor cell in the hippocampus. *Development*. 2012;139(21):3938–3949.
- [432] Bergsland M, Werme M, Malewicz M, et al. The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes and Development*. 2006;20(24):3475–3486.
- [433] Graham V, Khudyakov J, Ellis P, et al. SOX2 functions to maintain neural progenitor identity. *Neuron*. 2003;39(5):749–765.
- [434] Ellis P, Fagan BM, Magness ST, et al. SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult. *Developmental Neuroscience*. 2004;26(2-4):148–165.
- [435] Wang TW, Stromberg GP, Whitney JT, et al. Sox3 expression identifies neural progenitors in persistent neonatal and adult mouse forebrain germinative zones. *Journal of Comparative Neurology*. 2006;497(1):88–100.
- [436] Scott CE, Wynn SL, Sesay A, et al. SOX9 induces and maintains neural stem cells. *Nature Neuroscience*. 2010;13(10):1181–1189.
- [437] Uittenbogaard M, Chiaramello a. Expression of the bHLH transcription factor Tcf12 (ME1) gene is linked to the expansion of precursor cell populations during neurogenesis. *Brain research Gene expression patterns*. 2002;1(2):115–121.
- [438] Nakagawa T, Miyazaki T, Miyamoto O, et al. Regional expression of the radial glial marker vimentin at different stages of the kindling process. *Epilepsy Research*. 2004;61(1-3):141–151.
- [439] Arai F, Ohneda O, Miyamoto T, et al. Mesenchymal Stem Cells in Perichondrium Express Activated Leukocyte Cell Adhesion Molecule and Participate in Bone Marrow Formation. *The Journal of Experimental Medicine*. 2002;195(12):1549–1563.
- [440] Frobel J, Hemeda H, Lenz M, et al. Epigenetic rejuvenation of mesenchymal stromal cells derived from induced pluripotent stem cells. *Stem Cell Reports*. 2014;3(3):414–

422.

- [441] Obara C, Takizawa K, Tomiyama K, et al. Differentiation and molecular properties of mesenchymal stem cells derived from murine induced pluripotent stem cells derived on gelatin or collagen. *Stem Cells International*. 2016;9013089.
- [442] Abdallah BM, Boissy P, Tan Q, et al. dlk1/FA1 regulates the function of human bone marrow mesenchymal stem cells by modulating gene expression of pro-inflammatory cytokines and immune response-related factors. *The Journal of biological chemistry*. 2007;282(10):7339–51.
- [443] Abdallah BM, Jensen CH, Gutierrez G, et al. Regulation of human skeletal stem cells differentiation by Dlk1/Pref-1. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*. 2004;19(5):841–852.
- [444] Kubo H, Shimizu M, Taya Y, et al. Identification of mesenchymal stem cell (MSC)-transcription factors by microarray and knockdown analyses, and signature molecule-marked MSC in bone marrow by immunohistochemistry. *Genes to Cells*. 2009;14(3):407–424.
- [445] Almalki SG, Agrawal DK. Key transcription factors in the differentiation of mesenchymal stem cells. *Differentiation*. 2016;92(1-2):41–51.
- [446] Cui LL, Nitzsche F, Pryazhnikov E, et al. Integrin $\alpha 4$ Overexpression on Rat Mesenchymal Stem Cells Enhances Transmigration and Reduces Cerebral Embolism After Intracarotid Injection. *Stroke*. 2017;48(10):2895–2900.
- [447] Ball SG, Shuttleworth A, Kielty CM. Inhibition of platelet-derived growth factor receptor signaling regulates Oct4 and Nanog expression, cell shape, and mesenchymal stem cell potency. *Stem Cells*. 2012;30(3):548–560.
- [448] Farahani RM, Xaymardan M. Platelet-Derived Growth Factor Receptor Alpha as a Marker of Mesenchymal Stem Cells in Development and Stem Cell Biology. *Stem Cells International*. 2015;362753:1–8.
- [449] Han SM, Han SH, Coh YR, et al. Enhanced proliferation and differentiation of Oct4- And Sox2-overexpressing human adipose tissue mesenchymal stem cells. *Experimental and Molecular Medicine*. 2014;46(6).
- [450] Matic I, Antunovic M, Brkic S, et al. Expression of OCT-4 and SOX-2 in bone marrow-derived human mesenchymal stem cells during osteogenic differentiation. *Macedonian Journal of Medical Sciences*. 2016;4(1).
- [451] Park SB, Seo KW, So AY, et al. SOX2 has a crucial role in the lineage determination and proliferation of mesenchymal stem cells through Dickkopf-1 and c-MYC. *Cell Death and Differentiation*. 2012;19(3):534–545.
- [452] Tiwari N, Tiwari VK, Waldmeier L, et al. Sox4 Is a Master Regulator of Epithelial-Mesenchymal Transition by Controlling Ezh2 Expression and Epigenetic Reprogramming. *Cancer Cell*. 2013;23(6):768–783.
- [453] Bradshaw AD, Sage EH. SPARC, a matricellular protein that functions in cellular differentiation and tissue response to injury. *Journal of Clinical Investigation*. 2001;107(9):1049–1054.
- [454] Ivaska J, Pallari HM, Nevo J, et al. Novel functions of vimentin in cell adhesion, migration, and signaling. *Experimental Cell Research*. 2007;313(10):2050–2062.
- [455] Rhee C, Lee BK, Beck S, et al. Mechanisms of transcription factor-mediated direct reprogramming of mouse embryonic stem cells to trophoblast stem-like cells. *Nucleic acids research*. 2017;45(17):10103–10114.
- [456] Kubaczka C, Senner C, Araúzo-Bravo MJ, et al. Derivation and maintenance of murine trophoblast stem cells under defined conditions. *Stem Cell Reports*. 2014;2(2):232–242.

- [457] Douglas GC, VandeVoort CA, Kumar P, et al. Trophoblast stem cells: Models for investigating trophectoderm differentiation and placental development. *Endocrine Reviews*. 2009;30(3):228–240.
- [458] Ohinata Y, Tsukiyama T. Establishment of trophoblast stem cells under defined culture conditions in mice. *PLoS ONE*. 2014;9(9).
- [459] Peiffer I, Belhomme D, Barbet R, et al. Simultaneous differentiation of endothelial and trophoblastic cells derived from human embryonic stem cells. *Stem cells and development*. 2007;16(3):393–402.
- [460] Strumpf D. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development*. 2005;132(9):2093–2102.
- [461] Chen Y, Wang K, Gong YG, et al. Roles of CDX2 and EOMES in human induced trophoblast progenitor cells. *Biochemical and Biophysical Research Communications*. 2013;431(2):197–202.
- [462] Li Y, Moretto-Zita M, Soncin F, et al. BMP4-directed trophoblast differentiation of human embryonic stem cells is mediated through a ΔNp63+ cytотrophoblast stem cell state. *Development*. 2013;140(19):3965–76.
- [463] Schulz LC, Ezashi T, Das P, et al. Human Embryonic Stem Cells as Models for Trophoblast Differentiation. *Placenta*. 2008;29(Suppl A):S10–6.
- [464] Lee CQE, Gardner L, Turco M, et al. What Is Trophoblast? A Combination of Criteria Define Human First-Trimester Trophoblast. *Stem Cell Reports*. 2016;6(2):257–272.
- [465] Kidder BL, Palmer S. Examination of transcriptional networks reveals an important role for TCFAP2C, SMARCA4, and EOMES in trophoblast stem cell maintenance. *Genome Research*. 2010;20(4):458–472.
- [466] Tanaka S, Kunath T, Hadjantonakis AK, et al. Promotion of trophoblast stem cell proliferation by FGF4. *Science*. 1998;282(5396):2072–5.
- [467] Haffner-Krausz R, Gorivodsky M, Chen Y, et al. Expression of Fgfr2 in the early mouse embryo indicates its involvement in preimplantation development. *Mechanisms of Development*. 1999;85(1-2):167–172.
- [468] Selesniemi K, Reedy M, Gultice A, et al. Transforming growth factor-beta induces differentiation of the labyrinthine trophoblast stem cell line SM10. *Stem cells and development*. 2005;14(6):697–711.
- [469] Selesniemi K, Albers RE, Brown TL. Id2 Mediates Differentiation of Labyrinthine Placental Progenitor Cell Line, SM10. *Stem Cells and Development*. 2016;25(13):959–974.
- [470] Liang H, Zhang Q, Lu J, et al. MSX2 Induces Trophoblast Invasion in Human Placenta. *PLoS ONE*. 2016;11(4).
- [471] Yagi R, Kohn MJ, Karavanova I, et al. Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development*. 2007;134(21):3827–36.
- [472] Rubio-Perez C, Tamborero D, Schroeder MP, et al. In Silico Prescription of Anticancer Drugs to Cohorts of 28 Tumor Types Reveals Targeting Opportunities. *Cancer Cell*. 2015;27(3):382–396.
- [473] Wang L, Lawrence MS, Wan Y, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *The New England journal of medicine*. 2011;365(26):2497–506.
- [474] Guarini A, Marinelli M, Tavolaro S, et al. Atm gene alterations in chronic lymphocytic leukemia patients induce a distinct gene expression profile and predict disease progression. *Haematologica*. 2012;97(1):47–55.

- [475] Landau DA, Carter SL, Stojanov P, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*. 2013;152(4):714–726.
- [476] Baliakas P, Hadzidimitriou A, Sutton LA, et al. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia*. 2015;29(2):329–36.
- [477] Alhourani E, Othman MAK, Melo JB, et al. BIRC3 alterations in chronic and B-cell acute lymphocytic leukemia patients. *Oncology letters*. 2016;11(5):3240–3246.
- [478] Landau DA, Tausch E, Taylor-Weiner AN, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature*. 2015;526(7574):525–530.
- [479] Jebaraj BMC, Kienle D, Bühler A, et al. BRAF mutations in chronic lymphocytic leukemia. *Leukemia & lymphoma*. 2013;54(6):1177–82.
- [480] Rodríguez D, Bretones G, Quesada V, et al. Mutations in CHD2 cause defective association with active chromatin in chronic lymphocytic leukemia. *Blood*. 2015;126(2):195–202.
- [481] Quesada V, Conde L, Villamor N, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature Genetics*. 2011;44(1):47–52.
- [482] Ghobrial IM, Bone ND, Stenson MJ, et al. Expression of the chemokine receptors CXCR4 and CCR7 and disease progression in B-cell chronic lymphocytic leukemia/small lymphocytic lymphoma. *Mayo Clin Proc*. 2004;79(3):318–25.
- [483] Möhle R, Failenschmid C, Bautz F, et al. Overexpression of the chemokine receptor CXCR4 in B cell chronic lymphocytic leukemia is associated with increased functional response to stromal cell-derived factor-1 (SDF-1). *Leukemia*. 1999;13(12):1954–9.
- [484] Barretina J, Juncà J, Llano A, et al. CXCR4 and SDF-1 expression in B-cell chronic lymphocytic leukemia and stage of the disease. *Annals of Hematology*. 2003;82(8):500–505.
- [485] Crowther-Swanepoel D, Qureshi M, Dyer MJS, et al. Genetic variation in CXCR4 and risk of chronic lymphocytic leukemia. *Blood*. 2009;114(23):4843–6.
- [486] Ojha J, Secreto CR, Rabe KG, et al. Identification of recurrent truncated DDX3X mutations in chronic lymphocytic leukaemia. *British Journal of Haematology*. 2015;169(3):445–448.
- [487] Young E, Noerenberg D, Mansouri L, et al. EGR2 mutations define a new clinically aggressive subgroup of chronic lymphocytic leukemia. *Leukemia*. 2017;31(7):1547–1554.
- [488] Jeromin S, Weissmann S, Haferlach C, et al. SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. *Leukemia*. 2014;28(1):108–117.
- [489] Havelange V, Pekarsky Y, Nakamura T, et al. IRF4 mutations in chronic lymphocytic leukemia. *Blood*. 2011;118(10):2827–2829.
- [490] Martínez-Trillo A, Navarro A, Aymerich M, et al. Clinical impact of MYD88 mutations in chronic lymphocytic leukemia. *Blood*. 2016;127(12):1611–1613.
- [491] Rossi D, Rasi S, Fabbri G, et al. Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood*. 2012;119(2):521–529.
- [492] Clifford R, Louis T, Robbe P, et al. SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. *Blood*. 2014;123(7):1021–1031.
- [493] Rossi D. SAMHD1: A new gene for CLL. *Blood*. 2014;123(7):951–952.
- [494] Buchner M, Fuchs S, Prinz G, et al. Spleen tyrosine kinase is overexpressed and represents a potential therapeutic target in chronic lymphocytic leukemia. *Cancer*

Research. 2009;69(13):5424–5432.

- [495] Baudot AD, Jeandel PY, Mouska X, et al. The tyrosine kinase Syk regulates the survival of chronic lymphocytic leukemia B cells through PKCdelta and proteasome-dependent regulation of Mcl-1 expression. *Oncogene*. 2009;28(37):3261–3273.
- [496] Hoellenriegel J, Coffey GP, Sinha U, et al. Selective, novel spleen tyrosine kinase (Syk) inhibitors suppress chronic lymphocytic leukemia B-cell activation and migration. *Leukemia*. 2012;26(7):1576–1583.
- [497] Wiestner A. ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile. *Blood*. 2003;101(12):4944–4951.
- [498] Chen L, Widhopf G, Huynh L, et al. Expression of ZAP-70 is associated with increased B-cell receptor signaling in chronic lymphocytic leukemia. *Blood*. 2002;100(13):4609–4614.
- [499] Sausen M, Leary RJ, Jones S, et al. Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. *Nature Genetics*. 2013;45(1):12–17.
- [500] Wiestler B, Capper D, Holland-Letz T, et al. ATRX loss refines the classification of anaplastic gliomas and identifies a subgroup of IDH mutant astrocytic tumors with better prognosis. *Acta Neuropathologica*. 2013;126(3):443–451.
- [501] Jiao Y, Killela PJ, Reitman ZJ, et al. Frequent ATRX, CIC, FUBP1 and IDH1 mutations refine the classification of malignant gliomas. *Oncotarget*. 2012;3(7):709–22.
- [502] Dahiya S, Emmett RJ, Haydon DH, et al. BRAF-V600E mutation in pediatric and adult glioblastoma. *Neuro-Oncology*. 2014;16(2):318–319.
- [503] et al McLendon, R, Friedman, A, Bigner, D, Van Meir, EG, Brat, DJ, Mastrogiannakis, GM, Olson, JJ. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008;455(7216):1061–1068.
- [504] Knobbe CB, Reifenberger J, Reifenberger G. Mutation analysis of the Ras pathway genes NRAS, HRAS, KRAS and BRAF in glioblastomas. *Acta Neuropathologica*. 2004;108(6):467–470.
- [505] Toth J, Egervari K, Klekner A, et al. Analysis of EGFR gene amplification, protein over-expression and tyrosine kinase domain mutation in recurrent glioblastoma. *Pathology oncology research*. 2009;15(2):225–229.
- [506] Mukasa A, Wykosky J, Ligon KL, et al. Mutant EGFR is required for maintenance of glioma growth in vivo, and its ablation leads to escape from receptor dependence. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107(6):2616–2621.
- [507] Baumgarten P, Harter PN, Tönjes M, et al. Loss of FUBP1 expression in gliomas predicts FUBP1 mutation and is associated with oligodendroglial differentiation, IDH1 mutation and 1p/19q loss of heterozygosity. *Neuropathology and Applied Neurobiology*. 2014;40(2):205–216.
- [508] Parsons DW, Jones S, Zhang X, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science*. 2008;321(5897):1807–1812.
- [509] Cohen AL, Holmen SL, Colman H. IDH1 and IDH2 mutations in gliomas. *Current neurology and neuroscience reports*. 2013;13(5):345.
- [510] Xu P, Zhang A, Jiang R, et al. The Different Role of Notch1 and Notch2 in Astrocytic Gliomas. *PLoS ONE*. 2013;8(1).
- [511] Gallia GL, Rand V, Siu IM, et al. PIK3CA Gene Mutations in Pediatric and Adult Glioblastoma Multiforme. *Molecular Cancer Research*. 2006;4(10):709–714.

- [512] Kita D, Yonekawa Y, Weller M, et al. PIK3CA alterations in primary (de novo) and secondary glioblastomas. *Acta Neuropathologica*. 2007;113(3):295–302.
- [513] Ozawa T, Brennan CW, Wang L, et al. PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas. *Genes and Development*. 2010;24(19):2205–2218.
- [514] Chakravarty D, Pedraza AM, Cotari J, et al. EGFR and PDGFRA co-expression and heterodimerization in glioblastoma tumor sphere lines. *Scientific Reports*. 2017;7(1):9043.
- [515] Quayle SN, Lee JY, Cheung LWT, et al. Somatic Mutations of PIK3R1 Promote Gliomagenesis. *PLoS ONE*. 2012;7(11).
- [516] Wang L, He S, Yuan J, et al. Oncogenic role of SOX9 expression in human malignant glioma. *Medical Oncology*. 2012;29(5):3484–3490.
- [517] Gao J, Zhang JY, Li YH, et al. Decreased expression of SOX9 indicates a better prognosis and inhibits the growth of glioma cells by inducing cell cycle arrest. *International Journal of Clinical and Experimental Pathology*. 2015;8(9):10130–10138.
- [518] Labreche K, Simeonova I, Kamoun A, et al. TCF12 is mutated in anaplastic oligodendrogloma. *Nature Communications*. 2015;6:7207.
- [519] Cancer Genom Atlas. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487(7407):330–337.
- [520] Berg M, Danielsen SA, Ahlquist T, et al. DNA sequence profiles of the colorectal cancer critical gene set KRAS-BRAF-PIK3CA-PTEN-TP53 related to age at disease onset. *PloS one*. 2010;5(11):e13978.
- [521] De Roock W, Claes B, Bernasconi D, et al. Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: A retrospective consortium analysis. *The Lancet Oncology*. 2010;11(8):753–762.
- [522] Mao C, Wu XY, Yang ZY, et al. Concordant analysis of KRAS, BRAF, PIK3CA mutations, and PTEN expression between primary colorectal cancer and matched metastases. *Scientific Reports*. 2015;5(1):8065.
- [523] Zhang X, Nagahara H, Mimori K, et al. Mutations of epidermal growth factor receptor in colon cancer indicate susceptibility or resistance to gefitinib. *Oncology reports*. 2008;19(6):1541–4.
- [524] Oh BY, Lee RA, Chung SS, et al. Epidermal growth factor receptor mutations in colorectal cancer patients. *Journal of the Korean Society of Coloproctology*. 2011;27(3):127–32.
- [525] Gross ME, Zorbas MA, Danels YJ, et al. Cellular Growth Response to Epidermal Growth Factor in Colon Carcinoma Cells with an Amplified Epidermal Growth Factor Receptor Derived from a Familial Adenomatous Polyposis Patient. *Cancer Research*. 1991;51(5):1452–1459.
- [526] Bos JL, Fearon ER, Hamilton SR, et al. Prevalence of ras gene mutations in human colorectal cancers. *Nature*. 1987;327(6120):293–7.
- [527] Hao Y, Samuels Y, Li Q, et al. Oncogenic PIK3CA mutations reprogram glutamine metabolism in colorectal cancer. *Nature Communications*. 2016;7:11971.
- [528] Roper J, Hung KE. Molecular Mechanisms of Colorectal Carcinogenesis. In: *Molecular Pathogenesis of Colorectal Cancer*. New York, NY: Springer New York; 2013. p. 25–65.
- [529] Ngeow J, Heald B, Rybicki LA, et al. Prevalence of germline PTEN, BMPR1A, SMAD4, STK11, and ENG mutations in patients with moderate-load colorectal polyps. *Gastroenterology*. 2013;144(7).

- [530] Molinari F, Frattini M. Functions and Regulation of the PTEN Gene in Colorectal Cancer. *Frontiers in oncology*. 2013;3:326.
- [531] Miyaki M, Iijima T, Konishi M, et al. Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis. *Oncogene*. 1999;18(20):3098–3103.
- [532] Lü B, Fang Y, Xu J, et al. Analysis of SOX9 expression in colorectal cancer. *American journal of clinical pathology*. 2008;130(6):897–904.
- [533] Folsom AR, Pankow JS, Peacock JM, et al. Variation in TCF7L2 and increased risk of colon cancer: the Atherosclerosis Risk in Communities (ARIC) Study. *Diabetes care*. 2008;31(5):905–9.
- [534] Xu Y, Pasche B. TGF- β signaling alterations and susceptibility to colorectal cancer. *Human Molecular Genetics*. 2007;16(1):R14–20.
- [535] Xing M. Molecular pathogenesis and mechanisms of thyroid cancer. *Nature Reviews Cancer*. 2013;13(3):184–199.
- [536] Agrawal N, Akbani R, Aksoy BA, et al. Integrated Genomic Characterization of Papillary Thyroid Carcinoma. *Cell*. 2014;159(3):676–690.
- [537] Landa I, Ibrahimasic T, Boucail L, et al. Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. *Journal of Clinical Investigation*. 2016;126(3):1052–1066.
- [538] Kimura ET, Nikiforova MN, Zhu Z, et al. High prevalence of BRAF mutations in thyroid cancer: Genetic evidence for constitutive activation of the RET/PTC-RAS-BRAF signaling pathway in papillary thyroid carcinoma. *Cancer Research*. 2003;63(7):1454–1457.
- [539] Yoo SK, Lee S, Kim SJ, et al. Comprehensive Analysis of the Transcriptional and Mutational Landscape of Follicular and Papillary Thyroid Cancers. *PLoS Genetics*. 2016;12(8):e1006239.
- [540] Howell GM, Hodak SP, Yip L. RAS mutations in thyroid cancer. *The Oncologist*. 2013;18(8):926–932.
- [541] Nagy R, Ganapathi S, Comeras I, et al. Frequency of germline PTEN mutations in differentiated thyroid cancer. *Thyroid : official journal of the American Thyroid Association*. 2011;21(5):505–510.

Appendix

Table S1: Oncogenes, tumor suppressor genes, and housekeeping genes used in the analysis

Oncogenes					
ABL1	ABL2	AKT1	AKT2	ATF1	BCL11A
BCL2	BCL3	BCL6	BCR	BRAF	CARD11
CBLB	CBLC	CCND1	CCND2	CCND3	CDX2
CTNNB1	DDB2	DDIT3	DDX6	DEK	EGFR
ELK4	ERBB2	ETV4	ETV6	EWSR1	FEV
FGFR1	FGFR1OP	FGFR2	FUS	GOLGA5	GOPC
HMGA1	HMGA2	HRAS	IRF4	JUN	KIT
KMT2A	KRAS	LCK	LMO2	MAF	MAFB
MAML2	MDM2	MECOM	MET	MITF	MPL
MYB	MYC	MYCL	MYCN	NCOA4	NFKB2
NRAS	NTRK1	NUP214	PAX8	PDGFB	PIK3CA
PIM1	PLAG1	PPARG	PTPN11	RAF1	REL
RET	ROS1	SMO	SS18	TCL1A	TET2
TFG	TLX1	TPR	USP6		
Tumor suppressor genes					
APC	ARHGEF12	ATM	BCL11B	BLM	BMPR1A
BRCA1	BRCA2	CARS	CBFA2T3	CDH1	CDH11
CDK6	CDKN2C	CEBPA	CHEK2	CREB1	CREBBP
CYLD	DDX5	EXT1	EXT2	FBXW7	FH
FLT3	FOXP1	GPC3	IDH1	IL2	JAK2
MAP2K4	MDM4	MEN1	MLH1	MSH2	NF1
NF2	NOTCH1	NPM1	NR4A3	NUP98	PALB2
PML	PTEN	RB1	RUNX1	SDHB	SDHD
SMARCA4	SMARCB1	SOCS1	STK11	SUFU	SUZ12
SYK	TCF3	TNFAIP3	TP53	TSC1	TSC2
VHL	WRN	WT1			
Housekeeping genes					
C1orf43	CHMP2A	EMC7	GPI	PSMB2	PSMB4
RAB7A	REEP5	SNRPD3	VCP	VPS29	

Table S2: Stem cell differentiation marker genes

Neural Progenitor Cells (NPC)			
ABCG2 [403]	ASCL1 [404]	BMI1 [405]	CD133 [406]
CXCR4 [348]	FOXA2 [407, 408]	FOXO1 [409]	FZD9 [410]
GAP43 [411, 412]	GFAP [413, 414]	GLUT1 [415]	HES1 [416, 417]
MAP2 [408]	MSI1 [418]	NES [419]	NEUROD1 [420]
NFIX [421]	NOTCH1 [422, 423]	NTN1 [424]	OTX2 [425]
PAX3 [426, 427]	PAX5 [427]	PAX6 [408, 427, 428]	PAX7 [427]
PAX8 [427]	S100B [429]	SMARCA4 [430]	SOX1 [428, 431]
SOX11 [432]	SOX2 [433, 434]	SOX3 [435]	SOX4 [432]
SOX9 [436]	SYP [408]	TCF12 [437]	VIM [438]
Mesenchymal Stem Cells (MSC)			
ALCAM [439]	ANPEP [440]	CD44 [440, 441]	CD70 [347]
DLK1 [442, 443]	ENG [347, 441]	ETV1 [444]	ETV5 [444]
FOXP1 [444]	GATA4 [445]	GATA6 [444]	HMGA2 [444]
ITGA4 [446]	ITGB1 [441]	MYOD1 [445]	NANOG [440, 447]
NCAM1 [440]	NT5E [440]	OCT4 [447]	PDGFRA [447, 448]
POU5F1 [449, 450]	PPARG [445]	RUNX2 [440, 445]	SIM2 [444]
SOX11 [444]	SOX2 [449–451]	SOX4 [452]	SOX9 [440, 445]
SPARC [440, 453]	THY1 [347, 441]	VIM [454]	
Trophoblast Stem Cells (TSC)			
ARID3A [455]	BMP4 [456]	CD9 [457]	CDH1 [458]
CDX1 [459]	CDX2 [456, 459–461]	CGA [459, 462, 463]	CGB [457, 462, 463]
ELF5 [458, 464]	EOMES [456, 458, 462, 465]	ESRRB [458]	ETS2 [456]
FGF4 [466]	FGFR2 [456, 458, 467]	FURIN [456]	GATA2 [459, 463]
GATA3 [456]	GCM1 [463]	HAND1 [459]	ID2 [468, 469]
IGFBP3 [463]	KRT7 [457, 463]	MMP9 [463]	MSX2 [463, 470]
SMARCA4 [465]	SOX2 [458, 463]	TEAD4 [471]	TFAP2C [456]
TFAP2C [456, 458, 465]			

Table S3: Cancer marker genes

Chronic Lymphocytic Leukemia (CLL)	
ARID1A (2.41 [387, 472])	ATM (9 [473], 4.14 [472, 474])
BCOR (1.72 [475])	BIRC3 (2.5 [476], 19.7 [477])
BRAF (3.7 [478], 2.8 [387, 479])	CHD2 (5.3 [480], 4.8 [478, 481])
CXCR4 (OE [482–485])	DDX3X (1.03 [472], 2.4 [473], 1.72 [486])
EGR2 (3.8 [478, 487])	FBXW7 (1.03 [472], 2.5 [473, 488])
IRF4 (1.5 [387, 489])	MYD88 (2.2 [476], 4 [490], 8 [473], 5.17 [472, 488])
PAX5 [387]	NOTCH1 (3.1 [472], 4 [473], 8 [476], 11.3 [487, 488, 491])
SAMHD1 (11 [492, 493])	SF3B1 (11.2 [476], 15 [473], 7.93 [472, 473, 488])
SYK (OE [494–496])	TP53 (10.4 [476], 15 [473], 7.1 [488], 8.62 [472])
XPO1 (2.76 [472], 3.4 [488])	ZAP70 (OE [497, 498])
Lower Grade Glioma (LGG)	
ARID1A (11 [499], 5.92 [472])	ARID1B (11 [499], 2.37 [472])
ATRX (42.6 [500, 501])	BRAF (15 [502] , 1.85 [503, 504])
CIC (20.12 [501])	EGFR (OE [505], A [506], 23.22 [503], 4.14 [472])
FUBP1 (10.65 [507])	IDH1 (77.51 [508, 509])
IDH2 (3.55 [509])	NF1 (5.92 [472, 503, 508])
NOTCH1 (7.69 [472] , OE [510])	PIK3CA (6.51 [472], 10.03 [503, 508, 511, 512])
PDGFRA [513, 514]	PIK3R1 (5.92 [472, 503, 508, 515])
PTEN (4.14 [472], 30.34 [503, 508])	RB1 (1.78 [472, 503, 508])
SOX9 (OE [516, 517])	TCF12 (3.55 [518])
TP53 (50.89 [472], 30.61 [503, 508])	
Colorectal Cancer (CRC)	
APC (79.04 [519])	BRAF (16 [520], 4.7 [521], 3 [519, 522])
EGFR (12-22 [523–525])	KRAS (40 [521], 43 [519, 520, 522, 526])
FBXW7 (10 [519])	PIK3CA (14.5 [521], 15 [519, 520, 522, 527])
SMAD2 (3.4 [519, 528])	PTEN (14 [520], 4 [519, 521, 522, 529, 530])
SMAD3 (4.3 [519, 528])	SMAD4 (8.6 [519, 528, 531])
SOX9 (3.49 [472], 4 [519, 532])	TCF7L2 (9.17 [472, 533], 12 [519])
TGFBTR2 (3.49 [472], 2 [519, 534])	TP53 (59 [519])
Papillary Thyroid Cancer (PTC)	
AKT1 (15 [535])	ALK (10 [535])
ARID1B (1 [536] [537])	BRAF (35.8 [538], 56.52 [472])
CTNNB1 (25 [535])	EGFR (5 [535])
EIF1AX (1.5 [536, 539])	HRAS (20-40 [535, 540])
KMT2C (1 [536] [537])	KRAS (20-40 [535, 540])
NDUFA13 (15 [535])	NRAS (20-40 [535], 8.07 [472, 540])
PIK3CA (1–2 [535])	PTEN (4.8 [535, 541])
TG (2.7 [536])	TP53 (25 [535])
ZFHX3 (1.7 [536])	

Numbers in the brackets represent the expression (OE stands for overexpression) or mutation (A: amplification, number: percentage of mutation rate) of the gene.

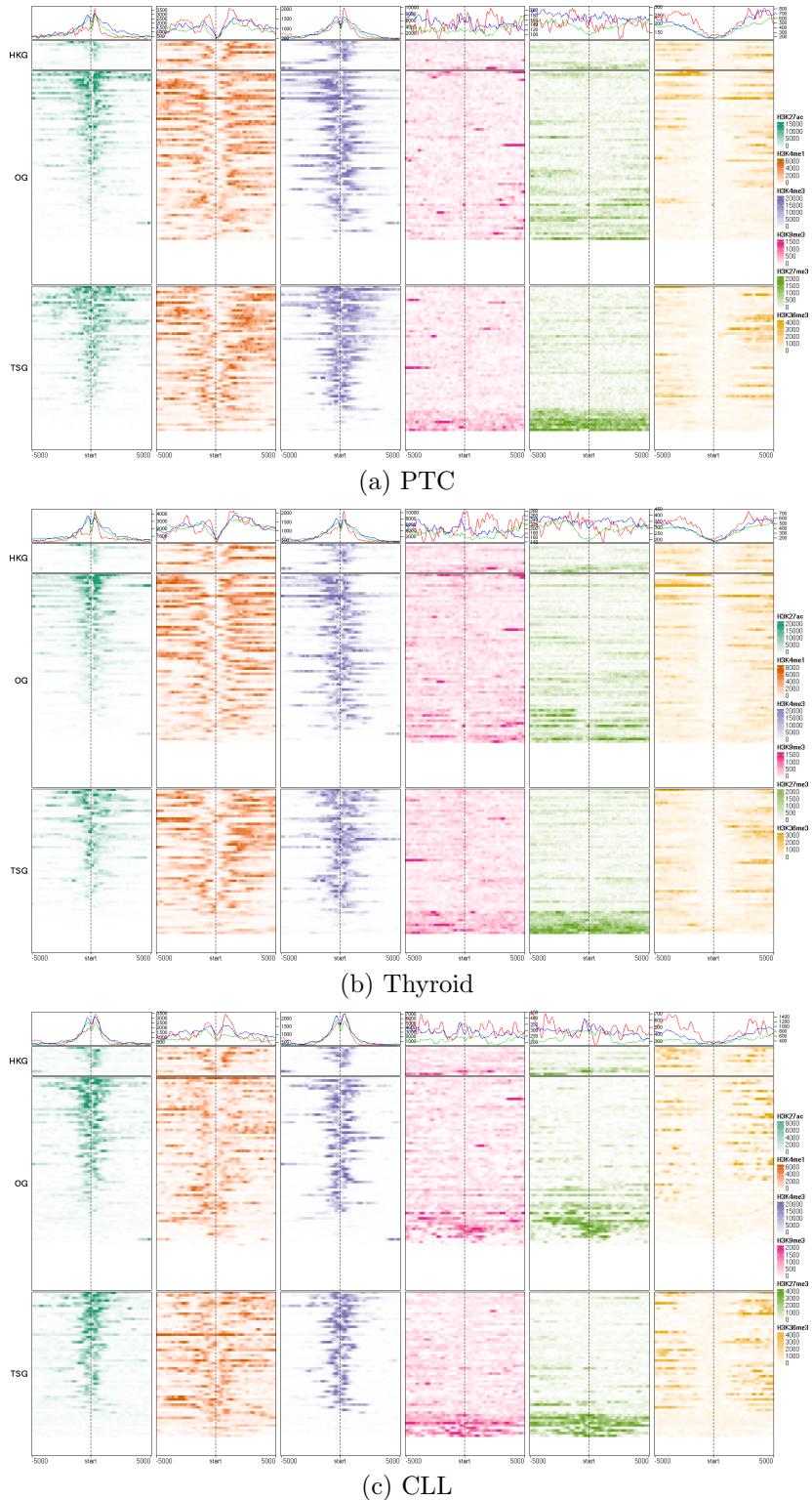
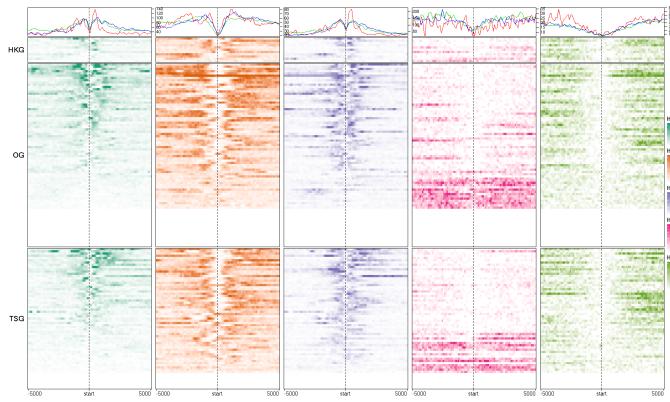
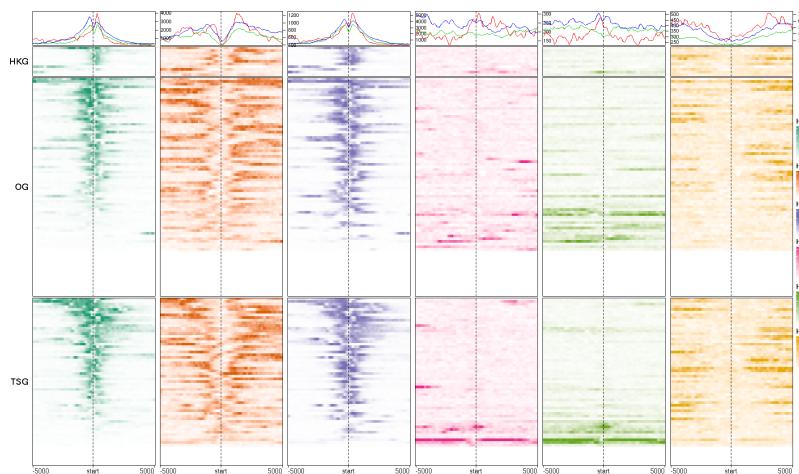


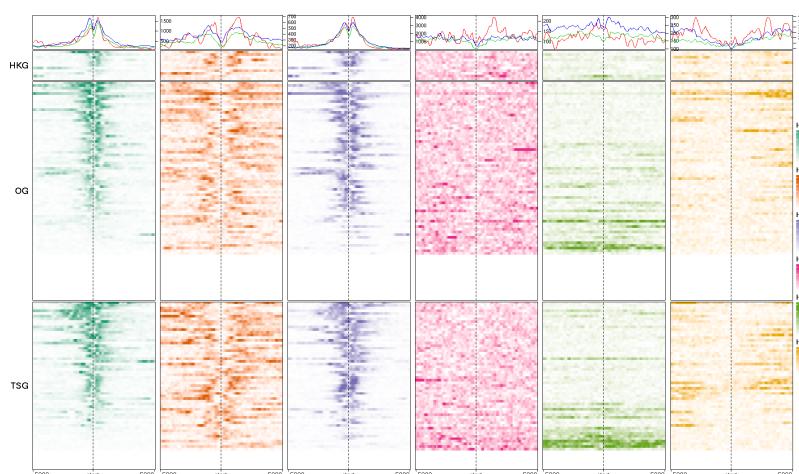
Figure S1: Continued on next page.



(d) B cell

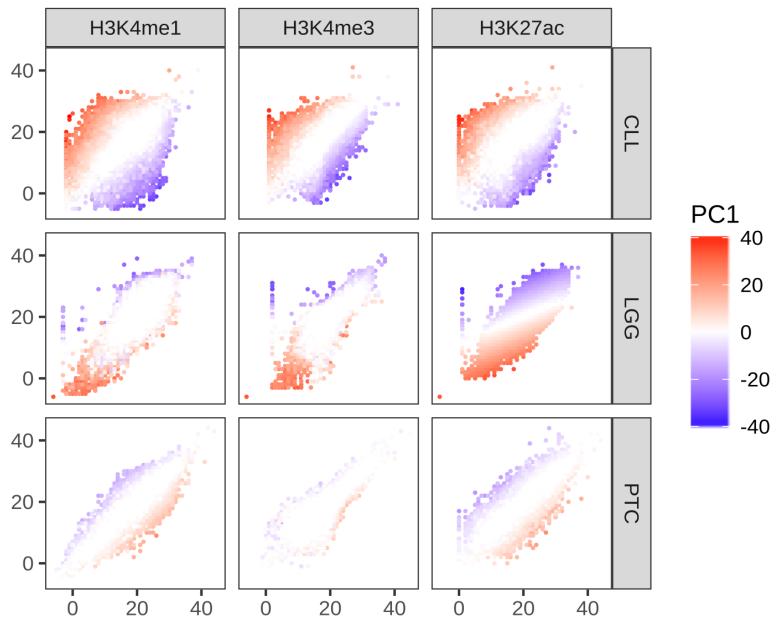


(e) LGG

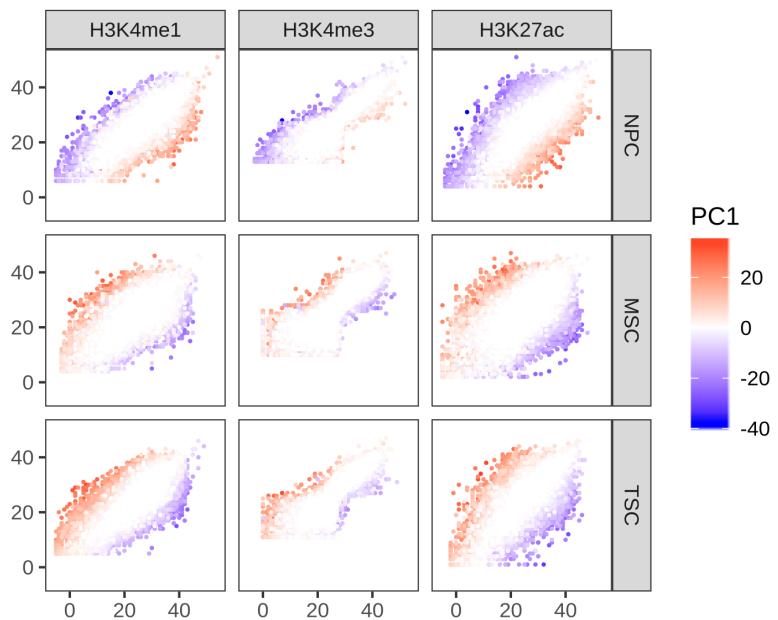


(f) Normal brain

Figure S1: Histone mark signals around oncogenes (OG), tumor suppressor genes (TSG), housekeeping genes (HKG).

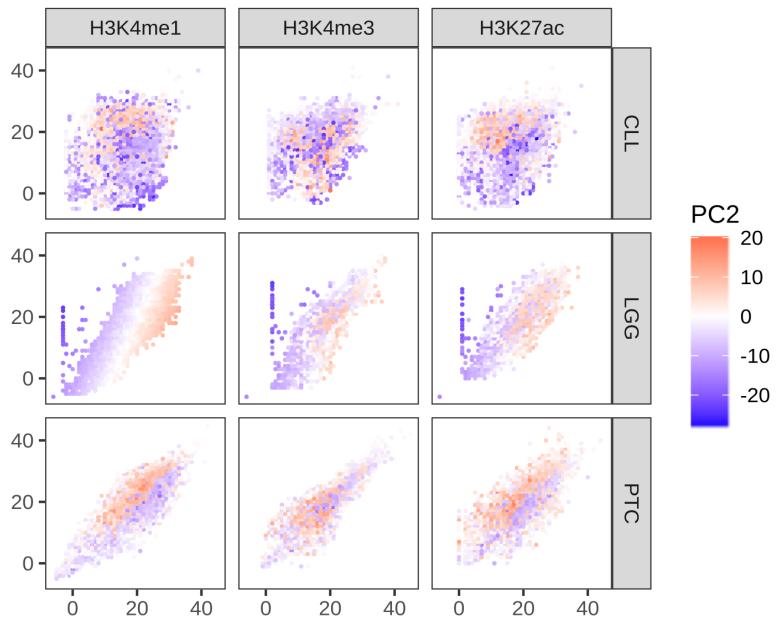


(a)

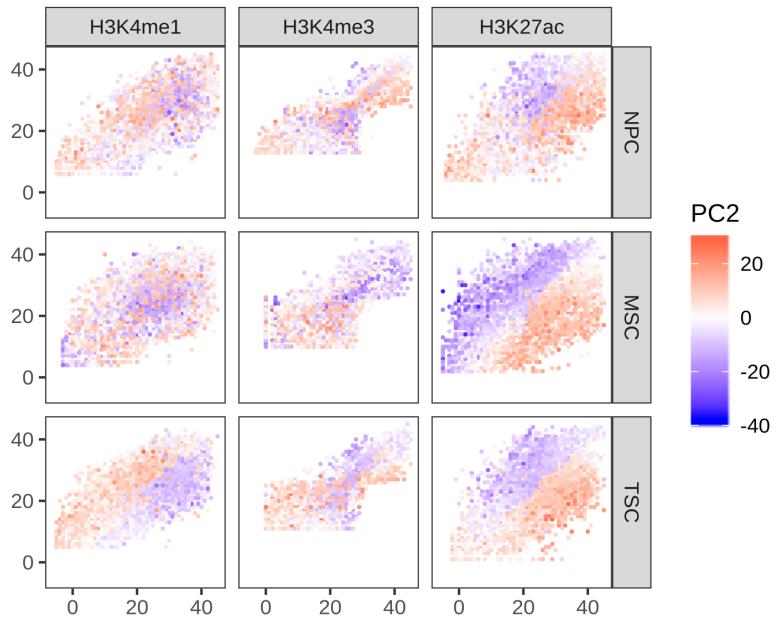


(b)

Figure S4: Continued on the next page.



(c)



(d)

Figure S4: dPC1 and dPC2 for three histone marks. The mean log intensities for every genomic locus in group one and group two from six test cases are plotted against each other, and the nodes are colored according to their dPC scales.