UNIVERSITÀ DEGLI STUDI
DI MILANO

European School of Molecular Medicine

PhD in Computational Biology

Diet-specific epigenetic signature revealed by H3K4me3 and H3K27me3 data analysis in C57BL6 mice

PhD candidate: Anna Russo

Supervisor: Pier Giuseppe Pelicci
Added supervisors: Lucilla Luzi
Marco Giorgio
External supervisor: Martin Vingron
Internal supervisor: Cesare Furlanello

Academic Year 2015/2016

To my nieces and nephew
and to my sister and brothers
for making them.
Please make more!

“Research is what I'm doing when
I don't know what I'm doing”

Wernher von Braun

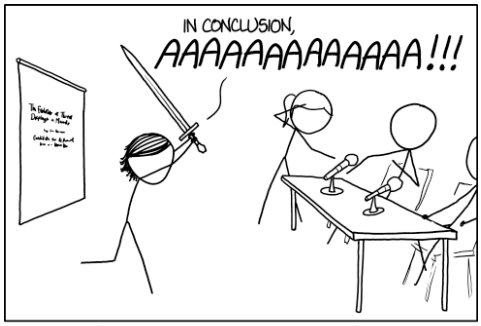


Table of contents

List of abbreviations

The following table describes the significance of various abbreviations and acronyms used throughout the thesis.

Abbreviation	Meaning
2DG	2-deoxy-D-glucose
ANOVA	Analysis of variance test
bp	Base pairs
BMI	Body Mass Index
WT	Wild type
ChIP	Chromatin immunoprecipitation
CR	Calorie Restriction
DB	Differentially bound
HDM	Histone demethylase
FC	Fold Change
FFPE	Paraffin embedded
FRiP	Fraction of reads in peaks
GO	Gene Ontology
H3K27me3	Trimethylation of Lysine 27 on histone 3
H3K4me3	Trimethylation of Lysine 4 on histone 3
HAT	Histone Acetyl transferase
HDAC	Histone deacetylase
HF	High Fat
HMT	Histone Methyltransferase
IDF	International Diabetes Federation
Kb	Kilo bases
KEGG	Kyoto Encyclopedia of Genes and Genomes
MetS	Metabolic syndrome
NK	Natural Killer
NRSF/REST	Neuron Restrictive Silencer Factor
PAT-ChIP	Patology Tissue ChIP
PCA	Principal Component
PCR	Polymerase Chain reaction
PPAR	Peroxisome proliferator-activated receptors
qCML	quantile-adjusted conditional maximum likelihood method
RE1	Repressor Element 1
RPKM	reads per kilobase per million
SCN	Super Chiasmatic Nucleus
SD	Standard diet
Std dev	Standard deviation
T2DM	Type II diabetes mellitus
TMM	Trimmed mean of M-values
TSS	Transcription Start Site
UTR	Untranslated region

List of Figures

Introduction

Results

Abstract

Increasing evidences demonstrate that adapting to different environmental conditions is mediated by epigenetic changes, which can participate in cellular processes. In particular, the adaptation to the different caloric intakes is of great relevance as it is crucial for the organism's fitness. Moreover, the phenotypic remodeling induced by different diets determine the susceptibility to life-threatening diseases. For example, refined sugar, fat and meat enriched diet, typical of Western countries, is thought to be responsible for about 30-35% of cancer cases, in addition to increased incidence of type 2 diabetes and cardiovascular diseases. On the other hand, caloric restriction has been shown to be the most powerful way to prolong lifespan and reduce cancer incidence in different experimental models.

Based on the hypothesis that epigenetic changes represents the mechanistic link between diet and disease risk, the aim of this work is to investigate chromatin modifications induced by different diets in murine models to identify specific epigenetic profiles associated with fat enriched diets and caloric restriction.

For this purpose, 8 weeks old C57Bl/6 female mice were divided in three groups and fed for 10 months with 3 different diets: Standard laboratory mouse Diet, Calorie Restriction without malnutrition, High Fat Diet.

Then, livers were extracted and investigated by chromatin immunoprecipitation (anti-H3K4me3, anti-H3K27me3) and transcriptomic approach for gene expression analysis.

Despite the presence of moderate technical and biological variability, data analysis demonstrated that specific epigenetic profiles were associated to different diets. In

particular, the distribution and frequency of H3K4me3 enabled the clustering of samples by diet-group.

Moreover, functional annotation of genes showing an increased signal of H3K4me3 for HF or CR respect to SD on their promoter regions, resulted in significantly enriched “Type II diabetes mellitus”, for which obesity represents a critical risk factor, and “Circadian Rhythm” pathways, whose known to affect longevity.

At mechanistic level, two DNA motifs related to the transcription and chromatin regulators ZSCAN4 and REST/NRSF were found enriched in correspondence of the regulative regions of the genes of the aforementioned pathways, suggesting these factors mediate the effects of diet on chromatin and gene expression.

1. Introduction

To sustain our body's energy needs and functions, we need food: each single cell requires a constant supply of calories and nutrients. Moreover, eating and food are also associated to other, yet important, needs: we use to eat to bond with loved ones, friends, family or coworkers; food inspires a sense of community, it is used as a source of comfort/reward or as a way to reduce stress in difficult moments of our life.

For these reasons, we have witnessed to a rising interest and curiosity towards food habits, the impact and the role that food have on the quality of our life and, especially, on our health.

Increasing number of studies show a correlation between Western-style diet (rich in fats, carbohydrates, proteins) and incidence or worsening of malignancies. The number of obese adults and, especially, children is constantly increasing worldwide, as reported by the World Health Organization (**WHO, Obesity and Overweight, 2015**), rising concerns in the healthcare systems and awareness in the population. In the last decades, this frame promoted the flourishing of new medical sciences that, placed side by side to biochemistry, investigate the impact that food has on our organism at the molecular level adding a new perspective.

This is the case, for example, of *Nutrigenomics* (studying the effects of foods and food constituents on gene expression, **Müller and Kersten, 2003**) and *Nutrigenetics* (studying the effect of genetic variations on the interaction between diet and health with implications to susceptible subgroups, **Mutch et al, 2005**).

In particular, a recently born discipline is *Nutri-epigenomics,* focused on the effects of food nutrients on human health exerted through *epigenetic modifications* (cellular and physiological variations not caused by changes in DNA sequences) (**Gallou-Kabani et al, 2007**).

In this thesis I report the results of our studies focused on investigating the impact of different diets on the mouse epigenome, starting from the hypothesis that food adaptation entails reprogramming of different cell functions, which might be maintained and/or executed through changes in chromatin.

The *aims* of our studies were to characterize the epigenetic changes induced by nutritional regimens that have been associated to either a higher risk of developing cancer or cardiovascular diseases *(high fat diet, HFD*), or to a protective effect against aging and diseases (*caloric restriction, CR*), and to possibly identify diet-specific epigenetic signatures. To this end, Next-Generation Sequencing (NGS) technologies were combined with Chromatin Immunoprecipitation (ChIP-seq) and transcriptional expression analyses (RNA-seq) of a big collection of *in vivo* samples of *liver* tissues.

In this first chapter we will provide fundamental concepts related to diet, metabolism, epigenetics and Next Generation Sequencing data creation and analyses, in order to proper illustrate the results reported in Chapter 3 and discussed in Chapter 4; all the materials and methods used in this study are described in Chapter 2.

1.1. Diet, metabolism and disease

"*We are what we eat*", it is often said, but what does this sentence really mean? From a general point of view, we can easily state that food affects every aspect of our life (mood, body functions, relationships), and that some kinds of food are even considered symbols of entire countries (as for example, pasta and pizza for Italy). In this perspective, we can say that food creates identity, defining us with respect to ourselves and the others.

From a scientific point of view it is known that, when it comes to food, bad habits often produce health problems: a diet rich in fat can be the starting point for metabolic disorders, overweight and obesity, all risk factors for cardiovascular diseases and different type of cancers.

In particular, it has been estimated by **Anand et al, 2008** that 90%-95% of US cancer cases are due to environmental and lifestyle factors like smoke, exposure to radiations and/or pollutants, alcohol consumption and, of these, 30-35% are related to unhealthy diet. Conditions like these have a huge impact, not only on the individual health status, but also on the overall healthcare system and on the whole society. On the contrary**,** in animal models, a low caloric diet is associated to a reduced cancer incidence, and moderate calorie restriction (without malnutrition) has emerged as the most potent dietary intervention for preventing cancer and prolonging life span (**Hursting et al, 2009**)**.** Although both these dietary conditions have been extensively studied, the mechanisms linking diet, metabolism and disease development/prevention are still unclear.

In this section we will first briefly summarize i) basic concepts of cellular metabolism and ii) roles of liver in controlling energy transformation and utilization, and then introduce iii) actual knowledge about the impact of high fat and calorie restriction diets on health and iv) how our biological internal clock, the circadian clock, is linked to the metabolic processes of the entire body, due to its relevance in our study results.

1.1.1. Nutrients metabolism

Living organisms need energy on a daily basis and food represents the fuel to this engine. The biochemical reactions needed to obtain and use energy at cellular level define the general process called *metabolism* (from the Greek, μεταβολή, metabolē, which means “transformation, change”).

Carbohydrates, lipids, and proteins represent the principal energetic components of diet. After digestion in the gastrointestinal tract and successive absorption, through the bloodstream, they reach every tissue and cell of the body, where their chemical energy content are further transformed and utilized. Monosaccharides (mainly glucose) - from carbohydrates -, monoacylglycerol and long-chain fatty acids - from lipids - and small peptides and amino acids - from proteins - are the ultimate compounds of the digestion activity. The energy released by breaking chemical bonds of these substrates is then transferred to high-energy compounds, which work as repositories and energy carriers in cells. One of these keys molecules is the adenosine triphosphate (ATP).

ATP is produced, in mitochondria, during the tricarboxylic acid (TCA) cycle (also known as Krebs cycle) - the central metabolic pathway where all products of nutrients' degradation (glycolysis, fatty acid oxidation and transamination/deamination of some amino acids) converge (Figure 1.1) - and

mainly during oxidative phosphorylation. All these metabolic energy-transducing events in this process are made possible by oxidation-reduction reactions: the electrons removed by the oxidation of nutrient molecules are transferred to two major electron carrier coenzymes, nicotinamide adenine dinucleotide (NAD^+) and flavin adenine dinucleotide (FAD), then converted to their reduced forms, NADH and $FADH_2$. These reduced electron carriers are themselves oxidized via the electron transport system (ETS) - a modular set of protein complexes which constitutes a chain of electron accepting/donating factors – which in turn allows the distribution of the free energy between the reduced coenzymes and the O_2 resulting in ATP synthesis.

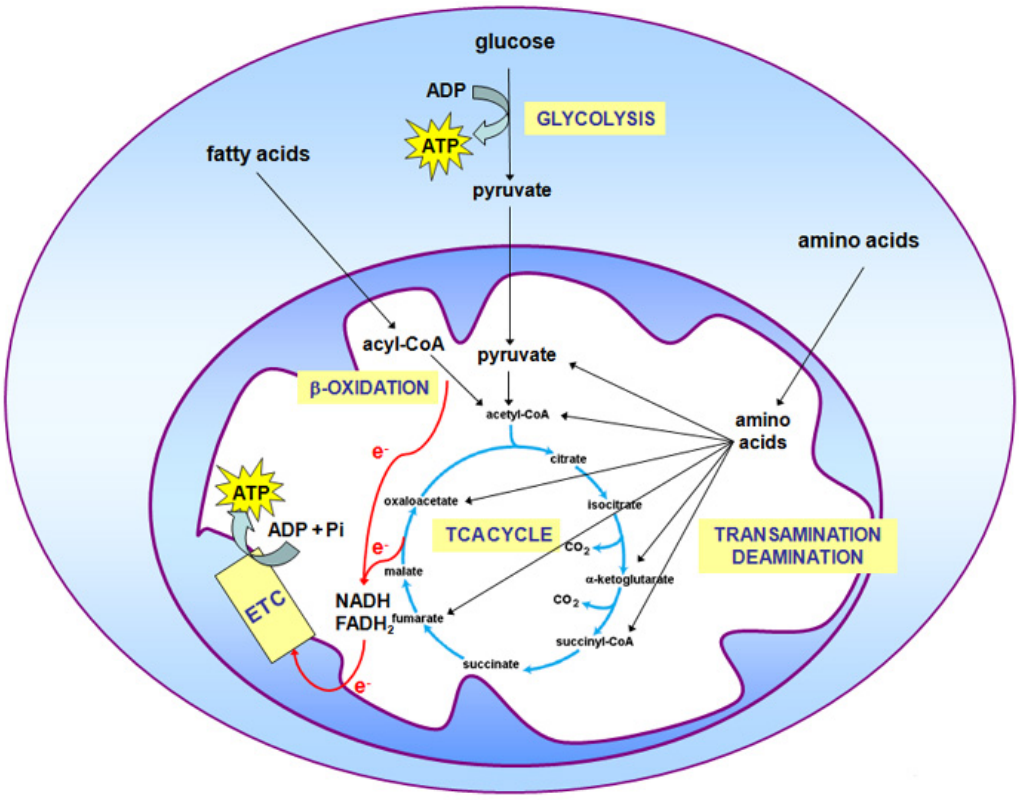


Figure 1.1 Metabolism activity scheme

Schematic representation of energy metabolism relationship: the degradation of lipids, proteins and carbohydrates produces fatty acid, amino acids and pyruvate, respectively and they all converge to TCA cycle. The electrons are transported from reduced coenzymes to O_2 in the electron transport system, resulting in ATP synthesis.

(Adapted from El Bacha et al, Nature Education, 2010)

Different cells may exhibit specific and unique metabolic profile, not only in a fixed tissue-specific context, but also according to different physiological conditions, such as the fed or fasting states (Figure 1.2). For example red blood cells, not having mitochondria, use only glucose as source of energy, and convert it into lactate; the brain relies on glucose and ketone bodies that are generated in the liver in case of starvation; adipose tissues uses fatty acids and glucose, while muscle cells use also amino acids; the liver uses fatty acid oxidation as energy sources **(El Bacha et al, 2010; Berg et al, 2002)**.

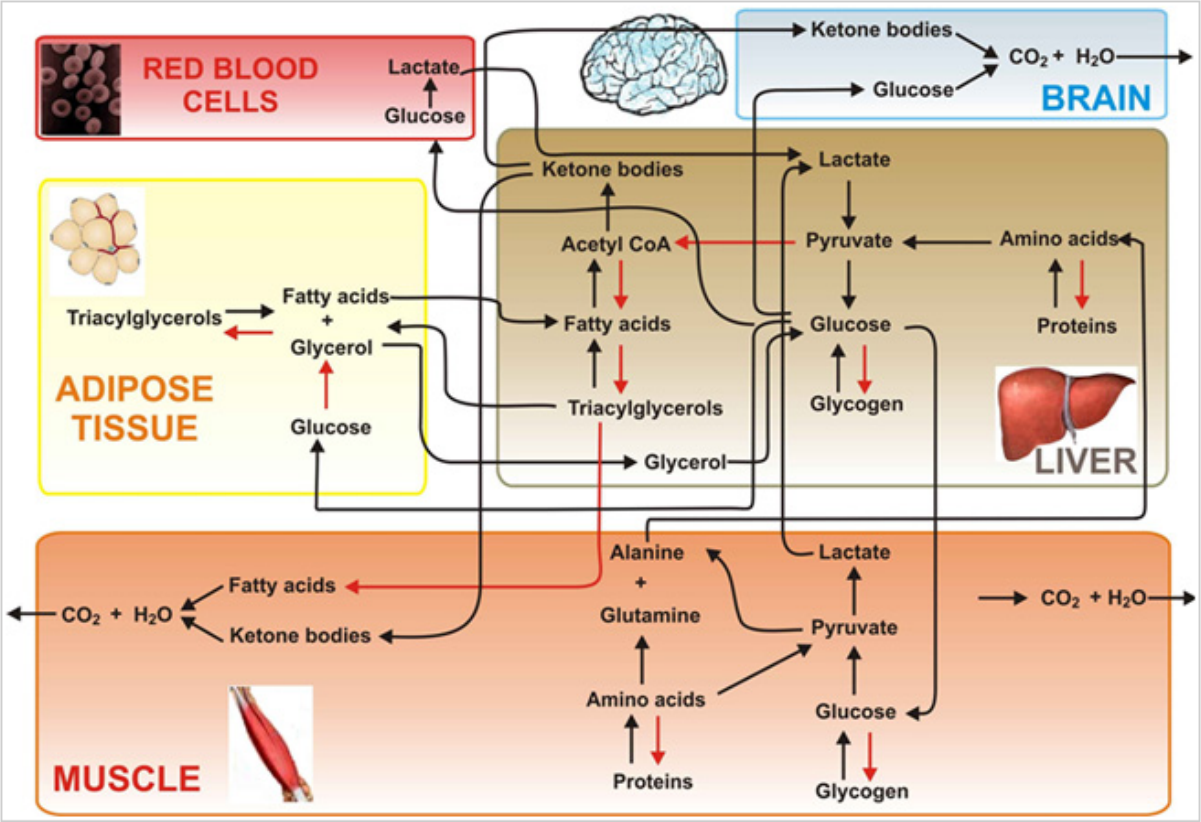


Figure 1.2 Different cell types manage differently to satisfy energetic needs

Red blood cells rely on glucose for energy and convert glucose to lactate. The brain uses glucose and ketone bodies for energy. The liver primarily uses fatty acid oxidation, while muscle cells use fatty acid, glucose and amino as energy sources.

(Adapted from El Bacha et al, Nature Education, 2010)

The comprehension of metabolic pathways therefore can be achieved, by and large, only considering all the integrative events, which contribute to energy regulations and their adaptation to various internal or environmental changes.

1.1.2. Liver anatomy and its role in metabolism

Hepatocytes, liver basic cells, have a major role in the synthesis of molecules utilized to sustain whole body *homeostasis* (the property of a system in which variables are regulated so that internal conditions remain stable and relatively constant), in converting molecules of one type to another, and in regulating energy balances. Compounds absorbed by the intestine are then processed by liver that, acting as a metabolic hub, provides fuel to muscles, brain and peripheral organs (**Berg et al, 2002**).

One the main function of the liver is to maintain normal the blood glucose levels for both short and long periods of time: hepatocytes employ many enzymes that alternatively switch on or off depending on fluctuations of blood glucose levels. For example, the excess of glucose entering in the blood after eating is taken up by liver and in particular by glycogen (*glycogenesis*), that later, when blood concentrations of glucose start to go down, will be depolymerized (*glycogenolysis*) to release glucose back into the blood for transport to all other tissues (cf. Figure 1.3).

If hepatic glycogen reserves end, as happens when an animal do not eat for many hours, hepatocytes activate other groups of enzymes that synthesize glucose from amino acids and non-hexose carbohydrates (*gluconeogenesis*).

This process is massively regulated by hormones, in particular, insulin and glucagon, that have opposing actions. Insulin levels rise in response to a meal, promoting nutrient storage and glycogenesis, whereas glucagon levels rise with fasting, promoting glycogenolysis and gluconeogenesis.

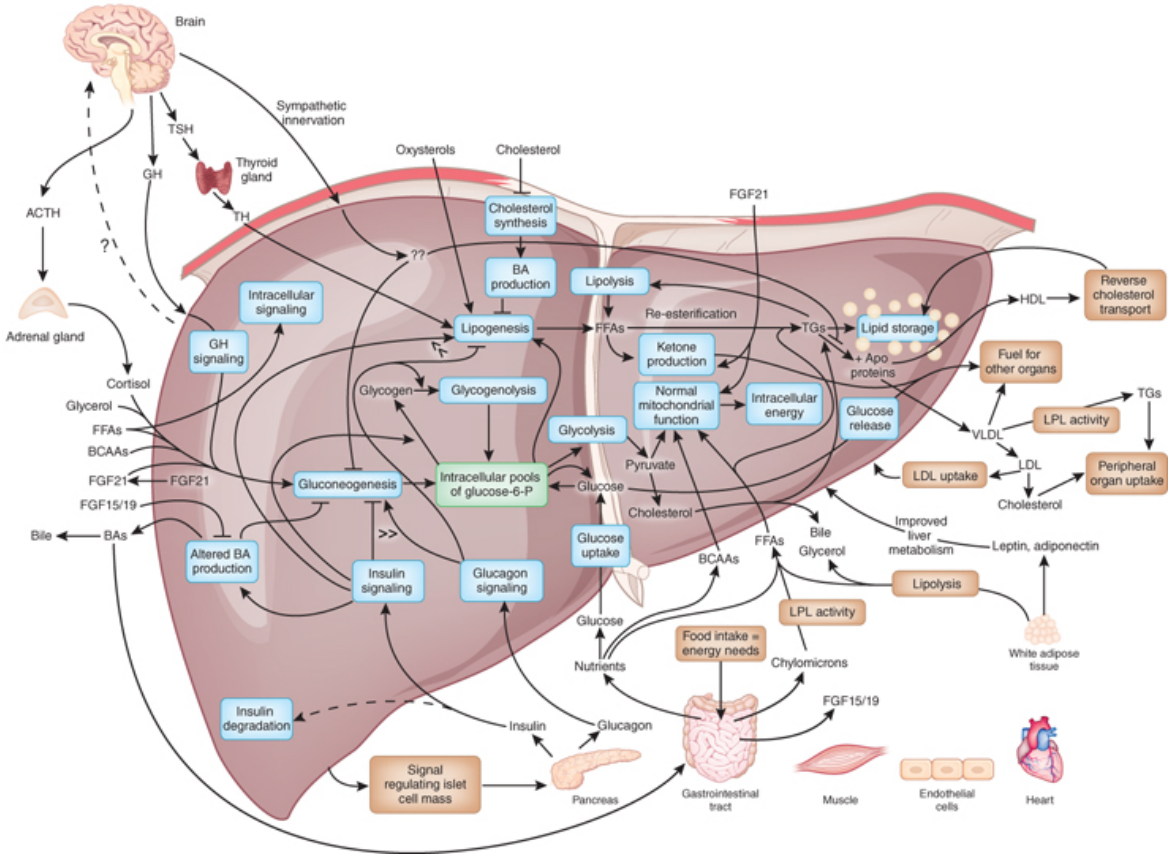


Figure 1.3 Metabolic homeostasis driven by the liver on the organ level
The liver integrates nutritional, neural and endocrine signals to store or mobilize nutrients, and to control carbohydrate, lipid and amino acid usage. Its main role is to maintain metabolic homeostasis.

(Adapted from Metabolic Syndrome eposter, Nature Medicine 2012)

Fatty acids in the blood passing through the liver are absorbed by hepatocytes and metabolized to produce energy in the form of ATP. Hepatocytes can also produce lipids like cholesterol, phospholipids, and lipoproteins that are used by other cells throughout the body. Much of the cholesterol produced by hepatocytes is excreted from the body as a component of bile. In addition, dietary proteins are broken down into their component amino acids by the digestive system and then being passed on to the hepatic portal vein. Amino acids entering the liver require metabolic processing before they can be used as an energy source. Hepatocytes first remove the amine groups of the amino acids and convert them into ammonia and eventually urea. Urea is less toxic than ammonia and can be excreted in urine

as a waste product of digestion. The remaining parts of the amino acids can be broken down into ATP or converted into new glucose molecules through the process of gluconeogenesis (**Lewis et al, 1997; Cherrington et al, 1999; Obici and Rossetti, 2003; Lin and Accili, 2011**).

1.1.3. High fat diet and (poor) health

More than 60 years ago, a first study by **Samuels et al, 1942** reported that rats, subjected to a regimen containing 70% energy as fat, became obese and showed higher basal and postprandial blood sugar values. Similar results were obtained with diets containing well above 30% energy for different animal models and diet lengths (**Budohoski et al, 1993; Harris et al, 1993**). Hyperglycemia and obesity induced by this so called *High Fat Diet* (or HFD) in rats, are also present in humans in a group of metabolic disorders called *metabolic syndrome* (MetS), that includes abdominal and visceral obesity, dyslipidemia, insulin resistance, hypertension and abnormal glucose metabolism. This syndrome is widely diffuse nowadays, spread worldwide and counting one-quarter of the world's adult population, as reported by the International Diabetes Federation (**IDF, 2015**). In turn, MetS is known to play a role in the development of cardiovascular diseases, diabetes mellitus (**Gami et al, 2007**) and a variety of tumor types (**Esposito et al, 2012**; **Giovannucci et al, 2007; Pais et al 2009; Aleksandrova et al, 2011**).

Moreover, Western-style diet, rich in fat, refined carbohydrates and animal proteins, is unanimously recognized as the main cause of overweight and obesity, which are not only related to metabolic syndrome, insulin resistance and cardiovascular risk, but are also major risk factors (as important as tobacco smoking) for the development of some types of tumour (breast, colorectal), as reviewed by **Berrino et al, 2006**. Notably, breast and colorectal cancer patients

with MetS have also increased risk of developing metastasis (**Shen et al, 2010**). Nevertheless, the molecular mechanism linking (hypercaloric) diet with disease risk is far from being clearly understood.

1.1.4. Calorie restriction, the anti-aging and health promoting effect

Since the beginning of the last century, moderate calorie restriction (CR) without malnutrition (defined as an experimental setting in which test animals receive a 30-70% less calories than *ad libitum*-fed controls) has emerged as the most potent dietary intervention for loss weight and preventing age-associated diseases (**McCay et al, 1935**).

In fact, a collection of studies show that CR extends lifespan in a variety of experimental models, like yeasts, worms, flies, spiders, rotifers, fish and rodents (**Chapman and Partridge, 1996; Fontana et al, 2010; Greer and Brunet, 2009; Kennedy et al, 2007; Mair and Dillin, 2008; Masoro, 2005; Weindruch et al, 1988**) and slows age-related chronic diseases. It is also known that CR reduces metabolic rate and oxidative stress, improves insulin sensitivity, and alters neuroendocrine and sympathetic nervous system function in animals (**Heilbronn et al, 2003**).

Moreover, as summarized by **Hursting et al, 2010**, calorie restriction in experimental tumour models inhibits cancer: **Mattison et al, 2012** and **Colman et al, 2009**, reported that rhesus monkeys, subjected to CR, showed a decreased risk of diabetes, neurological degeneration and cancer, **Harvie et al, 2012** and **Imayama et al, 2012** observed, in women fed with a CR regimen, a decreasing of inflammatory and endocrine markers that are associated with breast cancer risk, suggesting that CR beneficial effects on metabolism and chronic disease risk known for experimental models could actually apply also to human beings.

Even in this case, the mechanisms through which CR improves tumour suppression are still largely unclear.

1.1.5. The circadian clock and its connection with metabolism

It is easy to notice that feeding follows a certain rhythmicity: for instance, humans, that are daily organisms, feed during the day, while nocturnal organisms eat predominantly at night. In fact, food acts as an external stimulus for our internal biological clock that is called *circadian clock.*

In mammals the central circadian clock is located in the suprachiasmatic nuclei (SCN), a particular bilateral group of cells located in the anterior hypothalamus in the brain. This internal biological timer allows the daily coordination of biological and behavioural activities of an organism and, as suggested by the name (from Latin, *circa diem*, “about a day”), oscillates with a period of 24 hours regulating the day-night cycle and depends from external cues, like sunlight (review by **Welsh et al, 2010**).

It is important to notice that similar clock oscillators have been found in many tissues, such as the liver, intestine, heart, adipose tissue, retina and in various regions of the brain and that these oscillators are synchronized through both endogenous and external signals to regulate transcriptional activity throughout the day in a tissue-specific manner (**Balsalobre et al, 1998; Yamazaki et al, 2000; Yoo et al, 2004**). Moreover, the clock can be modified through environmental changes depending on the organism's ability to detect external time cues, like light. The circadian rhythms, driven by the circadian clocks, display three main characteristics:

The existence of an endogenous free-running period lasting approximately 24 hours. In animals active during daylight hours, in general it is a bit greater than 24 hours,while for nocturnal animals is shorter than 24 hours.

They have to be “entrainable”. meaning that it is possible to reset them through external stimuli (as, for example, light and heat). This process is called entrainment and it happens, for example, traveling across different time zones, since our biological clock needs to adjust to the local time.

They maintain circadian periodicity over a range of physiological temperatures. Since differences in thermal energy will affect the kinetics of all molecular processes in their cells of an organism, in order to keep track of time, the organism's circadian clock must maintain roughly a 24-hour periodicity despite the changing kinetics. This is known as temperature compensation.

Clock components are mostly transcriptional activators or repressors involved in the onset of two linked feedback loops: in the first one, CLOCK and BMAL1/ARNTL form a complex that, moving from cytoplasm to nucleus, starts transcription of target genes that are known as *period* genes (PER1, PER2, PER3) and *cryptochrome* genes (CRY1, CRY2); in the second one, PERs and CRYs form complexes that, travelling to nucleus, represses CLOCK:BMAL1 complex, thus blocking their own transcription (**Alberts et al, 2008**) as schematized in Figure 1.4.

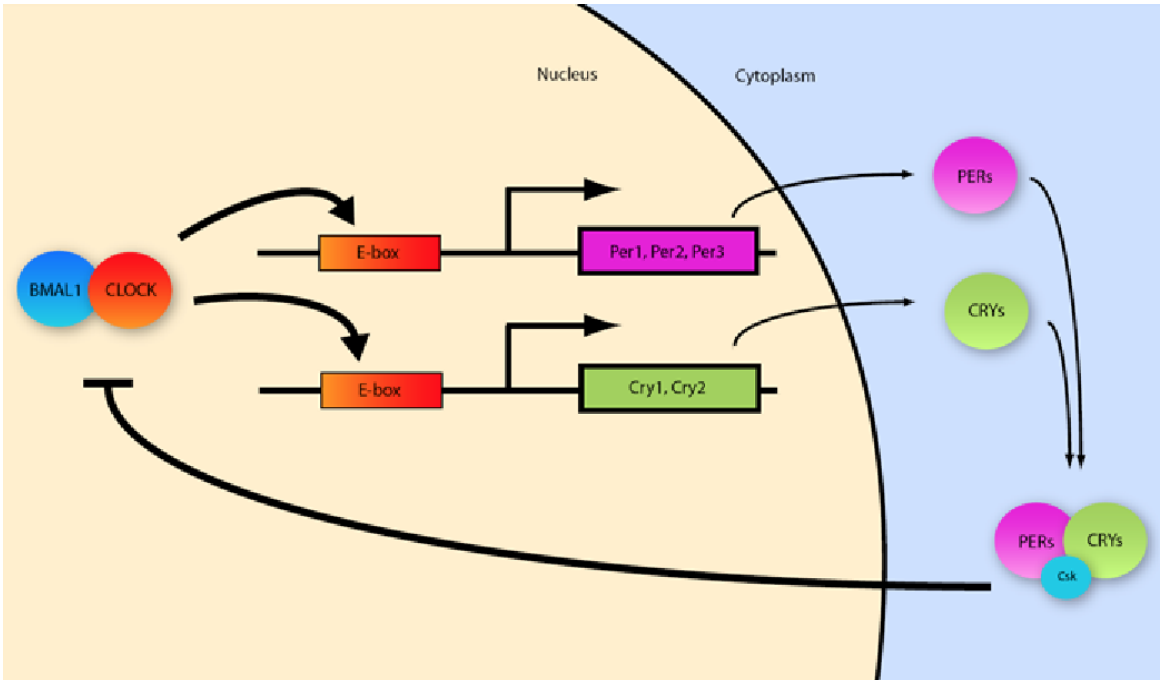


Figure 1.4 Feedback loops of the central circadian clock

BMAL1 and CLOCK heterodimerize in the nucleus, promoting the transcription of PER and CRY genes. The PERs and CRYs proteins from the cytoplasm, enter in the nucleus to repress BMAL1 and CLOCK activity.

(Adapted from Bernard et al, PLoS Comput. Biol. 2007)

In the last decades, several studies supported a unique role for circadian rhythm in metabolism. In fact, **Di Lorenzo et al, 2003** showed that disruption of the circadian cycle correlates with metabolic imbalance in individuals working night or rotating shifts: the prevalence of obesity was higher among shift workers compared to day workers, moreover shift workers showed higher BMI than day workers, and shift working was associated with BMI, independently of age and work duration. **Pitts et al, 2003; Pendergast et al, 2009** made similar observations in rodent models of circadian arrhythmia.

In conclusion, feeding is to be considered a circadian event, not only because it is an output of the clock, but also as a clock input mechanism. From metabolites to transcription factors, circadian clock and feeding intertwine in a crucial manner for the maintaining of metabolic homeostasis (**Eckel-Mahan, Sassone Corsi, 2013**).

1.2. The epigenetic link between metabolism and disease risk/prevention

In biology the term adaptation is referred to the ability of adjust in structure or habits, often occurring through natural selection, by which a species or individual becomes better able to function in their environment.

Many studies proved the ability of individuals responding to their environment by changing their own shape, as for example, leaf-mimicking insects that change color depending from the season and leafs that change shape depending from the conditions of soil, water and chemistry (**Laland et al, 2014**). Organisms are subjected to frequent environmental changes within their lifetime and natural selection responds inefficiently to these continuous immediate changes. Fitness to a fluctuating environment requires stable and reversible adaptation that involves the tuning of the genetic information by the soma. Physiological systems can respond and adapt to new changes in real time: the ongoing process by which internal body functions are regulated and adjusted to maintain homeostasis in the internal environment is called physiological adaptation. In particular, the physiological adaptation to the different caloric intakes is of great relevance as it is crucial for the organism's fitness. As proved by studies related to CR and HF diets in different organisms, dramatic changes in dietary regimens provoke a phenotypic remodeling, determining the susceptibility to life-threatening diseases (cf. Paragraphs 1.1.3 and 1.1.4). Recent studies in honeybees, mice and humans have shown that food can affect the activity of several chromatin-modifying enzymes, producing epigenetic traits that can be passed from generation to generation.

Moreover epigenetic modifications were proven to be alternative to genetic defects and sufficient to initiate tumorigenesis (when induced in animal models by genetic approaches) and may represent a common pathway of tumour progression.

In this context, epigenetics could represent the missing mechanistic link between diet and disease risk.

For these reasons, our working hypothesis in this study is that *food adaptation entails reprogramming of different cell functions, which might be executed and maintained through changes in chromatin*.

In this section we are going to focus on the main concepts related to epigenetics and its links with diet and disease development, to better set the frame in which our work is built.

1.2.1. Chromatin and epigenome

Chromatin is composed by naked DNA wrapped around *nucleosomes* - organized structures of specialized proteins called *histones* - and then packed to form chromosomes (Figure 1.5). Chromatin is not a mere depository of the genomic content but rather a signal transduction platform for extracellular or intracellular signals that regulates all genome functions, including gene expression, DNA replication and genome stability.

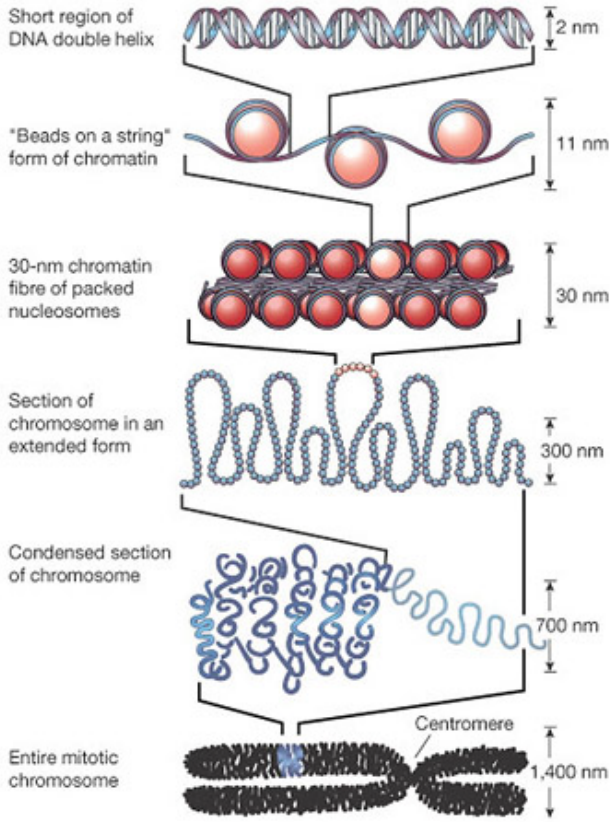


Figure 1.5 Chromatin structure

DNA is wrapped around nucleosomes, agglomerate of specific proteins called histones, creating a fiber called chromatin that is packed and condensed to form chromosomes.

(Adapted from Felsenfeld et al, Nature, 2003)

Upstream signals are translated by chromatin into either transient or long-lasting (and heritable during cell division) changes or modifications, thereby allowing chromatin to serve the double function of adapting cells to the environment changes while maintaining their lineage and/or identity. These modifications forms the *epigenome*: they are not modifications of DNA sequences but chemical changes happening on DNA (*DNA methylation)* or on specific regions of histone proteins called *tails* (*histone modifications*) that modify chromatin structure in different conformations as more open (*euchromatin)* and available to be bound by other proteins, or closed, compacted and then repressed (*heterochromatin*) (**Felsenfeld et al, 2003**).

1.2.2. Histone modifications

Histones are the structural units of the nucleosomes (the “beads” around which the DNA wraps to form chromatin fibers), they are very important proteins involved especially in gene regulation. Five are the major groups of histones: H1/H5, H2A, H2B, H3 and H4. Histones H2A, H2B, H3 and H4 are known as the *core histones*, while H1 and H5 are called the *linker histones*. Two of each of the core histones are needed to create an octameric nucleosome core and approximately 150 base pairs of DNA wrap around this core particle, while the linker histone H1 binds the nucleosome at entry and exit sites of the DNA, blocking it in place. The 4 core histones (H2A, H2B, H3 and H4) are relatively similar in structure and are highly conserved through evolution, having long *tails* on the N-terminal end which are more exposed, protruding from the center of the nucleosome core.

This tail is the location in which post-translational modification appears, altering the interaction of histones with DNA and nuclear proteins.

Many different modifications of the tail exist (**Kundaje et al, 2015**) of which, the more studied include methylation, acetylation, phosphorylation and ubiquitination (Figure 1.6). *Histone modifications* have a huge role in several biological processes such as gene regulation, DNA repair, chromosome condensation (mitosis) and spermatogenesis (meiosis). For these reasons, they often are present in specific genomic regions, like *promoters* (regions of DNA essential for the transcriptional regulation of genes; they are bound by both the basic transcription machinery complex and by a bunch of ancillary proteins (transcription factors, cofactors and chromatin modifiers) that all together impose a strict time- (cell cycle or development) and tissue-specific transcriptional program to their proximal target gene; they locate mainly upstream and around the Transcription Start Site - TSS - of genes) and *enhancers* (regions of DNA that can be bound by

proteins to activate transcription of a gene, that can be located nearby or far away from the activated gene, **Maston et al, 2006**).

The reaction of Lys-methylation is catalyzed by lysine (K) methyltransferases (KMTs) that uses S-adenosylmethionine (SAM or AdoMet) as a donor of methyl groups (**Smith BC et al, 2009**). Lysine residues may accept from 1 to 3 methyl groups and mono-, di-, or tri- methylated Lys are indeed observed (**Grant et al, 2001**).

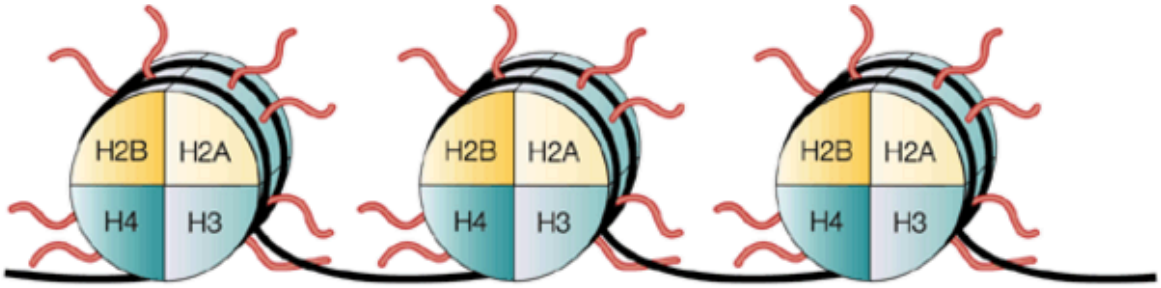
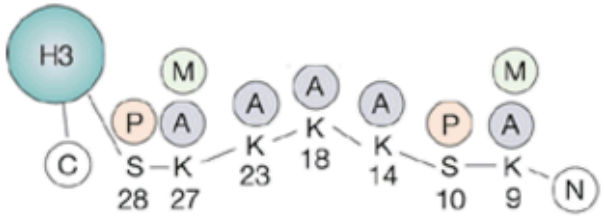


Figure 1.6 Nucleosomes, histones and tails modifications

Nucleosomes are constitued by two copies of core histones H2A, H2B, H3 and H4. Each histone has a tail protruding from the nucleosome, formed by an amino acid sequence that can be covalently modified adding, for example, methyl groups or acetyl groups in specific regions of the tail. Histone modifications are strongly related to gene transcription and regulation. Trimethylation of H3 Lysine 4, for example, correlates with active promoters.

(Adapted from Marks et al, Nature Rev. Cancer, 2001)



The tri-methylation of different Lys residues of the H3 histone is associated with different extent of gene transcription. In general, the tri-methylation of Lys 4, 36, 79 (H3K4me3, H3K36me3, H3K79me3) are found predominantly in the euchromatin (**Li, Carey et al, 2007**). In particular, H3K4me3 is localized in the proximity of the transcription start sites (TSSs) of many actively transcribed genes (**Kim et al,**

2005) or in promoters region bound by RNA polymerase II and others transcriptional factors (**Guenther et al, 2007**). These observations suggest that H3K4me3 is important to make chromatin accessible for transcription (**Li, Carey et al, 2007**).

H3K27me3, contrarily, is a marker of heterochromatin, found predominantly in repressed genes and usually where the H3K4me3 mark is absent (**Bernstein et al, 2005**). The only exceptions to this rule are the *bivalent promoters,* regions in which H3K27me3 colocalizes with H3K4me3 usually lying near genes that are poised for transcription, but needed to be rapidly expressed (**Bernstein et al, 2006**) (Figure 1.7 b-c).

Moreover, H3K4me1 is known to be especially associated to enhancer regions that usually are functionally active only when H3K4me1 it is also accompanied by an H3K27ac enrichment (**Shlyueva et al, 2014**) (Figure 1.7 a-d).

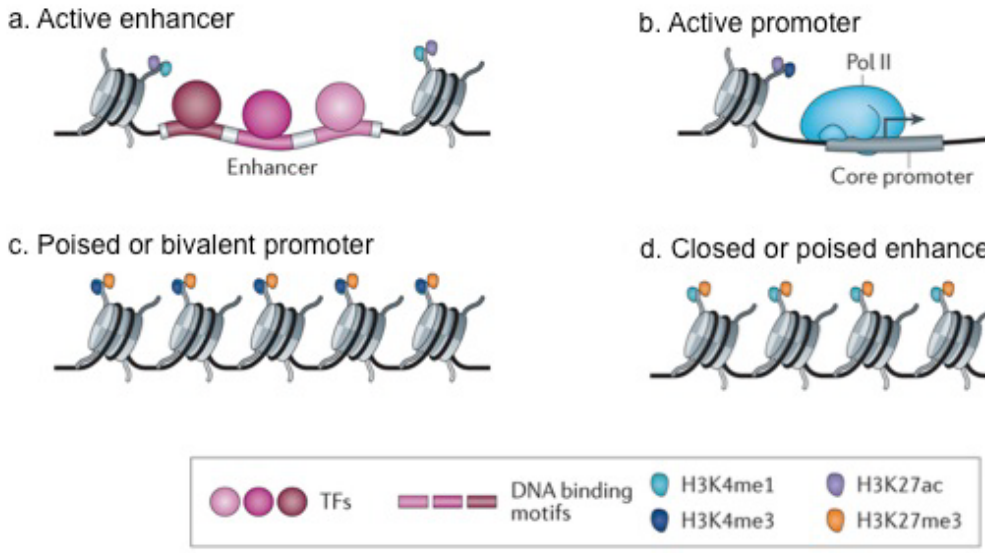


Figure 1.7 Histone H3 modifications and their role in gene regulation

Presence of H3K4me1 and H3K27ac (a) identifies active enhancers while H3K4me1 and H3K27me3 (d) turn off the region, defining a closed or poised enhancer. Active promoters, the ones in which Polymerase II can easily bind to initiate transcription of the genes, are identified by presence of H3K4me3 and H3K27ac (b), while poised genes are identified by presence of H3K4me3 and H3K27me3 (c).

(Adapted from Shlyueva et al, Nature Rev. Genetics, 2014)

1.2.3. DNA methylation

DNA methylation is a post-replication modification found at cytosines in any context of the genome (**Lister, Pellizzola, Dowen et al, 2009**). DNA methylation, is catalyzed by DNA methyltransferase genes and it is known to act as a repressor of gene transcription (**Kass et al, 1997**). It is fundamental for development being involved in genomic imprinting (phenomenon by which certain genes are expressed in a parent-of-origin-specific manner), X-chromosome inactivation (**Smith & Meissner, 2013**), and alteration of DNA methylation profile is a feature present in several diseases, including cancer (**Bergman & Cedar**, **2013**).

Differently from other modifications, DNA methylation can permanently alter the expression of genes in cells during cell division and differentiation from embryonic stem cells into specific tissues. This means that the resulting change is not reversible and permanent, in order to avoid that a differentiated cell could revert to a stem cell and then convert in another cell type. However, DNA methylation can be removed either passively, by dilution as cells divide, or by a faster, active, process. The latter process occurs via hydroxylation of the methyl groups that are to be removed, rather than by complete removal of methyl groups (**Iqbal et al, 2011; Wossidlo et al, 2011**).

1.2.4. Transcription Factors and Chromatin modifiers

Transcription initiation in Eukaryotes requires the activation of many proteins. In particular, the RNA polymerase enzyme, that actually transcribes genomic content from DNA to RNA, requires help to correctly positioning on gene promoters or pulling apart the two DNA strands. This kind of tasks are executed by *transcription factors,* proteins that, binding to specific DNA sequences, control the rate of

transcription of genetic information, activating or repressing RNA polymerase (**Latchman, 1997**).

These peculiar proteins are essential for gene expression regulation and, for this reason, massively present in all living organisms (in humans there are approximately 2,000 TFs). The distinguishing feature of TFs is the presence of one or more DNA binding domains, useful to recognize and bind only specific sequences next to the gene that has to be regulated (**Mitchell et al, 1989**).

Furthermore, genes often present flanking regions containing several binding sites for numerous transcription factors that all together work to properly regulate the targeted gene expression. All the possible combination of the ~2,000 human TFs, allow the unique regulation of each gene in the human genome during development (**Brivanlou, Darnell, 2002**).

Transcription factors bind to either enhancer or DNA promoter regions adjacent to their target genes and they can use different mechanisms to execute their task, as, for example, i) directly blocking or stabilizing RNA polymerase to the DNA, ii) recruiting histone acetylation/deacetylation (HAT/HDAC) proteins to produce the opening/closing of the chromatin fibers in specific regions (**Narlikar et al, 2002**) or iii) recruiting coactivator or corepressor proteins to the transcription factor DNA complex (**Xu et al, 1999**).

Transcription factors are usually recruited/activated downstream of signalling cascades triggered by internal or external stimuli. It is worth to notice that, among the external stimuli, the environmental conditions, also in higher organisms, play a very important role. For example, it is the case of the sterol regulatory element binding protein (SREBP), which helps maintain proper lipid levels in the cell (**Weber et al, 2004**).

1.2.5. Epigenetics alterations in disease

Histone modifications or DNA methylation are sufficient to initiate tumorigenesis and may represent a common pathway of tumour progression.

Indeed, a key feature of the cancer epigenome is the presence of a number of altered epigenetic traits (i.e., a reduction of the global content of methylated DNA; DNA hypermethylation at specific loci; increased methylation of lysine 4 of histone H3; decreased methylation of lysine 27 of histone H3), which are common to virtually all cancers, regardless their histological origin or stage of development (**Esteller, 2007; Hansen et al, 2011**). For example, **Ke et al, 2009** found that loss and/or gain of H3K4me3 and/or H3K27me3 in prostate cancer were strongly associated with differential gene expression in tumour samples compared to primary cells, thus indicating the presence of a H3K4me3/H3K27me3 epigenetic signature of prostate carcinogenesis. **He C et al, 2012** showed that high levels of H3K4me3 are associated with poor prognosis in hepatocellular carcinoma, while the change in H3K27me3 levels and the increased expression of H3K27me3 methyltransferase EZH2 leads to the silencing of tumour suppressor genes (e.g., *GAS2* and *ADRB2*) in prostate cancer patients. (**Chen Z et al, 2010**). Moreover, aberrant patterns of histone modifications, due to malfunctioning of both histone methyltransferases (HMTs) and histone demethylases (HDMs), were found in other conditions as diabetes (**Raciti et al, 2014**), cardiovascular disease (**Mathiyalagan et al, 2014**) and neurological diseases as Huntington's (**Urdinguio et al, 2009**).

Since histone modifications have been identified as possible predictive markers of disease, increasing attention is focused towards creating epigenetic drugs, such

as histone methyltransferase inhibitors for treatment, especially in cancer research (**Ngollo et al, 2014**).

1.2.6. Epigenetics and Diet

There are evidences that food can affect the activity of several chromatin-modifying enzymes. In fact, methyl groups derived from foods (i.e. fish, legumes, eggs, fruit, cereals) can favor histone methylation by increasing the cellular levels of the methyl donor S-adenosylmethionine or by regulating directly the activity of HMTs (**Park et al, 2012**).

The methylation of lys 4 residue in histone H3 (i.e mono-, di-, trimethylation) is induced by consumption of a high-starch/low-fat diet intake in rat jejunum. This trimethylation alters the gene expression of Si (sucrose-isomaltase) and Sglt1 (sodium-dependent glucose cotransporter) involved in carbohydrates metabolism. The levels of H3K4me1, H3K4me2 and H3K4me3, on the promoter and transcribed region of Si and Sglt1 genes were significantly higher in rats fed a high-starch/low-fat diet than in those fed a low-starch/high fat diet. On the contrary, the levels of H3K9me1, H3K9me2 and H3K9me3 (associated with heterochromatin) on the promoter and transcribed region of Si and Sglt1 genes were not significantly higher in rats fed a high-starch/low-fat diet than in those fed a low-starch/high fat diet (**Inoue et al, 2015**).

In addition, histone trimethylation of Lys residues changes in response to hyper-lipidemic diet that induces an increase of H3K9me3 and H3K4me3 in mouse primary hepatocytes. In this case, the high levels of H3K9me3 and H3K4me3 mark the promoters of many genes involved in biological pathways responsible for the development of hepatic steatosis and nonalcoholic fatty liver disease (**Jun et al**

2012). Moreover, it has been shown by **Kucharski et al, 2008** that bees larvae, producing queen and worker phenotypes, are genetically identical; the royal jelly silences Dnmt3 and activates genes needed to develop functional ovaries, egg laying abdomen and the necessary behavior to produce the queen phenotypic traits. Moreover, in “agouti viable yellow” (Avy) mice strain, which are prone to obesity, diabetes and cancer, mom’s diet can reverse in newborns the effect of unmethylated agouti gene, one of the gene that contribute to coat color: in fact, feeding with a methyl-rich diet a pregnant “yellow mouse”, the pups born with brown fur and stayed healthy for life (**Wolff et al, 1998**).

In humans, **Heijmans et al, 2008** reported that babies prenatally exposed to the Dutch hunger famine at the end of World War II, had lower level of DNA methylation of the imprinted IGF2 gene compared with their unexposed, same-sex siblings. Another interesting example is represented by SIRT1. This protein is a NAD+-dependent protein deacetylase, known to operate as a key nuclear metabolic sensor and as a mediator of the homeostatic responses to nutrient availability. Evidence indicates that SIRT1 may exploit these functions working as a master effector linking the metabolic status of a cell with the chromatin structure. In fact, by the deacetylation of histones, transcription factors and transcriptional co-factors, it is capable to regulate gene expression, thus influencing several fundamental cellular processes (**Brooks et al, 2009**). In addition, it has been recently demonstrated that the human epigenome contains hundreds of regions with high and stable inter-individual variability in DNA methylation, some of which correlate with the Body Mass Index (BMI) (**Dick et al, 2014**).

Moreover, **Eckel-Mahan et al, 2013** revealed, through analysis of H3K4me3 profiles in specific genes’ loci and expression data, that high-fat diet produces a

remodelling of the liver clock, disrupting the normal circadian cycle, impairing BMAL1 recruitment to target chromatin sites, and that these effects are reversible.

Later **Leung et al, 2014** showed that high-fat diet leads to chromatin remodelling in livers of C57BL6 mice, respect to mice fed with a control diet, and that these changes are associated with changes in gene expression.

These evidences confirm the interplay between diet and the epigenome, revealing the true potential in terms of possible therapeutic strategies for metabolic disease and cancer.

1.3. Next Generation Sequencing approach

The sequencing of the human genome and related organisms represents one of the most amazing scientific achievements in the history of mankind. From the discovery of DNA double helix in 1953 to the first DNA sequencing produced, 15 years passed and we had to wait until 1977 to watch the beginning of the modern sequencing (**Sanger et al, 1977**). Sanger DNA sequencing technology has allowed to advance enormously in molecular biology and genetics and several large projects have been successfully completed using this technology, as for example the **Human Genome Project**, **Rice Genome Project** and **Swine Genome Project**. However, Sanger low throughput, high cost and operation difficulties limited its use and increased the urge of researcher for faster and less costly sequencing. This need let to the rise of "Next Generation Sequencing" technologies (NGS): millions or billions of DNA molecules can be sequenced in parallel, highly increasing the throughput and minimizing the need for the fragment-cloning used in Sanger sequencing (**Ronaghi et al, 1996; Adams and Kron, 1994; Farinelli et al, 1998; Mayer et al, 1998**).

NGS has enabled researchers to characterize the molecular landscape of diverse diseases and produced a phenomenal advancement, especially in cancer genomic studies. For example, through whole-genome (WGS) and whole-exome sequencing (WES), there was an explosion of data in the context and complexity of cancer genomic alterations (point mutations, small insertions or deletions, copy number alterations, somatic and germline variants) (**Samuel & Hudson, 2012; Almendro et al, 2013; Yancovitz et al, 2012; Curtis et al, 2012; Bodini et al, 2015; Riva et al, 2013**).

Through whole transcriptome approach (RNA-Seq) it is possible not only to quantify gene expression profiles, but also to detect alternative splicing events, RNA editing and fusion transcripts (**Maher et al, 2009**; **Trapnell et al, 2010; Curtis et al, 2012; Graw et al, 2015**).

Moreover, epigenetic alterations, DNA methylation changes and histone modifications can be studied using other sequencing approaches including Bisulfite-Seq and ChIP-seq (**Schones & Zhao, 2008**; **Eckel-Mahan et al, 2013; The mouse ENCODE consortium et al, 2014; Engelen et al, 2015**)

In this new era we see the birth of new huge consortia like The Encyclopedia of DNA Elements Consortium (ENCODE, https://genome.ucsc.edu/ENCODE/; that from 2003 is building a catalogue of functional elements in the human genome, producing massive amount of OMICS high-throughput sequencing data publicly available) and The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/abouttcga; that grouping together different institutions have been able to collect, sequence and analyze thousands of samples of different tumours types in order to better understand the molecular basis of cancer through the application of genome analysis technologies).

The huge amount of data gathered through the combination of these approaches created the necessity of specific tools and skills in order to translate data into information. This necessity brought to the appearance of a new interdisciplinary field called *bioinformatics*, that combines computer science, statistics, mathematics, and engineering to develop methods and software tools in order to analyze, understand and interpret biological data (**Hogeweg, 2011**).

1.3.1. Chromatin Immunoprecipitation sequencing (ChIPseq)

Specific DNA sites in direct physical interaction with transcription factors and other proteins can be isolated by chromatin immunoprecipitation or ChIP, an experimental technique used to investigate the interaction between proteins and DNA in the cell. It aims to determine whether specific proteins are associated with specific genomic regions (as, for example, transcription factors on promoters or other DNA binding sites), or specific location in the genome that various histone modifications are associated with, indicating the target of the histone modifiers.

The protocol method is briefly described in Figure 1.8 and involve the following steps (**Orlando, 2000**):

1. Crosslinking of DNA and associated proteins on chromatin in living cells or tissues;
2. Complexes constituted by chromatin and protein are then sheared into ~500 bp DNA fragments by sonication or nuclease digestion;
3. Using a protein-specific antibody, cross-linked DNA fragments associated with the protein(s) of interest are selectively immunoprecipitated from the cell debris;
4. The DNA associated with the complex is then purified and identified by polymerase chain reaction (PCR), microarrays (ChIP-on-chip), molecular cloning and sequencing, or direct high-throughput sequencing (**ChIP-Seq**).

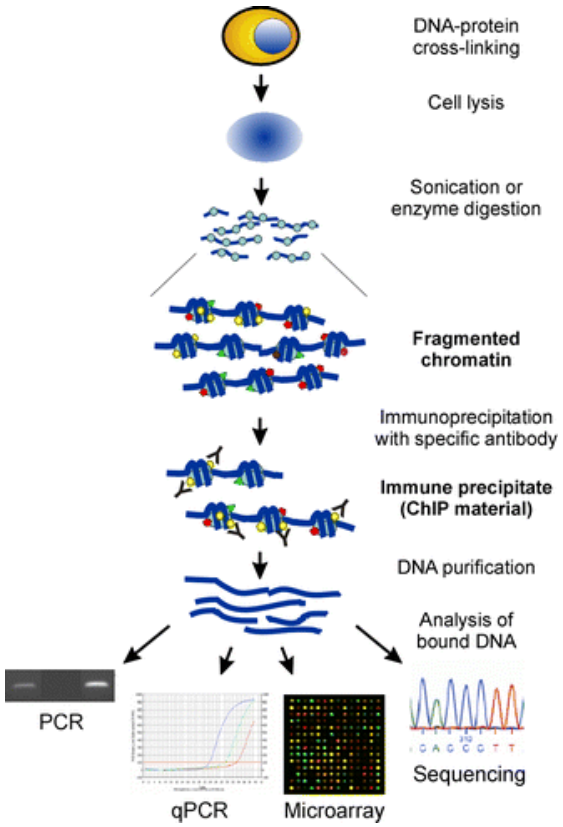


Figure 1.8 ChIP protocol steps

After crosslinking of DNA and associated proteins on chromatin in living cells or tissues, the DNA-protein complexes are then sheared into fragments by sonication or nuclease digestion. Using a protein-specific antibody, cross-linked DNA fragments bound with the protein(s) of interest are selectively immunoprecipitated. Finally, after purification, the DNA associated with the complex can be identified by polymerase chain reaction (PCR), microarrays (ChIP-on-chip), molecular cloning and sequencing, or direct high-throughput sequencing (ChIP-Seq).

(Adapted from Collas, Mol. Biotechnol. review, 2010)

In particular, in ChIP-seq, massively parallel sequence analyses are used in conjunction with whole-genome sequence databases to analyze the interaction pattern of any protein with DNA. The first studies using ChIPseq were published in 2007 (**Johnson et al, 2007; Barski et al, 2007; Robertson et al, 2007; Mikkelsen et al, 2007**) and many more followed; by now is one of the most used techniques for epigenomic studies.

1.3.1.1. Pathology Tissue Chromatin Immunoprecipitation (PAT-ChIP)

In general ChIP protocol is performed on chromatin obtained from cells cultured in vitro or from fresh tissues, but not always fresh samples are available for an immediate analysis, especially in clinical practice. In fact, clinical samples come often as paraffin-embedded tissues (commonly called FFPE - Formaldehyde Fixed-Paraffin Embedded) and large archives of these FFPE samples are present in most hospitals and have been extensively used for detailed case studies.

Pathologists use formalin to preserve biopsies and maintain intact their cellular structure, including cross-linked DNA and proteins as well. In our laboratory we developed a specific protocol called PAT-ChIP (pathology tissue chromatin immunoprecipitation) to utilize FFPE for chromatin preparations and analysis (**Fanelli et al, 2010**).

The setup of PAT-ChIP protocol was originally validated using spleen murine. First step of chromatin extraction is quite different from the classic ChIP protocol, according to which single-cell suspensions are cross-linked for short periods (10–15 min) using lower concentrations of formalin (1%). In PAT-ChIP protocol, tissues are treated overnight with high concentrations of formalin, followed by paraffin embedding. Chromatin preparation from FFPE-tissue started with the re-hydration and deparaffinization of 10-μm FFPE-tissue sections, followed by chromatin fragmentation and extraction. The sonication step was adapted to have comparable size of DNA fragments from chromatin of both Cells-ChIP and PAT-ChIP. Agarose gel electrophoresis demonstrated that the DNA fragments obtained after sonication have the same size (about 300 bps) both using cells or FFPE-spleen samples (Figure 1.9).

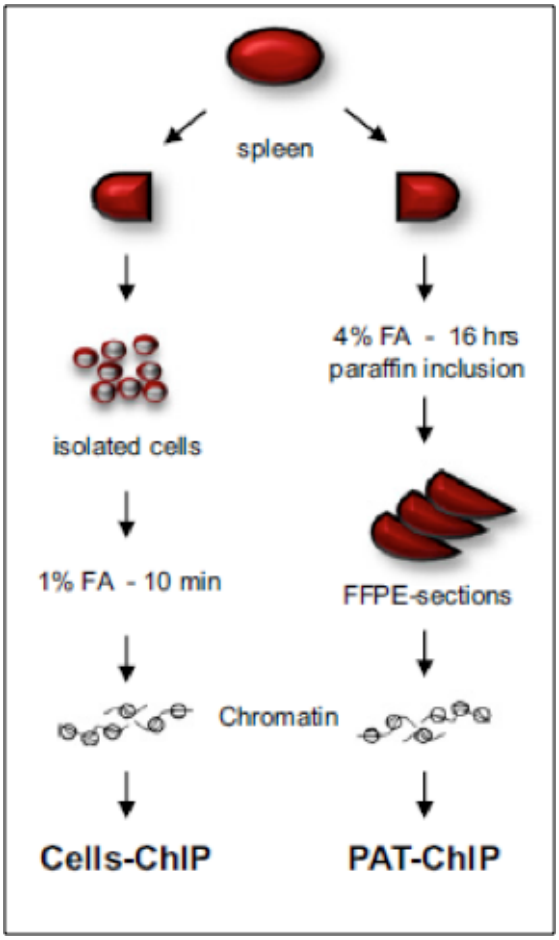


Figure 1.9 Comparison of Cells-ChIP and PAT-ChIP procedures
Schematic representation of canonical ChIP and PAT-ChIP procedures starting from the spleen of a mouse.
Results showed a robust overlap between results obtained with the two ChIP strategies.

(*Adapted from Fanelli et al, PNAS, 2010*)

Chromatin samples extracted by the two methods from mice spleens were compared using a set of specific antibodies for histone modifications such as H3K4me3 and H3K27me3 (respectively, associated to active and silent promoters) and hyper-acetylated histone H3 or H4 (H3ac, H4ac; associated with active regulatory regions). The amount of immunoprecipitated DNA obtained by Cells-ChIP or PAT-ChIP was similar for each of the used antibodies.

To compare these two approaches, the immunoprecipitated DNAs were analyzed by real-time quantitative PCR (qPCR) of four promoter regions of genes expressed (βActin and Gapdh) or silent (Crtl1 and Col2a) in spleen. βActin and Gapdh are

positive controls for H3K4me3, H3ac, H4ac and negative for H3K27me3, while Crtl1 and Col2a are negative controls for H3K4me3, H3ac, H4ac and positive for H3K27me3. Results showed a robust overlap between results obtained with the two ChIP strategies. The performance of PAT-ChIP and standard ChIP-protocols, were investigated also in high-throughput sequencing methodologies. Purified DNA was analyzed by ultra-sequencing using the Illumina Genome Analyzer II.

A dataset of H3K4me3-enriched genomic regions was generated for each experiment (Cells-ChIPSeq or PAT-ChIP-Seq) .

Also in this case, the data showed a very high correlation between Cells-ChIP and PAT-ChIP datasets, showing a substantial overlap of the two techniques. (**Fanelli M et al, 2011**).

1.3.2. Whole transcriptome sequencing (RNA-seq)

The evaluation of the gene expression profile of a cell or a tissue through the quantification of mRNA levels is a matter of great interest to researchers. In fact, measuring mRNA concentration levels is useful in order to understand how external cues can affect the transcriptional machinery of the cell or how transcriptome profiles differ between a healthy state and a diseased state.

One of the first experimental method introduced to reply to this kind of questions is represented by DNA microarrays: collections of microscopic DNA spots attached to a solid surface and each DNA spot contains small quantities of a specific DNA sequence, known as probes or oligos used to hybridize a cDNA or cRNA target sample. Probe-target hybridization is detected and then quantified to determine relative abundance of nucleic acid sequences in the target (**Baldi & Hatfield, 2002**). Although they are still largely used, microarrays require species- or transcript-specific probes, moreover background hybridization limits the accuracy

of expression measurements (**Zhao et al, 2014**). These limitations have been overcome by the introduction of RNA-seq (**Mortazavi et al, 2008**), in which a population of RNA (total or fractionated, such as poly(A)+) is converted to a library of cDNA fragments with adaptors attached to one or both ends (see Fig. 1.10).

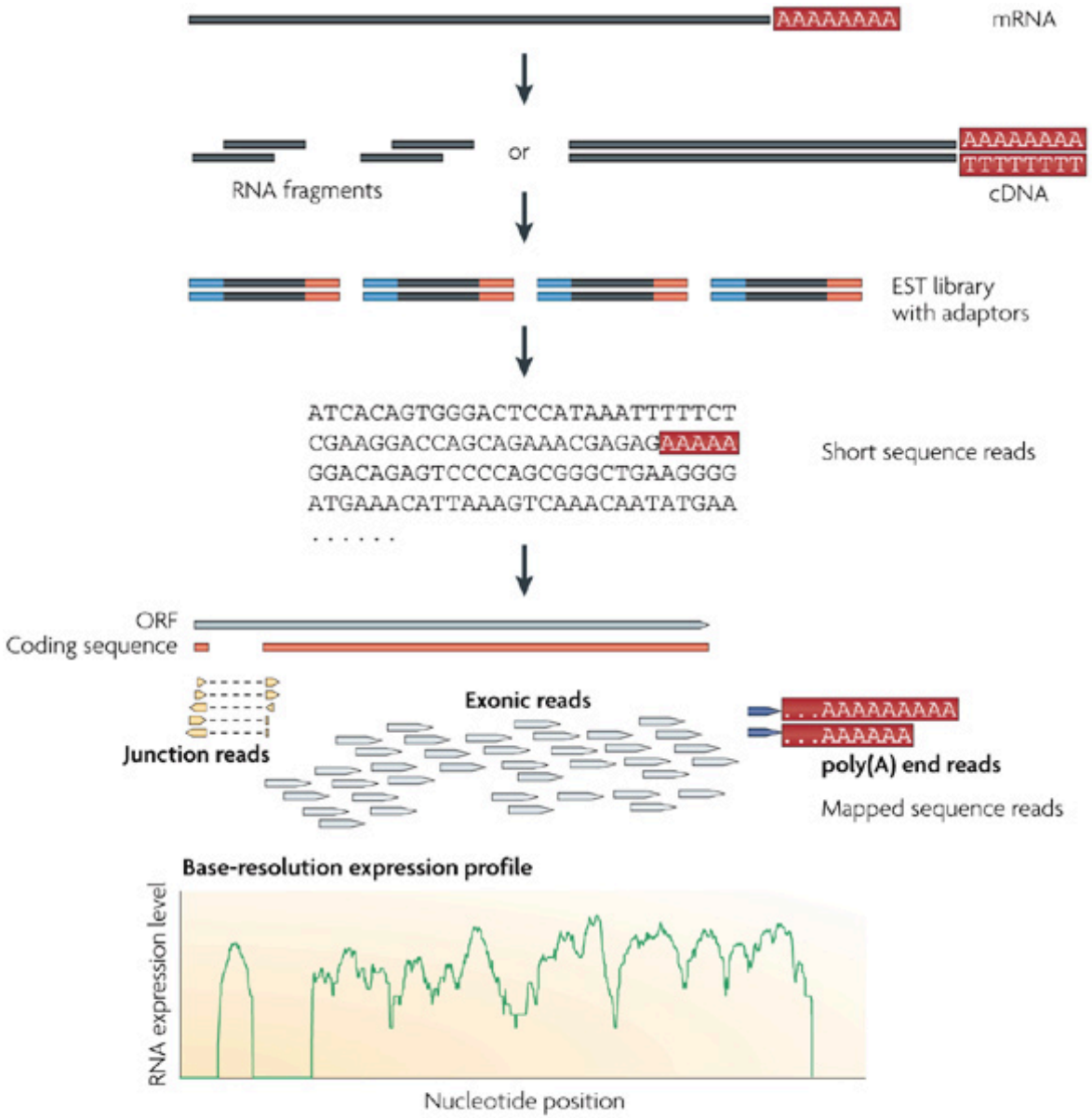


Figure 1.10 RNA-seq protocol steps

Long RNAs are converted into a library of cDNA fragments through either RNA fragmentation or DNA fragmentation. Then sequencing adaptors (in blue) are added to each cDNA fragment and short sequences are obtained from cDNAs using high-throughput sequencing technology.
Resulting sequence reads are aligned with the reference genome or transcriptome, and classified in three groups: exonic reads, junction reads and poly(A) end-reads. These three groups are then used to generate an expression profile for every gene, as illustrated at the bottom.

(Adapted from Wang et al, Nature Review Genetics, 2009)

Each molecule, with or without amplification, is then sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing). The reads have a length of 30–400 bp, depending on the DNA-sequencing technology used (**Wang et al, 2009**).

1.3.3. Bioinformatics Data format and general overview of analysis

Becoming cheaper and cheaper over the years, generating genomic data have made high-throughput sequencing an increasingly important part of biomedical research. This has created a new challenge of finding efficient and effective ways to analyze data and generate insights into the function of biological systems.

Raw data coming from the sequencing are contained in *FASTQ* files, a text-based format file that stores both the biological sequence and the corresponding quality scores, encoded with a single ASCII character (**Cock et al, 2009**). A FASTQ file normally uses four lines for each sequence:

- Line 1 begins with a '@' character and is followed by a sequence identifier and an optional description;
- Line 2 is the raw sequence letters (nucleotides);
- Line 3 begins with a '+' character and is optionally followed by the same sequence identifier (and any description) again;
- Line 4 encodes the quality values for the sequence in Line 2, and must contain the same number of symbols as letters in the sequence.

After filtering out low quality sequences (also called *reads*), the remaining are *aligned* against a reference genome organism-specific: each read sequence is compared to the reference genome sequences through a mapping algorithm and in this way a corresponding location is determined. The algorithm will try to find a

location, possibly unique, in the reference sequence that matches the read, tolerating a certain amount of mismatch to allow subsequence variation detection. Many different alignment algorithms exist; recently **Flicek & Birney, 2010** made a complete comparison of the most commonly used.

Many are the possible source for errors during the alignment step, for example PCR artifacts due to the PCR steps, will show up in multiple reads, or in *duplicates* (the same read occurs multiple times, skewing coverage calculations in the alignment). This kind of issues is taken into account in the following steps of the data processing in which, i.e. the removal of duplicated reads may be applied.

SAM and *BAM* files are, respectively, the text and binary format files usually produced as output by the aligner programs; beside the mapping location of each reads they contain a lot of other interesting information such as the quality of the alignment, presence, number and position of mismatches, and, for paired-end experiments, they keep info related to each read mate (**Li et al, 2009**). They are the starting-point data to flow in *ad hoc* analysis pipelines used for each specific type of experiment.

Importantly, data can be visualized through Genome Browsers. One of the most used is the UCSC Genome Browser (**Kent et al, 2002**): they are graphical interfaces in which it is possible to display and integrate information from several biological databases, enabling researchers to visualize and browse entire genomes.

Focusing in particular on ChIPseq and RNA-seq experiments, the most commonly used analysis pipelines follow:

- **ChIP-seq data analysis pipeline (Fig. 1.11):** after filtering out low quality reads, the remaining ones are mapped to the respective genome

sequences and they can be visualized on Genome Browser converting BAM files in *BIGWIG* files, a designed format for display of dense continuous data. The “peaks”, which correspond with regions of the genomes where ChIP-sequenced reads are overrepresented, and piling up, form a “bell-like shape” profile, represent the sought binding events and they are identified (in jargon, “called”) through a peak-caller program (see **Wilbanks and Facciotti, 2010** for a comparison of some of the most used peak-callers and algorithms which they employ). Other tools are then used to annotated these regions with respect to the reference genome and/or to other functional features, i.e. coding and noncoding genes. Then downstream analysis can include (but does not limit to) peak comparison among samples in different states to observe presence/absence of specific peaks, Gene Ontology or Pathway analysis, recurrent motif search, checking for quantitative significant changes in binding levels, peak shape analysis.

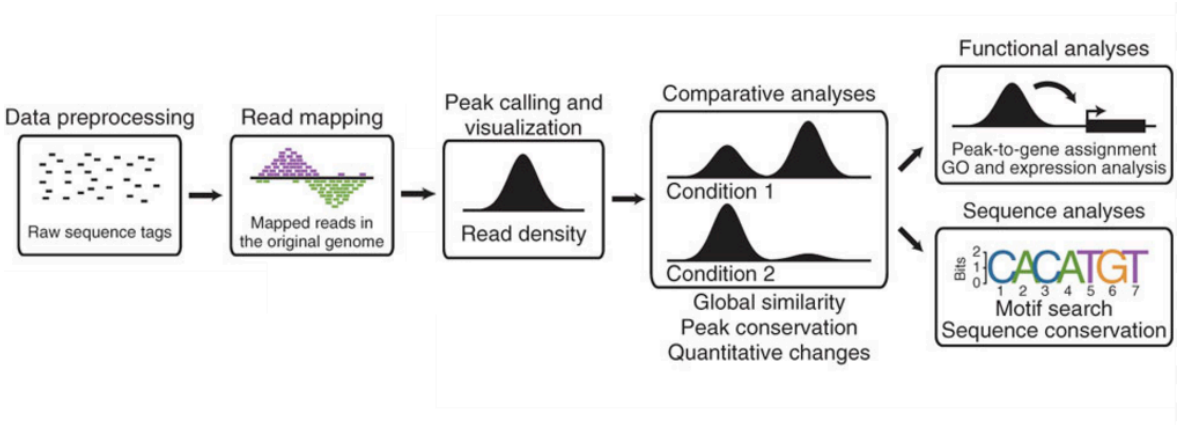


Figure 1.11 ChIPseq data analysis pipeline workflow

Raw data are preprocessed and mapped to the respective reference genome sequences; read densities can be visualized along the genome, and peaks representing binding events are called. Comparative analyses include a comparison of global binding similarity, analyses of presence/absence of peaks (i.e., peak conservation) and quantitative assessment of binding changes. Functional analysis such as Gene Ontology analysis of target genes, recurrent motif search and sequence conservation can then be conducted.

(Adapted from Bardet et al, Nature Prot. 2012)

- **RNA-seq data analysis pipeline (Fig. 1.12):** Once high-quality reads have been obtained, the first task of data analysis is to map the short reads from RNA-Seq to the reference genome using splicing-aware alignment tools like for example, TopHat or STAR (**Trapnell et al, 2009; Dobin et al, 2013**) which, contextually with the alignment process to the reference genomic sequence, can also take advantage of transcriptomic data to add information to the output data (i.e. the strand direction mapping both to genome and to parental gene). In fact RefSeq, UCSC Gene or GenCode transcript tables, in gtf format, are often also provided to the aligners. Other quality checks are needed to address RNA-seq-specific questions, such as exonic versus intronic alignments and transcript detection rates, duplication rates, GC bias, contaminating ribosomal RNA content, continuity of coverage, 3′/5′ biases and counts of detectable transcripts, among others (**DeLuca et al, 2012; Shen et al, 2014**). Then it is possible to summarize gene- or gene variant-level read counts using HTseq (**Anders et al, 2014**). Finally to identify differences in RNA expression levels of individual genes, or of individual splice variants of a single gene, between control and experimental samples, differential analysis can be performed and there are numerous tools available. Among the most popular ones are DESeq (**Anders & Huber, 2010**) and edgeR (**Robinson et al, 2010**), with both methods based on negative binomial testing, which provides an exact test (generalization of the Poisson distribution model) that is ideal for modeling biological variances of read count data.

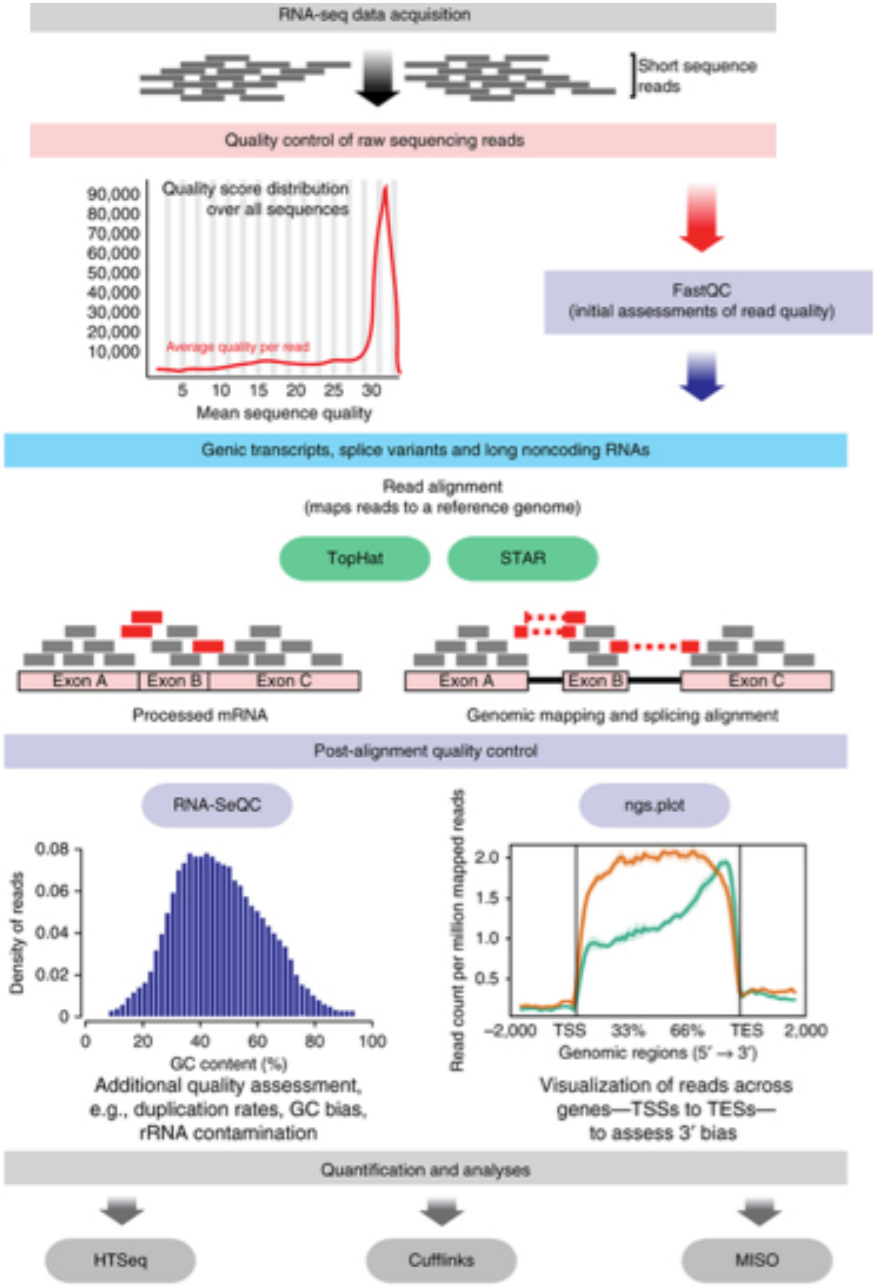


Figure 1.12 RNA-seq data analysis pipeline workflow

Following data acquisition, RNA-seq analyses typically begin with quality control assessments using analytical tools such as FastQC. Next, short sequencing reads can be aligned to a reference genome using programs such as TopHat or STAR. After alignment, additional quality control assessments can be made with RNA-SeQC and ngs.plot.
Finally, to quantify and analyze RNA-seq data, programs such as HTSeq, Cufflinks and MISO are typically used. Then for differential expression analysis in different conditions (for example, cancer tissue vs healthy one), edgeR or DESeq can be used.

(Adapted from Maze et al, Nature Neurosc., 2014)

2. Materials and methods

In this chapter we report experimental procedures used in this project and all bioinformatic methods.

Briefly, this project includes the study of 19 C57BL6 WT female mice divided in 3 groups. Each group was subjected to a different diet regimen: Standard Diet (SD - control diet), Calorie Restriction (CR) without malnutrition, High Fat Diet (HF). Mice were maintained on diet for 10 months and at the end of the treatment sacrificed; liver samples were collected, fixed in formaldehyde and embedded in paraffin (FFPE) for further analyses. Chromatin Immunoprecipitation for FFPE samples (PATChIP) anti-H3K4me3 (open chromatin - active transcription marker) and anti-H3K27me3 (suppressive marker) followed by deep sequencing were then conducted on liver samples. Sequencing data generated from Illumina HiSeq 2000 were then filtered to remove low quality reads, then the reads were aligned *versus* a reference mouse genome (UCSC mm9) with BWA and PCR artifacts were removed with SAMtools. Enriched regions or "peaks" were identified with SICER peak caller and annotated with ChIPseeker. We then applied two different methods for downstream analysis: i) a "positional" approach, using BEDtools, to identify specific regions present/absent specifically in each diet group; ii) a “quantitative” approach, based on the statistical comparison of read-density information of samples through DiffBind R package, to identify regions differentially enriched among different diet-groups.

Selection of regions by the two methods was followed by functional annotation of corresponding genes to KEGG pathways and GO analysis. Finally MEME suite was used to find recurrent motifs in these subsets of enriched regions. In parallel,

RNA-seq was performed starting from frozen livers of CR and SD female mice. After RNA extraction and Illumina HiSeq 2000 sequencing, low quality reads were filtered out and the remaining reads were aligned with TopHat versus the reference mouse genome (UCSC mm9). Gene-level read counts were calculated with HTSeq Count and further quality check were then performed to account for possible biases with the RSeQC set of tools. Finally differential expression analysis with edgeR was performed to find regulated genes in CR versus SD condition and genes lists were used for functional annotation to select specific GO biological processes and KEGG pathways.

2.1. Diet treatment of mice colonies and samples collection

C57BL6 8 weeks old female mice were generated at IEO facility and then divided in groups and maintained in the animal facility of the University of Milan with ad libitum standard (SD), restricted (CR) or high fat (HF) diets for 10 months.

All the experiments with mice were performed in accordance with the Italian Laws (D.L.vo 116/92 and following additions), which enforces EU86/609 directives (Council Directive 86/609/EEC of 24 November 1986 on the approximation of law regulations and administrative, provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes).

Mice were randomly divided into 3 groups (n = 10 per group): SD (control), HF and CR.

Standard diet (2018S Tekland 18% Protein Rodent Diet, provided by Harlan Tekland, Madison, WI, USA) is a fixed formula, non-autoclavable diet; energy provided by the macronutrients was approximately 30% of proteins, 15% of fats and 55% of carbohydrates, for a total of 3.3 kcal g^{-1}.

High fat diet (Diet Inducing Obesity D12492 provided by Brogaarden Aps, Denmark) contains 60% more fat ingredients than SD and kilocalories account for 20% from proteins, 60% from fat and 20% from carbohydrates, for a total of 5.24 kcal g^{-1}.

To calculate 30% caloric restriction starting at 8 weeks of age, daily food intake was measured in a subset of mice fed ad libitum. Food intake was determined by collecting and weighing all food remaining in the food hopper and cage at the same time each day for a week. Every day 20% or 30 less of amount of food, with respect to the average daily food intake observed ad libitum, was provided to the CR mice.

After 10 months of diet, all survived mice (8 for SD, 7 for CR and 5 for HF) have been sacrificed by cervical dislocation. Organs (Liver, Brain, Lungs, Kidneys, Intestine, Spleen, Abdominal Fat, Heart) were collected, rapidly washed in phosphate buffered saline (PBS) and incubate overnight at room-temperature in 4% formalin solution. Formalin-fixed samples were then routinely dehydrated by increasing concentrations of ethanol, starting from 70% through to 80%, 90% and 100% (absolute ethanol), and subsequently included in paraffin with use a tissue processor.

For this project only liver samples have been used, as: i) liver is a key metabolic organ; ii) it is easy to manipulate because of the organ dimension, iii) its histological structure is rather homogeneous, avoiding any issue on tissue dishomogeneity. Livers have been divided in halves; the first was used for FFPE samples for ChIPseq experiments, while the other half was flash frozen in liquid nitrogen and stored at -80°C for RNA-seq experiments.

2.2. Experimental procedures

In order to gain information on the possible impact of different diets on murine epigenome, we used collected liver paraffin-embedded and frozen tissues to perform respectively, Chromatin Immunoprecipitation from pathology tissues (PAT-ChIP) against two different histone modifications (H3K4me3 and H3K27me3) and RNA PolyA extraction, both followed by deep sequencing. The following paragraphs contain the details of the two experimental procedures protocols used. All the procedures were performed by wet lab biologists of our group, Costanza Savino, Valeriano Gentile (PAT-ChIP and library preparation) and post-doc, Elena Mylona (RNA extraction and library preparation).

2.2.1. PAT-ChIP from FFPE-liver samples and libraries

Chromatin extraction from FFPE-liver samples started with the rehydration and deparaffinization of 4 sections 10-μm of FFPE samples. Previously, samples were treated with histolemon to remove the paraffin and later treated using different decreasing concentration of ethanol and finally rinsed in water in order to rehydrate the tissue.

To extract chromatin from FFPE tissues, physical disruption of cell membranes and enzymatic digestion with micrococcal nuclease, were combined. Then, the extracted chromatin was fragmented through sonication.

The sonication step, which is one the more crucial and tricky, was adjusted, defining a quantity of chromatin to sonicate in order to achieve fragments of 300-150 bp in length and to immunoprecipitate the chromatin at a higher efficiency. Sonicated chromatin was immunoprecipitated using specific ChIP grade antibodies (H3K4me3 Rabbit pAb; Active motif and H3K27me3 Rabbit pAb;

Millipore). The obtained immunoprecipitated DNA was finally quantified by Qubit. Real time qPCR was performed to estimate the enrichment of H3K4me3 and H3K27me3 in promoter regions of known actively transcribed (Gapdh and βactin) and not transcribed genes (Crtl1 and Col2a1) in the liver. Finally, libraries were prepared following the HT-ChIPSeq library protocol (**Blecher-Gonen et al, 2013**).

2.2.2. RNA extraction from frozen liver samples and libraries

To isolate poly(A)+ mRNA, Qiagen Rneasy Mini Kit was used starting from 30 mg of frozen liver tissue per 600 uL of buffer RLT.

On-column DNase digestion was included and the RNA-seq library was built using the Illumina TruSeq version 2 kit (Low Sample protocol), starting from 1 ug total RNA (QC RNA: Bioanalyzer nano RNA kit, RIN minimum 8).

2.2.3. HiSeq2000 Illumina sequencing

ChIP-seq libraries and RNA-seq libraries were sequenced at the IEO NGS facility with Illumina HiSeq2000 and, in particular, for PAT-ChIPseq, 51 bp reads, 30 millions of reads depth of sequencing, single-end were used; while for RNA-seq we used 51 bp reads, 35 million reads depth, paired-end.

2.3. Bioinformatics methods

After sequencing we applied a quality filter to fastq files in order to remove low quality reads. Then we aligned the reads to the reference mouse genome. Enriched regions or "peaks" were identified and annotated; additional quality steps were performed to evaluate the samples enrichment**.** In order to both verify the extent of the variability among samples exposed to the same dietary regimens, at first, and finally to assess differences among groups of mice fed with different diets, we applied two different methods: i) a "positional" approach, where the

analysis was based on comparison of mapping information of the peaks to identify common or specific enriched regions; ii) a "quantitative" method, based on the statistical comparison of read-density information on a common list of peaks along different samples. Finally lists of genes obtained from the previous steps were characterized for downstream analysis including KEGG pathway analysis with ClusterProfiler, Gene ontology enrichment analysis and recurrent motif searching with MEME suite. Parallely, for RNA-seq samples from frozen livers of CR and SD males and females mice, low quality reads were filtered out and the remaining reads were aligned with TopHat versus a reference genome (UCSC mm9). Gene-level read counts were calculated with HTSeq Count and further quality check were then performed to account for possible biases with RNA-SeQC. Finally differential expression analysis with EdgeR was performed to find overexpressed and underexpressed genes in CR versus SD condition.

2.3.1. ChIPseq data analysis pipeline

The analysis of ChIPseq samples is divided in three main parts: Preprocessing analysis, Variability analysis, Downstream analysis.

Preprocessing analysis.

Samples were sequenced with Illumina HiSeq2000 with reads length of 51 bp. After applying a quality filter to remove low quality reads, we aligned reads to a reference mouse genome (UCSC mm9 assembly) with BWA **(Li and Durbin, 2009)**, and we removed duplicates, considered putative PCR artefacts, with SAMtools **(Li et al, 2009)**. Therefore we used SICER **(Zang et al, 2008),** considered more sensitive and specific for histone marks analysis than MACs1.4 **(Zhang et al, 2008)** to identify enriched regions or "peaks". Indeed we tested both programs on our samples confirming this broadly accepted cognition. Parameters for SICER runs were adjusted for each samples considering the their overall

quality: in particular we applied different thresholds according to the overall enrichment level of each samples [on the basis of the number of called peaks with the default threshold (E-value=100) and the relative Fraction of Reads in Peaks (FRiP) index calculated as reported in **(Landt et al, 2012**). Samples with FRiP<2%, as also suggested in **(Landt et al, 2012)**, were discarded.

We used a lenient limit (E-value=1000) for samples with low enrichment (FRiP≤5% and number of peaks ≤ 10,000 for H3K4me3 samples) and the default value (E-value=100) for all the other samples. Distribution of number of reads, peaks and peaks lenght of the samples were retrieved with to observe technical variability of the sequenced samples.

Peaks were finally annotated with an R package called ChIPseeker (**Guangchuang et al, 2015**) respect to the reference genome UCSC mm9.

Different genomic classes were considered to annotate peaks: **Promoter regions** (interval centered on Transcription Start Site of genes of 2kb, 4kb and 6 kb), **5' and 3' UTRs**, **Exonic regions** (divided in 1st Exon and Other Exon), **Intronic regions** (divided in 1st Intron and Other Intron), **Downstream** regions (less than 3Kb from the gene end), **Distal Intergenic** (outside genes and far from them).

Each peak is annotated to the nearest gene with priority given to the order of the above described genomic classes.

Variability analysis.

In order to perform the downstream analysis, we have first to take into account of possible differences due to biological variability.

To this aim we started analysing the within-diet biological variability, calculating for each diet group and for each histone modification the percentage of peaks in each genomic class and for each sample and the number of overlapping peaks respect

to all the other samples of the same diet group. Then, for each diet group, we computed a similarity matrix defined as:

$$J(d) = \begin{bmatrix} J_{11} & \dots & J_{1n_d} \\ & \vdots & \\ J_{n_d1} & \dots & J_{n_dn_d} \end{bmatrix}$$

where:

$$d = \mathrm{HF, SD, CR}$$
$$n_d = \text{number of samples of diet group d}$$
$$J_{ij} = 1 - \frac{|S_i \cap S_j|}{|S_i \cup S_j|} \; i = 1,\dots,n_d$$
$$\text{where } \mathrm{S_i} \text{ is the peak set of sample i of diet d}$$

and numerator is the cardinality of the set of peaks shared by sample i and j (the number of overlapping peaks between the two samples), while denominator is the cardinality of the union of peak sets i and j (the number of peaks of both samples minus the number of overlapping peaks). J_{ij} is also called Jaccard distance.

Then we represented the similarity matrices with heatmaps to observe concordance on peak calling for each diet group and assess intra-diet variability.

Finally, we perform ANOVA statistical test to investigate variability among diets (inter-diets).

Downstream analysis.

Once obtained information related to biological internal and inter-diets variability, we perform two different methods:

1. The “positional” method, focused on the comparison of peaks in different conditions in order to identify specific regions enriched (peaks presence/absence);
2. The “quantitative” method, based on the statistical comparison of read-density information on a list of shared peaks along different samples.

The positional method.

To take into account of the internal variability of each diet, first of all we used IntersectBed (**Quinlan et al, 2010**) and the Elbow method (**Thorndike, 1953**), plotting the number of overlapping peaks depending on the number of samples having those peaks in common, to select only reliable peaks for each diet group, i.e. peaks occurring in at least a certain number of replicas.

After creating these 3 lists of "*solid"* peaks (one for each diet-group) for each histone modification, we annotated them and, in order to retrieve peaks that are only occurring in a specific diet group respect to the control (*gained peaks*), we compared CR and HF solid peaks against all peaks present in SD samples.

Viceversa, comparing SD solid peaks with all peaks present in CR samples or in HF samples, we obtained *lost peaks* for CR and HF respectively.

Finally genes corresponding to gained and lost peaks, were used for functional annotation, retrieving information related to enriched KEGG pathways and gene ontologies biological processes with ClusterProfiler (**Guangchuang et al, 2012**), comparing the obtained results with up to date known literature.

The quantitative method.

This time we want to consider the quantitative differences (meaning the relative abundance of reads) in peaks for the different diet conditions. Specifically enriched regions for each diet group were derived by a differential enrichment analysis with DiffBind R package (**Diffbind, Stark et al**): DiffBind provides functions for processing ChIPseq data enriched for

genomic loci where specific protein/DNA binding occurs. It is designed to work simultaneously with multiple peak sets, representing different ChIP experiments. Starting from the original peak sets and from aligned reads files of each sample (bam files), DiffBind identifies a consensus peak set (peaks shared at least by a certain number of samples), it merges the initial peak sets and counts sequencing reads within the new intervals in the consensus peak set. To identify the best threshold to build the consensus peak set, as for the positional method, we used the "Elbow method" plotting the number of overlapping peaks depending on the number of samples having those peaks in common. After a normalization step, DiffBind identifies significantly differentially bound sites (DB sites) based on evidence of binding affinity using edgeR (**Robinson et al, 2010**) or DESeq (**Anders and Huber, 2010**), which are two widely used R statistical routines for RNA-seq differential expression data analysis. We used Trimmed Mean of M-values normalization (**Robinson and Oshlack, 2010**; best normalization method according to **Dillies et al, 2012**) and edgeR for the analysis.

Once retrieved statistically significant specific differentially bound sites for CR or HF *versus* SD, we annotated them and searched for possible enriched KEGG pathways and GO biological processes with clusterProfiler (**Guangchuang et al, 2012**) and recurring motifs with MEME suite (**Bailey et al, 2009**).

2.3.2. RNA-seq data analysis pipeline

RNA-seq samples coming from 3 CR (CR6, CR8, CR9) and 1 SD mice were analysed.

After filtering out low-quality reads, we mapped the short reads from RNA-Seq samples to the reference genome (mm9 UCSC) using TopHat **(Trapnell et al, 2009)**, while gene-level read counts were obtained using HTseq-Count (**Anders et al, 2014**).

Then quality checks, needed to address possible biases, were performed with RSeQC (**DeLuca et al, 2012**). Finally to identify differences in RNA expression levels of individual genes, between control and experimental samples, differential analysis was performed with edgeR (**Robinson et al, 2010**). To estimate samples variability, we calculated the similarity matrix starting from sample reads counts per gene among samples:

$$E = \begin{bmatrix} d_{11} & \cdots & d_{14} \\ & \vdots & \\ d_{41} & \cdots & d_{44} \end{bmatrix}$$

where:

$$d_{ij} = d(S_i, S_j), \quad i = 1,\ldots,4$$

and S_i is the array with the read counts for each gene and d is the euclidean distance between two vectors.

We used a heatmap with hierarchical clustering to represent the similarity matrix. A volcano plot representing $\log_2$(fold change) and $-\log_{10}$(p-value), for genes with RPKM value greater than 1 in at least one of the two conditions was used to define significance thresholds of regulated genes. Only entries with $|\log_2(FC)| \geq 1$ and a p-value$\leq$ 0.05 were considered to investigate possible enriched pathways and biological processes.

We used clusterProfiler R package to perform GO and KEGG enrichment analysis and graphically report the results.

Only GO BP terms with p-value smaller than 10^{-5} and q-value smaller than 0.05 were considered while threshold for KEGG enrichment was more lenient (p-value≤0.05). Finally results obtained are compared and integrated with ChIPseq data analysis.

3. Results

This chapter is divided in two main sections related to PAT-ChIPseq data analysis and RNA-seq data analysis. Here we summarize the main results that are then detailed and dis≥cussed in the following paragraphs.

1. **HF diet induces an overall significantly-higher mean-percentage of H3K4me3 peaks in promoter regions, as compared to CR** (ANOVA test, p-value0.05) and, at the same time, a significantly-lower mean percentage of peaks in distal intergenic regions. After identifying subsets of peaks commonly shared by biological replicas of the same diet group ("solid" peaks), **HF still showed a higher percentage of promoter peaks than CR and control groups (~90% versus ~75%) although the numerosity of solid peaks for the three groups was almost the same (~3000 peaks).** This means that **HF produces specific changes in chromatin conformation with respect to the other regimens, "opening", on average, more promoter regions than SD and CR.** This could result in an aberrant regulation of some genes since H3K4me3 signal correlates mostly with active transcription.
2. **Despite the presence of a moderate technical and biological variability of PAT-ChIPseq samples** (evaluated in terms of differences in final reads, number of called peaks, general enrichment, genomic localizations of peaks and intensity of the signal), after proper normalization, **data analyses supported the existence of diet-specific epigenetic signatures**,

detectable by differential analysis of H3K4me3 signal, allowing the clustering of samples by diet-group.

3. **Regions showing an increased level of H3K4me3 in CR** with respect to SD, **corresponds to genes involved in Circadian rhythmicity**. Moreover, the **motif** of a known chromatin modifier, **NRSF/REST is found enriched in these regions** (non significantly, because of the low number of sites). It has been shown that **higher REST levels** in brain **correlate with longevity and healthy aging, two features of CR beneficial effect**. Furthermore, after the analysis of anti-REST ChIPseq, **we showed that this factor binds, in liver, promoters of key genes involved in major metabolic processes**.
4. **Regions showing an increased level of H3K4me3 in HF with** respect to SD, **corresponds to genes involved in onset of Type II diabetes mellitus**. Moreover, the **motif** of a **ZSCAN4**, **a transcription factor involved in telomere elongation** in ES cells, **was found enriched in these regions** (non significantly, because of the low number of sites). **Telomere shortening** is known to be a characteristic of aging. In particular, it has been shown that telomere shortening **is a risk factor for type II diabetes mellitus.**
5. Although very preliminary and based on a smaller subset samples, the analysis of RNA-seq data reported **1181 genes significantly differentially expressed in CR vs SD**, almost **equally divided between over- and under-expressed.** All together they enrich specific pathways coherent with findings reported in literature including PPAR signaling pathway and Circadian rhythm. Moreover **all genes related to elevated level of H3K4me3 signal in CR are significantly overexpressed**, **confirming**

that CR modulates liver circadian clock through changes in H3K4me3 signal.

6. Through the identification and annotation of **H3K27me3 peaks present in a specific diet condition and absent in the control, data indicates** that **Olfactory transduction and Natural Killer mediated cytotoxicity pathway** result **impaired respectively in HF and CR**. These results are corroborated by literature findings.

3.1. PAT-ChIPseq data analysis

In this paragraph we report results obtained from the bioinformatic analyses of data produced by PAT-ChIPseq experiments with the pipeline described in the previous chapter.

The first part encloses a descriptive analysis of the collected datasets; PATChIP experimental variability is analysed in terms of number of final reads, number of called peaks and peaks genomic distribution.

Then two approaches are used: a “positional” method, focused on absence/presence of peaks in a diet condition respect to the control one, and a “quantitative” method, based on the statistical comparison of read-density information on a consensus peak-set along different samples.

Finally, to add insights on possible mechanisms through which diets modulate disease risk/prevention we performed pathway- and motif discovery analyses and we compared our results with state of the art literature.

3.1.1. Assessing biological and technical variability in PAT-ChIPseq replicas

3.1.1.1. Preprocessing and peak calling results

To gain insight into the effects of different diets in the epigenetic organization of the mouse genome we examined levels of trimethylated H3K4 and H3K27 in paraffin embedded liver samples using PAT-ChIP-seq.

So far we have analysed a total of:

- 19 samples for the anti-H3**K4**me3 PAT-ChIP
- 19 samples for the anti-H3**K27**me3 PAT-ChIP.

In particular, for each histone marks, we obtained 6 samples for the CR, 5 from the HF and 8 from the SD group of mice.

The pre-processing results are reported in Table 3.1: on average we obtained 9,803,500 (± 2,170,478) and 15,292,946 (± 3,382,450) final reads for H3K4me3 and H3K27me3 samples, respectively. After peak calling using SICER with default parameters, for each sample we calculated the fraction of all mapped reads that fall into peak regions as identified by the peak caller. The fraction of reads falling within peak regions is a first-cut metric to measure the success of the immunoprecipitation, since it is related to the signal-to-noise ratio of the experiment, and it is called FRiP (Fraction of Reads in Peaks). Generally, a FRiP value greater than 1-2% suggest that the ChIP-seq experiment lead to the enrichment of specific genomic regions (**Landt et al, 2012**).

A comparison of total mapped reads and called peaks distributions for both histone marks divided by diet group is plotted in Figure 3.1. We observed a slight trend to a higher number of final reads in the H3K27me3 experiments compared to those of the H3K4me3, but the number of final reads obtained for this set of

experiments is yet considered largely acceptable. The internal variability in each diet condition appears to be modest (from 2.7 to 4.0 millions reads std dev). However, these dissimilarities were taken in account in the subsequent steps of analyses. The number of peaks called by SICER seems to be more variable for the H3K4me3 samples than for the H3K27me3.

	H3K27me3				H3K4me3			
Sample name	H3K27me3 mapped reads	Evalue	H3K27me3 SICER peaks	FRiP	H3K4me3 mapped reads	Evalue	H3K4me3 SICER peaks	FRiP
CR4	9,617,067	100	4,030	19%	-	-	-	-
CR6	17,166,587	100	7,896	20%	7,702,439	100	13,931	6%
CR8	13,170,737	100	8,069	17%	7,896,093	**1000**	19,230	7%
CR9	12,032,726	100	108	0%	5,387,349	**1000**	15,633	6%
CR78	15,925,926	100	5,637	23%	9,483,331	**1000**	15,743	6%
CR79	14,051,089	100	7,592	21%	8,413,803	100	23,744	14%
CR81	-	-	-	-	7,573,374	**1000**	12,918	5%
mean	13,660,689		5,555	17%	7,742,732		16,867	8%
std dev	2,713,530		3,097	8%	1,347,715		3,997	3%
HF64	9,488,243	100	3,584	16%	7,252,524	100	13,708	9%
HF65	16,800,573	100	9,975	25%	9,439,119	100	17,843	11%
HF67	13,413,108	100	8,652	20%	9,902,328	100	15,566	6%
HF68	17,739,656	100	5,538	15%	13,125,661	**1000**	9,871	3%
HF70	13,970,800	100	6,504	17%	11,109,700	100	12,406	6%
mean	14,282,476		6,851	19%	10,165,866		13,879	7%
std dev	3,245,401		2,526	4%	2,164,277		3,034	3%
SD21	19,498,066	100	6,572	27%	8,756,851	100	15,497	5%
SD22	14,059,749	100	7,988	20%	11,191,360	100	11,198	3%
SD71	18,227,651	100	8,221	25%	12,892,666	100	26,869	9%
SD72	11,141,736	100	5,921	17%	10,306,570	100	14,512	5%
SD73	17,927,562	100	7,872	24%	9,244,469	100	11,072	4%
SD74	10,939,879	100	6,223	15%	11,700,447	100	12,608	5%
SD75	22,507,954	100	6,408	25%	12,733,191	100	10,609	3%
SD76	17,210,991	100	5,845	14%	12,155,232	100	12,078	5%
mean	16,439,199		6,881	21%	11,122,598		14,305	5%
std dev	4,073,667		982	5%	1,554,815		5,358	2%

Table 3.1 Sequencing and peak calling results of H3K27me3 and H3K4me3 ChIP-seqs by diet groups.
Different colours are identifying different diets regimen: green for Calorie Restriction, red for High Fat Diet and blue for Standard Diet samples; columns report (from left to right): number of mapped reads, E-value applied for peak calling with SICER, number of identified peaks using SICER and FRiP index.
E-values used for each sample are reported in Table 3.1 together with the corresponding final FRiP index. E-value was setted at 1000 for those samples that initially, with an E-value=100, displayed a FRIP ~1%.
The CR9 anti-H3K27me3 sample, denoted in red in Table 3.1, because of its very poor enrichment (108 peaks, 0% FRiP), has been discarded from downstream analysis.

The slight differences in total reads, called peaks and signal enrichment (FRiP) among biological replicas can be explained by intrinsic experimental complexities (i.e. adapted PAT-ChIP protocol for the liver tissue, differences in liver histology

among different diet groups) and yet indicate that we reached a satisfactory yield of ChIPped sequenceable material and the quality of the enrichment overall is adequate.

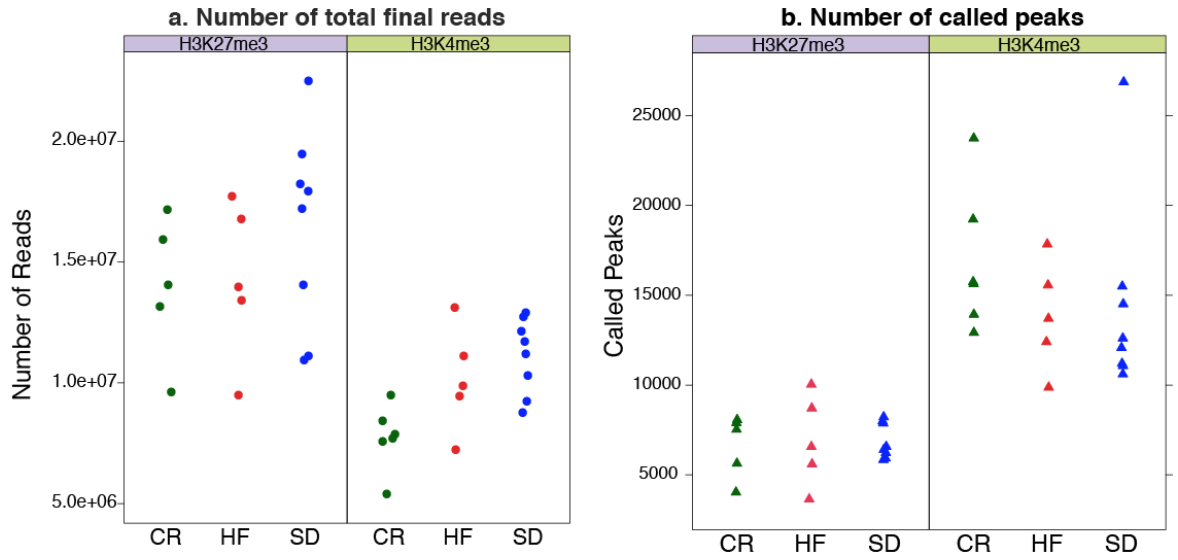


Figure 3.1: Features distribution of H3K27me3 and H3K4me3 Histone Marks by Diet groups.

a) Number of total final reads, b) number of called peaks. In green the Calorie Restriction samples, in Red the High Fat Diet samples, in Blue the Standard Diet samples

In particular:

- in the H3K27me3 dataset, there are no relevant differences in read counts among the three diet groups (Fig. 3.1, panel a, left section), all showing an overall high-number of total reads; SD and HF groups roughly exhibit the same distribution of numbers of called peaks (Fig. 3.1, panel b, left section), while for CR group the variance is higher.
- in the H3K4me3 samples, the distribution of numbers of total reads is similar for HF and SD samples, while for CR group variance and overall numbers are lower than in the other groups (Fig. 3.1, panel a, right section). The distributions of numbers of called peaks, not considering some outliers, are comparable among the three diet-groups, even if the peaks identified for the HF group were less, on average, than in the other two groups. On

the other hand the variance in HF samples is lower with respect to the others (Fig. 3.1, panel b, right section).

We then analysed the length (in base pairs) of the identified peaks, for each histone modification and diet group (Fig. 3.2). Length distributions are quite similar among diet groups for both markers (average of 1,800 and 35,000 bp for H3K4me3 and H3K27me3 peaks, respectively), with large ranges (from ~200 to ~3500 bp for H3K4me3 and from ~500 bp to almost 100 Kbp for H3K27me3 peaks). The marked difference in average length between the H3K4me3 and H3K27me3 peaks is due to the different genomic regions targeted by the two histone markers (genes' transcription start sites *vs.* intergenic regions, cf. paragraph 1.2.2).

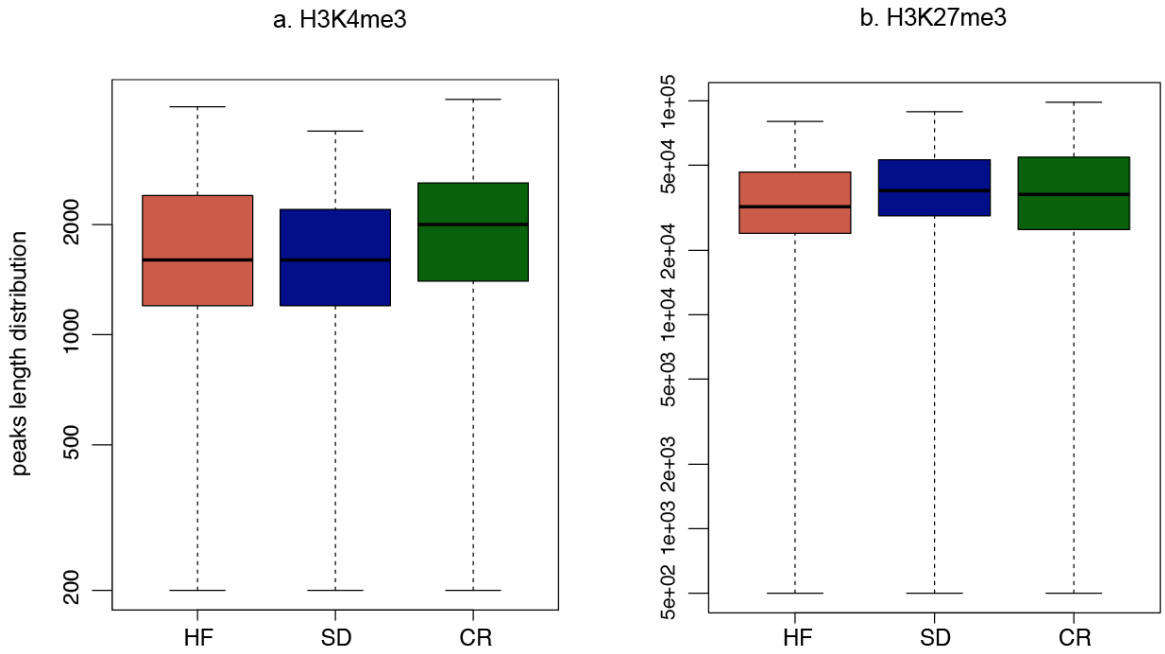


Figure 3.2 Peaks length distribution by histone modification and by diet group.

Peaks length distributions for each histone modification are quite similar among diet groups, with large ranges (~200-3500 bp for H3K4me3 peaks and ~500 bp-100 Kbp for H3K27me3 peaks). In particular the median value is ~1800 bp for H3K4me3 peaks while it is much higher for H3K27me3, being ~35 Kbp. This difference is partially expected since the two markers target mainly different regions of the genome (genes' TSSs vs intergenic regions).

3.1.1.2. H3K4me3: diet-group internal and inter-group variability analyses

Internal and inter-diets variability of H3K4me3 peaks is reported in Table 3.2 and Fig. 3.3, which shows the genomic peak-distribution for each biological replica, obtained annotating peaks with ChIPseeker (**Guangchuang et al, 2015**): diet groups are denoted with different colors (green for CR, red for HF and blue for SD) and different genomic regions are individually considered: Promoter regions (interval of 5 Kb centered on TSS), 5' and 3' UTRs, Exonic regions (divided in 1st Exon and Other Exon), Intronic regions (divided in 1st Intron and Other Intron), Downstream regions (less than 3 Kb from the gene end), Distal Intergenic (outside genes, more than 3 Kb far from them).

Sample	Promoter	5UTR	3UTR	1st Exon	Other Exon	1st Intron	Other Intron	Downstream	Distal Intergenic
CR6	15.90	1.26	1.37	0.32	2.28	3.68	11.28	1.01	62.90
CR78	60.55	1.00	1.19	0.60	1.97	2.91	6.90	0.80	24.08
CR79	35.90	1.49	1.31	0.88	1.83	2.70	8.26	0.76	46.88
CR81	53.05	1.55	1.80	0.83	2.55	2.88	9.14	0.88	27.32
CR8	11.15	0.82	1.21	0.35	1.97	3.48	12.16	0.70	68.16
CR9	44.33	0.79	1.20	0.39	2.07	3.54	10.09	0.81	36.77
mean	36.81	1.15	1.35	0.56	2.11	3.20	9.64	0.83	44.35
std dev	19.90	0.33	0.23	0.25	0.26	0.41	1.94	0.11	18.31
HF64	67.59	1.67	1.34	0.80	1.61	2.31	5.08	0.69	18.92
HF65	60.70	1.63	1.53	0.87	1.98	3.09	7.31	0.80	22.09
HF67	56.99	1.19	1.56	0.58	1.88	2.93	7.17	0.87	26.84
HF68	60.76	1.40	1.56	0.60	2.29	2.80	6.57	0.80	23.23
HF70	79.73	1.17	1.00	0.85	1.71	2.06	4.24	0.67	8.58
mean	65.15	1.41	1.40	0.74	1.89	2.64	6.07	0.76	19.93
std dev	9.00	0.24	0.24	0.14	0.27	0.44	1.35	0.08	6.95
SD21	19.71	0.68	1.27	0.17	2.03	3.63	12.22	0.83	59.45
SD22	26.87	1.25	1.84	0.38	2.48	3.98	11.99	0.88	50.32
SD71	25.20	0.80	0.89	0.45	1.63	2.54	9.48	0.62	58.40
SD72	51.57	1.32	1.29	0.59	1.72	2.85	7.66	0.64	32.35
SD73	68.91	0.94	1.23	0.51	1.73	2.62	6.49	0.79	16.77
SD74	66.16	1.33	1.56	0.77	2.03	2.44	6.50	0.63	18.58
SD75	40.46	1.34	1.62	0.52	2.13	3.14	9.02	0.68	41.09
SD76	64.08	1.56	1.61	0.99	2.04	2.84	6.92	0.79	19.16
mean	45.37	1.15	1.42	0.55	1.97	3.01	8.79	0.73	37.01
std dev	20.02	0.31	0.30	0.25	0.28	0.55	2.32	0.10	17.90

Table 3.2 H3K4me3 peaks genomic distribution divided by diet group

In green CR samples, in red HF samples and in blue SD samples. The promoter class includes all peaks falling in the interval [TSS-2.5kb,TSS+2.5kb]. In Promoter and Distal Intergenic classes, for HF group the variability seems to be modest (Std. dev. respectively ~9%-7%) while for CR and SD groups we have much higher standard deviations (~20% for Promoter region - ~18% for Distal Intergenic) and the mean percentage of peaks in Promoter class is much higher HF group than for CR and SD (65% vs 37% and 45%).

For each sample and each genomic region, the percentage of peaks falling in the specific class is reported. In particular the promoter class includes, for each gene, a genomic interval around the Transcription Start Site [$TSS - 2.5Kb, TSS + 2.5Kb$]. From the graphical description of the same data (Fig 3.3), it is possible to notice that the highest variability is concentrated in Promoter and Distal Intergenic classes. In particular, for HF (first panel) the variability seems much lower than CR (second panel) and SD (third panel).

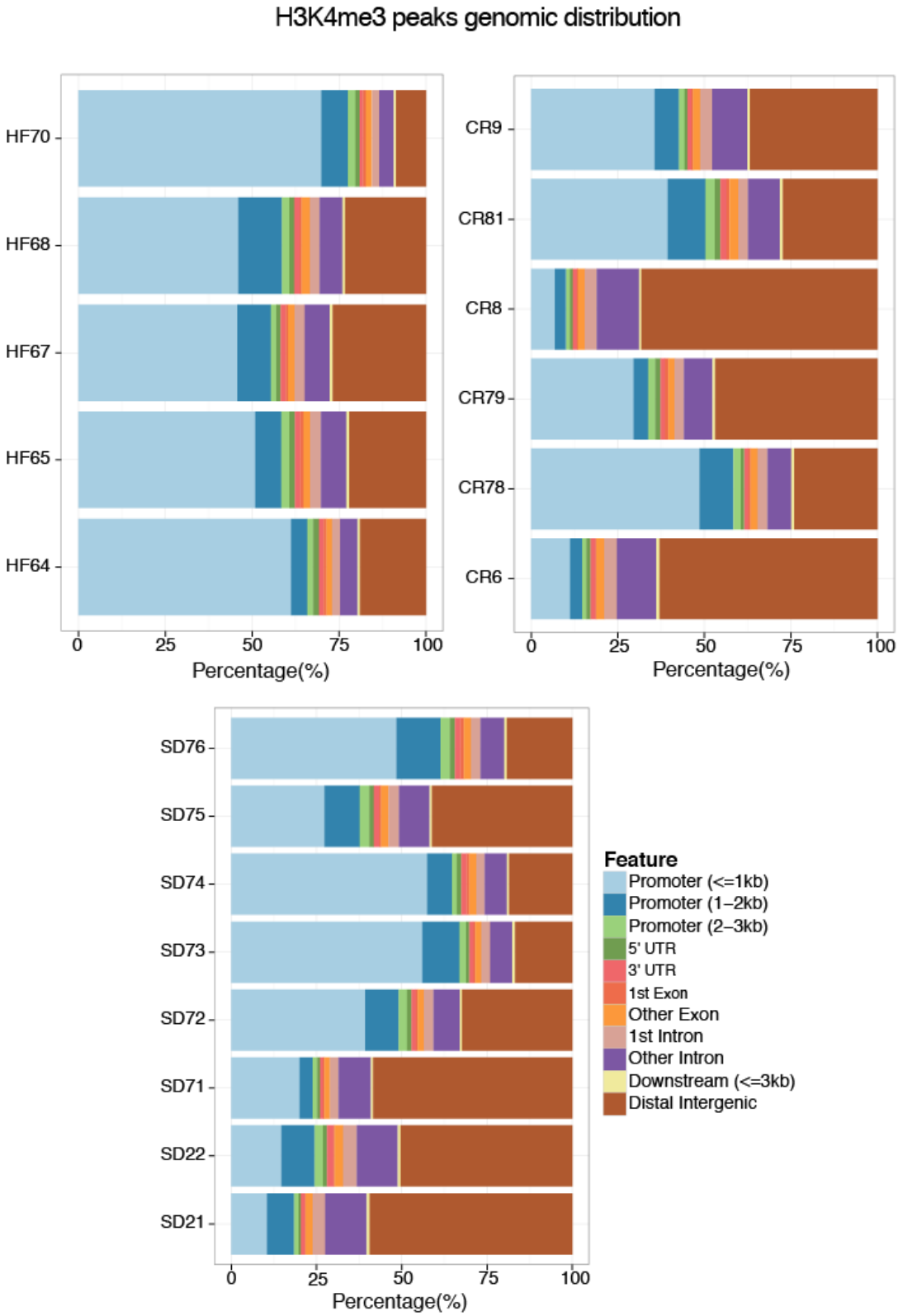


Figure 3.3 H3K4me3 peaks Genomic Annotation distribution by sample.

First panel - HF samples; Second panel - CR samples; Third panel - SD samples.
HF group seems to be much less variable than the other diet groups, showing for all replicas more or less the same percentage of peaks in each genomic class. In particular, focusing on Promoter regions, HF replicas range is 57%-79%, while CR goes from 11% to 60% and SD from 20% to 68%. These differences could be due to technical issues due to PATChIP protocol applied to liver tissues or they could reflect a real biological variability among replicas.

We then computed for each diet group the similarity distance (i-by-k) matrix defined as:

$$J(d) = \begin{bmatrix} J_{11} & \cdots & J_{1n_d} \\ & \vdots & \\ J_{n_d1} & \cdots & J_{n_dn_d} \end{bmatrix}$$

where:

$$d = \mathrm{HF, SD, CR}$$
$$n_d = \text{number of samples of diet group d}$$
$$J_{ik} = 1 - \frac{|S_i \cap S_k|}{|S_i \cup S_k|}$$
$$i,k = 1,\ldots,n_d$$

where i and k denote *i-th* and *k-th* sample of diet *d*, S_i represents the set of peaks of *i-th* sample of diet *d* and, in $J_{ik,}$ the numerator is the cardinality of the set of peaks shared by samples *i* and *k* (the number of overlapping peaks between the two samples), while the denominator is the cardinality of the union of peak sets *i* and *k* (the sum of the number of peaks of the two samples, minus the number of the common peaks). J_{ik} is called "Jaccard similarity distance" between *i-th* and *k-th* sample.

Smaller is this distance between two samples, more similar are the two samples (since they share more peaks). To better appreciate internal variability, we plotted the heatmaps of the three diet-group similarity matrices in Figure 3.4.

According to our results, the HF group (Fig. 3.4-a) seems to be the more stable dataset (max dist. value is 0.6; 4 out of 5 are similar between each other) while CR and SD (Fig.3.4-b and c) groups of samples are more variable (max dist. value is greater than 0.8; for CR 4 out of 6 are more similar between each other; for SD 5 out of 8 are more similar between each other).

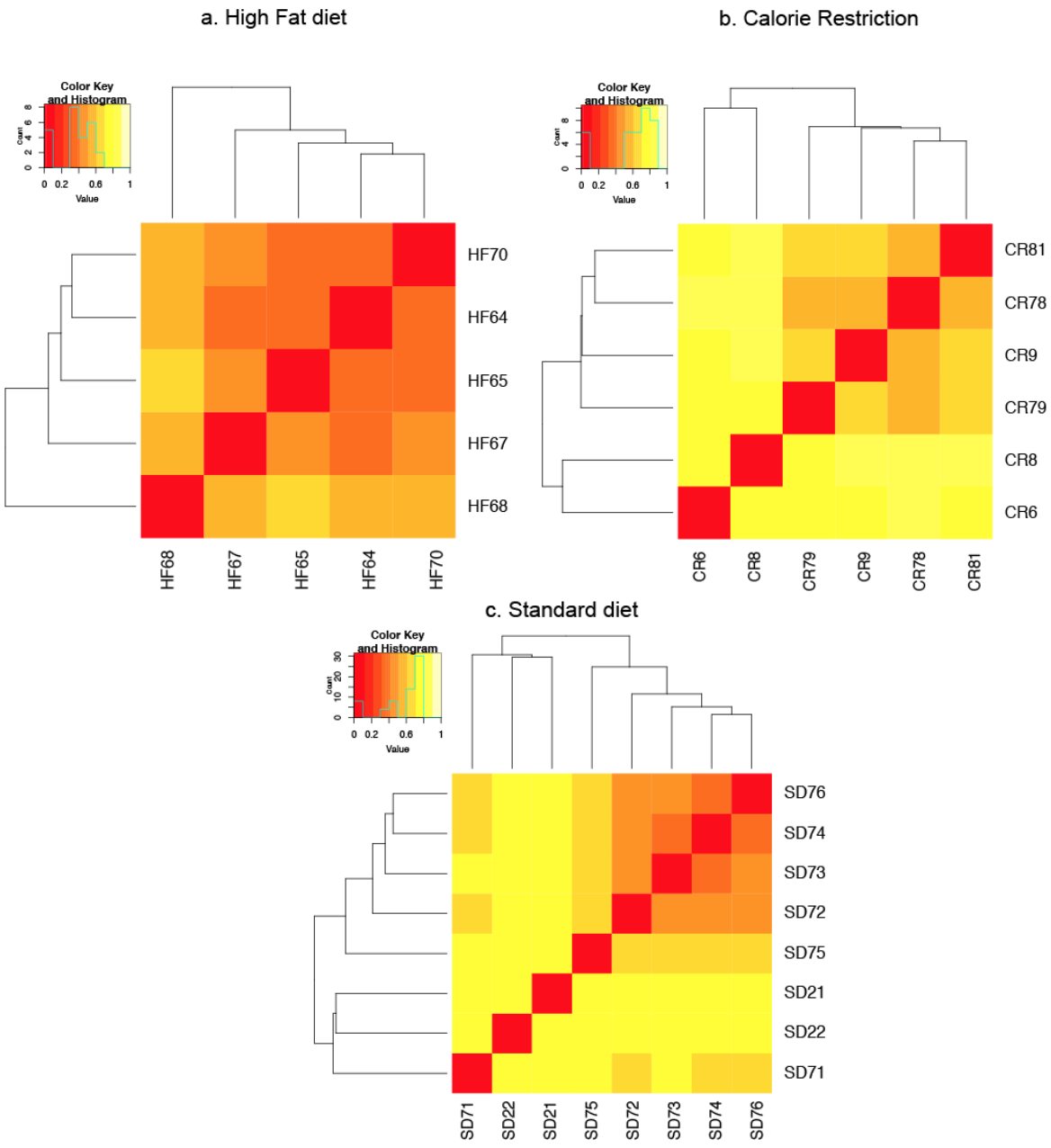


Figure 3.4 Jaccard similarity matrix heatmaps for diet group - H3K4me3

For each diet, we computed Jaccard similarity distance based on the peaks shared between samples of the same diet group. Smaller is the distance, grater is the peak overlap between the two samples.
HF group (a) seems to be the less variable (max dist. value is 0.6; 4 out of 5 are similar between each other) while CR and SD are more variable (max dist. value is greater than 0.8; for CR group, 4 out of 6 are more similar between each other; for SD 5 out of 8 are more similar between each other).

Focusing on Promoter regions, to assess the statistical significance of inter-groups mean-differences, we used the ANOVA test (that compares diet groups internal variability with the variability across the groups).

First, we performed the Shapiro test to assess the normality of peaks percentage-distributions for the promoter regions in each diet group:

```
> shapiro.test(allstats$allProm[allstats$group=="SD"])

        Shapiro-Wilk normality test

data:  allstats$allProm[allstats$group == "SD"]
W = 0.88609, p-value = 0.2151

> shapiro.test(allstats$allProm[allstats$group=="HF"])

        Shapiro-Wilk normality test

data:  allstats$allProm[allstats$group == "HF"]
W = 0.8682, p-value = 0.2592

> shapiro.test(allstats$allProm[allstats$group=="CR"])

        Shapiro-Wilk normality test

data:  allstats$allProm[allstats$group == "CR"]
W = 0.93086, p-value = 0.5868
```

Being the p-values all greater than 0.05, we failed to reject the null hypothesis, and assume that the distributions are normal.

We then performed the Bartlett's test to check homogeneity of variances, in which the null hypothesis assumes that the variances are equal across groups:

```
> bartlett.test(allstats$allProm~allstats$group)

        Bartlett test of homogeneity of variances

data:  allstats$allProm by allstats$group
Bartlett's K-squared = 2.59, df = 2, p-value = 0.2739
```

Being the p-value greater than 0.05, we failed to reject the null hypothesis and we assumed the homogeneity of variances.

Finally, we performed the ANOVA test:

```
> anova(lm(allstats$allProm~allstats$group))
Analysis of Variance Table

Response: allstats$allProm
               Df Sum Sq Mean Sq F value  Pr(>F)
allstats$group  2 2276.8 1138.42  3.5638 0.05247 .
Residuals      16 5111.0  319.44
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

which showed that indeed there is a statistically significant difference between groups (one-way ANOVA (F(2,16) = 3.5638, p = 0.05247).

To identify which were the groups showing significant difference, we performed the Tukey test:

```
> model<-lm(allstats$allProm ~ allstats$group)
> model

Call:
lm(formula = allstats$allProm ~ allstats$group)

Coefficients:
     (Intercept)  allstats$groupHF  allstats$groupSD
          36.814            28.339             8.558

> df<-df.residual(model)
> MSerror<-deviance(model)/df

> HSD.test(allstats$allProm,allstats$group, df, MSerror, group=FALSE,console=T)

Study: allstats$allProm ~ allstats$group

HSD Test for allstats$allProm

Mean Square Error:  319.4362

allstats$group,  means

   allstats.allProm       std r      Min      Max
CR         36.81422 19.902637 6 11.14983 60.55139
HF         65.15321  9.004773 5 56.98683 79.73398
SD         45.37242 20.021633 8 19.71477 68.90976

alpha: 0.05 ; Df Error: 16
Critical Value of Studentized Range: 3.649139

Harmonic Mean of Cell Sizes  6.101695
Comparison between treatments means

        Difference   pvalue sig.        LCL       UCL
CR - HF -28.338996 0.046448    * -56.264638 -0.413354
CR - SD  -8.558201 0.656152      -33.464595 16.348193
HF - SD  19.780795 0.159447       -6.510328 46.071918
```

A significant difference is scored between the CR and HF groups (p-value=0.04), suggesting that differences between these two groups, in the promoter regions, are higher than the internal variability.

We repeated these statistical analyses considering the Distal Intergenic class of peaks of H3K4me3 and we found the same results (not shown).

ANOVA test showed that HF diet induces an overall significantly-higher mean-percentage of H3K4me3 peaks in promoter regions, as compared to CR (p-value≤0.05) and, at the same time, a significantly-lower mean percentage of peaks in distal intergenic regions. No statistically significant differences in peaks genomic localization were scored between HF and SD and between CR and SD. Probably these differences exist but do not emerge for statistical reasons (low number of samples).

3.1.1.3. H3K27me3: diet-group internal and inter-group variability analysis

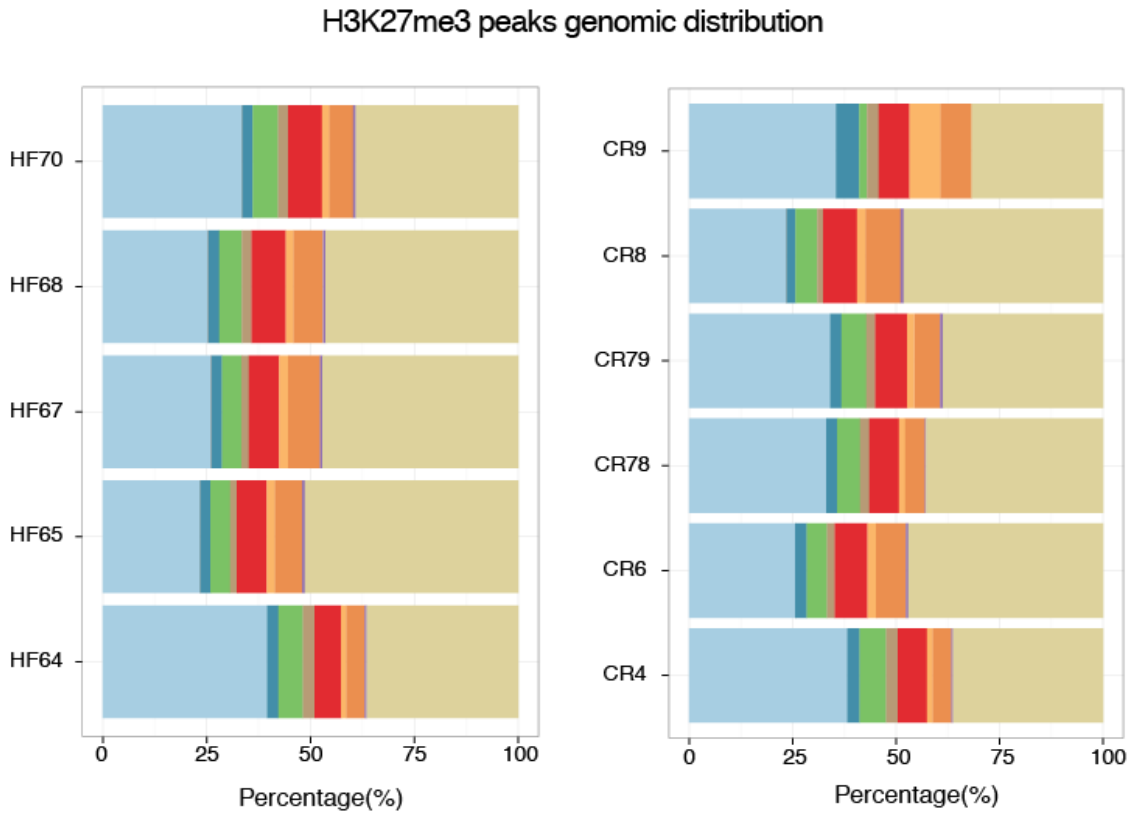
To analyse the internal variability in each diet group for H3K27me3, we report the genomic distribution of peaks for each biological replica in Table 3.3, and plotted results in Figure 3.5.

Sample	Promoter	5UTR	3UTR	1st Exon	Other Exon	1st Intron	Other Intron	Downstream	Distal Intergenic
CR8	23.59	2.16	5.21	1.43	8.18	2.11	8.40	0.76	48.18
CR6	25.62	2.79	4.94	1.87	7.85	2.08	7.19	0.61	47.04
CR78	33.13	2.63	5.57	2.25	7.15	1.45	4.77	0.35	42.69
CR79	34.04	2.71	6.02	2.16	7.75	1.73	6.26	0.69	38.65
CR4	38.17	3.00	6.43	2.80	7.05	1.44	4.52	0.35	36.24
CR9	35.51	5.61	1.87	2.80	7.48	7.48	7.48	0.00	31.78
mean	31.68	3.15	5.01	2.22	7.58	2.71	6.44	0.46	40.76
std dev	5.77	1.24	1.63	0.54	0.43	2.35	1.55	0.28	6.38
HF65	23.47	2.41	4.79	1.57	7.28	1.89	6.62	0.58	51.38
HF64	39.58	2.74	5.83	2.82	6.42	1.31	4.38	0.50	36.42
HF67	26.23	2.49	4.67	1.78	7.33	2.14	7.64	0.60	47.13
HF68	25.48	2.56	5.45	2.29	8.31	1.86	7.04	0.63	46.36
HF70	33.62	2.46	6.12	2.41	8.13	1.81	5.71	0.69	39.04
mean	29.67	2.53	5.37	2.18	7.49	1.80	6.28	0.60	44.07
std dev	6.74	0.13	0.63	0.50	0.76	0.30	1.27	0.07	6.16
SD72	32.91	2.47	5.81	2.45	7.67	1.93	6.22	0.63	39.93
SD21	32.51	2.71	4.98	2.12	6.50	1.22	4.64	0.49	44.85
SD22	24.04	2.29	5.22	1.58	7.59	1.89	6.92	0.81	49.66
SD71	24.49	2.54	5.36	1.79	7.91	1.84	6.39	0.62	49.06
SD73	30.69	2.68	5.64	2.19	7.22	1.51	6.12	0.55	43.40
SD74	22.92	2.60	5.06	1.86	7.92	2.22	7.86	0.66	48.89
SD75	38.93	2.67	5.96	2.58	6.09	1.11	4.01	0.64	38.02
SD76	30.56	2.36	6.25	2.14	8.37	2.09	6.23	0.55	41.46
mean	29.63	2.54	5.54	2.09	7.41	1.72	6.05	0.62	44.41
std dev	5.48	0.16	0.45	0.33	0.77	0.40	1.22	0.10	4.47

Table 3.3 H3K27me3 peaks genomic distribution divided by diet group

In green CR samples, in red HF samples and in blue SD samples. The promoter class includes all peaks falling in the interval [TSS-2.5kb,TSS+2.5kb]. The diet internal variability for each diet group and each genomic class is low and the same is true for the variability among diet groups.

It can be noticed from standard deviations and means values in Table 3.3 (and visually in Figure 3.5) that both internal and inter-diets variability is very low.



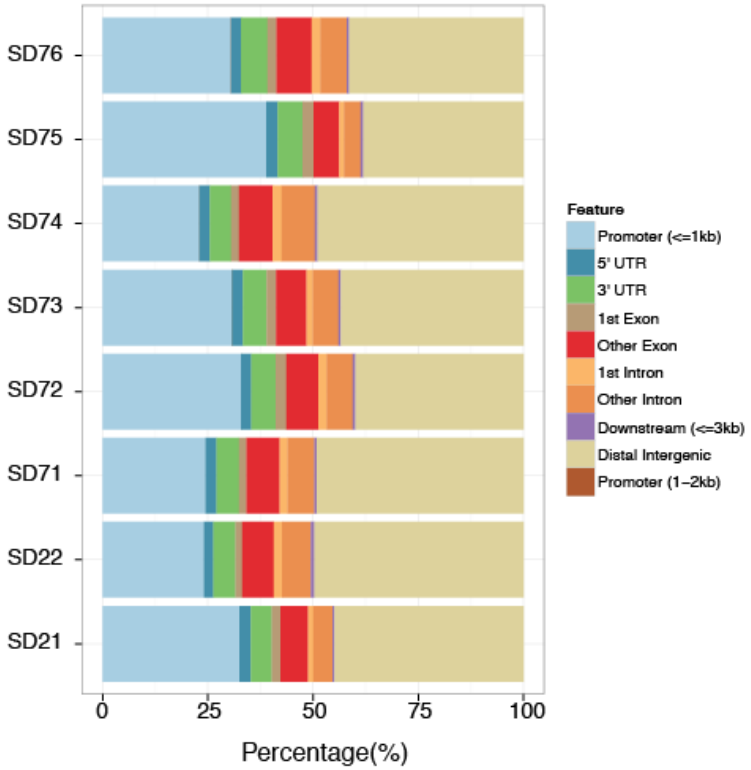


Fig. 3.5 H3K27me3 peaks Genomic Annotation distribution by sample.

First panel - HF samples; Second panel - CR samples; Third panel - SD samples.
Compared to H3K4me3, the distribution of peaks in the genomic classes for H3K27me3 peaks is much less variable both intra-diet that inter-diets.Most of the peaks are located in intergenic regions as expected (~30%-50%) and promoter regions (~20%-40%) as often H3K27me3 signal co-localizes with H3K4me3 in Polycomb targets.

For each diet group, we then computed the similarity distance matrix *J(d)* and plotted the relative heatmap (Figure 3.6), as described in paragraph 3.2.1.

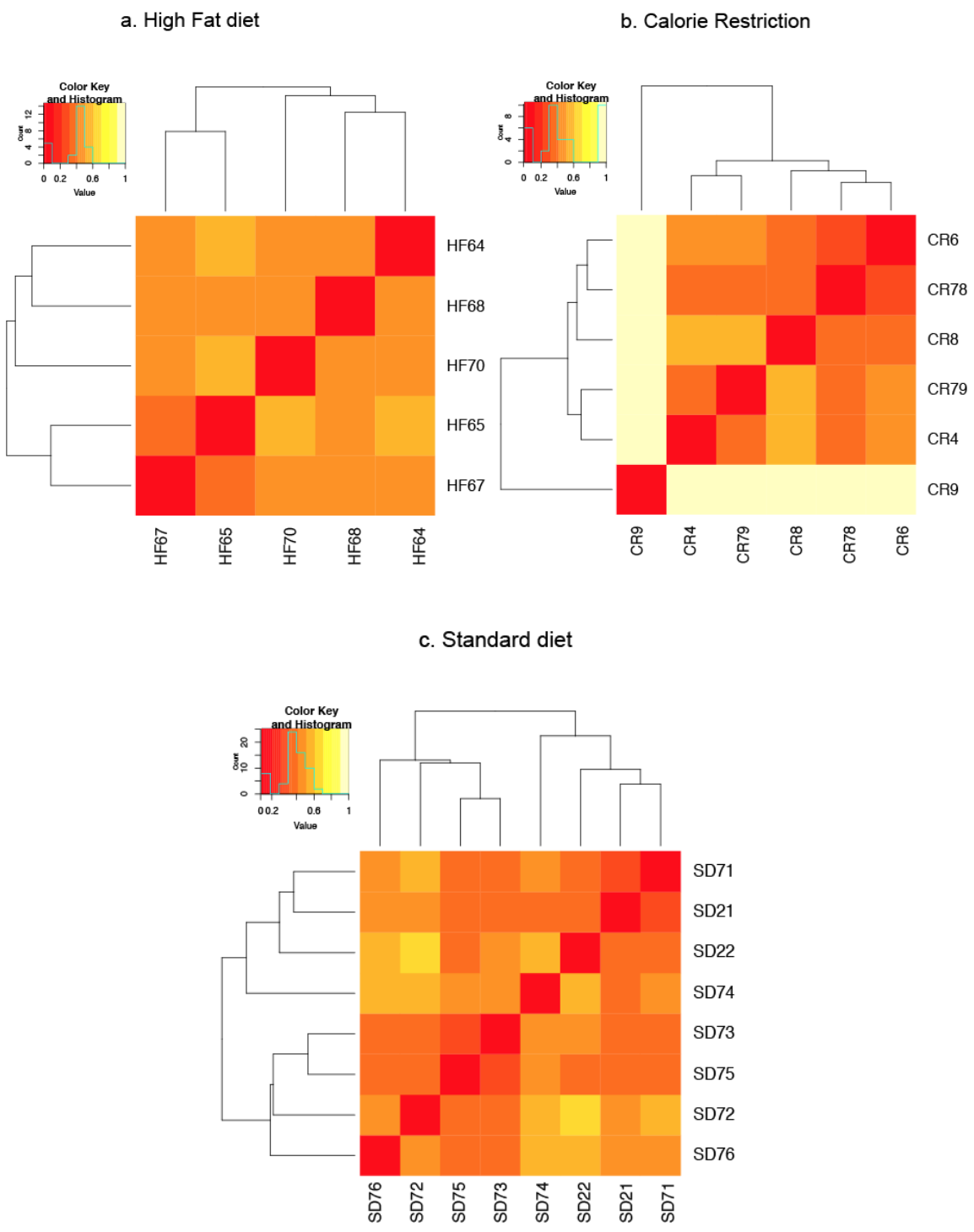


Figure 3.6 Jaccard similarity matrix heatmaps for diet group - H3K27me3

For each diet, we computed Jaccard similarity distance based on the peaks shared between samples of the same diet group. Smaller is the distance, grater is the peak overlap between the samples. In all diet groups the concordance on called peaks seems to be very high. The only exception is CR9 that seems to be very different from all the other CR samples: this sample has very few peaks and it will be discarded in the downstream analysis.

Results showed a much higher concordance of the called H3K27me3 peaks among samples for each diet group, as compared to the H3K4me3 peaks. The

only exception is the CR9 samples, which was very different from all others CR samples. However, this sample showed a very low number of called peaks and was not included in the subsequent analyses. Since the H3K27me3 marks are known to be enriched in heterochromatin regions (cf. paragraph 1.2.2), we performed the ANOVA test prioritizing analyses of Distal Intergenic Regions:

```
> shapiro.test(allstatsK27[allstatsK27$group=='CR',]$DistalIntergenic)

        Shapiro-Wilk normality test

data:  allstatsK27[allstatsK27$group == "CR", ]$DistalIntergenic
W = 0.92904, p-value = 0.5727

> shapiro.test(allstatsK27[allstatsK27$group=='SD',]$DistalIntergenic)

        Shapiro-Wilk normality test

data:  allstatsK27[allstatsK27$group == "SD", ]$DistalIntergenic
W = 0.91363, p-value = 0.3803

> shapiro.test(allstatsK27[allstatsK27$group=='HF',]$DistalIntergenic)

        Shapiro-Wilk normality test

data:  allstatsK27[allstatsK27$group == "HF", ]$DistalIntergenic
W = 0.93186, p-value = 0.6091


> bartlett.test(allstatsK27$DistalIntergenic~allstatsK27$group)

        Bartlett test of homogeneity of variances

data:  allstatsK27$DistalIntergenic by allstatsK27$group
Bartlett's K-squared = 2.1289, df = 2, p-value = 0.3449

> anova(lm(allstatsK27$DistalIntergenic~allstatsK27$group))
Analysis of Variance Table

Response: allstatsK27$DistalIntergenic
                  Df Sum Sq Mean Sq F value Pr(>F)
allstatsK27$group  2 103.05  51.525  1.2828 0.3043
Residuals         16 642.66  40.166
```

No differences among diet group means were scored. Analysis on Promoters was also performed (not shown) and gave the similar results.

Together, these data suggest that probably diet regimens does not affect H3K27me3 peaks genomic distribution.

3.1.2. Downstream analysis of H3K4me3 and H3K27me3 signals

We then analysed H3K4me3 ChIP-seq datasets to identify differences among peaks across diet conditions. We used two approaches:

I. a "positional" approach, where analyses are based on comparison of peak-mapping information to identify common or specific enriched-regions;
II. a "quantitative" method, based on the statistical comparison of read-density information.

3.1.2.1. The “positional” approach

Since we have estimated a certain degree of internal variability, we cannot trust all peaks identified in all samples. Thus, we first generated a unique not-redundant peak-dataset from the pool of the peaks identified in all samples, for each diet group (SD, HF and CR) using BEDtools (**Quinlan et al, 2010**). Then, to measure peak concordance among replicas, we plotted the number of common peaks as a function of the number of samples sharing those peaks (Fig. 3.7). Finally, we consider a peak to be "**solid**", for a given diet group, only if common to at least a certain number of samples of the same group. This number was calculated, for each diet group, with the Elbow method (**Thorndike, 1953**), often used to calculate the number of clusters to perform cluster analysis. Accordingly, the threshold is chosen at the angle point of the curve, point in which the variance reaches a plateau; Fig. 3.7). For the H3K4me3 datasets (panel a), the plateau is reached at 4 samples for all diet groups, allowing unambiguous assigning of 3,703 peaks for HF, 3,202 peaks for CR and 3,517 peaks for SD. For the H3K27me3 datasets (Fig. 3.7, panel b) the plateau is instead reached at 3 samples for HF (3,092 peaks) and CR (3,243 peaks) diet groups, and at 4 samples for SD (3,505 peaks). Now on, otherwise specified, this set of “solid” peaks was considered in

the all analyses. We annotated all regions with the ChIPseeker R package, using the RefSeq table. The genomic distribution of the identified K4 and K27 peaks in each diet group is reported in Fig. 3.8.

We first analysed the genomic distribution of the called peaks in the three diet-groups. With respect to the H3K4me3, we noticed an higher percentage of peaks on promoter regions in the HF diet-group (~90% of the total, as compared to ~72% in SD and ~76% in CR). Notably, the number of H3K4me3 “solid” peaks identified in each diet group is similar (~3,000), strengthening the relevance of the larger percentage of promoter peaks in the HF group, and suggesting that HF produces specific changes in chromatin conformation, “opening”, on average, more promoter regions, as compared to SD and CR. This effect could produce altered and aberrant levels of transcription of specific genes, thus contributing to the development and progression of different cancers or other diseases like diabetes and cardiovascular diseases (**Ke et al, 2009; He C et al, 2012**; **Chen Z et al, 2010**; **Raciti et al, 2014**; **Mathiyalagan et al, 2014,** cf. 1.2.5. in the introduction).

For H3K27me3, promoter and Intergenic regions were the most enriched (~50% and ~30%, respectively). However, we observed no differences for each genomic regions analysed among the different diet group (Fig. 3.8, panel b).

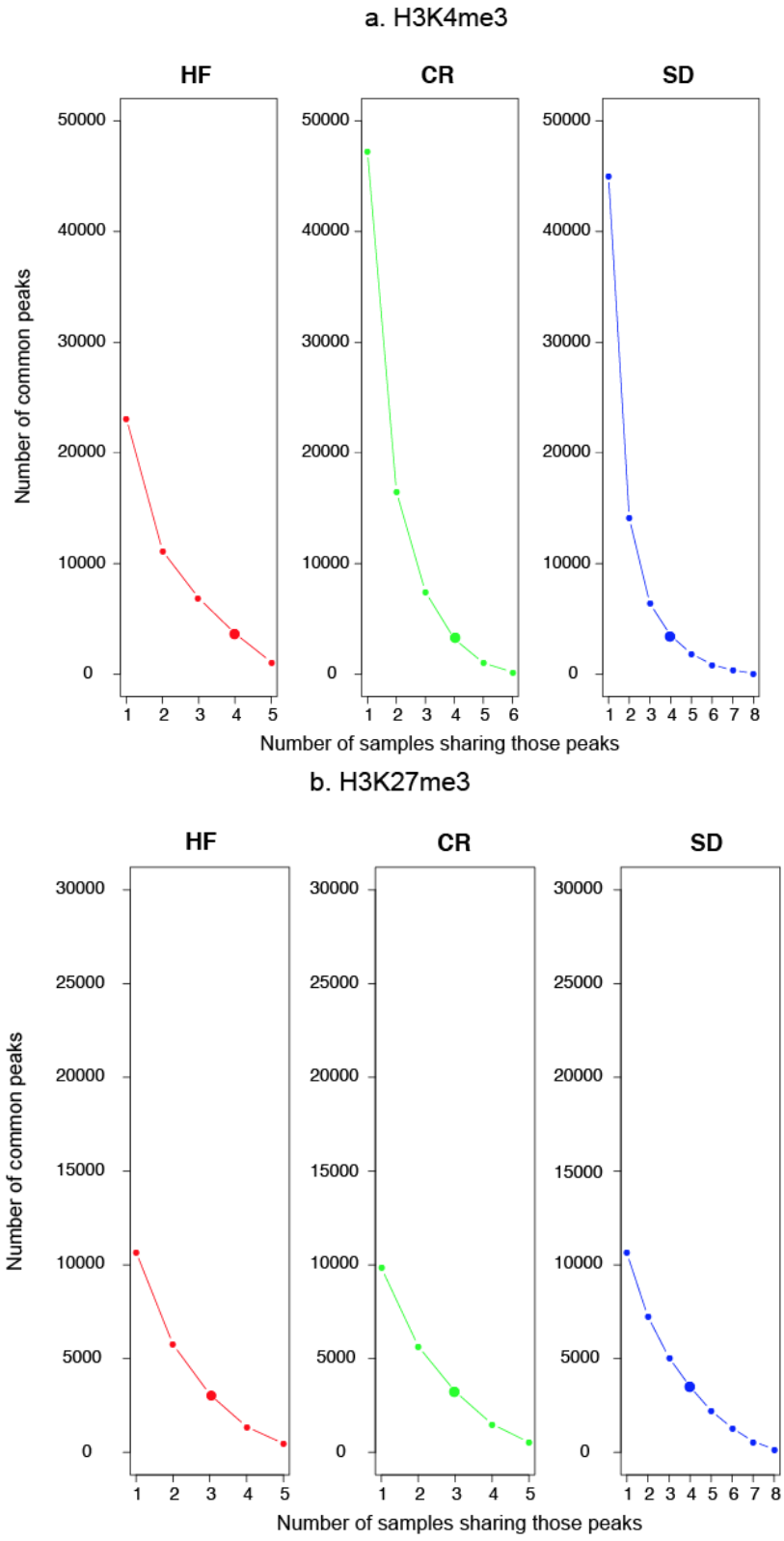


Figure 3.7 Peak calling concordance at different values of samples intersection.

For both histone modifications (H3K4me3 and H3K27me3) and each diet group (HF, CR and SD) and for each value of the intersection (1-8), the number of peaks common to that number of samples is reported. Each curve represent the variance in terms of shared peaks as function of the number of the samples that have those peaks in common.The chosen threshold in each plot is indicated with a bigger dot and it is chosen when there is an angle in the curve, meaning that we reached a plateau in the variance.

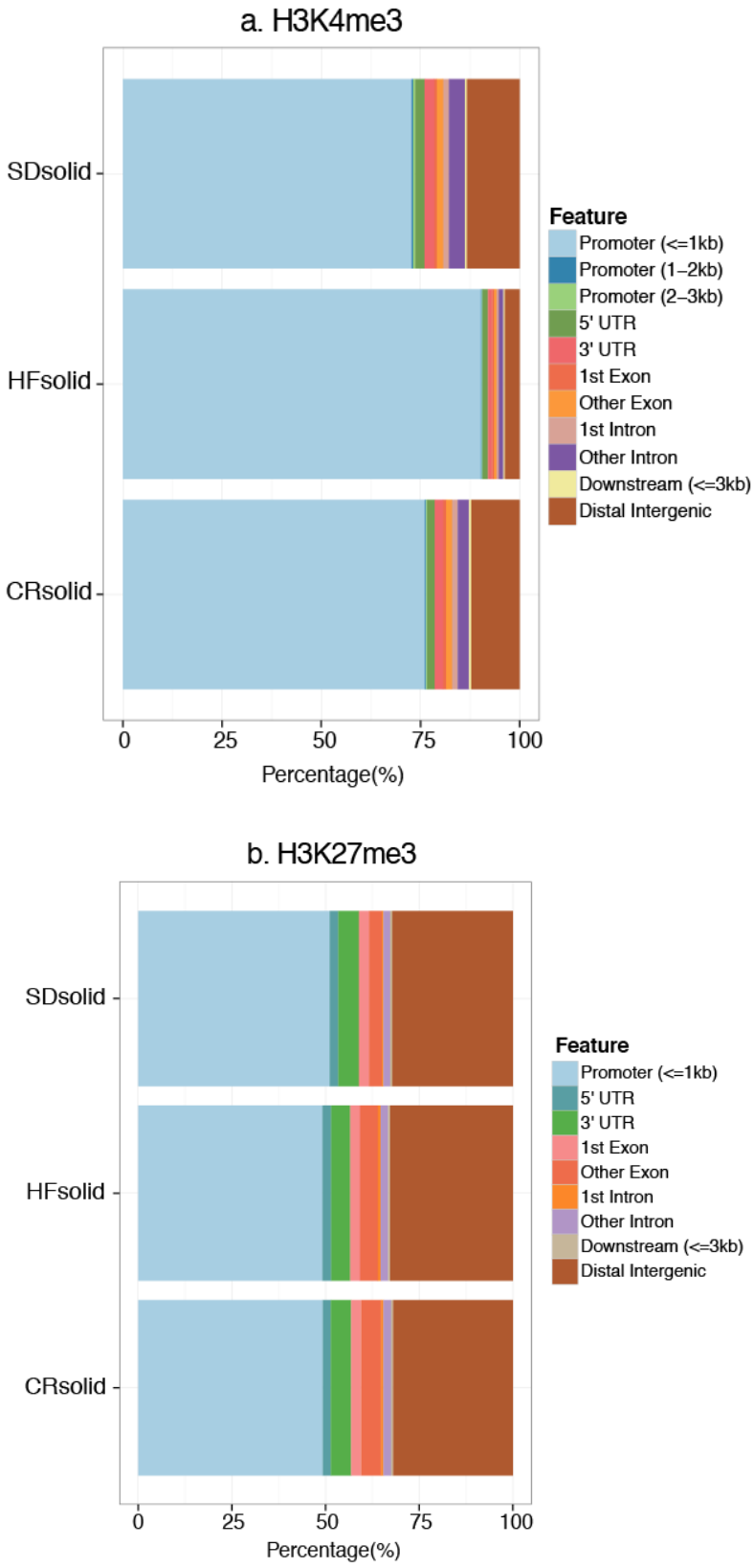


Figure 3.8 Genomic distribution of solid peaks divided by histone modification and diet group.

(a) Percentage of H3K4me3 solid peaks in Promoter classes for HF (90%) is significantly higher than SD (72%) and the same is true for CR (76%). (b) Percentage of H3K27me3 solid peaks in each genomic class is almost identical through diet groups, in particular K27 peaks are concentrated in Promoter (~50%) and Distal Intergenic (~35%) classes.

We then analysed K4 and K27 datasets to identify **lost peaks** (e.g. present in SD, but not in HF or CR) or **gained peaks** (e.g. present in HF or CR but not in SD). First, we generated a unique not-redundant peak dataset from the pool of peaks identified in all samples, for each diet group (SD, HF and CR) using BEDtools. Then, starting from the solid peaks, we used BEDtools to identify, by intersection, diet-specifically enriched regions. To find HF and CR **gained peaks**, to be more conservative, we intersected the "solid" peaks lists for CR and HF groups with the total non-redundant list of peaks identified in SD samples (that is the initial pooled list of peaks identified in at least one SD sample). Likewise, to find HF and CR **lost peaks**, we intersected the SD solid peaks list with the total not-redundant list of peaks identified in each CR or HF sample. Solid peaks were then annotated for their genomic position, by RefSeq and UCSC genes mapping.

Gained and lost peaks and corresponding genes for H3K4me3 and H3K27me3 datasets are reported, respectively, in Fig. 3.9 and 3.10:

For H3K4me3:

- HF gains 29 peaks, corresponding to 29 genes and loses 64 peaks corresponding to 1 gene only;
- CR gains 31 peaks, corresponding to 4 genes and loses 7 peaks, corresponding to no genes.

For H3K27me3:

- HF gains 27 peaks, corresponding to 16 genes and loses 130 peaks corresponding to 64 genes;
- CR gains 13 peaks, corresponding to 12 genes and loses 89 peaks, corresponding to 52 genes.

To make Pathway analysis on these sets of regions only peaks falling on gene promoter regions [TSS±2.5 Kb] were further considered.

A numerical summary of this analysis is reported on third and sixth columns in Table 3.4).

	H3K4me3			H3K27me3		
Diet group	solid peaks	genes' TSSs covered by solid peaks	genes included in pathways	solid peaks	genes' TSSs covered by solid peaks	genes included in pathways
SD	3517	2590	1094	3505	1790	582
CR	3202	2456	1014	3243	1598	515
HF	3703	3358	1380	3092	1522	481

Table 3.4 Solid peaks and solid genes involved in KEGG pathways enrichment

In the table are reported for each histone modification and each diet group, the number of solid peaks, the number of genes having their promoter region covered by a solid peak, the number of genes that were involved in significantly enriched pathways.

No specific pathways were found enriched for gained or lost H3K4me3 peaks in HF and CR, nor for lost H3K27me3 peaks.

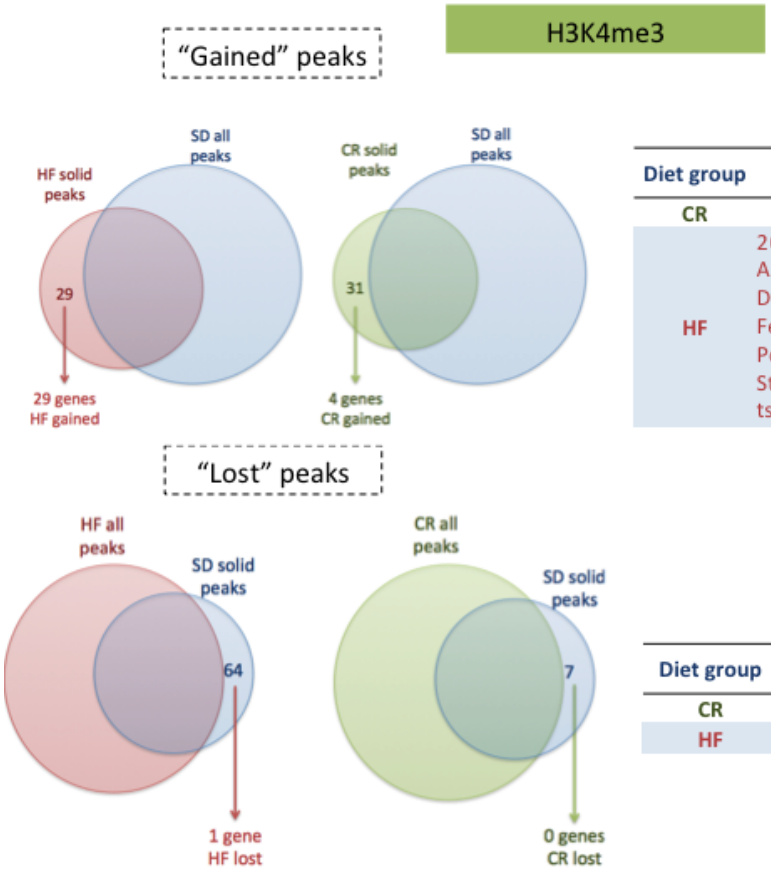
Instead for H3K27me3 gained peaks, “Natural killer cell mediated cytotoxicity” pathway was significantly enriched by 12 gained genes (p-value adjusted <0.05) and “Olfactory transduction” pathway was significantly enriched by 16 gained genes (p-value adjusted <0.05).

Since H3K7me3 is a repressive marker, having a certain pathway “gained” could probably mean that the pathway is “switched off” in the specific dietary condition.

The olfactory receptor system is used by animals to track chemical environment for molecules revealing the presence of food or toxic substances and to sense predators’ presence (**Zhang et al, 2004**).

Numerous evidences showed that the olfactory system is a target for hormones related to metabolism and food-intake regulation; moreover it adapts its function to

nutritional needs by promoting or inhibiting food foraging (**Palouzier-Paulignan et al, 2012**).



Diet group	Gained Genes
CR	Ak4, Pkdrej, Sult2a7, Usp2
HF	2010003K11Rik, AK039820, Ap1s2, Apoa2, Cabin1, Camk2g, Cenpt, Col27a1, D630039A03Rik, Disc1, Dkc1, Dpy19l3, Fes, Gk5, Il1r1, Mcm9, Nav1, Npr1, Pde7b, Pik3r1, Pxdn, Rab40b, St6galnac4, Stap2, Tex9, Thap11, Tomm40l, Ubqln2, tsec-1

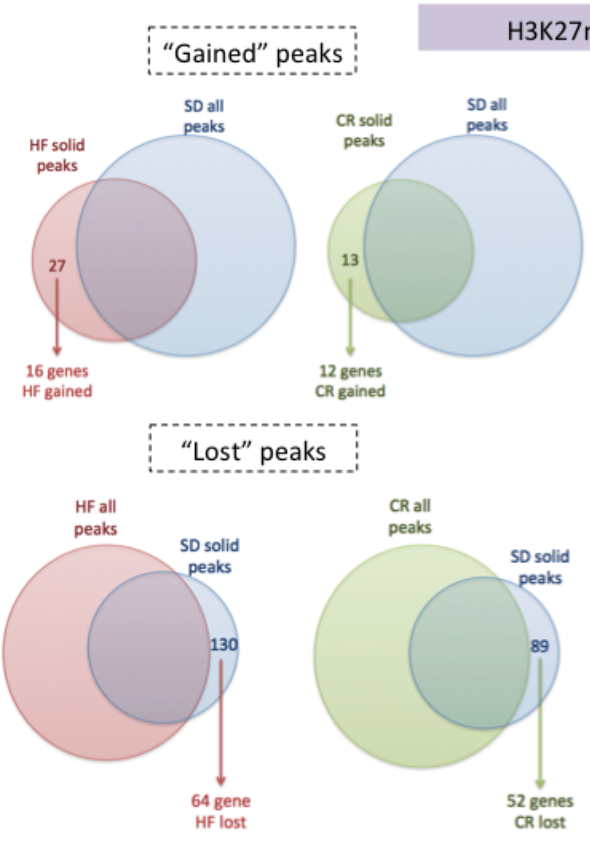
Diet group	Lost Genes
CR	-
HF	Smad9

Figure 3.9 Gained and Lost peaks for H3K4me3 dataset

Regions acquired bona fide specifically by HF or CR ("gained" peaks) are retrieved comparing diet solid peaks with all peaks found in at least one sample of SD. On the contrary, regions lost by HF and CR ("lost" peaks) are obtained comparing SD solid peaks with all peaks found in at least one sample of HF or CR. Peaks are then annotated and genes beneath peaks are reported in related flanking tables.
No significantly enriched pathway were found for H3K4me3 gained or lost genes.

Richardson et al, 2004 and **2012** showed that obese patients display decreased olfactory acuity and are significantly more likely to have absolute olfactory dysfunction or anosmia. Moreover, **Simchen et al, 2006** showed that the olfactory reception abilities decreases as body mass index (BMI) increases in subjects less than 65 years old, independent of any linkage to food odor or gender.

Recently, the elements of olfactory-like chemosensory signaling have been found also present in non-olfactory tissues such as testis (**Parmentier et al, 1992**), brain (**Mombaerts, 1999**), heart (Young et al, 2002), fat and muscles (**Choi et al, 2013**). These results, together with our evidence that the olfactory transduction pathway is switched off in HF fed mice liver, imply that the olfactory receptors and the molecules involved in olfactory transduction might be among the mediators of HFD-induced obesity progression in peripheral tissues.



Diet group	Gained Genes
CR	AK049871, Ache, Aplp1, DQ714940, Hcst, Kcnk10, Kirrel2, Nfkbid, Nphs1, Nphs1as, Tyrobp, Ufsp1
HF	1700081H04Rik, AK049871, AK148130, Abcc1, Dab1, Eml1, Gm10510, Gm5091, Lztfl1, Olfr1425, Olfr1426, Olfr1427, Olfr1428, Olfr76, Shisa7, Tmem174

Diet group	Lost Genes
CR	1810007D17Rik, 5830403L16Rik, 9530026F06Rik, AB349959, AK021075, AK029630, AK029958, AK035006, AK143160, Abca12, Agr2, Alox5, Bbs2, Bend7, Catsperb, Elovl4, Esr2, Estrb, Fa2h, Fam162b, Fbn2, Glra2, Gm5095, Gm5531, Gpr85, Gprc6a, Gyk, Mrgprb5, Mt3, Mt4, Myh1, Myh2, Myh8, Olfr341, Olfr381, Olfr382, Pnlip, Rfpl3s, Rfx6, Rgsl1, Slc22a3, Sphkap, TMEFF2, Tlr4, Tmeff2, Vav3, Vmn1r228, Vmn1r229, Vmn2r27, Wdr64, Zbbx, Zdhhc2
HF	1700049E17Rik1, 4930402F06Rik, 4930568D16Rik, A130077B15Rik, AIGP-2, AK029834, AK131712, AK139634, AK143160, Arhgef38, BV12S11J2S3, Bbs2, Bc1, C330011F03Rik, Cntn3, Crnn, Ctnna3, Dgkb, Dio2, Eltd1, Esr2, Estrb, Fam105a, Fam13c, Fam162b, Gbp1, Gbp2, Gpr85, Gprc6a, Hyal4, Hyal6, Mc5r, Mdm1, Mt3, Mt4, Myl1, Myo3b, Olfr371, Olfr516, Olfr517, Olfr518, Olfr519, Pglyrp3, Pip, Pou1f1, Rfx6, Serinc1, Slc24a3, Slc26a5, Sval1, Sval2, Sval3, TCR-beta, TCRB, TCRB13-1-1, TCRB13-1-3, TCRB13-3-3, TCRBVbeta5.1/Jbeta1.5, TRB, Tcb, Tcrb, Tmem74, Uaca, mKIAA1796

Figure 3.10 Gained and lost peaks and genes for H3K27me3 dataset

Regions acquired bona fide specifically by HF or CR ("gained" peaks) are retrieved comparing diet solid peaks with all peaks found in at least one sample of SD. On the contrary, regions lost by HF and CR ("lost" peaks) are obtained comparing SD solid peaks with all peaks found in at least one sample of HF or CR. Peaks are then annotated and genes beneath peaks are reported in related flanking tables.
No significantly enriched pathway were found for H3K27me3 lost genes, while for CR and HF gained H3K27me3 genes, "Natural killer mediated cytotoxicity" and "Olfactory transduction" pathways, respectively, were enriched.

For what concerns our finding that “Natural killer cell mediated cytotoxicity” pathway could be switched off in CR fed mice, we can find a confirm in **Clinthorne et al, 2013** in which it is proved that NK cells are reduced in frequency and numbers in most peripheral tissues of CR mice and that generation and/or maintenance of NK cells in peripheral tissues, such as the spleen, appear most affected.

3.1.2.2. The “quantitative” approach for H3K4me3 dataset

In order to identify regions specifically enriched in each diet group we performed a differential enrichment analysis using the DiffBind R computational tool (**Stark and Brown, 2011**). DiffBind provides functions for processing ChIPseq data obtained with antibodies specific for DNA-binding proteins, and is designed to work simultaneously with multiple peak sets from different ChIP experiments. We started from the peak sets identified by SICER and from aligned reads files of each sample (bam files), and applied DiffBind to identify a consensus peak set (peaks shared by a minimum number of samples) and merge the initial peak sets and counts sequencing reads within the new intervals in the consensus peak set. To identify the best threshold to build the consensus peak set, as for the positional study, we used the "elbow method", plotting the number of overlapping peaks depending on the number of samples having those peaks in common (Figure 3.11).

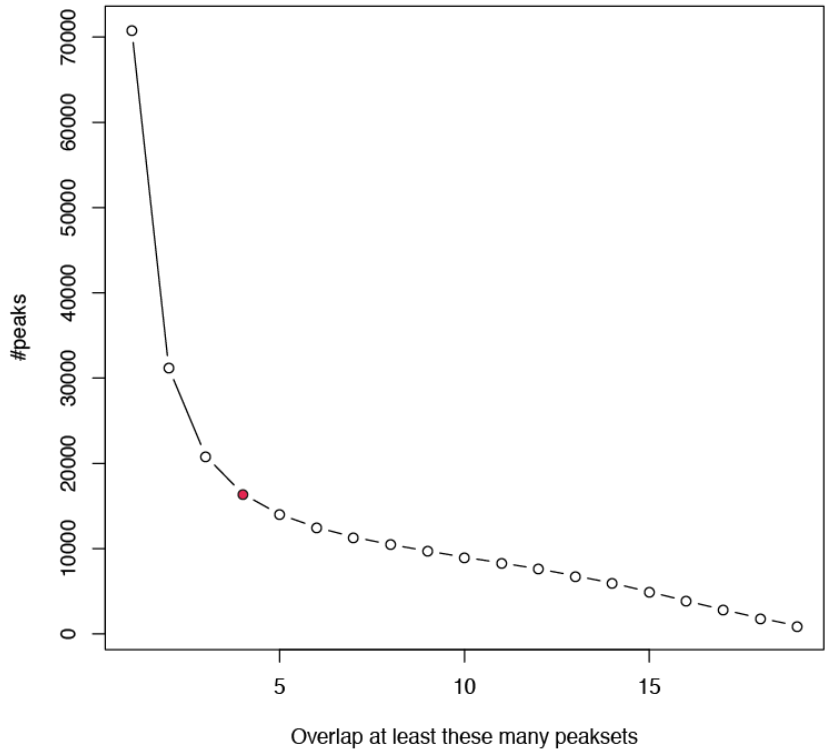


Figure 3.11 Number of overlapping peaks in all samples

On the y axis the number of peaks present in at least the number of samples represented on the x axis. Since we had to compare reads abundance for each region in different conditions, this time we take into account all the samples together without dividing them for diet condition.
In red we denoted the chosen threshold through the Elbow method (peaks common to at least 4 samples).
Accordingly we chose 4 as the minimum number of samples to consider while building the consensus peak set, which finally accounts for 16,424 regions.

After performing the TMM normalization step, we used edgeR to identify significantly differentially bound sites (DB sites), based on evidence of binding affinity (**Robinson et al, 2010**). For each possible combination of coupled group comparisons, DiffBind produced a different report of DB sites. In particular comparing HF vs SD we obtained 564 DB sites (p-value≤0.05 and FDR≤0.01), while comparing CR vs SD we obtained 59 DB sites (p-value≤0.05 and FDR≤0.05). PCA plots in Figure 3.12 show complete separation from CR and HF versus SD samples (using the relative set of DB sites).

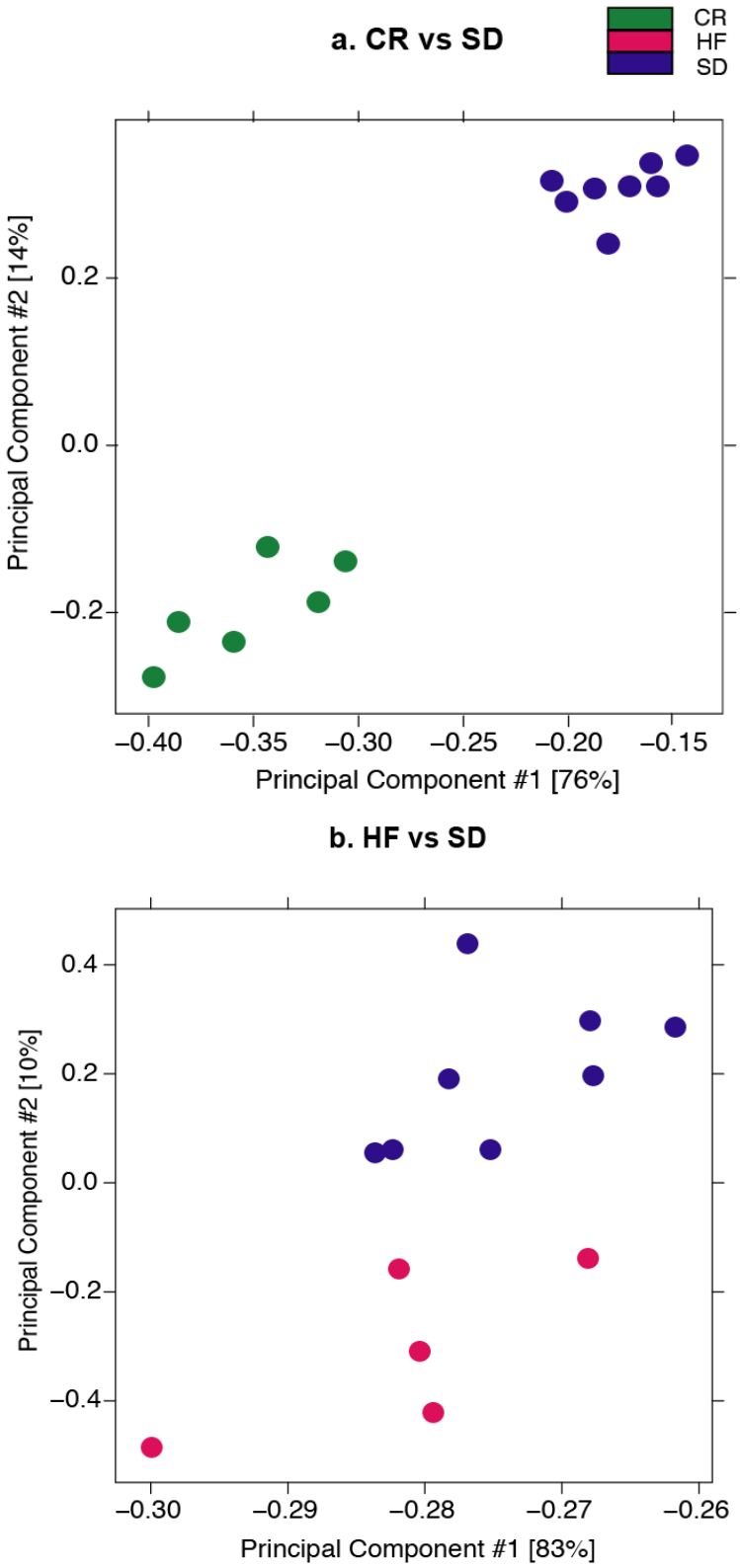


Figure 3.12 PCA analysis on statistically significant DB sites (p-value<0.05, FDR<0.01)

(a) PCA plot of CR vs SD samples on 59 DB sites, green dots represent CR samples while blue dots represent SD samples, they form two perfectly separate clusters; (b) PCA plot of HF vs SD samples on 564 DB sites, red dots represent HF samples while blue dots represent SD samples, they form two separate clusters although it seems that there is much more variability and the two clusters seem to be closer respect to CR and SD.

Results are confirmed by the heatmaps of Figure 3.13, which shows the details of correlation values for each sample. In particular, it is clearly evident how the two clusters of CR and SD are perfectly separated (Fig. 3.13, a) and correlation values histograms create two different distributions; while (Fig. 3.13, b) HF and SD seem to be much more similar to each other, despite the separation in two clusters (correlation values show a right-skewed distribution towards 1). This resemblance between SD and HF diet can be explained by the fact that ad libitum standard diet induces mild obesity.

These results give us a first hint regarding the capability of our experimental system to ascertain the existence of epigenomic features that are able to distinguish different diet conditions.

Boxplots in Figure 3.14 describe the distributions of log2 normalized reads in the regions found differentially enriched for H3K4m3 for the CR vs SD (a) and the HF vs SD (b) groups, also reporting numbers of sites at increased or decreased H3K4me3 levels for HF or CR respect to SD. The overall mean and the variance of signal intensity for CR group is higher than in SD group (Fig.3.14 (a), first panel), the number of total DB sites is very small and the number of regions where the level of H3K4me3 decreased in the CR group are higher than those where the intensity increased (33 vs 26, Fig.3.14 (a), second and third panel).

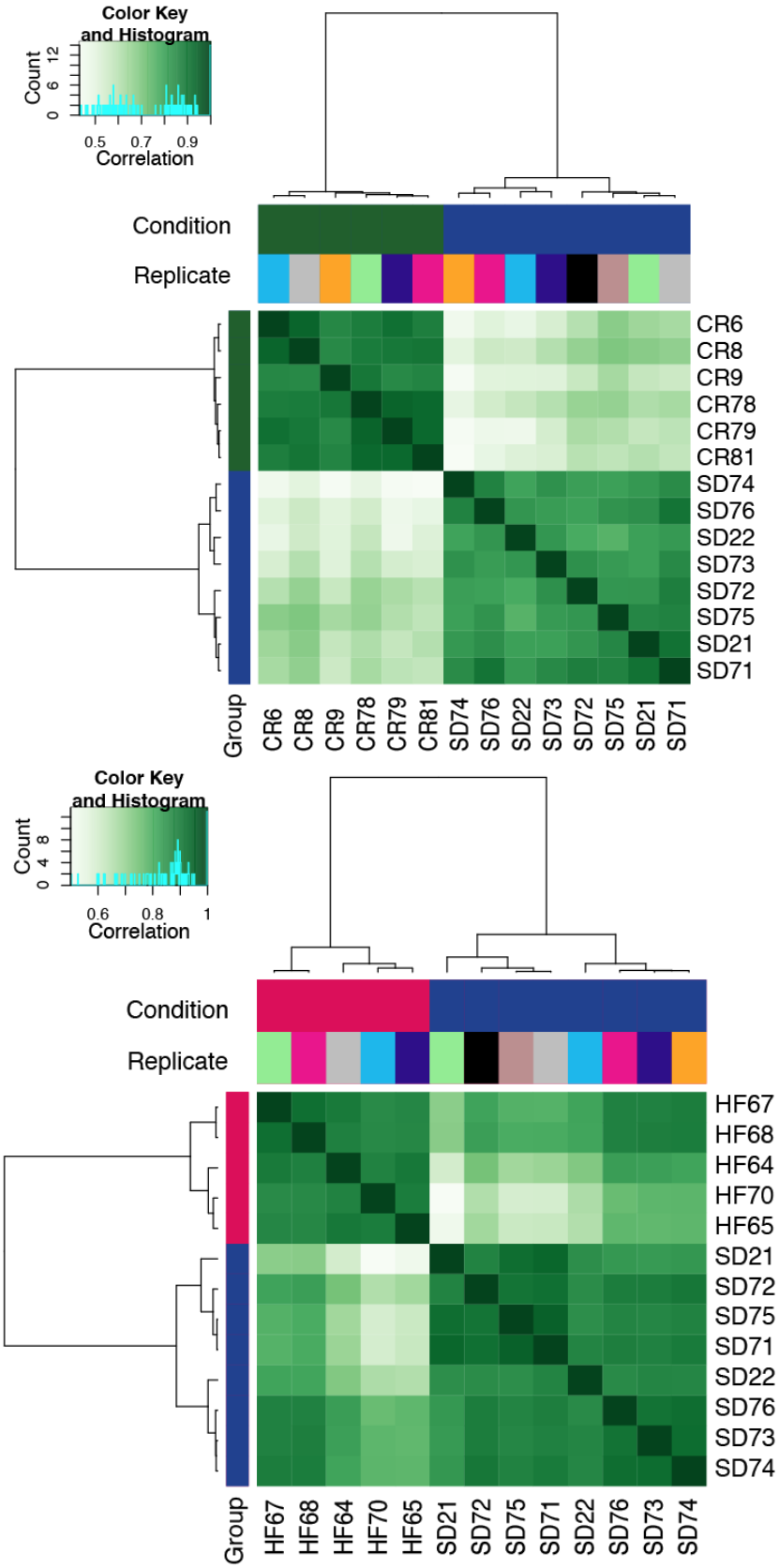


Figure 3.13 Heatmaps of correlation matrix of normalized signal in DB sites obtained by edgeR

Dark green bars denote CR samples, blue bars SD samples and red bars represent HF samples. a) CR samples cluster together respect to SD samples; b) HF and SD samples form separate clusters but the distance between HF and SD is lower than in (a).

Genomic distribution of CR DB sites (Figure 3.15, third and fourth barplots), show that, despite displaying approximately the same numerosity, increased CR regions correspond to genes TSSs in more than 75% of the cases, while CR H3K4me3-decreased regions only to ~50%.

For HF group the overall mean of reads concentration is lower with respect to SD, and the variance is almost the same for both groups (Fig.3.14(b), first panel); the number of decreased DB sites for HF is more than 6 times higher than the sites at increased level (492 vs 72, Fig.3.14(b), second and third panel). However, these few increased DB sites, which account only for 13% of the total DB sites, are all located at genes' TSSs, while only a very small number of the decreased DB sites localize on TSSs (Fig. 3.15, first and second barplots.).

Finally, we used genes corresponding to increased and decreased binding level for either HF or CR groups as input for KEGG pathway and Gene ontology Biological Processes analyses, using the clusterProfiler program. The lists of significantly enriched terms (Benjamini p-values<0.05 for pathways, q-value$\leq 10^{-5}$ for GO terms) obtained from this study are reported in Table 3.5. Strikingly, CR samples showed a statistically higher H3K4me3 signal on crucial circadian clock genes, suggesting a direct impact of the diet on the accessibility of these genes for transcription factors.

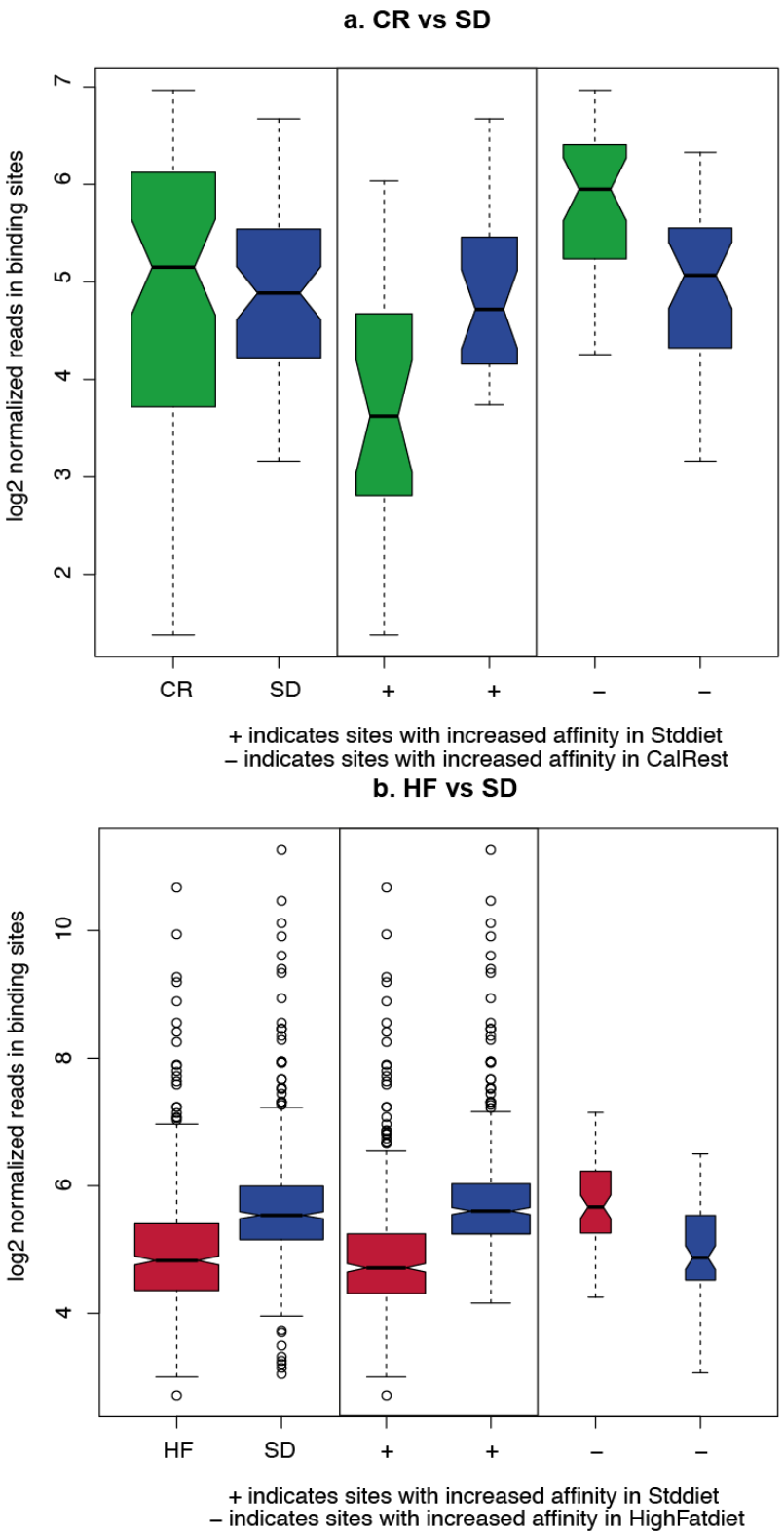


Figure 3.14 Log2 normalized reads coverage in DB sites for CR and HF vs SD.

(a) CR vs SD: first coupled boxplots are relative to all the 59 differentially bound sites between CR and SD groups, second coupled boxplots are relative to the 26 regions where we recorded a significant decrease of H3K4me3 signal in CR with respect to the SD, third coupled boxplots report the distributions of signal in the 33 regions where instead we found an increase; (b) HF vs SD: first coupled boxplots are relative to all the 546 differentially bound sites between HF and SD groups, second coupled boxplots are relative to the 492 sites discovered to be decreased in HF samples, third coupled boxplots reports the 72 sites found increased in HF.

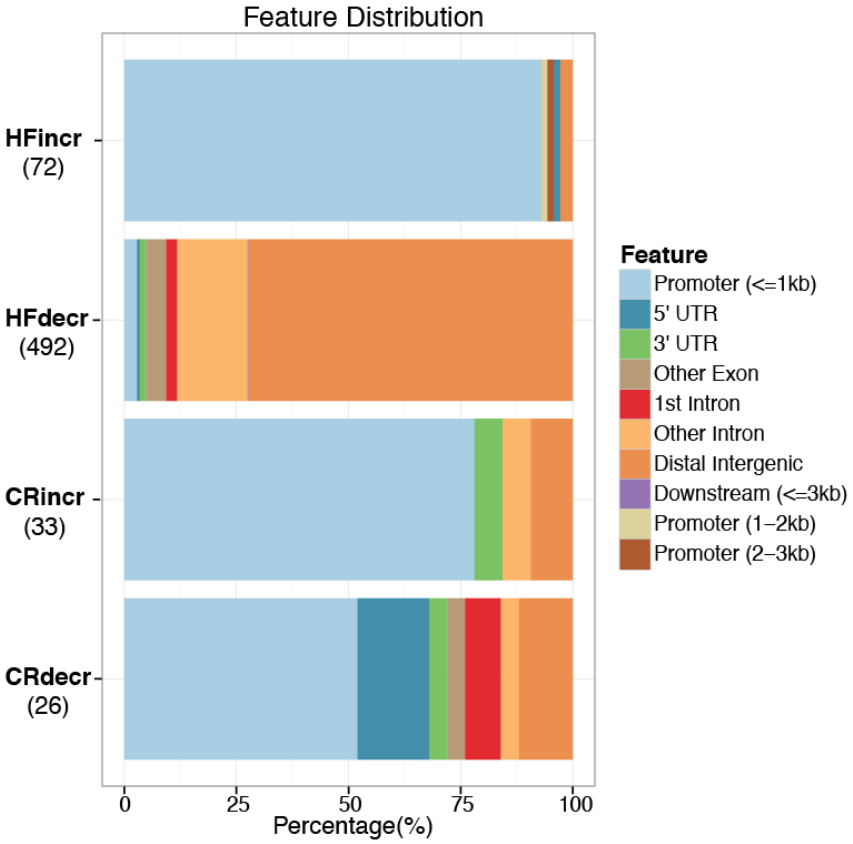


Figure 3.15 Genomic Distribution of DB sites

Differentially bound sites genomic annotation is reported for both CR and HF vs SD divided in sites at increased level and decreased level.
Increased level DB sites for both HF and CR are mostly on promoter regions (94% and 75%) while just a small percentage of decreased level DB sites are on promoter regions.

In Figure 3.16, a combined screenshot of the Genome Browser of the genomic regions around the TSS of seven representative genes (Socs3, Orm1, Pik3r, Gck, Usp2, Ciart and Per2) shows the density tracks of the H3K4me3 signal for all the 19 PAT-ChIPseq samples. The first four regions (highlighted in red) display a higher level of H3K4me3 in HF than in SD; the last three (highlighted in green) show the same behavior in CR *versus* SD.

	Description	Class	Genes
CR incr.	Rhythmic process	GO BP	Ccrn4l, Dbp, Per1, Per2, Tef, Ciart, Ahcy, Usp2
	Circadian rhythm	GO BP	Ccrn4l, Dbp, Per1, Per2, Ciart, Ahcy, Usp2
	Circadian regulation of gene expression	GO BP	Ccrn4l, Per1, Ciart, Usp2
	Circadian behavior	GO BP	Ciart, Ahcy, Usp2
	Rhythmic behavior	GO BP	Ciart, Ahcy, Usp2
	Circadian rhythm	KEGG	Per1, Per2
	Metabolic pathways	KEGG	Ces1d, Ahcy, Cyp2c54, Gpt2, Prodh, Aldh1a1, Hsd11b1
HF incr.	Response to stress	GO BP	Gck, Lgr4, Actb, Bcl3, Btg2, Socs3, Flt1, Gas6, Glul, Hk2, Hp, Il15, Il1r1, Il4ra, Orm1, Pik3r1, Prkcd, Sbno2, Thbd, Ucp2, Kdm2a, C8a, Trafd1, Mir21a, Irf7, Unc93b1, Scamp5, Grina
	Immune system process	GO BP	Mtus1, Lgr4, Bcl3, C4b, Cebpd, Flt1, Gas6, Hp, Il15, Il4ra, Psmb9, Orm1, Pik3r1, Prkcd, Sbno2, C8a, Trafd1, Mir21a, Irf7, Unc93b1
	Defense process	GO BP	Lgr4, Bcl3, Socs3, Hp, Il15, Il1r1, Il4ra, Orm1, Prkcd, Sbno2, C8a, Trafd1, Mir21a, Irf7, Unc93b1
	Regulation of protein secretion	GO BP	Gck, Lgr4, Gas6, Glul, Il4ra, Rhbdf2, Ucp2, Mir21a, Scamp5
	Type II diabetes mellitus	KEGG	Pik3r1, Socs3, Hk2, Gck, Prkcd
	Central carbon metabolism in cancer	KEGG	Pik3r1, Fgfr1, Hk2, Gck

Table 3.5 Gene ontology and KEGG pathway enrichment for genes correspondent to DB sites

Genes corresponding to DB sites localized on TSSs were retrieved and used for GO Biological Processes and KEGG pathways enrichment. In the table only significantly enriched terms are reported (p-value < 0.05 and q-value <0.0001). In particular, CR shows a higher level of H3K4me3 signal on sites all related to genes involved in Circadian processes while HF shows a higher level of H3K4me3 signal on sites of genes involved in Type II diabetes mellitus.

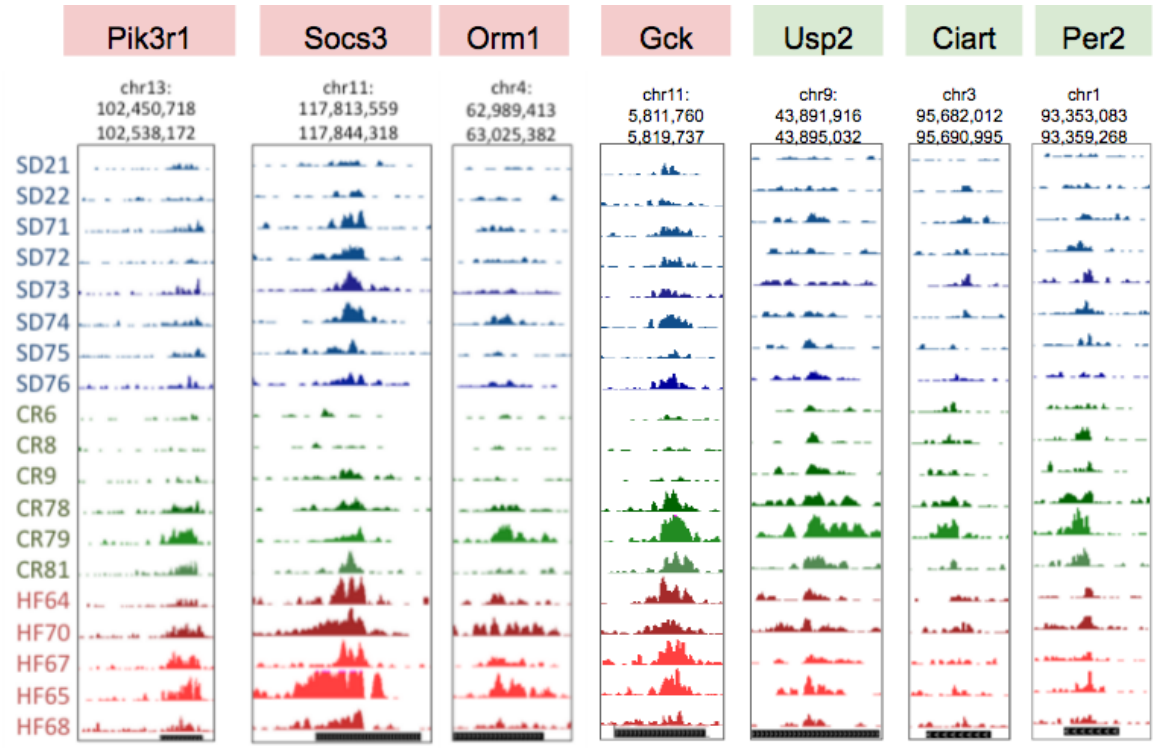


Figure 3.16 H3K4me3 signal density on TSSs of some genes.

UCSC Genome Browser screenshots of SD, CR and HF tracks samples in regions near the TSS of a subset of genes: in blue SD , in green CR and in red HF samples. For Pik3r1, Socs3, Orm1 and Gck there is a visible higher enrichment for HF samples compared to SD, on top of their promoter region (indicated with a black bar); the same is true for CR samples on TSSs of Usp2, Ciart and Per2.

Motif searching for the "Quantitative Method"

Starting from the regions found through the quantitative method, we used MEME suite to search for recurrent motifs (MEME-ChIP, **Machanick and Bailey, 2011**) and identify possible transcription factors or chromatin modifiers (TOMTOM, **Gupta et al, 2007**) involved in these diet-induced changes in chromatin.

We analyzed four classes of regions:

1. 33 regions with increased levels of H3K4me3 in CR versus SD (*CR increased*);
2. 26 regions with decreased level of H3K4me3 in CR versus SD (*CR decreased*);
3. 72 regions with increased level of H3K4me3 in HF versus SD (*HF increased*);
4. 492 regions with decreased level of H3K4me3 in HF versus SD (*HF decreased*).

1. Motifs in the H3K4me3 sites increased in CR.

In this very small set of sequences, MEME identified enrichment of the RE1/NRSE motif (Repressor Element 1/Neuron-Restrictive Silencer Element, p-value~0.08).

The RE1/NRSE is a 21 bp-motif that represents the transcription factor binding site for the chromatin modifier called REST/NRSF (RE1-silencing transcription factor or Neuron Restrictive Silencer Factor), originally identified through a bioinformatic genome-wide analysis by **Bruce et al, 2004**. In Figure 3.17 we report the motif discovered by MEME-ChIP analysis of the regions with increased level of H3K4me3 in CR vs SD (denoted with green boxplot), the comparison with the

REST motif by Tomtom and the associated p-value (measure of similarity between input motif and database motif).

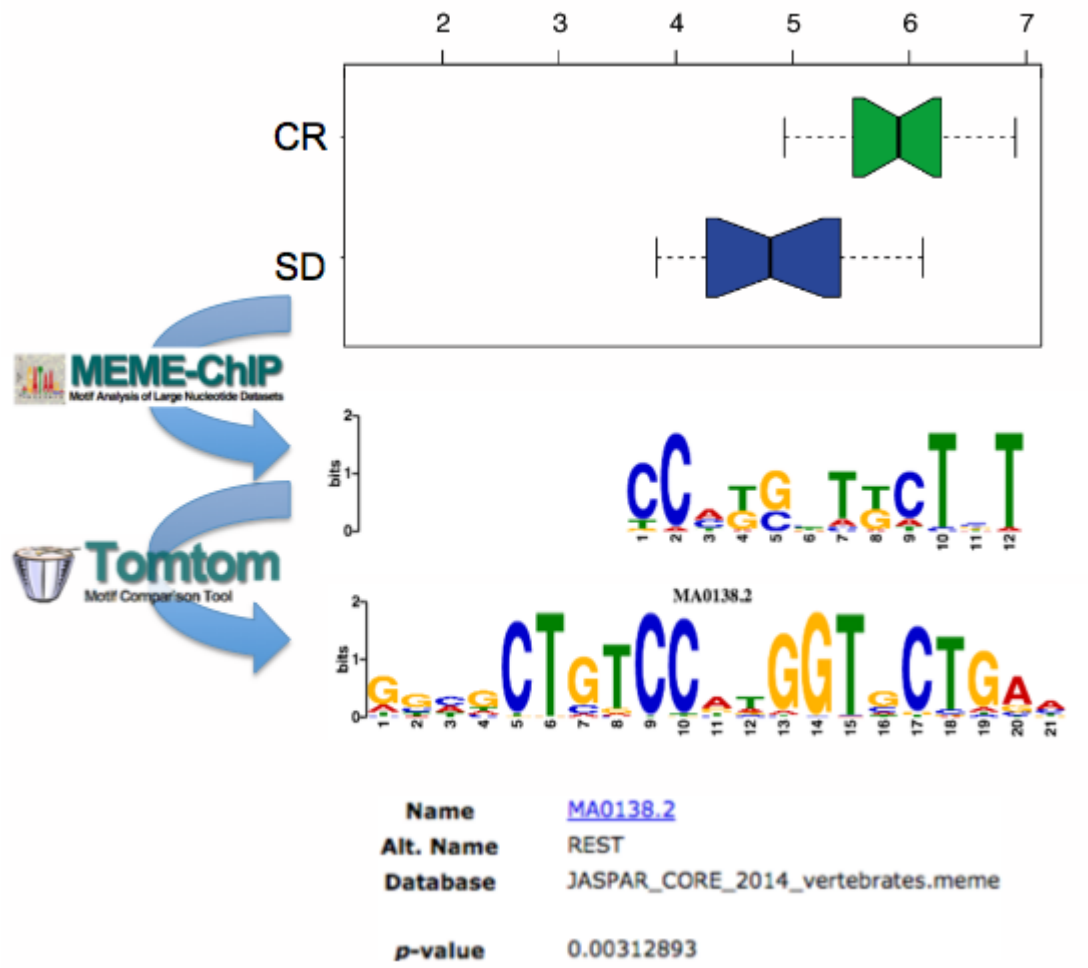


Figure 3.17 REST motif is found enriched in regions that are showing an increased level of H3K4me3 signal in CR samples

Using MEME-ChIP for motif discovery and Tomtom to compare found motifs to known TFs binding sites, we found REST motif enriched for CR at increased H3K4me3 signal sites.

Widely studied in brain, REST is involved in neuronal differentiation and silences gene transcription through recruitment of multiple chromatin-modifying partners like coREST, G9a, LSD1, mSin3, CtBP (**Anders et al, 1999; Grimes et al, 2000; Huang et al, 1999; Naruse et al, 1999; Roopra et al, 2000**).

Although it was initially thought only to repress neuronal genes in non-neuronal cells, recent evidences suggest that its role is tissue dependent and definitively more complex. In particular, it has been shown that REST interacts with CtBP in a NADH-dependent manner: NADH is the metabolite detected by the NRSF complex as a readout, or proxy, for metabolic state in rat lung fibroblastic cell line JTC-19 treated with glycolytic inhibitor 2-deoxy-D-glucose (2DG) (**Garriga-Canut et al, 2006**). CtBP homo- and hetero-dimerize in the presence of NADH to recruit various chromatin modifying complexes including HDACs and HDMs (i.e. LSD1) (as summarized by **Hayakawa et al, 2011**) (Figure 3.18).

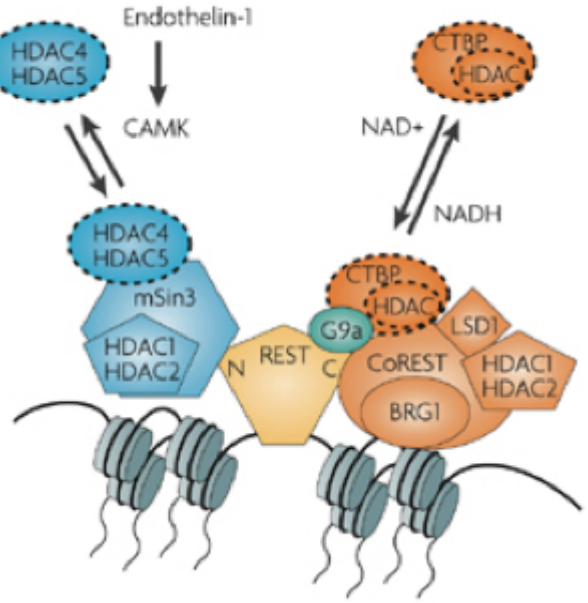


Figure 3.18 REST interacts with CtBP in a NADH-dependent manner and CtBP recruits HDACs and HDMs

REST interacts with CtBP in a NADH-labile manner: NADH is the metabolite detected by the NRSF complex as a readout, or proxy, for metabolic state. CtBP homo- and hetero-dimerize in the presence of NADH to recruit various chromatin modifying complexes including HDACs and HDMs (LSD1).

(*Adapted from Ooi and Wood, Nature Genetics Review, 2007*)

Furthermore, REST has a documented role in aging: it is down-regulated in elderly people with Alzheimer's disease and its levels are highest in the brains of people

who lived up to be 90 - 100s and remained cognitively proficient. In particular, in this group, REST levels remain specifically high in those brain areas that are more vulnerable to Alzheimer's harms, suggesting that REST might protect from dementia. Indeed, it is assumed that REST represses genes that promote cell death and Alzheimer's disease pathology, and induces the expression of stress response genes. Moreover, REST potently protects neurons from oxidative stress and amyloid β-protein (**Lu et al, 2014**).

Therefore, as an exploratory test, we used a publicly available ChIPseq anti-REST in liver of adult C57BL6 mouse (**Faure et al, 2012**; ArrayExpress accession number E-MTAB-941) to compare REST actual binding sites with regions in which we found H3K4me3 signal increased for CR respect to SD. In Figure 3.19 we report: on the left, the set of total REST binding sites (REST peaks) retrieved by analysis of anti-REST ChIP-seq (conducted as for H3K4me3 samples, SICER E-value threshold=100) together with the subset of REST peaks falling on gene promoters region; on the right the total number of genes present in UCSC mm9 assembly, together with the number of genes covered by REST peaks.

After peak calling and annotation, we found that 2,177 (26% of the total) REST peaks were localized on TSSs of ~3000 genes. These genes were then used to perform functional enrichment analysis for KEGG pathways and we discovered that many of the major metabolic processes were enriched: remarkably we noticed that Circadian rhythm, PPAR and Insulin signaling pathway are among the most enriched. These are by themself novel results, since none before characterized the REST functional role in liver.

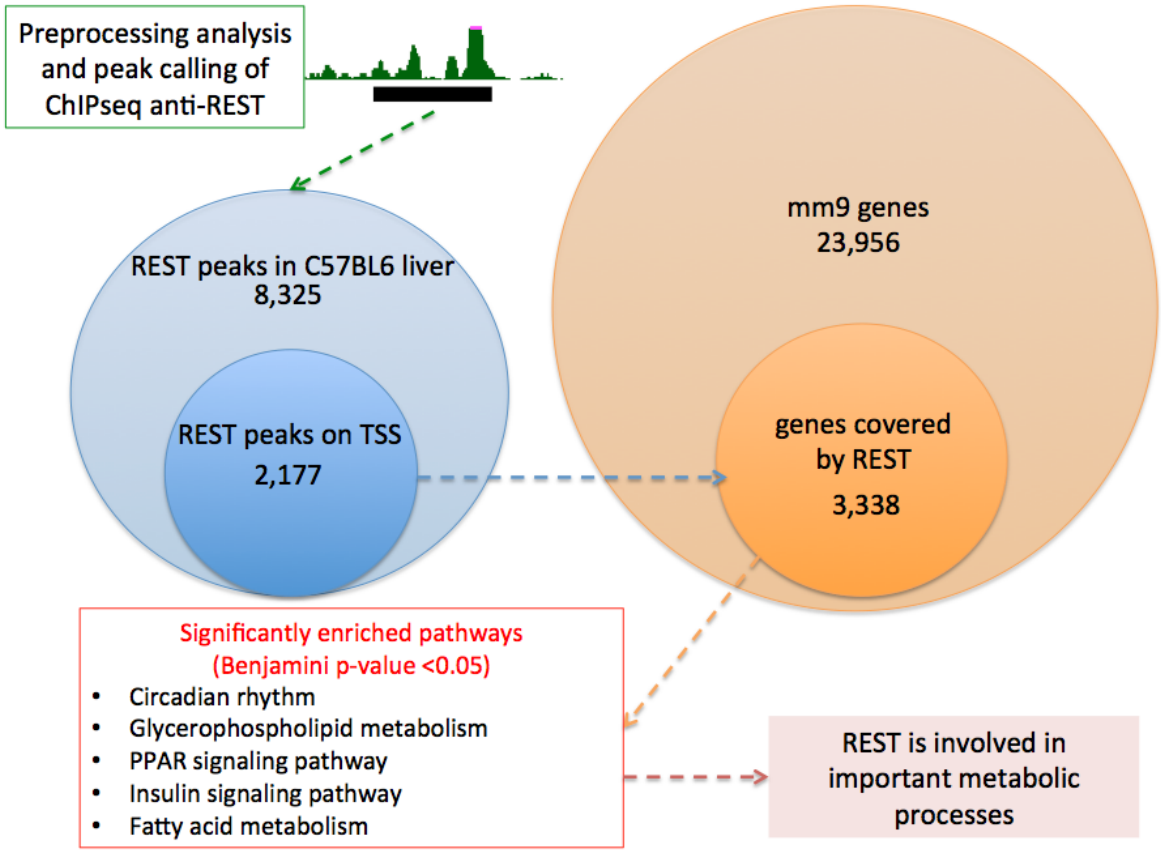


Figure 3.19 Anti-REST ChIPseq sample analysis and genomic annotation

After calling REST peaks, we annotated them and only 2177 (26% of the total) were localized on TSSs of ~3000 genes. Performing functional annotation on KEGG pathways, many of the major metabolic processes were enriched, in particular circadian rhythm, Ppar signaling and insulin signaling pathway. This is by itself a novel result, since none before characterised REST functional role in liver.

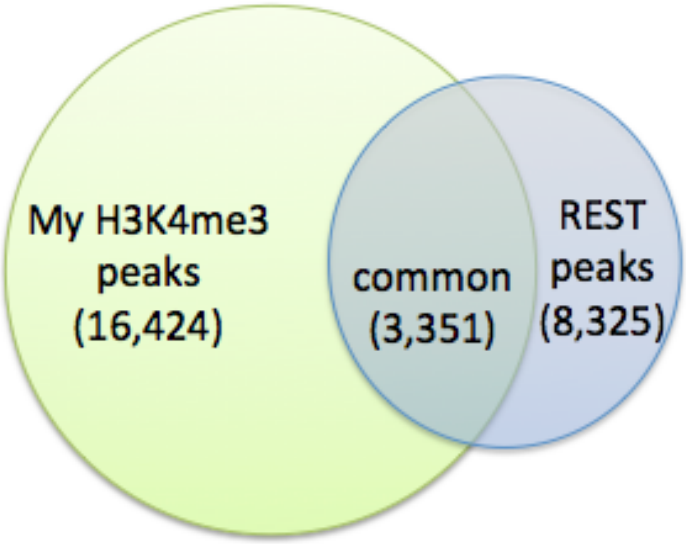
Comparing the REST binding sites (derived from ChIP-seq analysis) with the H3K4me3 consensus peak-set (derived from the DiffBind analysis), we estimated the overall probability that H3K4me3 and REST co-localizes and it score to be 0.20 (Figure 3.20).

In the subset of the 33 regions where we observed increased levels of H3K4me3, in CR samples, 15 also showed signal of REST motif binding. We used the binomial test to assess the significance of this overlap:

```
> binom.test(15,33,0.20)

        Exact binomial test

data:  15 and 33
number of successes = 15, number of trials = 33, p-value = 0.0008393
alternative hypothesis: true probability of success is not equal to 0.2
95 percent confidence interval:
 0.2810662 0.6364935
sample estimates:
probability of success
             0.4545455
```



$$p = |A \wedge B| / |A| = 3{,}351/16{,}424 = 0.20$$

Figure 3.20 REST peak-set and H3K4me3 consensus peak-set comparison

Comparing the two datasets, we found that Rest and H3K4me3 colocalize 20% of the times.

Being the p-value less than 0.05, this overlap is statistically significant, meaning that this overlap is not happening by chance.

Overall our preliminary observations suggest a mechanistic link between epigenetics modifications induced by CR and the regulation of circadian genes and REST. CR, decreasing NADH level in liver, inhibits the homo- and hetero-dimerization of CtBP, a REST cofactor, with the consequent inhibition of recruiting histone demethylases. In this way, CtBP has the potential to link a metabolic

status to specific changes in the epigenetic landscape of the nucleus and play a dominant role in determining cellular behavior.

3. H3K4me3 increased in HF

In this group we found enrichment of a recurrent motif in 12 out of 72 regions (p-value~0.09) that TOMTOM recognized as slightly similar to the Zinc Finger and SCAN Domain Containing 4 (Zscan4) transcription factor binding site (Figure 3.21).

It is worth to notice that 4 of the 12 regions in which Zscan4 motif was found, correspond to genes belonging to the “Type II diabetes mellitus” pathway found significantly enriched (Pik3r1, Hk2, Gck, Prkcd).

Zscan4 is known to have a role in telomere elongation in ES cells and genomic stability (**Zalzman et al, 2010**).

Telomere shortening in peripheral blood cells has been shown to correlate with weight gain and an increased Body Mass Index (**Kim et al, 2009**). Moreover, the average telomere length of type 2 diabetic patients was found significantly shorter than in control subjects in a cohort of 930 patients and 867 controls (**Xiao et al, 2010**). Indeed, experimental evidence suggests that telomerase is important in maintaining glucose homeostasis in mice (**Kuhlow, Florian, von Figura et al, 2010**). Conversely, elevated blood glucose levels increase oxidative stress, potentially interfering with telomerase function and resulting in shortened telomeres (**Serra et al, 2000**). Moreover, **Zhao et al, 2013** demonstrated that short telomere length is associated with future development of type 2 diabetes independently of known type 2 diabetes risk factors.

These evidences, together with our finding, suggest a possible epigenetic regulation of Zscan4 activity and of its roles in telomere maintenance and its direct

transcriptional action on specific genes' promoters involved in the onset of type 2 diabetes.

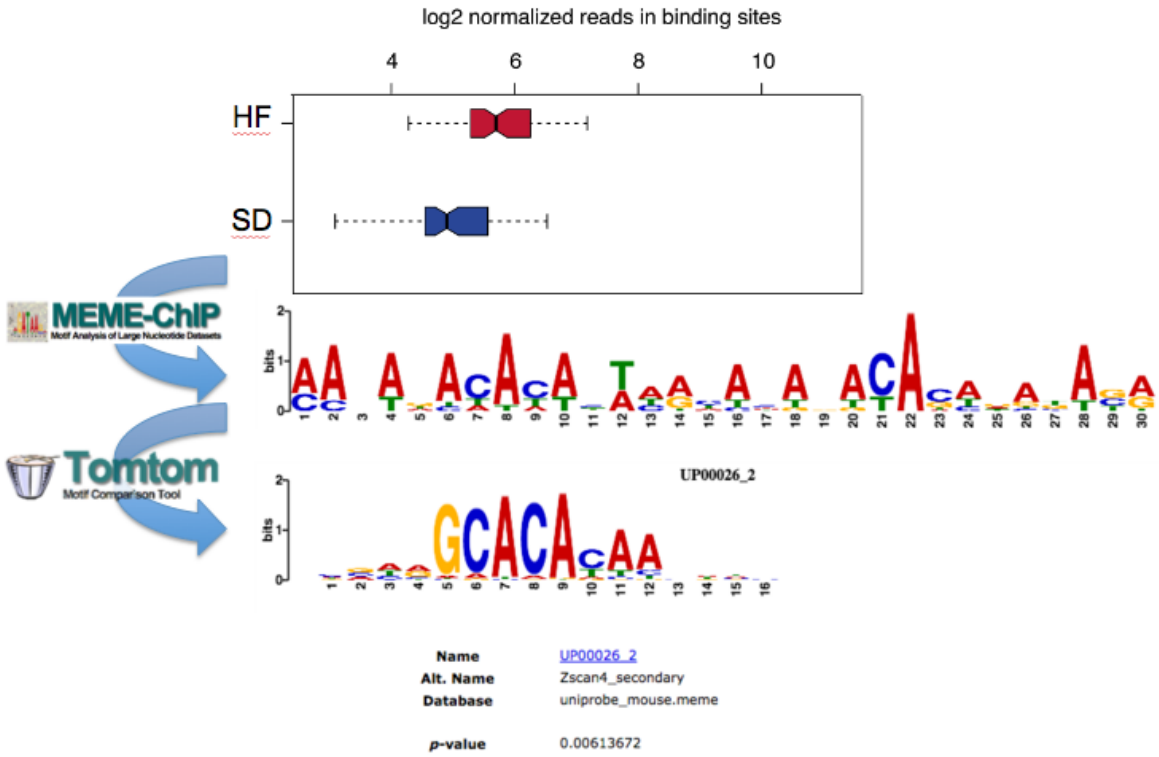


Figure 3.21 Zscan4 motif is found in sites at increased level of H3K4me3 signal for HF

Using MEME-ChIP for motif discovery and Tomtom to compare found motifs to known TFs binding sites motifs, we found Zscan4 motif enriched for DB sites with increased H3K4me3 signal sites in HF, in particular around 4 genes involved in the onset of type 2 diabetes mellitus.

No motifs were found in decreased regions of both diets (DB sites in (2) and (4)).

3.1.2.3. The "quantitative" approach for H3K27me3 dataset

Starting from the peak sets identified by SICER and from the aligned-reads files of each sample (bam files), to identify the best threshold to build the consensus peak set, we draw the number of overlapping peaks as a function of the number of samples having those peaks in common (Figure 3.22). Accordingly, we chose 5 as the minimum number of samples to consider while building the consensus peak set, which was finally composed by 5,746 regions.

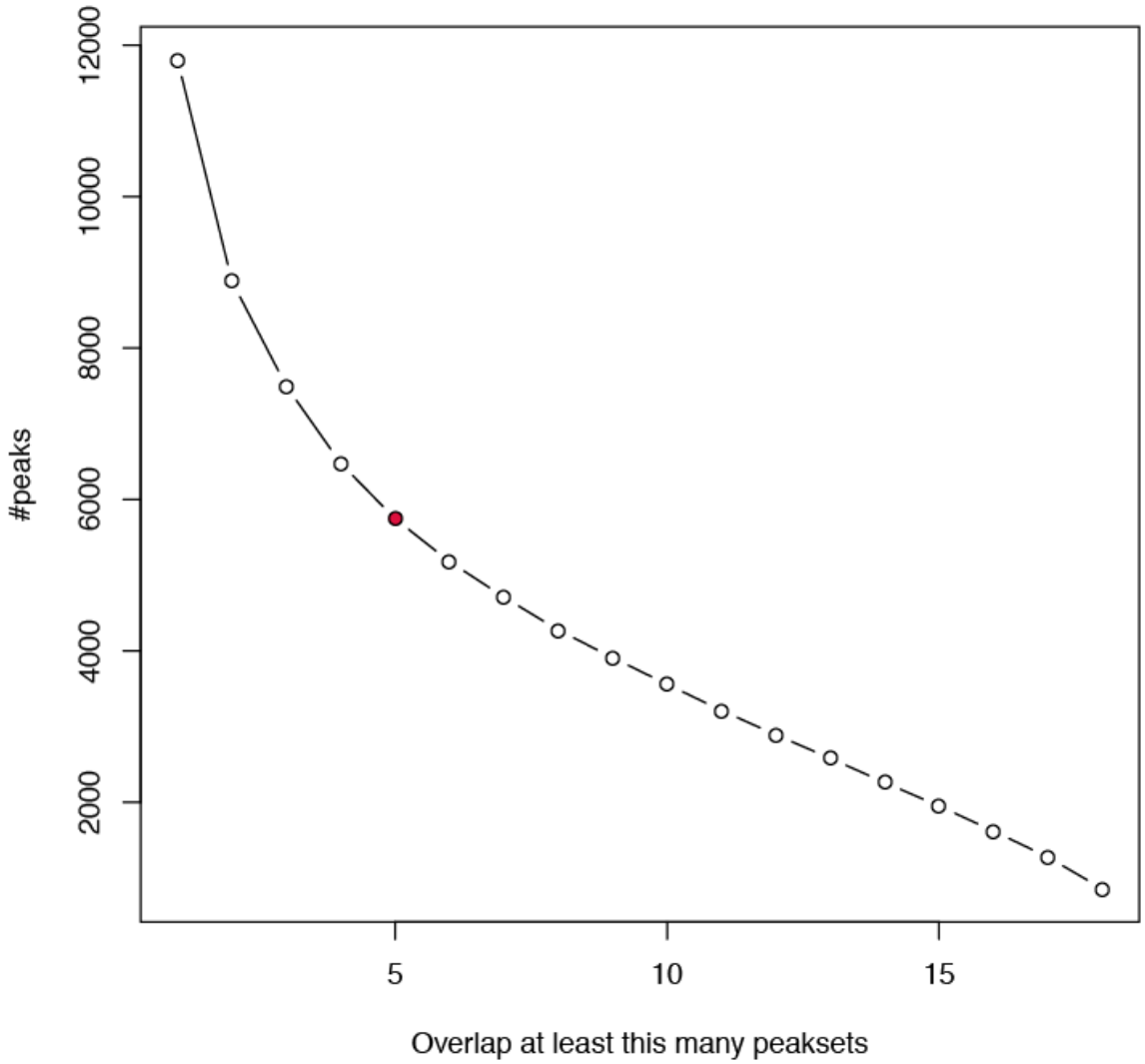


Figure 3.22 Number of overlapping peaks in all samples - H3K27me3 dataset

On the y axis the number of peaks present in at least the number of samples represented on the x axis. Since we had to compare reads abundance for each region in different conditions, we take into account all the samples together without dividing them for diet condition.
In red we denoted the chosen threshold through the Elbow method (peaks common to at least 5 samples).

Similarly to what was done for H3K4me3 dataset, after performing the TMM normalization step, we used edgeR in order to identify significantly differentially

enriched sites. For both diet conditions no trustable significantly DB sites were found.

We tried to lower the minimum overlap threshold but very few sites were identified and the fold change between experimental conditions and control was always so small that made them not trustable. This is probably due to the inefficacy of properly identification H3K27me3 broad peaks.

3.2. RNA-seq data analysis

In this paragraph we report results obtained from the bioinformatic analyses of data produced by RNA-seq experiments with the pipeline described in the previous chapter.

The first part encloses a descriptive analysis of the collected datasets in which we report preprocessing information related to reads abundance, samples variability and quality controls.

The second part is focused on the differential expression analysis and the functional annotation of regulated genes in KEGG pathways and GO biological processes.

3.2.1. Preprocessing, variability analysis and quality check

RNA-seq samples were derived from 3 CR mice (CR6, CR8, CR9) and 1 SD mouse (SD1, different from the previous cohort). As described in paragraph 2.2, after filtering out low-quality reads, we mapped the short reads to the reference genome (mm9 UCSC) using TopHat **(Trapnell et al, 2009)**, while gene-level read counts were obtained using HTseq-Count (**Anders et al, 2014**).

In table 3.6 the number of accepted hits and unmapped hits are reported together with mean and standard deviation values. For all four samples we reached a

satisfactory amount of mapped fragments, although some differences appear between CR9 and SD1.

Sample ID	Mapped fragments	Unmapped fragments
SD1	38,365,177	3,715,838
CR6	42,000,394	3,186,394
CR8	49,364,306	3,327,045
CR9	64,082,059	3,622,055
mean	48,452,984	3,462,833
std dev	11,379,781	247,804

Table 3.6: RNA-seq reads counts for each sample

The counts of mapped and unmapped reads are reported for each sample with the mean and standard deviation for both classes. We reached an acceptable yield of sequences for each sample, although there is a certain difference especially between CR9 and SD1.

This difference is also reflected in the gene body coverage curves calculated with RSeQC (**DeLuca et al, 2012**) and represented in Figure 3.23, in which is clearly visible that a 3' bias is present especially in SD1 and CR6 samples, while CR8 and CR9 have a more uniform distribution of reads along the gene body.

We also checked splice junctions saturation level: since for a well-annotated organism the number of expressed genes in a specific tissue is almost fixed, the number of splice junctions is also invariant. All known splice junctions should be rediscovered from a saturated RNA-seq data, otherwise, downstream alternative splicing analysis is problematic because low abundance splice junctions are missing. This analysis checks for saturation by resampling 5%, 10%, 15%, and so on until 95% of total alignments from each sample BAM file of mapped reads, and then detects splice junctions from each subset and compares them to reference gene model (Figure 3.24).

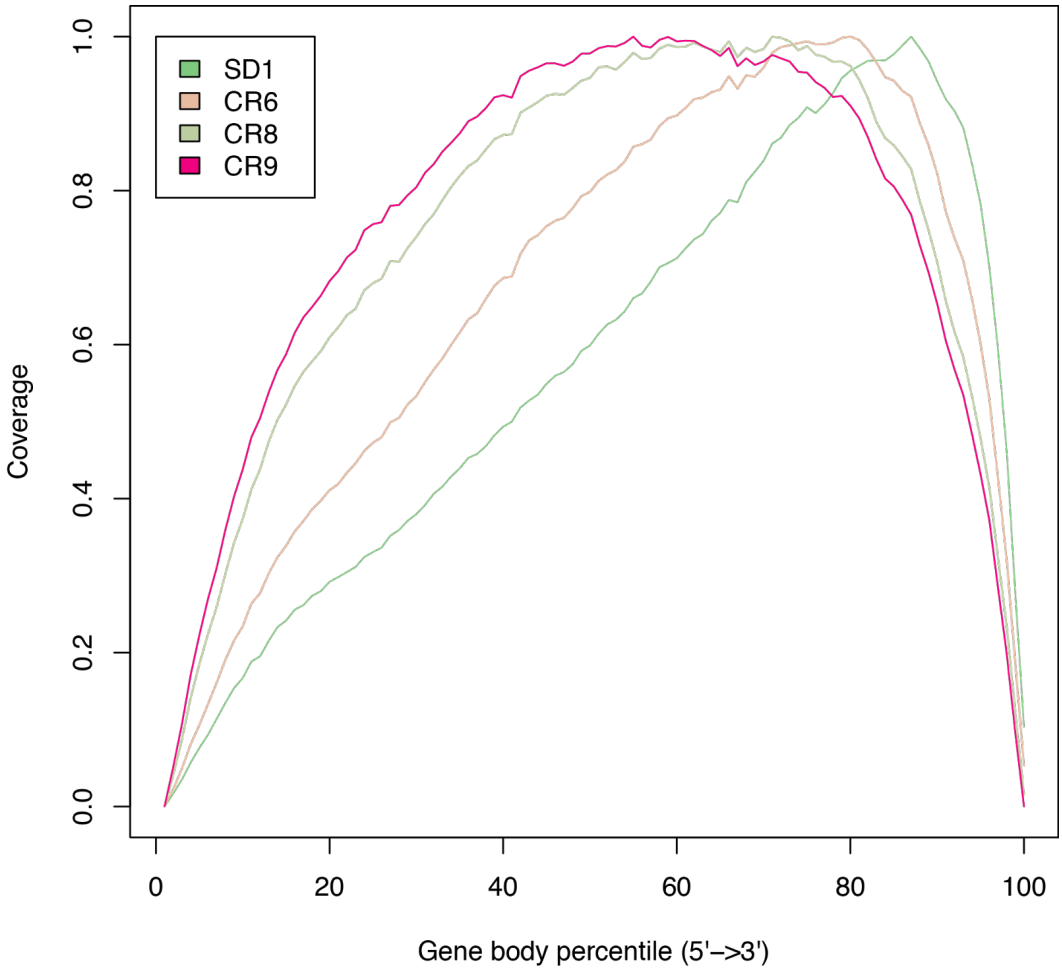


Figure 3.23 Gene body coverage curves for each sample

For each sample, a curve representing the gene body coverage from 5' to 3' is reported. While CR8 and CR9 samples seem to have an equal distribution of the coverage on the whole gene body, SD and CR6 clearly show a bias on 3' being more covered than 5' regions.

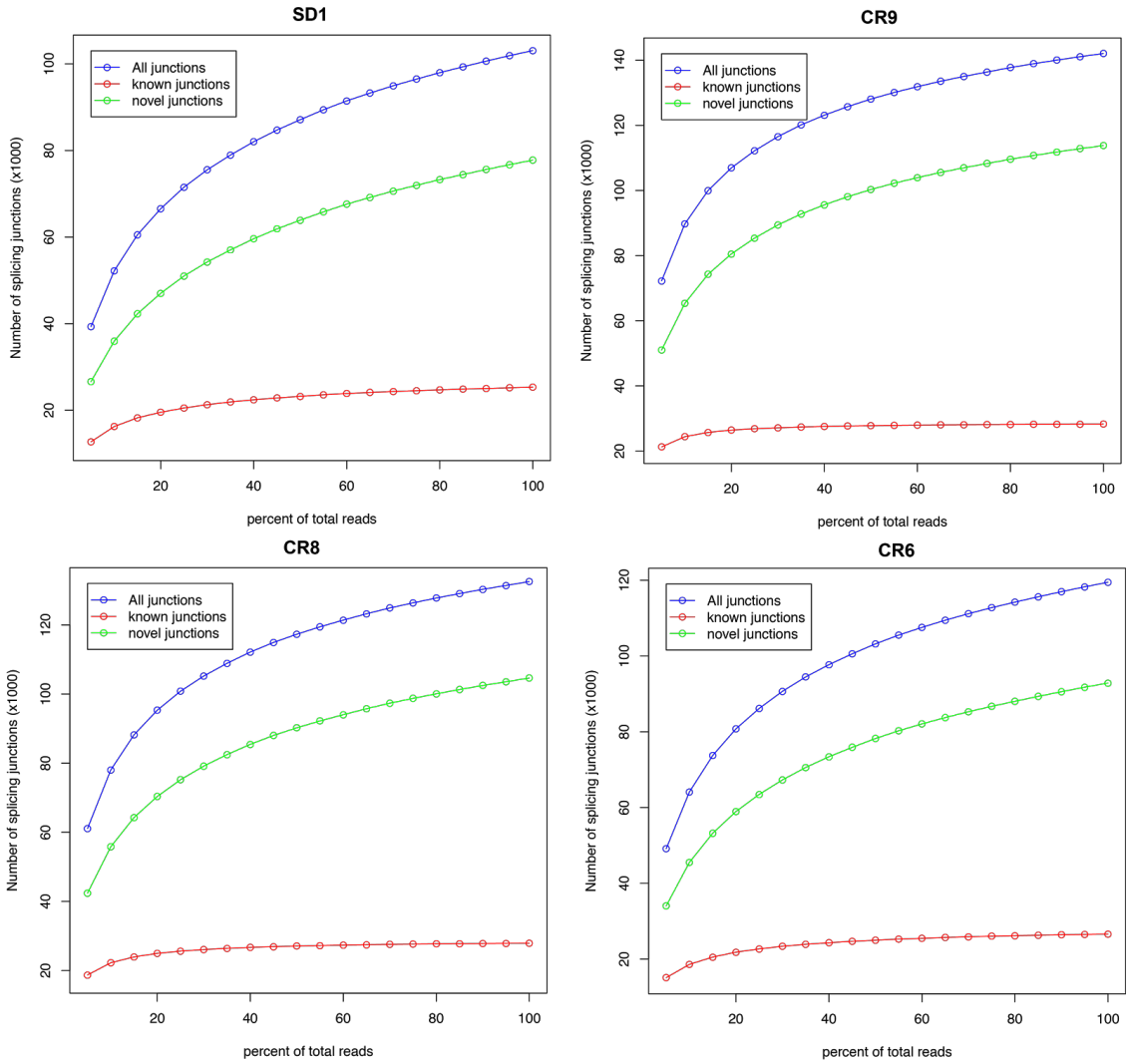


Figure 3.24 Junctions saturation analysis for each sample

CR6 and SD1 known junction curves (red curves) are not completely at saturation, differently from CR8 and CR9. For all samples, the novel junction curve (green) is not at saturation.

To estimate samples variability, we calculated the similarity matrix starting from sample reads counts per gene among samples:

$$E = \begin{bmatrix} d_{11} & \cdots & d_{14} \\ & \vdots & \\ d_{41} & \cdots & d_{44} \end{bmatrix}$$

where $d_{ij} = d(S_i, S_j)$, with i,j represent two samples in {SD1, CR6, CR8, CR9) and S_i is the array with the read counts for each gene of i-th sample and d is the euclidean distance.

In Figure 3.25 the heatmap with hierarchical clustering represents the similarity matrix calculated: the three CR samples are much similar and are clustering together, while the SD1, being different from the others, remains apart.

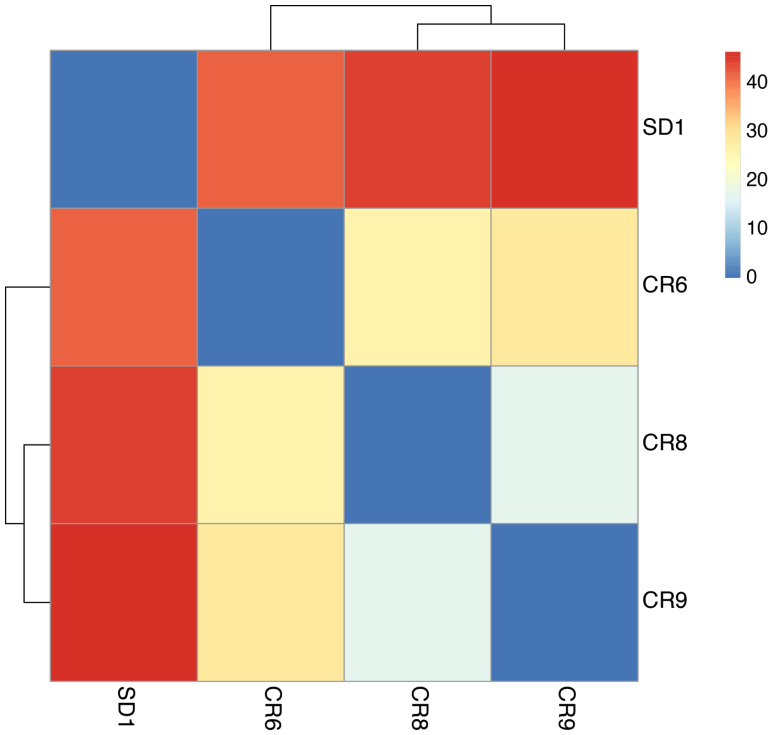


Figure 3.25 Heatmap of similarity distance matrix.

Higher distance between two samples is denoted with red colors, while lighter colors as blue or yellow denote lower distances, as reported by the legend.

CR samples cluster among themselves (in particular CR8 and CR9 are much similar respect to the CR6), being different from the SD sample.

3.2.2. Differential Expression analysis and functional enrichment

In order to identify differences in RNA expression levels of individual genesbetween control and experimental samples, differential analysis was

performed using the edgeR (**Robinson et al, 2010**), after a step of TMM normalization to correct for different library sizes and to reduce RNA composition effect. Moreover, we used the biological coefficient of variation (a measure of dispersion) estimated by edgeR that was lower than 0.2 (this means that genes expression typically differs from replicate to replicate less than 20%) using the quantile-adjusted conditional maximum likelihood method (qCML).

To define significance thresholds of regulated genes we used a volcano plot representing log2(fold change) and -log10(p-value) for genes with RPKM (reads per kilobase per million) greater or equal to 1 in at least one of the two conditions.

Using a tresholds of $|\log2(FC)| \geq 1$ with a p-value≤ 0.05 (green dots in Figure 3.26), we identified a total of 1,181 genes significantly regulated, of which 597 upregulated and 584 downregulated in CR versus SD.

We then used DAVID to perform KEGG enrichment analysis: pathways significantly enriched (p-value adjusted <0.05) are "*Drug metabolism*", "*Retinol metabolism", "Metabolism of xenobiotics by cytochrome P450", "Prion diseases", "Circadian rhythm", "Alanine, aspartate and glutamate metabolism*" and *"Steroid hormone biosynthesis*", as reported in Table 3.7, together with the genes involved and the relative adjusted p-value.

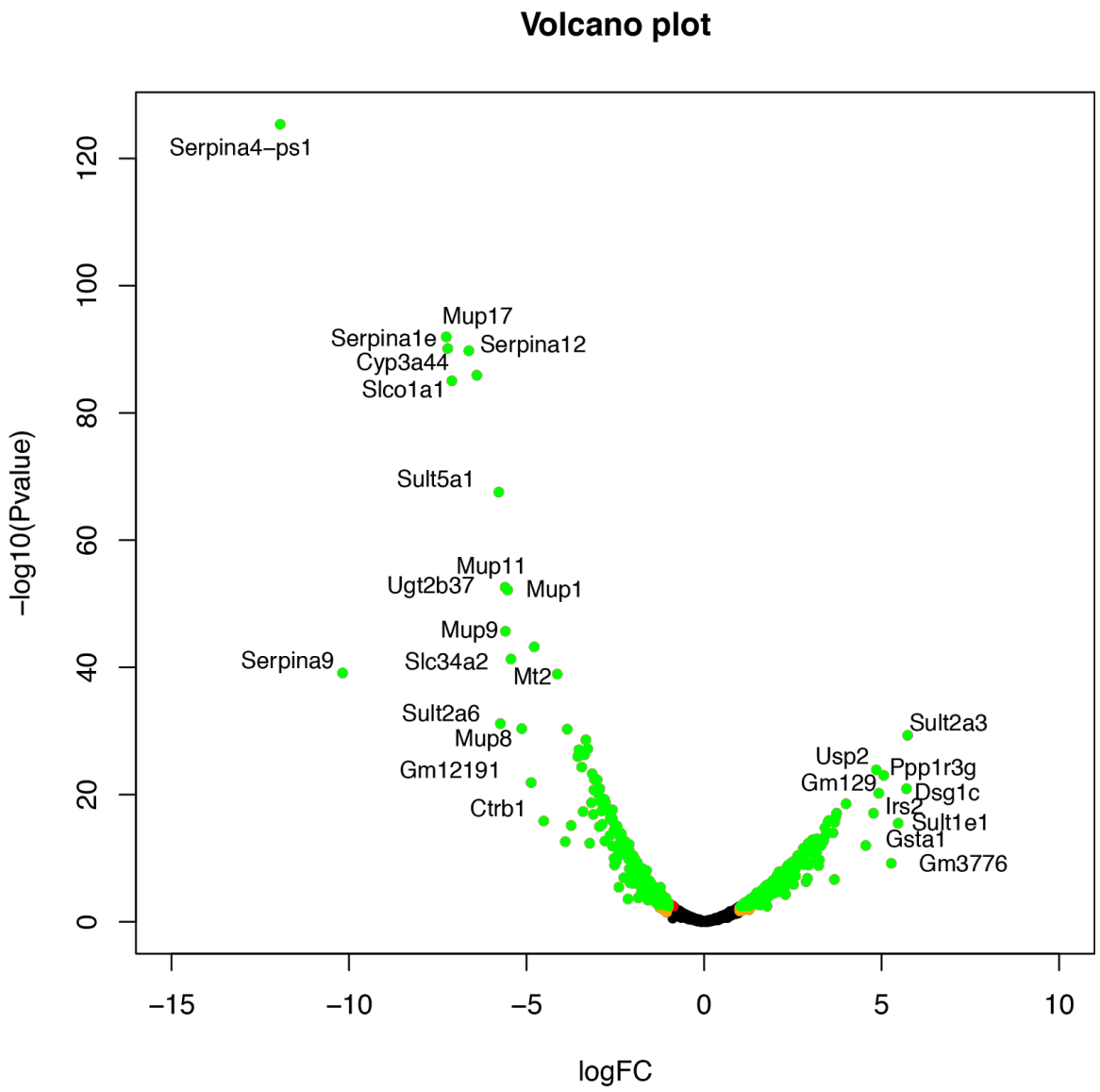


Figure 3.26. Volcano plot of genes with RPKM>1 in at least one experimental condition

Red dots represent genes with p-value less than 0.05, orange dots represent genes with log2(FoldChange) greater than 1, while green dots are genes satisfying both conditions. Top regulated genes' names are reported in the plot. Genes with log2(FC) less than zero are underexpressed in CR vs SD while, on the opposite, genes with log2(FC) greater than zero are upregulated in CR vs SD. A total of 1181 genes are significantly (p-value<=0.05 and log2(FC)>=1) regulated of which 597 upregulated in CR vs SD and 584 downregulated.

Term	Genes	Pvalue adj
Drug metabolism	CYP2D9, CYP2C69, CYP2C40, FMO4, FMO5, UGT1A9, GSTM3, UGT1A6B, FMO2, UGT1A6A, UGT1A5, CYP3A41A, FMO3, CYP3A44, CYP2C70, GSTA1, GSTA2, CYP2C55, GSTA3, CYP3A16, CYP3A11, CYP2C29, CYP2B13, GSTT2, CYP2B10, UGT1A1, CYP2A22, CYP2A5, CYP3A59, CYP2C39	1.31E-08
Retinol metabolism	BCMO1, POLR2L, CYP2C69, CYP2C40, UGT1A9, UGT1A6B, UGT1A6A, UGT1A5, CYP3A41A, CYP3A44, CYP2C70, CYP2C55, CYP3A16, CYP3A11, CYP2C29, CYP2B13, CYP26A1, CYP2B10, UGT1A1, CYP4A10, CYP2A22, DHRS3, CYP4A32, CYP4A31, CYP2A5, CYP3A59, CYP4A14, RDH16, CYP2C39, RETSAT	2.16E-07
Metabolism of xenobiotics by cytochrome P450	CYP2C70, GSTA1, GSTA2, GSTA3, CYP2C55, CYP3A16, CYP2F2, CYP2C69, CYP3A11, CYP2C29, CYP2C40, CYP2B13, GSTT2, CYP2B10, UGT1A1, DHDH, GSTM3, UGT1A9, UGT1A6B, UGT1A5, UGT1A6A, CYP3A41A, CYP3A59, CYP2C39, CYP3A44	3.03E-06
PPAR signaling pathway	PPARA, LPL, ACOX1, POLR2L, EHHADH, PPARG, FADS2, CPT1A, PCK1, CYP4A10, APOA1, CD36, CYP4A32, CYP4A31, APOC3, FABP3, FABP4, FABP1, FABP7, CYP4A14, FABP5, ANGPTL4	2.21E-04
Prion diseases	C1QA, C8A, C1QB, C8B, C9, FYN, IL1B, HSPA1B, HSPA5, PRNP, C1QC	0.00306202
Circadian rhythm	NR1D1, PER2, PER1, PER3, ARNTL, CRY1, CLOCK	0.00352466
Alanine, aspartate and glutamate metabolism	ADSSL1, GOT1, ASS1, ABAT, ALDH4A1, AGXT2, GPT, CPS1, GPT2	0.01862406
Steroid hormone biosynthesis	CYP3A16, HSD17B2, CYP3A11, UGT1A1, CYP7B1, UGT1A9, UGT1A6B, UGT1A6A, UGT1A5, CYP3A41A, CYP3A59, SRD5A1, SULT1E1, AKR1D1, CYP3A44	0.01765231
Drug metabolism other enzymes	CYP3A16, CYP3A11, UPP2, UGT1A1, CYP2A22, TYMP, UGT1A9, UGT1A6B, UGT1A5, UGT1A6A, CYP3A41A, CYP2A5, CYP3A59, CDA, CYP3A44	0.02656602

Table 3.7 Significantly enriched KEGG pathways for differentially regulated genes in CR vs SD

For each pathway described in the first column, the genes found differentially regulated in CR vs SD are reported together with the Benjamini- Hochberg adjusted p-value in second and third column respectively.

In Figure 3.27 we report, for each pathway, genes upregulated or downregulated with their Log Fold Change.

Alanine, aspartate and glutamate metabolism pathway upregulation is consistent with **Hagopian et al, 2003** work in which they found that mice on CR showed significant increases in the activities of alanine and aspartate transaminases, and of malate and glutamate dehydrogenases. This is an effect of an increased gluconeogenic activity in CR mice correlating with a state of increased hepatic gluconeogenesis and protein turnover during CR.

The same is true for Prion diseases pathway: these are protein misfolding disorders of the central nervous system with many similarities to other neurodegenerative diseases, as, for example, deposition of aggregated protein,

gliosis, and loss of synapses and neurons. **Chen et al, 2008** showed that CR delays onset of Prion diseases and this beneficial effect has also been proven to happen in other neurodegenerative disorders like Huntington's (**Duan et al, 2003**) and Alzheimer's (**Patel et al, 2005**).

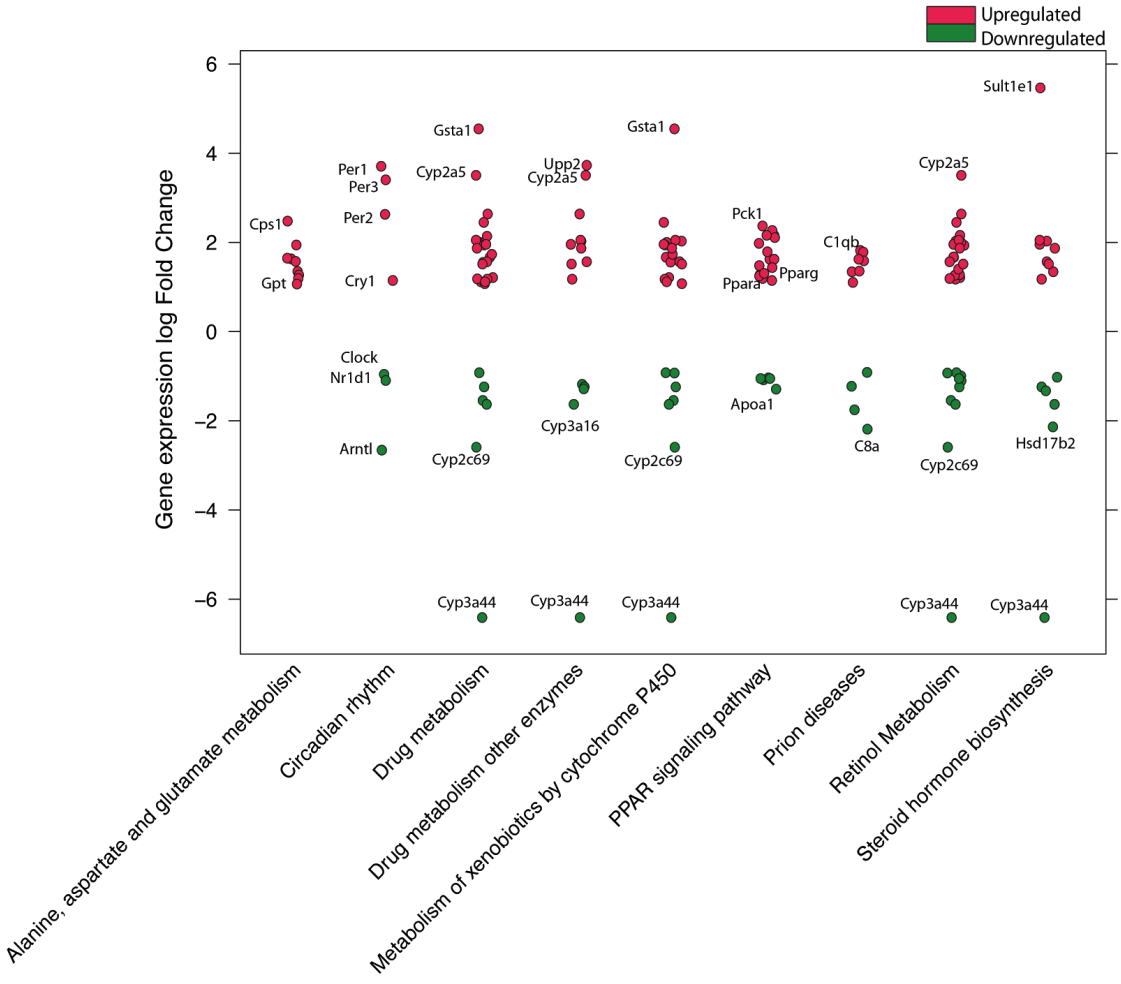


Figure 3.27 Upregulated and downregulated genes for enriched pathways in CR vs SD

For each pathway, corresponding genes log fold change are reported. Green dots represent downregulated genes, while red dot upregulated ones. Names of genes corresponding to highest or lowest log fold change values are reported for each pathway, together with interesting ones. For example, Ppara and Pparg result upregulated together with Per1, Per2, Per3 and Cry1. As expected, Arntl is downregulated as it should be since Per and Cry inhibit Arntl transcription.

Alterations of hepatic retinoid metabolism in long term dietary restricted old rats were also already reported (**Chevalier et al, 1999**). Although Calorie Restriction is known to extend lifespan, it is often accompanied by impaired reproductive function. The differential regulation of Steroid hormone biosynthesis in CR mice

can be involved in this kind of mechanisms. In fact **Thondamal et al, 2014** show in worms that the steroid signaling pathway, which regulates reproduction, is activated in response to dietary restriction (DR) and is required for DR-induced lifespan extension. It would be very interesting to investigate such relationship in mouse model.

Since calorie-restricted mice exhibit increased antioxidative defenses, and they have a slower rate of accumulation of tissue oxidative damage with age (**Merry, 2004; Hunt et al, 2006**), the up-regulation of xenobiotic metabolism (composed by many metabolizing enzymes and transporters that together work for the detoxification and elimination of potentially poisonous compounds) could be viewed as another form of enhanced stress resistance. In fact, other studies of gene expression analysis in Caenorhabditis elegans, in Ames dwarf mice, Little mice and calorie-restricted Snell dwarf mice (**McElwee et al, 2004**; **Amador-Noguez et al, 2004; Amador-Noguez D et al, 2007; Tsuchiya et al, 2004**) suggest a role for the up-regulation of xenobiotic detoxification genes as an important mechanism for longevity assurance.

It has been shown that Calorie Restriction entrains the clock in the SCN, affecting during daytime, the temporal organization of the SCN clockwork and circadian processes in mice, under light-dark cycle (**Challet et al, 1998; Challet et al, 2003; Mendoza et al 2005**). Moreover, through gene expression data comparison in seven different tissues, “circadian rhythms” is identified among the most over-expressed biological processes in mice subjected to CR (**Swindell et al, 2008**). This suggests that synchronization of peripheral oscillators during CR could be achieved directly by synchronizing the SCN, which, in turn, sends humoral or

neuronal signals to entrain the peripheral tissues (**Resuehr & Olcese, 2005; Froy et al, 2006; Froy et al, 2007**). Our findings perfectly fits in this frame; moreover, we can add that the CR impact on circadian rhythms is epigenetic-mediated, since genes involved in circadian processes also showed a significantly higher level of H3K4me3 in promoter region and we can also hypothesize that this effect could be mediated by NRSF/REST. In particular, except for Ccrn4l, all the genes that we found at increased level of H3K4me3 in CR are significantly upregulated. Fold changes and p-values are reported in Table 3.8

Gene	logFC	FoldChange	logCPM	P-value	RPKM_SD	RPKM_CR
Ciart	4.918036	30.23266538	2.726176	3.04E-13	0.205210093	6.2398442
Per1	3.712817	13.112013	4.392074	4.49E-15	0.439595512	5.8056194
Per2	2.632502	6.2010048	4.547945	9.06E-10	0.830239785	5.0599098
Dbp	3.194269	9.153152486	5.017977	1.67E-09	2.827098346	24.781006
Tef	1.915611	3.772736519	6.930385	4.36E-07	9.022132147	33.463581
Usp2	4.848302	28.80608592	5.226633	8.34E-27	0.396273642	11.237668
Ahcy	1.238026	2.358754894	10.91623	0.000551	381.8114788	886.73626

Table 3.8 Focus on expression levels of genes found with elevated level of H3K4me3 in CR respect to SD and involved in circadian rhythmic processes

All the genes in the table result overexpressed (p-value$\leq$0.05 and RPKM$\geq$1 in at least one condition and |logFC|>1). In particular Ciart and Usp2 are ~30 times more expressed in CR respect to SD, followed by Per1 and Dbp (~10 times).

Peroxisome proliferator-activated receptors (PPARα, PPARγ, and PPARβ/δ) are members of the nuclear receptors superfamily and are found expressed in multiple organs. As summarised by **Masternak & Bartke, 2006** these transcription factors regulate many physiological functions such as energy metabolism, insulin action, immunity and inflammation. In particular, PPARα regulates lipid metabolism and binds to the *Bmal1* promoter to modulate its expression. Moreover, its own expression is regulated by CLOCK–BMAL1 through E-boxes present in its promoter region (**Canaple et al, 2006; Oishi et al, 2005**). Calorie restriction is

known to act on PPARs (**Corton et al, 2005**) but the effects are strikingly organ dependent (Figure 3.28).

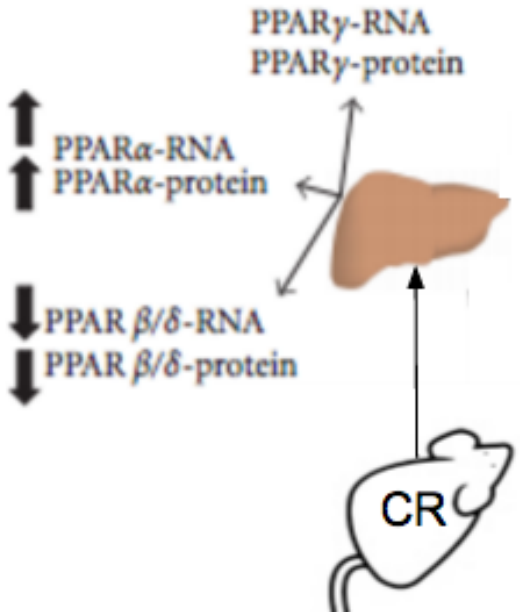


Figure 3.28 PPARs expression levels in liver of CR mouse

Scheme of the effects of calorie restriction (CR) on the expression of PPARs family genes in mouse liver. PPARg levels of mRNA and protein are not altered, as indicated by the absence of arrows.

(Adapted from Masternak & Bartke, PPAR Research, 2007)

Masternak et al, 2005 proved that the hepatic expression of PPARα and PPARβ/δ genes from the PPAR family is differentially altered by CR, while no alteration of mRNA or protein levels of hepatic PPARγ were found. PPARα increases to facilitate glucose homeostasis maintenance during periods of food scarcity. Furthermore, **Corton et al, 2005** indicated that ~20% of hepatic genes involved in lipid metabolism, inflammation, and cell growth altered by CR, were dependent on PPARα. These evidences support PPARa as a mediator for CR (**Guerre-Millo et al, 2001; Kersten et al, 1999; Leone et al, 1999; Masternak et al, 2004; Masternak et al, 2005**). Finally the expression of PPARβ/δ in the liver was significantly decreased by CR at both mRNA and protein levels **Masternak et al, 2005**.

4. Discussion

Over the last decades we witnessed an increasing attention towards food habits, especially for their putative role in disease prevention/risk. It is of the last days the large attention of the media on the report of International Agency for Research on Cancer about a link between processed and red meat consumption and higher risk of colorectal cancer (**Bouvard et al, 2015**). This does not come out, actually, as a novel finding, as many previous studies had reported the detrimental effects of diets rich in refined sugar, fat and meat, typical of Western countries. This is thought to be responsible for increased incidence of of metabolic disorders, type 2 diabetes, various types of cancer and cardiovascular diseases (**Gami et al, 2007; Giovannucci et al, 2007; Pais et al 2009; Aleksandrova et al, 2011).**

On the other hand, caloric restriction without malnutrition has been shown to have beneficial health effects: it prolongs lifespan (**Chapman and Partridge, 1996; Fontana et al, 2010; Greer and Brunet, 2009; Kennedy et al, 2007; Mair and Dillin, 2008; Masoro, 2005; Weindruch et al, 1988**) and reduces age-associated diseases, including cancer, in different experimental models (**Mattison et al, 2012; Colman et al, 2009; Harvie et al, 2012; Imayama et al, 2012).**

However, the molecular mechanisms behind the observed associations between diet and disease risk are still unknown.

On another side, increasing body of evidences suggest a role of epigenomic dynamics in the adaptation to different environmental cues, including food. In particular, numerous studies highlighted the role of DNA methylation in shaping chromatin structure in different organisms, silencing specific regions of the genome and producing precise phenotypic effects (**Wolff et al, 1998; Kucharski**

et al, 2008; Heijmans et al, 2008); while relationships between histone modifications and diet are much less explored.

For these reasons, based on the hypothesis that food adaptation entails reprogramming of different cell functions, which might be executed and maintained through changes in chromatin, in this study we investigated the impact of different diet regimens on histone modifications H3K4me3 and H3K27me3 in murine model, in order to identify a diet-specific signature and a set of potential clinical markers.

4.1. H3K4me3 profile variability

We expected to observe a certain degree of variability in chromatin features among biological replicas. Besides being a possible limiting factor for the feasibility of this study, the issue of inter-individual variability of chromatin patterns represents an unsolved issue per se that we tried to adress. We recognize, in fact, that differences in the variables we measured (i.e. final mapped reads, called peaks and global level of enrichment) might not only be ascribable to biological diversity, but also to experimental complexities (i.e. adapted PAT-ChIP protocol for liver tissue, different liver histology among different diet groups). We evaluated these specific differences as modest, in the light of ENCODE guidelines for quality check on ChIPseq experiments (**Landt et al, 2012**).

Moreover, differences in peaks genomic localization within the same diet group could represent both a real biological difference and a technical matter (e.g. underestimation of number of peaks due to lower global enrichment of the replica). In this case we had to minimize the possibility of underscoring peaks, using different threshold in the peak calling step, based on the global enrichment value (FRiP).

Patterns of H3K27me3 distribution resulted quite stable for all three diet groups, with very small fluctuation of the signal localization in genomic classes (cf. Fig. 3.5) and an index of similarity among replicas between 0.2 and 0.6 for all diet groups (meaning that replicas share from 40% to 80% of the peaks, cf. Fig. 3.6).

H3K4me3 differences among biological replicas were much higher: fluctuation of signal genomic localization in HF group was very low, compared to CR and SD (Fig.3.3). Moreover the similarity index among HF samples is around 0.3 while in CR, it is around 0.6 - 0.7 and for SD it goes from 0.3 to 0.6, in both cases with the exception of some samples recorded to be totally different from the core clusters and representing the source of major variability (Fig. 3.4).

These diversities represented an issue to downstream analysis so we reduced it through normalization and increasing the stringency of the analyses, even at the expense of obtaining a reduction of the peak sets considered significant for each diet group.

Since we finally were able to obtain meaningful and coherent results that fit with both known and expected outcomes reported in literature, we believe that, in spite of the variability of the experimental setting we used, overall this work makes an interesting and original contribution to the understanding of the field.

4.2. H3K4me3 signal reveals the presence of diet-specific epigenetic signature

We analyzed statistically significant quantitative-differences of reads abundance in CR and HF respect to SD samples. Our differential data analysis of genome-wide H3K4me3 profile supports the existence of diet-specific epigenetic signatures.

In fact, a relatively small number of sites with significant different abundance of H3K4me3 signal is able to separate and cluster CR and HF from control diet samples (Fig. 3.12 and 3.13).

4.2.1. Calorie restriction acts on circadian clock through epigenetic mechanisms, shaping chromatin conformation and altering gene expression of specific regulators

Genomic regions showing an increased level of H3K4me3 in CR respect to SD, corresponds to genes involved in Circadian rhythmic processes (Per1, Per2, Tef, Ciart, Ahcy, Usp2, Dbp, Ccrn4l). The same genes were also found significantly overexpressed in CR (Table 3.8).

The circadian clock is in charge of biological timekeeping, on a systemic level. The central clock situated in the SCN in the brain, communicates and regulates local peripheral clocks, present in other tissues, synchronizing them as a unique system (cf. paragraph 1.1.5, **Froy, 2011**). It has been shown that Calorie Restriction entrains the clock in the SCN, affecting during daytime, the temporal organization of the SCN clockwork and circadian processes in mice, under light-dark cycle (**Challet et al, 1998; Challet et al, 2003; Mendoza et al 2005**). Moreover, through gene expression data comparison in seven different tissues, “circadian rhythms” was identified among the most over-expressed biological processes in mice subjected to CR (**Swindell et al, 2008**). This suggests that synchronization of peripheral oscillators during CR could be achieved directly by synchronizing the SCN, which, in turn, sends humoral or neuronal signals to entrain the peripheral tissues (**Resuehr & Olcese, 2005; Froy et al, 2006; Froy et al, 2007**). Our findings are coherent with what was previously found; moreover, our observations support a new mechanistic theory by which the CR impact on circadian rhythms is

epigenetic-mediated, since genes involved in circadian processes also showed a significantly higher level of H3K4me3 in promoter region.

4.2.2. NRSF/REST could be the mediator of CR induced beneficial effects acting on chromatin remodeling and transcription of circadian genes

Starting from the regions obtained through quantitative differential analysis, we performed motif discovery analysis and comparison with known transcription factor motif databases. The motif of a known chromatin modifier, NRSF/REST is found enriched in regions with increased level of H3K4me3 in CR (even if just below the threshold we considered significant; pvalue=0.003128). Moreover, through the analysis of anti-REST ChIPseq in liver of adult mouse publicly available, we assessed the actual presence of REST peaks on these regions and on other genes promoters involved in many important metabolic processes (**Chong et al, 1995**).

REST is a protein, member of the Kruppel-type zinc finger transcription factor family that represses transcription by binding a DNA sequence element called the neuron-restrictive silencer element. It acts as a master negative regulator of neurogenesis and it is expressed in different tissues, including brain, liver, stomach and spleen.

Widely studied in brain, REST is involved in neuronal differentiation and it silences gene transcription through the recruitment of multiple chromatin-modifying partners like coREST, G9a, Lsd1, mSin3, CtBP (**Anders et al, 1999; Grimes et al, 2000; Huang et al, 1999; Naruse et al, 1999; Roopra et al, 2000**).

Although it was initially thought only to repress neuronal genes in non-neuronal cells, evidences are more recently suggesting that its role is tissue dependent and definitively more complex.

In particular, it has been shown that REST interacts with CtBP in a NADH-dependent manner: NADH is the metabolite detected by the NRSF complex as a readout, or proxy, for metabolic state in rat lung fibroblastic cell line JTC-19 treated with glycolytic inhibitor 2-deoxy-D-glucose (2DG) (**Garriga-Canut et al, 2006**).

CtBP homo- and hetero-dimerize in the presence of NADH to recruit various chromatin modifying complexes including HDACs and HDMs (i.e. Lsd1) (as summarized by **Hayakawa et al, 2011**).

Furthermore, higher REST levels in brain of old people protect from Alzheimer's and correlate with longevity and healthy aging, two features of CR beneficial effect (**Lu et al, 2014**)

Calorie restriction is known to decrease NADH levels and this particular effect has been correlated with the increase in lifespan in yeast and mammals (**Lin et al, 2000** and **2004**).

Of course, it needs experimental proofs, but these evidences together with our data are compatible with the model in which CR, decreasing NADH levels, impairs REST recruitment of CtBP, and consequently of Histone Demethylases on its targets (including the circadian genes), that in turn produces an increase of H3K4me3 levels on their promoters and results in their overexpression.

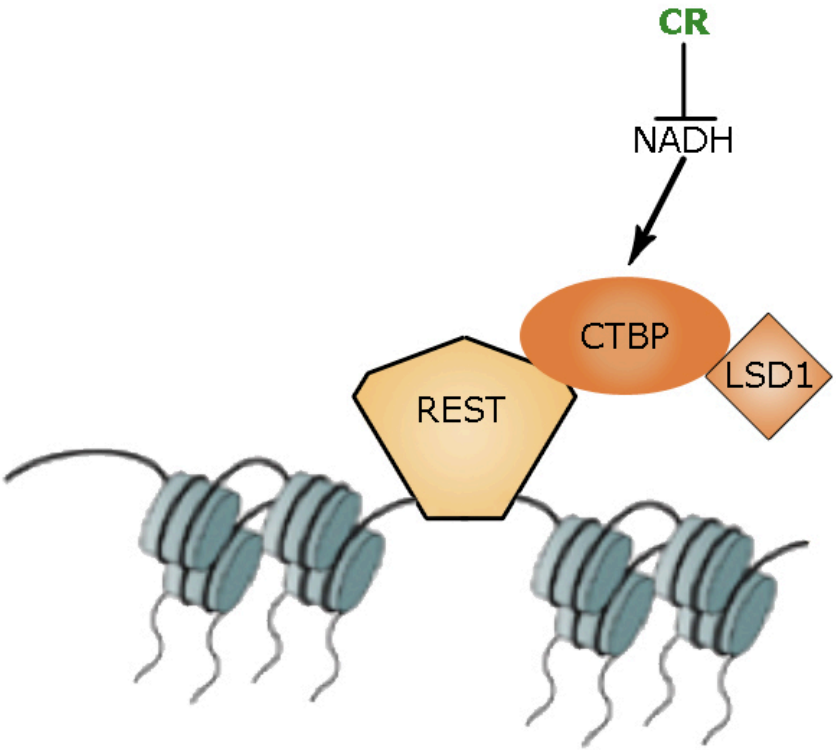


Figure 4.1 REST could be the mediator of CR-induced beneficial effects acting on chromatin remodeling and transcription of circadian genes

CR, decreasing NADH levels, impairs REST recruitment of CtBP, and consequently of Histone Demethylases on its targets (including the circadian genes), that in turn produces an increase of H3K4me3 levels on their promoters and results in their overexpression.

4.2.3. The High Fat diet shapes chromatin configuration, favouring a higher "opening state" at gene promoters

Testing the significance of differences in genomic localization of H3K4me3 signal among different diet groups, the HF diet results having an overall significantly higher mean percentage of H3K4me3 peaks in promoter regions respect to CR and, at the same time, a significant lower mean percentage of peaks in distal intergenic regions. Even increasing the stringency of our analysis, HF showed a higher percentage of promoter peaks than the other two groups (~90% versus ~75%) although the number of the peak-set for the three groups was almost the same (~3000 peaks). This clearly indicates that HF produces specific changes in chromatin conformation, driving the "opening", on average, of more promoter

regions respect to SD and CR. This epigenetic state could result in an aberrant regulation of some genes since H3K4me3 signal correlates mostly with active transcription, contributing to the development and progression of diseases like diabetes and cardiovascular diseases through the activation or suppression of gene functions (**Ke et al, 2009; He C et al, 2012**; **Chen Z et al, 2010**; **Raciti et al, 2014**; **Mathiyalagan et al, 2014**). This global result is consistent with local findings shown by **Inoue et al, 2014** and **Jun et al, 2012** that HF produced elevated levels of H3Kme3 signal on promoters of specific genes altering their expression in rats and mouse hepatocytes.

4.2.4. High fat diet induces changes in liver H3K4me3 profile promoting the onset of Type 2 Diabetes Mellitus

Noteworthy some of the regions showing an increased level of H3K4me3 in HF respect to SD correspond to genes involved in onset of Type II diabetes mellitus (Pik3r1, Socs3, Gck, Hk2, Prkcd).

Type II diabetes mellitus (T2DM) is a metabolic disorder characterized by high blood sugar due to pancreatic beta-cell functional impairment and insulin resistance in different tissues, including liver (**Dayeh et al, 2014**).

Though genetic variants are known to have a role in the development of T2DM (**Zeggini, 2007**), different lifestyle factors, including obesity, overweight, lack of physical activity, are reported as risk factors for T2DM onset (**Abdullah et al, 2010**). Analysis of DNA methylation status in pancreatic beta cells from diabetic and healthy individuals revealed epigenetic changes in approximately 850 genes confirming the presence of diabetes associated epigenetic modifications producing impaired insulin release (**Dayeh et al, 2014**).

MicroRNAs (miRNA) are also shown to be involved in glucose homeostasis and diabetes. For example, miRNA 21a has been shown to reverse high glucose and high insulin induced resistance in adipocytes modulating PTEN-AKT pathway (**Ling et al, 2012**) and to be over-expressed in diabetes patients (**Zeng et al, 2013**). Furthermore, hyperglycemia induced histone modifications and DNA methylation of pro-inflammatory genes triggering the vascular inflammation (**Villeneuve et al, 2010**). Lastly, **Jufvas et al, 2013** observed that adipocytes from type 2 diabetic and non-diabetic overweight subjects exhibited level of trimethylation at lysine 4 was 40% higher in adipocytes from overweight diabetic subjects compared with normal-weight and overweight non-diabetic subjects.

In this perspective, our findings are adding novel information regarding the role that HF-induced H3K4me3 liver profile may have in the onset of T2DM, since changes of histone modifications can result in aberrant gene expression.

Moreover, we found mir-21a with increased level of H3K4me3 and this result is coherent with its overexpression described in **Ling et al, 2012 and Zeng et al, 2013.**

4.2.5. ZSCAN4 could be the mediator of the detrimental effects of High Fat diet, acting on telomere shortening increasing the risk of T2DM development

We showed (even with low significant statistical value) that regions with elevated level of H3K4me3 in HF and involved in type 2 diabetes mellitus pathway are enriched for the Zinc Finger and SCAN Domain Containing 4 (Zscan4) transcription factor motif. Zscan4 has been demonstrated to be in charge of telomere elongation in ES cells and maintenance of genomic stability (**Zalzman et al, 2010**).

Kim et al, 2009 proved that weight gain and increased BMI positively correlate with telomere shortening in peripheral blood cells.

Interestingly, **Xiao et al, 2010** found that the average telomere length of type 2 diabetic patients was significantly shorter than the one of control subjects in a cohort of 930 patients and 867 controls. Indeed, experimental evidences suggest that telomerase is important in maintaining glucose homeostasis in mice (**Kuhlow, Florian, von Figura et al, 2010**). Conversely, elevated blood glucose levels increase oxidative stress, potentially interfering with telomerase function and resulting in shortened telomeres (**Serra et al, 2000**). Moreover, **Zhao et al, 2013** demonstrated that short telomere length is associated with future development of type 2 diabetes independently of known type 2 diabetes risk factors.

These evidences, together with our finding, seem to suggest a possible involvement of Zscan4 in HF induced detrimental effect through epigenetic regulation and of its role in telomere maintenance and its direct transcriptional action on specific genes' promoters involved in the onset of type 2 diabetes that were not previously proposed.

4.3. Conclusion and future perspectives

In conclusion, our study further elucidate the epigenetic link between caloric intake and disease risk/prevention, highlighting the impact of Calorie Restriction on circadian clock activity in liver and the detrimental effect of High Fat diet in the onset of Type II diabetes mellitus.

In particular, we hypothesized the involvement of two factors, REST and ZSCAN4, in mediating diet effects on chromatin remodeling, which in turn may results in changing of transcriptional regulation of specific genes.

Further directions will be to prove these hypotheses including analyses of ChIPseq

profiles anti-REST (in CR and SD samples) and anti-ZSCAN4 (in HF and SD samples) evaluating the activity and the differences of these two transcription factors in different diet conditions and to enlarge the number of RNA-seq samples and include also HF transcriptional profiles.

Moreover it would be interesting to investigate whether these diet-induced liver signatures are permanent or reversible, long lasting or transient and if they can be imprinted passed to following generations.

Bibliography

1. Abdullah, Asnawi et al. “The Magnitude of Association between Overweight and Obesity and the Risk of Diabetes: A Meta-Analysis of Prospective Cohort Studies.” Diabetes Research and Clinical Practice 89.3 (2010): 309–319. Web.

2. Alberts, Bruce. Molecular Biology of the Cell, 5th Edition. New York: Garland Science, 2008. Print.

3. Aleksandrova, K. et al. “Metabolic Syndrome And Risks of Colon and Rectal Cancer: The European Prospective Investigation into Cancer and Nutrition Study.” Cancer Prevention Research 4.11 (2011): 1873–1883.

4. Almendro, Vanessa, and Gemma Fuster. “Heterogeneity Of Breast Cancer: Etiology and Clinical Relevance.” Clin Transl Oncol Clinical and Translational Oncology 13.11 (2011): 767–773. Web.

5. Amador-Noguez, Daniel et al. “Alterations In Xenobiotic Metabolism in the Long-Lived Little Mice.” Aging Cell 6.4 (2007): 453–470. Web.

6. Amador-Noguez, Daniel et al. “Gene Expression Profile of Long-Lived Ames Dwarf Mice and Little Mice.” Aging Cell 3.6 (2004): 423–441. Web.

7. Anand, Preetha et al. “Cancer Is a Preventable Disease That Requires Major Lifestyle Changes.” Pharm Res Pharmaceutical Research 25.9 (2008): 2097–2116.

8. Anders, S., P. T. Pyl, and W. Huber. “HTSeq - A Python Framework to Work with High-Throughput Sequencing Data.” (2014): n. pag. Web.

9. Anders, Simon et al. “Count-Based Differential Expression Analysis of RNA Sequencing Data Using R and Bioconductor.” Nat Protoc Nature Protocols 8.9 (2013): 1765–1786. Web.

10. Anders, Simon, and Wolfgang Huber. “Differential Expression Analysis for Sequence Count Data.” Genome Biol Genome Biology 11.10 (2010): n. pag. Web.

11. Andres, M. E. et al. “CoREST: A Functional Corepressor Required for Regulation of Neural-Specific Gene Expression.” Proceedings of the National Academy of Sciences 96.17 (1999): 9873–9878. Web.

12. Bailey, T. L. et al. “MEME SUITE: Tools for Motif Discovery and Searching.” Nucleic Acids Research 37.Web Server (2009): n. pag. Web.

13. Baldi, Pierre, and G. Wesley Hatfield. “DNA Microarrays And Gene Expression.” (2002): n. pag. Web.

14. Balsalobre, Aurélio, Francesca Damiola, and Ueli Schibler. “A Serum Shock Induces Circadian Gene Expression In Mammalian Tissue Culture Cells.” Cell 93.6 (1998): 929–937. Web.

15. Bardet, Anaïs F et al. “A Computational Pipeline for Comparative ChIP-Seq Analyses.” Nat Protoc Nature Protocols 7.1 (2011): 45–61. Web.

16. Berg, Jeremy M., John L. Tymoczko, and Lubert Stryer. Biochemistry. New York: W. H. Freeman and Co., 2002. Print

17. Bergman, Yehudit, and Howard Cedar. “DNA Methylation Dynamics in Health and Disease.” Nat Struct Mol Biol Nature Structural & Molecular Biology 20.10 (2013): 1236–1236. Web.

18. Bernard, Samuel et al. “Synchronization-Induced Rhythmicity Of Circadian Oscillators in the Suprachiasmatic Nucleus.” PLoS Comput Biol PLoS Computational Biology preprint.2007 (2005): n. pag.

19. Bernstein, Bradley E. et al. “A Bivalent Chromatin Structure Marks Key Developmental Genes In Embryonic Stem Cells.” Cell 125.2 (2006): 315–326.

20. Bernstein, Bradley E. et al. “Genomic Maps And Comparative Analysis of Histone Modifications in Human and Mouse.” Cell 120.2 (2005): 169–181.

21. Berrino, F. et al. “Adjuvant Diet To Improve Hormonal and Metabolic Factors Affecting Breast Cancer Prognosis.” Annals of the New York Academy of Sciences 1089.1 (2006): 110–118.

22. Blecher-Gonen, Ronnie et al. “High-Throughput Chromatin Immunoprecipitation for Genome-Wide Mapping of in Vivo Protein-DNA Interactions and Epigenomic States.” Nat Protoc Nature Protocols 8.3 (2013): 539–554. Web.

23. Bodini, M. et al. “The Hidden Genomic Landscape of Acute Myeloid Leukemia: Subclonal Structure Revealed by Undetected Mutations.” Blood 125.4 (2014): 600–605. Web.

24. Bouvard, Véronique et al. “Carcinogenicity Of Consumption of Red and

Processed Meat." The Lancet Oncology (2015).

25. Brivanlou, A. H. "Signal Transduction And the Control of Gene Expression." Science 295.5556 (2002): 813–818.

26. Brooks, Christopher L., and Wei Gu. "How Does SIRT1 Affect Metabolism, Senescence and Cancer?" Nature Reviews Cancer Nat Rev Cancer 9.2 (2008): 123–128.

27. Bruce, A. W. et al. "Genome-Wide Analysis of Repressor Element 1 Silencing Transcription Factor/Neuron-Restrictive Silencing Factor (REST/NRSF) Target Genes." Proceedings of the National Academy of Sciences 101.28 (2004): 10458–10463. Web.

28. Budohoski, L., Panczenko-Kresowska, B., Langfort, J., et al "Effects of saturated and polyunsaturated fat enriched diet on the skeletal muscle insulin sensitivity in young rats." Journal of Physiology and Pharmacology 44 (1993): 391–398

29. Canaple, Laurence et al. "Reciprocal Regulation Of Brain and Muscle Arnt-Like Protein 1 and Peroxisome Proliferator-Activated Receptor α Defines a Novel Positive Feedback Loop in the Rodent Liver Circadian Clock." Molecular Endocrinology 20.8 (2006): 1715–1727. Web.

30. Challet E, Solberg LC, Turek FW. Entrainment in calorie-restricted mice: conflicting zeitgebers and free-running conditions. Am J Physiol. 1998; 274:R1751-R1761.

31. Challet, E. et al. "Synchronization Of the Molecular Clockwork by Light- and Food-Related Cues in Mammals." Biological Chemistry 384.5 (2003): n. pag. Web.

32. Chapman, T., and L. Partridge. "Female Fitness In Drosophila Melanogaster: An Interaction between the Effect of Nutrition and of

Encounter Rate with Males." Proceedings of the Royal Society B: Biological Sciences 263.1371 (1996): 755–759. Web.

33. Chen, Danica et al. "The Role of Calorie Restriction and SIRT1 in Prion-Mediated Neurodegeneration." Experimental Gerontology 43.12 (2008): 1086–1093. Web.

34. Chen, Zhong et al. "Histone Modifications and Chromatin Organization in Prostate Cancer." Epigenomics 2.4 (2010): 551–560.

35. Cherrington, A. D. "Banting Lecture 1997. Control Of Glucose Uptake and Release by the Liver in Vivo." Diabetes 48.5 (1999): 1198–1214. Web.

36. Chevalier, S. et al. "Dietary Restriction Alters Retinol And Retinol-Binding Protein Metabolism in Aging Rats." The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 54.9 (1999): n. pag. Web.

37. Choi, Youngshim, Cheol-Goo Hur, and Taesun Park. "Induction Of Olfaction and Cancer-Related Genes in Mice Fed a High-Fat Diet as Assessed through the Mode-of-Action by Network Identification Analysis." PLoS ONE 8.3 (2013): n. pag. Web.

38. Chong, Jayhong A et al. "REST: A Mammalian Silencer Protein That Restricts Sodium Channel Gene Expression to Neurons." Cell 80.6 (1995): 949–957. Web.

39. Christopher P. Adams, Stephen Joseph Kron, "Method for performing amplification of nucleic acid with two primers bound to a single solid support. Inventors", US Patent 5,641,658, (1994)

40. Clinthorne, J. F. et al. "NK Cell Maturation And Function in C57BL/6 Mice Are Altered by Caloric Restriction." The Journal of Immunology 190.2 (2012): 712–722. Web.

41. Cock, P. J. A. et al. “The Sanger FASTQ File Format for Sequences with Quality Scores, and the Solexa/Illumina FASTQ Variants.” Nucleic Acids Research 38.6 (2009): 1767–1771. Web.

42. Colman, R. J. et al. “Caloric Restriction Delays Disease Onset And Mortality in Rhesus Monkeys.” Science 325.5937 (2009): 201–204.

43. Corton, J. C., and H. M. Brown-Borg. “Peroxisome Proliferator-Activated Receptor Coactivator 1 In Caloric Restriction and Other Models of Longevity.” The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 60.12 (2005): 1494–1509. Web.

44. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Langerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Borresen-Dale AL, Brenton JD, Tavare S, Caldas C, Aparicio S: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012, 486:346–352.

45. Damiola, F. “Restricted Feeding Uncouples Circadian Oscillators in Peripheral Tissues from the Central Pacemaker in the Suprachiasmatic Nucleus.” Genes and Development 14.23 (2000): 2950–2961.

46. Dayeh, Tasnim et al. “Genome-Wide DNA Methylation Analysis Of Human Pancreatic Islets from Type 2 Diabetic and Non-Diabetic Donors Identifies Candidate Genes That Influence Insulin Secretion.” PLoS Genetics PLoS Genet 10.3 (2014): n. pag. Web.

47. Deluca, D. S. et al. “RNA-SeQC: RNA-Seq Metrics for Quality Control and Process Optimization.” Bioinformatics 28.11 (2012): 1530–1532. Web.

48. Di Lorenzo, L et al. “Effect Of Shift Work on Body Mass Index: Results of a Study Performed in 319 Glucose-Tolerant Men Working in a Southern Italian Industry.” Int J Obes Relat Metab Disord International Journal of Obesity 27.11 (2003): 1353–1358.

49. Dick, Katherine J et al. “DNA Methylation and Body-Mass Index: a Genome-Wide Analysis.” The Lancet 383.9933 (2014): 1990–1998.

50. Dobin, A. et al. “STAR: Ultrafast Universal RNA-Seq Aligner.” Bioinformatics 29.1 (2012): 15–21. Web.

51. Duan, W. et al. “Dietary Restriction Normalizes Glucose Metabolism and BDNF Levels, Slows Disease Progression, and Increases Survival in Huntingtin Mutant Mice.” Proceedings of the National Academy of Sciences 100.5 (2003): 2911–2916. Web.

52. Duffy, P et al. “Effect Of Chronic Caloric Restriction on Physiological Variables Related to Energy Metabolism in the Male Fischer 344 Rat.” Mechanisms of Ageing and Development 48.2 (1989): 117–133. Web.

53. Eckel-Mahan, K., and P. Sassone-Corsi. “Metabolism And the Circadian Clock Converge.” Physiological Reviews 93.1 (2013): 107–135.

54. Eckel-Mahan, Kristin L. et al. “Reprogramming Of the Circadian Clock by Nutritional Challenge.” Cell 155.7 (2013): 1464–1478.

55. El Bacha, T., Luz, M., Da Poian, A. “Dynamic Adaptation of Nutrient Utilization in Humans.” Nature Education 3.9 (2010): 8

56. Engelen, Erik et al. “Proteins That Bind Regulatory Regions Identified by Histone Modification Chromatin Immunoprecipitations and Mass Spectrometry.” Nature Communications Nat Comms 6 (2015): 7155. Web.

57. Esposito, K. et al. “Metabolic Syndrome And Risk of Cancer: A Systematic Review and Meta-Analysis.” *Diabetes Care* 35.11 (2012): 2402–2411. Web.

58. Esteller, Manel. “Cancer Epigenomics: DNA Methylomes and Histone-Modification Maps.” Nat Rev Genet Nature Reviews Genetics 8.4 (2007): 286–298.

59. Fanelli, M. et al. “Pathology Tissue-Chromatin Immunoprecipitation, Coupled with High-Throughput Sequencing, Allows the Epigenetic Profiling of Patient Samples.” Proceedings of the National Academy of Sciences 107.50 (2010): 21535–21540. Web.

60. Fanelli, Mirco et al. “Chromatin Immunoprecipitation and High-Throughput Sequencing from Paraffin-Embedded Pathology Tissue.” Nat Protoc Nature Protocols 6.12 (2011): 1905–1919. Web.

61. Faure, A. J. et al. “Cohesin Regulates Tissue-Specific Expression by Stabilizing Highly Occupied Cis-Regulatory Modules.” Genome Research 22.11 (2012): 2163–2175. Web.

62. Felsenfeld, Gary, and Mark Groudine. “Controlling The Double Helix.” Nature 421.6921 (2003): 448–453.

63. Ferland, G. et al. “Effect Of Dietary Restriction on Hepatic Vitamin a Content in Aging Rats.” Journal of Gerontology 47.1 (1992): n. pag. Web.

64. Flicek, Paul, and Ewan Birney. “Sense From Sequence Reads: Methods for Alignment and Assembly.” Nature Methods Nat Meth 7.6 (2010): 479–479. Web.

65. Fontana, L., L. Partridge, and V. D. Longo. “Extending Healthy Life Span--From Yeast To Humans.” Science 328.5976 (2010): 321–326. Web.

66. Froy, O. “Circadian Rhythms, Aging, And Life Span in Mammals.” Physiology 26.4 (2011): 225–235. Web.

67. Froy, O., N. Chapnik, and R. Miskin. “Long-Lived MUPA Transgenic Mice Exhibit Pronounced Circadian Rhythms.” AJP: Endocrinology and Metabolism 291.5 (2006): n. pag. Web.

68. Froy, Oren, and Ruth Miskin. “The Interrelations among Feeding, Circadian Rhythms and Ageing.” Progress in Neurobiology 82.3 (2007): 142–150. Web.

69. Gallou-Kabani, Catherine et al. “Nutri-Epigenomics: Lifelong Remodelling of Our Epigenomes by Nutritional and Metabolic Factors and Beyond.” Clinical Chemical Laboratory Medicine 45.3 (2007)

70. Gami, Apoor S. et al. “Metabolic Syndrome And Risk of Incident Cardiovascular Events and Death.” Journal of the American College of Cardiology 49.4 (2007): 403–414.

71. Garriga-Canut, Mireia et al. “2-Deoxy-D-Glucose Reduces Epilepsy Progression by NRSF-CtBP–Dependent Metabolic Regulation of Chromatin Structure.” Nature Neuroscience Nat Neurosci 9.11 (2006): 1382–1387. Web.

72. Giovannucci E, “Metabolic syndrome, hyperinsulinemia, and colon cancer: a review.” American Journal of Clinical Nutrition 86(2007): 836-42

73. Grant P.A., “A tale of histone modifications”, Genome Biology 2.4(2001)

74. Graw, Stefan et al. “Robust Gene Expression and Mutation Analyses of RNA-Sequencing of Formalin-Fixed Diagnostic Tumor Samples.” Sci. Rep. Scientific Reports 5 (2015): 12335. Web.

75. Greer, Eric L., and Anne Brunet. “Different Dietary Restriction Regimens Extend Lifespan by Both Independent and Overlapping Genetic Pathways in C. Elegans.” Aging Cell 8.2 (2009): 113–127. Web.

76. Grimes, J. A. “The Co-Repressor mSin3A Is a Functional Component of the REST-CoREST Repressor Complex.” Journal of Biological Chemistry 275.13 (2000): 9461–9467. Web.

77. Guenther, Matthew G. et al. “A Chromatin Landmark And Transcription Initiation at Most Promoters in Human Cells.” Cell 130.1 (2007): 77–88.

78. Guerre-Millo, M. et al. “PPAR- -Null Mice Are Protected From High-Fat Diet-Induced Insulin Resistance.” Diabetes 50.12 (2001): 2809–2814. Web.

79. Gupta, Shobhit et al. “Quantifying Similarity between Motifs.” Genome Biol Genome Biology 8.2 (2007): n. pag. Web.

80. Hagopian, K. “Caloric Restriction Increases Gluconeogenic and Transaminase Enzyme Activities in Mouse Liver.” Experimental Gerontology 38.3 (2003): 267–278. Web.

81. Hansen, Kasper Daniel et al. “Increased Methylation Variation in Epigenetic Domains across Cancer Types.” Nature Genetics Nat Genet 43.8 (2011): 768–775.

82. Harris, R. B., Kor, H., “Insulin insensitivity is rapidly reversed in rats by reducing dietary fat from 40 to 30% of energy”, Nutrition 122 (1992): 1811–1822.

83. Harvie, Michelle, and Anthony Howell. “Energy Restriction and the Prevention of Breast Cancer.” Proceedings of the Nutrition Society Proc. Nutr. Soc. 71.02 (2012): 263–275.

84. Hayakawa, Tomohiro, and Jun-Ichi Nakayama. “Physiological Roles Of Class I HDAC Complex and Histone Demethylase.” Journal of Biomedicine and Biotechnology 2011 (2011): 1–10. Web.

85. He, Chuanchao et al. “High Expression of Trimethylated Histone H3 Lysine 4 Is Associated with Poor Prognosis in Hepatocellular Carcinoma.” Human Pathology 43.9 (2012): 1425–1435.

86. Heijmans, B. T. et al. “Persistent Epigenetic Differences Associated with Prenatal Exposure to Famine in Humans.” Proceedings of the National Academy of Sciences 105.44 (2008): 17046–17049.

87. Heilbronn LK, Ravussin E., “Calorie restriction and aging: review of the literature and implications for studies in humans”, American Journal of Clinical Nutrition 78.3 (2003): 361-9.

88. Hogeweg, Paulien. “The Roots Of Bioinformatics in Theoretical Biology.” PLoS Comput Biol PLoS Computational Biology 7.3 (2011): n. pag. Web.

89. Huang, Y., Myers, S. J. & Dingledine, R. Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes. Nature Neurosci. 2, 867–872 (1999).

90. Human Genome project, http://www.genome.gov/10001772

91. Hunt, Nicole D. et al. “Bioenergetics Of Aging and Calorie Restriction.” Ageing Research Reviews 5.2 (2006): 125–143. Web.

92. Hursting, S. D. et al. “Calories And Carcinogenesis: Lessons Learned from 30 Years of Calorie Restriction Research.” Carcinogenesis 31.1 (2009): 83–89.

93. Hursting, Stephen D, and Sarah M Dunlap. “Nutrition And Physical Activity in Aging, Obesity, and Cancer.” Annals of the New York Academy of Sciences. Blackwell Publishing Inc, n.d. 6 Aug. 2015. <http://www.ncbi.nlm.nih.gov/pmc/articles/pmc3485672/>

94. IDF "Worldwide Definition Of the Metabolic Syndrome." International Diabetes Federation. 6 Aug. 2015. <http://www.idf.org/metabolic-syndrome>

95. Imayama, I. et al. "Effects Of a Caloric Restriction Weight Loss Diet and Exercise on Inflammatory Biomarkers in Overweight/Obese Postmenopausal Women: A Randomized Controlled Trial." Cancer Research 72.9 (2012): 2314–2326.

96. Inoue, S. et al. " Induction of histone H3K4 methylation at the promoter, enhancer, and transcribed regions of the Si and Sglt1 genes in rat jejunum in response to a high-starch/low-fat diet", Nutrition 2015 Feb; 31(2):366-72

97. Iqbal, K. et al. "Reprogramming Of the Paternal Genome upon Fertilization Involves Genome-Wide Oxidation of 5-Methylcytosine."Proceedings of the National Academy of Sciences 108.9 (2011): 3642–3647. Web

98. Jones, J. R. et al. "Deletion Of PPAR in Adipose Tissues of Mice Protects against High Fat Diet-Induced Obesity and Insulin Resistance." Proceedings of the National Academy of Sciences 102.17 (2005): 6207–6212. Web.

99. Jun, HJ et al. " Hepatic Lipid Accumulation Alters Global Histone H3 Lysine 9 and 4 Trimethylation in the Peroxisome Proliferator-Activated Receptor Alpha Network". PLoS One. 2012;7(9):e44345.

100. Jufvas, Åsa et al. "Global Differences in Specific Histone H3 Methylation Are Associated with Overweight and Type 2 Diabetes." Clin Epigenetics Clinical Epigenetics 5.1 (2013): 15. Web.

101. Kass, Stefan U., Dmitry Pruss, and Alan P. Wolffe. "How Does DNA Methylation Repress Transcription?"Trends in Genetics 13.11 (1997): 444–449. Web.

102. Ke, Xi-Song et al. “Genome-Wide Profiling Of Histone H3 Lysine 4 and Lysine 27 Trimethylation Reveals an Epigenetic Signature in Prostate Carcinogenesis.” PLoS ONE 4.3 (2009): n. pag.

103. Kennedy, B. K., K. K. Steffen, and M. Kaeberlein. “Ruminations On Dietary Restriction and Aging.” Cell. Mol. Life Sci. Cellular and Molecular Life Sciences 64.11 (2007): 1323–1328. Web.

104. Kent, W. J. et al. “The Human Genome Browser At UCSC.” Genome Research 12.6 (2002): 996–1006. Web.

105. Kersten, Sander et al. “Peroxisome Proliferator–Activated Receptor α Mediates the Adaptive Response to Fasting.” Journal of Clinical Investigation J. Clin. Invest. 103.11 (1999): 1489–1498. Web.

106. Kim, S. et al. “Obesity And Weight Gain in Adulthood and Telomere Length.” Cancer Epidemiology Biomarkers & Prevention 18.3 (2009): 816–820. Web.

107. Kim, Tae Hoon et al. “A High-Resolution Map of Active Promoters in the Human Genome.” Nature 436.7052 (2005): 876–880.

108. Kucharski, R. et al. “Nutritional Control Of Reproductive Status in Honeybees via DNA Methylation.” Science 319.5871 (2008): 1827–1830.

109. Kuhlow D, Florian S, von Figura G, et al. Telomerase deficiency impairs glucose metabolism and insulin secretion. Aging (Albany, NY Online) 2010; 2:650–658

110. Kundaje et al, “Integrative analysis of 111 reference human epigenomes”, Nature 518, 317–330 (2015)

111. Laland, Kevin et al. “Does Evolutionary Theory Need a Rethink?” Nature 514.7521 (2014): 161–164. Web.

112. Latchman, David S. “Transcription Factors: An Overview.” The International Journal of Biochemistry & Cell Biology 29.12 (1997): 1305–1312.

113. Laurent Farinelli, Eric Kawashima, Pascal Mayer, "Method of nucleic acid amplification", published 1998-10-08

114. Leone, T. C., C. J. Weinheimer, and D. P. Kelly. “A Critical Role for the Peroxisome Proliferator-Activated Receptor (PPAR) in the Cellular Fasting Response: The PPAR -Null Mouse as a Model of Fatty Acid Oxidation Disorders.” Proceedings of the National Academy of Sciences 96.13 (1999): 7473–7478. Web.

115. Leung, A. et al. “Open Chromatin Profiling In Mice Livers Reveals Unique Chromatin Variations Induced by High Fat Diet.” Journal of Biological Chemistry 289.34 (2014): 23557–23567.

116. Lewis, G. F. et al. “Fatty Acids Mediate the Acute Extrahepatic Effects of Insulin on Hepatic Glucose Production in Humans.” Diabetes 46.7 (1997): 1111–1119. Web.

117. Li, Bing, Michael Carey, and Jerry L. Workman. “The Role Of Chromatin during Transcription.” Cell 128.4 (2007): 707–719.

118. Li, H. et al. “The Sequence Alignment/Map Format and SAMtools.” Bioinformatics 25.16 (2009): 2078–2079. Web.

119. Li, H., and R. Durbin. “Fast And Accurate Short Read Alignment with Burrows-Wheeler Transform.” Bioinformatics 25.14 (2009): 1754–1760. Web.

120. Lin, Hua V., and Domenico Accili. “Hormonal Regulation Of Hepatic Glucose Production in Health and Disease.” Cell Metabolism 14.1 (2011): 9–19. Web.

121. Lin, S.-J. “Calorie Restriction Extends Yeast Life Span by Lowering the Level of NADH.” Genes & Development 18.1 (2004): 12–16. Web.

122. Lin, S.-J. “Requirement Of NAD and SIR2 for Life-Span Extension by Calorie Restriction in Saccharomyces Cerevisiae.” Science 289.5487 (2000): 2126–2128. Web.

123. Ling, H.-Y. et al. “MiRNA-21 Reverses High Glucose And High Insulin Induced Insulin Resistance in 3T3-L1 Adipocytes through Targeting Phosphatase and Tensin Homologue.” Exp Clin Endocrinol Diabetes Experimental and Clinical Endocrinology & Diabetes 120.09 (2012): 553–559. Web.

124. Lister R, Pelizzola M, Dowen RH, et al. "Human DNA methylomes at base resolution show widespread epigenomic differences". Nature. 2009; 462(7271):315-22

125. Lu, Tao et al. “REST And Stress Resistance in Ageing and Alzheimer’s Disease.” Nature 507.7493 (2014): 448–454. Web.

126. Machanick, P., and T. L. Bailey. “MEME-ChIP: Motif Analysis of Large DNA Datasets.” Bioinformatics 27.12 (2011): 1696–1697. Web.

127. Maher, Christopher A. et al. “Transcriptome Sequencing to Detect Gene Fusions in Cancer.” Nature 458.7234 (2009): 97–101. Web.

128. Mair, William, and Andrew Dillin. “Aging And Survival: The Genetics of Life Span Extension by Dietary Restriction.” Annu. Rev. Biochem. Annual Review of Biochemistry 77.1 (2008): 727–754. Web.

129. Marks, Paul A. et al. “Histone Deacetylases And Cancer: Causes And Therapies.” Nature Reviews Cancer Nat. Rev. Cancer. 1.3 (2001): 194–202.

130. Masoro, Edward J. “Overview Of Caloric Restriction and Ageing.” Mechanisms of Ageing and Development 126.9 (2005): 913–922. Web.

131. Masternak, M. M. et al. “Divergent Effects Of Caloric Restriction on Gene Expression in Normal and Long-Lived Mice.” The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 59.8 (2004): n. pag. Web.

132. Masternak, M. M. et al. “Effects Of Caloric Restriction and Growth Hormone Resistance on the Expression Level of Peroxisome Proliferator-Activated Receptors Superfamily in Liver of Normal and Long-Lived Growth Hormone Receptor/Binding Protein Knockout Mice.” The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 60.11 (2005): 1394–1398. Web.

133. Masternak, Michal M., and Andrzej Bartke. “PPARs In Calorie Restricted and Genetically Long-Lived Mice.” PPAR Research 2007 (2007): 1–7. Web.

134. Maston, Glenn A., Sara K. Evans, and Michael R. Green. “Transcriptional Regulatory Elements In the Human Genome.” Annual Review of Genomics and Human Genetics Annu. Rev. Genom. Human Genet. 7.1 (2006): 29–59. Web.

135. Mathiyalagan, P. et al. “Chromatin Modifications Remodel Cardiac Gene Expression.” Cardiovascular Research 103.1 (2014): 7–16.

136. Mattison, Julie A. et al. “Impact Of Caloric Restriction on Health and Survival in Rhesus Monkeys from the NIA Study.” Nature 489.7415 (2012): 318–321.

137. Mayer P et al., presented at the Fifth International Automation in Mapping and DNA Sequencing Conference, St. Louis, MO, USA (October 7–10, 1998). DNA colony massively parallel sequencing ams98 presentation "A

very large scale, high throughput and low cost DNA sequencing method based on a new 2-dimensional DNA auto-patterning process"

138. Maze, Ian et al. “Analytical Tools and Current Challenges in the Modern Era of Neuroepigenomics.” Nature Neuroscience Nat Neurosci 17.11 (2014): 1476–1490. Web.

139. McCay C.M., Crowel M.F., Maynard L.A. “The effect of retarded growth upon the length of the life span and upon the ultimate body size.” Nutrition 10 (1935): 63–79

140. Mcelwee, Joshua J. et al. “Shared Transcriptional Signature In Caenorhabditis Elegans Dauer Larvae and Long-Lived Daf-2 Mutants Implicates Detoxification System in Longevity Assurance.” Journal of Biological Chemistry J. Biol. Chem. 279.43 (2004): 44533–44543. Web.

141. Mendoza, J. “Feeding Cues Alter Clock Gene Oscillations And Photic Responses in the Suprachiasmatic Nuclei of Mice Exposed to a Light/Dark Cycle.” Journal of Neuroscience 25.6 (2005): 1514–1522. Web.

142. Merry, B. J. “Oxidative Stress and Mitochondrial Function with Aging - the Effects of Calorie Restriction.” Aging Cell 3.1 (2004): 7–12. Web.

143. Mitchell, P., and R Tjian. “Transcriptional Regulation in Mammalian Cells by Sequence-Specific DNA Binding Proteins.” Science 245.4916 (1989): 371–378.

144. Mombaerts, Peter. “Molecular Biology Of Odorant Receptors In Vertebrates.” Annu. Rev. Neurosci. Annual Review of Neuroscience 22.1 (1999): 487–509. Web.

145. Müller, Michael, and Sander Kersten. “Opinion: Nutrigenomics: Goals and Strategies.” Nat Rev Genet Nature Reviews Genetics 4.4 (2003): 315–322. Web.

146. Mutch, D. M. “Nutrigenomics And Nutrigenetics: the Emerging Faces of Nutrition.” The FASEB Journal 19.12 (2005): 1602–1616. Web.

147. Narlikar, Geeta J., Hua-Ying Fan, and Robert E. Kingston. “Cooperation Between Complexes That Regulate Chromatin Structure and Transcription.” Cell 108.4 (2002): 475–487.

148. Naruse, Y. et al. “Neural Restrictive Silencer Factor Recruits mSin3 and Histone Deacetylase Complex to Repress Neuron-Specific Target Genes.” Proceedings of the National Academy of Sciences 96.24 (1999): 13691–13696. Web.

149. Ngollo, Marjolaine et al. “Epigenetic Modifications in Prostate Cancer.” Epigenomics 6.4 (2014): 415–426.

150. Obici, Silvana, and Luciano Rossetti. “Minireview: Nutrient Sensing And the Regulation of Insulin Action and Energy Balance.” Endocrinology 144.12 (2003): 5172–5178. Web.

151. Oishi, Katsutaka, Hidenori Shirai, and Norio Ishida. “CLOCK Is Involved in the Circadian Transactivation of Peroxisome-Proliferator-Activated Receptor α (PPAR α) in Mice.” Biochem. J. Biochemical Journal 386.3 (2005): 575–581. Web.

152. Ooi, Lezanne, and Ian C. Wood. “Chromatin Crosstalk in Development and Disease: Lessons from REST.” Nat Rev Genet Nature Reviews Genetics 8.7 (2007): 544–554. Web.

153. Orlando, V. “Mapping Chromosomal Proteins in Vivo by Formaldehyde-Crosslinked-Chromatin Immunoprecipitation.” Trends in Biochemical Sciences 25.3 (2000): 99–104. Web.

154. Pais, Raluca. “Metabolic Syndrome and Risk of Subsequent Colorectal Cancer.” World Journal of Gastroenterology WJG 15.41 (2009): 5141.

155. Palouzier-Paulignan, B. et al. “Olfaction Under Metabolic Influences.” Chemical Senses 37.9 (2012): 769–797. Web.

156. Park, Lara K. et al. " Nutritional influences on epigenetics and age-related disease", Proceedings of the Nutrition Society, Vol. 71, iss. 01, February 2012, pp 75-83

157. Parmentier, Marc et al. “Expression Of Members of the Putative Olfactory Receptor Gene Family in Mammalian Germ Cells.” Nature 355.6359 (1992): 453–455. Web.

158. Patel, Nilay V. et al. “Caloric Restriction Attenuates Aβ-Deposition in Alzheimer Transgenic Models.” Neurobiology of Aging 26.7 (2005): 995–1000. Web.

159. Pendergast, Julie S. et al. “Robust Food Anticipatory Activity In BMAL1-Deficient Mice.” PLoS ONE 4.3 (2009): n. pag.

160. Pitts, Sinae, Elizabeth Perone, and Rae Silver. “Food-Entrained Circadian Rhythms Are Sustained in Arrhythmic Clk/Clk Mutant Mice.” American Journal of Physiology - Regulatory, Integrative and Comparative Physiology Am J Physiol Regul Integr Comp Physiol 285.1 (2003): n. pag.

161. Raciti, Gregory Alexander et al. “Personalized Medicine and Type 2 Diabetes: Lesson from Epigenetics.” Epigenomics 6.2 (2014): 229–238.

162. Randy S. Levinson,C. Ron Kahn, Domenico Accili, “Metabolic Syndrom ePoster”, Nature Medicine, http://www.nature.com/nm/e-poster/eposter_full.html

163. Resuehr, David, and James Olcese. “Caloric Restriction and Melatonin Substitution: Effects on Murine Circadian Parameters.” Brain Research 1048.1-2 (2005): 146–152. Web.

164. Rice Genome project, http://rice.plantbiology.msu.edu/

165. Richardson, Brynn E. et al. “Altered Olfactory Acuity In the Morbidly Obese.” Obesity Surgery Obes Surg 14.7 (2004): 967–969. Web.

166. Richardson, Brynn E. et al. “Gastric Bypass Does Not Influence Olfactory Function In Obese Patients.” Obesity Surgery OBES SURG 22.2 (2011): 283–286. Web.

167. Riera, Celine E., and Andrew Dillin. “Tipping The Metabolic Scales towards Increased Longevity in Mammals.” Nature Cell Biology Nat Cell Biol 17.3 (2015): 196–203.

168. Riva, L et al. “Acute Promyelocytic Leukemias Share Cooperative Mutations with Other Myeloid-Leukemia Subgroups.” Blood Cancer J Blood Cancer Journal 4.3 (2014): n. pag. Web.

169. Robinson, M. D., D. J. Mccarthy, and G. K. Smyth. “EdgeR: a Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data.” Bioinformatics 26.1 (2009): 139–140. Web.

170. Robinson, Mark D, and Alicia Oshlack. “A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data.” Genome Biol Genome Biology 11.3 (2010): n. pag. Web.

171. Ronaghi, Mostafa et al. “Real-Time DNA Sequencing Using Detection Of Pyrophosphate Release.” Analytical Biochemistry 242.1 (1996): 84–89. Web.

172. Roopra, A. et al. “Transcriptional Repression By Neuron-Restrictive Silencer Factor Is Mediated via the Sin3-Histone Deacetylase Complex.” Molecular and Cellular Biology 20.6 (2000): 2147–2157. Web.

173. Samuel, N., and T. J. Hudson. “Translating Genomics To the Clinic: Implications of Cancer Heterogeneity.” Clinical Chemistry 59.1 (2012): 127–137. Web.

174. Samuels, Leo T., Roger M. Reinecke, and Howard A. Ball. “Effect Of Diet On Glucose Tolerance And Liver And Muscle Glycogen Of Hypophysectomized And Normal Rats 1 , 2.” Endocrinology 31.1 (1942): 42–45.

175. Sanger, F., S. Nicklen, and A. R. Coulson. “DNA Sequencing with Chain-Terminating Inhibitors.” Proceedings of the National Academy of Sciences 74.12 (1977): 5463–5467. Web.

176. Schones, Dustin E., and Keji Zhao. “Genome-Wide Approaches to Studying Chromatin Modifications.” Nat Rev Genet Nature Reviews Genetics 9.3 (2008): 179–191. Web.

177. Serra, Violeta et al. “Telomere Length As a Marker Of Oxidative Stress in Primary Human Fibroblast Cultures.” Annals of the New York Academy of Sciences 908.1 (2000): 327–330. Web.

178. Shen, Li et al. “Ngs.Plot: Quick Mining and Visualization of next-Generation Sequencing Data by Integrating Genomic Databases.” BMC Genomics 15.1 (2014): 284. Web.

179. Shen, Zhanlong et al. “Metabolic Syndrome Is an Important Factor for the Evolution of Prognosis of Colorectal Cancer: Survival, Recurrence, and Liver Metastasis.” The American Journal of Surgery 200.1 (2010): 59–63.

180. Shlyueva, Daria, Gerald Stampfel, and Alexander Stark. “Transcriptional Enhancers: from Properties to Genome-Wide Predictions.” Nat Rev Genet Nature Reviews Genetics 15.4 (2014): 272–286.

181. Simchen, U et al. “Odour And Taste Sensitivity Is Associated with Body Weight and Extent of Misreporting of Body Weight.” European Journal of Clinical Nutrition Eur J Clin Nutr 60.6 (2006): 698–705. Web.

182. Smith, Brian C., and John M. Denu. “Chemical Mechanisms of Histone Lysine and Arginine Modifications.” Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1789.1 (2009): 45–57.

183. Smith, E., and H. J. Morowitz. “Universality In Intermediary Metabolism.” Proceedings of the National Academy of Sciences 101.36 (2004): 13168–13173.

184. Smith, Zachary D., and Alexander Meissner. “DNA Methylation: Roles in Mammalian Development.”Nat Rev Genet Nature Reviews Genetics 14.3 (2013): 204–220. Web.

185. Smith, Zachary D., and Alexander Meissner. “DNA Methylation: Roles in Mammalian Development.”Nat Rev Genet Nature Reviews Genetics 14.3 (2013): 204–220. Web.

186. Swindell, William R. “Comparative Analysis of Microarray Data Identifies Common Responses to Caloric Restriction among Mouse Tissues.” Mechanisms of Ageing and Development 129.3 (2008): 138–153. Web.

187. Swine Genome sequencing project, https://www.sanger.ac.uk/resources/downloads/othervertebrates/pig.html

188. Takamura, Toshinari et al. “Obesity Upregulates Genes Involved In Oxidative Phosphorylation in Livers of Diabetic Patients.” Obesity 16.12 (2008): 2601–2609. Web.

189. The mouse ENCODE consortium et al. “A Comparative Encyclopedia of DNA Elements in the Mouse Genome” Nature 515, 355–364 (2014)

190. Thondamal, Manjunatha et al. “Steroid Hormone Signalling Links Reproduction to Lifespan in Dietary-Restricted Caenorhabditis Elegans.” Nature Communications Nat Comms 5 (2014): 4879. Web.

191. Trapnell, C., L. Pachter, and S. L. Salzberg. “TopHat: Discovering Splice Junctions with RNA-Seq.” Bioinformatics 25.9 (2009): 1105–1111. Web.

192. Trapnell, Cole et al. “Transcript Assembly and Quantification by RNA-Seq Reveals Unannotated Transcripts and Isoform Switching during Cell Differentiation.” Nat Biotechnol Nature Biotechnology 28.5 (2010): 511–515. Web.

193. Tsuchiya, T. “Additive Regulation of Hepatic Gene Expression by Dwarfism and Caloric Restriction.” Physiological Genomics 17.3 (2004): 307–315. Web.

194. Urdinguio, Rocio G, Jose V Sanchez-Mut, and Manel Esteller. “Epigenetic Mechanisms in Neurological Diseases: Genes, Syndromes, and Therapies.” The Lancet Neurology 8.11 (2009): 1056–1072.

195. Villeneuve, L. M., and R. Natarajan. “The Role of Epigenetics in the Pathology of Diabetic Complications.” AJP: Renal Physiology 299.1 (2010): n. pag. Web.

196. Wang, Zhong, Mark Gerstein, and Michael Snyder. “RNA-Seq: a Revolutionary Tool for Transcriptomics.” Nat Rev Genet Nature Reviews Genetics 10.1 (2009): 57–63. Web.

197. Weber LW, Boll M, Stampfl A. "Maintaining cholesterol homeostasis: sterol regulatory element-binding proteins". World Journal of Gastroenterology 10.21 (2004): 3081–7

198. Weindruch, R. et al. “Influences Of Aging and Dietary Restriction on Serum Thymosin I Levels in Mice.” Journal of Gerontology 43.2 (1988): n. pag. Web.

199. Welsh, David K., Joseph S. Takahashi, and Steve A. Kay. “Suprachiasmatic Nucleus: Cell Autonomy And Network Properties.” Annual Review of Physiology Annu. Rev. Physiol. 72.1 (2010): 551–577. Web.

200. WHO, “Obesity And Overweight.”. 6 Aug. 2015. <http://www.who.int/mediacentre/factsheets/fs311/en/>

201. Wilbanks, Elizabeth G., and Marc T. Facciotti. “Evaluation Of Algorithm Performance in ChIP-Seq Peak Detection.” PLoS ONE 5.7 (2010): n. pag. Web.

202. Wolff GL, Kodell RL, Moore SR, Cooney CA, “Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice”, FASEB Journal 12.11 (1998):949-57.

203. Xiao, F. et al. “Telomere Dysfunction-Related Serological Markers Are Associated With Type 2 Diabetes.” Diabetes Care 34.10 (2011): 2273–2278. Web.

204. Xu, Lan, Christopher K Glass, and Michael G Rosenfeld. “Coactivator And Corepressor Complexes in Nuclear Receptor Function.” Current Opinion in Genetics and Development 9.2 (1999): 140–147.

205. Yamazaki, S. “Resetting Central And Peripheral Circadian Oscillators in Transgenic Rats.” Science 288.5466 (2000): 682–685. Web.

206. Yancovitz, Molly et al. “Intra- And Inter-Tumor Heterogeneity of BRAF(V600E) Mutations in Primary and Metastatic Melanoma.” PLoS ONE 7.1 (2012): n. pag. Web.

207. Yoo, S.-H. et al. “PERIOD2::LUCIFERASE Real-Time Reporting of Circadian Dynamics Reveals Persistent Circadian Oscillations in Mouse Peripheral Tissues.” Proceedings of the National Academy of Sciences 101.15 (2004): 5339–5346. Web.

208. Young, J. M. “The Sense of Smell: Genomics of Vertebrate Odorant Receptors.” Human Molecular Genetics 11.10 (2002): 1153–1160. Web.

209. Yu, Guangchuang et al. “ClusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters.” OMICS: A Journal of Integrative Biology 16.5 (2012): 284–287. Web.

210. Yu, Guangchuang, Li-Gen Wang, and Qing-Yu He. “ChIPseeker: An R/Bioconductor Package for ChIP Peak Annotation, Comparison and Visualization.” Bioinformatics 31.14 (2015): 2382–2383. Web.

211. Zalzman, Michal et al. “Zscan4 Regulates Telomere Elongation and Genomic Stability in ES Cells.” Nature 464.7290 (2010): 858–863. Web.

212. Zeggini, E. “A New Era for Type 2 Diabetes Genetics.” Diabetic Medicine Diabetic Med 24.11 (2007): 1181–1186. Web.

213. Zeng, J. et al. “MiR-21 Is Overexpressed in Response to High Glucose and Protects Endothelial Cells from Apoptosis.” Exp Clin Endocrinol Diabetes Experimental and Clinical Endocrinology & Diabetes 121.07 (2013): 425–430. Web.

214. Zhang, X. et al. “High-Throughput Microarray Detection of Olfactory Receptor Gene Expression in the Mouse.” Proceedings of the National Academy of Sciences 101.39 (2004): 14168–14173. Web.

215. Zhao, J. et al. “Short Leukocyte Telomere Length Predicts Risk Of Diabetes in American Indians: the Strong Heart Family Study.” Diabetes 63.1 (2013): 354–362. Web.

216. Zhao, Shanrong et al. “Comparison Of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells.” PLoS ONE 9.1 (2014)