Table of Contents

Table of Contents i
Abstract 1
Riassunto del lavoro svolto 3
Introduction 5
Epigenetics: non-genetics factors determine cell identity 5
Chromatin: functional DNA packaging 5
DNA methylation 8
Next-generation sequencing 9
SOLiD (Applied Biosystems, Life Technologies) 9
Solexa/Illumina Genome Analyzer 10
Ion Proton 10
Building skeletal muscle: adult myogenesis 10
Transcriptional regulation of myogenesis 10
Epigenetic regulation of muscle differentiation 11
Aim of the thesis 13
Materials and Methods 15
Composition of the buffers 15
Antibodies used 16
Cells 16
Mouse splenocytes 16
C2C12 cell culture 16
CHQ5B primary cell culture 17
Cell counting with Neubauer chamber or Hemocytometer 18
DNA extraction 18
Phenol-chloroform nucleic acid extraction 18
Isopropanol/salt or ethanol/salt nucleic acid precipitation 19
Lithium Chloride RNA precipitation 19
DNA and chromatin fragmentation 19
Agencourt® AMPure® XP beads DNA purification 21
Nucleic acid quantification and quality evaluation 22
NanoDrop™ spectrophotometer 22
Qubit® 2.0 fluorometer 22
Agarose gel electrophoresis 23
XChIP-seq homemade protocol 24
Set-up of chromatin preparation on mouse splenocytes 24
Set-up of chromatin preparation on C2C12 24
Home-made protocol for ChIP-seq on CHQ5B 25
Diagenode True MicroChIP kit 30
Native ChIP on CHQ5B 35
Ion Proton low input protocol 37
DNA methylation 39
Sodium Bisulfite DNA conversion 39
A preliminary protocol: MeSS v1 40
The definitive protocol: MeSS v2 41
SureSelect Methyl-Seq Target enrichment 43
Protocol 44
Ion Proton BS-seq 48
Bioinformatics analysis 50

Results and Discussion of the preliminary protocols 53
Chromatin Immunoprecipitation (ChIP) 53
Crosslinked ChIP (X-ChIP) 54
Native ChIP (nChIP) 61
DNA methylation 63
Methylome SOLiD™ sequencing preliminary protocol (MeSS v1) 64
Methylome Proton Sequencing 65
Results and discussion 67
Is DNA methylation involved in myogenesis? 67
MeSS v2 for whole-methylome: an expensive gold-standard! 67
SureSelect Methyl-Seq Target enrichment 69
ChIP-seq 73
Read alignment and mapping 74
Conclusions 77
References 79

Abstract

Epigenetics is subjected to a pressing attention from the scientific community, because of its potential to explain the mechanisms of gene activation or repression. In this thesis I present a discovery-driven project aimed to the investigation of the epigenetic role in human myogenesis (and in particular the differentiation of myoblasts in myotubes).
Studying epigenetics still presents significant hurdles, both experimental and computational. Therefore my first task was the establishment of robust protocols for investigating the role of epigenetics players during skeletal muscle differentiation. In particular, I focused on setting up the tools for studying DNA methylation and protein-DNA interactions, through bisulfite sequencing and chromatin immunoprecipitation. In this thesis, together with the description of the protocols that I developed, I report the first results that were obtained on myogenic cells.

DNA methylation was investigated with bisulfite treatment of the DNA coupled with next generation sequencing. A novel method for studying the whole methylome was conceived and applied on one myoblast sample using SOLiD 5500xl platform giving reliable results.
However, since the cost of methylome sequencing is still very high, instead of producing data for the whole methylome, I decided to focus on selected regions that could be relevant in methylation studies. For this reason I used the SureSelect MethylSeq Target Enrichment kit (Agilent) that effectively captures more than 2,700,000 CpG sites in the human genome.

We identified less than 600 differentially methylated sites (DMS) in myoblasts compared to myotubes, however we observed that the activation of muscle specific genes seem to be poorly correlated with DNA methylation changes. Therefore, we argue that DNA methylation does not show a major role in the control of muscle specific genes. Interestingly, further analysis revealed that a high percentage of differentially methylated regions (DMR) localizes near novel non-coding RNA genes. On the one hand, this observation suggests a role for novel regulatory RNAs in the epigenetics of muscle differentiation; on the other hand, DNA methylation might have a role in the regulation of these RNAs.

Together with DNA methylation, chromatin compaction is a major epigenetics player. In order to describe the epigenetic landscape of muscle differentiation, I optimized the ChIP-seq approach to define the localization of specific histone modifications on DNA. After setting-up the protocol for ChIP-seq, H3K4me3, H3K27me3 and H3K9ac histone modifications were mapped on myoblast DNA. As I started integrating gene expression and ChIP-seq data, I verified that a great concordance exists between gene expression and ChIP-seq results. In particular, euchromatin-associated histone modifications are found in transcription start sites of active genes, and heterochromatic signature spans promoters and bodies of inactive genes. Further investigation show that genes for non-coding RNAs have an euchromatic signature. This observation, together with the

findings of DMRs in novel non-coding RNA genes, endorses the hypothesis of a role for novel regulatory RNAs in myogenic differentiation.

In conclusion, the integration of BS-seq, ChIP-seq and RNA-seq data are opening interesting scenarios concerning the involvement of regulatory RNAs, while recent reports are suggesting to extend our investigation even to DNA hydroxymethylation in the epigenetics of muscle differentiation.

Riassunto del lavoro svolto

Negli ultimi anni l'epigenetica ha raccolto un sempre crescente interesse da parte della comunità scientifica, grazie al suo potenziale di spiegare i meccanismi di attivazione e repressione dell'espressione genica. In questa tesi si presentano i risultati di un progetto di analisi del ruolo dell'epigenetica nella miogenesi umana, mediante approcci genomici. Gli studi di epigenetica presentano tuttora ostacoli significativi, sia dal punto di vista sperimentale che di analisi computazionale del dato prodotto. Il mio primo obiettivo è stato quindi la messa a punto di un protocolli robusti per l'analisi del ruolo dei meccanismi epigenetici durante il differenziamento del muscolo scheletrico, in particolare la metilazione del DNA e le interazioni DNA-proteine (mediante il sequenziamento di DNA trattato con bisulfito e immunoprecipitazione della cromatina). In questa tesi, oltre alla descrizione dei protocolli sviluppati, sono riportati i primi risultati ottenuti applicando i suddetti protocolli alle cellule miogeniche.

La metilazione è stata analizzata mediante trattamento con bisulfito del DNA, sequenziato successivamente con tecniche di nuova generazione (NGS). A tal riguardo è stato messo a punto un nuovo metodo per lo studio del metiloma intero, che è stato applicato ad un campione di mioblasti e successivamente sequenziato con la piattaforma SOLiD 5500xl.
Questo approccio richiede tuttavia una quantità massiva di sequenze, che ad oggi risultano ancora eccessivamente costose. È stato quindi affiancato allo studio del metiloma il sequenziamento delle regioni più comunemente analizzate in studi di metilazione (ovvero regioni promotoriali ed isole CpG) con un kit di arricchimento di regioni target che cattura selettivamente più di 2.700.000 siti CpG nel genoma umano. Con questo approccio sono stati identificati meno di 600 siti differenzialmente metilati (DMS) nei mioblasti confrontati con i miotubi. Questo studio ha permesso di osservare che l'attivazione di geni muscolari sembra poco correlata con cambiamenti nella metilazione del DNA, permettendo di ipotizzare che la metilazione del DNA non abbia un ruolo centrale nel controllo dell'attivazione dei geni muscolo-specifici.
L'analisi di questi dati ha inoltre permesso di rilevare che una significativa frazione di regioni differenzialmente metilate (DMR) localizza in prossimità di geni codificanti non-coding RNA. Da un lato quest'osservazione suggerisce che gli RNA regolatori potrebber avere un ruolo nell'epigenetica nel differenziamento muscolare, e inoltre che la metilazione del DNA potrebbe avere un ruolo nella regolazione di questi RNA.

Un altro aspetto importante della regolazione epigenetica, oltre alla metilazione del DNA, è lo stato di condensazione della cromatina. Per vagliarne il contributo nell'ambito del differenziamento muscolare, è stato ottimizzato un approccio di immunoprecipitazione della cromatina seguito dal sequenziamento (ChIP-Seq) per definire la localizzazione nel DNA di specifiche modificazioni istoniche: H3K4me3, H3K27me3 and H3K9ac.
I risultati di questi esperimenti di ChIP-Seq sono stati confrontati con quelli di analisi del profilo di espressione (RNA-Seq), permettendo di verificare un ampio margine di

sovrapposizione fra il livello di espressione ed i risultati di ChIP-Seq. In particolare le modificazioni istoniche associate all'eucromatina localizzano nei pressi del sito di inizio di trascrizione di geni espressi, mentre i *marker* di eterocromatina fiancheggiano i promotori dei geni non attivi. Questa osservazione, assieme all'osservazione di DMR in nuovi RNA non codificanti, supporta l'ipotesi di un ruolo per nuovi circuiti regolatori di RNA nella differenziazione miogenica.

Per concludere, l'integrazione dei dati di BS-Seq, ChIP-Seq e RNA-Seq apre interessanti scenari riguardo il ruolo di RNA regolatori. La letteratura più recente suggerisce inoltre di estendere le nostre analisi anche all'idrossimetilazione del DNA per comprendere appieno il ruolo delle modificazioni epigenetiche nel differenziamento muscolare.

Introduction

During my PhD experience I focused on both setting-up the tools to study epigenetics and applying these techniques on human myogenic cells. In order to investigate the epigenetic changes during adult myogenesis, which is the process of building functional muscle from satellite cells, I developed and used ChIP-seq and BS-seq techniques. Coupling chromatin immunoprecipitation and bisulfite DNA treatment with next generation sequencing permits to investigate at high resolution the genome-wide distribution of epigenetic players, both histone modifications and DNA methylation.

Epigenetics: non-genetics factors determine cell identity

There are more than 200 cell types in human body, each containing the same DNA sequence, but showing different phenotypes and functions. Cell identity is determined by the quality and quantity of genes that it expresses and the mechanisms by which cells establish, maintain and transmit gene expression profiles to progeny are referred to as epigenetic mechanisms and memory. Epigenetics is the study of heritable changes in gene expression that cannot be explained by changes in DNA sequence, and unravelling the basis of epigenetics is a current challenge in modern biology (Hemberger M *et al.*, 2009).

Transcription factors play a crucial role in the definition of cell identity by determining the lineage-specific gene expression profile of a cell, but they are not the only players. In fact, their accessibility to DNA is influenced by chromatin compaction that is in turn determined by epigenetic mechanisms such as DNA methylation, chromatin modifiers enzymes, noncoding RNAs, histone variants and histone modifications (Murrell A *et al.*, 2005; Guttman M et al., 2009; Henikoff S, 2008; Klose RJ and Bird AP, 2006; Rando OJ and Chang HY, 2009).

C.H. Waddington defined the epigenetic landscape as the “interactions of genes with their environment, which bring phenotype into being”. Environmental factors can alter the epigenetic state of a cell, allowing the cell to respond and adapt to the environment with a rapid and stable modification in gene expression (Bjornsson HT *et al*., 2008), and epigenetics alterations last through multiple cell generations.

There are two major known epigenetic mechanisms: cytosine methylation in cytosine-guanine dinucleotide (CpG) in the DNA, and histone modifications, especially acetylation and methylation.

CpG DNA methylation, when happening in CpG-rich promoter regions (CpG islands) represses gene transcription (Ribel-Madsen R *et al*., 2012), while histone modifications may result in both activation and silencing of genes (Cheung V *et al*., 2000).

Chromatin: functional DNA packaging

In eukaryote cells, DNA is not naked in the nucleus, but it is wrapped in nucleosomes and hierarchically packed in highly ordered structures that constitute the chromatin. The nucleosome is the primary chromatin compaction structure and is composed of 147bp of DNA wrapped around an octamer of histones (two dimers H2A-H2B, and a tetramer

H3-H4) that are basic proteins with a globular core and a protruding N-terminal histone tail. Nucleosomes are bonded together by the linker histone H1.
Chromatin compaction influences transcription factors accessibility to the DNA, thus affecting gene expression through epigenetic phenomena such as DNA methylation, covalent histone modifications, incorporation of histone variants, nucleosome remodelling, and noncoding RNAs (Sharma S *et al.*, 2010). The accessibility of transcription machinery to promoters is influenced by nucleosome occupancy (Li B *et al.*, 2007) and transcription is allowed only when DNA-histone interactions are weak (Baar K, 2010).

Histone modifications: actors and promoters for chromatin structure

Post-translational modifications (PTMs) of amino acids in the N-terminal tail of histones can modify chromatin condensation, thus influencing the accessibility of DNA to the transcriptional machinery. Histone PTMs can directly affect chromatin structure due to their chemical properties or can recruit co-factors that in turn modify chromatin compaction and accessibility (Kouzarides T, 2007). Spatial and combinatorial relations exist between different histone modifications and, by examining the distribution of many types of histone modifications along the genome, an association emerges between specific genomic features and distinct types of chromatin signature (Barski A *et al.*, 2007; Mikkelsen TS *et al.*, 2007).
Histone PTMs can occur on all histones and consist mainly of acetylation and methylation, but even phosphorylation, ubiquitination, and sumoylation.

Histone acetylation is associated with euchromatin

Enzymatic acetylation of lysine residues is the most common PTM of histones and has the effect of neutralizing the positive charge of the lysine amino group. This decreased affinity between the histone tail and the negatively charged DNA results in the opening of the chromatin structure that becomes transcriptionally active (euchromatin). Deacetylation reverses this process, by closing chromatin structure and blocking transcription (heterochromatin). Acetyl-lysine residues may also regulate protein-protein interactions by acting as docking sites for other co-factors that further activate transcription (Alamdari N *et al.*, 2013).

Histone methylation as docking site for chromatin modifiers

The enzymatic methylation of lysine or arginine residues in the N-terminal histone tails does not modify the amino acids charge and so it is unlikely to alter chromatin compaction. This epigenetic modification associates either with euchromatic or heterochromatic regions, depending either on the recruited proteins or on the position of the histone modification within a genetic region.
Lysine and arginine can respectively accept up to three and two methyl-groups, and effector proteins can recognize mono-, di-, and trimethylated epitopes with different affinities so that the degrees of histone methylation influence gene expression. For example, trimethylated H3K9 is exclusively associated with silenced genes whereas monomethylated H3K9 has been associated with both expressed and repressed genes.
On the other hand, the position of the histone methylation in a genomic region is important in determining its effect on gene transcription: for example, for the same

gene, H3K9 methylation correlates with transcription if it is in the coding region of a gene, but with silencing if it is in the promoter.

ChIP-seq

DNA-binding proteins, both transcription factors and histone proteins, play crucial roles in cellular processes, such as transcription, splicing, replication and DNA repair. Their interactions with DNA can be studied using a technique called ChIP-seq (Furey TS, 2012).

Chromatin immunoprecipitation coupled with NGS allow mapping multiple modifications in an unbiased fashion and therefore it is the most widely used technique to identify genome-wide binding patterns of proteins of interest (transcription factors, histone variants, histone modifications, RNA Pol II) (Valouev A *et al.*, 2008; Kharchenko PV *et al*., 2008).

ChIP is the most direct way to identify the binding sites of a single DNA-binding protein or the location of modified histones. In principle, ChIP is based on the selective isolation of a chromatin fraction containing a specific antigen. Typically, antibodies that recognize a protein of interest are used to "capture" the protein coupled with its associated genomic DNA. Determining the sequence of the co-immunoprecipitated DNA permits to identify the positions where the protein was bound in the genome. After isolating the DNA from the immunoprecipitated chromatin fraction, it is directly sequenced and the reads are mapped to a reference genome. The mapping of reads mapping on a specific genomic location is directly proportional with the probability of the DNA-protein association in that region.

Table 1| Representative sudies in which ChIP-seq was used to analyse the interaction between proteins and nucleic acids.

Transcription factor mapping	Johnson DS *et al*.	Science 2007
Histone mapping	Barski A *et al*.	Cell 2007
Enhancer identification	Visel A *et al*.	Nature 2009
Study DNA methylation	Bock C *et al*.	Nat. Biotechnol. 2010
Study protein-RNA interactions	Sanford JR *et al*.	Genome Res. 2009

ChIP-seq can be used to map transcription factors DNA-binding sites or for analysing histone modifications; some examples of the first studies unravelling the potentiality of this technique are included in Table 1.

ChIP is the ideal method for learning how a specific protein interacts with the genome in its "natural" state and allows the analysis of chromatin dynamic changes during physiological perturbation or cellular differentiation.

The work of the ENCODE consortium (The ENCODE project consortium, 2012) in ChIP-seq experiments, gave to the scientific community both an enormous amount of epigenetic data and a set of guidelines for successfully performing further ChIP-seq experiments. The success of this approach largely depends on the optimization of good protocols and validation of highly specific antibodies, because antibody quality is sometimes variable and unpredictable (Egelhofer TA *et al.*, 2011), also because multiple histone modifications can alter the efficacy of certain antibodies (Fuchs SM *et al*., 2011).

There are two general procedures for ChIP experiments: native ChIP (nChIP) and cross-linking ChIP (XChIP). The most appropriate procedure depends on experimental aims and starting material.
In crosslinking ChIP, formaldehyde is typically used to cross-link proteins to the DNA, so a broad range of chromatin-associated factors can be analysed. After formaldehyde treatment, chromatin is randomly fragmented by sonication, and the protein of interest is immunoprecipitated using a specific antibody. Crosslinks are then reversed and the associated DNA purified and sequenced.
In this ChIP experiment, the substrate is native chromatin (proteins are not cross-linked to the DNA) and fragmentation is achieved by micrococcal nuclease digestion. N-ChIP applications are suitable for studying proteins that are very tightly associated with chromatin, typically histones and their modifications (O'Neill and Turner, 2003).

DNA methylation

DNA methylation consists in the enzymatic addition of a methyl group to the 5N position of the cytosine, mainly in CpG dinucleotide of CpG islands, but recent evidence showed that one quarter of methylated cytosines occurs in non-CpG sites in human embryonic stem cells, particularly in gene bodies (Lister R *et al*., 2009; Bernstein BE *et al*., 2007; Bird AP, 1986).
Cytosine methylation of the DNA was the first epigenetic mark discovered in human and has a role in regulating, especially in suppressing, gene expression, and in silencing transposons and repetitive sequences (Beck S *et al*., 2008, Gibney ER *et al*., 2010). About 70-80% of human CpG dinucleotides are methylated in the human genome, especially in the repetitive regions, but 88% of active promoters are associated with unmethylated CpGs.
In mammals DNA methylation is involved in physiological processes such as embryogenesis, genomic imprinting and X inactivation (Bestor TH, 2000; Bird A, 2002; Lippman Z *et al*., 2004; Rhee I *et al*., 2002), and abnormal DNA methylation affects human health leading to, for example, chromosomal instability, loss of imprinting and tumorigenesis (Feinberg AP, 2007).
The relationship between gene-promoter hypermethylation and transcriptional inactivation has been described, but the mechanisms through which DNA methylation regulates genome function are still unclear (Murrel A *et al*., 2005; Jaenisch R *et al*., 2003).

BS-seq to decipher DNA methylation

Among the methods to study epigenetics, reviewed in Laird PW, (Laird PW, 2010), BS-seq was the one chosen for this project. Investigating the role of DNA methylation in biological process at genome-wide level requires accurate and sensitive methods and to date there are a few techniques, different in cost, resolution and throughput. The gold standard to study DNA methylation at single-base resolution is bisulfite sequencing, as sodium bisulfite DNA treatment converts unmethylated cytosines to uracils and leaves methylated Cs unchanged (Frommer M *et al*., 1992). It allows genome-wide single-base resolution by mapping the bisulfite-converted sequenced reads to an in silico converted reference sequence (Cokus SJ *et al*., 2008; Meissner A *et al*., 2008).

There are both enrichment-based and whole-genome approaches for BS-seq. To study the entire methylome, a whole genome resequencing is performed after bisulfite treatment and, except for repetitive regions; this technique can virtually determine the methylation state of the all cytosines in the genome. The first single-base resolution whole methylome was published in 2009 and was performed on the genome of Arabidopsis thaliana (Lister R *et al*., 2008; Lister R *et al*., 2009).
Sequencing capacities and costs for whole methylome analyses are still high, so BS-seq studies are often limited to parts of the genome. Among these methods, reduced representation bisulfite sequencing (RRBS) is the most widely used (Gu H. *et al*., 2010) and interrogates approximately 10% of the genome.
I used another enrichment system in which regions of interest are captured by hybridization (SureSelect), bisulfite treated and sequenced. I report in this thesis the successful use of this system, which captures 87Mb of the human genome, mainly gene promoters, CpG islands and their flanking regions (shelves).
The main challenge in BS-seq is the bioinformatics analysis, due to the reduced complexity of the bisulfite converted reads, and the resulting difficulties in correctly mapping the reads on the in silico bisulfite-converted reference genome.

Next-generation sequencing

Thanks to next generation sequencing technologies we can now investigate the relationships that exist between gene expression and epigenetic players, such as histone modifications and DNA methylation, during cell life and tissue differentiation.
Sequencing technology, efficiency and cost have been improving since 1977, when Maxam and Gilbert, and Sanger first described DNA sequencing (Maxam AM and Gilbert W, 1977; Sanger F *et al*., 1977). Advances in automatic capillary sequencing technology permitted in 1990 to launch the “human genome project” aimed to sequence the entire human genome and whose completeness was announced in 2001(Lander ES *et al*., 2001; The International Human Genome Sequencing Consortium, 2004; Venter JC *et al*., 2001).
NGS was the answer to the ever-growing demand for sequencing better accuracy and higher throughput, thus starting a revolution in genomics analysis. Current technologies allow the resequencing of the human genome in few days and offer the possibility of systematic investigations.
The most used NGS technologies today available are characterized by physical immobilization of the short DNA fragments to be sequenced and their massively parallel sequencing (Metzker ML, 2010).

SOLiD (Applied Biosystems, Life Technologies)

The SOLiD sequencing technology is based on the immobilization of DNA fragments on small beads followed by their amplification by emulsion PCR. To determine the sequence of the clonal DNA fragment on each bead, SOLiD uses multiple cycles of hybridization, ligation and cleavage.
The sequencing reaction starts when a universal anchoring primer ligates to a sequencing primer that is a fluorescently labelled 8-mer sequence. After the ligation of the 8-mer the emitted fluorescence is registered, then the three last nucleotides of the 8-mer are cleaved and another cycle of ligation starts (Shendure J *et al*., 2005). The

main advantage of the SOLiD technology is that each base is interrogated twice; so SOLiD sequencing is very accurate (>99.94%). This re-reading of sequence minimizes base-calling errors and makes the SOLiD well suited to high accuracy sampling applications such as genome resequencing, polymorphism analysis and BS-seq.

Solexa/Illumina Genome Analyzer

In the Solexa/Illumina system, clonal DNA clusters are generated by bridge amplification onto the glass surface of the instrument flow cell. It results in a dense array of non-overlapping fragment sequencing colonies. Each fragment colony is then sequenced one base at a time by reversible termination with the cyclical addition of fluorescently labelled nucleotides (Bentley DR *et al*., 2008).

Ion Proton

Ion Torrent sequencing technology is based on the release of an hydrogen ion whenever a nucleotide is added to a growing DNA strand. DNA fragments are immobilized in tiny wells with a semiconductor chip; the four nucleotides floods one by one over the wells and the semiconductor registers a pH variation when a nucleotide is incorporated in the sequence. The major disadvantage of this technology is the susceptibility to the homopolymeric stretch problem, as it was for 454 technology. The main advantages are low running cost and sequencing speed.

Building skeletal muscle: adult myogenesis

Myogenesis is the process of building functional muscle from precursor cells: in adult individuals it occurs when satellite cells, a population of myogenic precursors, provide regeneration and muscle growth (Bentzinger CF *et al*., 2012).

Transcriptional regulation of myogenesis

Pax3 and Pax7 transcription factors have a predominant role in the embryonic myogenesis, whereas in postnatal life satellite cells respond to environmental stimuli supplied by adhesion molecules, growth factors or cytokines. Extracellular stimuli activate muscle-specific gene transcription through signalling cascades, mainly p38 MAPK pathway.

Quiescent satellite cells express Pax7 and, upon activation and asymmetric division, start expressing Myf5. When MyoD expression is upregulated in Pax7+ Myf5+ committed satellite cells, they enter a proliferative state and become myoblasts (Mb). The differentiation can proceed through myocyte till myotube when MyoD associates to E-proteins forming a heterodimer that activates muscle-specific genes. The proliferative state of Mb population is maintained thanks to high level of a protein called Inhibitor of differentiation (Id), which associates to MyoD and to E-proteins, thus giving origin to inactive Id/MyoD and Id/E-protein heterodimers. When the terminal differentiation signal reaches the proliferating Mb, MyoD can associate with E-proteins and activate the other Muscle regulatory factors (MRFs): early during differentiation Mb express MyoG and, later, MRF4 (Berkes CS and Tapscott SJ, 2005). p38 MAPK pathway is essential in the onset of this terminal differentiation by mediating (1) phosphorylation of E47, thus inducing the MyoD/E47 heterodimer formation, (2) phosphorylation of Muscle enhancer factor 2 (MEF2D) and inducing its transcriptional activity (McKinsey TA *et al*., 2002).

In late differentiation, when myotubes are forming, JNK pathway is inhibited; cells exit cell cycle and express skeletal muscle structural genes (Brand-Saberi B and Bodo C, 1999).

Epigenetic regulation of muscle differentiation

Mechanisms other than the classic transcription factors control the myogenesis gene expression program: epigenetic regulation plays a crucial role in the maintenance of quiescence and proliferation states in muscle satellite cells, preventing their differentiation.

DNA methylation appears to be a major repressive mechanism of muscle satellite cell differentiation. The demethylation of MyoD promoter and distal enhancer, and of MyoG appear to be necessary for starting the differentiation program, but the precise timing and mechanism that regulate DNA methylation in adult myogenesis are still unclear. In particular, it is debated if methylation/demethylation events cause or follow gene-expression changes (Barreiro E and Sznaider JI, 2013)

The in vitro differentiation of Mb to Mt is an excellent model for adult myogenesis. Tsumagari K *et al.* (Tsumagari K *et* al., 2013) studied DNA methylation of Mb, Mt and adult skeletal muscle in order to search for changes in DNA methylation correlated to myogenesis. They found that differences between Mb and Mt DNA methylation are less than between Mb or Mt and skeletal muscle. In particular, they hypothesize that DNA demethylation during muscle differentiation is more likely a consequence of gene expression than the cause.

In quiescent satellite cells and in proliferating myoblasts, PcG proteins maintain muscle differentiation specific genes in heterochromatic state with hypoacetylated H3K27me3 and H3K9me2 histone modifications. H3K27 trimethylation is catalysed by Ying Yang protein 1 (YY1), which recruits the Polycomb Repressive Complex to inactive muscle gene promoters. Moreover, MyoD and MEF2 interact with histone deacetylases (HDACs) in proliferating Mb.

When myoblasts enter terminal differentiation program, HDACs are removed from muscle specific gene promoters, thus allowing histone acetylases (HATs), p300 and PCAF to interact with transcription factors and start gene expression. Moreover, MRF2 recruits TrGX histone methyltransferase that trimethylate H3K4 histone, H3K9me2 and H3K27me3 are removed (Perdiguero E *et al*., 2009), and muscle specific genes can be transcribed.

Aim of the thesis

Epigenetics is essential for gene regulation during cell life and organism development: the network of chromatin-associated proteins and DNA methylation is the basis of chromatin dynamics and leads to gene activation or repression. The aim of my PhD project was to create the bases to study the role of epigenetics in human muscle differentiation, using primary human myoblasts and myotubes.

When I began my PhD, epigenetics was still a relatively new field, however it had already shown its potential in unravelling the secrets of gene expression. Nevertheless, alongside the undisputed worldwide interest on this topic, there was a general lack of established procedures. Therefore my initial task was to set-up suitable methods to study epigenetics. In particular I wanted to create robust and reliable protocols for ChIP-seq and BS-seq that would allow the investigation of the role of epigenetics in myogenesis.

As a further aim I wanted to validate and apply these new procedures to investigate the epigenetics involvement in skeletal muscle differentiation and plasticity. Moreover, a final goal of my PhD thesis was the integration of the epigenetics results with the findings obtained from other lines of research carried out in the laboratory, concerning mRNA and miRNA expression in the same experimental model of myogenic cells.

Materials and Methods

Composition of the buffers

General solutions

1X PBS: 137mM NaCl, 2.7mM KCl, 10mM $Na_2HPO_4*2H_2O$, 2mM KH_2PO_4, pH 7.4

1X TAE: 40mM Tris, 20mM acetic acid, 1mM EDTA

0.5X TBE: 45mM Tris-borate, 1mM EDTA
HBSS: 1.26mM $CaCl_2$, 0.49mM $MgCl_2*6H_2O$, 0.41mM $MgSO_4*7H_2O$, 5.33mM KCl, 0.44mM KH_2PO_4, 138mM NaCl, 0.34mM $Na_2HPO_4*7H_2O$, 5.56mM D-glucose.

ChIP buffers

XChIP

XChIP Buffer A: 11% HCOH, 0.1M NaCl, 1mM EDTA, 0.5mM EGTA pH8.0, 20mM HEPES pH 7.9

XChIP Buffer B: 0.25% triton-x100, 10mM EDTA, 0.5mM EGTA pH8.0, 20mM HEPES pH7.9

XChIP Buffer C: 0.15M NaCl, 1mM EDTA, 0.5mM EGTA pH8.0, 20mM HEPES pH7.9

XChIP Sonication Buffer: 0.3% SDS, 1% triton-x100, 150mM NaCl, 1mM EDTA, 0.5mM EGTA pH8.0, 20mM HEPES pH7.9

XChIP dilution buffer: 1% Triton-X 100, 150 mM NaCl, 1 mM EDTA, 20 mM HEPES pH7.9
XChIP Wash Buffer A: 0.1% SDS, 0.1% NaDOC, 1% Triton-X 100, 150mM NaCl, 1mM EDTA, 0.5mM EGTA pH 8.0, 20mM HEPES pH7.9

XChIP Wash Buffer B: 0.1% SDS, 0.1% NaDOC, 1% Triton-X 100, 500mM NaCl, 1mM EDTA, 0.5mM EGTA pH 8.0, 20mM HEPES pH7.9

XChIP Wash Buffer C: 0.25M LiCl, 0.5% NaDOC, 0.5% NP-40, 1mM EDTA, 0.5mM EGTA pH 8.0, 20mM HEPES pH7.9

XChIP Wash Buffer D: 1mM EDTA, 0.5mM EGTA pH 8.0, 20mM HEPES pH7.9

XChIP Elution Buffer: 1%SDS, 0.1M NaHCO3

nChIP

2X base buffer: 120mM KCl, 30mM NaCl, 10mM $MgCl_2$, 0.2mM EDTA, 30mM Tris/Cl pH7.5

nChIP Buffer 1: 0.3M sucrose, 5mM Na Butyrate, 0.1mM PMSF, 0.5mM DTT, 60mM KCl, 15mM NaCl, 5mM $MgCl_2$, 0.1mM EDTA, 15mM Tris/Cl pH7.5

nChIP Buffer 2: 0.3M sucrose, 5mM Na Butyrate, 0.1mM PMSF, 0.5mM DTT, 60mM KCl, 15mM NaCl, 5mM $MgCl_2$, 0.1mM EDTA, 15mM Tris/Cl pH7.5 + 0.8% V/V NP-40

nChIP Buffer 3: 1.2M sucrose, 5mM Na Butyrate, 0.1mM PMSF, 0.5mM DTT, 60mM KCl, 15mM NaCl, 5mM $MgCl_2$, 0.1mM EDTA, 15mM Tris/Cl pH7.5

MNase digestion buffer: 0.3M sucrose, 5mM Na Butyrate, 0.2mM PMSF, 4mM $MgCl_2$, 1mM $CaCl_2$, 50mM Tris/Cl pH7.5
nChIP Incubation buffer: 150mM NaCl, 20mM Na Butyrate, 5mM EDTA, 0.2mM PMSF, 20mM Tris/Cl pH7.4

nChIP Washing buffer 1: 5mM Na Butyrate, 10mM EDTA, 75mM NaCl, 50mM Tris/Cl pH7.5

nChIP Washing buffer 2: 5mM Na Butyrate, 10mM EDTA, 125mM NaCl, 50mM Tris/Cl pH7.5

nChIP Washing buffer 3: 5mM Na Butyrate, 10mM EDTA, 175mM NaCl, 50mM Tris/Cl pH7.5

nChIP Elution buffer: 1% SDS, 50mM NaCl, 5mM EDTA, 20mM Na Butyrate, 0.1mM PMSF, 20mM Tris/Cl pH7.5

Antibodies used

Antibody	Source	Company
Anti-TBP (cat. no sc-273 X)	Rabbit	Santa Cruz Biotechnology, Dallas, TX
Anti-H3K9me3 (cat. no C15410069)	Rabbit	Diagenode, Liege, Belgium
Anti-H3K9me3 (cat. no C15410003)	Rabbit	Diagenode
Anti-H3K27me3 (cat. no sc.130356 X)	Mouse	Diagenode
Normal mouse IgG (cat. no I5381)	Mouse	Sigma-Aldrich, Saint Louis, MO
Normal rabbit IgG (included in True MicroChIP kit)	Rabbit	Diagenode

Cells

Mouse splenocytes

One of the aims of this project was to set-up a protocol for chromatin immunoprecipitation experiments. Initially in order to perform and assess different ChIP protocols splenocytes were used due to the ease of obtaining high number of cells. The number of splenocytes obtained from a single mouse spleen is ~100million.

Mouse dissection and spleen isolation

Fresh splenocytes were isolated from the spleen of a healthy mouse (kindly provided by Dr. P. Braghetta, Dept. Biomedical Science, Univ. Padua).

1. Put the mouse on the bench with the abdomen facing up.
2. Grab hold the skin at the urethral opening using a forceps, and cut the skin along the ventral midline till the groin.
3. Remove the peritoneal muscle wall without touching the internal organs.
4. The spleen is a dark-red organ in the left side of the abdomen, close to the pancreas and behind the intestine. After finding the spleen, cut its connections with internal organs and remove it.
5. Prepare a single-cell suspension of splenocytes in 3ml of PBS and count the cells in using a Neubauer counting chamber.
6. Divide the cell suspension in several aliquots containing 20million cells each and centrifuge for 8 min at 200rcf, RT.
7. Aspirate all the supernatant, snap-freeze the aliquots in liquid nitrogen and store at -80°C for up to 1 year.

C2C12 cell culture

The immortalised C2C12 cell line is a subclone (Blau HM *et al*., 1985) of the mouse myoblast cells, C2. These cells are a model for studying skeletal muscle differentiation due to their capacity of in vitro myogenic differentiation.

The cells grow in adhesion on a plastic surface and are maintained at 37°C in a humidified incubator with 6% CO_2. Undifferentiated proliferating Mb grow in DMEM (GIBCO, Life Technologies, Carlsbad, CA) with 10% FBS (GIBCO) and 50µg/ml Gentamycin (GIBCO) (referred to as “growth medium”). Differentiation of Mb to myotubes (Mt) is triggered by low serum conditions. Differentiation medium: DMEM with 2% horse serum (GIBCO) and 50µg/ml Gentamycin.

Cell culture and passage

Proliferating Mb are maintained in 175cm^2 flasks (SARSTEDT, Nümbrecht, Germany) with 25ml of growth medium. When growing in these conditions cells reach confluence at ~3*10^6 cells.

When the Mb in a flask reach 50% of confluence (no contact or very few contacts between cells) they were used for ChIP experiments as Mb, or the cells were seeded into new flasks following the procedure reported here.

1. Remove and discard growth medium and rinse twice with 10ml of Dulbecco's PBS (calcium and magnesium free) at RT.
2. Detach cells by trypsinization: add 2.5ml of 1X Trypsin-EDTA solution (GIBCO) in each flask and wait until cells detach, usually 2-3'.
3. Add 5ml of FBS to the flask and collect cells in a 15ml tube.
4. Centrifuge the cells at 150rcf for 8', discard supernatant and suspend cells in 10ml of growth medium.
5. Count the cells using a Neubauer chamber and plate 3x10^5 cells in new flasks with 25ml of growth medium.

Instead of dividing the cells to maintain the cell culture, the cells can be collected and frozen for future use. Mb dry pellets can be snap-frozen in liquid nitrogen and stored at -80°C for up to 1 year for future applications.

C2C12 myogenic differentiation

When C2C12 cells in a flask are 70-80% confluent, the substitution of the growth medium with the low-serum differentiation medium induces differentiation.

1. Remove and discard the growth medium and rinse the cells once with Dulbecco's PBS (without calcium and magnesium) at RT.
2. Add 25ml of differentiation medium and differentiate for 9 days, replacing medium every 3 days. Myotubes at this stage can be used immediately for ChIP experiments or can be detached and stored at -80°C for future experiments.
3. Detach cells by trypsinization: add 2.5ml of 1X Trypsin-EDTA solution to each flask and wait until cells detach.
4. Add 5ml of FBS to the flask and collect cells in a 15ml tube.
5. Centrifuge the cells at 150rcf for 8', snap-freeze the dry pellet in liquid nitrogen and store at -80°C for up to 1 year for future applications.

CHQ5B primary cell culture

CHQ5B is a primary cell line of human Mb initially provided by Dr. V. Mouly (URA, CNRS, Paris, France). Accordingly to French legislation and ethical rules CHQ5B human Mb were isolated from the quadriceps of a 5-day infant female without any sign of neuromuscular disorders. CHQ5B cells reach proliferative senescence after approximately 50 divisions. Mb can differentiate in vitro by changing the medium composition (**Figure 1**); the cells grow in adhesion on plastic surfaces and are maintained at 37°C in a humidified incubator with 6% CO_2.

Undifferentiated proliferating Mb grow in a medium composed of DMEM with 20% FBS and 50µg/ml Gentamycin, and to induce differentiation growth medium is switched with differentiation medium: DMEM with 2% horse serum and 50µg/ml Gentamycin.

Both cell culture and cell differentiation conditions are analogous to that described for C2C12 cells, except for growth medium composition.

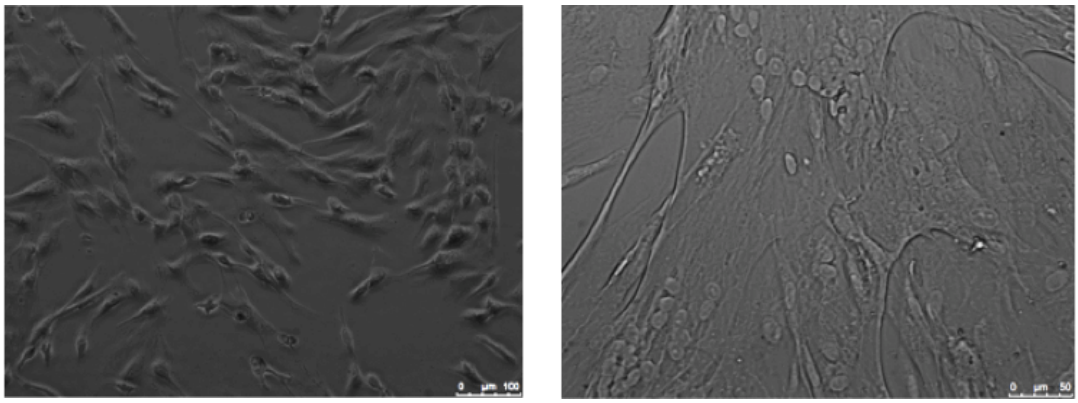

Figure 1| CHQ5B is a human primary cell culture of myoblasts (on the left panel) that can differentiate *in vitro* into myotubes (right panel), by serum limitation in the culture medium.

Cell counting with Neubauer chamber or Hemocytometer

Cell counting is required to monitor cell growth and ensure experimental reproducibility in the applications that require a precise initial number of cells, such as ChIP experiments.

The hemocytometer is a thick glass slide containing 2 counting chambers with laser-etched grids of perpendicular lines. Volumes of both entire grid and single quadrants are known and this allows the estimation of cell concentration in the initial cell suspension.

The number of cells in the chamber is determined by direct counting under an inverted microscope and this number is used to calculate the concentration of the initial cell suspension. After counting the cells in the 4 corners, calculate the average number of cell per quadrant by dividing the sum four times; multiply by 10,000 and then multiply by the dilution used to obtain, cell/ml.

DNA extraction

If starting with whole cells:

1. Resuspend 3million cells (as dry-pellet, both fresh or -80°C stored) in 1ml PBS with 2% SDS (Sigma-Aldrich) and 1mM EDTA (Sigma-Aldrich). If starting with more cells, scale accordingly the volume of PBS-SDS-EDTA buffer.
2. Aliquot 200µl of the cell suspension in 1.5ml microcentrifuge tubes.

Phenol-chloroform nucleic acid extraction

1. Add 1 volume of TE saturated phenol pH10.5 (Sigma-Aldrich) mix vigorously by vortexing, and incubate at RT till the formation of two phases;
2. Centrifuge 5’ 16,100rcf at RT, then transfer the upper aqueous phase in a new 1.5ml microcentrifuge tube.
3. Add 1 volume of chloroform (Carlo Erba Reagents, Arese-Milano, Italy) to the aqueous phase and mix by inversion.

4. Centrifuge 5', 16,100rcf at RT and transfer the upper aqueous phase in a new 1.5ml microcentrifuge tube;
5. Add 1 volume of chloroform and repeat the centrifugation step.
6. Transfer the upper phase into a new 1.5ml microcentrifuge tube and proceed with isopropanol/salt or ethanol/salt nucleic acid precipitation.

Isopropanol/salt or ethanol/salt nucleic acid precipitation

1. Add 5M NaCl (Sigma-Aldrich) to a final concentration of 250mM and 1 volume of isopropanol (Carlo Erba Reagents) or, alternatively, 2 volumes of absolute ethanol (Carlo Erba Reagents). The use of isopropanol or ethanol will be specified case by case, but if not specified, 1 volume of isopropanol is used.
2. Incubate 4-16h at -20°C, then centrifuge 30' 16,100rcf 4°C and discard the supernatant.
3. Wash the pellet with 500µl of freshly prepared 70% EtOH.
4. Centrifuge 20' 16,100rcf 4°C and discard supernatant.
5. Repeat the washing and centrifugation steps.
6. Aspirate all the supernatant with a micropipette tip and leave the pellet to air-dry.
7. Resuspend the pellet in a suitable volume of Low TE buffer or ultrapure water (Sigma-Aldrich). Volume and buffer for resuspending the DNA pellet will be specified case by case.

Lithium Chloride RNA precipitation

When extracting DNA from cells RNA contamination is always present and high concentration of lithium chloride allows the selective precipitation of RNA, while DNA remains in solution.

1. Add 1/3 of the initial volume of 8M LiCl (Sigma-Aldrich) in the nucleic-acid containing solution and incubate on ice for 4h.
2. Centrifuge 40' 16,100rcf 4°C.
3. Transfer the DNA-containing supernatant in a new microcentrifuge tube and discard the RNA-containing pellet.
4. Repeat the isopropanol/salt or ethanol/salt DNA precipitation to concentrate the DNA.

The quantification of the extracted DNA is performed both with NanoDrop™1000, Thermo Fisher Scientific, Waltham, MA) and Qubit® 2.0 fluorometer (Invitrogen, Life Technologies, Carlsbad, CA). Quality evaluation is performed on a 0.6% w/v agarose gel electrophoresis in 1X TAE buffer.

DNA and chromatin fragmentation

For both ChIP-seq and BS-seq applications the COVARIS™ S2 Ultrasonicator (Covaris Inc., Woburn, MA) was used to efficiently fragment chromatin and DNA.
Below are reported the chromatin and DNA fragmentation protocols used for this project.

Mouse splenocytes chromatin shearing

Sample volume: 2ml
Sample Buffer: XChIP sonication Buffer
Water level: 12
Water bath temperature: 4-8°C
Tubes: 16×65mm TC13 glass tubes
Mode: frequency sweeping
Duty cycle: 20%
Intensity: 10
Cycles per burst: 1000
Time: 60s
Number of cycles: 10

C2C12 chromatin shearing

Sample volume: 1ml
Sample Buffer: XChIP sonication Buffer
Water level: 12
Water bath temperature: 4-8°C
Tubes: 16×65mm TC13 glass tubes
Mode: frequency sweeping
Duty cycle: 20%
Intensity: 10
Cycles per burst: 1000
Time: 60s
Number of cycles: 8

CHQ5B nuclei chromatin shearing

Sample volume: 1ml
Sample Buffer: XChIP sonication Buffer
Water level: 12
Water bath temperature: 4-8°C
Tubes: 16×65mm TC13 glass tubes
Mode: frequency sweeping
Duty cycle: 20%
Intensity: max
Cycles per burst: 1000
Time: 60s
Number of cycles: 8

CHQ5B chromatin shearing for True MicroChIP kit

Sample volume: 100µl
Sample buffer: 25% Lysis Buffer tL1 + 75% HBSS
Water level: 15
Water bath temperature: 3-5°C
Tubes: AFA™ fiber microtube
Mode: frequency sweeping
Duty cycle: 5%
Intensity: 2
Cycles per burst: 200
Time: 60s
Number of cycles: 10

DNA fragmentation for SureSelect Methyl-Seq Target Enrichment

Sample volume: 130µl
Sample buffer: 1X Low TE
Water level: 15
Water bath temperature: 4-7°C
Tubes: AFA™ fiber microtubes
Mode: frequency sweeping
Duty cycle: 20%
Intensity: 10
Cycles per burst: 1000
Time: 60s
Number of cycles: 10

DNA fragmentation after bisulfite treatment (MeSS v2)

Sample volume: 130µl
Sample buffer: 1X Low TE
Water level: 15
Water bath temperature: 4-6°C
Tubes: AFA™ fiber microtubes
Mode: frequency sweeping
Duty cycle: 10%
Cycles per burst: 200
Intensity: 4
Time: 60”
Number of cycles: 2

DNA fragmentation before bisulfite treatment (MeSS v1 and Ion Proton BS-seq)

Sample volume: 130µl
Sample buffer: 1X Low TE
Water level: 12
Water bath temperature: 6-8°C
Tubes: AFA™ fiber microtubes
Mode: frequency sweeping
Duty cycle: 10%
Cycles per burst: 200
Intensity: 5
Time: 60”
Number of cycles: 3

XChIP-DNA fragmentation for SOLiD fragment library preparation

Sample volume: 130µl
Sample Buffer: 1X Low TE Buffer
Water level: 15
Water bath temperature: 4-6°C
Tubes: AFA™ fiber microtube
Mode: frequency sweeping
Duty cycle: 20%
Intensity: 5
Cycles per burst: 200
Time: 60s
Number of cycles: 6

nChIP-DNA fragmentation for Ion Proton fragment library preparation

Sample volume: 50µl
Sample Buffer: 1X Low TE Buffer
Water level: 12
Water bath temperature: 6-8°C
Tubes: AFA™ fiber microtube
Mode: frequency sweeping
Duty cycle: 10%
Intensity: 2
Cycles per burst: 200
Time: 40s
Number of cycles: 1

Agencourt® AMPure® XP beads DNA purification

Fragmented DNA purification was used either to change buffer composition or to purify libraries from adapters.

1. Add a suitable volume of Agencourt® AMPure® XP (Beckman Coulter, Brea, CA) beads were added to the sample. The ratio between the volume of beads used and the initial volume of the sample will be specified every time.
2. Pipet to mix, pulse-spin and incubate at RT for 5’.
3. Pulse-spin and place the tube in a magnetic rack for 3’, then remove and discard the supernatant without disturbing the beads. Do not remove the tube from the magnet.
4. Without removing the tube from the magnet, wash the beads with 500µl of freshly prepared 70% EtOH. Incubate for 30”, turning the tube in the magnet to move the beads. When the solution is clear, remove and discard the SN without disturbing the pellet.
5. Repeat the washing step.
6. To remove residual ethanol: pulse-spin the tube, place it back on the magnet and carefully remove the remaining supernatant.
7. Without removing the tube from the magnetic rack, leave the beads to air-dry at RT for ≤5’.

8. Remove the tube from the magnetic rack and add a suitable volume of H_2O or LowTE buffer directly on the pellet to dissolve the beads. The volume and buffer for resuspending the DNA pellet will be specified case by case.
9. Pipet 5 times to mix, vortex for 10”
10. Pulse-spin and place the tube back on the magnetic rack for 2’.
11. Transfer the supernatant in a new 1.5ml LoBind (Eppendorf, Hamburg, Germany) tube without disturbing the pellet.

Nucleic acid quantification and quality evaluation

Quality and concentration of DNA were determined using both NanoDrop™ spectrophotometer and agarose gel electrophoresis. The concentration was then more accurately determined using the Qubit® 2.0 fluorometer.

NanoDrop™ spectrophotometer

NanoDrop™ spectrophotometer (NanoDrop™1000, Thermo Fisher Scientific) can measure the absorbance of the nucleic acid solution at 260nm (A260).
An A260 of 1 is equivalent to a concentration of approximately 50µg/ml for double stranded DNA and of 33µg/ml for single stranded DNA. The settings for single stranded DNA are used for the quantification of bisulfite-treated DNA.
The presence of contaminants in the solution can be estimated by observing the A260/280 and the A260/230 ratios. As a general statement we consider a sample clean if these values are above 1.80; lower values indicates the presence of contaminants. All RNA, salts, proteins and solvents alter the quantification of the DNA.

Qubit® 2.0 fluorometer

The Qubit® 2.0 fluorometer (Life Technologies, Carlsbad, CA) permits an accurate, specific and high sensitive quantification of even low concentration of DNA. It uses a fluorimetric technology in which fluorescent probes emit a signal only when bound to their specific target molecule. The presence of proteins, RNA, salts or detergent contaminants in the solution does not interfere with the accuracy of the DNA quantification.
The dsDNA high sensitivity kit used during this project is highly selective for double-stranded DNA and is accurate for initial sample concentrations between 10pg⁄µl and 100ng⁄µl. Highly concentrated samples must be properly diluted.
Samples were prepared for the assay following instructions:

1. Prepare the working solution by diluting the DNA HS fluorescent dye 1:200 in DNA HS buffer. 200µL of working solution are required for each sample.
2. Set-up one thin-well, clear, 0.5ml assay tube (Qubit® assay tube) for each sample and mix reagents following the table here reported:

Working solution	180-199µl
DNA sample	1-20µl
Total	200µl

3. Vortex the tubes for 2-3", then pulse-spin.
4. Incubate in the dark for 2' at RT, then take the read on the Qubit® fluorometer with dsDNA HS settings.
5. Calculate the concentration of the initial sample using the dilution calculator feature of the Qubit® 2.0 fluorometer.
6. Fluorometric assay yielded a quantification of the DNA in the samples and could be compared with the data obtained using NanoDrop™.

Agarose gel electrophoresis

Agarose gel electrophoresis permits the separation of DNA fragments depending on their length and can be used to estimate quantity, quality and dimensions of DNA.
Depending on buffer composition and agarose concentration different information can be obtained. The electrophoretic run is performed at 100V for 40' in a 10cm long gel. The signal can be seen on a transilluminator (a UV light box), due to the fluorescent dye added to the agarose. Ethidium bromide can be added to the agarose before pouring at a final concentration of 0.5ug/ml. However, as an alternative SYBR® Green can be used.

DNA integrity check

To evaluate the integrity and purity of a DNA extraction the agarose gel electrophoresis is performed in 1X TAE buffer. The agarose gel is 0.6% w/v agarose (Sigma-Aldrich) in 1X TAE with 0.5µg/ml of Ethidium Bromide (Sigma-Aldrich).
This method permits to determine the integrity of the DNA and allows its indirect quantification by comparing the intensity of the DNA signal with the intensity of the ladder, whose concentration is known. Agarose gel electrophoresis permits to see the presence of RNA, proteins, salts or detergent contaminations.

Estimation of DNA dimensions

After sonication steps of at the end of a library preparation, a preliminary idea on the dimensions of the DNA is given by gel electrophoresis. Depending on the expected fragment dimensions the electrophoresis can be performed in TAE or in TBE buffer, with the last giving the best resolution for low molecular weight DNA.
To estimate DNA dimensions I used the 1.8% w/v agarose gel in 1X TAE or 0.5X TBE buffer with 0.5µg/ml Ethidium Bromide. The buffer choice is specified every time an electrophoresis is cited in the protocol.
For precise evaluation and visualization of DNA dimensions the tool of choice is the Agilent 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany).

Agilent 2100 Bioanalyzer

Agilent 2100 Bioanalyzer is a microfluidic-based platforms used for sizing, quantification and quality evaluation of nucleic acid preparation.
Readings of DNA after sonication steps or after library preparation were done using the Agilent 2100 Bioanalyzer by BMR genomics using DNA HS kit and chip.
Agilent Bioanalyzer DNA HS kit can accurately analyze samples with a DNA concentration between 10 and 800pg/µl. For this thesis, samples were quantified

using the Qubit® 2.0 fluorometer, diluted in Low TE buffer till the concentration of 500pg/µl and then submitted for Agilent analysis.

XChIP-seq homemade protocol

The set-up of the chromatin preparation protocol was conducted on the basis of a protocol for ChIP-seq kindly provided by Dr. Mokry (Mokry M *et al*., 2010).

Set-up of chromatin preparation on mouse splenocytes

The initial setup of the crosslinking and shearing conditions for ChIP experiments was conducted on mice splenocytes.

1. Defrost one ice one 20million cell-aliquot of mice splenocytes.
2. Resuspend the pellet in 1ml of RT PBS and transfer the cell suspension in a 15ml falcon tube containing 9ml of RT PBS.
3. Add 1ml (1/10 V_i) of a XChIP Buffer A and incubate RT for 20' with occasional mixing by inversion.
4. Add glycine to a final concentration of 0.125M in order to block cross-linking and incubate RT for 5'.
5. Centrifuge 5' 1,200rcf 4°C and discard the supernatant.
6. Resuspend pellet in 10ml of ice-cold XChIP Buffer B + PIC and incubate on ice for 10'. This step allows the nuclei isolation.
7. Centrifuge 5' 1200 rcf 4°C and discard the supernatant.
8. Resuspend the pellet in 10ml of ice-cold XChIP Buffer C + PIC and incubate on ice for 10'.
9. Centrifuge 10' 2500rcf 4°C and aspirate the supernatant.
10. Resuspend the pellet in 2ml (1ml every 10million cells) of ice-cold XChIP Sonication Buffer + PIC.
11. Incubate on ice for 5' and then sonicate using the COVARIS with the program specified below in the Covaris section.
12. Centrifuge 20' 4°C 13200rcf, and transfer the SN in a new microcentrifuge tube (use 75µl to check the fragmentation).
13. Sheared chromatin can be used immediately or can be snap-frozen in liquid nitrogen and stored at -80°C for up to 6 months.

Set-up of chromatin preparation on C2C12

If starting from fresh not frozen C2C2 cells, either Mb or Mt, remove and discard the culture medium and add 10ml of DMEM + 10% FBS directly onto the cell layer.
If starting with a frozen pellet, defrost cell pellet on ice, add 1ml of DMEM + 10% FBS and transfer the cell suspension in a 15ml falcon tube containing 9ml of DMEM + 10% FBS.
Each chromatin preparation is performed on 10million cells.

1. Add 270µl of 37% formaldehyde (Sigma-Aldrich), immediately swirl gently and incubate RT for 12' for cross-linking.

2. Add Glycine (Sigma-Aldrich) to a final concentration of 0.125M to stop cross-linking, swirl gently to mix and incubate RT for 5'.
3. Remove the PBS/Formaldehyde/glycine solution and from this point onward work on ice.
4. Rince the flask twice with 10ml of ice-cold PBS.
5. Detach the cells by scraping in 5ml of ice-cold PBS and transfer the fixed-cell suspension in a 15ml falcon tube.
6. Rinse the flask with other 5ml of ice-cold PBS + 1X Protease Inhibitor Cocktail (Roche, Basel, Switzerland) and transfer the remaining cells in the 15ml falcon tube.
7. Centrifuge 400rcf 4°C 8'.
8. Remove and aspirate as much supernatant as you can, paying attention not to disturb the cell pellet.
   At this point you can proceed immediately with cell lysis and chromatin shearing or you can snap-freeze the cross-linked cells in liquid nitrogen and store them at -80°C for up to 6 months.
9. Resuspend pellet in 10ml of ice-cold XChIP Buffer B + PIC and incubate on ice for 10'.
10. Centrifuge 5' 1200 rcf 4°C and discard the supernatant.
11. Aspirate the supernatant.
12. Resuspend the pellet in 10ml of ice-cold XChIP Buffer C + PIC and incubate on ice for 10'.
13. Centrifuge 10' 2500rcf 4°C and aspirate the supernatant.
14. Resuspend the pellet in 1ml of ice-cold XChIP Sonication Buffer + PIC, incubate on ice for 5' and then sonicate using the COVARIS as specified in the dedicated section.
15. Centrifuge 20' 4°C 13200rcf, and transfer the SN in a new microcentrifuge tube (use 75µl to check the fragmentation).
16. Sheared chromatin can be used immediately or can be snap-frozen in liquid nitrogen and stored at -80°C for up to 6 months.

Home-made protocol for ChIP-seq on CHQ5B

Cross-linking and harvest CHQ5B for ChIP-experiments

For ChIP experiments CHQ5B Mb cells are collected at ~50% of confluency, and Mt after 9 days of differentiation. Each chromatin preparation is performed using ~10 million of fresh cells.

1. Discard the medium and wash each bottle twice with RT PBS.
2. Scrape cells in 10 ml of RT PBS and collect cell suspension in 50ml Falcon tubes.
3. Centrifuge 10' 150rcf 4°C and discard the SN.
4. Resuspend cell pellet in 5ml of XChIP buffer B + PIC (Roche) and incubate 10' on ice with occasional mixing by inversion.

5. Centrifuge 30' at 2,500rcf 4°C, and carefully discard the SN.
6. Resuspend pellet in 10 ml of PBS + PIC, add 1ml of XChIP buffer A + PIC, and incubate 10' on ice and 5' RT with occasional mixing by inversion.
7. Centrifuge 30' 2500rcf 4°C, and discard SN.
8. Resuspend pellet in 10ml XChIP buffer C + PIC, incubate 10' on ice with occasional mix by inversion.
9. Centrifuge 10' 2000rcf 4°C, discard SN.
10. Resuspend pellet in 1ml of XChIP sonication buffer + PIC (use 75µl to check chromatin integrity).
11. Sonicate with Covaris S2 for 8' with the settings specified in the Covaris section.
12. Centrifuge 20' 4°C 13200rcf, and transfer the SN in a new microcentrifuge tube (use 75µl to check the fragmentation).

Sheared chromatin can be used immediately or can be snap-frozen in liquid nitrogen and stored at -80°C for up to 6 months.

Chromatin integrity check

1. Add an equal volume (75µl) of XChIP dilution buffer to the 75µl chromatin samples.
2. Add 5M NaCl to a final concentration of 200mM and incubate 5h at 65°C with occasional mixing to reverse crosslinking.
3. Extract once with phenol and twice with chloroform. Briefly:
4. add 1 volume (150µl) of TE-saturated phenol pH10.5 to the sample and mix 10" by vortexing;
5. centrifuge 13,800rcf for 5' and the transfer the upper aqueous phase in a new microcentrifuge 1.5ml tube;
6. add 1 volume (150µl) of chloroform and mix by inversion;
7. centrifuge 13,800rcf for 5' and the transfer the upper aqueous phase in a new microcentrifuge 1.5ml tube;
8. repeat once the previuos two steps.
9. Proceed with an Ethanol/salts DNA precipitation.
10. Quantify DNA at NanoDrop and load 500 ng of DNA in a 1% w/v agarose/TAE gel.
11. Check the pre-sheared chromatin integrity and the correct fragmentation.

Protein A/G preparation

Each Immunoprecipitation experiment required 5 million cells.

1. For 5million cells mix 75µL proteinA and 75µL proteinG sepharose(Pharmacia, Upsala, Sweden) in a 2ml LoBind Eppendorf tube.
2. Centrifuge 1,200rcf 4°C 5' and carefully remove SN.
3. Wash the pellet with 1.4ml of XChIP incubation buffer + 0.1% BSA.
4. Centrifuge 1,200rcf 4°C 5' and carefully remove SN.
5. Repeat the last two washing steps twice.

6. Add 1.5ml of ChIP incubation buffer + 0.1% BSA to the beads and incubate 12-16h at 4°C on a rotating wheel.

Immunoprecipitation

1. Dilute the chromatin 1:1 with XChIP dilution buffer.
2. Add 3µg of the antibody of interest (TBP and H3K9me3) and incubate for 3h at 4°C on a rotating wheel.
3. At the end of the 3h-incubation, centrifuge the Protein A/G sepharose mixture for 5' 1,200rcf 4°C, remove the supernatant and add the ChIP sample (chromatin + antibody of interest) to the beads-containing tube.
4. Incubate 12-16h at 4°C on a rotating wheel.

Washing and elution

1. At the end of the 16h incubation, centrifuge the samples 5' 1,200rcf 4°C and remove the supernatant.
2. Wash the pellet twice with 1.7ml of ice-cold XChIP Wash Buffer A and incubate at 4°C on a rotating wheel for 7'.
3. Centrifuge 5' 1,200rcf 4°C and remove the supernatant.
4. Wash the pellet once with 1.7ml of ice-cold XChIP Wash Buffer B and incubate at 4°C on a rotating wheel for 7'.
5. Centrifuge 5' 1,200rcf 4°C and remove the supernatant.
6. Wash the pellet once with 1.7ml of ice-cold XChIP Wash Buffer C and incubate at 4°C on a rotating wheel for 7'.
7. Centrifuge 5' 1,200rcf 4°C and remove the supernatant.
8. Wash the pellet twice with 1.7ml of ice-cold XChIP Wash Buffer D and incubate at 4°C on a rotating wheel for 7'.
9. Centrifuge 5' 1,200 rcf 4°C and remove the supernatant.
10. Remove all the supernatant and resuspend the pellet in 1.2ml of RT XChIP Elution Buffer and incubate for 30' on a rotating wheel at RT.
11. Centrifuge 2,000rcf 10' RT and collect the supernatant in 3 new 1.5ml LoBind Eppendorf tubes (400µl each). Discard the pellet.
12. Add 5M NaCl to a final concentration of 250mM and incubate at 65°C for 5h with occasional mixing to reverse crosslinking.

Phenol/chloroform DNA extraction and Ethanol/salt concentration

1. After the 5h-incubation add 1 volume of TE saturated phenol pH10.5, mix by vortexing and centrifuge 16,100rcf 5' RT.
2. Transfer the upper aqueous phase in a new 1.5ml LoBind Eppendorf tube and add 1 volume of chloroform.
3. Mix by inversion and centrifuge 15,600rcf 5' RT.
4. Repeat the previous two steps.
5. Transfer the upper aqueous phase in a new 1.5ml LoBind Eppendorf tube.
6. Add 5µg of Glycogen, 3M NaAc to 300mM final concentration and 1ml of ice-cold absolute Ethanol.
7. Mix by vortexing and incubate 12-16h at -20°C.

8. Centrifuge 16,100rcf 30' 4°C, discard the supernatant and wash the pellet with 70% Ethanol.
9. Repeat the previous step once and centrifuge 16,100rcf 20' 4°C.
10. Remove all the supernatant and leave the pellet to air-dry.
11. Dissolve the pellet in 30µl of SIGMA H_2O and collect the resuspended pellets of the same sample in one tube (90µl).

Semi-quantitative PCR

To evaluate the enrichment of immunoprecipitated samples I performed a semiquantitative PCR.
The primers I used:

SAT2 FOR	5'–CTGCAATCATCCAATGGTCG–3'
SAT2 REV	5'–GATTCCATTCGGGTCCATTC–3'
FOS FOR	5'–CATCTGCGTCAGCAGGTTTCCAC–3'
FOS REV	5'–TGAGCATTTCCCAGTTCCTGTCT–3'

1µl of immunoprecipitated DNA was used as template for semiquantitative PCR.
PCR Reaction Mix (for 1 sample):

H_2O	13.6µl
5X GoTaq Reaction Buffer	4µl
Template	1µl
10mM dNTPs	0.4µl
10µM FOR Primer	0.4µl
10µM REV Primer	0.4µl
GoTaq [5U/µl]	0.2µl
Total	20µl

Thermal cycler program:

95°C	4'	
95°C	20"	x40times
54°C	30"	
72°C	30"	
72°C	7'	
4°C	hold	

Every 2 cycles, starting from 26 till 40, a sample was removed from the thermal cycler and the reaction was stopped on ice, to check the amplification. Load the sample on a 1% w/v TAE-agarose gel to check amplification.

Library preparation

I prepared SOLiD sequencing libraries from four immunoprecipitated samples: Mb TFIID, Mt TFIID, Mb H3K9me3, Mt H3K9me3.
Buffers and enzymes that are used in this section are all included in the SOLiD Fragment Library Preparation Kit.

Sonication and end-repair of the immunoprecipitated DNA

1. The immunoprecipitated DNA was fragmented as reported in the section "Fragmentation of the IP DNA for SOLiD fragment library preparation".
2. Quantitate the DNA using the Qubit® fluorometer and check the fragment size with Agilent Bioanalyzer.
3. End-polish the DNA to obtain 5'-phosphorilated, blunt-ended dsDNA fragments. For each immunoprecipitated sample set-up the reaction as follows:

Sheared DNA	130µl
5X Reaction Buffer	40µl
10mM dNTPs	8µl
End Polishing Enzyme E1	4µl
End Polishing Enzyme E2	16µl
H_2O	2µl
Total	200µl

4. Vortex to mix and spin.
5. Incubate RT for 30'.
6. Purify the DNA with 1.5volumes of Agencourt® AMPure® XP beads and elute in 60µl of Low TE buffer.

Adapters ligation

1. Set-up the adapters ligation mix:

End-Polished DNA	60µl
5X Ligation Buffer	20µl
Ligase	9µl
50µM P1 adaptor	0.5µl
50µM P2 barcoded adaptor	0.5µl
H_2O	9µl
Total	100µl

2. Vortex and spin.
3. Incubate RT for 15', then proceed with an Agencourt® AMPure® XP beads DNA purification using 1.5 volumes of beads.
4. Elute the DNA in 12µl of Low TE buffer.

Library amplification

Before emulsion-PCR, bead-enrichment and sequencing, the library is PCR-amplified in order to obtain an amount of DNA suitable for sequencing procedures.

1. Library amplification mix:

Adapters-ligated DNA	12µl
50µM P1 primer	1.8µl
50µM P2 primer	1.8µl
Platinum® PCR Amplification Mix	72µl
H_2O	2.4µl
Total	90µl

2. Vortex and spin.
3. Incubate in the thermal cycler with the following program:

72°C	20’	
95°C	5’	
95°C	15”	x10 times
62°C	15”	
70°C	1’	
70°C	5’	
4°C	hold	

4. Purify the amplified library using 1.5v of Agencourt® AMPure® XP beads. Elute in 20µl of Low TE Buffer.
5. Quantitate the amplified library with the DNA HS kit on the Qubit® fluorometer.
6. Check the quality of the library with Agilent Bioanalyzer. If adapters are visible on the Agilent profile, repeat the AMPure® XP beads purification step.

The sequencing core of Prof. Valle Group performed emulsion PCR, bead-enrichment and SOLiD sequencing.

Diagenode True MicroChIP kit

For this ChIP-seq experiment I used 100,000 CHQ5B Mb for each immunoprecipitation. In particular, I performed ChIP-seq with Anti-H3K4me3, -H3K9ac and -H3K27me3 antibodies and a control.

Buffers and enzymes whose company is not specified in this section are included in the Diagenode True MicroChIP kit.

Crosslinking protocol

Use only fresh cell flasks.

1. Place PBS, cell culture medium and Trypsin-EDTA at room temperature.

2. Since using adherent cells, discard culture medium to get rid of dead cells and was the cell layer by adding 10ml PBS.
3. Detach cells by tripsinization and collect by adding medium (DMEM + 10% FBS).
4. Transfer the cell suspension in a 15ml Falcon tube and centrifuge 150rcf 8' RT.
5. Discard the supernatant and resuspend the cells in 3ml DMEM+10% culture medium and count the cells using Neubauer chamber. Cells must be in single-cell suspension.
   The amount of cells to use for fixation is calculated as "number of IP to perform + 1 control" x "100,000cells/IP" and a unique batch of chromatin was fixed.
6. Aliquot the number of cells to be fixed as a unique batch in a 1.5ml tube and add DMEM + 10% FBS to 1ml final volume.
7. Add 27µl of 37% Formaldehyde to the 1ml-sample and invert the tube immediately two or three times to ensure complete mixing.
8. Incubate 10' RT with occasional manual agitation.
9. Add 115µl of 1.25M Glycine to the sample and immediately mix by inversion 4-5 times.
10. Incubate RT for 5'.
    a. From this point onward work on ice!
11. Centrifuge 300rcf 10' 4°C.
12. Aspirate the supernatant slowly and leave approximately 30µl of the solution in order not to remove the crosslinked cells.
13. Wash the crosslinked cells with 1ml of ice-cold HBSS containing PIC:
    a. add 1ml of HBSS + PIC and gently vortex the tube to completely resuspend he cells;
    b. centrifuge 300rcf 10' 4°C
14. Aspirate the supernatant and keep cell pellet on ice. Proceed directly to cell lysis or, if desired, the cell pellet can be snap-frozen and stored at -80°C for up to 2 months.

Cell lysis and chromatin shearing

1. Prepare Lysis Buffer:
   a. warm Lysis Buffer to RT (make sure there are no crystals in the Lysis Buffer before using);
   b. add PIC to Lysis Buffer: this is the complete Lysis Buffer tL1.
2. Add Lysis Buffer tL1 to the cells: use 25µl of complete Lysis Buffer tL1 every 100,000 cells.
3. Agitate manually the bottom of the tube to resuspend the cells and allow bubbles to form.
4. Incubate on ice for 5'.
5. Add HBSS + PIC to the cell lysate: use 75µl of HBSS + PIC every 100,000 cells.

6. Dispense 100µl of cell lysate in Covaris AFA microtubes. Make sure there are no precipitates before splitting cell lysate in Covaris microtubes, otherwise gently heat until crystals disappear.
7. Sonicate samples using COVARIS S2 with the settings described in the dedicated section.
8. After sonication, transfer the sheared chromatin to a new 1.5ml LoBind Eppendorf tube.
9. Centrifuge 14,000rcf 10' 4°C and collect supernatant in a new 1.5ml LoBind tube.
10. Snap-freeze the chromatin and analyse chromatin shearing using 50µl of the sheared cells:
    a. transfer 50µl of the cell lysate to a new 1.5ml LoBind tube, add 5M NaCl to 250mM final concentration;
    b. incubate 4h at 65°C;
    c. perform standard phenol:chloroform DNA extraction and isopropanol/salts DNA concentration;
    d. analyse chromatin shearing on a 2% 0.5X TBE agarose gel electrophoresis.
11. The sheared chromatin can be used immediately in ChIP or, if desired, it can be snap-frozen and stored at -20°C for up to 2 months.

Magnetic immunoprecipitation

1. Prepare the complete ChIP buffer tC1 by adding protease inhibitor complex to ChIP Buffer tC1.
2. Add 100µl of complete ChIP buffer tC1 per 100µl of sheared chromatin.
3. Add 200µl of diluted sheared chromatin per tube for each IP.
4. Add antibodies:

H3K27me3	1µg
H3K9ac	1µg
H3K4me3	1µg
Normal IgG Rabbit	2µg

5. Incubate 16h 4°C on a rotating wheel (40rpm).
6. After the 16h-incubation prepare magnetic beads. Each IP requires 11µl of beads.
    a. Add 55µl of Beads Wash Buffer tBW1 to 11µl stock solution of beads for each IP and scale accordingly;
    b. Resuspend the beads and place them in a magnetic rack;
    c. Repeat the previous two steps once;
    d. Resuspend the bead pellet in 11µl of Beads Wash Buffer tBW1 per IP reaction.
7. Add 10µl of prewashed Protein A-coated beads to each IP tube.
8. Incubate the IP tubes for 2h on a rotating wheel (40rpm) at 4°C.
9. Briefly spin the tubes and place them in the magnetic rack.
10. Wait for 1' and remove the supernatant.

11. Wash the beads with Wash Buffer tW1:
    a. add 100µl of Wash Buffer tW1 and gently shake the tubes to resuspend the beads;
    b. incubate 4' on a rotating wheel at 4°C.
12. Wash the beads with Wash Buffer tW2 following the procedure described for Wash Buffer tW1.
13. Wash the beads with Wash Buffer tW3 in the same way.
14. Wash the beads with Wash Buffer tW4 in the same way.

DNA decrosslinking and purification

1. After removing the last wash buffer, add 400µl of Elution Buffer tE1 to the beads and incubate for 30' on a rotating wheel at RT.
2. Briefly spin the tubes and place them in the magnetic rack for 2 minutes.
3. Transfer the supernatant to a new tube and add to this supernatant 16µl of Elution Buffer tE2.
4. Incubate for 4h in a thermomixer at 1,300rpm at 65°C.
5. Briefly spin the tubes.
6. Add 400µl of phenol:chloroform:isoamyl alcohol 25:24:1 to each sample and mix by vortexing.
7. Centrifuge samples 10' 14,000rcf RT, and transfer the upper aqueous phase into new 1.5ml LoBind Eppendorf tubes.
8. Precipitate the DNA by adding 40µl of Precipitant tP1, 2µl of co-precipitant tCP2 and 1ml of ice-cold absolute ethanol to the samples.
9. Vortex and incubate at -20°C for 16h.
10. Centrifuge 25' 14,000rcf 4°, remove the supernatant and add 1ml of ice-cold 70% ethanol to the pellet.
11. Centrifuge 10' 14,000rcf 4°C, remove the supernatant and leave the tubes to air-dry
12. Resuspend the pellet in 20µl of Low TE buffer. That corresponds to the purified immunoprecipitated DNA.
13. Quantitate the DNA using the Qubit® fluorometer.

SOLiD Fragment Library construction

DNA end repair

1. For each immunoprecipitated DNA, prepare the reaction mix in a 1.5ml LoBind Eppendorf tube:

ChIP DNA	18µl
5X Reaction Buffer	10µl
10mM dNTP Mix	2µl
End Polishing Enzyme 1	0.5µl
End Polishing Enzyme 2	4µl
Nuclease-free H_2O	15.5µl
Total	50µl

2. Vortex briefly, pulse-spin and incubate at room temperature for 30'.
3. Purify the end repaired DNA using 1.8 volumes of Agencourt® AMPure® XP beads. Elute in 30µl of Low TE buffer.

P1 and P2 adaptors ligation

1. Dilute the P1 and P2 adaptors 1:20 in nuclease-free water [2.5pmol/µl].
2. Prepare the reaction mix:

P1 Adaptor [2.5pmol/µl]	1µl
P2 Barcoded Adaptor [2.5pmol/µl]	1µl
5X T4 Ligase Buffer	20µl
End-repaired DNA	30µl
T4 DNA ligase [5U/µl]	5µl
Nuclease-free water	43µl
Total	100µl

3. Incubate 10' RT.
4. Purify the end repaired DNA using 1.8 volumes of Agencourt® AMPure® XP beads. Elute in 20µl of Low TE buffer.

Nick Translation and library amplification

1. For each sample prepare the reaction mix:

Platinum® PCR Amplification Mix	100µl
Library PCR Primer 1 [50µM]	2.5µl
Library PCR Primer 2 [50µM]	2.5µl
Adaptor-Ligated DNA	20µl
Total	125µl

2. Divide the mix into two separate PCR tubes (125µl each) and run the following program in a thermal cycler:

72°C	20'	
95°C	5'	
95°C	15"	x15 times
62°C	15"	
70°C	1'	
70°C	5'	
4°C	hold	

3. After amplification, combine the samples in one 1.5ml LoBind tube.
4. Purify the amplified DNA with 1.8 volumes of Agencourt® AMPure® XP beads. Elute in 30µl of Low TE Buffer.
5. Quantitate the amplified library using Qubit® fluorometer.
6. Assess the quality of the library using Agilent Bioanalyzer.

Native ChIP on CHQ5B

Cell lysis and nuclei purification

2 x 10^6 cells were used for chromatin preparation, and 1 x 10^6 cells where then used for each immunoprepcipitation experiment.

1. Resuspend cells in 1ml of PBS and transfer the sample in a 2ml eppendorf tube.
2. Centrifuge 600rcf 10', then discard the supernatant.
3. Resuspend cell pellet completely in 1ml of ice-cold nChIP Buffer1 [0.3M sucrose, 5mM Na Butyrate, 0.1mM PMSF, 0.5mM DTT, 60mM KCl, 15mM NaCl, 5mM $MgCl_2$, 0.1mM EDTA, 15mM Tris/Cl pH 7.5].
   NOTE: cells must be in a single cell suspension.
4. Add 1ml nChIP Buffer 2 and homogenize on ice in a Dounce homegeniser for 3minutes (90 strokes) using the pestle A.
5. Incubate on ice in the Dounce homogeniser for other 7 minutes (no more than 7minutes).
6. During the 7-minute incubation fill 8ml nChIP Buffer 3 into a 15ml Falcon tube (use 2 Falcon tubes for each sample).
7. After the 7-minute incubation, overlay the 8ml nChIP Buffer 3 with 1ml cell suspension. Disturb the interface between nChIP Buffer 3 and cell suspension with the pipette tip.
8. Centrifuge 3,500rcf 4°C 30' in a swing-out rotor.
9. Remove the supernatant carefully and completely. Pay attention not to touch the pellet.

MNase chromatin digestion

1. Resuspend the two nuclei pellets for each sample in 1ml MNase digestion buffer and aliquot 500µl of this suspension into two 1.5ml microcentrifuge tubes.
2. Add 15U of MNase and incubate 5' at 37°C.
3. Stop the reaction by adding 20µl 0.5M EDTA to each 500µl MNase digested sample and put the tube on ice.
4. Centrifuge 13,000rcf 10' 4°C.
5. Transfer the supernatant to a new tube (S1).
6. Quantify chromatin in S1 by measuring OD_{260} against MNase digestion buffer (OD_{260}=1 => [DNA]= 50ng/µl). Since lot of protein is present in the solution, this DNA quantification is not precise, but is sufficient for reproducibility.
   CHECKPOINT: phenol:chloroform DNA extraction from a 50µl aliquot of S1 and gel electrophoresis of 20µl of this extract in a 2% 0.5X TBE agarose gel to verify the digestion. Expected result: mono–pentanucleosomes.
   STOPPING POINT (optional): S1 can be stored over night at -20°C.
7. Prepare a 2% 0.5X TBE agarose gel.
8. Thaw S1 and centrifuge 13,000rcf 10' 4°C.

9. Transfer the supernatant in a new 1.5ml Eppendorf tube.
10. Centrifuge 13,000rcf 10' 4°C and transfer the supernatant in a new 1.5ml Eppendorf tube.
11. Centrifuge 13,000rcf 10' 4°C and transfer the supernatant in a new 1.5ml Eppendorf tube. These triple centrifugations are important to reduce the aspecific immunoprecipitation background. If a pellet is still visible after the third centrifugation, centrifuge again.
12. Use 50µl of S1 for a phenol/chloroform extracted DNA and load 20µl on the gel (95V, 30'). Mono to pentanucleosomes should be visible.

Incubation with the antibody

400µl of the sample are used for each immunoprecipitation: one ChIP uses the H3K9me3 and another one uses a normal mouse IgG as a control.

1. Add appropriate amounts of stock solutions to generate the antibody nChIP Incubation Buffer (150mM NaCl, 20mM Na Butyrate, 5mM EDTA, 0.2mM PMSF, 20mM Tris/Cl pH 7.4). Final volume is 1ml for each immunoprecipitation sample. Take into consideration the volume of S1 (in this case 400µl) and its buffer.
2. Dilute S1 in 1ml final volume of nChIP incubation buffer.
3. Add 2µg of antibody (2µg H3K9me3 in the sample tube, 2µg of normal mouse IgG in the control tube).
4. Incubate overnight on a slowly rotating wheel at 4°C.

Immunoprecipitation and DNA extraction

1. Prepare 50µl of sepharose protein A + sepharose protein G (25µl sepharose protein A + 25µl sepharose protein G) for each sample.
2. Wash the beads:
3. Short spin;
4. Carefully remove the supernatant;
5. Replace the supernatant with an equal volume of sterile H_2O;
6. Add 50µl of the mixed sepharose protein A + G to each tube.
7. Incubate for 5 hours at 4°C on a slowly rotating wheel.
8. Prepare nChIP washing buffers (10ml of each buffer for each sample, including the control without antibody) and cool down to 4°C.
9. After the 5h-incubation, centrifuge the chromatin-antibody-beads mixture 10' 4°C 11,600rcf.
10. Keep the supernatant in a 2ml LoBind tube (this is the unbound fraction).
11. Resuspend the pellet in 1ml nChIP Washing buffer 1 and transfer into a 15ml Falcon tube containing 9ml nChIP washing buffer 1.
12. Mix for 10' on a rotating wheel at 4°C.
13. Centrifuge 10' 4,000rcf 4°C and pour off supernatant.
14. Add 10ml nChIP washing buffer 2, mix for 10' on a rotating wheel at 4°C and centrifuge 10' 4,000rcf 4°C.
15. Pour off supernatant.

16. Add 10ml nChIP washing buffer 3, mix for 10 min on a rotating wheel at 4°C, and centrifuge 10' 4,000 rpm 4°C.
17. Pour off supernatant.
18. Centrifuge 10' 4,000rcf 4°C.
19. Remove remaining supernatant completely.
20. Resuspend pellet in 500µl nChIP elution buffer at room temperature.
21. Transfer to a 1.5ml LoBind tube and incubate 15' RT on a rotating wheel.
22. Centrifuge 10' 11,600rcf 20°C.
23. Transfer the supernatant to a new 1.5ml LoBind tube. This is the bound fraction.

DNA extraction

Use a standard phenol/chloroform DNA extraction. Briefly:

1. Add an equal volume of Phenol TE saturated pH 10.5. Mix by inversion and centrifuge 16,100rcf RT 5'.
2. Transfer the aqueous phase into a new microcentrifuge tube, add an equal volume of Chloroform, and mix by inversion. Centrifuge 16,100rcf RT 5'.
3. Repeat the previous step.
4. Transfer the aqueous phase into a new LoBind 1.5ml microcentrifuge tube.
5. Add 4µl of a 5µg/µl glycogen stock solution.
6. Add 5M NaCl to 250mM final concentration, and add 1 volume isopropanol.
7. Incubate 12-16h at -20°C.
8. Precipitate by centrifugation: 30' 16,100rcf 4°C and wash the pellet twice with 70% ethanol.
9. Dry the pellet and resuspend in 20µl 10mM Tris/Cl pH 7.4.
10. Quantitate the DNA using the Qubit® 2.0 fluorometer (DNA HS kit).

Ion Proton low input protocol

The nChIP on 1million cells yielded 2ng of immunoprecipitated DNA. This DNA is quite heterogeneous in its dimensions: in fact it shows a nucleosome-ladder pattern. To prepare a sequencing library, the DNA needs to be sonicated in order to bring the average DNA fragment size to approximately 170bp.

All the Buffer and enzymes used here are part of the Ion Proton Fragment Library Preparation kit.

1. Sonicate the nChIP immunoprecipitated DNA using the covaris conditions described in the section "Fragmentation of the nChIP DNA for Ion Proton fragment library preparation".
2. End-repair the sonicated DNA fragments. In a 1.5ml LoBind tube mix:

1-10ng di ChIP DNA	~50-79µl
H_2O	to 79µl
5X End Repair Buffer	20µl
End Repair Enzyme	1µl
Total	100µl

3. Mix well by pipetting and incubate RT for 20'.
4. Purify the End-Repaired DNA with 1.8 volumes of Agencourt® AMPure® XP beads and elute in 25µl of Low TE buffer.

Barcoded adapters ligation

1. Dilute 1:16 both Ion P1 Adapter and Barcoded adapters in H_2O.
2. In a 0.2ml PCR tube prepare the ligation mixture. Add the reagents in the order reporter below:

End-repaired ChIP DNA	25µl
10X Ligase Buffer	10µl
Ion P1 adapter, 1:16	1µl
Ion Barcode adapter 1:16	1µl
dNTP mix	2µl
H_2O	51µl
DNA ligase	2µl
Nick repair Polymerase	8µl
Total	100µl

3. Incubate samples in a thermal cycler with the following program:

25°C 15'
72°C 5'
4°C Hold (it is not a stopping point: proceed as soon as possible with purification and size selection)

4. Purify the adapter-ligated DNA using 1.8volumes of Agencourt® AMPure® XP beads. Elute in 25µl of Low TE buffer.

Library amplification

1. Prepare the library amplification mixture as follows (recipe for 1 sample):

Platinum® PCR SuperMix High Fidelity	100µl
Library Amplification Primer Mix	5µl
ChIP DNA	25µl
Total	130µl

2. Split the 130µl-reaction volume in two 0.2ml PCR tube and run the following program on a thermal cycler:

95°C 5'

95°C 15"
62°C 15"
70°C 1'
} x15 times

70°C 5'
4°C hold

3. Collect the same sample in a unique 1.5ml LoBind tube

4. Purify the amplified library with 1.5 volumes of AMPure® XP beads and elute in 50µl of Low TE buffer.
5. Purify the library another time with 1.5 volumes of AMPure® XP beads and elute in 25µl of Low TE buffer.
6. Quantitate the DNA using the Qubit® fluorometer and assess the library quality on Agilent Bioanalyzer.

DNA methylation

Sodium Bisulfite DNA conversion

Bisulphite conversion of the genomic DNA was conducted using the EZ DNA Methylation-Lightning™ Kit (Zymo Research, Irvine, CA) following manufacturer's instructions.

1. Dilute 1µg of NanoDrop™-quantified DNA in 20µl of water.
2. Add 130µl of Lightnin Conversion Reagent to the 20µl of DNA from the previous step in a 0.2ml PCR tube.
3. Mix by pipetting and spin to ensure no droplets are present on the surface of the tube.
4. Place the PCR tube in a thermal cycler with the program: 98°C for 8', 54°C for 60' and 4°C up to 20h.
5. Add 600µl of M-Binding Buffer to a Zymo-Spin™ IC Column and place it into a provided Collection Tube.
6. Load the 150µl sample (after the thermal cycler step) into the M-Binding Buffer containing column. Close the cap and mix by inverting the tube several times.
7. Centrifuge 15,000rcf 30" RT, then discard the flow-through.
8. Add 100µl of M-Wash Buffer to the column.
9. Centrifuge 15,000rcf 1' RT.
10. Add 200µl of L-Desulphonation Buffer to the column close the cap and incubate RT for 20'. After the incubation centrifuge 15,000g RT 1'.
11. Add 200µl M-Wash Buffer to the column, then centrifuge 15,000rcf 30" RT. Repeat this wash step.
12. Let the column stand at RT for 2'.
13. Place the column on a new 1.5ml microcentrifuge tube and add 10µl M-Elution Buffer to the column matrix.
14. Incubate RT for 2', then centrifuge for 1' at 15,000rcf
15. Add other 10µl of Elution buffer directly to the column matrix and incubate RT for 2'.
16. Centrifuge 15,000rcf 1' RT to elute the DNA.

A preliminary protocol: MeSS v1

Adapter sequences for this protocol:

Oligo P1 UP:
5'–CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGA–3'

Oligo P1 DOWN
5'–
NNNNNNTCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGC
C–3'

Oligo P2 UP
5'–(P)AGAGAATGAGGAACCCGGGGCAGTT–3'

Oligo P2 DOWN
5'–CTGCCCCGGGTTCCTCATTCTCTNNNNNN–3'

Protocol

1. Bisulfite-treat 1µg of DNA.
2. Sonicate the bisulfite treated DNA as specified in the dedicated Covaris section.
3. In a 0.2ml PCR tube mix:

Bisulfite-treated and sonicated DNA	500ng in 6.4µl
10X T4 DNA ligase buffer	0.8µl
T4 Polynucleotide kinase [10U/µl]	0.8µl
Total	8µl

4. Incubate the following program in a pre-heated thermal cycler:

37°C	30'
65°C	20'
65°C	Hold

During the 65°C-hold phase add 1µl of "P1 adapter" and 1µl of "P2 adapter". Mix well by pipetting. These adapters were obtained by mixing100µM Oligo (P1 or P2) UP and 100µM Oligo (P1 or P2) DOWN in a 1:5 ratio.

5. Incubate 65°C 10' and immediately put on ice for 5'.
6. Incubate at 16°C for 10' and, in the meanwhile, prepare the ligation mixture:

10X T4 DNA ligase buffer	1µl
10mM ATP	1µl
T4 DNA ligase [10U/µl]	7µl
40% PEG8000	
Total	10µl

Since PEG8000 greatly increases the viscosity of the mixture, mix by pipetting several times.

7. After the 10' incubation add the ligation mixture and incubate at 16°C for 16h.
8. Add 0.66µl of 10 mM dNTPs mix and 0.5µl of E.coli DNA Polymerase Klenow fragment (5U/µl).
9. Incubate at 25°C for 30'.
10. Purify the sample twice with 1.8 volumes of Agencourt® AMPure® XP beads. Elute in 20µl of Low TE buffer.
11. Incubate in a thermal cycler with the following program:

95°C	5'	
95°C	15"	x15 times
62°C	15"	
70°C	1'	
70°C	5'	
4°C	hold	

12. Purify the sample using 1.5 volumes of Agencourt® AMPure® XP beads.
13. Elute in 20µl Low TE buffer.
14. Quantitate the DNA using the Qubit® fluorometer and assess the library quality with Agilent Bioanalyzer.

The definitive protocol: MeSS v2

300ng of Mb NanoDrop™-quantified Mb DNA were sonicated and then bisulfite-treated.

Adapters:

MeSS blocked P2:
5' - CTGCCCCGGGTTCCTCATTCTCTGTGTAAGAGGCTGCTGTACGGCCAAG GCGTNNNNNN-(P) - 3'

Oligo P1 UP:
5'–CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGA–3'

Oligo P1 DOWN
5'–
NNNNNNTCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGC C–3'

Template phosphorylation

1. Prepare the reaction mix in a 0.2ml PCR tube:

300ng DNA	15µl
T4 DNA ligase (Invitrogen)	4µl
PNK (NEB)	1µl
_____	
Total	20µl

2. Mix well by pipetting and pulse-spin.
3. Incubate in a thermal cycler with heated lid with the following program:

22.5°C	30’
80°C	20’
4°C	Hold

Oligonucleotide hybridization for template switching

1. Add to the previous reaction tube:

i.	50µM MeSS Blocked P2	1µl
ii.	10mM dNTPs	1.5µl
iii.	NEB Buffer 2	1µl
iv.	H_2O	5.5µl
v.	_____	
vi.	Total	29µl

2. Incubate for 5’ at 90°C in a thermal cycler, and then put immediately on ice for 5’.

Template switching

1. Add to the previous reaction tube:

DNA polymerase I (NEB)	1µl
_____	
Total	30µl

2. Incubate in a thermal cycler with the following program:

16°C	30’
80°C	5’

3. Transfer the sample in a 1.5ml LoBind tube.
4. Purify the sample using 1.5 volumes of Agencourt® AMPure® XP beads. Elute in 30µl of H_2O.

P1 adapter hybridization and ligation

1. Prepare the hybridization mix in a 0.2ml PCR tube:

DNA from previous step	30µl
50µM MeSS P1 UP	1µl
50µM MeSS P1 DOWN	1µl
H2O	6µl
_____	
Total	38µl

2. Incubate in a thermal cycler for 5’ at 90°C, then put immediately on ice for 5’.
3. Add to the previous tube:

5X T4 DNA ligase buffer (Invitrogen)	10µl
DNA ligase [5U/µl]	2µl
Total	50µl

4. Incubate 16h in a thermal cycler without heated lid.
5. Inactivate the ligase for 20' at 80°C in a thermal cycler.
6. Purify the adapter-ligated DNA with 1.5 volumes of Agencourt® AMPure® XP beads. Elute in 50µl H_2O.
7. Purify again the sample with 1.5 volumes of Agencourt® AMPure® XP beads. Elute in 20µl H_2O.

Library amplification PCR

1. Prepare the amplification mix:

5X GoTaq Buffer	4µl
10mM dNTPs	0.4µl
50µM SOLiD PCR Primer 1	0.5µl
50µM SOLiD PCR Primer 2	0.5µl
DNA	14.4µl
GoTaq2 [5U/µl]	0.2µl
Total	20µl

2. Incubate in a thermal cycler with the following program:

95°C	5'	
95°C	15"	x15 times
62°C	15"	
70°C	1'	
70°C	5'	
4°C	hold	

3. Purify the sample using 1.5 volumes of Agencourt® AMPure® XP beads. Elute in 20µl Low TE buffer.
4. Repeat the previous step once.
5. Quantitate the DNA using the Qubit® fluorometer and assess the library quality with Agilent Bioanalyzer.

SureSelect Methyl-Seq Target enrichment

The kit is a target enrichment system, which can be used to focus on regions where methylation is known to impact gene regulation. It can target CpG islands, DNaseI hypersensitive sites and can reveal methylated regions that are not able to be detected by reduced representation bisulfite sequencing (RRBS), and the

enrichment-based technique of methylated DNA immunoprecipitation sequencing (MeDIP-seq).

For this kit I used the DNA of 2 Mb samples (at division 34.5 and 36) and 2 Mt samples (9 days of differentiation).
Before starting the integrity of the starting DNA was evaluated on a 0.6% agarose gel electrophoresis in 1X TAE buffer.
Assess the quality of the DNA using NanoDrop™ 1000 (Thermo Scientific, Waltham, MA): the ratio A_{260}/A_{280} should be between 1.8 and 2.
Quantitate the DNA using the Qubit® 2.0 fluorometer (Invitrogen, Life Technologies, Carlsbad, CA).

Protocol

DNA shearing

1. Shear the DNA following the instruction reported in "1. Covaris sonication of the DNA before bisulfite treatment".
2. After shearing quantitate the DNA using Qubit® fluorometer (DNA HS kit): proceed with no less than 2µg of DNA for next steps.
3. Agilent 2100 BioAnalyzer to check the sonication (Agilent Technologies). DNA HS kit.

Repair the DNA ends

We proceed using 3µg of DNA in 50µl Low TE (desiccate the sample in a Speed Vac to reach this concentration).

1. Add 50µl of XT2 End-Repair Master Mix to the sample (on ice) and mix well by pipetting.
2. Incubate in the thermal cycler at 20°C for 30'. Do not use heated lid.
3. Purify sample using 1.8 volumes of Agencourt® AMPure® XP beads and elute the samples in 22µl of H_2O.

Adenylate the 3'end of the DNA fragments

1. 20µl of End-repaired DNA + 20µl of dA-tailing Master Mix.
2. Incubate 30' at 37°C in a thermal cycler. Do not use heated lid.

Ligate the methylated adapters

3. transfer the dA-tailed fragments ON ICE and add 5µl of XT2 Ligation Master Mix.
4. Add 5µl of SureSelect Methyl-Seq methylated Adaptor and mix by pipetting.
5. Incubate in a thermal cycler for 15' (no more than 15') at 20°C. Do not use heated lid.
6. Purify the adaptor ligated DNA using 1.2 volumes of Agencourt® AMPure® XP beads. Elute in 22µl H_2O.
7. Quantitate the DNA using Qubit® fluorometer. To proceed it is necessary to have, at least, 500ng of adaptor-ligated DNA.

Hybridization

1. Dry the sample in order to have a total volume of 3.4µl. After drying, vortex to recover even the DNA on the walls of the tube, then spin for 10".
2. Prepare the Hybridization Buffer in a PCR tube and store it RT.

SureSelect Hyb#1	25µl
SureSelect Hyb#2	1µl
SureSelect Hyb#3	10µl
SureSelect Hyb#4	13µl
Total	49µl

   If a precipitate forms, warm the solution for 5' at 65°C.
3. Prepare the SureSelect Methyl-Seq Block Mix in a PCR tube and keep it on ice.

SureSelect Indexing Block#1	2.5µl
SureSelect Block#2	2.5µl
SureSelect Methyl-Seq Block#3	0.6µl
Total	5.6µl

4. Prepare the RNase block in a PCR tube and keep it on ice.

RNase block	0.5µl
H_2O	1.5µl
Total	2µl

5. In the RNase block tube add 5µl of SureSelect Human Methyl-Seq Capture Library. Mix well by pipetting.
6. In the SureSelect Methyl-Seq Block Mix tube add the 3.4µl of the library. Mix well by pipetting.
7. Put the SureSelect Methyl-Seq Block + library tube in a thermal cycler with the following program (heated lid set at 105°C).

95°C	5'
65°C	hold

   At the end of the program open the lid, but don't remove the tube: it must stay at 65°C.
8. Put the Hybridization Buffer tube in the thermal cycler at 65°C and incubate for 5'.
9. After the 5'-incubation put the RNase block + Capture Library tube at 65°C in the thermal cycler.
10. Incubate all of the tubes at 65°C in the thermal cycler for 2' and, even after these 2', do not remove tubes from the thermal cycler.
11. Transfer 13µl of Hybridization Buffer in the RNase block + Capture Library tube.
12. Transfer the 9µl of the library tube in the RNAse + Capture + Hybr tube and mix by pipetting 8-10 times. At this point the volume of the hybridization mixture is 27-29µl.

13. Seal the caps and incubate for 24h at 65°C in the thermal cycler with heated lid at 105°C.

Prepare the streptavidin beads

1. Set the water bath at 65°C and pre-warm the SureSelect Wash Buffer #2 at 65°C.
2. Vortex the Streptavidin beads (Dynabeads C1, Life technologies) and prepare 50µl of beads for each sample in a 1.5ml eppendorf tube.
3. Wash the beads:
4. add 200µl of SureSelect binding buffer;
5. vortex for 5" and spin;
6. put on a magnetic rack till the solution clarify;
7. remove and discard SN.
8. Repeat the washing step two other times (for a total of three washes).
9. Resuspend the beads in 200µl of SureSelect Binding Buffer.

Capture hybrids using streptavidin beads

1. At the end of the 24h-incubation, measure and register the remaining volume of the reaction.
2. Keep the tube at 65°C while adding the reaction volume directly from the thermal cycler to the streptavidin beads solution.
3. Incubate on a wheel rotator at RT for 30' (22rpm), then pulse-spin and magnetize. Remove the supernatant.
4. Resuspend the beads in 500µl of SureSelect Wash Buffer #1 by vortexing for 5" and incubate RT for 15', then magnetize and remove the supernatant.
5. Wash the beads:
    a. resuspend the beads in 500µl of SureSelect Wash Buffer #2 and vortex for 5";
    b. incubate 10' in the water bath at 65°C;
    c. spin and magnetize;
    d. remove the supernatant.
6. Repeat the washing step two other times, for a total of 3 washes. Make sure the entire wash buffer has been removed.
7. Mix the beads in 20µl of SureSelect Elution Buffer on a vortex mixer for 5" to resuspend the beads and incubate the sample for 20' RT.
8. Magnetize and transfer the SN to a new 1.5ml LoBind Eppendorf tube. Discard the beads.
9. Proceed immediately to bisulfite conversion.

Bisulfite conversion

EZ DNA Methylation-Lightning kit was used as previously described. Avoid the first denaturation step. Briefly:

1. Transfer 130µl of Lightning conversion reagent in a PCR tube.
2. Add 20µl of the eluted captured library and mix by pipetting.
3. Thermal cycler:

54°C	60’
4°C	hold

4. Follow manufacturer’s instructions for desulfonation and purification.
5. For the elution:
    a. put the column into a new 1.5ml eppendorf tube and let the tubes stand at RT for 2’ before proceeding;
    b. add 10µl of M-elution buffer directly to the column matrix and incubate RT for 3’;
    c. centrifuge 60” 13,000rcf;
    d. add other 10µl M-elution buffer to the column matrix and incubate RT for 3’;
    e. centrifuge 60” 13,000rcf.
6. The final volume is ~ 20µl.

PCR amplification of the bisulfite-treated libraries

1. Prepare the reaction mix:

H_2O	30µl
SureSelect 2X Taq2000	50µl
MethySeq PCR Primer F	1µl
MethySeq PCR Primer R	1µl
Total	82µl

2. Add 18µl of the converted DNA. Mix well by pipetting.
3. Thermal cycler:

95°C	2’	
95°C	30”	] x10
60°C	30”	
72°C	30”	
72°C	7’	
4°C	hold	

4. Purify the PCR product using 1.8 volumes of Agencourt ® AMPure® XP beads. Elute in 25µl of H_2O.

Sample indexing

1. In a 0.2ml PCR tube prepare the PCR mix (for 1 sample):

SureSelect Methyl-Seq 2X Taq2000 MM	25µl
SureSelect Methyl-Seq indexing primer common	0.5µl
PCR primer Index of choice	0.5µl
Amplified bisulfite-treated library	24µl
Total	50µl

2. Mix well by pipetting and incubate in a thermal cycler with the following program:

95°C 2’

95°C 30”
60°C 30” } x6
72°C 30”

72°C 7’
4°C hold

3. Purify the indexed libraries using 1.8 volumes of Agencourt® AMPure® XP beads. Elute in 24µl of H_2O
4. Quantitate the DNA using the Qubit® fluorometer.
5. Assess the quality of the library with the Agilent bioanalyzer.
6. Mix the sample in order to have 15µl of a 5nM sequencing pool.
7. Bridge PCR and Illumina® HiSeq1000 Sequencing reactions were performed by the sequencing service of the University of Verona (Prof. Delle Donne Group).

Ion Proton BS-seq

After Bisulfite conversion the DNA was treated as follows.

Second strand synthesis

1. Prepare the reaction mix:

Bisulfite-treated DNA	200ng in 10µl
10 mM dNTPs	1µl
10X NEBuffer 2	2µl
100µM Random Hexamer Primers	2µl
Nuclease-free Water	3µl
Total	18µl

2. Incubate the mix in a thermal cycler with the following program:

75°C 5’
5°C 10’ During this 10’ 5°C incubation add 2µl of E. coli DNAPolymerase I (NEB, 10U/µl)
25°C 10’
37°C 10’
75°C 20’
12°C ∞

At the end of the program the DNA is double stranded, blunt ended, but unphosphorylated.

3. Purify the dsDNA with 1.5 volumes of Agencourt® AMPure® XP beads. Elute in 20µl of Low TE.

4. Transfer the supernatant in a new 1.5ml LoBind tube without disturbing the pellet.

Library construction protocol

End-repair of the DNA

1. Prepare the end-repair mix in a 1.5ml LoBind tube:

dsDNA	20µl
H_2O	59µl
5X End-Repair Buffer	20µl
End Repair Enzyme Mix	1µl
Total	100µl

2. Mix well by pipetting and incubate RT for 20'.

Agencourt® AMPure® XP beads size selection

Two rounds with different concentrations of beads are used to size-select the DNA. In the first round DNA >250bp is captured on the beads, and DNA ≤250bp is retained in the supernatant. In the second round DNA >100bp is captured from the first-round supernatant. The size-selected DNA is then 100-250bp in length.

First round

1. Add H_2O to bring to 250µl the final volume of the DNA solution.
2. Add 225µl (0.9 volumes) of Agencourt® AMPure® XP beads to the sample, mix by pipetting, then pulse-spin and incubate at RT for 5'
3. Pulse-spin and place the sample in a magnetic rack for 3'
4. Carefully transfer the supernatant to a new 1.5ml LoBind tube. Do not disturb the pellet

Second round

1. Add 100µl (0.2 volumes) of Agencourt® AMPure® XP beads to the supernatant from the first round, pipet to mix, and proceed as a normal Agencourt® AMPure® XP beads DNA purification. Elute in 25µl of Low TE buffer.

Adapters ligation, nick-repair and ligated-DNA purification

1. Prepare the reaction mix in a 0.2ml PCR tube:

End-repaired DNA	25µl
10X Ligase Buffer	10µl
Proton adapters	2µl
10mM dNTP Mix	2µl
H_2O	51µl
DNA ligase	2µl
Nick-repair polymerase	8µl
Total	100µl

2. Incubate in a thermal cycler with the following program:

25°C	15’	
72°C	5’	
4°C	∞	This is not a stopping point, proceed as soon as possible.

3. Purify the sample using 1.4 volumes of Agencourt® AMPure® XP beads. Elute in 20µl of Low TE buffer.

Library amplification

1. Prepare the library amplification mixture as follows (recipe for 1 sample):

Platinum® PCR SuperMix High Fidelity	100µl
Library Amplification Primer Mix	5µl
ChIP DNA	25µl
Total	130µl

2. Split the 130µl-reaction volume in two 0.2ml PCR tube and run the following program on a thermal cycler:

95°C	5’	
95°C	15”	] x10 times
62°C	15”	
70°C	1’	
70°C	5’	
4°C	hold	

3. Collect the same sample in a unique 1.5ml LoBind tube.
4. Purify the library with 1.4 volumes of AMPure® XP beads and elute in 20µl of Low TE buffer.
5. Quantitate the DNA using the Qubit® 2.0 fluorometer and assess the quality of the library on the Agilent Bioanalyzer.

The sequencing core of Prof. Valle Group performed emulsion PCR, beads enrichment and Ion Proton Sequencing.

Bioinformatics analysis

In advanced laboratory methodologies (HT-NGS) the need for a new generation of bioinformatic tools is an essential prerequisite to accommodate further strategic development and improvement of the output results.

ChIP-seq analysis

The sequencing reads produced were imported in CLC Genomics Workbench 6.5, a commercial software for DNA sequence analysis, and mapped against the reference genome (hg19) with the build in aligner. The uniquely aligned reads are

then used as input for the “ChIP Seq” analysis module of CLC, using default parameters.
CLC allows to find regions that are enriched in the ChIP-sample to identify peaks of protein-DNA binding.
Once the enriched regions are established, a threshold or minimum peak height is calculated based on the false discovery rate. The goal is to determine whether a region of enrichment, or peak, is significant over background.
Experimental design, exhaustive antibody testing and careful sample preparation to minimize background are more critical to success of an experiment than the exact peak calling algorithm used for analysis.

BS-seq

After sequencing the bisulfite-treated DNA, the reads were aligned against the reference genome using PASS-bis software, which allow for mapping the reads tolerating the reduced complexity resulting from bisulfite treatment. Sequencing clonality has been removed from the mapping output using a PASS supplied module. Moreover, the set of tools of PASS-bis was even used to pair the paired-end reads obtained with Illumina Sequencing.
PASS-bis software (Campagna D et al., 2013) performed the methylation calling of CG, CHG and CHH locations along the genome or along selected regions. After this passage the MethylKit package (Akalin A, et al., 2012), written for the R CRAN statistics software, was used to discover the differentially methylated regions and sites of the Illumina sequenced samples.

Data integration

Data were annotated and integrated in pathways using DAVID 6.7 (Database for Annotation, Visualization and Integrated Discovery) webserver (Huang DW et al., Nature Protoc. 2009; Huang DW et al., Nucleic Acids Res. 2009), selecting the gene ontology and kegg pathways as data sources. Data were visualized in IGV (Robinson JT et al., 2011).
Graphite web, a public web server for the analysis and visualization of biological pathways using high-throughput gene expression data, has been used to highlight relevant pathways in gene expression data.

Results and Discussion of the preliminary protocols

Next Generation Sequencing represents a revolution in modern biology. Apart from genomics and transcriptomics, it is opening new opportunities also in the field of high-resolution genome-wide epigenetics. The continuous technology improvements and the concurrent cost reduction are now allowing in depth epigenomics investigations, thus unravelling chromatin state and rearrangements during cell life. The activation or repression of genes controlling human skeletal muscle differentiation requires an extensive crosstalk between transcription factors, DNA methylation and histone modifications. The exact sequence of chromatin events underlying myogenesis is far from being characterized, but a role for epigenetics both in preventing and maintaining myogenic terminal differentiation has been postulated.

To investigate the epigenetic landscape of human adult myogenesis I was particularly interested in two methodological approaches: chromatin immunoprecipitation and DNA bisulfite treatment followed by deep sequencing (ChIP-seq and BS-seq).

A difficulty in studying epigenetics comes from the novelty of most procedures and the consequent susceptibility of the methods to be very little reproducible; therefore, the experimental plan must be designed with great attention and the protocols must be carefully optimized on the specific experimental model.

No techniques for studying epigenetics were available in the laboratory at the beginning of my PhD study, so I had to critically evaluate existing methods to decide if any were suitable for my research. All the methods required an extensive work of optimization, but sometimes, since this field of research is very new, there were no methods available and I had to challenge myself in the development of new techniques.

The results I obtained during the first part of my PhD project are mainly negative results; nevertheless they were essential for the further development of my studies. Many technical problems made difficult the establishment of robust methods for ChIP-seq and BS-seq and this chapter summarizes step-by-step these developments.

Chromatin Immunoprecipitation (ChIP)

One of the best ways to take a comprehensive look at the epigenome is chromatin ChIP-seq, as it allows profiling the genome-wide distribution of a protein of interest. DNA-protein complexes can be isolated using highly specific antibodies, then the DNA fragments associated to the protein of interest are extracted and massively sequenced.

General guidelines for ChIP experiments recommend the optimization of the chromatin preparation procedure and the evaluation of antibody specificity; in fact, crosslinking and sonication conditions vary between cell types, and half of the ChIP-grade antibodies are not able to recognize their antigen in chromatin (Egelhofer TA *et al.*, 2011).
Preliminary ChIP tests were conducted either on crosslinked or native chromatin, and the results are presented in this section with a brief discussion.

Crosslinked ChIP (X-ChIP)

Crosslinked ChIP is suitable for mapping proteins that are both weakly and strongly associated to the DNA as the crosslinking minimizes chromatin rearrangements during the experiment. An overview of a general XChIP-seq experiment is reported in Figure 2. After shearing the formaldehyde-crosslinked chromatin in 300–1,000bp-long fragments, antibodies are used to enrich the proteins of interest by immunoprecipitation. Finally, the crosslinking is reverted and the DNA is isolated and sequenced.

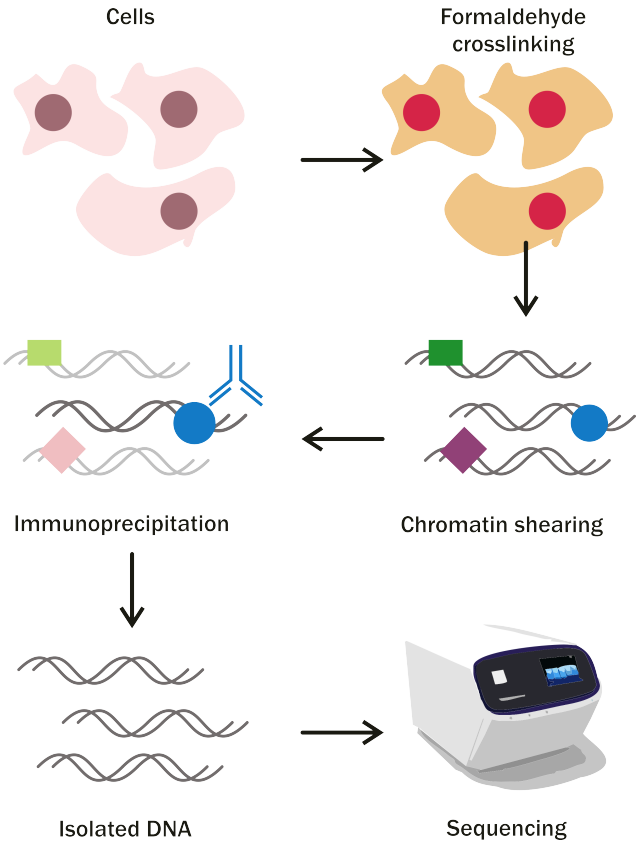

Figure 2| Standart outline of a crosslinked ChIP-seq experiment: cells are crosslinked with formaldehyde, which introduces covalent bonds between proteins and DNA. Chromatin is then fragmented in pieces suitable for immunoprecipitation and protein-DNA complexes of interest are isolated by the use of specific antibodies. DNA is then extracted and sequenced with next generation sequencing platforms.

Chromatin preparation: from mouse dissection to nuclei extraction.

XChIP optimal conditions can be very different between cell types: the perfect balance between crosslinking and sonication is a prerequisite to perform good and reproducible experiments.

Since different cell types can be more or less difficult to obtain, I firstly set up some preliminary steps on mouse splenocytes that were readily available from mice that had been sacrificed for other experiments. A further step of the optimization was done on C2C12 cells, a murine myoblast immortalized cell line that is easy to maintain in culture. Finally, the experiments were performed using CHQ5B cells, a primary myoblast cell culture derived from human muscle biopsies.

Mice splenocytes: crosslinking and shearing

Mice splenocytes were initially chosen to set-up the chromatin preparation method. An adult healthy mouse spleen contains about 100 million splenocytes that are adequate for 4-5 assays. Approximately 20million cells at a time were processed to set-up the chromatin preparation conditions and formaldehyde was chosen as crosslinking reagents due to the possibility of reversing crosslinks by heating the sample.

Several conditions were tested in order to optimize the formaldehyde crosslinking reaction; in fact, both excessive and insufficient crosslinking impair the experiment through altering sonication efficiency and epitope exposure and abundance in the chromatin sample.

Formaldehyde crosslinking, performed with 1% final concentration of formaldehyde in PBS, was optimized through a time-course experiment spanning from 5 to 30 minutes. In parallel, several sonication conditions were tested for obtaining the best compromise between crosslinking and sonication conditions. At the end, the definitive protocol for the preparation of chromatin from mice splenocytes chromatin preparation contemplate 20’ PBS crosslinking with formaldehyde at 1% of final concentration, and 10’ sonication in 16×65mm COVARIS glass tubes with maximum settings at 4-8°C. Sonication results of mice splenocytes are reported in Figure 3.

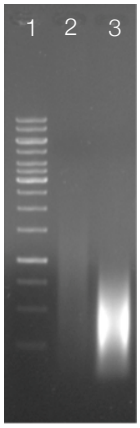

Figure 3| DNA from murine splenocytes after crosslinking and sonication. Lane 1: 1kb GeneRuler DNA ladder; Lane 2: DNA after crosslinking and 5’ of sonication; Lane 3: DNA after 10’ of sonication.

C2C12 crosslinking and shearing

C2C12 is a murine myoblast cell line that has the ability to differentiate in low serum conditions and is a model for studying muscle differentiation in mammals. Since it is easier to obtain large numbers of murine rather than human myoblasts I chose to test chromatin preparation conditions using the C2C12 murine immortalized muscle cells before moving to the human model. The conditions used for splenocytes were inadequate for the new model, system so I repeated the time-course experiments and found new conditions: crosslinking was performed for 12min in DMEM + 10% FBS and 10million cells were sonicated at a time, for 8min with maximum settings of the power (Figure 4).

Figure 4 | Agarose gel electrophoresis of C2C12 DNA after crosslinking and sonication: (*left*) myoblasts and (*right*) myotubes.

Left: first lane 1kb GeneRuler DNA ladder, second lane C2C12 Mb DNA after sonication.

Right: first lane LowRange DNA ladder, second lane C2C12 Mt DNA after 4' sonication, third lane C2C12 Mt DNA after 8'sonication.

Further experiments revealed that there were no significant differences in crosslinking cells in adhesion, or in suspension.

CHQ5B: nuclei extraction before crosslinking

Although C2C12 and CHQ5B are both similar in that they are myoblasts, chromatin preparation turned out to be significantly different. In fact, on applying to CHQ5B the crosslinking conditions used for C2C12, the chromatin appeared to be completely disrupted even before sonication (Figure 5). The attempts to improve this by modifying crosslinking time, buffer and temperature failed, the chromatin always appeared broken before sonication.

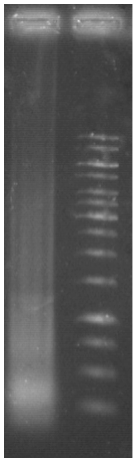

Figure 5 | Agarose gel electrophoresis of CHQ5B Mb DNA (first lane) after whole-cell crosslinking, but before sonication. Notably the chromatin appeared to be disrupted before shearing. 1kb GeneRuler molecular weight marker is loaded in the second lane.

We hypothesized that this problem might be caused by the release of DNase from cytosolic components during the cell lysis step. To try to overcome this problem we isolated the nuclei before crosslinking and we found that the problem was solved (Figure 6).

Chromatin preparation experiments were performed using 10 million cells at a time; nuclei were crosslinked with formaldehyde in PBS for 10min on ice + 5min at room temperature in PBS and sonicated for 8min at the maximum settings (Covaris sonicator).

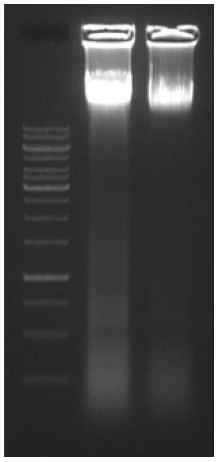

Figure 6 | Agarose gel electrophoresis of CHQ5B Mb (second lane) and Mt (third lane) DNA after nuclei extraction and crosslinking and before sonication. 1kb GeneRuler DNA ladder is loades on the first lane.

TBP and H3K9me3: immunoprecipitation and SOLiD™ sequencing

Sheared chromatin from 5X 10^6 nuclei of both CHQ5B myoblasts (Mb) and myotubes (Mt) was immunoprecipitated with antibodies (3µg) to either TBP or

H3K9me3. TBP (TATA-box Binding protein) specifically marks actively transcribed genes, whereas H3K9me3 is a marker of constitutive heterochromatin.
Each immunoprecipitation yielded about 20ng of DNA (Mb TBP: 25.06ng, Mb H3K9me3: 21.48ng, Mt TBP: 14.20ng, Mt H3K9me3: 25.54ng).
To test the specificity of the immunoprecipitation, an aliquot of the immunoprecipitated DNA was used for semi-quantitative PCR: accordingly to SOLiD™ protocol for ChIP-seq experiments, two primer pairs were chosen to quantitate and amplify an heterochromatic region (SAT2) and a constitutively expressed gene (FOS). With these antibodies we would expected to see enrichment in the SAT2 and FOS regions, respectively for the H3K9me3 and TBP immunoprecipitated sample.

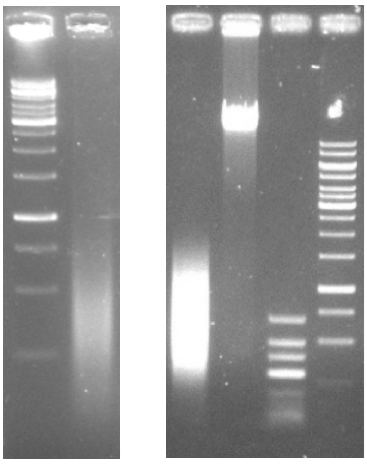

Figure 7 | Agarose gel electrophoresis of CHQ5B DNA after nuclei crosslinking and sonication.

Left: 1kb GeneRuler DNA ladder on the first lane and post-sonication CHQ5B Mb DNA on the second lane.

Right: post-sonication CHQ5B Mt DNA on the first lane, pre-sonication CHQ5B Mt DNA ladder on the second and 1kb Gene Ruler DNA marker on the third.

Figure 8 | Agarose gel electrophoresis of SAT2 semiquantitative PCR in Mb TB- and H3K9me3-immunoprecipitated samples. SAT2 band becomes visible in the H3K9me3 immunoprecipitated sample at the 32nd PCR cycle. On the external left and external right lanes 1kb GeneRuler DNA marker was loaded.

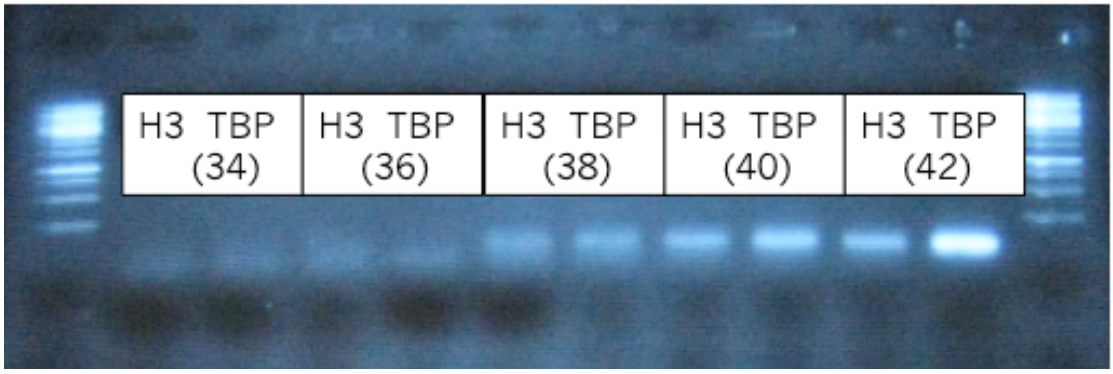

Figure 9 | Agarose gel electrophoresis of FOS semiquantitative PCR in Mb TB- and H3K9me3-immunoprecipitated samples. FOS band becomes visible both in H3K9me3- and TBP-immunoprecipitated sample at the 38th PCR cycle. On the external left and external right lanes 1kb GeneRuler DNA marker was loaded.

The PCR results were not satisfactory, as it can be seen in Figure 8 and Figure 9. Bands were detected only after a high number of PCR cycles and the enrichment was equivalent to two PCR cycles only. Nevertheless, considering both the limiting amount of template DNA and the limits of the semi-quantitative PCR, we decided to prepare the sequencing library to better understand the ChIP experiment.

SOLiD sequencing

DNA for sequencing was prepared according to the Mokry's protocol (Morky M et al., 2010), immunoprecipitated DNA was sonicated to produce fragments 150-250bp in length, which is compatible with SOLiD library preparation. Sequencing libraries were prepared following the SOLiD™ manual and sequenced in a lane of SOLiD™ 5500xl. As expected, the resulted coverage was low, but enough to make some considerations on the experimental results.

PASS software was used to align the reads on the reference genome and the results are reported (Table 2).

Table 2 – Technical data on the SOLiD sequencing of H3K9me3- and TBP- immunoprecipitated DNA in CHQ5B Mb and Mt.

Reads	Mb H3K9me3	Mb TBP	Mt H3K9me3	Mt H3K9me3
Total	21.675,623	37,037,894	24,862,470	23,218,274
Filtered-out	1,806,614	2,929,205	2,012,372	1,946,813
Aligned	17,506,469	29,360,242	20,208,011	19,357,009
Aligned/Filtered	81 %	79 %	81 %	83 %
Unique/Filtered	72 %	71 %	73 %	75 %

A quick look at the reads aligned in the genome browser confirmed the results of the semiquantitative PCR: there is a high background noise (Figure 10).

Figure 10 | Screenshot of the IGV view of ChIP-seq aligned reads. From the top to the bottom the samples are: Mt TBP, Mt H3K9me3, Mb TBP and Mb H3K9me3.

This was the first sequenced ChIP-seq experiment in the lab and, in order to learn how to manage the data, a test with CLC software was performed to search for enriched peaks. The results are reported in the Table 3.

Table 3 – Matrix of peaks in common among ChIP-seq experiments. For example, 15 peaks were found by CLC in the TBP-immunoprecipitated Mb sample; of these peaks 15 were present only in the Mb TBP sample, 4 were shared between Mb TBP and Mb H3K9me3, 6 were shared between Mb TBP and Mt H3K9me3 and 5 between Mb TBP and Mt TBP.

	Mb H3K9me3			
Mb H3K9me3	7 (0 unique peaks)	Mb TBP		
Mb TBP	4 overlapping peaks	22 (15)	Mt H3K9me3	
Mt H3K9me3	5	6	15 (5)	Mt TBP
Mt TBP	4	5	7	16 (7)

Normally, ChIP-seq experiments show thousands of enriched regions, instead this experiment detected only 7-22 peaks per sample. Even if it is known that low coverage can flatten the signal-to-noise ratio thus impairing peaks identification, in this case coverage cannot justify the results. Moreover, when considering the position of the few peaks in the genome, overlaps between euchromatic- and heterochromatic-peaks are observed: as reported in Table 3, TBP- and H3K9me3-immunoprecipitated samples share 50% of the peaks. In conclusion, immunoprecipitation specificity should be improved in future experiments.

Native ChIP (nChIP)

The absence of cross-linking makes native chromatin suitable for studying proteins that are tightly associated to the DNA, such as histones. An overview of the experiment is reported in

Figure 11.

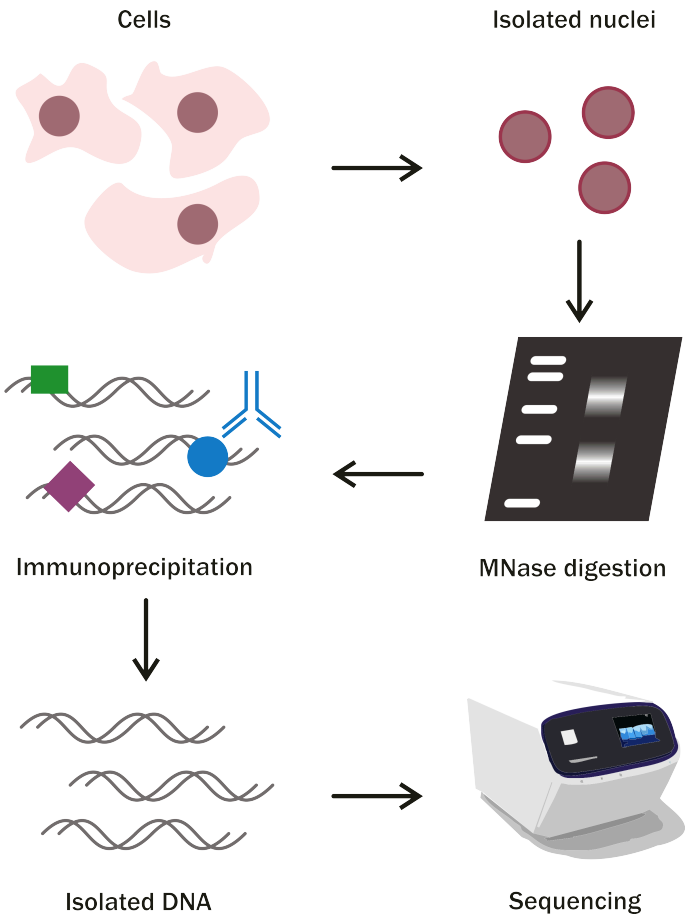

Figure 11 | native ChIP outline: after nuclei isolation chromatin is digested with micrococcal nuclase to produce fragments suitable for immunoprecipitation. After the immunoprecipitation of DNA-protein coplexes of interest, DNA is isolated and sequenced.

After micrococcal nuclease digestion of the nuclei, native ChIP was performed on $1x10^6$ CHQ5B myoblasts, using an antibody against H3K9me3 (2µg) and a normal mouse IgG (2µg) as a negative control. This experiment yielded 2ng of immunoprecipitated DNA for the H3K9me3 sample and 1ng for negative control.
Since no standard protocols for sequencing library preparations were available, I optimized an Ion PGM™ protocol for Ion Proton™ sequencing. The main difference between Ion PGM™ and Proton™ sequencing library preparation is the initial DNA fragment size. In the immunoprecipitated sample the DNA is a nucleosome-ladder of fragments 150-850bp in length. Micrococcal nuclease digested chromatin (50 ng) was used to set-up the sonication conditions in order to have a 100-150bp

smear, and 2ng of this sheared DNA were used to optimize the low-input Ion PGM™ protocol for Ion Proton™ sequencing (Figure 12).

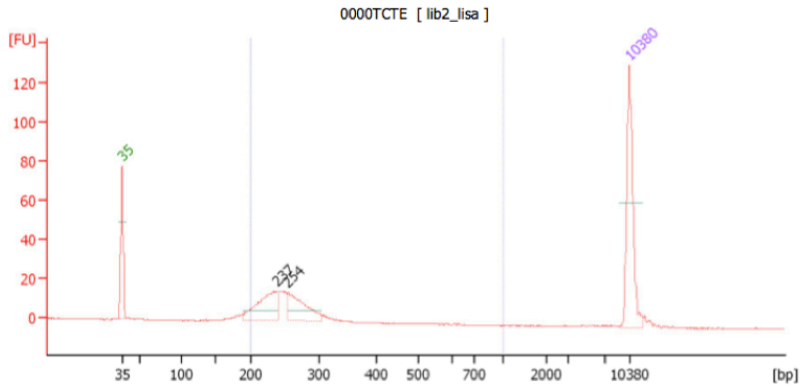

Figure 12| Agilent bioanalyzer trace of the Ion Proton sequencing library prepared from 2ng of starting DNA.

Nevertheless, the sequencing library preparation using the nChIP-DNA sample was critical and adapters were present in the library only at the end of the protocol (Figure 13).

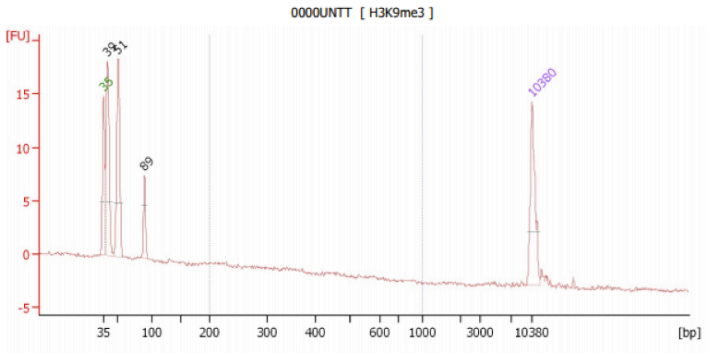

Figure 13 | Agilent Bioanalyzer trace of the Ion Proton sequencing library prepared from the 2ng of H3K9me3 immunoprecipitated DNA sample. Only adapters are visible as high peaks in the left part of the trace.

We found the problem to be in the sonication step: in fact, even if the immunoprecipitated DNA yields were reproducible between both the technical and biological replicas, we could not perform a proper nChIP-seq experiment because of the impossibility of efficiently shearing 2ng of DNA (Figure 14).

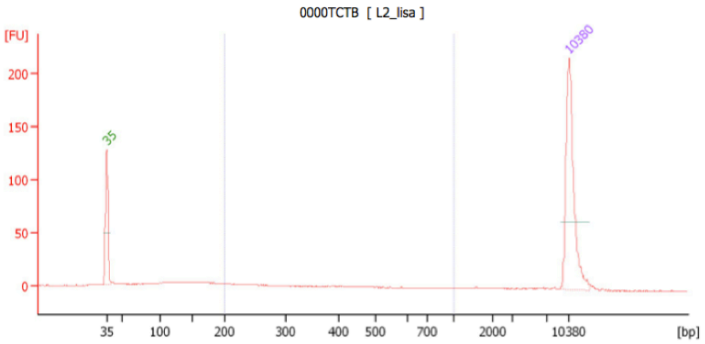

Figure 14| Agilent Bioanalyzer trace of 2ng of micrococcal digested DNA after sonication.

For crosslinked chromatin immunoprecipitation, our results suggest that a better evaluation of the immunoprecipitation conditions is highly recommended for future experiments. For what concern native ChIP, even if the immunoprecipitation protocol seems to work, in our hand it was impossible to construct sequencing libraries from such a small amount of initial DNA. Therefore we decided to test a commercially available kit, which claim to perform ChIP from 100,000 cells. The results of this definitive method will be presented and discussed in the "Results and discussion" chapter.

DNA methylation

DNA cytosine methylation in CpG dinucleotide context has a significant role in downregulating gene expression. One of the aims of this PhD project was to investigate the contribution of DNA methylation on differential gene-expression profiles observed between Mb and Mt. Transcriptomics analysis was performed in the laboratory on CHQ5B Mb and Mt by Rusha Guha, another PhD student, and the data are available in her PhD thesis.

To study DNA methylation we chose a BS-seq approach. Sodium bisulfite DNA treatment chemically converts the unmethylated cytosines to uracils, while leaving methylated cytosine unaltered; this treatment, together with next generation sequencing, offers single-base resolution of methylated sites. To be reliable, the conversion should be complete, so that every C sequenced was really methylated in the original sample.

Sodium bisulfite treatment is harsh since it can completely degrade the nucleic acid during the conversion step. We encountered this problem when the DNA used for the experiment was not sufficiently pure. Therefore the NanoDrop™ spectrophotometer was used to assess DNA quality and only samples with A260/230 ≥ 2 and A260/280 ≥ 1.8 were used for BS-seq protocol analyses.

A novel method for studying DNA methylation using the SOLiD™ 5500xl platform was conceived and developed in collaboration with Robin Targon, another PhD student in Prof. Valle group. The preliminary method is described in this section, the

definitive one in the “Results and discussion” section. In order to improve the time taken for BS-seq, Ion Proton™ platform was also tested for DNA methylation analysis.

Methylome SOLiD™ sequencing preliminary protocol (MeSS v1)

Next generation sequencing of the whole human methylome at single base resolution is expensive and the data analysis is tricky, due to the reduced complexity of bisulfite-converted genome. In fact, after bisulfite conversion, the two DNA original strands are no more complementary, and the same is for the reverse complementary of converted strands (Figure 15).

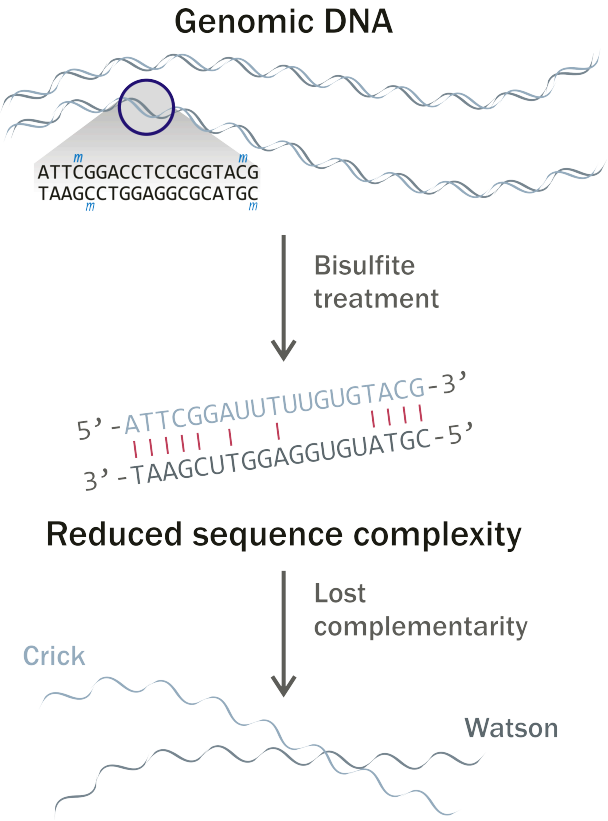

Figure 15 | Effects of sodium bisulfite DNA treatment. After bisulfite conversion the two strands of DNA are no more complementary and the complexity of the genome is highly reduced.

The alignment of sequencing reads on an in silico bisulfite-converted reference genome is a challenging bioinformatics task. To facilitate data analyses, researchers are now producing strand-specific BS-seq libraries by the use of methylated adapters, but the method is costly and time-consuming. In collaboration with Robin Targon we conceived a novel approach to produce strand-specific libraries for BS-seq on the SOLiD™ platform. The method should be easily adapted to every type of NGS platform. In this section I report the results of the preliminary protocol for methylome SOLiD™ sequencing (MeSS), while the results of the definitive protocol will be discussed in the Results chapter.

MeSS v1 was designed to be a fast and cheap single-tube protocol for constructing BS-seq strand-specific libraries; it was conceived to perform with 50ng initial DNA

and to avoid intermediate purifications and buffer exchanges. The method is based on the hybridization of protruding adapters suitable for SOLiD sequencing on bisulfite-converted, sheared and single-stranded DNA (Figure 16). These protruding adapters are designed to hybridize the DNA in a unique direction so that, after PCR library amplification, only the original bisulfite converted DNA strand would be sequenced. Only 10 cycles of PCR amplification were required to obtain the sequencing library, and this should have reduced read clonality, another advantage of using MeSS.

MeSS v1 protocol was applied to 50ng of CHQ5B Mb DNA, and the sample was initially sequenced in one lane of SOLiD™ 5500xl platform. The produced number of reads was slightly under the average output for SOLiD sequencing, with 110M reads of output instead of 120M. Data analysis was performed with the PASS-bis software and only 11M reads were assignable unambiguously to the human genome; of these unique reads, 7M were identical and mapped uniquely on the subtelomeric region of Chr10. We analysed this sequence, but we were not able to define its origin and we decided not to investigate further and proceed with the analysis of the remaining reads. On the 4M of unique and non-repeated reads mapping on the genome we assessed a percentage of directionality that was higher than 90%.

The high percentage of non-mapping reads suggested that further optimization of the MeSS technique was required, and we came out with MeSS v2 protocol, whose success and results will be discussed later in this thesis.

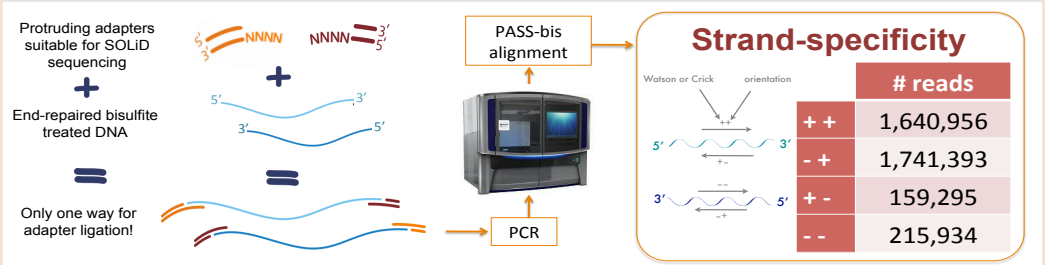

Figure 16| MeSS v1 workflow and results: the hybridization of protruding adapters on bisulfite treated DNA allow to construct strand-specific seqeuncing libraries.

Methylome Proton Sequencing

The sequencing rapidity of Ion Proton™ is unmatched and, in order to investigate its performance in methylation studies, we performed a BS-seq test without checking directionality (Figure 17). Double-stranded DNA was obtained by DNA polymerase I-extension of random hexamers on the bisulfite-treated single-stranded DNA and the sequencing library was prepared following standard protocol.

In our hands Ion Proton™ did not seem to work perfectly with BS-seq: sequencing output was 1.1Gb with an average read length of 35bp, standard Ion Proton™ output >10Gb with a 120bp the read-length.

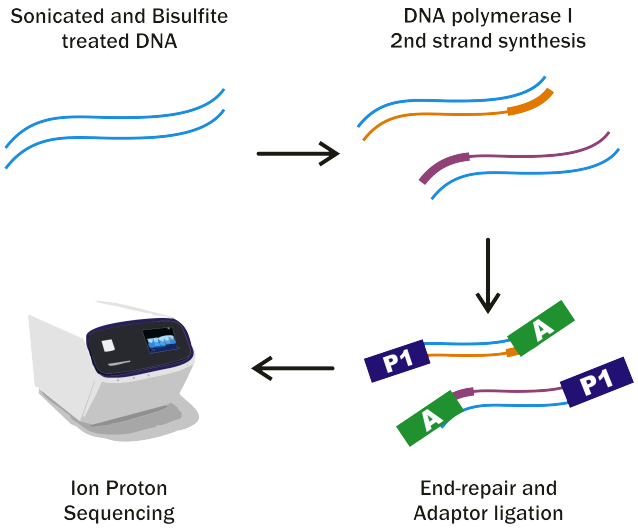

Figure 17 | Outline of the Ion Proton BS-seq protocol.

It is known that Ion-Proton does not perform very well on homopolymeric stretches. Therefore we hypothesize that the increased number of homopolymeric stretches derived from bisulfite treatment impaired the Ion Proton™ sequencing. Nevertheless we did not prove this hypothesis and developed the definitive protocol for Methylome SOLiD™ sequencing.

Results and discussion

Is DNA methylation involved in myogenesis?

MeSS v2 for whole-methylome: an expensive gold-standard!
Easy, innovative, exhaustive...but still too expensive

Sodium bisulfite treatment converts all unmethylated cytosines of the DNA into uracils, while leaving methylated C unaltered. This allows the genome-wide study of DNA methylation at single-base resolution.
At the beginning of this project only few methods were available for studying whole-genome DNA methylation, all being very expensive and requiring huge amount of starting DNA. In order to simplify bioinformatics analysis, the available techniques use methylated adapters to construct strand-specific sequencing libraries before sodium bisulfite DNA treatment. Sodium bisulfite typically degrades 80-90% of the starting DNA during conversion. Because of this degradation, the initial amount of DNA is 2-4µg, and many pre-amplification PCR cycles are required for sequencing. Moreover, when di-tagged DNA sequencing libraries are treated with sodium bisulfite, the vast majority of fragments will be broken, resulting in mono-tagged fragments that impair enrichment and sequencing results.
In order to optimize DNA methylation analysis, in collaboration with Robin Targon who is another PhD student, we developed a novel approach in which bisulfite conversion is performed before library preparation. This method, that we call MeSS v2, is rapid and cost-effective, and notably reduced the required starting amount of DNA. In particular, we used 300ng of NanoDrop-quantified DNA to successfully carry out the protocol that requires only one intermediate purification step. Last September, as we were analysing the first MeSS v2 results, Khanna A et al. (Epicentre® company - Nat. Methods, 2013) published a very similar method.

Technical details of MeSS v2 protocol

300ng of Nanodrop-quantified Mb DNA were sonicated and then bisulfite treated: after considering the degradation that follows bisulfite treatment we found the right DNA sonication conditions (Figure 18).

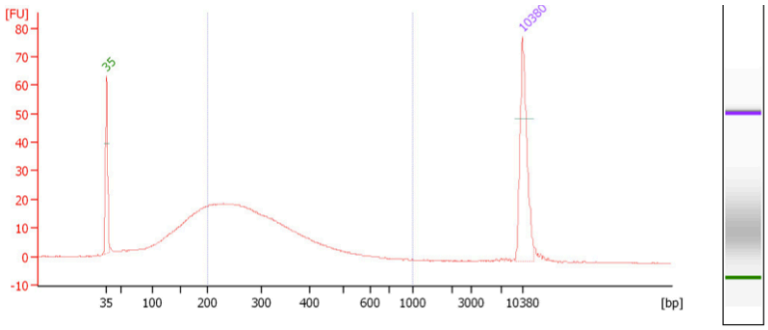

Figure 18 | Agilent profile of the sonicated and bisulfite-treated Mb DNA used for MeSS v2 protocol.

At first, the annealing of a 3'-blocked single-stranded adapter followed by a template-switching reaction produced the mono-tagged DNA fragment (with the sequencing tag at the 3' end of the DNA). Then, after the ligation of the second adapter at the 5' end, 15 PCR cycles produced the sequencing library for SOLiD 5500xl platform (Figure 19).

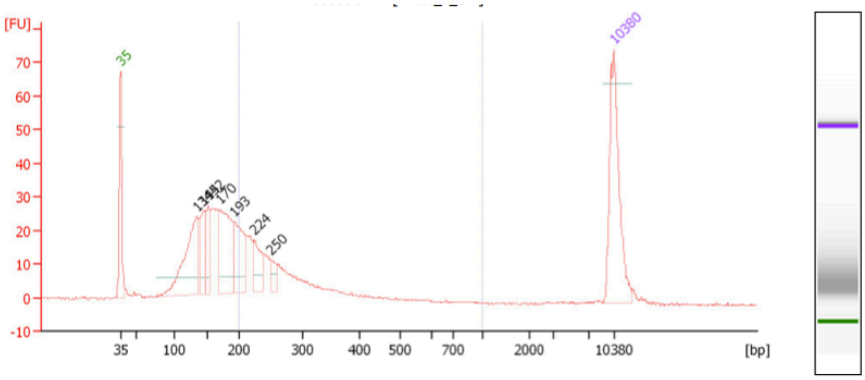

Figure 19 | MeSS v2 library constructed from 300ng of Mb DNA.

MeSS v2 protocol is summarized in Figure 20.

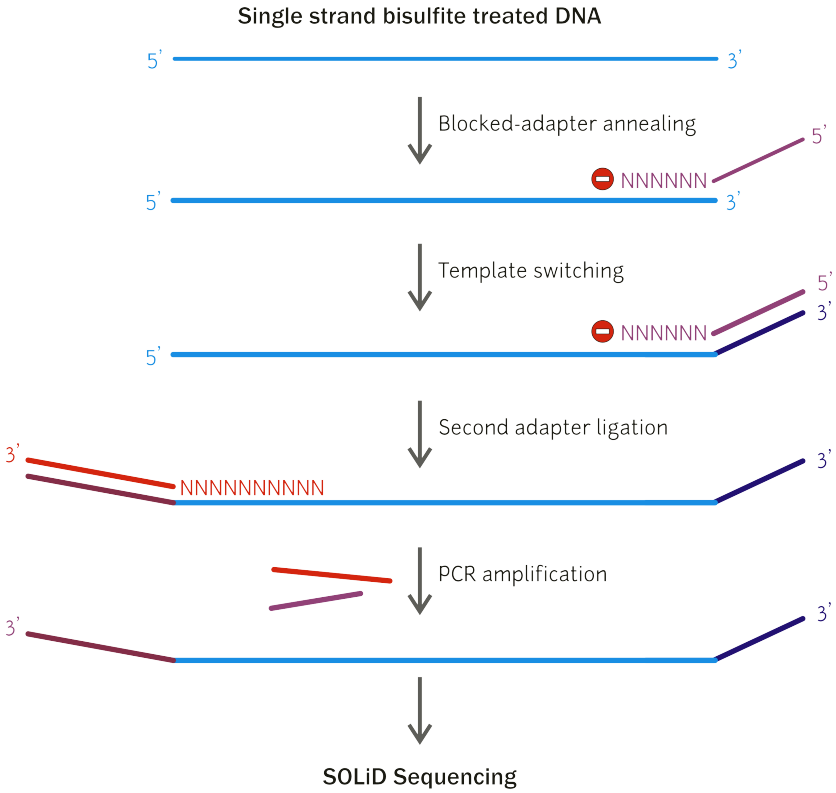

Figure 20| MeSS v2 outline

Data analysis was performed using the analysis pipeline based on Pass-bis software. In Table 4 the technical aspects of the sequencing trial lane are reported.

Table 4 – Technical aspects of MeSS v2 protocol

total number of reads	104M
mapping efficiency	60%
unique	> 80% of the mapped
directionality	> 90%

The whole-methylome analysis of one Mb sample required the sequencing of 12 SOLiD lanes, in order to obtain realistic methylation levels. Bioinformatics analyses are ongoing, but preliminary results obtained by comparing MeSS with SureSelect method (discussed below) show that the methylation levels are comparable in the shared regions. Further investigation is required for mapping all CG, CHG, CHH methylation positions along the genome, in order to evaluate the existence of methylation patterns, the presence of extra-CpG methylation in muscle, the role of CHG and CHH in myogenesis.
The method we propose allows the investigation of cytosine methylation at single base resolution in CG, CHG and CHH contexts, as other methods do, but with the advantage of saving DNA and time. The problem of all kinds of human whole-methylome analysis is the cost, which is around 10,000$ for obtaining 20X coverage of the genome, therefore we decided not to investigate the role of DNA methylation in adult myogenesis with the whole-genome approach. Although, the working protocol we conceived is now available in the laboratory and can be used, for example, for studying the methylome of smaller genomes, or as an alternative to RRBS after enzymatic digestion of DNA.
In conclusion, we developed a method for methylome studies that can compete with the recent published epicentre kit, with the advantage for MeSS to be publicly available and easily customizable for every sequencing platform, by simple changing the adapters sequences.

SureSelect Methyl-Seq Target enrichment
A valid alternative to whole-methylome BS-seq

One of the objectives of my research was to study the effects of methylation on the differentiation process in human muscle, from myoblasts to myotubes. To do this I tested an enrichment kit, the SureSelect Human MethylSeq target enrichment kit (Agilent). This kit is designed to selectively capture by hybridization 84Mbp of the human genome; in particular, it captures regions containing 3.7M CpG sites mainly in CpG islands, shores and shelves, and in gene promoters and enhancers. Since methylation of these regions impairs gene transcription, this method can be used to obtain preliminary, but highly informative data on the role of CpG methylation in

adult myogenesis. It also has the added advantage of being quicker and cheaper than performing whole genome sequencing. The DNA methylation data obtained with the kit could be integrated with the ChIP-seq and also with the transcriptomic data produced in the lab by Rusha Guha, another PhD student working in the laboratory.

The SureSelect Human MethylSeq kit was successfully used for the enrichment of 2 Mb 2 Mt samples before bisulfite conversion with the procedure illustrated in Figure 21. Briefly, methylated adapters are ligated to sonicated DNA before capturing CpG rich regions with biotinylated RNA probes. Then, captured DNA was isolated and bisulfite treated before PCR amplification. All the libraries were sequenced as paired-end on an Illumina HiSeq1000 platform and the data were analysed with Pass-bis software.

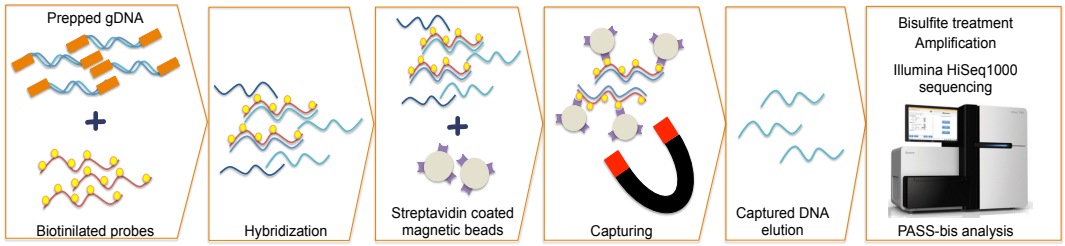

Figure 21 | Outline of SureSelect MethylSeq Target enrichment kit.

At the time we bought the kit there was no literature on it, therefore we had to validate its technical performance before analysing the sequencing data.

Kit performance: technical validation

To test the performance of the kit the percentage of in-target reads and of regions that are effectively captured, as well as the average coverage were calculated. Mapping efficiency was very high (Figure 22) and more than 85% of the reads were found to be in-target and more than 90% of the regions that were supposed to be captured were actually captured with an average coverage of at least, 20X.

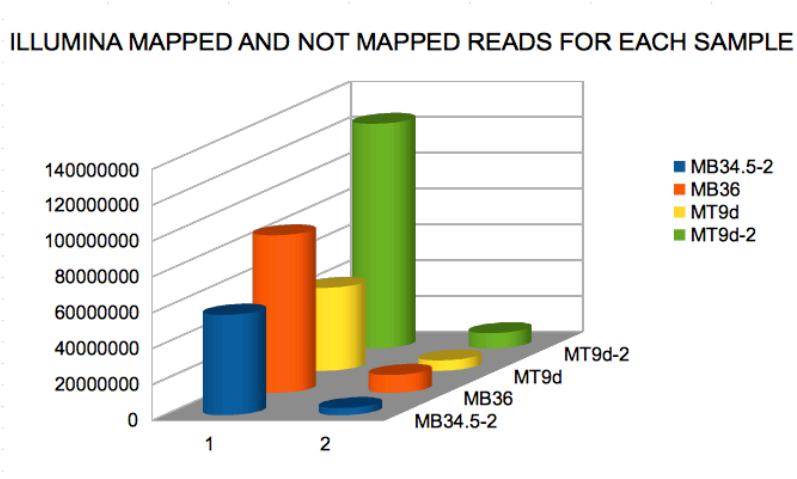

Figure 22 | Mapping efficiency of SureSelct sequenced reads: 1= mapped reads, 2= unmapped reads. The vast majority of reads has been mapped.

These data confirmed the SureSelect kit to be a valid alternative to RRBS: both the methods has the disadvantage of bisulfite treatment after library construction and necessitate a huge amount (µg) of starting DNA. Compared to RRBS the SureSelect kit had the advantage that no enzymatic digestion and gel excision were required. In the literature RRBS is cited (Stirzaker C *et al*., 2014) to capture less than 5000, 000 CpG sites, whereas the SureSelect kit method captured at least 2,700,000.

Differentially methylated regions and sites

After sequencing the four samples (2Mb and 2Mt) the Pass-bis software was then used to map the reads for methylation calling after pairing. Methylation data were used as input for an R methyl kit (Akalin A *et al*., 2012), which calculated both differentially, methylated sites (DMS) and regions (DMR).

There was very little difference in DNA methylation between Mb and Mt. In fact there are 577 DMS between Mb and Mt samples, located near 442 genes. In a recently published paper Tsumagari K et al. (Tsumagari K et al. 2013) concluded that DNA methylation only has a minor role in regulating adult myogenesis, being more a consequence than the cause of altered gene expression during the differentiation of Mb into Mt.

Nevertheless, we checked for DNA methylation changes in regions, instead of single sites, by adopting a 100bp window sliding every 10bp along the genome. R-methyl kit detected 1,917different DMR between Mb and Mt, near 1,404 genes. In particular 1,019 DMR were hypermethylated and 385 hypomethylated in Mt compared to Mb.

The comparison between DMR and differentially expressed gene (DEG) data are summarized in Table 5.

Table 5 – Classification of DMRs. The two rows (hypometylated and hypermethylated DMRs) have compared with gene expression profiling of Mt vs Mb cells.

DMR	no change in gene expr.	overexpressed in Mt compared to Mb	underexpressed in Mt compared to Mb	very low expression	ncRNA genes
Hypomethylated (385)	201	38	14	132	60
Hypermethylated (1,019)	397	77	80	465	204

After obtaining ChIP-seq results (see next paragraph) I observed that, among the DMR, which do not have a correspondently expressed gene, there is a high percentage of genes with H3K27me3, therefore I thought that hypermethylated regions may be already heterochromatic or that the process of heterochromatinization had already started before DNA methylation. From the ChIP-

seq results I hope to see if hypomethylation results in chromatin activation of miRNA.

Analysis of results

DMR, DMS and gene expression data were analysed both for the Kegg pathway and Gene ontology (Table 6 and Table 7).

Table 6 | DAVID Gene ontology: molecular function results with associated p-values

DEG (1074/2255 DAVID IDs)	DMR (572/118 DAVID IDs)	DMS (122/365 DAVID IDs)
cytoskeletal protein binding (2.6E-16)	sequence-specific DNA binding (1.9E-7)	transcription regulator activity (1.4E-4)
structural constituent of muscle (2.8E-15)	Ras guanyl-nucleotide exchange factor activity (8.7E-7)	transcription factor activity (1.7E-4)
actin binding (3.7E-12)	Rho guanyl-nucleotide exchange factor activity (6.2E-6)	sequence-specific DNA binding (3.9E-4)

Table 7 | Gene ontology: biological pathway results for the first five muscle-related pathways with associated p-value.

DEG (1408/2255 DAVID IDs)	DMR (587/118 DAVID IDs)	DMS (182/365 DAVID IDs)
muscle system process (5.4E-21)	embryonic skeletal muscle development (3.8E-7)	heart development (1.3E-3)
muscle contraction (8.7E-21)	skeletal system morphogenesis (2.1E-6)	heart morphogenesis (3.3E-3)
cytoskeleton organization (2.0E-20)	skeletal system development (2.1E-6)	regulation of skeletal muscle tissue development (1.7E-2)
muscle organ development (2.2E-18)	embryonic skeletal system morphogenesis (2.1E-5)	regulation of muscle development (1.9E-2)
muscle tissue development (4.2E-11)	heart development (4.2E-3)	cytoskeleton organization (2.9E-2)

DMR and DMS KEGG pathway results do not highlight any pathways specific for myogenesis, even if there are pathways for cell-cell junctions, for cancer and Wnt and MAPK signalling pathways.

In the KEGG pathways obtained from gene expression data, as expected the cell cycle and DNA replication appear to be inhibited during the differentiation of Mb to Mt, however muscle-related pathways are activated, as well as DNA repair mechanisms. Taken together, the gene ontology and Kegg pathways show the expression of muscle specific genes during differentiation, as structural components of the sarcomere and machinery for muscle contraction. Moreover although the cell cycle is inhibited during differentiation, the DNA repair mechanisms still remain active in the form of mismatch, base excision and nucleotide excision repair.

On the other hand, DNA methylation is linked to pathways that are more likely related to DNA-binding activities rather than muscle-specific genes. Even if gene ontology detected a significant number of terms associated with muscle

development, the significance is less than that of same terms in gene expression data.
Our results suggest that the involvement of DNA methylation during Mb differentiation to Mt does not show a direct, predominant role and, as Tsumagari (Tsumagari K et al, 2013) suggested, the demethylation observed between Mb-Mt and adult skeletal muscle seem to be more likely a consequence of altered gene expression, rather than the cause. Further investigation is required to clarify the role of DNA methylation during myogenesis, which is far from being completely understood. The presence of DMR in non-coding genes leads me to formulate a novel hypothesis on the epigenetic events during myogenesis.
Epigenetic mechanisms in addition to- or other than DNA methylation are certainly involved in activating muscle-specific genes and in repressing the cell cycle and, in particular, we hypothesise a role for histone modification complexes, DNA hydroxymethylation and regulatory RNAs. At first, chromatin compaction impairs gene activity by preventing the access of transcription factors to the promoters: MyoD itself recruits histone acetylases on muscle-specific gene promoters, thus enhancing the formation of euchromatin.
Moreover, we observed that many DMR map near non-coding RNA genes, so it would be interesting to investigate the link between methylation levels, gene expression and biological function of the variety of miRNA, miscRNA, lincRNA regions that emerge from our data.
Eventually, as recent literature suggests, we can hypothesise a role for DNA hydroxymethylation in activating methylated promoters: on the one side, if hydroxymethylation is involved in muscle differentiation, this could explain, at least in part, the small number of observed DMR. Supporting this hypothesis, we observed the expression of TET1 and the activation of DNA excision repair pathways, which can demethylate the DNA to the hypomethylated state described in adult skeletal muscle by Tsumagari K et al. (Tsumagari K et al., 2013).
In order to start unravelling the epigenetic mechanisms other than DNA methylation in myogenesis, I focused on chromatin state, thus setting-up chromatin immunoprecipitation. In conclusion I think it would be worthwhile to correlate histone modifications, DNA methylation and gene expression data.

ChIP-seq

ChIP-seq was performed on 100,000 Mb cells for each antibody and immunoprecipitated DNA was sequenced on the SOLiD 5500xl platform. For studying gene activity we chose 2 euchromatic markers (H3K4me3 and H3K9ac) and an heterochromatin marker (H3K27me3).
Trypsinized Mb were crosslinked in solution, and then sonicated to fragment chromatin in 100-250bp long DNA fragments; chromatin was then immunoprecipitated and the immunoprecipitated DNA was used for preparing a fragment sequencing library.

Read alignment and mapping

After sequencing, reads were aligned using Pass software, and only uniquely mapped reads were used for peak-calling and annotation using commercial CLC software for peak calling.

In particular, CLC analysis identified 2,201 peaks for H3K4me3 (assigned to 1,996 genes), and 1,608 peaks for H3K9ac (assigned to 1486 genes). An example is shown in Figure 23.

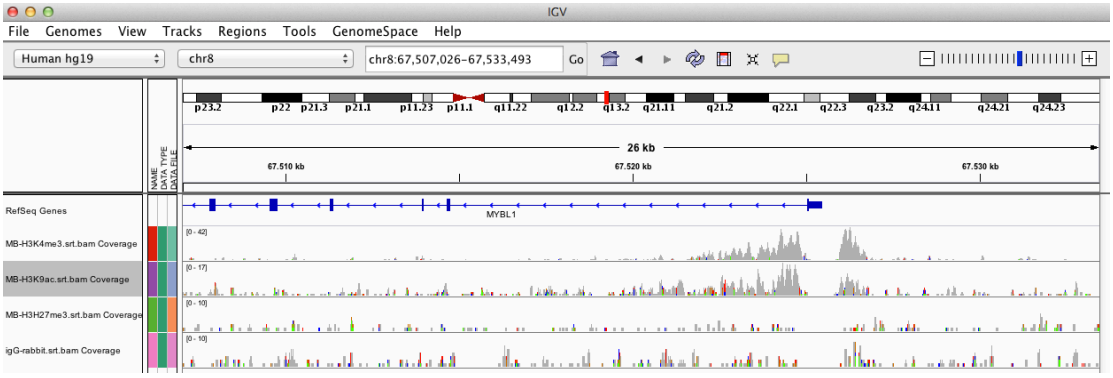

Figure 23| IGV view of the ChIP-seq profile of the transcription start site of MYBL1, an actively transcribed gene in Mb. From top to bottom, the four alignement tracks from ChIP-seq experiment are: H3K4me3 (euchromatin marker), H3K9ac (euchromatin marker), H3K27me3 (heterochromatin marker), and normal rabbit IgG (control).

We observed an impressive concordance between ChIP-seq and gene expression data and, in fact more than 80% of the peaks assigned to effectively transcribed genes. 14% of the peaks mapped in the TSS of novel non-coding RNAs (miRNA, lincRNA, miscRNA, that show very low expression in transcriptomic analysis). There are some possible explanations for this event: on the one hand, these RNAs might be expressed at very low level and the normalization of RNA-seq data masks their presence. On the other hand, since these non-coding genes are classified as “novel” they might be absent in the annotation system used for gene expression data. If the second hypothesis is true, the update of the annotation can solve the situation. Of course, the presence of H3K4me3 and H3K9ac on non-coding RNAs open a fascinating scenario on the epigenetic role of these transcripts in myogenesis, especially if considering the findings about DMRs in miRNA, miscRNA, snRNA and lincRNA genes.

Less than 2% of the euchromatin peaks co-localized with H3K27me3 and the presence of the heterochromatin marker can be, at least in part, the reason why the genes are very poorly expressed. An example is repored in Figure 24.

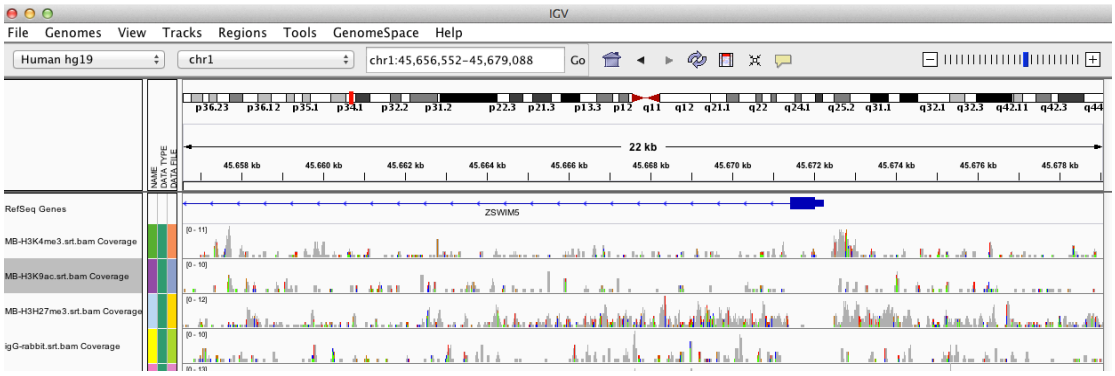

Figure 24 | IGV view of the ChIP-seq profile of the transcription start site of ZSWIM5, a gene non transcribed in Mb. From top to bottom, the four alignement tracks from ChIP-seq experiment are: H3K4me3 (euchromatin marker), H3K9ac (euchromatin marker), H3K27me3 (heterochromatin marker), and normal rabbit IgG (control).

Less than 2% of euchromatin peaks are probably mis-assigned: in fact, these genes are very close to other genes, which are expressed.

Finally, about 2% of the peaks show both the euchromatin markers, but they are not transcribed.

After correlating ChIP-seq data with gene expression and DNA methylation data in Mb I am now producing the ChIP-seq data for Mt in order to define the existence of chromatin dynamics in controlling myogenesis.

Conclusions

The activation or repression of genes that control adult myogenesis requires an extensive crosstalk between transcription factors and epigenetics players, such as DNA methylation and histone modifications. The final goal of my PhD project is the description of the epigenetic mechanisms underlying human muscle differentiation.
When I started this PhD project, epigenetics was still an innovative field and I had to deal with the lack of tools for studying epigenetic mechanisms. Therefore, as a first task of my project, I set-up tools for investigating DNA methylation and protein-DNA interactions at high resolution, by the use of next generation sequencing technologies. In particular, I set-up BS-seq and ChIP-seq protocols.

The gold standard technique for DNA methylation studies is BS-seq (bisulfite treatment of DNA followed by deep sequencing), in which bisulfite converts all the unmethylated cytosines to uracils and leaves methylated ones unaltered, thus permitting single-base resolution. Next generation sequencing of the whole human methylome at single base resolution is expensive and the data analysis is challenging, due to the reduced complexity of bisulfite-converted genome. We developed a novel method, that we called MeSS (Methylome SOLiD Sequencing), to investigate whole-genome methylation. The approach and performance of this new method are similar to the ones recently proposed by the commercially available Epicentre® kit (Nat. Methods, 2013). The approach we propose relies on post-bisulfite library construction that avoids the degradation of the library during the conversion step and increases sequencing efficiency. Moreover, MeSS protocol permits the construction of strand-specific libraries from only 300ng of starting DNA. We used MeSS to produce the whole methylome of one myoblast sample.
Since we were interested in evaluating the involvement of DNA methylation in myogenesis, the whole-methylome analysis of myoblast and myotube samples was challenging and demanding. Therefore we opted for analysing only a subset of genomic regions using the Agilent SureSelect Methyl Seq Enrichment kit, which claims to capture by hybridization 3.7M CpG sites in gene promoters, CpG islands, shelves and shores. Since there were not published results obtained with this SureSelect kit, before analysing the data we evaluated the performance of the kit and we found that more than 2.7M CpG sites are effectively captured with a minimum coverage of 10X. Moreover we calculated that more than 85% of the reads were in-target and, among the regions that should be captured by the kit, more than 90% have an average coverage of at least 5X.
Since the positive results of the performance evaluation of the SureSelect kit, we successfully used it to determine the differentially methylated sites (DMS) and regions (DMR) between myoblasts and myotubes. 577 DMS and 1917 DMR were found between the two conditions and more than half of these sites did not show correlation with gene expression. Moreover, the significativity of muscle-related

terms in Gene Ontology is lower than the one emerging from RNA-seq data. Therefore we hypothesized that the DNA methylation role in myogenesis is not predominant, at least in the activation of muscle-specific genes during differentiation. However, many differentially methylated regions or sites were associated to genes coding for regulatory RNAs and, as a future perspective, a correlation between DNA methylation and expression level of corresponding regulatory RNAs should be further investigated.

Since chromatin compaction has an important role in regulating the accessibility of transcriptional machinery to the DNA, and we are interested in describing the epigenetic landscape of myogenesis, ChIP-seq was set-up in parallel with BS-seq.
I focused on ChIP-seq experiments to define the localization of H3K9ac, H3K4me3 and H3K27me3 histone modifications on DNA. A homemade method for crosslinked-ChIP (X-ChIP) has been widely tested and immunoprecipitation was performed both on myoblast and myotubes. Analysis of the sequencing results suggested that a careful evaluation of the immunoprecipitation conditions is highly recommended.
Native ChIP was also tested, and, in order to prepare sequencing libraries from less than 2ng of starting DNA, I adapted an Ion PGM protocol to the Ion Proton platform, but the low recovery of immunoprecipitated DNA (2ng) and the need for sonicating this DNA made the nChIP library preparation very problematic.
At the end, the True MicroChIP kit from Diagenode was used for immunoprecipitation experiments on myoblasts using two euchromatin markers (H3K9ac and H3K4me3) and an heterochromatin one (H3K27me3). By comparing ChIP-seq with RNA-seq data, we observed that more than 80% of H3K9ac and H3K4ac peaks are localized in the transcription start site of actively transcribed genes and H3K27me3 was found in transcriptionally inactive regions, as expected. An interesting finding was the localization of euchromatin-associated peaks in the TSS of novel non-coding RNAs. In the immediate future we are planning to perform the ChIP-seq of myotubes for mapping the same histone modifications. This will allow us to have an exhaustive view of the chromatin state during myogenic differentiation.
As I began the integration of BS-seq, ChIP-seq and gene expression data I noticed that both in Kegg pathway and Gene Ontology, excision repair mechanism and TET1 are active during differentiation. Considering that Tsumagari K et al. (2013) found high levels of cytosine hydroxymethylation in the promoter of muscle specific genes, the hypothesis of a role of hydroxymethylation in myogenesis would be one of our future priorities to investigate.

References

1. Alamdari N, Aversa Z, Castillero E, Hasselgren PO: **Acetylation and deacetylation – novel factors in muscle wasting**. Metabolism 2013; **62**: 1-11
2. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, *et al*.: **methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles**. Genome Biology 2012; **13**:R87.
3. Baar K: **Epigenetic control of skeletal muscle fibre type**. Acta Physiol. 2010; **199**: 477-87
4. Barreiro E, Sznaider JI: **Epigenetic regulation of muscle phenotype and adaptation: a potential role in COPD muscle dysfunction**. J. Appl. Physiol. 2013; **114**: 1263-72
5. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, *et al.*: **High-resolution profiling of histone methylations in the human genome**. Cell 2007; **129**: 823-37
6. Beck S, Rakyan VK: **The methylome: approaches for global DNA methylation profiling**. Trends Genet. 2008; **24**: 231-7
7. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, *et al.*: **Accurate whole human genome sequencing using reversible chain terminator chemistry**. Nature 2008; **456**: 53-9
8. Bentzinger CF, Wang YX, Rudnicki MA: **Building muscle: molecular regulation of myogenesis**. Cold Spring Harb. Perspect. Biol. 2012; **4**: a008342
9. Berkes CS, Tapscott SJ: **MyoD and the transcriptional control of myogenesis**. Semin. Cell. Dev. Biol. 2005; **16**: 585-95
10. Bernstein BE, Meissner A, Lander ES: **The mammalian epigenome**. Cell 2007; **128**: 669-81
11. Bestor TH: **The DNA methyltransferases of mammals**. Hum. Mol. Genet. 2000; **9**: 2395-402
12. Bird A: **DNA methylation patterns and epigenetic memory**. Genes Dev. 2002; **16**: 6-21
13. Bird AP: **CpG-rich islands and the function of DNA methylation**. Nature 1986; **321**: 226-32
14. Bjornsson HT, Sigurdsson MI, Fallin MD, *et al*.: **Intra-individual change over time in DNA methylation with familial clustering**. JAMA 2008: **299**: 2877-83
15. Blau HM, Pavlath GK, Hardeman EC, Chiu CP, *et al*.: **Plasticity of the differentiated state**. Science 1985; **230**: 758-66
16. Bock C, Tomazou EM, Brinkman AB, Müller F, *et al*.: **Quantitative comparison of genome-wide DNA methylation mapping technology**. Nat. Biotechnol. 2010; **28**: 1106-14

17. Brand-Saberi B, Bodo C: **Genetic and epigenetic control of muscle development in vertebrates**. Cell Tissue Res. 1999; **296**: 199-212
18. Campagna D, Albiero A, Bilardi A, Caniato E, *et al.*: **PASS: a program to align short sequences**. Bioinformatics 2009; **25**: 967-8
19. Campagna D, Telatin A, Forcato C, *et al*.: **PASS-bis: a bisulfite aligner suitable for whole methylome analysis of Illumina and SOLiD reads**. Bioinformatics 2013; **29**:268-70
20. Cheung V, Allis CD, Sassone-Corsi P: **Signaling to chromatin through histone modifications**. Cell 2000; **103**: 263-71
21. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, *et al.*: **Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning**. Nature 2008; **452**: 215-9
22. Egelhofer TA. Minoda A, Klugman S, Lee K, Kolasinska-Zwierz P, *et al.*: **An assessment of histone-modification antibody quality**. Nat. Struct. Mol. Biol. 2011; **18**: 91-3
23. Feinberg AP: **Phenotypic plasticity and the epigenetics of human disease**. Nature 2007; **447**: 433-40
24. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, *et al.*: **A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands**. Proc. Natl. Acad. Sci. USA 1992; **89**: 1827-31
25. Fuchs SM, Krajewski K, Baker RW, Miller VL, Strahl BD: **Influence of combinatorial histone modifications on antibody and effector protein recognition**. Curr. Biol. 2011; **21**: 53-8
26. Furey TS: **ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions**. Nat. Rev. Genet. 2012; **13**: 840-52
27. Gibney ER, Nolan CM: **Epigenetics and gene expression**. Heredity 2010; **105**: 4-13
28. Gu H, Bock C, Mikkelsen TS, Jäger N, Smith ZD, *et al.*: **Genome-scale DNA methylation mapping of clinical samples at single base resolution**. Nat. Methods 2010; **38**: 133-6
29. Guttman M, Amit I, Garber M, French C, Lin MF, *et al.*: **Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals**. Nature 2009; **458**: 223-7
30. Hemberger M, Dean W, Reik W: **Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal**. Nat. Rev. Mol. Cell Biol. 2009; **10**: 526-37
31. Henikoff S: **Nucleosome destabilization in the epigenetic regulation of gene expression**. Nat. Rev. Genet. 2008; **9**: 15-26
32. Huang DW, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources**. Nature Protoc. 2009; **4**:44-57

33. Huang DW, Sherman BT, Lempicki RA. **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists**. Nucleic Acids Res. 2009; **37**:1-13
34. Jaenish R, Bird A: **Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals**. Nat. Genet. 2003; **33** (**Suppl**): 245-54
35. Johnson DS, Mortazavi A, Myers RM, Wold B: **Genome-wide mapping of in vivo protein-DNA interactions**. Science 2007; **316**: 1497-502
36. Khanna A, Czyz A, Syed F: **EpiGnome™ Methyl-Seq Kit: a novel post–bisulfite conversion library prep method for methylation analysis**. Nat. Methods 2013; **10**
37. Kharchenko PV, Tolsturukov MY, Park PJ: **Design and analysis of ChIP-seq experiments for DNA-binding proteins**. Nat. Biotechnol. 2008; **26**: 1351-59
38. Klose RJ, Bird AP: **Genomic DNA methylation: the mark and its mediators**. Trends Biochem. Sci. 2006; **31**: 89-97
39. Kouzarides T: **Chromatin modifications and their function**. Cell 2007; **128**: 721-33
40. Laird PW: **Principles and challenges of genomewide DNA methylation analysis**. Nat. Rev. Genet. 2010; **11**: 191-203
41. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, *et al.*: **Initial sequencing and analysis of the human genome**. Nature 2001; **409**: 860-921
42. Li B, Carey M, Workman JL: **The role of chromatin during transcription**. Cell 2007; **128**: 707-19
43. Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, *et al.*: **Role of transposable elements in heterochromatin and epigenetic control**. Nature 2004; **430**: 471-6
44. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, *et al.*: **Highly integrated single-base resolution maps of the epigenome in Arabidopsis**. Cell 2008; **133**: 523-36
45. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, *et al.*: **Human DNA methylomes at base resolution show widespread epigenomic differences**. Nature 2009; **462**: 315-22
46. Maxam AM, Gilbert W: **A new method for sequencing DNA**. Proc. Natl. Acad. Sci. USA 1977; **74**: 560-4
47. McKinsey TA, Zhang CL, Olson EN: **Signalling chromatin to make muscle**. Curr. Opin. Cell. Biol. 2002; **14**: 763-72
48. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, *et al.*: **Genome-scale DNA methylation maps of pluripotent and differentiated cells**. Nature 2008; **454**: 766-70

49. Mikkelsen TS, Ku M, Jaffe DB, Isaac B, Lieberman E, *et al.*: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells**. Nature 2007; **448**: 553-60
50. Metzker ML: **Sequencing technologies – the next generation**. Nat. Rev. Genet. 2010; **11**: 31-46
51. Mokry M, Hatzis P, de Bruijn E, *et al.*: **Efficient Double Fragmentation ChIP-seq Provides Nucleotide Resolution Protein-DNA Binding Profiles**. PLoS One 2010; **5**: e15092
52. Murrel A, Rakyan VK, Beck S: **From genome to epigenome**. Hum. Mol. Genet. 2005; **14** (**Spec. No 1**): R3-R10
53. O'Neill LP, Turner BM: **Immunoprecipitation of native chromatin: NChIP**. Methods 2003; **31**: 76-82
54. Perdiguero E, Sousa-Victor P, Ballestar E, Munoz-Canoves P: **Epigenetic regulation of myogenesis**. Epigenetics 2009; **8**: 541-50
55. Rhee I, Bachman KE, Park BH, Jair KW, Yen RW, *et al.*: **DNMT1 and DNMT3b cooperate to silence genes in human cancer cells**. Nature 2002; **416**: 552-6
56. Rando OJ, Chang HY: **Genome-wide views of chromatin structure**. Annu. Rev. Biochem. 2009; **78**: 245-71
57. Ribel-Madsen R, Fraga MF, Jacobsen S, Bork-Jensen J, *et al*.: **Genome-Wide Analysis of DNA Methylation Differences in Muscle and Fat from Monozygotic Twins Discordant for Type 2 Diabetes**. PLoS One, 2012; **12**: e52302
58. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, *et al*.: **Integrative Genomics Viewer**. Nature Biotechnol. 2011; **29**: 24-6
59. Sanford JR, Wang X, Mort M, Vanduyn N, *et al*.: **Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts**. Genome Res. 2009; **19**: 381-94
60. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors**. Proc. Natl. Acad. Sci. USA 1977; **74**: 5463-7
61. Sharma S, Kelly TK, Jones PA: **Epigenetics in cancer**. Carcinogenesis 2010; **31**: 27-36
62. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, *et al.*: **Accurate multiplex polony sequencing of an evolved bacterial genome**. Science 2005; **309**: 1728-32
63. Stirzaker C, Taberlay PC, Statham AL, and Clark SJ: **Mining cancer methylomes: prospects and challenges**. Trends Genet. 2014; **30**: 75-84
64. The ENCODE project consortium: **An integrated encyclopedia of DNA elements in the human genome**. Nature 2012; **489**: 57-74

65. The International Human Genome Sequencing Consortium: **Finishing the euchromatic sequence of the human genome**. Nature 2004; **431**: 931-45
66. Tsumagari K, Baribault C, Terragni J, Varley KE, Gertz J, *et al.*: **Early de novo DNA methylation and prolonged demethylation in the muscle lineage**. Epigenetics 2013; **8**: 317-32
67. Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, *et al.*: **Genome-wide analysis of transcription factor binding sites based on ChIP-seq data**. Nat. Methods 2008; **5**: 829-34
68. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, *et al.*: **The sequence of the human genome**. Science 2001; **291**: 1304-51
69. Visel A, Blow MJ, Li Z, Zhang T, *et al*.: **ChIP-seq accurately predicts tissue-specific activity of enhancers**. Nature 2009; **457**: 854-8