

PhD degree in Molecular Medicine (curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples “Federico II”

Settore disciplinare: BIO/10

## **ChroP approach dissects the dynamic profiling of chromatome at enhancers of inflammatory genes**

*Gianluca Sigismondo*

IEO, Milan

Matricola n. R09407

*Supervisor:* PhD, Tiziana Bonaldi

IEO, Milan

*Internal Advisor:* PhD, Diego Pasini

IEO, Milan

*External Advisor:* Prof. Matthias Selbach

Max-Delbrück Center for Molecular Medicine, Berlin

Anno accademico 2013-2014

## Table of contents

### *Figure and Table index*

<i>Acknowledgements</i> .....	iii
<b>LIST OF ABBREVIATIONS</b> .....	v
<b>1. ABSTRACT</b> .....	1
<b>2. INTRODUCTION</b> .....	3
2.1 Epigenetics, chromatin composition and hPTMs.....	3
2.1.1 Technical approaches to study hPTMs.....	6
2.2 Mass Spectrometry analysis and MS based-proteomics.....	7
2.2.1 Protein identification from MS data.....	14
2.2.2 Biochemical enrichment strategies and mass spectrometry-based identification of post translational modifications.....	15
2.2.3 Analytical implementation to boost hPTMs identification .....	17
2.3 Relative quantitation strategies in MS-based proteomics.....	19
2.3.1 SILAC-based approaches for the relative protein quantitation.....	22
2.4 hPTM readers' screening through qProteomics approaches.....	25
2.5 Enhancers' elements and transcriptional regulation as prototype for dynamically regulated biological process.....	27
<b>3. AIM OF THE WORK</b> .....	35
<b>4. MATERIAL AND METHODS</b> .....	35
4.1 General biochemistry and molecular biology buffers.....	37
4.2 Cell Culture, SILAC and triple-SILAC labeling of cells.....	40
4.3 Bone marrow-derived macrophages in vitro differentiation.....	41
4.4 Cell treatment with lipopolysaccharides.....	42
4.5 Cross-linking Chromatin immunoprecipitation (X-ChroP).....	42
4.6 Native Chromatin immunoprecipitation (N-ChroP).....	43
4.6.1 Native chromatin immunoprecipitation (N-ChIP) followed by MNase digestion.....	43
4.6.2 Cross-linked chromatin immunoprecipitation (X-ChIP) followed by MNase digestion and sonication to use a transcription factor as bait in ChroP experiment.....	45
4.7 Protein co-immunoprecipitation from nuclear soluble extract.....	46
4.8 In-gel chemical alkylation and digestion of histones prior to MS analysis.....	47
4.9 Standard protein in-gel digestion for MS-based protein identification.....	48

4.10 Nano Liquid Chromatography and Tandem Mass Spectrometry .....	49
4.11 Analysis of Mass Spectrometric data.....	50
4.11.1 Identification and quantitation of hPTMs.....	50
4.11.2 Visualization of hPTMs.....	52
4.12 Analysis of SILAC data for interactome and dynamic study.....	52
4.13 Quantitative RT-PCR of immunopurified DNA.....	54
4.14 RNA extraction and cDNA synthesis.....	54
4.15 Immunoblot analysis.....	55
4.16 Retroviral vector oligonucleotide cloning.....	55
4.16.1 Retroviral Constructs and Production of Retroviruses.....	55
4.16.2 Retroviral Transduction of Bone Marrow-Derived Macrophages.....	56
<b>5. RESULTS.....</b>	<b>57</b>
5.1 Employment of X-ChroP approach in RAW 264.7 cells.....	57
5.1.1 SILAC labeling of RAW 264.7 cells.....	58
5.1.2 Optimization of chromatin cross-linking conditions for the characterization of inflammatory genes' enhancers.....	60
5.1.3 Optimization of the antibody/peptide ratio and of the peptide competition in the H3K4me1 X-ChroP.....	63
5.1.4 Setup of Pu.1 X-ChroP.....	69
5.2 Dissection of enhancers' determinants in resting macrophages by X-ChroP.....	72
5.2.1 Characterization of enhancers' composition by the intersection of H3K4me1 and Pu.1 interactomes.....	75
5.3 Profiling chromatin determinants of enhancers during inflammatory stimulus.....	81
5.3.1 Triple SILAC labeling, for time-course X-ChroP analysis.....	81
5.3.2 Follow-up validation of Mpeg1 protein as a novel enhancer determinant.....	89
5.4 Dissection of hPTMs associated with enhancers of inflammatory genes in macrophages.....	95
5.4.1 N-ChroP strategy to characterize the H3K4me1- modifcome.....	95
5.4.2 Setup and initial assessment of a new strategy combining formaldehyde cross-linking with MNase digestion to analyse the modifcome of Pu.1- chromatin.....	104
<b>5. Discussion.....</b>	<b>109</b>
<b>References .....</b>	<b>117</b>
<b>Appendix .....</b>	<b>125</b>

## Figure and table index

Fig.1: Histone post-translational modifications.....	4
Fig.2: Histone writers, readers and eraser of hPTMs important in transcription regulation	5
Fig.3: Determinants of chromatin landscapes.....	6
Fig.4. Classical workflow for bottom-up MS analysis.....	8
Fig.5: Schematic representation of sample ionization in MALDI ion source.....	10
Fig.6: Schematization of electrospray-ionization process.....	11
Fig.7: Scheme of Q Exactive geometry.....	14
Fig.8: Protein identification by MS/MS database searching.....	15
Fig.9: Schematic representation of distinct relative quantitation strategies.....	22
Fig.10: Schematic view of SILAC labeling and of a prototypic experimental design.....	23
Fig.11: Regulatory chromosomal looping .....	29
Fig.12: Dynamics of chromatin determinants at enhancers.....	31
Fig.13: Hallmarks of enhancers of ES and macrophages cells at different developmental stages.....	32
Fig.14: Experimental design of the SILAC-based X-ChroP strategy.....	58
Fig.15: Establishment of the SILAC labeling in RAW 264.7 cells.....	59
Fig.16: Balance between formaldehyde cross-linking and DNA fragmentation to reach the optimal DNA length.....	61
Fig.17: Optimization of cross-linking percentages for H3K4me1 X-ChroP experiment....	62
Fig.18: Optimization of antibody/input ratio and peptide fold excess for H3K4me1 X-ChroP experiment.....	64
Fig.19: Readout of H3K4me1 X-ChroP SILAC experiment.....	67
Fig.20: qPCR and WB validation of peptide competition in H3K4me1 X-ChroP experiment.....	68
Fig.21: Optimization of the X-ChroP conditions using Pu.1 as bait.....	70
Fig.22: qPCR and WB validation of peptide competition in Pu.1 X-ChroP experiment....	71
Fig.23: Dissection of the H3K4me1 interactome.....	73
Fig.24: Dissection of the Pu.1 interactome.....	74
Fig.25: Dissection of the basal enhancers' interactome.....	76
Fig.26: Protein complexes significantly enriched at enhancers' of resting macrophages...	78
Fig.27: Enrichment of players of the enhancer/promoter looping in the basal enhancers' interactome.....	80
Fig.28: Schematic representation of triple-SILAC labeling combined with X-ChroP	81
Fig.29: Incorporation rate of RAW264.7 cells in heavy and medium SILAC media.....	82
Fig.30: Analysis of the inputs from tc-ChroP experiments.....	84
Fig.31: Analysis of tc-X-ChroP forward 2 and reverse 2 experiments upon LPS stimulus.	86

Fig.32: Dynamic profiling of enhancer' determinants during LPS stimulus and proteins newly recruited at these cis-regulatory regions.....	88
Fig.33: Characterization of Mpeg1 localization and dissection of chromatin-independent Pu.1 interaction.....	92
Fig.34: Functional characterization of Mpeg1 depletion in macrophages.....	94
Fig.35: The N-ChroP strategy for the characterization of the modifecome of H3K4me1-chromatin.....	96
Fig.36: Intensity-based quantification of peptide 3 – 8 of histone H3 in H3K4me1 N-ChroP.....	98
Fig.37: Relative enrichment of hPTMs in H3K4me1 CHIP.....	101
Fig.38: Histone PTMs within H3K4me1-regions profiled at basal conditions and upon inflammation.....	103
Fig.39: X-ChroP/MNase combined strategy for the characterization of the modifecome of Pu.1-mononucleosomes.....	106
Appendix 1 Protein identified through the intersection of H3K4me1 and Pu.1 basal interactomes.....	125
Appendix 2 Dynamic profiling of enhancers' determinants in tc-ChroP.....	133
Appendix 3 Proteins newly recruited at enhancers in tc-ChroP.....	141
Appendix 4hPTMs identified and quantified at enhancers.....	147

## **LIST OF ABBREVIATIONS**

<b>Ac</b>	Acetylation
<b>ACN</b>	Acetonitrile
<b>APS</b>	Ammonium persulfate
<b>AUC</b>	Area under the curve
<b>BMDMs</b>	Bone marrow-derived macrophages
<b>ChIP</b>	Chromatin immunoprecipitation
<b>ChIP-Seq</b>	ChIP-Sequencing
<b>ChroP</b>	Chromatin Proteomics
<b>CID</b>	Collision induced dissociation
<b>DDA</b>	Data-dependent acquisition
<b>di-Ac</b>	Di-acetylation
<b>DMSO</b>	Dimethyl sulfoxide
<b>eRNA</b>	Enhancer RNA
<b>ESI</b>	Electrospray ionization
<b>ETD</b>	Electron transfer dissociation
<b>FDR</b>	False discovery rate
<b>Fwd</b>	Forward experiment
<b>H3K4me1</b>	Histone H3 lysine 4 mono-methylation
<b>H3K4me3</b>	Histone H3 lysine 4 tri-methylation
<b>H3K9me3</b>	Histone H3 lysine 9 tri-methylation
<b>HCD</b>	Higher energy collisional dissociation
<b>HPLC</b>	High pressure liquid chromatography
<b>hPTM</b>	Histone post-translational modification
<b>ICAT</b>	Isotope coded affinity tag
<b>iTRAQ</b>	Isobaric tag for relative and absolute quantification
<b>LC</b>	Liquid chromatography
<b>LC MS/MS</b>	Liquid chromatography and tandem MS
<b>LCR</b>	Locus control region
<b>LDS</b>	Lithium dodecyl sulfate
<b>LPS</b>	Lypopolysaccharides
<b>LTQ</b>	Linear trap quadrupole
<b>m/z</b>	Mass-to-charge ratio
<b>MALDI</b>	Matrix assisted laser desorption ionization
<b>me1</b>	Mono-methylation
<b>me2</b>	Di-methylation
<b>me3</b>	Tri-methylation
<b>MNase</b>	Micrococcal nuclease S7

*List of abbreviations*

---

<b>MS</b>	Mass spectrometry
<b>MS/MS</b>	Tandem mass spectrometry
<b>N-ChIP</b>	Native chromatin immunoprecipitation
<b>NDR o NFR</b>	Nucleosome-depleted or nucleosome-free region
<b>RA</b>	Relative abundance
<b>RE</b>	Relative enrichment
<b>Rev</b>	Reverse experiment
<b>RT</b>	Room temperature
<b>SCX</b>	Strong cation exchange
<b>SDS</b>	Sodium dodecyl sulfate
<b>SDS-PAGE</b>	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
<b>SILAC</b>	Stable isotope labeling by amino acids in cell culture
<b>TEMED</b>	Tetramethylethylenediamine
<b>tetra-Ac</b>	Tetra-acetylation
<b>TF</b>	Transcription factor
<b>tri-Ac</b>	Tri-acetylation
<b>TSS</b>	Transcription start site
<b>UHPLC</b>	Ultra performance liquid chromatography
<b>WB</b>	Western blot
<b>X-ChIP</b>	Crosslinking chromatin immunoprecipitation
<b>XIC</b>	Extracted ion chromatogram

## **ABSTRACT**

In macrophages, the inflammatory stimulus drives a fine rearrangement of cell-specific chromatin regulatory regions to orchestrate the adaptive immune response and activate inflammatory genes transcription. The fine-tuning of gene expression in time and space is controlled by the synergistic cooperation among different chromatin determinants, thus suggesting a strong link between the composition of these chromatin regions and their function. While promoter regions are well characterized, the enhancers of macrophage are largely unexplored; indeed the only known markers are the enrichment of mono-methylation on lysine 4 of histone H3 (H3K4me1) and the binding of the master regulator Pu.1.

The aim of this thesis is therefore the global characterization of the enhancers' determinants and their dynamic profiling during the inflammatory response. We addressed this issue by adopting the ChroP approach recently established in our laboratory; this strategy combines chromatin immunoprecipitation (ChIP) and mass spectrometry-based proteomics for the comprehensive dissection of histone post-translational modifications (modificome) and chromatin-binding proteins (interactome), associated with a specific chromatin region.

ChroP approach was employed in RAW 264.7 macrophage-like cell line and optimized to be amenable to the immunoprecipitation not only of H3K4me1 but also of the transcription factor Pu.1 used as bait in ChIP. In order to profile the dynamics of the enhancers' determinants during the early phases of the inflammatory response, we performed ChroP experiments in time-course (tc-ChoP), in cells untreated or treated for 1h or 4hrs with lipopolysaccharide (LPS). We characterized three different scenarios: proteins stably associated to enhancers during the inflammatory response and proteins evicted or recruited to these regulatory regions upon the stimulus.



Our findings suggest a general model where at macrophage enhancers a high-ordered structure pre-exists to the stimulus and consists of both positive and negative transcriptional regulators in association with hPTM patterns typically marking a more open chromatin state. During the inflammatory response, the general high-ordered structure of enhancers is overall maintained but we determined a dynamic behaviors of a set of proteins which are either recruited (as the PBAF complex) or evicted (like Dnmt1) from enhancers and synergize with the enrichment (hyper-acetylation of histone tails) or depletion (H3K9me3) of specific histone marks, to set an environment permissive to transcriptional activation.

Interestingly, some proteins are newly recruited at enhancers after the inflammatory stimulus to fine-tune the appropriate gene expression profile; among these we identified some proteins already known to be involved in the inflammatory response, such as Junb and Stat1, while others were completely unexpected, thus very interesting to be further investigated.

Finally, among the enhancers' determinants that show a stable enrichment we focused our attention on Mpeg1 protein, which has been described as macrophage-specific but whose functional role is almost unexplored.

ChroP strategy allowed the dissection of enhancers' determinants in basal as well as upon massive transcriptional activation; through this approach, we were able to corroborate already known findings and dynamics but -more interestingly- we could identify novel potential markers for enhancer, which will be further investigated through ChIP-seq. The combination of these two ChIP-based strategies indeed allows a comprehensive characterization of the same chromatin region from two different, but complementary, perspectives.

## **INTRODUCTION**

In this introductory section, I will begin with the description of the general concept of epigenetic regulation mediated by chromatin, with a focus on histone post-translational modifications (hPTMs) and their interplay with other chromatin determinants.

I will then introduce the two main approaches to study hPTMs, with more focus on mass spectrometry (MS) methods and MS-based proteomics, which are the strategies that I employed extensively in my study.

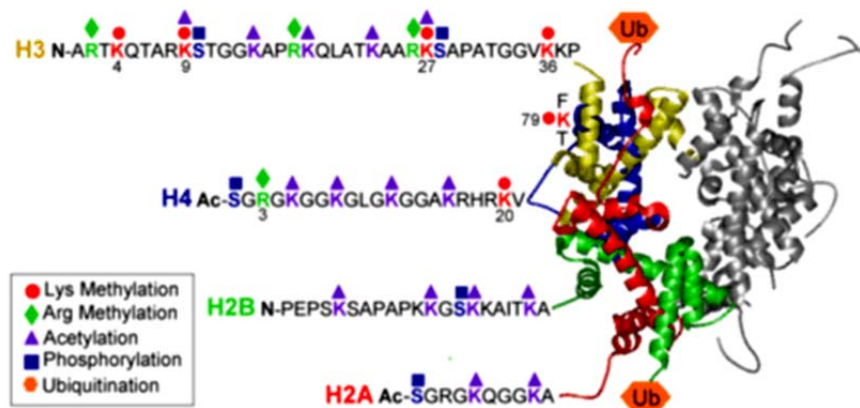
I will then describe the most common MS-approaches to characterize the protein determinants of chromatin, such as hPTMs and chromatin interactors and modifiers, by means of the combination of biochemistry and MS-based proteomics strategies.

Finally, I will conclude with a short description of macrophages' enhancers of inflammatory genes as prototypic example of genomic regulatory regions where an in-depth dissection of its chromatin determinants is crucial to understand the molecular mechanism of their function.

### **2.1 Epigenetics, chromatin composition and hPTMs**

The concept of epigenetics was introduced by Waddington in 1942 (Waddington 1942) to define the molecular pathways that modulate the expression of the same genotype to produce different phenotypes. Literally, epigenetics means “above” or “on top” of genetics and in the last years the meaning of this word has gradually changed toward a definition that refers to all the inheritable but reversible phenomena that modulate gene expression without altering the DNA sequence (Wu et al. 2001). A univocal definition of epigenetic is still under debate but it is generally accepted to consider as epigenetic marks many chromatin determinants like DNA methylation, modifications on histone proteins, microRNA (miRNA) and chromatin remodeling complexes, even though heritability has not been demonstrated for all of them (Tammen et al. 2013). 2013). All these epigenetic

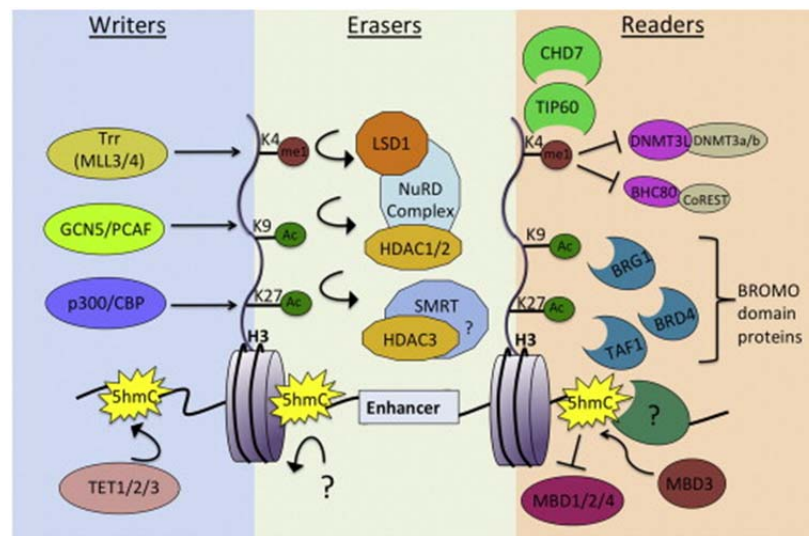
determinants impinge on and modulate chromatin which is composed at the structural level of nucleosomes: repeated units of 147 bp of DNA wrapped around the octamer core constituted by one H3-H4 tetramer and two H2A-H2B dimers (Kornberg 1974). Chromatin is a dynamic and highly structured nucleoprotein complex that in eukaryotes mediates both the DNA compaction into the nucleus and the regulation of different nuclear functions such as transcription, replication and DNA repair. In this thesis we will focus on the proteinaceous components of chromatin, namely transcription factors, chromatin-binding proteins, histones and their post-translational modifications. Part of the epigenetic modulation mediated by reversible hPTMs, which mainly occur at the protruding N-terminal regions of histones (the so-called “histone tails”) and are recognized as key regulators of chromatin structure and function. Distinct enzymatic families add a large panel of different covalent PTMs on histones: acetylation on lysines (K-Ac), methylation on arginines or lysines (R or K-me), phosphorylation on serine, threonine and tyrosine (S, T or Y-ph), ubiquitination and sumoylation on lysines (K-ub and K-su) (Kouzarides 2007) and lysine crotonylation (K-cr) (Tan, Luo et al. 2011) (Fig. 1).



**Fig. 1: Histone post-translational modifications.** Nucleosome and core histone N-terminal tails with annotation of sites of post-translational modifications. The most common types of hPTMs are listed in the box with the legend and indicated by colored symbols. Monomers of histones are represented: H3 (yellow), H4 (blue), H2A (red) and H2B (green); the other half of the nucleosome is in grey. (Adapted from Briggs’s personal page <https://ag.purdue.edu/biochem/Pages/Profile.aspx?strAlias=sdbriggs&intDirDeptID=9>)

The repertoire, location and combination of these hPTMs are extremely variegated and the cell uses combinations of such modifications to code for different gene expression patterns

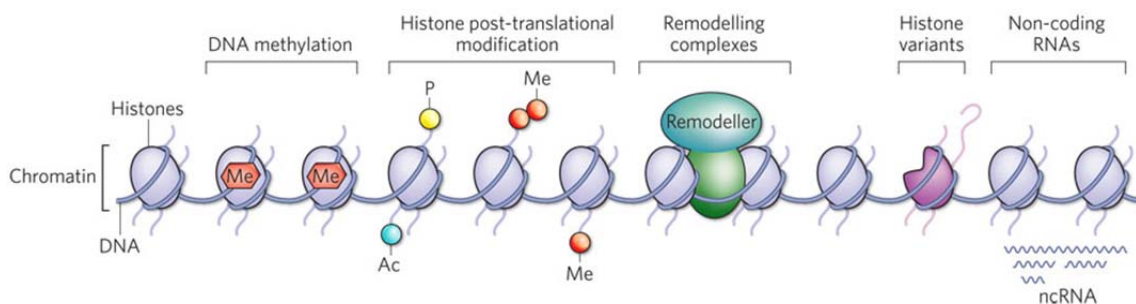
and thus different functional effects. The concept of the specific functional outcome associated with a distinct modification pattern has led to the elaboration of the “histone code hypothesis” that suggests that hPTMs, singly or in combination, can define a sort of molecular language that modulates the functional state of the underlying DNA. Indeed, hPTMs are placed and removed by enzymes known as “writers” and “erasers”, respectively (Fig. 2).



**Fig. 2: Histone writers, readers and eraser of hPTMs important in transcription regulation.** Example of enzymes able to add (writers), remove (erasers) and bind (readers) to hPTMs to determine the functional role of these modifications. (Adapted from Calo and Wysocka, 2013)

Post-translational modifications may exert their function through two distinct mechanisms: either the chromatin compaction is altered *via* changes in inter-nucleosomal or histone-DNA interactions, thus controlling the accessibility of DNA-binding proteins such as transcription factors or co-regulators (*cis* mechanism). Alternatively, hPTMs can generate binding platforms for the recruitment of macromolecular complexes, named “readers”, containing specialized domains (*trans* mechanism) able to “read” these marks and translate them into a specific process on the DNA (Nightingale, Gendreizig et al. 2007; Turner 2007). Alteration of chromatin architecture directly affects gene expression and other cellular processes; as a consequence, aberrant hPTMs patterns have been observed in several diseases, among them cancer. For this reason, decrypting this “code” would boost

the comprehension of physiological and pathophysiological conditions based on chromatin. In addition to hPTMs, chromatin is also characterised by local enrichment of histone variants, DNA methylation, chromatin binding proteins and differential nucleosome density and positioning; all these features synergize together to create a specific “chromatin landscape” and modulate gene expression at specific loci (Margueron and Reinberg 2010) (Fig. 3).



**Fig. 3: Determinants of chromatin landscapes.** (Adapted from Dulac C. 2010)

### 2.1.1 Technical approaches to study hPTMs

During the past years, two main approaches have been successfully employed to characterise chromatin composition and architecture: assays based on antibodies specifically raised against hPTMs or proteins of interest, and mass spectrometry. Antibodies are traditionally used to study the hPTMs in different assays; they can be employed to analyse modifications at a single cell level (i.e. immunofluorescence, IF), or for the profiling of PTMs in different samples and/or conditions (western blot, WB). Antibodies can also be used in chromatin immunoprecipitation (ChIP) to map a specific modification, or a protein, at genome-wide level and with the resolution of a few nucleosomes (Barski, Cuddapah et al. 2007). ChIP is nowadays combined with massive DNA sequencing (ChIP-seq), which enables to observe that hPTMs are not uniformly distributed along the genome but tend to cluster within defined chromatin regions in distinct patterns.

Despite their high specificity, antibodies have some limitations: they cannot determine the

modification status of different histones within the same nucleosome thus antibodies fall short in studying the combinatorial view of the code (Garcia, Shabanowitz et al. 2007); moreover, antibody binding may be reduced by steric effect of modifications occurring on adjacent or closely spaced residues, thus causing epitope-masking.

In this context, MS has proved to be an effective alternative to characterise known modification but also to identify novel and unpredicted hPTMs. Furthermore, thanks to the development of high resolution MS-instruments, it allows distinguishing between nearly isobaric modifications and mapping very complex combinatorial patterns of hPTMs, thus overcoming antibodies limitations (Beck 2010). Finally, the use of chemical and metabolic labeling strategies has enabled the accurate quantification of modifications in a relative and absolute manner. However, the major limitation of MS lies on the fact that it has been employed mainly for the analysis of bulk chromatin, which provides only a global view of the modification state of a system, with no locus-specific information. An important implementation would therefore consist in the characterization of hPTMs pattern specifically associated with defined chromatin landscapes.

## 2.2 Mass Spectrometry analysis and MS based-proteomics

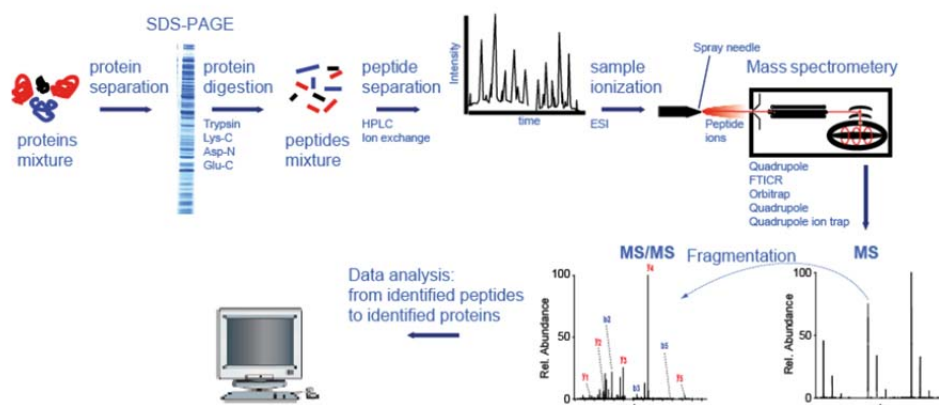
In the last decade, technological improvements generated mass spectrometers characterized by constantly higher resolution, sensitivity and performance. Nowadays, the most widely used approach for proteomic analysis is the liquid-chromatography coupled with tandem mass-spectrometry (LC-MS/MS).

The classical workflow for the identification and characterization of complex mixture of proteins is the so-called “bottom-up” approach, a peptide-centric strategy where proteins are eventually separated and then digested with proteases into peptides. The most common protease used in bottom-up approach is trypsin, which cleaves at the C-terminus of lysine and arginine residues unless they are followed by proline residue. The large use of this

enzyme in bottom-up is due to the fact that it generates peptides about 10-12 amino acids long, which are then easily detectable by MS.

In order to decrease the complexity of the mixture, peptides are separated by liquid chromatography typically in reversed phase mode (RP-LC), according to their hydrophobicity. Moreover, in RP-LC the polar and volatile mobile phase favors the ion transferring in the mass analyzer which measures the mass-to-charge ratio ( $m/z$ ) of freely moving gas-phase ions in electric and/or magnetic fields. In the mass spectrometer precursor peptides are isolated and scanned (full scan MS or survey scan MS or MS or MS1) to identify the  $m/z$  that provides information on the elemental composition of the peptide. The most intense of these precursors are then fragmented in the constituent fragment ions through tandem MS events (MS/MS or MS<sub>n</sub>) to retrieve their amino acid sequence.

Peptide identification is achieved by comparing the tandem mass spectra derived from peptide fragmentation with theoretical MS/MS spectra generated from *in silico* digestion of a protein database. Protein identification is then inferred by assigning peptide sequences to proteins (Fig. 4).



**Fig. 4: Classical workflow for bottom-up MS analysis.** A protein mixture from a biological sample is separated by SDS-PAGE and enzymatically digested in the corresponding peptides, which are separated by HPLC. Peptides are then ionized and analysed by tandem mass spectrometry generating MS and MS/MS spectra. MS/MS spectra are searched against protein databases to obtain the protein identification. (Adapted from [www.biochem.mpg.de](http://www.biochem.mpg.de))

A mass spectrometer consists of three main parts:

1. *The ion source*: it converts the peptides into gas-phase ions.
2. *The mass analyzer*: it separates the ions according to their mass/charge ratio ( $m/z$ ).
3. *The detector*: it records the number of ions at each  $m/z$  value.

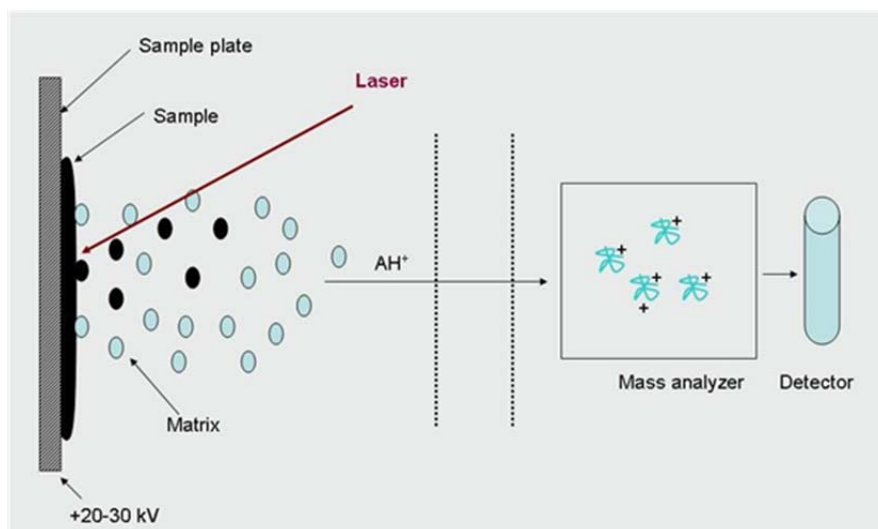
### 1. Ion Source and ionization of peptides

Proteins and peptides are non-volatile and polar compounds as such their transferring into the gas phase without causing extensive degradation, required special vehicles.

Mass spectrometers can measure  $m/z$  values of ionized molecules in gas-phase but since proteins and peptides are non-volatile and polar molecules, their analysis require an ionization method to transfer them into the gas-phase. One of the most important developments in instrumentation has been the introduction of “soft-ionization” technology, which allows for the ionization of proteins and peptides without causing extensive degradation. Two techniques are very widely used in the modern bench-top MS devices: matrix-assisted laser desorption ionization (MALDI) (Hillenkamp and Karas 1990; Hillenkamp, Karas et al. 1991) and electrospray ionization (ESI) (Fenn, Mann et al. 1989). In a MALDI source, peptides are co-crystallized with a solid-phase matrix on a metal plate. The matrix typically consists of a small organic molecule and the most extensively used for the analysis of peptides and proteins are  $\alpha$ -cyano-4-hydroxycinnamic acid and di-hydroxy benzoic acid (DHB). When laser pulses irradiate the resulting solid mixture, this absorbs the laser energy and transfers it to the acidified peptides. At the same time, the rapid heating causes desorption of both matrix and newly formed  $[M+H]^+$  protonated peptides into the gas-phase (Fig. 5). This ionization process generates ions in packets rather than beam-type, thus it requires an MS analyzer capable either to acquire a wide range of



masses in a single scan or to trap all the ions simultaneously. MALDI ionization can support different types of mass analyzers, but the most common combination for proteomics studies is the MALDI/time-of-flight (TOF) setup (Song and Lee 2001); in this analyzer, ions generated in the source are accelerated to a fixed amount of kinetic energy and travel down a flight tube. Smaller ions will have a higher velocity and will be recorded by the detector before the larger ones. In a TOF spectrum the  $m/z$  value for a given analyte is proportional to the time required to reach the detector.

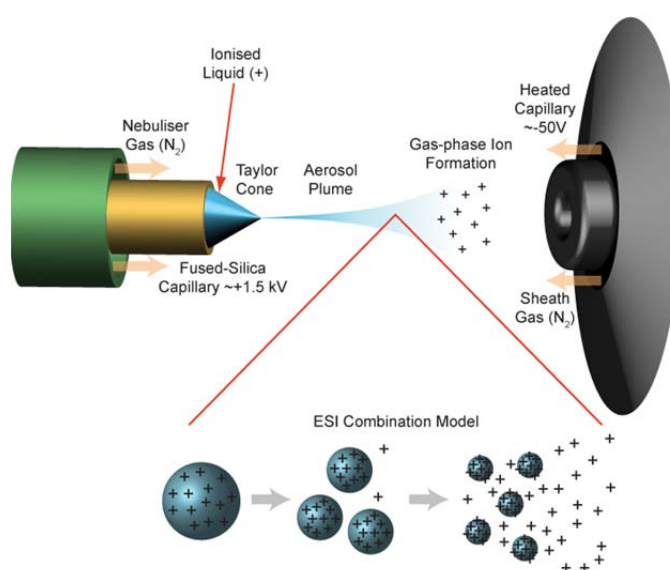


**Fig. 5: Schematic representation of sample ionization in MALDI ion source.** Sample is co-crystallized with matrix solution on a sample plate; laser irradiation transfer energy to the sample thus promoting its conversion in to gas-phase. Ions are then separated in the mass analyzer and detected by the detector. (Adapted from <http://www.tankonyvtar.hu>)

In contrast to MALDI, the peptides or proteins analysed by ESI source are in aqueous solution and their charge is controlled by the pH of the solution. Peptides contain functional groups and at acidic pH values, protonation of the amines confers an overall net positive charge to the analyte; on the contrary, basic pH causes de-protonation of the amine and carboxyl groups thus conferring a negative charge. ESI sources produce ions from solution; the electrospray phenomenon is driven by applying a high voltage (2-6kV) between the emitter at the end of the LC-column and the inlet of the mass spectrometer. This electrically charged spray is followed by creation and desolvation of analyte-solvent droplets aided by a heated capillary and in some cases by sheath gas flow at the inlet

(Fig.6). Different theoretical models were formulated to explain the mechanism of ions formation in ESI source, but the features of this ionization technique are: the formation of multi-charged ions and the sensitivity to analyte concentration and flow rate.

Typically, ESI source are usually coupled “on-line” with the LC instruments to achieve continuous analysis; an important development in ESI includes the micro and nano-ESI source where the flow rates are lowered to nanoliter/min, in order to improve the sensitivity (Griffin, Coffman et al. 1991; Emmett and Caprioli 1994) and increase the concentration of the analyte.



**Fig. 6: Schematization of electrospray-ionization process.** Peptides eluted from the chromatographic column are ionized by applying a high voltage between the capillary and mass spectrometer. Charged liquid forms a cone shape (known as Taylor cone) and the analyte-solvent droplets burst away in to a spray. (Adapted from <http://www.lamondlab.com>)

## 2. Mass Analyzer and dissociation methods

The mass analyzer corresponds to the central core of the mass spectrometer and they can store ions and separate them based on the  $m/z$ . Common mass analyzers that have proven useful for peptide analysis from complex mixture are the linear ion trap (LIT), the Orbitrap, the FT-ICR, the quadrupole (Q) and the TOF above-mentioned. Since each analyzer isolates and measures peptides masses using different mechanisms, each mass

spectrometer represents a balance between sensitivity speed and accuracy. These mass analyzers fall in two broad categories: the scanning and ion-beam mass spectrometers like the TOF and Q; and the trapping mass spectrometers such as IT, Orbitrap and FT-ICR. Although these mass analyzers differ in the details of how they work, they all select a single  $m/z$  species from a mixture of peptide ions generated by the source and fragment them to obtain the MS/MS spectrum.

In proteomics, one of the most frequently used fragmentation techniques is the collision-induced dissociation (CID) (Shukla and Futrell 2000) that breaks the peptide bond in fragment ions by multiple collisions with rare gas atoms. Upon fragmentation the characteristic  $b$ - and  $y$ -ions (at N- and C-terminus, respectively) are generated which are then used for the univocal peptide identification. In addition to these standard backbone ions, MS/MS spectra contain also other fragment ions like internal ions, immonium ions or neutral losses of thereof.

In the last few years the higher energy collisional dissociation (HCD) has been adopted to high-resolution mass analyzers (Olsen, Macek et al. 2007). This dissociation resembles all the characteristics of CID and the combination with orbitrap analyzer allows for the in-time analysis of ion beams generated by the ESI source; moreover both the full MS and the MS/MS scans may be acquired at high resolution and with high mass accuracy (“high-high” acquisition mode) (Mann and Kelleher 2008) thus helping the peptide identification. In light of the type of dissociation, HCD ion spectra cover a wide mass range and are therefore particularly useful for the inspection of immonium ions as well as reporter ions resulting from isobaric-tag labeling strategies (*see paragraph 2.3*).

These dissociation methods generate limited information for peptide longer than 15 amino acids coming from the digestion with other proteases such as Glu-C and Asp-N that cut at less frequent amino acids. This limitation has been addressed by *ad-hoc* dissociation strategies to carry out peptide fragmentation through electron transfer to multi-protonated

longer peptides enabling their sequencing: electron capture dissociation (ECD) (Kelleher, Zubarev et al. 1999), electron transfer dissociation (ETD) (Syka, Coon et al. 2004) and ETD at HCD (hcETD) (Frese, Nolting et al. 2013).

In a typical shot-gun proteomics workflow, peptides are analysed in a data-dependent acquisition mode (DDA) (Sandhu, Hewel et al. 2008) where the mass analyzer selects the most intense ions in a time-window and subject them to fragmentation; the analyzer then automatically switches back to the full scan MS mode and selects the next most intense to be fragmented. DDA is therefore biased towards the most abundant species, thus more recently have been employed other acquisition methods that in a time window tend to generate as many MS/MS spectra as possible without intensity threshold (named data independent acquisition or DIA) (Gillet, Navarro et al. 2012) or methods that alternate different CID and ETD dissociation strategies according to the charge state of the precursor ion (data-dependent decision tree or ddDT) (Swaney, McAlister et al. 2008).

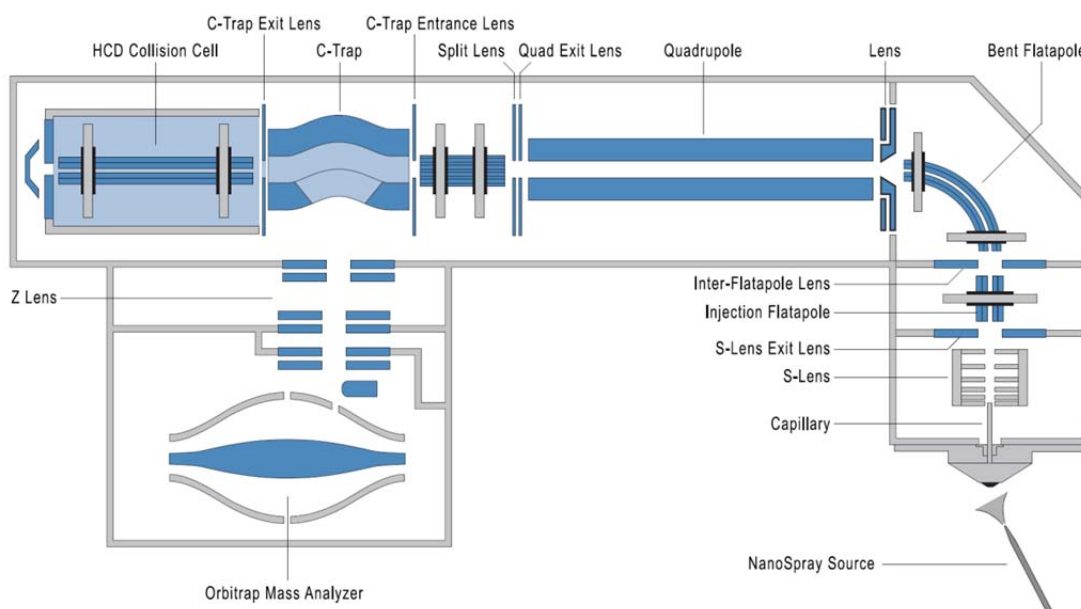
### 3. Detector

The role of the detector, as “Channeltron” or “Electron multiplier tubes (EMT)”, is to record the number of ions at each  $m/z$  value, thus it is placed at the end of mass spectrometer. “Channeltron” is a dynode coated with semiconductor material. The ions that strike the inner walls cause the emission of electrons, which are recorded in a counting system. “EMT” consists in a series of dynodes that cause the same effect of “Channeltron”, but it is able to amplify the signal of electric current until it is quantifiable.

In the last years the combination of linear ion trap with orbitrap analyzer has become very popular in bench-top instruments giving rise to the Orbitrap family of hybrid mass spectrometers. More recently the linear trap quadrupole (LTQ) Orbitrap Velos instrument has been developed; this device has a linear geometry constituted by the s-lens for the ions transmission, the LTQ for isolation but also fragmentation and scanning at low

resolution, the c-trap for the ion transferring to the HCD collision cell or to the orbitrap analyzer. This instrument allows both CID and HCD fragmentations and MS acquisition at high resolution while MS/MS can be acquired at high or low resolution.

Nowadays a new hybrid instrument has become popular in mass spectrometry-based proteomics, the Q Exactive. This instrument combines the selectivity and the speed of the quadrupole mass filter with the high resolution of the orbitrap analyzer. Thanks to the implemented instrument geometry and the presence of the quadrupole, Q Exactive selects virtually instantaneously the ions and is able to fragment them in HCD allowing the so-called “hi-hi” acquisition modality in a very short duty cycle (Fig. 7).

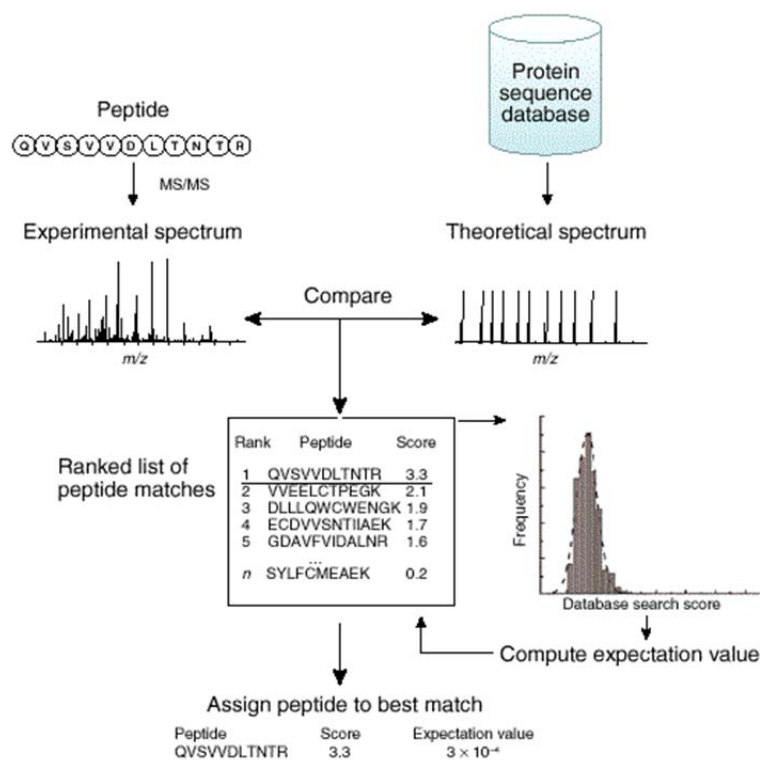


**Fig. 7: Scheme of Q Exactive geometry.** Construction details of the hybrid instrument constituted by the quadrupole mass filter HCD collision cell and orbitrap mass analyzer closely interfaced by the c-trap (Adapted from Michalski et al 2011).

### 2.2.1 Protein identification from MS data

In shotgun proteomics, the identification of a protein is reduced to the identification of the constituent peptides generated by enzymatic digestion. In MS, the analysis of peptides has some advantages: first they are easily eluted compared to intact proteins; second MS has a higher sensitivity for small molecules in respect to proteins and finally peptides are easily fragmented (Rappsilber and Mann 2002).

The identification of a peptide is performed by using its tandem MS-spectrum; several MS/MS databases have been developed but they all follow a similar procedure. Briefly, every ion MS/MS spectrum is scored against a database containing all the theoretical fragmentation patterns constructed for peptides from the searched database. User-specified criteria like the mass tolerance, the proteolytic enzyme adopted and the post-translational modification allowed, restrict the pool of candidate peptides. The output results is a list of fragment ion spectra matched to the peptide, ranked according to the search score; this value measures the similarity between the experimental and the theoretical spectrum and allows a first separation between true or false identifications (Fig. 8).



**Fig. 8: Protein identification by MS/MS database searching.** Classical workflow for protein identification. (Adapted from Nesvizhskii A.I. et al. 2007).

### 2.2.2 Biochemical enrichment strategies and mass spectrometry-based identification of post translational modifications

Generally, the workflow for proteins identification is also applicable to PTMs although this analysis is more difficult due to three characteristics of post-translation modifications: first PTMs are typically low abundant, hence high sensitivity of detection is

required; second the covalent bond between the PTM and the amino acid side chain of the peptide is frequently labile, thus it is difficult to maintain the peptide in its modified form during sample preparation and ionization. Third, some PTM are often transient, thus their capture is extremely difficult.

Due to this complexity, common strategies for the enrichment of modified peptides have been developed; examples of these approaches are: immunopurification (IP), immobilized metal affinity chromatography (IMAC), titanium dioxide (TiO<sub>2</sub>) metal-based chromatography, ion exchange (strong cation/anion exchange, SCX/SAX) chromatography and hydrophilic interaction liquid chromatography (HILIC).

Immunopurification relies on the use of specific antibodies able to recognize specifically a certain modification to enrich for it prior to MS analysis. SCX and SAX are separation strategies orthogonal respect to the RP (Giddings 1984) and are used in highly complex samples to decrease the complexity. The other three methods, namely IMAC (Ficarro, McClelland et al. 2002), TiO<sub>2</sub>(Pinkse, Uitto et al. 2004) and HILIC (McNulty and Annan 2008) are all widely used in phosphoproteomics studies and are based on the principle of chelating phosphate group of phosphopeptides by using different metals like Fe<sub>3+</sub> or TiO<sub>2</sub>+

Conceptually, the identification of a post-translational modification by MS is rather straightforward, since each PTM generates a positive or negative delta mass ( $\Delta$  mass) relative to the unmodified form of the same peptide. However, the identification of post-translational modifications is particularly challenging because the search engine has to take into account all the different modified forms that might exist. Computational methods for PTM annotation analyse MS and MS/MS spectra taking into account both the delta-mass produced by the modification of interest and neutral losses or other diagnostic ions like the immonium ions (Matthiesen, Trelle et al. 2005). The statistical approaches can be broadly clustered in two groups: in the first, the user selects *a priori* a list of modification that will

be matched during the sequence database search; in this step, PTMs are assigned to the specific amino acid of a candidate peptide sequence. In a very large dataset, the number of modifications specified is exponentially correlated with the complexity of the search, thus a restriction is usually imposed in order to reduce the rate of false-positive assignments.

The second approach is based on a “blind” attribution of PTMs and it is composed of two sequential steps. In the first one, only unmodified peptides or recurring forms of thereof are searched, such as oxidation of methionine residues; the search of these standard modifications avoids the false-positive assignments of other PTMs later on. Due to the fact that every PTM correspond to a  $\Delta$  mass, in the second step unidentified MS/MS spectra are inspected as modified forms of an already assigned peptide by using a list of naturally occurring known modifications (Marshall and Hendrickson 2008).

The false-positive identification rate is in any case higher for computational methods that search for PTMs; in this regards, the development of high-resolution instruments have helped to partially address this issue, in fact they generate MS and/or MS/MS spectra with higher mass accuracy thus allowing the distinction between PTMs with similar delta-mass.

Some false-positives might also come from *in vitro* artifacts that are wrongly assigned as *in vivo* enzymatic modifications. A classic example is the di-glycine (-GG) tag on lysine amino acid: this delta-mass is used to determine ubiquitination of this residue but iodoacetamine (IAA), commonly used in shotgun proteomics for cytosine alkylation, has the same elemental composition (Nielsen, Vermeulen et al. 2008). Other examples of false assignment are amino acids substitutions that generate an identical delta-mass corresponding to PTM, like serine to glutamine substitution and glycine to alanine substitution that have the same  $\Delta$  mass as acetylation (+42.010) and mono-methylation (+14.015), respectively.



### 2.2.3 Analytical implementation to boost hPTMs identification

In the identification of PTM on histones, the peculiar amino acid composition of these proteins represents a further challenger: histones are in fact highly rich in arginine and lysine residues, especially at the N-terminus where the majority of the hPTMs are located; for this reason, trypsin digestion generates peptides too short to be retained by the HPLC and to be efficiently analysed *via* MS (Bonaldi, Imhof et al. 2004).

Another peculiarity of hPTM that affect MS-characterization is the fact that they often co-exist on the same histone, thus the PTMs combinatorics must be taken into consideration during the identification procedure.

In order to overcome these limitations, one way is to enzymatically digest histones with ArgC protease that cleaves at the C-terminus of arginine residues, thus generating longer and more ionizing histone peptides compared to trypsin (Bonaldi, Imhof et al. 2004; Jufvas, Stralfors et al. 2011). Moreover, these peptides will have a well-defined  $\gamma$ -ions series due to the fact that they retain a positive charge on the C-terminus of arginines (Smith, Haimberger et al. 2002; Bonaldi, Regula et al. 2004; Garcia, Mollah et al. 2007). Alternatively, histones can be first chemically derivatized with deuterated acetyl anhydride ( $D_6$ -acetic anhydride  $[(CD_3CO)_2O]$ ) and then digested with trypsin; this combination leads to an ArgC-like digestion because alkylated lysines are not digested (Shevchenko, Tomas et al. 2006). This approach has the same advantages as the classical ArgC strategy, furthermore can be employed in in-gel digestion which ensures the separation of histone at the level of individual molecules (Trelle, Salcedo-Amaya et al. 2009). An additional advantage of ArgC-like digestion is that the derivatization labels both unmodified and mono-methylated lysines by adding a deuterated acetyl moiety (delta mass or  $\Delta$  mass + 45.0294 Da) while this reaction does not occur at di-/tri-methylated or acetylated lysines, thus leading to a better distinction of peptides bearing isobaric modifications. Finally, the chemical alkylation slightly increases the retention time of the peptide, thus contributing to

the unambiguous assignment of the modifications to specific residues.

Improved approaches for hPTM identification include also the optimization of LC-MS strategies for a better separation and site-specific attribution especially of hyper-modified histone peptides. Very recently we could demonstrate that ultra-high performance liquid chromatography (UHPLC) separation outcompetes the classical high performance liquid chromatography (HPLC) in separation efficiency and this effect is particularly important for hyper-modified histone peptides bearing isobaric modifications (Soldi, Cuomo et al. 2014). Indeed, the use of columns packed with 1.9 $\mu$ m C<sub>18</sub> beads in ultra-high-pressure regime confers very high peak capacity, thus increasing the peptide separation (Shen, Zhang et al. 2005; Nagaraj, Kulak et al. 2012).

A second improvement is represented by the Q Exactive mass spectrometer that -being very selective and fast- allows a very high identification rate (Michalski, Damoc et al. 2011). The Q Exactive fragments peptides in HCD mode, thus allows a “high-high” acquisition that helps the hPTMs assignment.

The platform that combines UHPLC and Q Exactive mass spectrometer out-competes previous analytical strategies (e.g. the HPLC coupled with LTQ Velos Orbitrap system) and boosts the dissection of hyper-modified peptides bearing isobaric modifications. ArgC-like digestion contributes to the better dissection of hPTMs, even though in solution Arg-C digestion may represent the ideal choice in peptides containing multiple lysines where the chemical alkylation can induce an “isobarization effect” (peptide 4 – 17 of histone H4) where peptide naturally acetylated at different degrees are equalized by the chemical acetylation (Soldi, Cuomo et al. 2014).

### 2.3 Relative quantitation strategies in MS-based proteomics

In the last few years, the concept that the mere qualitative data is not sufficient to describe the dynamics of the proteome is raising; thus, the quantitative information is

needed, especially for the comparative analysis of proteins among distinct functional states.

Mass spectrometry is not intrinsically a quantitative technique for different reasons: first, the ion intensity is proportional to its amount but it is also dependent on its chemical-physical properties like molecular composition, charge and hydrophobicity. Second, the comparative analysis between two different LC-MS/MS runs is influenced by external variations like temperature and chromatography reproducibility. In order to overcome this limitation, two conceptually distinct solutions are adopted: strategies based on *a posteriori* comparison among samples (*label-free* approaches) and quantification techniques that rely on peptide/protein isotopic labeling (Fig. 9).

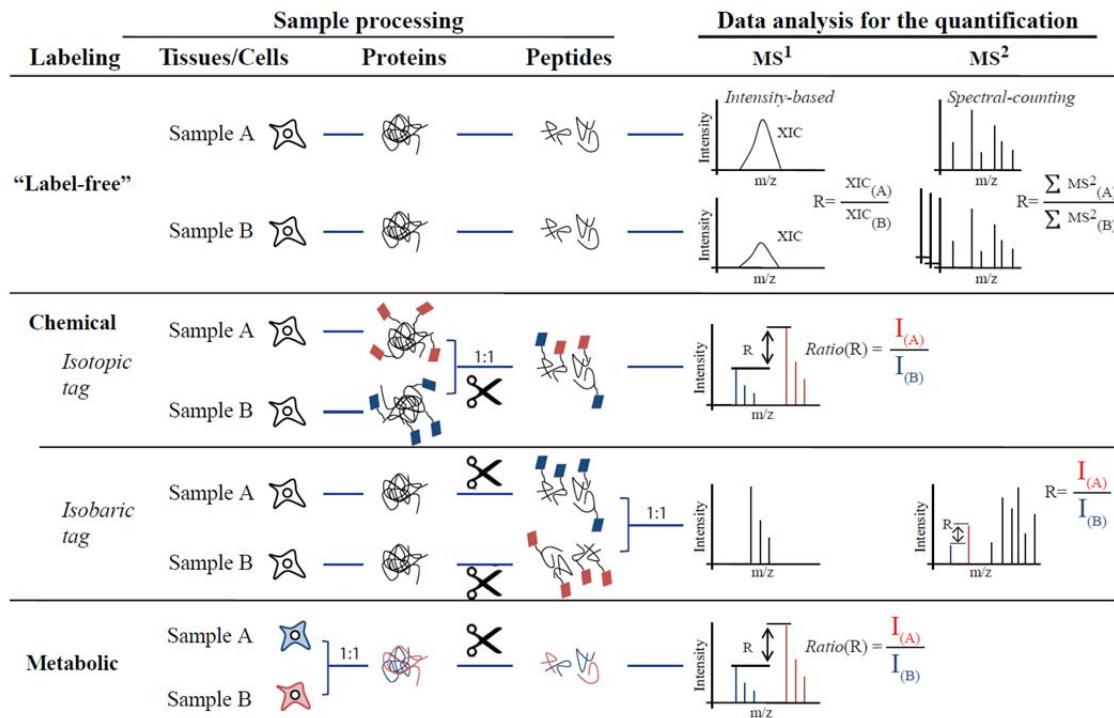
*Label-free* strategies take advantage of peptide intrinsic characteristics and the quantification is made post-acquisition. These approaches can be grouped in: 1) intensity-based, where the quantification relies on the comparison of one or more peptide intensities of the same protein in different samples and 2) spectra counting quantification strategies, based on the number of MS/MS spectra for the same protein in different samples.

The intensity-based approaches rely on the concept that the extracted ion chromatography (XIC or area under the curve, AUC) for a given peptide is linearly correlated with its abundance. XICs of the same peptide in different samples are extracted and compared for the peptide quantification. These approaches require extremely reproducible chromatography among multiple runs and *ad hoc* software capable to perform retention time realignment and peptide intensity normalization over the global intensity.

Spectral counting strategies use MS/MS spectra both for identification and quantification; these approaches are based on the rationale that more abundant proteins are identified with a higher number of peptides. Limits of these techniques are that both the length and the amino acid composition influence this linear correlation; furthermore, a high number of spectra are needed for accurate protein quantification.

*Label-free* strategies offer the great advantage that can be extended to a theoretically infinite number of conditions and can be applied to every type of samples (cell lines, primary cells, tissues, etc.). Limits of these approaches are: the requirement of highest technical and experimental reproducibility and an overall lower accuracy in protein quantification in respect to isotope-based techniques.

All the different isotope-based quantification strategies follow an overall similar rationale: create a delta-mass that distinguishes peptide derived from different samples in a single MS analysis. In a classical workflow, samples isotopically light-(L) or heavy-(H) labeled are mixed and the H/L intensity ratio allows for an accurate peptide and protein relative quantification in the two samples. The isotopic tag can be introduced in the peptide at different stages during the sample processing and the general rule is that the early this step is performed, the less are the variations introduced between the two samples to be compared, so the more accurate the quantification is.



**Fig. 9: Schematic representation of distinct relative quantitation strategies.** The comparative analysis made by label-free or chemical/metabolic labeling approaches are summarized in the picture. Extracted proteins from cells are digested in to peptide before mass spectrometry analysis and quantification at MS<sup>1</sup> or MS<sup>2</sup> level. Labeling is represented by blue and red colors, scissors mean protein digestion and 1:1 the sample mixing. XIC = extracted ion chromatography, I = intensity, A and B refer to the corresponding samples

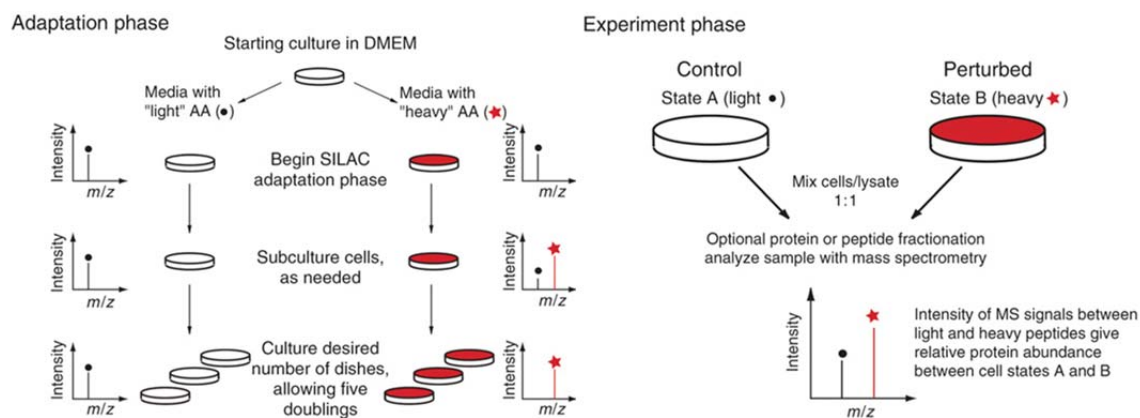
Based on how proteins or peptides are labeled, isotope-based approaches can be divided in chemical- and metabolic-labeling strategies. In the first group the tag is added covalently to the reactive side chains of amino acids before or after proteolysis, through a chemical reaction. The first example of chemical-labeling was the isotope-coded affinity tag (ICAT) (Gygi, Rist et al. 1999), where proteins are labeled with cysteine-specific tag; the principal limit of this approach is that only cysteine-containing proteins can be labeled and quantified at MS level. The isobaric tag for relative and absolute quantification (iTRAQ) (Ross, Huang et al. 2004) is the most successful example of chemical labeling. Side chains of lysine residues are labeled with an isobaric tag that allows the quantification at MS/MS level. The advantage of iTRAQ is the possibility to multiplexing, by using up to 8 different isobaric tags; however, this approach requires high-resolution instruments and possibly HCD fragmentation, due to the low  $m/z$  of the reporter ions of the tag.

Both ICAT and iTRAQ can be applied to different types of samples, such as tissue or clinical biopsies; among the limitations we mention the variable labeling efficiency that depends on sample complexity and the less accurate protein quantification due to the rather late step of labeling throughout the sample preparation protocol.

A valid alternative is the metabolic-labeling strategy, where stable isotopes are added to the growing cell medium as metabolic precursors in order to be incorporated in the proteome during protein biosynthesis. Major advantages of metabolic labeling approaches are that they can be applied to *in vivo* studies and that samples can be mixed at the beginning of the sample preparation. These strategies are therefore compatible with complex purification procedures and allow accurate protein quantitation. A clear drawback of these techniques is that, they can be generally applied only to cultured cells. The most successful example of metabolic-labeling strategies is the stable isotope labeling of amino acids in cell culture (SILAC) (*see paragraph 2.3.1*).

### 2.3.1 SILAC-based approaches for the relative protein quantitation

SILAC is a powerful and accurate procedure based on the use of essential amino acids where the naturally isotopes  $^{12}\text{C}$ ,  $^1\text{H}$  and  $^{14}\text{N}$  (or light, L) are substituted with heavy (or H) isotopes ( $^{13}\text{C}$ ,  $^2\text{H}$  and  $^{15}\text{N}$ , respectively). In a classical SILAC experiment two cell populations, representative of two different functional states, are grown in medium containing either with light or heavy isotope-coded essential amino acids for a sufficient number of cell doublings in order to assure the full incorporation of the isotopes in the proteins. After complete labeling, two identical proteomes are generated that only differ for protein isotope composition; the two samples might be directly mixed in equal amounts, thus generating a unique proteome that is processed *via* LC-MS/MS. As readout, every peptide will be represented as a pair where the  $\Delta$  mass is dependent on both the number and the type of the amino acids incorporated in the sequence (Ong, Blagoev et al. 2002) (Fig. 10). In SILAC labeling are generally used arginine and lysine isotope-coded amino acids in combination with trypsin digestion because this set up allow for the labeling of virtually every peptide.



**Fig. 10: Schematic view of SILAC labeling and of a prototypic experimental design.** SILAC experiment is composed of two different phases: in the adaptation phase cells are grown in medium containing either light- or heavy-isotope coded amino acids until fully incorporation (red star). In the experiment phase the two population are mixed, proteins are purified, digested and analysed by MS. (*Adapted from Ong S.E. & Mann M., 2007*).

The SILAC labeling has been successfully applied to a wide variety of studies: from protein expression profiling (de Godoy, Olsen et al. 2006) to global PTM analysis (Blagoev and Mann 2006), protein-protein and DNA-protein interaction analysis (Blagoev, Kratchmarova et al. 2003).

As further implementation, triple-SILAC setup has been proposed to discriminate among three distinct conditions within a single LC-MS/MS analysis. In this labeling strategy, a heavy-medium deuterated form of the above-mentioned amino acids are employed; as result each peptide is represented by peak-a triplet in MS. Due to its versatility, triple-SILAC has been successfully adopted to determine the stimulus specific interactions in the Wnt pathway (Hilger and Mann 2012). Triple SILAC was also employed in pulsed-SILAC (pSILAC) experiments, to quantitatively profile protein translation dynamics (Schwanhausser, Gossen et al. 2009).

The major limitation of SILAC is that can be applied only to cells growing in culture for a sufficient number replications that ensure the complete incorporation of the isotope-coded amino acids. A solution is offered by the SILAC spike-in strategy that allows the protein quantification of unlabeled samples by using a protein extract from SILAC-labeled cells as internal standard (Ishihama, Sato et al. 2005). This strategy has been further implemented in the super-SILAC approach that uses a mixture of isotopically heavy-labeled cell lines (super SILAC-mix) as internal standard. The super-SILAC mix is added as spike-in to samples of various origins, in order to quantify the protein amount through the H/L ratio(Geiger, Cox et al. 2010).

SILAC labeling can also be used to directly label PTMs: the heavy-methyl SILAC (hmSILAC) strategy implies the use of methionine-free media to which light- or heavy-labeled forms of this amino acid are added. In the cell, methionine is converted to S-adenosyl methionine (SAM), which represents the only donor of methyl groups in enzymatic methylation reaction. Proteins that contain methylations are therefore

enzymatically light- or heavy-methyl labeled and upon mixture the corresponding peptides are identified as a peak pair. The MS analysis allows for the unambiguous identification and quantification of *in vivo* methylated sites (Ong, Blagoev et al. 2002; Bremang, Cuomo et al. 2013).

## 2.4 Screening of hPTM readers by distinct qProteomics approaches

Histone PTMs are one of the mechanisms to fine-tune gene expression; in particular hPTMs function as docking sites for the recruitment of readers that translate a distinct modification pattern into a specific biological outcome. Due to their biological relevance, an increasing number of quantitative proteomics approaches have been developed to screen hPTM readers. These approaches can be classified into two main strategies: those based on affinity pull-down screening and those based on the isolation of native chromatin domains.

The first approach designed for the identification of hPTM readers relied on pull-down strategies using peptides bearing a certain modification, followed by MS-analysis of single bands; in this type of experiment, the specificity was assessed by using unmodified peptide as control (Wysocka, Swigut et al. 2005; Wysocka, Swigut et al. 2006). A step forward was represented by the SILAC-based peptide pull-down where the labeling allows the robust identification and quantification of hPTMs readers (Vermeulen, Mulder et al. 2007; Vermeulen, Eberl et al. 2010); as alternative the Cross-Linking-Assisted SILAC Protein Identification (CLASPI) approach adopts also the photo-cross-linking in order to stabilize the more labile interaction between hPTM and binders, to identify low affinity readers (Li, Foley et al. 2012). The SILAC Nucleosome Affinity Purification (SNAP) approach was developed to study the synergic contribution between DNA methylation and hPTMs in promoting the binding of chromatin factors; indeed recombinant nucleosomes bearing methylation on H3 residues, on DNA or combination of thereof were used to



assess the interplay between these epigenetic marks in recruitment or chromatin binders (Bartke, Vermeulen et al. 2010).

All these *in vitro* strategies, allow a precise assessment of proteins interacting with a specific chromatin modification; however, they suffer from the artificial conditions in which interactions are investigated and can address only limited combination of PTMs.

The second class of approaches includes techniques that try to reproduce more physiological conditions, through the proteomic characterization of enriched native chromatin regions.

The first strategy designed to purify a specific chromatin landscape was exploited through a DNA sequence-specific enrichment: in the Proteomics of Isolated Chromatin (PICh) approach, a nucleic acid probe is employed to isolate telomeric DNA, thus allowing the proteomic characterization of determinants and proteins associated to these chromatin regions (Dejardin and Kingston 2009). Similarly, the Chromatin Affinity Purification with Mass Spectrometry (ChAP-MS) technique allows a more focused analysis of hPTMs and proteins associated with a single genomic locus. In this strategy the GAL1 locus was engineered in budding yeast, enriched and proteins co-associated were quantified through isotope labeling (Byrum, Raman et al. 2012). The implementation of ChAP-MS approach relied on the engineering of the transcription activator-like (TAL) protein A (TAL-PrA) in order to make amenable to recognize GAL1 promoter. The TAL-PrA was used for the affinity purification of the locus followed label-free quantitation of interacting proteins (TAL-ChAP-MS) (Byrum, Taverna et al. 2013).

The enrichment of a specific chromatin region can be alternatively achieved by using a specific chromatin determinant as bait for the immunoprecipitation.

In the Modified ChIP (mChIP) approach Figeys and colleagues endogenously expressed TAP-tagged version of histone H2A (Hta2p) and its variant Htz1p in *S. cerevisiae*. After

affinity purification, they were able to reconstruct the network of these chromatin-associated proteins through identification *via* LC-MS/MS (Lambert, Mitchell et al. 2009).

A similar approach is the Chromatin Interacting Protein-Mass Spectrometry (ChIP-MS) where male-specific lethal proteins (MSL2 and 3) were expressed as biotin-tagged *in vivo* in *D. melanogaster*. The streptavidin chromatin enrichment allowed to identify both the MSL-enriched hPTMs and to characterize proteins interacting with the MSL complex (Wang, Alekseyenko et al. 2013).

An alternative is represented by the Chromatin Proteomics (ChroP) strategy, which was developed by our groups and combines SILAC-based quantitative proteomics together with ChIP technique. In this approach cells are first metabolically labeled and cross-linked; after that, markers of specific chromatin regions are used as bait in the ChIP, followed by SILAC-based quantitative proteomic analysis of the determinants at the enriched functional chromatin territories. The ChroP was successfully employed for the characterization of hetero-chromatin and eu-chromatin regions using as bait H3K9me3 and H3K4me3, respectively (Soldi and Bonaldi 2013). As future perspective, a very interesting development would be the dissection of more focused chromatin domains with a proved relationship between protein composition and function; in this view, ChroP could be used to analyse in-depth the proteomic composition of a very restricted chromatin regions. We elaborated on this concept in the work discussed in this thesis, using ChroP to dissect the structure of enhancers

## 2.5 Enhancers' elements and transcriptional regulation as prototype for dynamically regulated biological process

In biology there is a strong interest in understanding the mechanisms through which gene expression is established and fine-tuned at cellular level and specifically orchestrated in a time-specific fashion. The fundamental importance of this tight regulation relies on the

fact that deregulation of this process often leads to disease, such as cancer. Eukaryotic genes' transcription is a biological process controlled at different steps from transcription initiation, elongation, to transcriptional termination. However, most of the regulation is believed to occur at the level of transcription initiation. RNA polymerase II transcribed genes, are indeed controlled by two families of regulatory elements acting in *cis*: proximal and distal regulatory elements. In the first group we find promoter elements and transcription start sites (TSSs), while in the latter category there are enhancers, silencers, insulators and locus control regions (LCR), which all contain binding sites for both transcription activators and suppressors, to fine tune gene transcription.

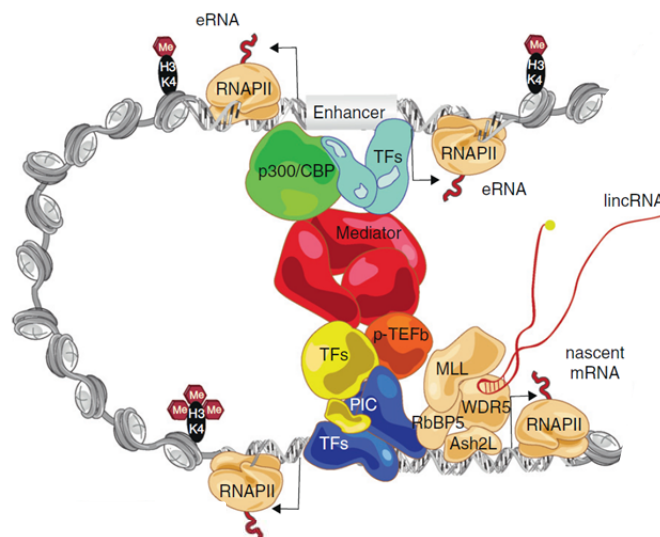
Enhancers are defined as genomic regions approximately few hundreds of base pairs long, which are able to increase the transcription of target genes and to regulate the transcription in a cell-specific and temporal-specific manner (Banerji, Rusconi et al. 1981): in fact, the same promoter can be activated by distinct enhancers in different times or tissues, or in response to different stimuli, thus regulating the time-and-space- specific gene expression.

Enhancers are characterised by two main features: first, they function in an orientation-independent manner relative to the promoter of the target gene. Enhancers are composed of clusters of transcription factor binding sites (TFBSs) and the feature of independency of orientation only applies to the enhancer as whole, indeed the organization of the TFBSs is tightly linked to precise enhancers' regulation.

TFs may bind enhancers in different ways: they can be recruited in clusters with no apparent protein-protein interaction (PPI) among them (enhanceosome model); alternatively, they may first interact with co-factors that promote enhancer binding (cooperative model). Otherwise, specific TFs, named pioneer TFs, may compete with nucleosomes for the binding to target enhancer regions, thus generating a nucleosome-depleted region (NDR) which allow the subsequent recruitment of other transcription

factors. Pu.1 and AP1 are two examples of pioneer transcription factors able to bind their target motives within a closed chromatin environment and to function as cell-specificity determinants and master regulators (Heinz, Benner et al. 2010; Serandour, Avner et al. 2011).

The second characteristic of enhancers is that they show a long-distance activity; indeed, they can be found frequently several hundreds of kilobases from the corresponding TSS. Moreover they can be located upstream or downstream to the promoter gene, and both in extragenic and intragenic regions. One of the most extreme examples is the enhancer of the *Sonic hedgehog* (*Shh*) gene, which is located more than 1 Mbp from its gene promoter (Lettice, Heaney et al. 2003). Due to this long-distance activity, the key question is how enhancers and promoters communicate and synergize to activate the transcription of the joint target gene; currently a widely accepted model proposes the existence of a chromosomal looping between these two transcriptional regulatory elements that allows their physical interaction and thus promotes transcription initiation (Fig. 11).



**Fig. 11: Regulatory chromosomal looping.** The enhancer region is marked by mono-methylation of H3K4 while promoter region is enriched in H3K4me3. The promoter – enhancer chromosomal looping is mediated by sequence-specific TFs that -together with the mediator complex- stabilize this long-range interaction. Pre-iniziation complex (PIC) and positive transcription-elongation factor b (p-TEFb) regulate the transcription initiation and elongation of the nascent RNA. Enhancer RNA (eRNA) and other lincRNA) are transcribed and may help stabilize these structure, even if their precise function is still controversial. (MLL complex: MLL, WDR5, RbBP5, Ash2L). (Adapted from Flynn R.A. and Chang H. Y., 2011).

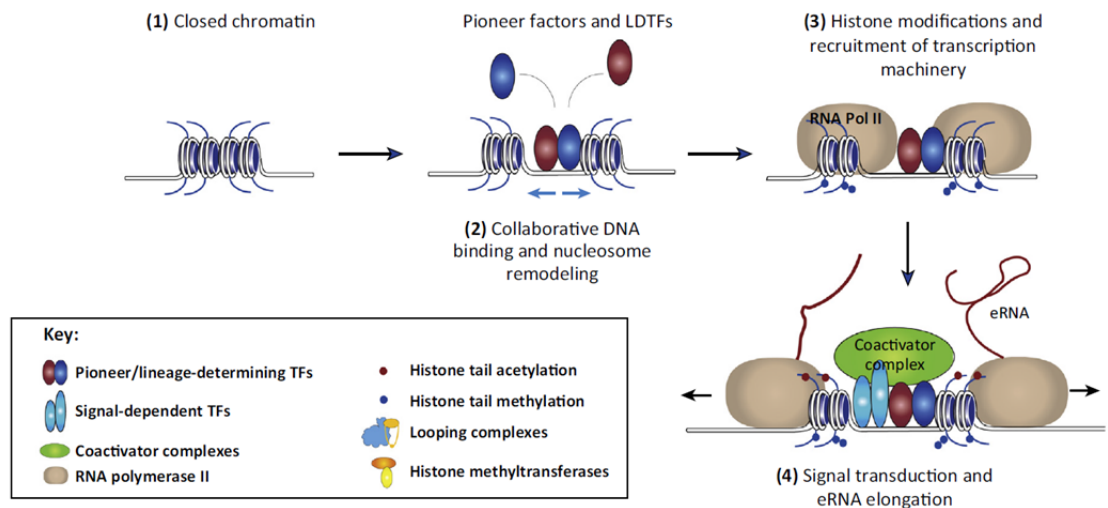
The interaction can also involve different components of the basal transcriptional machinery, as well as components of the mediator complex that acts as transcriptional co-activator (Koch, Fenouil et al. 2011). Also the cohesin complex seems to play an important role in the stabilization of the bridging between enhancers and promoters. The development of new technologies based on chromosome conformation capturing (3C) (Dekker, Rippe et al. 2002) and microscopy (Lieberman-Aiden, van Berkum et al. 2009) have allowed studying the physical interactions between enhancers and promoters, facilitating the understanding on this long-range communication. Interestingly, recent data showed that the formation of new enhancer loops precedes the transcription activation in the  $\beta$ -globin locus (Krivega, Dale et al. 2014), suggesting the pre-existence of a high-ordered structure, which is then dynamically remodeled in response to gene activation. These findings are in line with earlier data showing the co-localization of genes regulated by common TFs in nuclear foci termed “transcription factories” (Schoenfelder, Sexton et al. 2010; Razin, Gavrillov et al. 2011), even though with exceptions, since not all the genes in these clusters resulted to be active at the same time, which implies that other regulatory processes may exist.

At the molecular level, enhancers were initially defined as genomic regions hypersensitive to the DNase digestion due to more open chromatin conformation caused by the TFBSs; for this reason, techniques like the micrococcal nuclease (MNase) digestion and the formaldehyde-assisted isolation of regulation elements (FAIRE) can be used for the identification of these *cis*-regulatory regions.

The application of genome-wide strategies has added further knowledge on enhancers' molecular signature and suggested a characteristic enrichment of the mono-methylation of lysine 4 of histone H3 in the absence of tri-methylation of the same residue ( $H3K4me1^{Hi}/H3K4me3^{Low}$ ) (Heintzman, Stuart et al. 2007) to these *cis*-regulatory regions. More recently, by using CHIP-based approaches, the repertoire of enhancers has been

further subdivided in distinct functional types: active, poised or inactive *cis*-regulatory regions, marked by different and diagnostic hPTMs.

In recent years a new feature of enhancers has emerged: the presence of transcription at these *cis*-regulatory by RNA polymerase II producing long non-coding RNAs, named enhancer RNAs (eRNAs) (De Santa, Barozzi et al. 2010; Heinz, Benner et al. 2010; Kim, Hemberg et al. 2010). Many of these eRNAs are short, bidirectional, lack a poly-A tail and are present at very low copy number in the cell. The function of these non-coding transcripts is still not completely understood; in the past different models have been proposed to explain their activity, in line with that of other non-coding RNAs (lncRNAs).



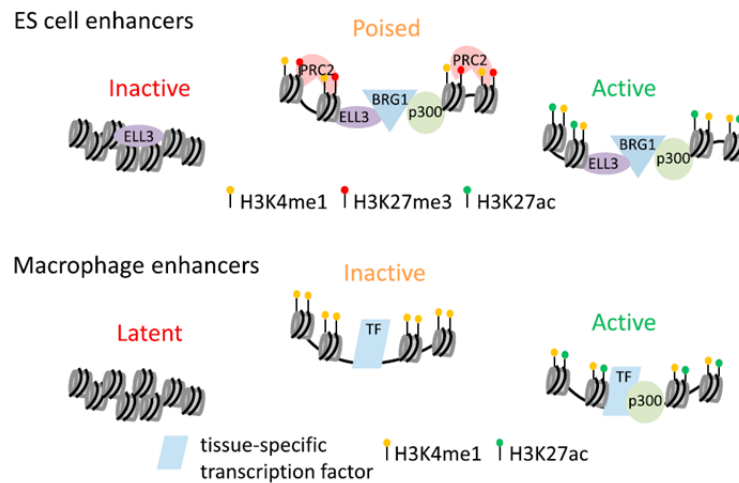
**Fig. 12: Dynamics of chromatin determinants at enhancers.** The collaborative binding of pioneer TFs and lineage-determining TFs (LDTFs) to closed chromatin (1) determine nucleosome remodeling and increased chromatin accessibility (2). This binding triggers mono-methylation of K4 of histone H3 (H3K4me1, blue circle) and the recruitment of RNA polymerase II (3). Upon stimulation, signal-dependent TFs (SDTFs) bind to recognition sequence at enhancers and recruit co-activator complexes leading to further epigenetic modifications (histone acetylation, red circles) and eventually activation of transcription at enhancers (eRNA). (Adapted from Lam M. T. Y., et al 2014).

lncRNAs have in fact been proposed to act as decoy molecules, by binding and displacing TFs or repressor from their target DNA region; alternatively they can play a role as scaffold by tethering together proteins in a complex. A third model proposes a bridging function between proteins and DNA, which favors the recruitment of specific factors to the DNA. Finally, a fourth hypothesis is a structural model where, they can bind and stabilize

the chromosomal looping 3D structures. In addition to all above-mentioned models that apply to lncRNAs in general and could be hypothesized to work also for eRNAs, these specific set of transcripts may contribute to gene expression regulation in a time- and cell-specific manner (Fig. 12).

Gene expression regulation plays crucial roles in different biological processes from the control of cell differentiation during embryonic development, to response to external stimuli. This tight regulation is achieved through the interplay among different chromatin determinants that modify enhancers' structure and composition to enforce specific transcriptional programs (Fig. 13).

In embryonic stem (ES) cells enhancers can be divided in three classes based on their functional state: poised, active and inactive. The first group is composed of enhancers of genes not actively transcribed but with the potential to be. They are marked by the enrichment of H3K4me1 and tri-methylation of lysine 27 of histone H3 (H3K27me3), together with the presence of polycomb repressive complex 2 (PRC2), Brg1 protein (also known as Smarca4), p300 and the RNA polymerase II elongation factor ELL3. During differentiation, the poised enhancers of newly expressed genes become active; the active state is characterized by the loss of PRC2 and H3K27me3, paralleled by the gain of acetylation of H3K27 (H3K27Ac) (Creyghton, Cheng et al. 2010; Rada-Iglesias, Bajpai et al. 2011). Pluripotency genes get instead inactivated during differentiation; this inactive state is features by the loss of all enhancers' hallmarks except for the ELL3 which is retained (Fig. 13).



**Fig. 13: Hallmarks of enhancers of ES and macrophages cells at different developmental stages.** In embryonic stem (ES) cells enhancer regions in a poised state are marked by the presence of the four DNA binding protein (ELL3, BRG1, p300 and PRC2) and two hPTMs (H3K4me1 and H3K27me3). Upon activation enhancers acquire H3K27Ac and lose both PRC2 and H3K27me3; while one inactivated loses all the enhancer determinants but ELL3. In macrophages inactive enhancers are marked by H3K4me1 and Pu.1; active enhancers acquire H3K27Ac and p300 while the latent enhancers seem to not express any determinant (*Adapted from Plank J. L. and Dean A., 2014*)

Gene expression is dynamically regulated in a very plastic fashion also in fully differentiated cells, like the macrophages. In response to inflammatory stimuli macrophages undergo a global gene expression reorganization, which included the activation of genes coding for effectors of the inflammatory response and the parallel inactivation of other sets of genes, for example the ones coding for proteins involved in proliferation.

Hallmarks of macrophages' enhancers are high H3K4me1/H3K4me3 ratio and enrichment of the master regulator pioneer TF Pu.1 (Fig. 13), which is constitutively expressed at high levels to induce and maintain the differentiated state. This TF binds almost all the genomic enhancer regions associated with H3K4me1 (about 35.000 – 40.000 genomic sites) (Ghisletti, Barozzi et al. 2010; Heinz, Benner et al. 2010) and is important for the local deposition of this mark.

Macrophages can be stimulated *in vitro* by lipopolysaccharide (LPS) to a pro-inflammatory commitment state (M1 macrophages). LPS is a constituent of the outer membrane of gram-negative bacteria and it stimulates a strong immune response in immune-competent cells,



indeed upon LPS treatment, enhancers of inflammatory genes rapidly acquire the histone acetyl-transferase p300 that mediates H3K27 acetylation (Creyghton, Cheng et al. 2010) and other stimulus-specific TFs.

Recently a new class of enhancers has been characterized in macrophages, the so-called “latent enhancers”. They correspond to a very minor fraction of the total and do not display the classical markers of all other macrophage enhancers, at basal conditions. Only upon inflammation they acquire Pu.1, H3K4me1 and H3K27Ac, which promote a chromatin conformation prone to the transcriptional activation of target genes (Ostuni, Piccolo et al. 2013) (Fig13).

All these evidences suggest that there is a strong link between the enhancers’ composition in term of both binding proteins and hPTMs and their functional state.

However, in spite of this limited set of determinants well-characterised, many more molecular hallmarks remain to be identified; thus a more comprehensive dissection of these *cis*-regulatory regions will allow for a better understanding of the mechanism through which gene expression is regulated in time and space.

## **AIM OF THE WORK**

Chromatin architecture is dynamically regulated by associated proteins and histones and DNA modifications that synergize to create distinct “chromatin landscapes”, with regulatory effects on the functional state of the underlying genes.

In recent years a number of MS-proteomics approaches have been developed to study the protein composition of physically or functionally distinct chromatin regions. In particular, our laboratory optimized the Chromatin Proteomics (ChroP) approach (Soldi, 2013), which combines chromatin immunoprecipitation (ChIP) and SILAC-based proteomics for the characterization of both protein binders and modifications specifically clustering at distinct chromatin regions.

The aim of my PhD project was to implement the pre-existing protocol into a new version of the ChroP method, which was used to study a more restricted chromatin region namely enhancers. I decided to focus on these *cis*-regulatory regions because they are characterized by a strong link between their molecular composition and the functional state of the corresponding target genes. Furthermore, while the definition of the molecular signature enabling large-scale identification of enhancers is an appealing goal, their characterization is largely unexplored.

Based on these considerations, we reasoned that ChroP approach could provide a valid analytical tool to expand the knowledge on their molecular composition and the consequent mechanism of regulation. To this aim, I worked to adapt ChroP strategy to the different model system under investigation and to better answer new open questions. As such, I employed a different cell line (RAW 264.7 macrophage-like cell line), I optimized the protocol to use a transcription factor that directly binds to nucleosome-free regions (NFR) as bait in the preparative ChIP and performed temporal analysis of the enhancers' chromatinome by coupling ChroP with triple-SILAC labelling. Results are illustrated in the following sections.



## MATERIALS AND METHODS

### 4.1 General biochemistry and molecular biology buffers

#### 1. Phosphate-Buffered Saline (PBS), pH 7.4:

137 mM NaCl

2.7 mM KCl

8.1 mM Na<sub>2</sub>HPO<sub>4</sub>(7H<sub>2</sub>O)

1.76 mM KH<sub>2</sub>PO<sub>4</sub>

#### 2. Tris-Buffered Saline (TBS):

150 mM NaCl

2.7 mM KCl

25 mM Tris base

#### 3. Tris-Buffered Saline (TBS)-Tween:

150 mM NaCl

2.7 mM KCl

25 mM Tris base

0.1% Tween 20

#### 4. Tris-EDTA (TE):

10 mM Tris-HCl pH 8.0

1 mM EDTA pH 8.0

#### 5. Tris-Acetate-EDTA (TAE):

40 mM Tris acetate

1 mM EDTA pH 8.0

#### 6. RIPA lysis buffer

20 mM Tris-HCl (pH 7.5)

150 mM NaCl

1 mM EDTA

1 mM EGTA

1% NP-40

1% sodium deoxycholate

7. Urea lysis buffer:

20 mM Hepes, pH 7.5

8 M Urea

8. Running gel mix (1 gel at 12%):

Acrylamide/Bis-acrylamide solution 30: 0.8 ratio (2.52 mL and 3.48mL for 17.5% 12% gel)

Tris-HCl, pH 8.8 (1.5 mL)

ddH<sub>2</sub>O (1.92 mL)

SDS 10% (60 µL)

APS 10% (60 µL)

TEMED (3 µL)

9. Stacking gel mix (1 gel):

Acrylamide/Bis-acrylamide stock solution - 30: 0.8 ratio at 30% (0.52 mL)

Tris-HCl, pH 6.8 (1.25 mL)

ddH<sub>2</sub>O (3.10 mL)

SDS 10% (50 µL)

APS 10% (50 µL)

TEMED (5 µL)

10. Laemmli loading buffer (5X):

100mM Tris-HCl pH 6.8

20% Glycerol

2% SDS

5% β-mercaptoethanol

11. LDS sample loading buffer (1X):

10% Glycerol

1% Lithium dodecyl sulfate (LDS)

1% Ficoll-400

0.2 M Triethanolamine-Cl pH 7.6

0.00625% Coomassie G250

0.5 mM EDTA disodium

12. SDS-PAGE running buffer:

192 mM Glycine  
25 mM Tris base  
2% SDS

13. NuPAGE ® MES SDS running buffer:

50mM 2-(N-morpholino) ethanesulfonic acid (MES)  
50 mM Tris base  
0.1% SDS  
1mM EDTA

14. Fixing gel solution:

50% Methanol  
10% Acetic acid

15. Coomassie staining gel solution

2.5g Coomassie brilliant blue R/250  
10% Acetic acid  
50% Methanol

16. Colloidal Blue Coomassie Staining gel solution (Invitrogen Kit):

20% Methanol  
20% Stainer A (Colloidal Blue Stain Invitrogen Kit)  
5% Stainer B (Colloidal Blue Stain Invitrogen Kit)

17. Western blot transfer buffer:

25 mM Tris base  
192 mM Glycine  
20% Methanol

18. Western blot Stripping buffer:

62.5 mM Tris-HCl, pH 7.6  
2% SDS  
100 mM  $\beta$ -mercaptoethanol

19. HEPES-buffered solution (HBS) pH 7.4

25mM HEPES

145 mM NaCl

1.25 mM CaCl<sub>2</sub>

2.6 mM KCl

1.2 mM MgCl<sub>2</sub>

1.2 mM KH<sub>2</sub>PO<sub>4</sub>

#### 4.2 Cell Culture, SILAC and triple-SILAC labeling of cells

RAW 264.7 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated Fetal Bovine Serum (FBS, Invitrogen 10270-106), 2mM glutamine, 100 U/ml penicillin and 100 mg/ml streptomycin and cultured at 37°C in a 5% CO<sub>2</sub> humidified atmosphere.

SILAC-labeled RAW 264.7 cells were grown in “Light”, “Medium” and “Heavy” SILAC media prepared supplementing SILAC DMEM (EuroClone B7511L), depleted of lysine and arginine, with 10% dialyzed FBS (Invitrogen, 26400-044) and isotope encoded amino acid with different stable isotopes composition More specifically: Light isotope-coded medium contains <sup>12</sup>C<sub>6</sub><sup>14</sup>N<sub>2</sub> L-lysine (Lys 0, Sigma L8662) and <sup>12</sup>C<sub>6</sub><sup>14</sup>N<sub>4</sub> L-arginine (Arg 0, Sigma A6969); <sup>12</sup>C<sub>6</sub><sup>14</sup>N<sub>2</sub>-d<sub>4</sub> L-lysine (Lys 4, Sigma 616192) and <sup>13</sup>C<sub>6</sub><sup>14</sup>N<sub>4</sub> L-arginine (Arg 6, Sigma 643440) were added to the medium isotope-coded medium; while Heavy medium is supplemented by <sup>13</sup>C<sub>6</sub><sup>15</sup>N<sub>2</sub> L-lysine (Lys 8, Sigma 68041) and <sup>13</sup>C<sub>6</sub><sup>15</sup>N<sub>4</sub> L-arginine (Arg10, Sigma 608033) always at a concentration of 146 mg/L and 84 mg/L, respectively. RAW 264.7 cells were grown in SILAC media for 8 generations at 37°C in a 5% CO<sub>2</sub> humidified atmosphere. Careful monitoring of growth rate, viability and overall morphology demonstrated no alteration from physiology compared to cells cultured in standard medium.

Cells were eventually frozen in freezing solution at  $5 \times 10^6$  cells/mL in cryovials, let for o.n. at  $-80^\circ\text{C}$  in isopropanol cryobox and then stored in liquid nitrogen.

*Freezing solution:* 10% DMSO

90% FBS serum (Invitrogen 10270-106 or Invitrogen, 26400-044 for unlabeled or SILAC labeled cells, respectively)

#### 4.3 Bone marrow-derived macrophages in vitro differentiation

Bone marrow (BM) cells were isolated from either Fvb/Hsd (C57BL/6-129P2) mice. Red blood cells were lysed by osmotic shock (1 min in 2 ml NaCl 0.2% followed by addition of the same volume of NaCl 1.6% to reconstitute the physiological osmotic pressure). BM cells were then washed twice in cold PBS, resuspended in BM medium and plated  $3 \times 10^6$  cells in 5.5 ml of BM-Medium per 100 mm petri dish. Differentiating cells were harvested at day 7-8 of differentiation (or as otherwise stated).

*BM medium:* DMEM supplemented with 30% L929-conditioned medium, 20% FBS, 2 mM glutamine, 100 U/ml penicillin, 100 mg/ml streptomycin, 0.5% sodium pyruvate, 0.1%  $\beta$ -mercaptoethanol

#### 4.4 Cell treatment with lypopolysaccharides

Lypopolysaccharide treatment has been performed by direct injection in the medium of 100ng/mL of LPS from *E. coli* serotype 055:B5 (Sigma) for the 1h or 4hrs.

#### 4.5 Cross-linking Chromatin immunoprecipitation (X-ChroP)

X-ChroP protocol applied to this cell line was very similar to the recently published one (Soldi and Bonaldi 2013), with minor modifications regarding cross-linking conditions and antibody concentration. SILAC-labelled RAW 264.7 cells were harvested and cross-linked in 1.5% formaldehyde for 10 min at room temperature (RT) on the orbital shaker in order



to stabilize protein-DNA and protein-protein interactions. Formaldehyde was quenched adding 125 mM Tris-HCl pH 7.6 for 5 min. After three washes with cold PBS, cells were suspended in Lysis Buffer1 (see above) for 10 min at 4°C. After centrifugation the nuclear pellets were washed once with Washing Buffer and then re-suspended in CHIP Incubation Buffer. Chromatin from nuclei was sonicated for 20 cycles (30 sec “on” and 1 min “off”) in a cooled water bath sonicator Bioruptor UCD 300 plus (Diagenode). After sonication, 1% of Triton-100 was added to sonicated chromatin to pellet debris. Nucleosomes were solubilized through centrifugation at 13000 rpm for 10 minutes at 4°C, soluble fraction was then immunoprecipitated by adding 20 µg of the H3K4me1 (AbCam 8895) or Pu.1 (sc-352 X) antibody. The reaction was carried out overnight on a rotating wheel. In parallel, 200 µl of G protein-coupled magnetic beads (Dynabeads, Invitrogen 100.04D) for each sample, were blocked in BSA 0.5% PBS for an overnight. Blocked beads were washed and added to chromatin and incubated for 3 h at 4°C on the wheel. Beads were washed with 20 mM Tris-HCl pH 7.6, 2 mM EDTA, 0.1% SDS, 1% Triton-100 and increasing NaCl concentration (150 and 300 mM). To reverse the crosslinking and elute the immunoprecipitated proteins, SDS-PAGE sample buffer (250 mM Tris-HCl pH 8.8, 0.5M β-mercaptoethanol, 2% SDS) was added to the beads for 25 min at 95 °C. Proteins were resolved on 4-12% Bis-Tris acrylamide SDS-PAGE pre-cast gradient gels Invitrogen system and visualized by Colloidal Coomassie staining kit (Invitrogen).

*Buffers for X-ChIP:*

1. *Lysis Buffer1*: 50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton-100, 0.5 mM PMSF, 5 mM NaF, 5 mM Na<sub>3</sub>VO<sub>4</sub>, 5mM NaButyrate, 5 mg/ml Aprotinin, 5 mg/ml Pepstatin A, 5 mg/ml Leupeptin.

2. *Washing Buffer*: 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.5 mM PMSF, 5 mM NaF, 5 mM Na<sub>3</sub>VO<sub>4</sub>, 5mM NaButyrate, 5 mg/ml Aprotinin, 5 mg/ml Pepstatin A, 5 mg/ml Leupeptin.

3. *ChIP Incubation Buffer*: 10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% sodium lauroylsarcoside, 0.5 mM PMSF, 5 mM NAF, 5 mM Na<sub>3</sub>VO<sub>4</sub>, 5mM NaButyrate, 5 mg/ml Aprotinin, 5 mg/ml Pepstatin A, 5 mg/ml Leupeptin.

## 4.6 Native Chromatin immunoprecipitation (N-ChroP)

### 4.6.1 Native chromatin immunoprecipitation (N-ChIP) followed by MNase digestion

Eighty millions RAW 264.7 cells were homogenized in Lysis Buffer2 and nuclei were separated from cytoplasm, by centrifugation at 3750 rpm (4°C) for 30 minutes, putting cellular lysate on sucrose cushions. Nuclear pellets were washed in PBS, re-suspended in Digestion Buffer and digested with micrococcal nuclease (MNase, Roche) at a final concentration of 0.005 U/ml, at 37 °C for 70 minutes. The reaction was stopped by adding 1 mM EDTA and chilling on ice. The soluble fraction of chromatin (S1), enriched in mono-nucleosomes, was collected as the supernatant obtained after centrifugation of re-suspended nuclei at 10000 rpm (4 °C) for 10 minutes. DNA extracted by Qiaquick columns (QIAGEN) was run on 1% agarose gel to evaluate fractions of chromatin. The S1 fraction was used as input for the immunoprecipitation with 10 µg of H3K4me1 (AbCam 8895). Antibody was incubated overnight with chromatin; in parallel, 100 µl of G protein-coupled magnetic beads (Dynabeads, Invitrogen 100.04D) were blocked in BSA 0.5% PBS for an overnight. Blocked beads were washed and added to chromatin and incubated for 3 h at 4°C on a rotating wheel. Beads were washed four times (50 mM Tris-HCl pH 7.6, 10 mM EDTA) at increasing salt concentration (75, 125 and 175 mM NaCl). LDS Sample

Buffer (Invitrogen NP0007) supplemented with 50 mM DTT was added to the beads for 5 min at 70 °C to elute the immunoprecipitated proteins from the beads. Proteins were resolved on 4-12% Bis-Tris acrylamide SDS-PAGE pre-cast gels (Invitrogen NP0335BOX) on an Invitrogen system and visualized on the gel using Colloidal Coomassie staining Kit (Invitrogen LC6025).

*Buffers for N-ChIP:*

1. *Lysis Buffer*<sup>2</sup>: 0.32 M saccharose, 0.5 mM EGTA pH 8.0, 15 mM NaCl, 60 mM KCl, 15 mM HEPES, 0.5% Triton, 0.5 mM PMSF, 1 mM DTT, 5 mM NaF, 5 mM Na<sub>3</sub>VO<sub>4</sub>, 5mM Na-butyrate, 5 mg/ml Aprotinin, 5 mg/ml Pepstatin A, 5 mg/ml Leupeptin.
2. *Digestion Buffer*: 0.32 M saccharose, 50 mM Tris-HCl pH 7.6, 4 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub>, 0.1 mM PMSF.
3. *Dialysis Buffer*: 10 mM Tris-HCl pH 7.6, 1 mM EDTA, 0.5 mM PMSF, 5 mM NaF, 5 mM Na<sub>3</sub>VO<sub>4</sub>, 5mM NaButyrate and protease inhibitors cocktail (complete EDTA-free, Roche, Cat. No. 11 873 580 001).

*4.6.2 Cross-linked chromatin immunoprecipitation (X-ChIP) followed by MNase digestion and sonication to use a transcription factor as bait in ChroP experiment*

Chromatin was first fixed with formaldehyde at a final concentration of 0.75% on an orbital shaker, the reaction was then quenched by adding 125 mM Tris-HCl pH 7.6 for 5 min. Cells were washed twice with cold PBS and centrifuged at 1000 rpm for 5min. Pellet from one hundred million cells was first resuspended in 5 mL of Buffer A, then one volume of Lysis buffer<sup>3</sup> was added and the sample was put in ice for 5 min. After centrifugation the nuclear pellet was washed with 10mL of buffer B to remove any residual detergent, then resuspended into 500 uL of MNase digestion buffer and digested with 6 U MNase (Roche, Catalog No. 10107921001) for about 1h at 37 °C in order to obtain a

fraction enriched in mono-nucleosomes. Reaction was stopped with 50 mM EDTA, DNA was then de-crosslinked at 65 °C for some hours and loaded onto 1% agarose gel. Three mL of lysis buffer 4 were added to the lysate and 1.5mL aliquots of the input were sonicated for 3 pulses (30 sec “on” and 1 min “off”) in a cooled water bath sonicator Bioruptor UCD 300 plus (Diagenode) to break the nuclei. NP-40 1% final was added and the sample was then pelleted at 13000 rpm for 10 min. The supernatant was then used as input for the immunoprecipitation with Pu.1 antibody (sc-352 X); the reaction was carried out overnight as already described for X-ChroP. Blocked beads were washed and added to chromatin and incubated for 3 h at 4°C on the wheel. Beads were washed with 20 mM Tris-HCl pH 7.6, 2 mM EDTA, 0.1% SDS, 1% NP40 and increasing NaCl concentration (150 and 300 mM). Protein decrosslinking - likewise the elution of immunoprecipitated proteins, the SDS-PAGE separation and the Coomassie staining of the gel, were carried out similarly as for X-ChroP (see paragraph 4.5).

*Buffers for X-ChIP + MNase and sonication:*

1. *Buffer A*: 15 mM NaCl, 15 mM Tris-HCl pH 7.6, 60 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 300 mM saccharose, 1 mM DTT, 0.2 mM spermine, 1 mM spermidine, 0.5 mM PMSF.
2. *Lysis buffer3*: 15 mM NaCl, 15 mM Tris-HCl pH 7.6, 60 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 300 mM saccharose, 0.4% NP40, 1 mM DTT, 0.2 mM spermine, 1 mM spermidine, 0.5 mM PMSF.
3. *Buffer B*: 15 mM NaCl, 15 mM Tris-HCl pH 7.6, 60 mM KCl, 300 mM saccharose, 1 mM DTT, 0.2 mM spermine, 1 mM spermidine, 0.5 mM PMSF.
4. *MNase digestion buffer*: 20 mM Tris-HCl pH 7.6, 5 mM CaCl<sub>2</sub>, 1 mM DTT, 0.2 mM spermine, 1 mM spermidine, 0.5 mM PMSF.

5. *Lysis Buffer 4* : 10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate, 0.5% Na-laurylsarcosine, 1 mM DTT, 0.2 mM spermine, 1 mM spermidine, 0.5 mM PMSF.

#### 4.7 Protein co-immunoprecipitation from nuclear soluble extract

RAW cells were collected, washed twice with cold PBS 1% and resuspended in Lysis buffer<sup>5</sup>. After 5 min on ice and centrifugation the pellet was re-suspended into a Nucleosol extraction buffer. The supernatant was diluted to decrease NaCl concentration for the subsequent immunoprecipitation. After overnight incubation of the antibody with the nuclear extract, 100 µl of G protein-coupled magnetic beads (Dynabeads, Invitrogen 100.04D), previously blocked with BSA 0.5% in PBS and washed three times with PBS were added and incubated for 3h on a rotating wheel at 4°C. After incubation, beads were pelleted, washed for four times (50 mM Tris-HCl pH 7.6, 10 mM EDTA) at increasing salt concentration (75, 125 and 175 mM NaCl) and finally incubated with LSD Sample Buffer -supplemented with 50 mM DTT- for elution of the co-immunoprecipitated proteins, which were then either resolved on a 4-12% Bis-Tris acrylamide SDS-PAGE pre-cast gel and visualized with Colloidal Coomassie staining Kit, or loaded on hand-made linear SDS-PAGE gels for western blot immunodetection.

*Lysis buffer<sup>5</sup>*: 2 mM EDTA pH 8.0, 50 mM Tris-HCl pH 8.0, 0.1% NP-40 and 10% glycerol).

*Nucleosol extraction buffer*: 250 mM NaCl, 0.2 % NP-40, 50 mM Tris-HCl, 0.5 mM EGTA and 0.5 mM EDTA

#### 4.8 In-gel chemical alkylation and digestion of histones prior to MS analysis

This protocol was specifically modified in order to maximize histone protein coverage and hPTMs detection (Bonaldi, Imhof et al. 2004, Soldi and Bonaldi 2013).

Bands corresponding to the core histones were excised from the gel, de-stained with repeated washes in 50% acetonitrile (ACN) in ddH<sub>2</sub>O, alternated with dehydration steps in 100% ACN. Gel pieces were in-gel chemically alkylated as previously described, by incubation with D<sub>6</sub>-acetic anhydride (Sigma 175641) 1:9 in 1 M NH<sub>4</sub>HCO<sub>3</sub> and CH<sub>3</sub>COONa solution as catalyzer (Bonaldi, Imhof et al. 2004). After 3h at 37 °C with strong shaking in a thermo mixer, gel slices were washed with buffer at increasing ACN concentration (50% and 100% ACN in water). In-gel digestion was performed with 100 ng/μl trypsin (Promega V5113) in 50 mM NH<sub>4</sub>HCO<sub>3</sub> at 37 °C overnight; the alkylation of lysine residues in combination with trypsin digestion generates an “in-gel” ArgC-like digestion because the enzyme will cleave only at the amide bond C-terminal of arginine residues. This digestion produces histone peptides with optimal length for MS analysis. Digested peptides were extracted, desalted and concentrated using a combination of reverse-phase C18/Carbon “sandwich” system and strong cation exchange (SCX) chromatography, on hand-made nano-columns (StageTips) (Rappsilber, Mann et al. 2007): digested peptides loaded on C<sub>18</sub>/C and SCX StageTips were then eluted with high organic solvent (80% ACN) and NH<sub>4</sub>OH, respectively. Eluted peptides were lyophilized, re-suspended in 0.1% TFA in ddH<sub>2</sub>O, pooled and subjected to LC-MS/MS.

#### 4.9 Standard protein in-gel digestion for MS-based protein identification

Processing of gel-separated proteins prior MS analysis was carried out as previously described, with minor modifications (Shevchenko, Tomas et al. 2006). About ten gel slices per lane were cut from gel and de-stained in 50% v/v acetonitrile (ACN)/50 mM NH<sub>4</sub>HCO<sub>3</sub>. Cysteine residues were reduced with 10 mM DTT in 50 mM NH<sub>4</sub>HCO<sub>3</sub> and were alkylated with 55 mM iodoacetamide in 50 mM NH<sub>4</sub>HCO<sub>3</sub> to avoid disulphide bond formation. Proteins were subjected to in-gel digestion with 12.5 ng/μL trypsin (Promega

V5113) in 50mM  $\text{NH}_4\text{HCO}_3$  for overnight at 37 °C. Digested peptide were first extracted from the gel with 3% TFA, 30% ACN and finally with 100% ACN; peptides were then lyophilized, desalted and concentrated on  $\text{C}_{18}$  Stage Tips (Rappsilber, Mann et al. 2007). Samples were loaded in 1% TFA and 5% ACN and eluted with high organic solvent (80% ACN). Eluted peptides were lyophilized, re-suspended in 0.1 % TFA in ddH<sub>2</sub>O and subjected to LC-MS/MS analysis.

#### 4.10 Nano Liquid Chromatography and Tandem Mass Spectrometry

Two different methods were adopted for either proteome analysis of protein interactors or for the MS-analysis of histone peptides, indeed the latter takes into account the peculiar chemical-physical properties of these proteins such as the basic composition of the peptides. To this aim different gradients were employed in the run.

Peptide mixtures were analyzed by online nano-flow liquid chromatography tandem mass spectrometry (LC-MS/MS) using an EASY-nLC™ 1000 (Thermo Fisher Scientific, Odense, Denmark) connected to a Q Exactive (Thermo Fisher Scientific) through a nano electrospray ion source. For nUHPLC, the nano LC system was operated in one column set-up with a 25 cm analytical column (75  $\mu\text{m}$  inner diameter, 350  $\mu\text{m}$  outer diameter) packed with  $\text{C}_{18}$  resin (ReproSil, Pur  $\text{C}_{18}\text{AQ}$  1.9  $\mu\text{m}$ , Dr.Maisch, Germany) configuration.

For the analysis of peptides from interactome studies solvent A was 0.1% FA, 2% ACN in ddH<sub>2</sub>O and solvent B was 80% ACN with 0.1% FA. Samples were injected in an aqueous 1% TFA solution at a flow rate of 500 nl/min and they were separated with a gradient of 5-30% solvent B over 90 min followed by a gradient of 30-60% in 5 min and 60-95% over 5 min at a flow rate of 250 nl/min in the EASY-nLC 1000 system.

For the analysis of histone peptide the solvent A was 0.1% FA in ddH<sub>2</sub>O and solvent B was 80% ACN with 0.1% FA. These samples were injected in an aqueous 1% TFA solution at a flow rate of 500 nl/min and they were separated with a gradient of 0-40% solvent B over

100 min followed by a gradient of 40-60% in 5 min and 60-95% over 5 min at a flow rate of 250 nl/min in the EASY-nLC 1000 system.

The Q Exactive instrument was operated in the data-dependent mode (DDA) to automatically switch between full scan MS and MSMS acquisition. Survey full scan MS spectra (from  $m/z$  300-1150) were analysed in the Orbitrap detector with resolution  $R=35,000$  at  $m/z$  400. The ten most intense peptide ions with charge states  $\geq 2$  were sequentially isolated to a target value of  $3e^6$  and fragmented by Higher Energy Collision Dissociation (HCD) with a normalized collision energy setting of 25%. The maximum allowed ion accumulation times were 20 ms for full scans and 50 ms for MSMS while the target value for MSMS was set to  $1e^6$ . The dynamic exclusion time was set to 20s. Standard ESI conditions for all experiments were: spray voltage, 2.4 kV; no sheath and auxiliary gas flow.

#### 4.11 Analysis of Mass Spectrometric data

Acquired Raw data were analyzed by the integrated MaxQuant software v.1.3.0.5, using the integrated Andromeda search engine (Cox, Neuhauser et al. 2011). This software was designed for the analysis of experiments performed in SILAC, therefore was optimized for the quantification SILAC pairs. The MOUSE 1401 database (51195 entries) was used for peptide identification. Enzyme specificity was set to Arg-C for in-gel chemical alkylation followed by trypsin digestion (“Arg-C-like” mode), while for the interactome studies we set trypsin as enzyme which cleaves at the C-terminus of lysine and arginine residues unless they are followed by proline residue. In MaxQuant, we fixed the estimated false discovery rate (FDR) of all peptide identifications at a maximum of 1%. The main search was performed with a mass tolerance of 6 ppm. A maximum of 3 missed cleavages were permitted, and the minimum peptide length was fixed at 6 amino acids.



#### 4.11.1 Identification and quantitation of hPTMs

In the search, variable modifications included: lysine D3-acetylation (+45.0294 Da), lysine mono-methylation (+59.045 Da) from the sum of D3-acetylation mass (+45.0294) plus the mass of mono-methylation (+14.016 Da)], arginine mono-methylation (+14.016 Da), lysine and arginine di-methylation (+28.031 Da), lysine tri-methylation (+42.046 Da), lysine acetylation (+42.010 Da), lysine crotonylation (+68.026 Da) and lysine ubiquitination (+118.127 Da or +383.228 Da). MaxQuant search results were exported and peptides with Andromeda score less than 60 or localization probability score less than 0.75 were removed, as identified with low-confidence.

MaxQuant is not an unbiased search tool as the user selects a set of PTMs of interest prior to employ the bioinformatic tool for peptide and protein ID. This option is applied during the sequence database search, when PTMs are assigned to the relevant amino acid of a candidate peptide sequence. Histones are characterized the co-occurrence of different modifications on the same peptide, but the increase of the number of modifications in the search enhances the complexity of the search itself; for this reason for very large datasets the search has to be restricted to a limited number of PTM. In our searches we used a maximum of five different modifications in each search and we run different MaxQuant jobs in parallel in order to cover theoretically all the possible different combinations among them. For peptides from ArgC-like digestion, we applied as constant modification the D3-acetylation on lysine residues in every MaxQuant job.

Filtered data were subjected to manual inspection and validation using QualBrowser version 2.0.7 (ThermoFisher Scientific) where theoretical mass of a specific modification and its MS profile are used to reconstruct the eXtracted Ion Chromatogram (XIC), while MS/MS spectrum is used for the robust identification and the assignment of the modification. XICs were constructed for each precursor based on the m/z value, using a mass tolerance 10 ppm and a mass precision up to 4 decimals. Histone PTMs were first

quantified by calculating the area under the curve (AUC) of each peak corresponding to every specific modified peptide. Then, their relative abundance was estimated by dividing the area under the curve (AUC) of each peptide by the sum of the areas corresponding to all observed modified forms of that peptide, in percentage as already described in (Jung, Pasini et al. 2010, Soldi and Bonaldi 2013). For the same modifications, we also calculated the relative enrichment as the ratio between the relative abundance of the same modification in the ChIP and in the corresponding chromatin input.

#### 4.11.2 Visualization of hPTMs

hPTMs in basal state we represented through barr charts where we divided the analysis in H3, H4 and H2A peptides generated upon ArgC-like digestion. We expressed the log<sub>2</sub> of the relative enrichment for all the modifications robustly identified on the same peptide.

Visualization of hPTMs clustering was performed using Perseus program (<http://www.perseus-framework.org/>). Here we represented in a heat map the log<sub>2</sub> value of the different hPTMs robustly identified in the histone H3 and H4 peptides ordered according to peptide and histone they belong to. Maximum and minimum relative enrichment values are used as extremes in the colour code.

#### 4.12 Analysis of SILAC data for interactome and dynamic study

Carbamidomethylation of cysteine was selected as a fixed modification whereas oxidation of methionine, and acetylation of protein N-terminus were included as variable modifications. The modifications corresponding to arginine and lysine labeled with heavy and medium stable isotopes were treated as fixed modifications in the Andromeda search. Additional peptides were identified by "the match between run" option in MaxQuant, which matches precursor masses in a 2-min retention time window (after realignment of the runs) based on the accurate mass measurement. Proteins were accepted if identified

both in forward and reverse experiments with two peptides, at least one of which unique. For the quantification, we accepted proteins with at least 1 ratio count ( $RC > 0$ ) in forward and reverse experiments. Proteins that pass these filtering criteria in both H3K4me1 and Pu.1 ChIPs were subjected to statistical analysis and visualization using the open-source software package R and Perseus software (<http://www.perseus-framework.org/>). Based on the SILAC ratio distribution generated through R-script, we defined the statistical threshold to distinguish between real binders and background proteins for forward and reverse experiments; the two dataset were then visualized in a scatter plot (H/L ratio and L/H ratio for forward and reverse experiment, respectively) where we identified as specific binders only those proteins significantly enriched in both experiments (tops) that are represented in the upper right quadrant. To define the genuine enhancers' binders we intersect the tops subset of proteins from H3K4me1 and Pu.1.

For the dynamic analysis we generate SILAC ratio distribution plot for every time point (1hr/UT, 4hr/UT) and for both the experiments (forward and reverse) performed with the two baits, namely H3K4me1 and Pu.1. For all these ratio distributions we build a theoretical distribution of 90% of the entire set which correspond to proteins that are supposed to do not change upon the stimulus. We calculated the standard deviations (s.d.) of these 90% distributions e we applied them to the corresponding entire datasets to define proteins recruited (SILAC ratio  $> +s.d.$ ) or evicted (SILAC ratio  $< -s.d.$ ) in that specific time point. We profiled in time the list of genuine enhancers' determinants and we defined as recruited or evicted only those proteins that showed that specific trend in at least 3 out of 4 dataset (H3K4me1 forward, H3K4me1 reverse, Pu.1 forward, Pu.1 reverse) thus taking into account different kinetics of recruitment of eviction between the two baits used in the ChIP.

The list of newly recruited proteins at enhancers is composed of proteins showing a positive trend in at least 3 out of 4 abovementioned datasets taking into account different

kinetics of recruitment between H3K4me1 and Pu.1.

#### 4.13 Quantitative RT-PCR of immunopurified DNA

DNA from CHIP-ed material was eluted in TE (Tris-HCl pH 7.5, EDTA) containing 2% SDS for 15 minutes at 65°C (for X-CHIP experiments DNA was also de-crosslinked at 65°C overnight) and DNA was then purified through Qiaquick columns (QIAGEN). 1 µl of purified DNA was used for substrate for amplification on Applied Biosystems 7500 Fast Real-time PCR system applying Biosystem Sybr-green.

#### 4.14 RNA extraction and cDNA synthesis

Total RNA was extracted using RNeasy kit (Qiagen), according to manufacturer instruction. RNA was quantified by ND-1000 spectrophotometer (NanoDrop Technologies) and its quality was assessed by measuring A260/A280 and A260/A230 ratios.

Complementary DNA (cDNA) was obtained by reverse transcription with the following protocol: 0.5 µg of RNA were mixed with 0.5 µg/reaction of Oligo(dT) and 0.5 µg/reaction of Random primers into a final volume of 5 µL. Then were added 4 µL of 5x Reaction buffer, 1.5 mM final concentration of MgCl<sub>2</sub>, 0.5 mM/dNTP of dNTP mix, and 1 µL of Reverse Transcriptase, in a total volume of 20 µL. Enzymatic reactions were performed in the shaking thermoblock: 10 min 25°C (annealing), 60 min 42°C (elongation) and 5 min 95°C (enzyme inactivation).

#### 4.15 Immunoblot analysis

Input chromatin and immuno precipitated histone octamers were separated in 12% or 17.5% SDS-PAGE and transferred to PVDF membranes (Millipore). Membranes were blocked 1 h in 5% BSA in TBS-T. After blocking, membranes were incubated at 4°C for

an overnight with primary antibodies diluted in TBS-T 5% BSA. After three washes in TBS-T, binding was revealed by ECL western blot detection reagent and hyperfilm ECL (GE Healthcare) or by Clarity western blot ECL Substrate and Chemidoc XRS with Image Lab software (Biorad). For Western blot analysis the following antibodies were used, according to the manufacturer's instructions: H3K4me1 (Abcam, dilution 1:2000), Pu.1 (Santa Cruz Biotechnology, dilution 1: 1000), H3K27Ac (Active Motif, 1:2000), Mpeg1 (Genetex, dilution 1:1000), MafG (SantaCruz Biotechnology, dilution 1:1000), H3K9me3 (Abcam, dilution 1:2000), vinculin (Upstate Millipore 06-866, dilution 1:10000), H3 (Abcam, dilution 1:5000).

#### 4.16 Retroviral vector oligonucleotide cloning

##### *4.16.1 Retroviral Constructs and Production of Retroviruses*

All shRNA constructs were prepared in the MSCV-based pLMP retroviral vector (Dickins, Hemann et al. 2005). The sequences of the shRNAs used in the experiments are:

- KD1-Mpg1: GCTTAATTGTCTCCAAAGGAAA
  
- KD2-Mpg1: TGCATTTAGTCTGAATTTGAAA

Phoenix-ECO packaging cells were plated into producer cell growth media (DMEM, 10% heat-inactivated FBS, 2mM glutamine, 100 U/ml penicillin and 100 mg/ml streptomycin) co-transfected with 2.8 µg pCLEco and 10 µg of either LMP empty vector or the same vector carrying the directed shRNA sequences, using the CaCl<sub>2</sub> 12.2 mM and Hebes Buffered Saline (HBS). After 15 min 20 µM final of cloroquine was added to inhibit lysosomal degradation of exogenous DNA. After few hours the medium of phoenix-ECO packaging cells was replaced with target cell medium. Supernatants from transfected Phoenix-Eco were collected at 48 hours post-transfection and used for BM infections.

#### *4.16.2 Retroviral Transduction of Bone Marrow-Derived Macrophages*

On day 0, bone marrow cells were isolated and  $3 \times 10^6$  cells/dish were seeded on 10 cm plates in BM medium. Infection was performed on day 1. In details, cells were first spun at 2500 rpm for 5 min, the BM medium was removed, and the retroviral supernatants, supplemented with 8  $\mu\text{g/ml}$  polybrene and 8  $\mu\text{l/ml}$  of 1 M HEPES pH 7.5, were added. After 12 hours the virus-containing medium was gently removed and replaced by fresh BM medium. Puromycin selection (3  $\mu\text{g/ml}$ ) started on day 3. Assays were carried out at days 7–8.



## RESULTS

### 5.1 Employment of X-ChroP approach in RAW 264.7 cells

For the interactome dissection, we needed to grow the cells in the appropriate medium for seven to nine replications (see below), thus we used as model system a cell line, namely the murine macrophage cell line RAW 264.7. These cells were obtained by transformation of mouse macrophages from male BALB/c mice with Abelson murine leukemia retrovirus (Ab-MLV) and Moloney MLV (M-MLV) as helper virus (Raschke et al. 1978). RAW 264.7 cells are macrophage-derived, therefore they respond to inflammatory stimuli like interferon-gamma (IFN- $\gamma$ ) and lipopolysaccharide (LPS), mounting inflammatory response and producing pro-inflammatory cytokines. The LPS-induced activation of RAW 264.7 cells is a very well characterized mechanism thus making this macrophage-like cell line the most widely employed model to study the inflammatory response (Chang et al 2010, Shih et al 2010, Min et al 2012, Ying et al 2013). Two complementary processes characterize their activation: macroscopically, they stop growing and acquire a macrophage-like morphology characterized by the formation of the cellular protrusions; at the molecular level they activate different pathways, such as the toll-like receptor (TLR) pathway and the tumor-necrosis factor (TNF) receptor pathway, that in turn elicit a massive transcriptional activation of inflammatory genes, like those coding for chemokines and cytokines.

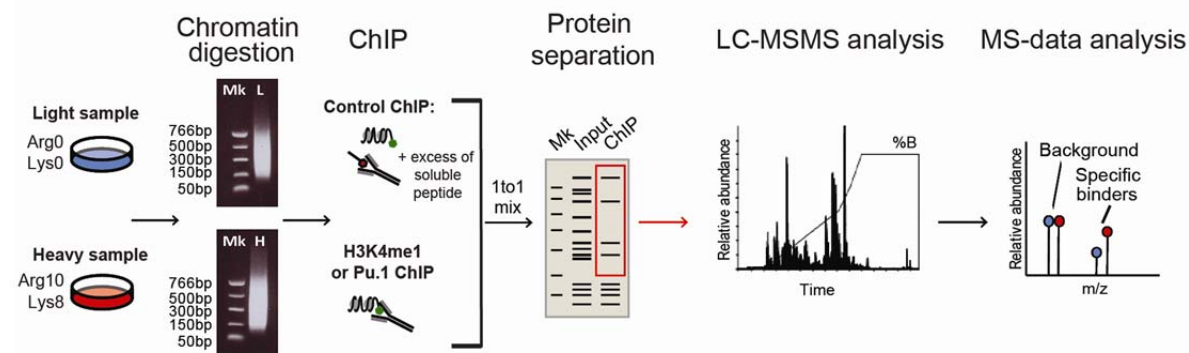
In order to dissect the enhancers' determinants at both basal and inflamed states, we set up the Chromatin Proteomics (ChroP) approach, recently developed in our laboratory (Soldi and Bonaldi 2013), in this model system.

The cross-linked ChroP (X-ChroP) approach has been developed to dissect the interactors and histone variants specifically associated with a chromatin region of interest, the so-called chromatin interactome. Following this approach, a specific histone PTM is



used as bait in the chromatin immunoprecipitation (ChIP) and the proteins co-enriched together with the chromatin stretch are analysed by mass spectrometry (MS) (Fig.14).

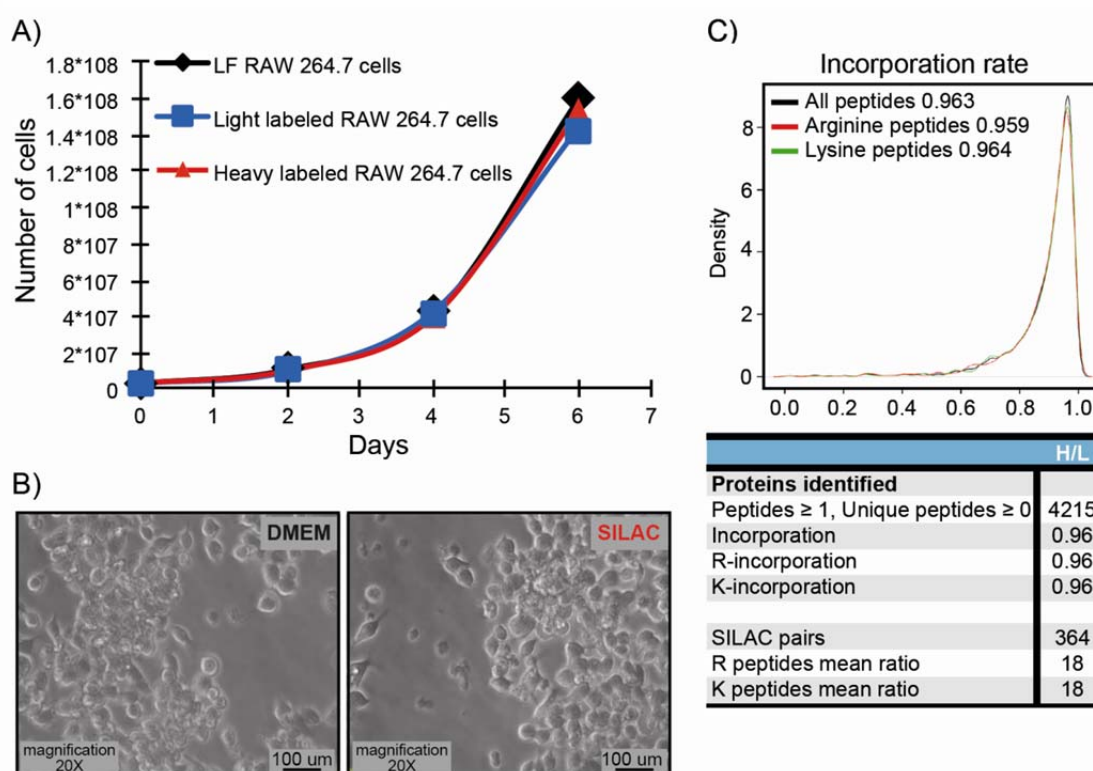
The X-ChroP approach takes advantage of SILAC metabolic labeling in order to discriminate between the background proteins and the specific interactors: chromatin is prepared from heavy- and light- labeled cells and fragmented in nucleosome stretches of chosen length by sonication; poly-nucleosomes are used as input for the ChIP. In the immunoprecipitation step, the antibody against the modification used as bait is added in both light and heavy SILAC channels, whereas in only one of the two chromatin preparations (es. light channel in the forward experiment) an excess of soluble peptide bearing the epitope is added, to compete the native mark and displace it from the antibody. Upon H/L cells mixing in 1-to-1 ratio, proteins are eluted from the beads and analysed by LC-MS/MS. Due to the competition with the soluble peptide, the bait is displaced from the antibody together with all the co-associated interactors, thus generating a specific SILAC ratio different from 1 ( $H/L > 1$  in the forward experiment). The binding of the unspecific contaminants is instead unaffected by the competition, leading to a SILAC H/L ratio equal to 1 (Fig. 14).



**Fig. 14: Experimental design of the SILAC-based X-ChroP strategy.** Purified nuclei from SILAC-labeled and formaldehyde fixed RAW 264.7 cells are fragmented through sonication and the chromatin inputs are subjected to ChIP. An excess of soluble peptide displaces partially the antibody in the light SILAC channel so that, upon mixing in 1to1 H/L ratio, separation of eluted proteins *via* SDS-PAGE and LC-MS/MS analyses, interactors will show a high H/L SILAC ratio while contaminants will have an H/L ratio equal to 1. (Light = blue, heavy = red).

### 5.1.1 SILAC labeling of RAW 264.7 cells

We cultured RAW 264.7 cells in parallel, in normal and in SILAC DMEM medium to assess whether the different composition between them may affect cell growth. The SILAC medium contains heat-inactivated and dialyzed fetal bovine serum (FBS), to ensure that the isotope-coded light (Arg0, Lys0) and heavy (Arg10, Lys8) amino acids added represent the only source of the amino acids used for new protein synthesis. Growth curves of RAW264.7 cells cultured in SILAC medium containing light or heavy amino acids are



**Fig. 15: Establishment of the SILAC labeling in RAW 264.7 cells.** **A)** Comparison of the growth curves between RAW 264.7 cells cultured in DMEM (black) or SILAC medium (both light and heavy isotope coded media, blue and red, respectively). **B)** Evaluation of the morphology in RAW 264.7 cells cultured in DMEM or in SILAC medium through phase contrast microscopy. **C)** Incorporation rate of RAW 264.7 cells grown in heavy SILAC medium (Arg10, Lys8). Table summarizes the number of peptides identified and the incorporation percentages of arginine- or lysine-containing peptides.

identical to the one of cells cultured in normal DMEM, thus indicating that the growth is not affected by the poorer composition of the SILAC medium (Fig. 15A).

To assess whether the SILAC serum induces an inflammatory state, we evaluated the morphology of RAW 264.7 cells cultured in both conditions, which represent an early

prognostic feature of macrophage activation (Fig. 15B). We did not appreciate any significant change in the morphology of the cells culture in the two conditions, thus we concluded that RAW 264.7 cells are not induced or activated by the SILAC serum.

We cultured RAW 264.7 cells in heavy SILAC medium for eight replications to determine the incorporation rate; upon cell lysis, we separated the total extract on SDS-PAGE gel and evaluated by quantitative MS the percentage of heavy peptides from 2 slices of the gel lanes. We identified 4215 peptides in total and estimated that the incorporation rate of arginine- and lysine- containing peptides was 96 and 96.4%, respectively. In fact, we identified 10 peptides existing in the only-light form, which corresponds the 0.23% of the whole peptide population; the peptides present as H/L doublets were 364, corresponding to about 8% and their average SILAC H/L ratio was 18 (Fig. 15C), while all remaining 3843 peptides were present only in the heavy-labeled form, accounting for 92% of the total. Altogether these results demonstrate that we could reach an almost complete incorporation of heavy Arg10 and Lys8 in RAW 264.7 cells.

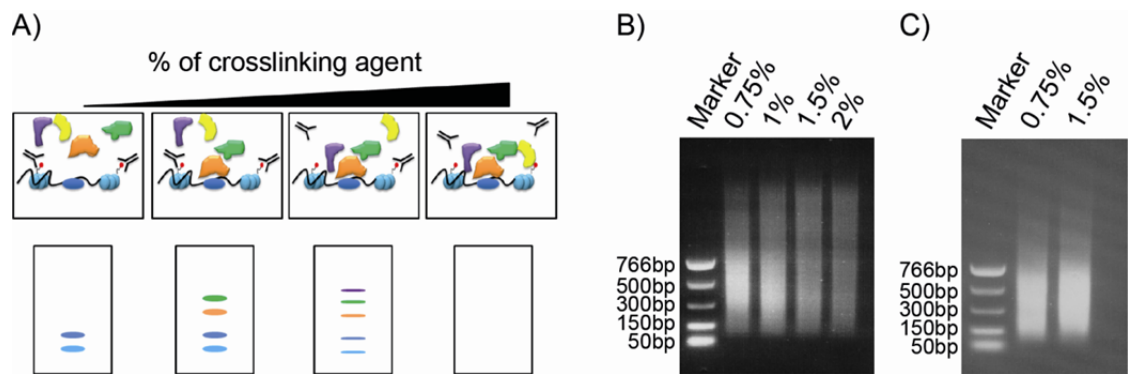
### *5.1.2 Optimization of chromatin cross-linking conditions for the characterization of inflammatory genes' enhancers*

We carried out ChIPs for ChroP experiments with two well-described markers of enhancers: one is a histone post-translational modification namely the mono-methylation of lysine 4 of histone H3 (H3K4me1) and the other is a transcription factor and macrophage master regulator Pu.1, which binds specifically to enhancers of inflammatory genes.

The cross-linking conditions and the length of the DNA stretch are the two critical aspects of the protocol for the preparation of the chromatin input for ChIP.

Formaldehyde is used in ChIP to stabilize DNA-protein and protein-protein interactions (PPI): while too low concentration of formaldehyde only traps the more stable interactions,

but missing the more dynamic ones, an excess of formaldehyde may mask the epitope. As such, an optimal concentration of this reagent should be defined to ensure the right balance between PPI stability and the efficiency of the immunoprecipitation of the bait (Fig. 16A). Also the length of the chromatin stretch must be defined precisely: we chose a stretch of 2 – 3 nucleosomes that corresponds to an overall DNA length between 300bp and 500bp.



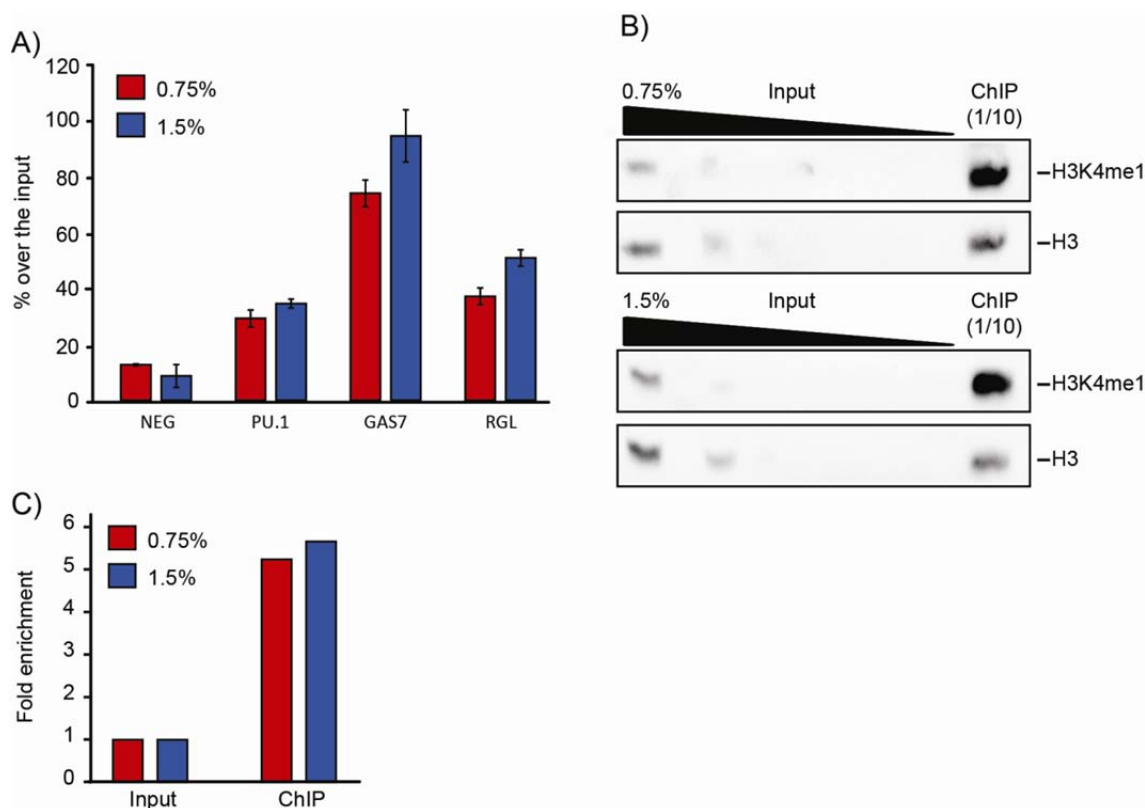
**Fig.16: Balance between formaldehyde cross-linking and DNA fragmentation to reach the optimal DNA length.** **A)** Schematic representation of the effect of the increase of cross-linking percentage on the ChIP efficiency. **B)** Comparison among DNA from 0.75%, 1%, 1.5% and 2% formaldehyde cross-linked chromatin upon 10, 15, 18 and 24 cycles of sonication (30sec ON/60secOFF), respectively. **C)** DNA from 0.75% and 1.5% of formaldehyde cross-linked chromatin upon 10 and 20 cycles of sonication, respectively.

Such nucleosomal stretch can be used as input for both immunoprecipitations, since the transcription factor binds to the nucleosome-free region (NFR) in-between two nucleosomes.

We tested different concentrations of formaldehyde and we tried to balance the concentration with the number of sonication cycles needed to reach the chosen DNA length, taking into account that increasing sonication cycles can lead to overheating of the sample and protein denaturation.

As expected, at increasing formaldehyde concentrations, a higher number of sonication cycles were needed to reach a fixed and established nucleosome length (Fig.16B). We could obtain an optimal DNA length using 1.5% formaldehyde cross-linked chromatin with 20 cycles of sonication, thus we focused our investigation around this percentage of cross-linking agent (Fig.16C).

We analysed by qPCR the DNA from 0.75 or 1.5% formaldehyde cross-linked chromatin in the input and in ChIPs using H3K4me1 as bait. We then evaluated the fold enrichment of three genomic regions with enhancer function (GAS7, PU.1, RGL) and one hetero-chromatin region as negative control (NEG) in the ChIP and in the respective input, in order to test whether the increase of formaldehyde reduced the epitope accessibility (Fig. 17A). The three target regions were 2 to 5 fold more enriched in the H3K4me1 ChIP, normalised over the input, then the negative control; we also observed a slightly higher efficiency upon 1.5% formaldehyde cross-linking, suggesting that this experimental condition increases the binding and immunoprecipitation yield (Fig. 17).



**Fig. 17: Optimization of cross-linking percentages for H3K4me1 X-ChroP experiment.** **A)** qPCR measures the level of target genomic regions and negative region in the ChIP using 0.75% (red) or 1.5% (blue) formaldehyde cross-linked chromatin input. **B)** WB validation of H3K4me1 enrichment in the ChIP compared to the input, using 0.75% or 1.5% formaldehyde cross-linked chromatin; total histone H3 level is used as normalizer. **C)** Quantification of WB in B) obtained by calculating H3K4me1 enrichment in the ChIP over the input; total H3 used as normalizer for fold enrichment calculation [H3K4me1(ChIP/Inp)/H3(ChIP/Inp)].

We also measured the amount of immunoprecipitated H3K4me1 by western blot, estimating the enrichment of the signal for mono-methyl K4 normalized over the total H3

amount in ChIP and input (Fig. 17B). Fig. 17C shows that the enrichment of the bait is comparable between the experiment carried out with 0.75 or 1.5% formaldehyde cross-linking, confirming similar efficiency in the immunoprecipitation of the bait. We therefore carried out ChroP protocol with 1.5% formaldehyde with the aim of increasing the stabilization of more dynamic DNA-protein and protein-protein interactions.

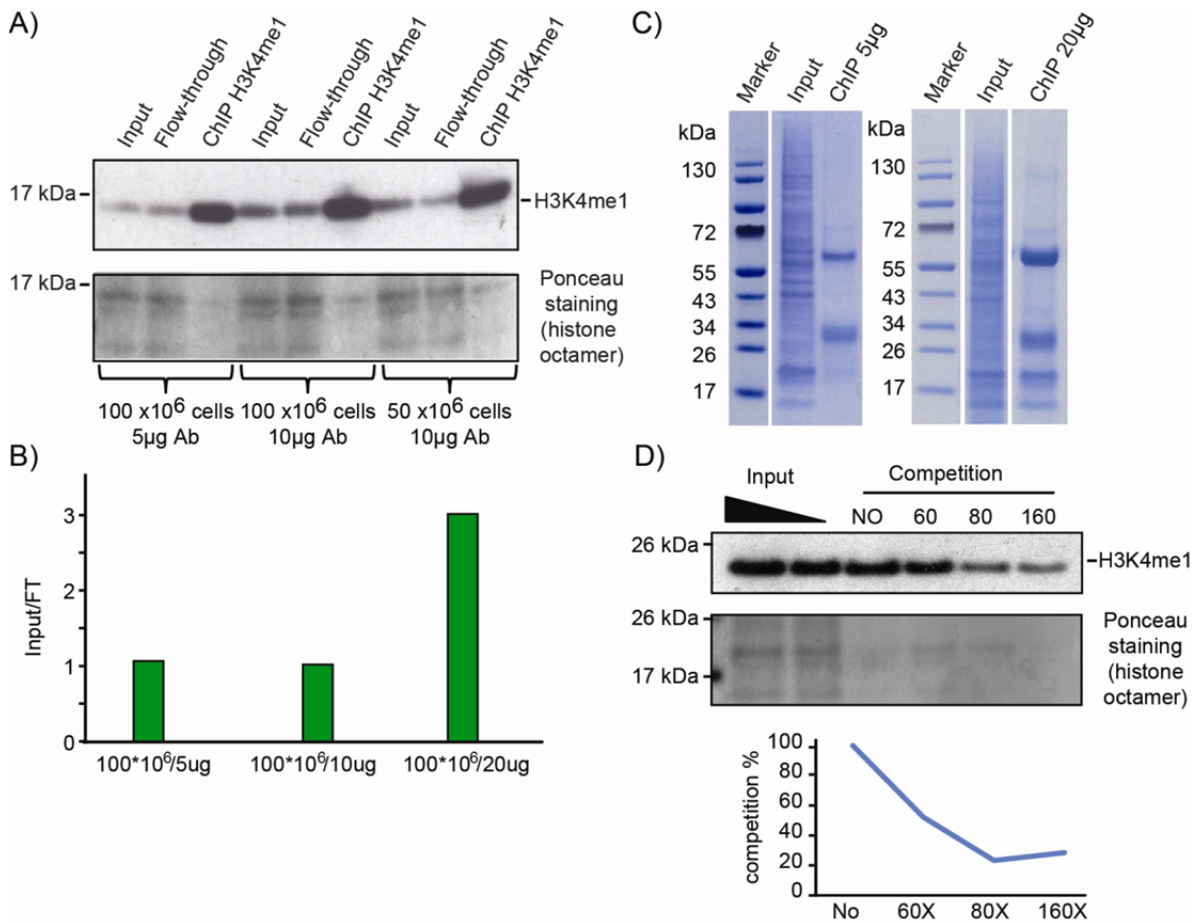
### *5.1.3 Optimization of the antibody/peptide ratio and of the peptide competition in the H3K4me1 X-ChroP*

We next optimized the ratio between chromatin input and the amount of antibody to be used in the ChIP in order to immunoprecipitate a representative proportion of the bait. We can extrapolate this by evaluating the depletion of bait in the unbound fraction (flow-through, FT). Chromatin isolated from 100 million RAW 264.7 cells was used as input in ChIPs carried out with increasing ratios of antibody/input, starting from 5 $\mu$ g of anti-H3K4me1 per 100 million cells up to 20 $\mu$ g (Fig. 18A). As expected, the increase of antibody/input ratio is mirrored by a progressive H3K4me1 depletion in the corresponding FT; in Fig. 18B the relative amount of bait in the input compared to the FT is plotted at increasing antibody/input ratio. It is worth mentioning that we immunoprecipitated only about 65% of H3K4me1 even when we use the highest antibody/input ratio; this is likely due to the relatively high abundance of this modification in bulk RAW 264.7 chromatin. Based on MS data, we estimated that about 20% of peptide H3 (3 – 8) is mono-methylated at lysine 4 (*see also below and paragraph 5.4.1*), therefore is quite difficult to completely deplete this modification; in any case, we consider that 65% of the total will allow achieving a good representation of the composition of H3K4me1-bearing chromatin.

We compared the lowest and the highest antibody/input ratio in a Coomassie-stained SDS-PAGE to assess whether the amount of histone octamer immunoprecipitated is sufficient

for the MS-analysis: the highest ratio yields an amount of stoichiometric histone octamer sufficient for the subsequent mass spectrometry investigation (Fig. 18C).

Based on all these tests we concluded that our optimal experimental conditions were the following: 80 million RAW 264.7 cells as starting material, 1.5% formaldehyde concentration and 20 $\mu$ g of anti-H3K4me1 antibody.



**Fig.18: Optimization of antibody/input ratio and peptide fold excess for H3K4me1 X-ChroP experiment.** **A)** H3K4me1 levels in input, flow-through and ChIP using different antibody/input conditions (chromatin from 100\*10<sup>6</sup> cells and 5 $\mu$ g of antibody, 100\*10<sup>6</sup> cells and 10 $\mu$ g of antibody or 50\*10<sup>6</sup> cells and 10 $\mu$ g of antibody); Ponceau staining of the histone octamer is shown as normalizer. **B)** WB quantification of flow-through (FT) depletion calculated as Input/FT in H3K4me1 and normalized on the same ratio in histone octamer ponceau staining. **C)** Coomassie staining of SDS-PAGE showing the ChIP efficiency using either 5 $\mu$ g (left) or 20 $\mu$ g (right) of H3K4me1 antibody on the same chromatin input. **D)** H3K4me1 levels in the inputs and in the ChIPs not competed (NO) or competed with 60, 80 or 160 fold molar excess of peptide bearing H3K4me1; Ponceau staining of histone octamers used as loading control for normalization of the bait PTM over the total histone amount.

In the X-ChroP experiment, the combination of the ChIP with a SILAC-based peptide competition assays allows to robustly distinguish between specific and unspecific chromatin interactors. A molar excess of a peptide bearing the epitope recognized by the antibody is added in one of the two SILAC-ChIP channels, to displace the bait with all co-associated proteins, thus generating a H/L ratio differing from 1; background proteins maintain a H/L ratio equal to 1. The excess of soluble peptide added in the light channel must be titrated to reach an optimal SILAC H/L ratio; the rationale is in fact to get the highest SILAC ratio for the bait without reaching the point of a complete loss of the signal in the competed channel, which would generate an infinite ratio. A high SILAC ratio enables to better discern the specific binders from background.

For the competition, we employed the same commercial peptide used as epitope to generate the antibody; we performed four ChIPs in parallel, using the same number of cells and the same amount of antibody but with increasing doses of soluble peptide: 60, 80, 160 fold molar excess (Fig. 18D). As expected, the bait is enriched in the non-competed ChIP compared to the input but the efficiency of the ChIP significantly drops when the excess of peptide is added. In particular, 60-fold peptide molar excess decreases the bait of about 50%, while with 80-fold molar excess the ChIP efficiency is about 20%. At 160-fold peptide molar excess, we abolished the binding of H3K4me1 in the ChIP. Hence, we chose 120-fold molar excess of the peptide as a compromise to ensure strong competition of the bait without abolishing completely the binding; this setting is expected to produce an optimal H/L SILAC ratio of the bait and co-associated proteins.

We thus performed a forward X-ChroP experiment using 80 million RAW 264.7 cells 1.5% formaldehyde cross-linked and 20 $\mu$ g of antibody against H3K4me1 for each heavy and light chromatin input, plus a 120-fold molar excess of soluble peptide in the light SILAC channel. After protein separation in SDS-PAGE, we excised the histone bands

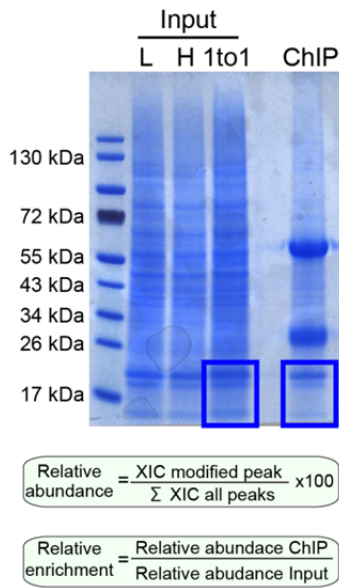


from both the H/L 1-to-1 mix input and from the ChIP (Fig. 19A) and processed them using an ArgC-like in-gel digestion protocol (*see material and methods for details*).

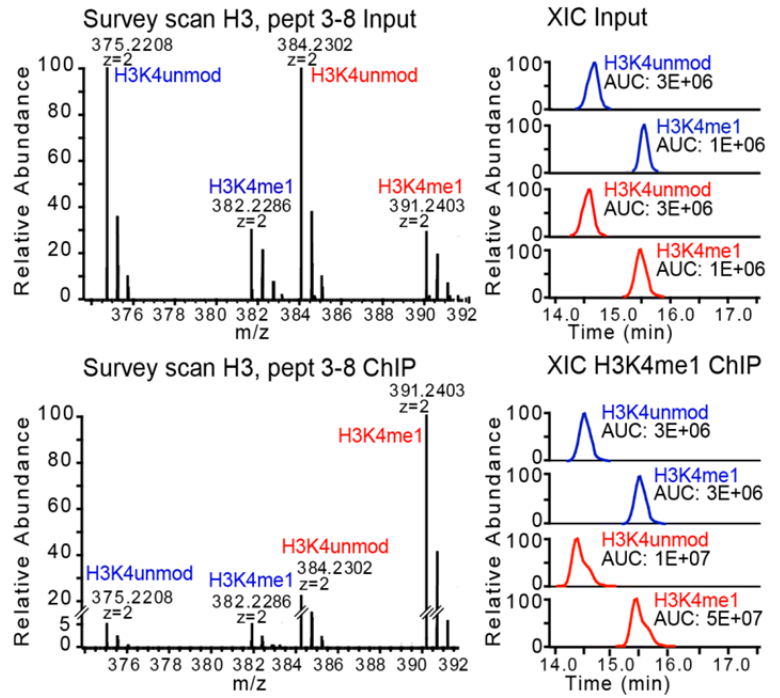
Upon LC-MS/MS analysis we focused the analysis on the peptide H3 (3 – 8) and calculated the extracted ion chromatography (XICs) for all the hPTMs on this peptide based on their exact mass, for both heavy and light sample, (*see also 5.4.1*) (Fig. 19B). We then calculated the percentage relative abundances (RA%) of each modification as the XIC of the specific modification over the sum of all the XICs of all the modifications on the same peptide expressed in percentage. Next, we estimated the relative enrichment (RE) of each modification as the ratio between its relative abundance in the ChIP and in the corresponding input.

As expected, the H3K4me1 RA% increases upon ChIP (80% vs. 25% in ChIP and input, respectively) which corresponds to 3.2-fold relative enrichment (1.7 fold in log<sub>2</sub>) (Fig. 19C). Such enrichment is less marked after competition with the soluble peptide (RA: 53% vs 23% in ChIP and input, respectively) leading to about 2 fold RE (Fig. 19D). Of note, the abundance of this modification in bulk chromatin (around 20%) limits the efficiency of the enrichment; as a matter of fact, 5 folds is the maximum enrichment that we can expect. The MS-readout shows a H/L SILAC ratio of 32 fold between the H3K4me1 in the heavy, relative to the light channel, thus confirming that the peptide competition works efficiently (Fig. 19E).

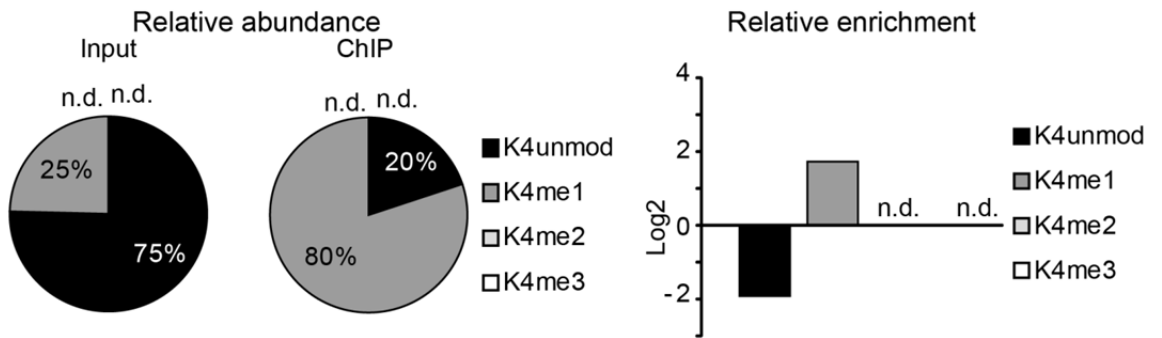
A) SILAC forward experiment (a-H3K4me1)



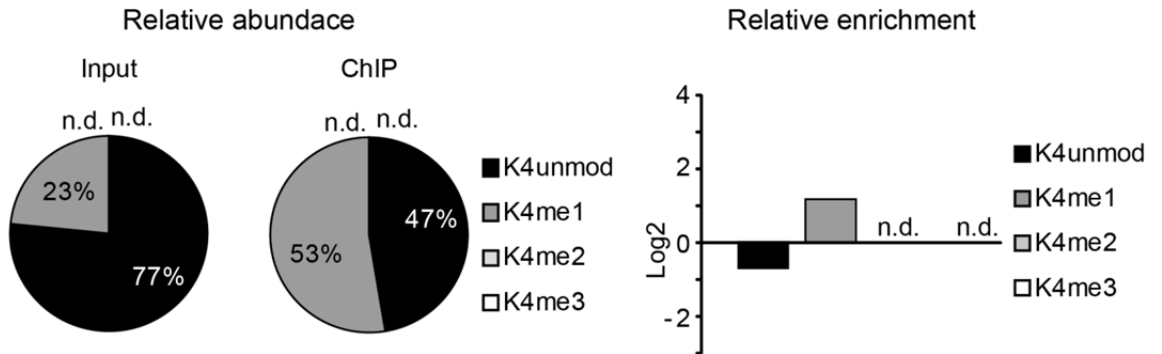
B) MS-based quantification of modified histone peaks

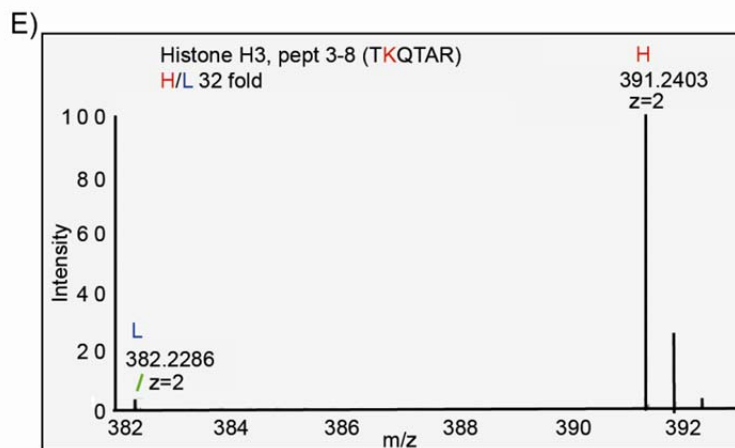


C) Heavy chromatin (not competed)



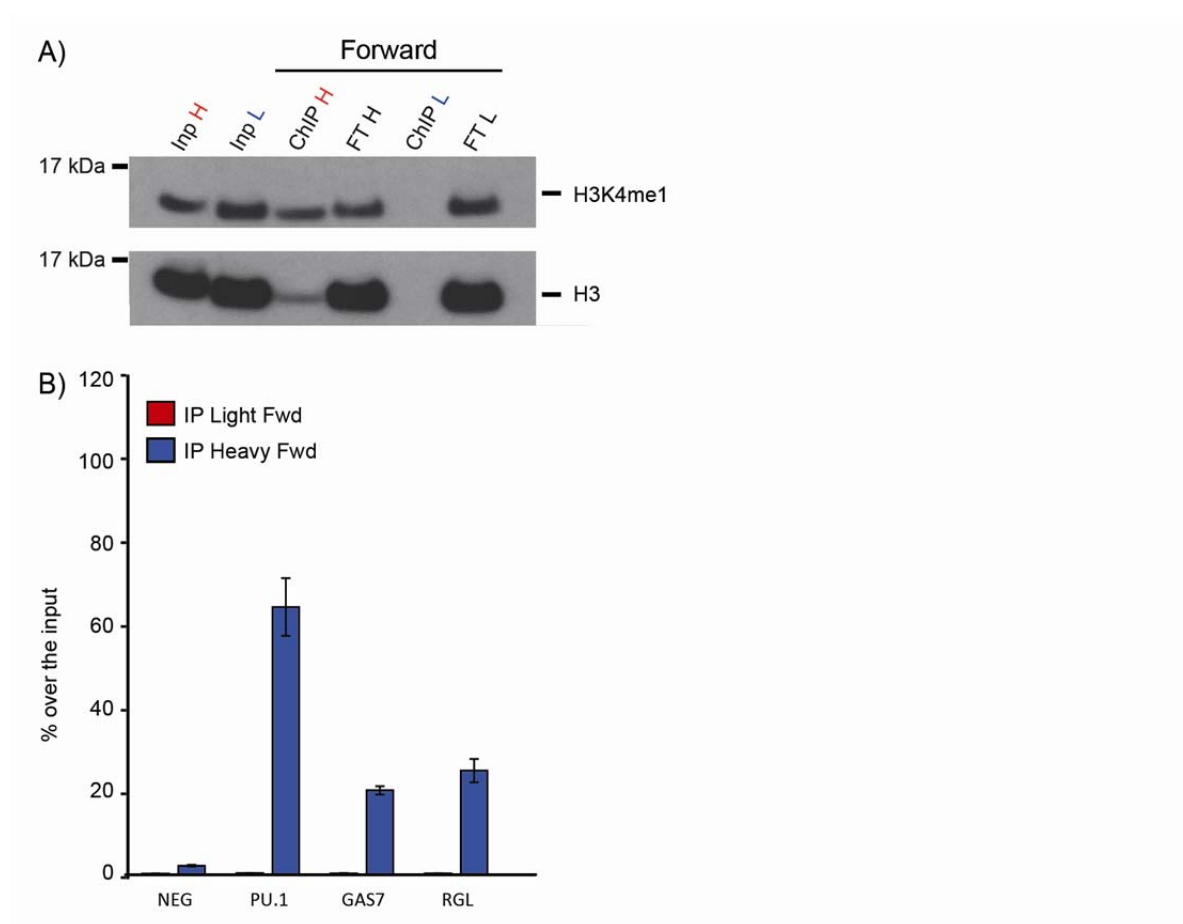
D) Light chromatin (competed)





**Fig. 19: Readout of H3K4me1 X-ChroP SILAC experiment.** **A)** Coomassie staining of SDS-PAGE of X-ChroP experiment; in blue boxes histone octamers analyzed. **B)** Survey scan (left) and XIC visualization (right) of peptide 3 – 8 of histone H3 in the chromatin input (up) or in the ChIP (bottom). **C)** Relative abundance of peptide 3 – 8 in the input and in the ChIP (left) and its relative enrichment (right) in heavy not competed chromatin. **D)** Relative abundance of peptide 3 – 8 in the input and in the ChIP (left) and its relative enrichment (right) in light peptide-competed chromatin. **E)** MS validation of peptide competition between not competed (heavy) and competed (light) ChIPs. (Blue = light; red = heavy).

The results obtained by mass spectrometry were also validated by western blot; Fig. 20 shows the amount of H3K4me1 in the input as well as in the ChIP before the 1-to-1 H/L mixing. As expected, the bait is enriched in the heavy ChIP (ChIP H) over its input, while the competed light ChIP (ChIP L) does not show a similar enrichment (Fig. 20A). Furthermore the heavy FT (FT H) is more depleted compared to light one (FT L). We also validated these results by qPCR analysis of the DNA extracted from the same samples: a strong enrichment the genomic regions with enhancer function (GAS7, PU.1, RGL) is observed in the heavy, not-competed ChIP but not in the light, competed channel. The negative control (NEG) did not show enrichment in any of the ChIPs (Fig. 20B).



**Fig. 20: WB and qPCR validation of peptide competition in H3K4me1 X-ChroP experiment.** **A)** WB analysis of Input, flow-through (FT) and ChIP in heavy and light SILAC channels in forward X-ChroP experiment. **B)** qPCR validation of the enrichment of the mRNA of positive (PU.1, GAS7 and RGL) and negative (NEG) regions in forward X-ChroP experiment. (Red boxes = ChIP light channel forward experiment, blue boxes = ChIP heavy channel forward).

#### 5.1.4 Setup of Pu.1 X-ChroP

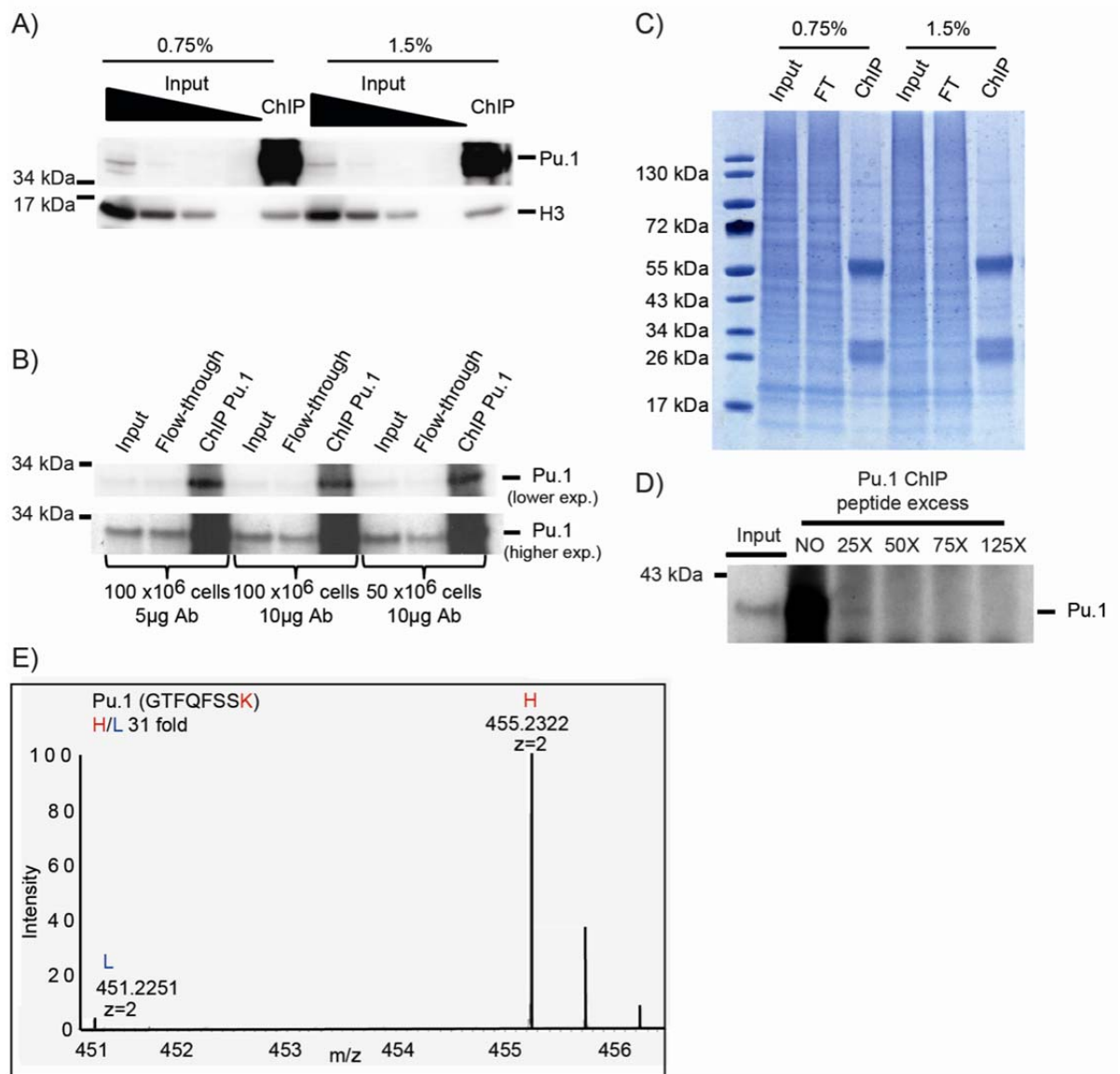
Following the same strategy as for H3K4me1, we optimized the immunoprecipitation for the transcription factor Pu.1, which binds enhancers' chromatin at the level of the NFRs. We compared the immunoprecipitation efficiency of the bait in the ChIP performed using either 0.75% or 1.5% of formaldehyde cross-linked chromatin. Western blot analysis indicated that the immunoprecipitation was not impeded by the higher concentration of cross-linking agent (Fig. 21A), as Pu.1 was immunoprecipitated in both conditions with similar efficiency. Hence, we adopted 1.5% formaldehyde cross-linked chromatin as input, to better stabilize the DNA-protein interactions. As such, we can

use the same chromatin input in the two X-ChroP experiments carried out with H3K4me1 and Pu.1 as baits.

We then optimized the input/antibody ratio by comparing three different conditions tested:  $100 \times 10^6$  cells as input for the ChIP and 5 $\mu$ g of antibody against Pu.1 and increasing twofold or fourfold the ratio (Fig. 21B). The progressive increase in the antibody to input ratio leads to a more efficient immunoprecipitation of the bait (Fig. 21B, lower exposure), mirrored also by its more pronounced depletion in each corresponding FT (Fig. 21B, higher exposure). Based on these findings, we chose as experimental conditions for the Pu.1 X-ChroP:  $150 \times 10^6$  RAW 264.7 cells and 20 $\mu$ g of antibody, a setting that could efficiently immunoprecipitate a representative portion of the bait.

We performed two parallel ChIPs using as input chromatin cross-linked with either 0.75% or 1.5% of formaldehyde and the chosen input/antibody ratio (Fig. 21C). Upon SDS-PAGE and coomassie staining, we observed that the immunoprecipitation efficiency of the histone octamer and the general pattern of the bands in the two ChIPs were comparable; since 1.5% formaldehyde does not seem to hamper IP efficiency and it is supposed to better stabilize the dynamic interaction, we chose 1.5% formaldehyde cross-linking as working condition.

The peptide fold excess in the Pu.1 X-ChroP was then optimized employing the commercial peptide used as epitope for the generation of the polyclonal anti-Pu.1 antibody. We tested four different peptide concentrations to evaluate the right molar excess capable of producing the optimal Pu.1 H/L ratio. Fig. 21D shows that a 25-fold molar excess of the Pu.1 peptide abolishes the immunoprecipitation of the bait; therefore we tested slightly lower fold molar excess: the effect of adding 18-peptide fold molar excesses was investigated by mass-spectrometry in a SILAC forward experiment, where the competition was carried out in the light channel. Fig. 21E shows the SILAC pair for the C-terminal

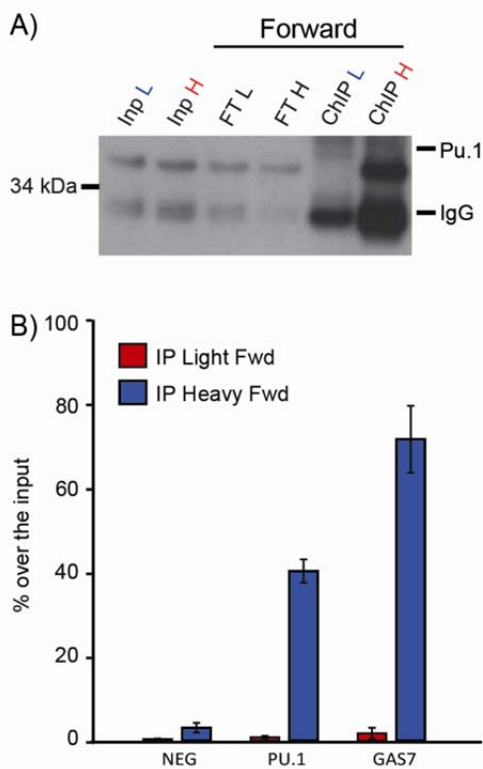


**Fig.21: Optimization of the X-ChroP conditions using Pu.1 as bait.** **A)** Pu.1 ChIP efficiency over the input using either 0.75% or 1.5% formaldehyde cross-linked chromatin; global histone H3 level shown as normalizer. **B)** Pu.1 levels in input, flow-through and ChIP using different antibody/input ratios (100\*10<sup>6</sup> cells and 5ug of antibody, 100\*10<sup>6</sup> cells and 10ug of antibody or 50\*10<sup>6</sup> cells and 10ug of antibody) (exp. = exposure). **C)** Coomassie stained SDS-PAGE showing Input, flow-through (FT) and ChIP using either 0.75% or 1.5% cross-linked chromatin input. **D)** Pu.1 enrichment in the not-competed ChIP (NO) over the input and decreased ChIP efficiency upon increasing Pu.1 peptide folds excess. **E)** MS validation of peptide competition between not-competed (heavy) and peptide-competed (light) ChIP.

peptide (GTFQFSSK) of Pu.1 with a H/L SILAC ratio of 31, which we evaluated adequate for the subsequent interatomics experiments.

MS-results were also validated by western blot; Fig. 22 shows the amount of Pu.1 in the input as well as in the ChIP before the 1-to-1 H/L mixing. The bait is enriched in the

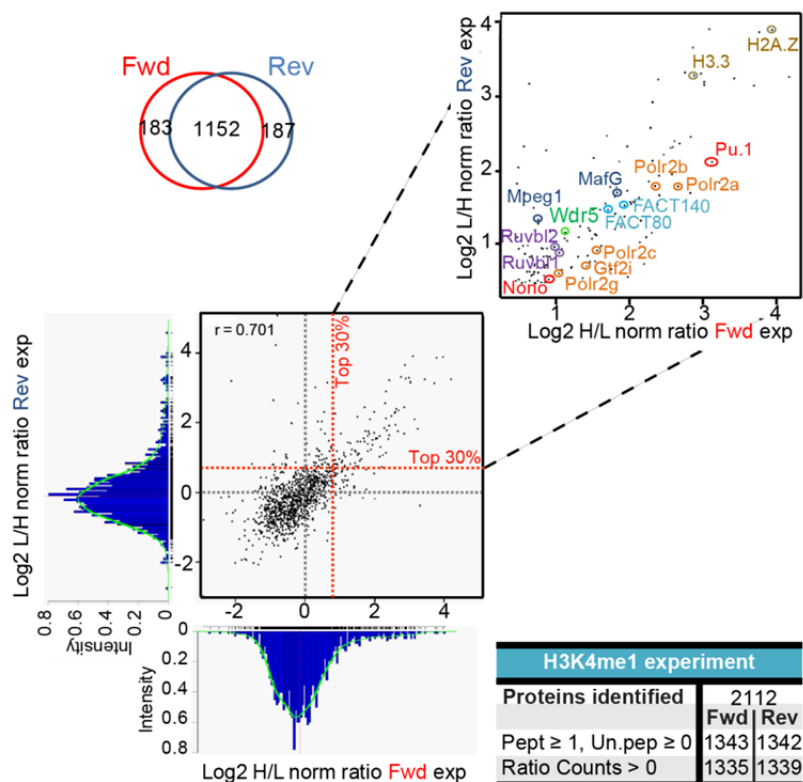
not competed ChIP (ChIP H) over its input, while Pu.1 does not show a similar enrichment in the light competed ChIP (ChIP L) (Fig. 22A). These results are confirmed by a higher depletion of the heavy FT (FT H), relative to the light one (FT L). qPCR analysis of the DNA extracted from the same samples displays a strong enrichment of the genomic regions with enhancer function (GAS7, PU.1) in the heavy, not-competed ChIP but not in the light, competed channel. The negative control (NEG) does not show enrichment in any of the ChIPs (Fig. 22B).



**Fig. 22: qPCR and WB validation of peptide competition in Pu.1 X-ChroP experiment.** **A)** WB analysis of Input, flow-through (FT) and ChIP in heavy and light SILAC channels in forward X-ChroP experiment. **B)** qPCR validation of the enrichment of the mRNA of positive (PU.1 and GAS7) and negative (NEG) regions in forward X-ChroP experiment. (Red = ChIP light channel forward experiment, blue = ChIP heavy channel forward).

## 5.2 Dissection of enhancers' determinants in resting macrophages by X-ChroP

We started by characterizing the proteins co-associated with the H3K4me1-bound chromatin. We performed forward and reverse X-ChroP experiments by swapping the SILAC channel in which the peptide competition was carried-out. We mixed the two SILAC-labeled channels in a combined sample that was then separated by SDS-PAGE. We excised 10 slices for gel lane and subjected them to trypsin digestion, followed by MS-analysis on Q Exactive instrument (Thermo Fisher Scientific) (*see material and methods for details*). Proteins were identified and quantified by using the MaxQuant software suite (Cox and Mann 2008; Cox, Neuhauser et al. 2011) setting a confidence level of 99% (protein and peptide FDR equal 1%).

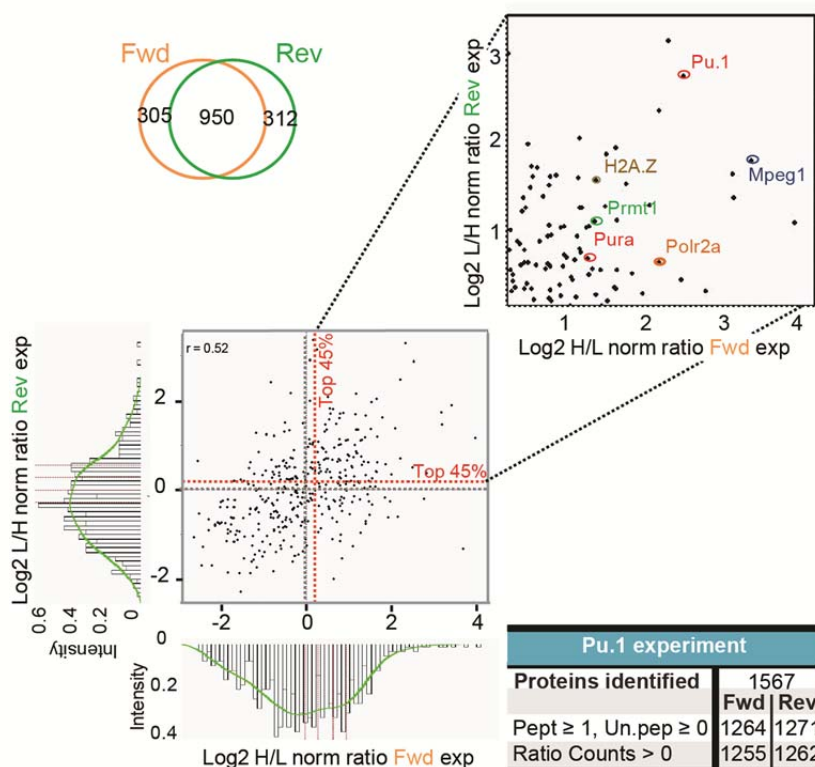


**Fig.23: Dissection of the H3K4me1 interactome.** Venn diagram shows the overlap between forward (Fwd) and reverse (Rev) X-ChroP experiments using H3K4me1 as bait. The scatterplot shows the log<sub>2</sub> SILAC distribution (H/L ratio and L/H ratio for the forward and reverse experiment, respectively) of common proteins. The zoom underlines significantly enriched proteins belonging to the top 30% distribution of both Fwd and Rev experiment. The table summarises proteins identified and quantified in the experiments.



We quantified 1335 and 1339 proteins in the forward and reverse experiment, respectively, of which 1152 were quantified with at least one ratio count in each experiment (Fig. 23).

Based on the protein ratio distributions ( $\log_2$  of H/L and L/H ratio for the forward and the reverse experiment, respectively) we defined as putative interactors those proteins within the Top 30% (70<sup>th</sup> percentile) of both distributions, which are located in the upper-right section of the scatterplot. Zooming into this area, we observed proteins specifically associated with H3K4me1-bound chromatin such as Pu.1 itself, several subunits of the RNA polymerase II complex, the facilitating of chromatin transcription (FACT) complex, components of diverse chromatin remodeling complexes and H3.3 and H2A.Z, two histone variants, traditionally associated with active chromatin (Fig. 23, zoomed upper panel).



**Fig.24: Dissection of the Pu.1 interactome.** Venn diagrams show the overlap between forward (Fwd) and reverse (Rev) X-ChroP experiments using Pu.1 as bait. The scatterplot shows the  $\log_2$  SILAC distribution (H/L ratio and L/H ratio for the forward and reverse experiment, respectively) of common proteins. The zoom underlines significantly enriched proteins belonging to the top 45% distribution of both Fwd and Rev experiment. The table summarizes proteins identified and quantified in the experiments.

The geLC-MS/MS analysis of the forward and reverse X-ChroP experiments using the Pu.1 as bait produced 1255 and 1262 proteins in the forward and reverse experiments, respectively, of which 950 were in common and they were identified with at least one ratio count in each experiment (Fig. 24).

Based on the log<sub>2</sub> of the two protein ratio distributions, we set as putative interactors of Pu.1-marked chromatin those proteins within the Top 45% (55<sup>th</sup> percentile) of both distributions (in the upper-right quadrant of the scatterplot). As expected, Pu.1 was the top scoring protein among them, but other interesting proteins such as the RNA polymerase II subunit Rbp1, the histone variant H2A.Z, the protein arginine methyl-transferase 1 (Prmt1) and the transcription activator Pura were found in this group (Fig. 24, zoomed upper panel).

#### *5.2.1 Characterization of enhancers' composition by the intersection of H3K4me1 and Pu.1 interactomes.*

We intersected the data coming from the two interactome studies (Fig. 25A) and identified 764 proteins in common (displayed in the heat map in Fig. 25B). Among these, we selected as genuine enhancers' determinants only those proteins belonging to the top distribution of each interactome, namely the Top 30% and Top 45% for H3K4me1 and Pu.1, respectively. Upon this filtering criterion, 148 proteins were selected, of which a subset, organized in distinct functional classes, is listed in Fig. 25C (*see also Appendix 1*).

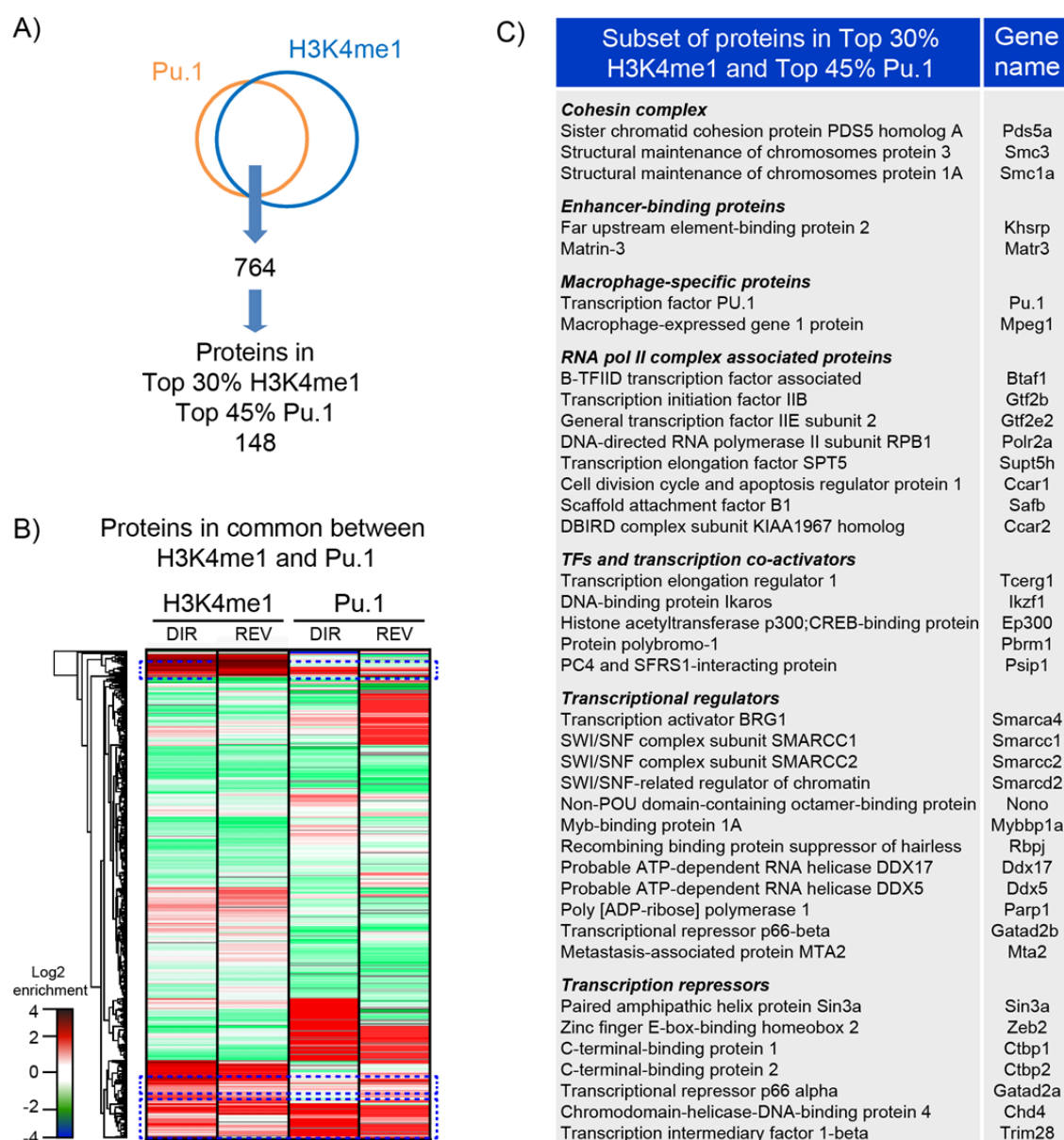
In this group we identified two core subunits (Smc1a and Smc3) and one accessory subunit (Pds5a) of the cohesin complex, important for chromatin structural stabilization and two enhancers binding proteins, namely the far upstream-elements binding protein 2 (Khsrp) and matrin3 (Matr3). Interestingly, matrin3 has been recently shown to be required for the effective enhancer's activation and subsequent transcription of the homeodomain transcription factors target genes (Skowronska-Krawczyk, Ma et al. 2014).

Among the potential enhancers' determinants we identified components of the RNA polymerase II complex, like the B-TFIID transcription factor associated protein (Btaf1), the general transcription factor II B (Gtf2b) and IIE (Gtf2e2) and the RPB1 subunit of the RNA polymerase II (Polr2a). Moreover, we also identified proteins associated with this complex like the Ccar2 protein and the scaffold attachment factor B1 (Safb), core components of the Dbc1-Zird (DBIRD) complex and the transcriptosomal complex, respectively, that both coordinate transcript elongation and RNA processing.

In the group of transcription factors or co-activator identified, interestingly we cite the histone acetyl-transferase p300 (Ep300) that is a key-mark of the LPS-inducible enhancers in macrophages (Ghisletti, Barozzi et al. 2010) and the Brg1-associated polybromo protein (Pbrm1), whose expression is a prognostic marker in renal cell carcinomas (Poirier, de Murcia et al. 1982).

At enhancers we also identified transcriptional regulator proteins that can play both a positive or negative role on transcriptional control, depending on the distinct set of binding partners associated: among those some subunits of the SWI/SNF ATP-dependent chromatin remodelers family (Smarcc1, Smarcc2 and Smarcd2); the Non-POU domain-containing protein (Nono, also known as p54(nrb)), which interact with both Matr3 and RNA-binding proteins; some ATP-dependent RNA helicases (Ddx17, Ddx21, Ddx9) already described to play important role in transcription as co-activator or co-repressors (Fuller-Pace 2006) and Parp1 that plays a dual role in transcription (Poirier, de Murcia et al. 1982).

Interestingly, we also identified as enhancers' determinants some transcriptional repressors like the protein Sin3a, the p66 alpha protein (Gatad2a) and the C-terminal binding proteins (Ctbp) 1 and 2.



**Fig. 25: Dissection of the basal enhancers' interactome.** **A)** Venn diagrams show the overlap between H3K4me1 and Pu.1 interactomes. **B)** Proteins in common between the two interactomes are subjected to hierarchical cluster analysis and represented as heatmap; dotted lines show proteins in the Top of both H3K4me1 and Pu.1 experiments. **C)** Subset of new potential enhancers' determinants classified according to their molecular function.

To understand whether these enhancers' determinants belong to specific complexes we used the "g:Profiler" tool (Reimand, Kull et al. 2007). This web-based toolset generates functional descriptions of gene lists derived from large-scale experiments, such descriptions include the presence of experimentally-characterized protein complexes, described in the CORUM database (Ruepp, Waegle et al. 2010). Using this feature, we were able to assign a statistical significance to a set of protein complexes through a

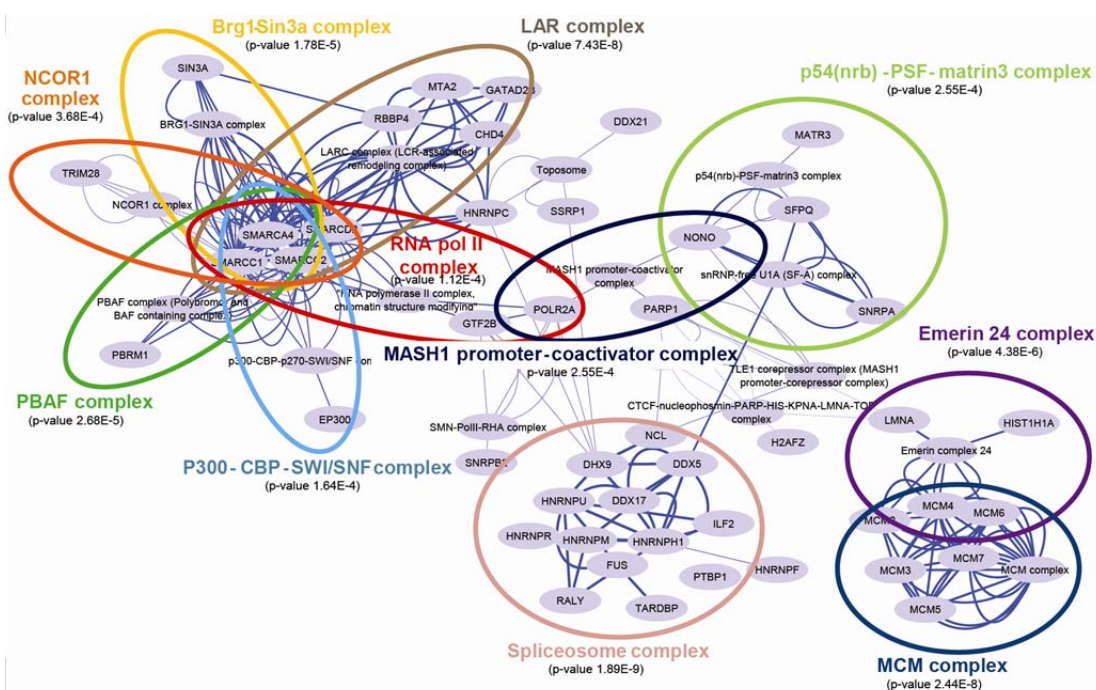
comparison of our list of proteins enriched at enhancers with a background that included the complete list of proteins identified and quantified in input (Fig. 26).

The results showed an enrichment of transcription regulator complexes associated with SWI/SNF family proteins and involved in both transcription activation and repression. Thanks to the identification of specific subunits, we were able to discern among different SWI/SNF-containing complexes.

Among the most significant complexes identified, we determined the locus control region (LCR)-associated remodeling (LAR) complex which fine-tunes the transcription of the

$\beta$ -globin locus through long-range interactions involving the enhancer region DNase hypersensitive 2 (HS2) and the beta  $\beta$ -globin promoter (Mahajan, Narlikar et al. 2005).

This complex consists of 18 subunits comprising components of the SWI/SNF family and the nucleosome remodeling and deacetylase (Mi2/NuRD) complex like Chd4, Mta2 and Rbbp4. Moreover, the heterogeneous nuclear ribonucleoproteins (hNRNPs) confer to the LAR complex the sequence-specific DNA recognition.



**Fig.26: Protein complexes significantly enriched at enhancers' of resting macrophages.** Visualization of transcriptional regulators and RNA processing protein complexes found significantly enriched at enhancers, using the entire set of proteins robustly identified in the chromatin input as background for the p-value calculation. (p-value reported in to parenthesis).

Thanks to the identification of the Sin3a protein, we could also confidently identify the Brg1-Sin3a complex (Sif, Saurin et al. 2001); this complex has been recently shown to regulate transcription in a context-dependent manner and to mediate efficient signal transducer and activator of transcription (STAT)-dependent transcriptional activities (Icardi, Mori et al. 2012).

As above-mentioned, in the cluster of the transcription co-activators we identified the polybromo protein 1, which is a unique component of the Polybromo, Brg1-Associated Factors (PBAF) complex. PBAF is the only SWI/SNF-containing complex that potentiates transcription and its role in gene transcriptional regulation during stress response in *Caenorhabditis elegans* has been recently described (Kuzmanov, Karina et al. 2014). Moreover, this complex is necessary for ligand-dependent transactivation *via* nuclear hormone receptor (Lemon, Inouye et al. 2001). Very interestingly PBAF is recruited to active chromatin through the bromo-domains of the Pbrm1 that recognize a specific pattern of lysine acetylation on the H3 tail (Chandrasekaran and Thompson 2007).

We identified as robustly associated with enhancers also two other SWI/SNF-related complexes, namely the p300-CBP-SWI/SNF complex and the RNA polymerase II complex. The p300-CBP-SWI/SNF complex brings together the histone acetyl-transferases activity of p300 and CBP with the chromatin remodeling function of SWI/SNF complex, thus representing a model of transcriptional activation function through histone acetylation (Dallas, Cheney et al. 1998). The RNA polymerase II complex was identified in our dataset by the specific enrichment of both the Polr2a protein and the general transcription factor II B (Gtf2b) (Cho, Orphanides et al. 1998). It promotes the activation of transcription by coupling chromatin remodeling and RNA polymerase II activities. Noteworthy, at enhancers we identified both the acetyl-transferase p300 and the RNA polymerase II which form the transcription initiation complex when the c-terminal domain (CTD) of the RNA polymerase II is non-phosphorylated.

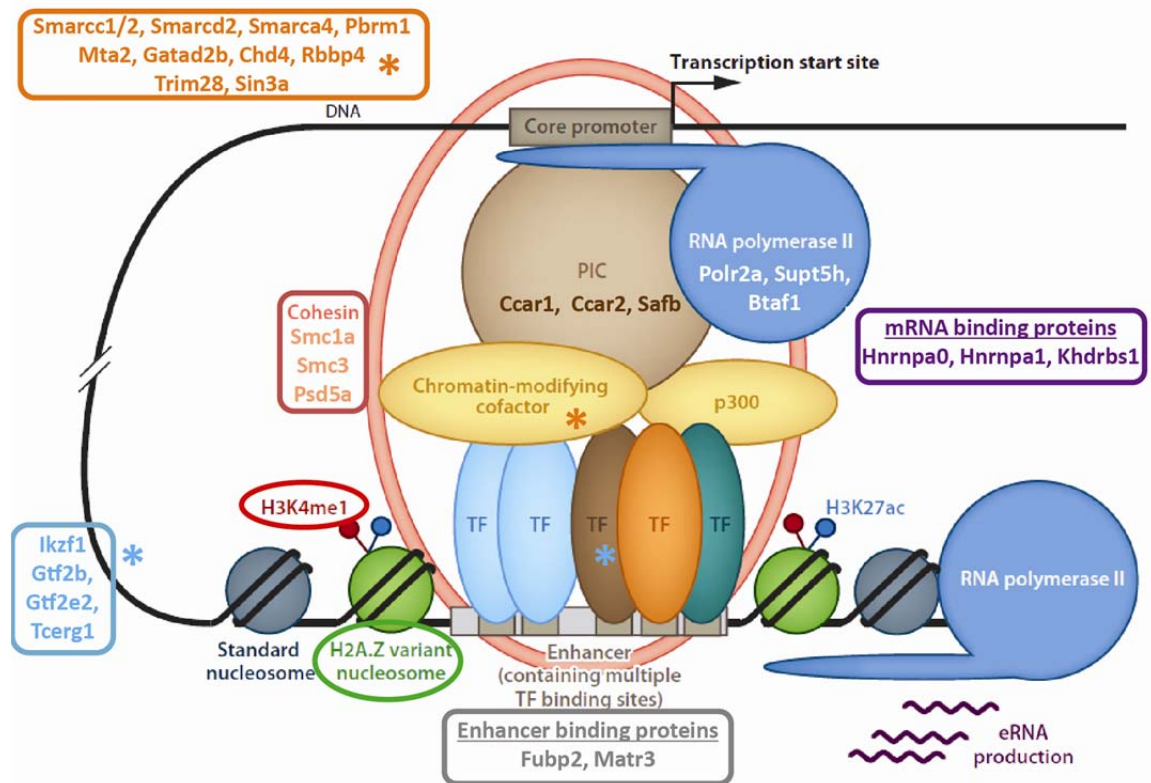
We also identified the MASH1 promoter-co-activator complex, which promotes transcription *via* Parp1-dependent co-repressor displacement (Ju, Solum et al. 2004).

At enhancers, the presence of two RNA binding complexes, namely the Emerin complex 24 and the p54(nrb)-PSF-matrin3 complex was also observed; while the former is associated with RNA processing (Holaska and Wilson 2007), the latter is involved in RNA binding (Zhang and Carmichael 2001) and it is composed by Matrin3 and Nono proteins, both previously described also as transcriptional regulators.

Interestingly, the spliceosome complex (Zhou, Licklider et al. 2002) seems enriched at enhancers; this finding corroborates the model that proposes co-occurrence between transcription and mRNA editing in both time and space.

The role of the minichromosome maintenance (MCM) complex in DNA replication is well established since about two decades (Ishimi, Ichinose et al. 1996); however more recently, it has been described that some subunits of this complex interact with histones and the RNA polymerase II (Holland, Gauthier et al. 2002) and that MCM complex may participate in the Stat1 target gene activation and RNA polymerase II-mediated transcription (Snyder, He et al. 2005). Our findings indicate a clear enrichment of this complex at enhancers thus corroborate the hypothesis of a novel function of MCM complex in transcriptional regulation.

Putting all our findings in a common view, at enhancers we observed the specific enrichment of chromatin-modifying co-factors associated with transcriptional regulation, together with mRNA binding proteins, transcription factors, enhancer binding proteins, RNA polymerase II complex, p300, components of the pre-initiation complex (PIC) and the cohesin complex. These data are reminiscent of a model whereby the enhancer-promoter chromosomal looping bridges together transcriptional regulator complexes and chromatin remodelers in order to tightly regulate the expression levels of relevant genes at the correct time and in the specific cellular context (Fig. 27).



**Fig. 27: Enrichment of players of the enhancer/promoter looping in the basal enhancers' interactome** Transcription regulators, chromatin modifiers and Cohesin complex, bridge together enhancer and promoter regions establishing a chromosomal looping that modulates the formation of the pre-initiation complex (PIC) and transcription by RNA polymerase II. (Adapted from Maston G A et al. *Annu Rev Gen Hum Gen* 2012).

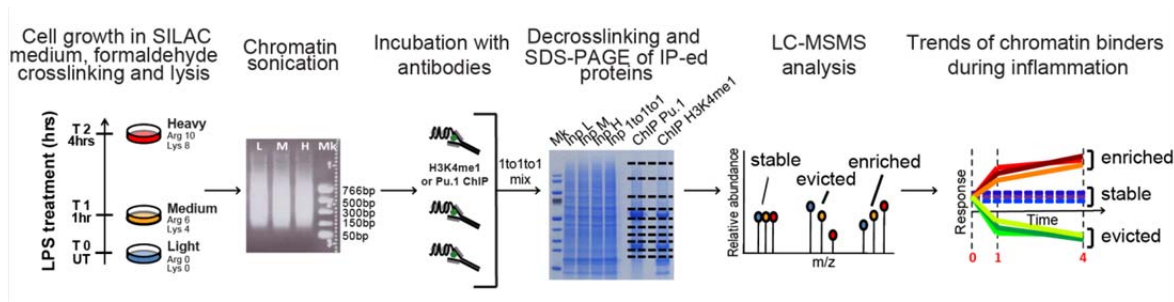
### 5.3 Profiling chromatin determinants of enhancers during inflammatory stimulus

#### 5.3.1 Triple SILAC labeling, for time-course X-ChroP analysis

Having determined the set of proteins that constitute the basal enhancers' proteome, we then want to profile the changes in chromatin composition in response to an inflammatory stimulus, such as LPS. In fact RAW 264.7 cells react to LPS by eliciting a massive transcriptional activation of inflammatory genes, accompanied by a global rearrangement of chromatin architecture.

We set up an X-ChroP experiment using triple SILAC labeling, to carry out the temporal analysis at three time-points during the inflammatory response.

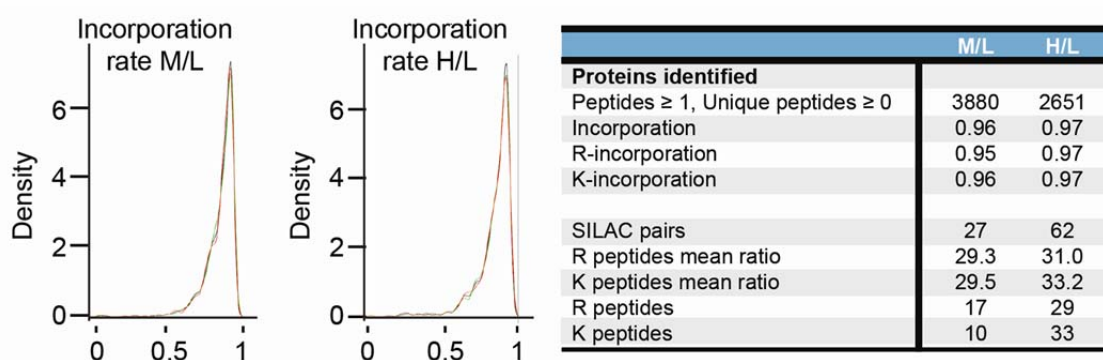




**Fig. 28: Schematic representation of triple-SILAC labeling combined with X-ChroP.**

Experimental design of time-course X-ChroP (tc-ChroP); cells are labeled in light, medium and heavy SILAC amino acids and left untreated, treated for 1h and 4hrs with LPS, respectively. Formaldehyde-fixed cells are lysed and fragmented and chromatin is used as input for ChIP, using H3K4me1 and Pu.1 as baits. Eluted proteins are mixed in a 1:1:1 H:M:L ratio, separated and analyzed by LC-MS/MS. Proteins stably associated, evicted and recruited to chromatin during the inflammatory response will display, a stable, decreasing or increasing protein ratio passing from light to medium up to heavy, respectively.

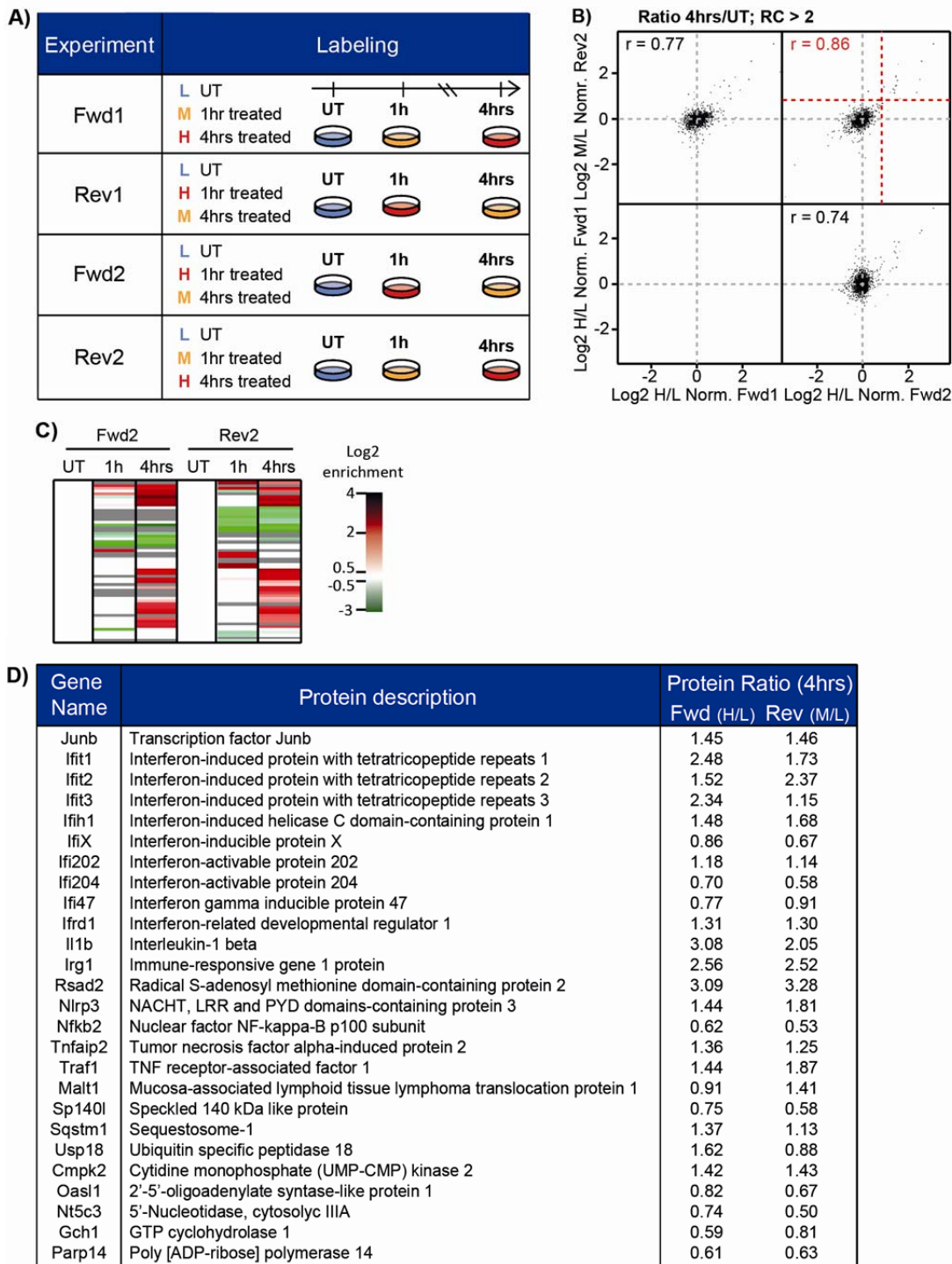
Cells were grown in SILAC medium containing light, medium or heavy isotope-coded lysine and arginine amino acid. In the forward experimental design, light, medium and heavy labeled cells were left untreated or treated for 1 or 4hrs with LPS, respectively; these two intervals of treatment recapitulate the early (1h) and acute (4hrs) phase of activation of the inflammatory genes. Cells were cross-linked and lysed and then chromatin from each preparation was fragmented through sonication to obtain 300-500bp nucleosomal stretches. This was used as input for three parallel ChIP experiments. The eluted proteins were mixed in a 1-to-1-to-1 ratio, de-cross-linked and separated *via* SDS-PAGE. Upon geLC-MS/MS analysis, proteins were identified and quantified by MaxQuant software (Cox and Mann 2008; Cox, Neuhauser et al. 2011). The final SILAC readout enables to discern proteins recruited, evicted or stably associated with enhancers during inflammation, as schematized in Figure 28. More specifically: proteins stably associated with these *cis*-regulatory regions are expected to display a similar intensity in all the three SILAC channels; while proteins evicted or recruited to enhancers, will show a progressive decrease or increase, respectively, in the intensity of the corresponding peaks from the medium and heavy channels relative to the light one, which represents the basal state.



**Fig. 29: Incorporation rate of RAW264.7 cells in heavy and medium SILAC media.** Density distributions show the percentages of incorporation of medium (left) or heavy (right). The table summarizes the peptides identified (at list 2 peptides, 1 of which unique) and the incorporation rate of arginine- and lysine containing peptides in both medium and heavy samples.

We evaluated the incorporation efficiency of RAW 264.7 cells cultured in SILAC DMEM containing Medium amino acids (Arg 6, Lys 4); upon 8 passages, we obtained 95% and 96% incorporation for Arg6- and Lys4-containing peptides, respectively (Fig. 29).

We performed four tc-ChroP replicate experiments, two in the forward and two in the reverse setups (Fig. 30A); we evaluated the global correlation of protein ratios among the input of the experiments, focusing on the H/L and M/L ratios for forward and reverse experiments respectively, which correspond to 4hrs of LPS treatment over the untreated state. The forward 2 (fwd2) and reverse 2 (rev2) experiments showed the highest correlation (Fig. 30B) and were selected as reference dataset for the analysis of the dynamic profiling of enhancers' determinants. We clusterized the proteins based on their SILAC ratio during the inflammatory stimulus, carrying out an unsupervised hierarchical clustering analysis of protein ratios in the fwd2 and rev2 chromatin inputs. Only a subset of proteins resulted up-regulated (SILAC ratio greater than +2 s.d.) during the stimulus (Fig. 30C), precisely 26 out of 2532 proteins.



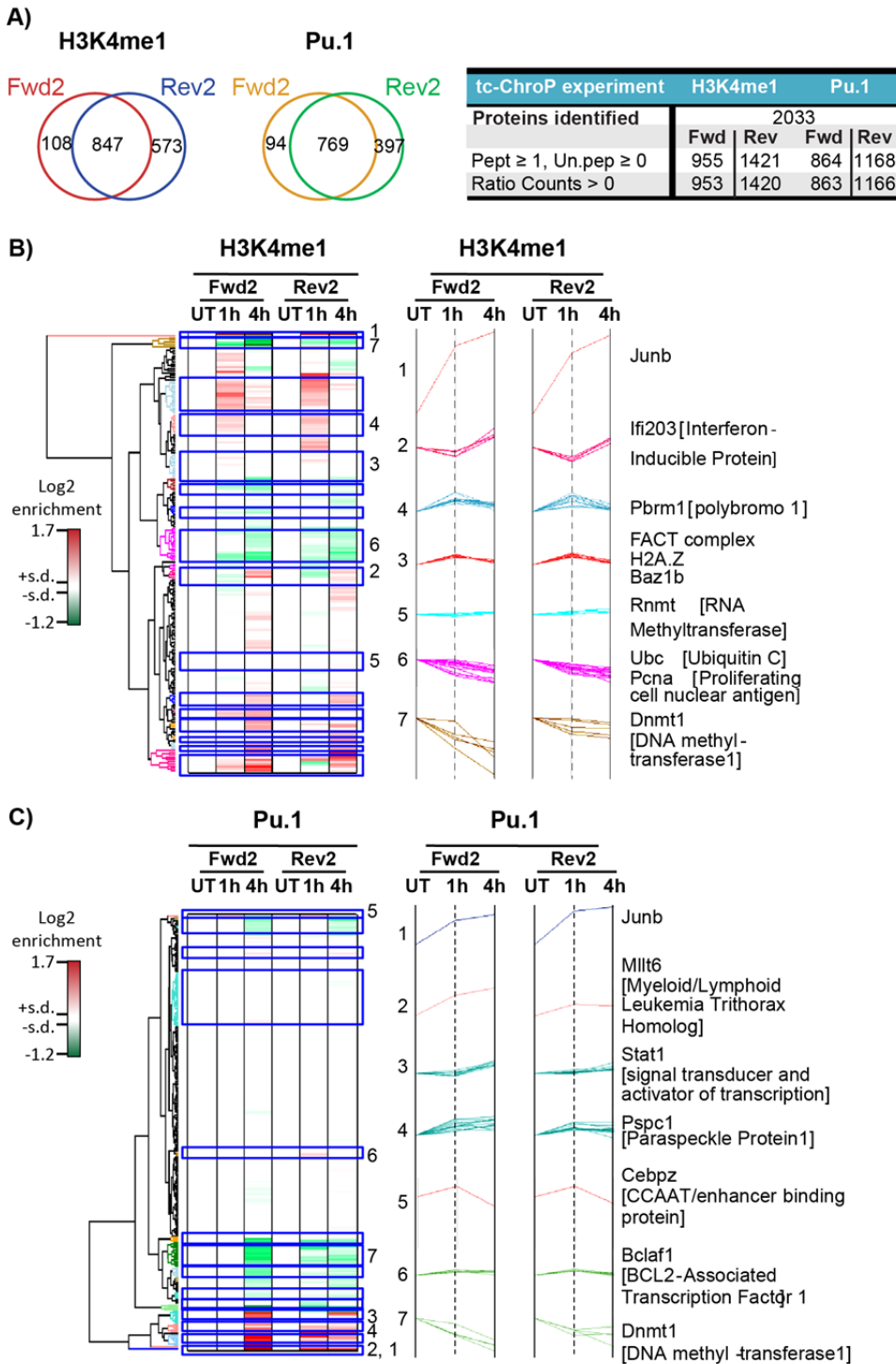
**Fig. 30: Analysis of the inputs from tc-ChroP experiments.** **A)** Experimental design of forward and reverse tc-ChroP experiments. Untreated cells (UT, blue) are always light-labeled; in forward experiment, medium (orange) and heavy-labeled (red) cells are treated for 1h and 4hrs with LPS, respectively (these correspond to 4hrs and 1hr treatment respectively in reverse experimental design). **B)** Multi scatterplot visualization of 4hrs/UT ratio in the inputs (red dotted-lines = threshold for significantly up-regulated proteins). **C)** Heat map representing the hierarchical unsupervised clustering of proteins enriched upon LPS in forward 2 (Fwd2) and reverse 2 (Rev2) experiments (red = enriched, green = depleted; white = within  $\pm 2$  standard deviations). **D)** Protein description and log<sub>2</sub> SILAC ratio of enriched proteins in Fwd2-Rev2 experiments.

The majority of these proteins have been previously described as involved in the inflammatory response, such as the transcription factor JunB (Fujihara, Muroi et al. 1993), the interferon-induced proteins (Ifit1, Ifit2, Ifit3, Ifih1) (Lee, Jenkins et al. 1995; Smith and Herschman 1996), the interleukin-1 beta (Il1b) (Wu, Chen et al. 2009) as well as proteins important for the toll-like receptor 4 (TLR4) and NF- $\kappa$ B cascades, like the Nod-like receptor (NLR) family member, Nlrp3.

The up-regulation of all these positive markers, suggested us that the model system adopted efficiently responds to the inflammatory stimulus and that the experimental approach and the technique employed are sensitive enough to reveal such biological effects.

Each tc-ChroP experiment was carried in duplicate using both H3K4me1 and Pu.1 as baits (see Fig. 28 for the scheme of the experimental design). In H3K4me1 tc-ChroP we identified and quantified 953 proteins and 1420 proteins in the forward and reverse experiments, respectively; in the Pu.1 tc-ChroP, we quantified 863 proteins in the forward and 1166 proteins in the reverse experiment. Upon intersection of the technical replicates we robustly identified 847 and 769 proteins in H3K4me1 and Pu.1 dataset, respectively (Fig. 31A).

The M/L and H/L SILAC ratios in the forward experiment correspond to the difference between the peak intensity of a given protein in the 1hr-stimulated sample (medium channel, M) and in the 4hrs-stimulated sample (heavy channel, H) respectively, relative to the untreated state (light channel, L). These ratios correspond to H/L and M/L in the reverse experiment. We applied an unsupervised hierarchical clustering based on the protein SILAC ratios, to classify proteins based on similar trends of response upon the inflammatory stimulus; these clusters are represented in the heatmaps in Fig. 31B and 31C, for H3K4me1 and Pu.1 respectively. Each cluster displays specific dynamics, represented as a profile plot (Fig. 31B and Fig. 31C, right).



**Fig.31: Analysis of tc-X-ChroP forward 2 and reverse 2 experiments upon LPS stimulus.** **A)** Venn-diagrams show the overlap between forward and reverse experiment for both H3K4me1 (right) and Pu.1 (left); table summarises protein identified and quantified in the experiments. **B)** Heatmap visualization of the unsupervised hierarchical clustering analysis of proteins in common between forward 2 (Fwd2) and reverse 2 (Rev2) H3K4me1 tc-ChroP experiments (left), line-trend visualization of protein with similar behaviors during inflammation (right). **C)** Heat map visualization of the unsupervised hierarchical clustering analysis of proteins in common between forward 2 (Fwd2) and reverse 2 (Rev2) Pu.1 tc-ChroP experiments (left), line-trend visualization of protein with different behavior at enhancers, during inflammation (right). (Red = recruited, green = evicted, white = within  $\pm 1$  s.d. of 90% distribution).

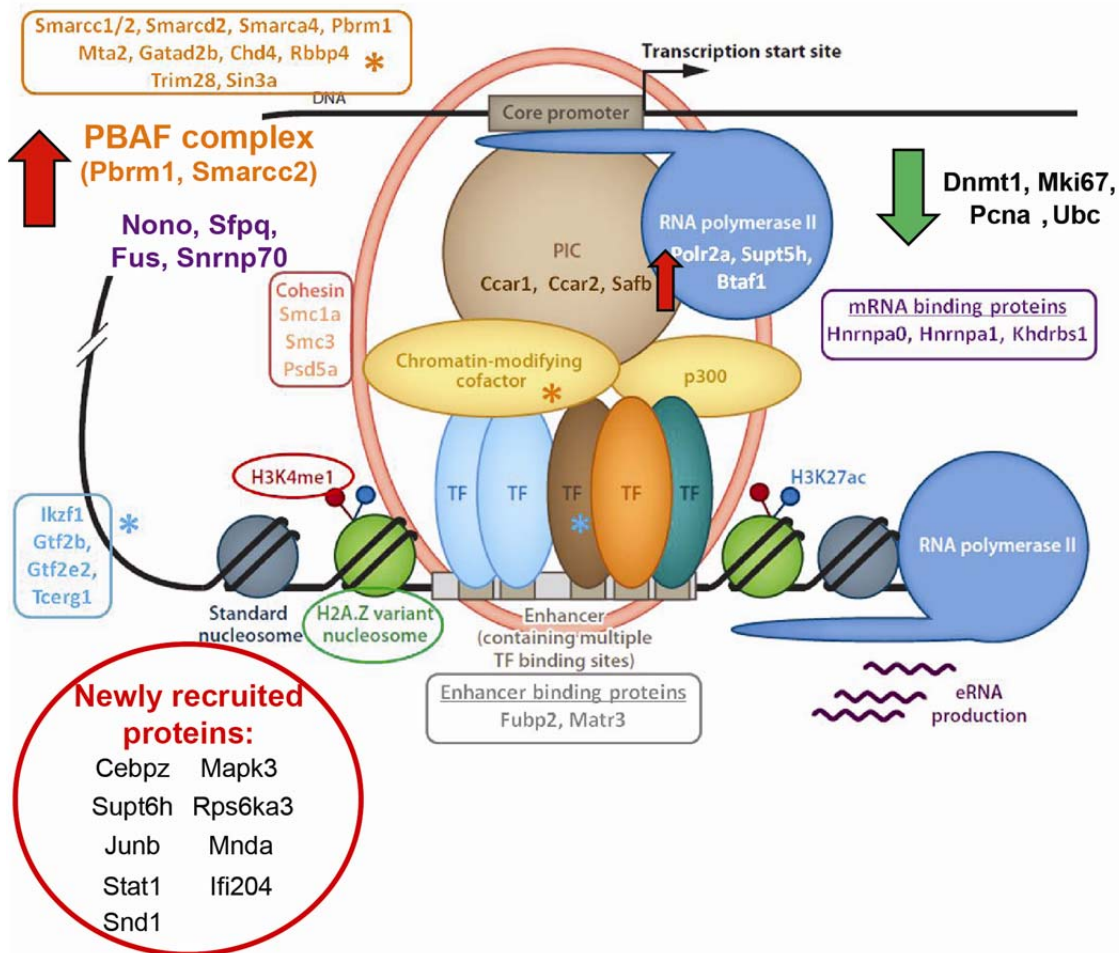
As expected, some proteins were not influenced by the LPS stimulus and their association with enhancers remains constant during inflammation, with a “Pu.1-like” behavior; other proteins were dynamically recruited to- or evicted from-enhancers upon LPS, with different kinetics.

Using these data, we wanted to profile the dynamic behavior of the proteins identified as specific macrophage enhancers’ determinants at basal conditions. We thus intersected the temporal protein profiles with the initial list of 148 static enhancer’s determinants (*See also Appendix 2*). Interestingly, we observed a specific recruitment at enhancers of the Smarcc2 protein and the Pbrm1 protein, two subunits of the SWI/SNF-related complex PBAF. As already mentioned, this complex is associated with transcriptional activation; indeed while the SWI-SNF subunits mediate the ATP-dependent chromatin remodeling activity, the bromo-domains of Pbrm1 regulate the site-specificity for the binding of the PBAF complex (Fig. 32).

In addition to PBAF, we observed the recruitment of Nono and Sfpq proteins, both involved in RNA processing and in the androgen receptor-mediated transcription (Kawahara, Ikei et al. 2006). Safb protein was also specifically recruited at enhancers; interestingly, this protein is a component of the transcriptosomal complex that couples transcription to RNA processing.

In parallel we observed the eviction from enhancers of some interesting proteins: the DNA (cytosine-5) methyl-transferase 1 (Dnmt1) protein, which promotes the methylation of hemi-methylated CpG islands at promoter regions, a well-established mechanism through which transcriptional silencing is achieved (Mohn, Weber et al. 2008); the eviction of Dnmt1 from enhancers upon LPS, is in accordance with the concept of a more accessible chromatin state which promotes TFs binding to DNA.

Upon LPS stimulus some proteins related to the processes of DNA replication and cell proliferation, like the Proliferating Cell Nuclear Antigen (PCNA) protein and the Marker



**Fig.32:Dynamic profiling of enhancer' determinants during LPS stimulus and proteins newly recruited at these *cis*-regulatory regions.** (Adapted from Maston G A et al. *Annu Rev Gen Hum Gen* 2012). Recruitment of PBAF complex at enhancers during the inflammatory stimulus, as well as Nono, Sfpq, Fus and Snrnp70. Displacement of Dnmt1, Mki67, PcnA and Ubc from enhancers along the LPS treatment. A class of proteins newly recruited at enhancers is displayed.

of proliferation Ki-67 (Mki67) protein, are evicted from enhancers; this result is in line with the evidence that activated macrophages stop dividing and focalize their activity on mounting the immunological response, to counteract the infection.

Our analysis also enables to identify those proteins which were undetectable at basal condition but resulted as newly recruited at enhancer only upon LPS (See also Appendix 3). In this class we found the CCAAT/enhancer binding protein zeta (Cebpz) that is a transcription factor important for myeloid differentiation (Zhang, Hetherington et al. 1996) and Junb, a transcription factor already proposed to be inducer of pro-inflammatory cytokines during LPS-dependent inflammation (Gomard, Michaud et al. 2010).

Interestingly, we also identified some transcriptional regulators involved in the response to inflammatory stimuli, such as the Signal Transducer and Activator of Transcription 1 (Stat1) protein, which activates genes involved in the inflammatory response (Luu, Greenhill et al. 2014) and two components of the HIN-200 family, namely the Myeloid cell Nuclear Differentiation Antigen-Like (Mndal) and the Interferon-activable protein 204 (Ifi204), both acting as transcriptional regulators (Dauffy, Mouchiroud et al. 2006; Zhang, Kagan et al. 2009). Furthermore, two proteins crucial for transcriptional activation were also newly recruited: the Staphylococcal Nuclease and tudor Domain containing 1 (Snd1) and the Suppressor of Ty6 Homolog (Supt6h). Snd1 is a DNA-binding protein, co-activator of Stat6 and bridging factor between this signal transducer and the basal transcription factor (Yang, Aittomaki et al. 2002); Supt6h is a member of the polymerase associated factor 1 (PAF1) complex, required for efficient transcriptional elongation (Willmann, Milosevic et al. 2012).

This temporal profiling highlighted a somehow expected dynamics of enhancers upon LPS illustrated by the recruitment of chromatin remodelers and the displacement of a DNA methyl-transferase enzyme, which together stimulate gene expression. Based on these findings, during the inflammatory response the majority of enhancers' determinants are stably associated, suggesting that enhancers are constituted by an overall high-ordered structure that pre-exist the transcription activation; after the LPS treatment, specific proteins are recruited in a time-specific fashion and serve to fine-tune gene expression.

### *5.3.2 Follow-up validation of Mpeg1 protein as a novel enhancer determinant*

In addition to chromatin determinants dynamically recruited or evicted from enhancers during the inflammatory response, we also identified a set of proteins that are stably associated with these regions, like Pu.1. They are potentially very interesting as they may either synergize with Pu.1, or act independently as novel transcription factors to



regulate the expression of inflammatory genes in macrophages. Within this list of candidates, we focused our attention on the Macrophage-Expressed gene 1 (Mpeg1) protein, because of its macrophage-specific expression and its functional dependence on Pu.1 (Zakrzewska, Cui et al. 2010). As a matter of fact Mpeg1 has been identified in mice and human as macrophage-specific, with an increasing expression during macrophage differentiation (Spilsbury, O'Mara et al. 1995). The *mpeg1* gene is evolutionary conserved from invertebrates to vertebrates; Mpeg1 is structurally composed of a N-terminal Membrane-Attack Complex and Perforin (MACPF) domain and a C-terminal region containing a single-pass transmembrane domain. Proteins belonging to the MACPF superfamily are involved in different biological functions, but almost all of them play crucial roles in the immune defense against both extracellular and intracellular infections by means of two different mechanisms: either they form pores in the outer membrane of Gram negative bacteria or they deliver cytotoxic granzyme into target cells, thus inducing cell death.

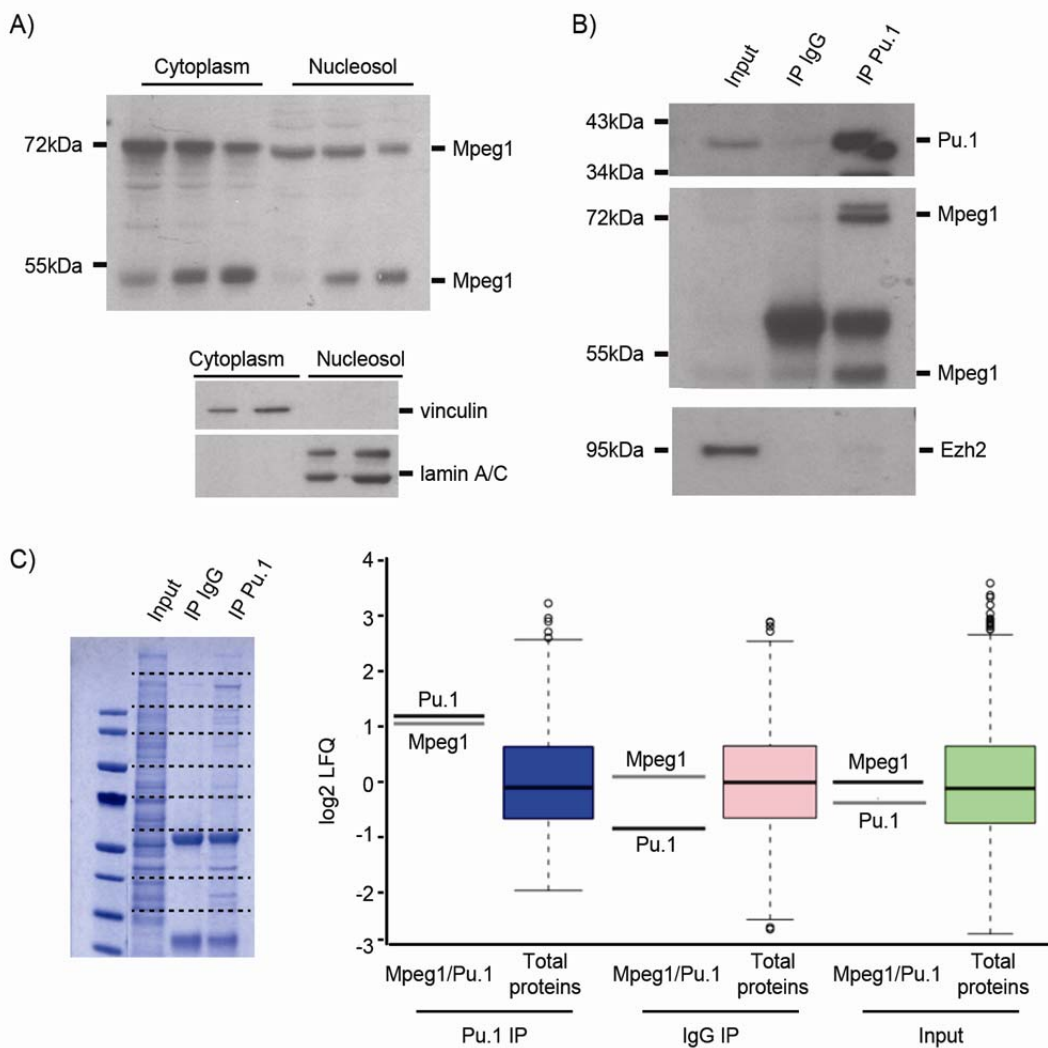
Mouse embryonic fibroblasts induce *mpeg1* expression in response to infection and their ability to kill the pathogen is lost upon *mpeg1* depletion (McCormack, de Armas et al. 2013). In human and murine cells, Mpeg1 is described to be either up- or down-regulated, depending on the different inflammatory stimulus; while in zebrafish two isoforms exist, Mpeg1 and Mpeg1.2, originated by gene duplication and typically regulated in opposite directions upon the same stimulus. The specific function of these two isoforms is still poorly characterized (Benard, Racz et al. 2014).

The reproducible enrichment of Mpeg1 in X-ChroP experiments carried out using both Pu.1 and H3K4me1 as baits was quite surprising and intriguing, because there were neither evidences in the literature concerning the involvement of this protein in the macrophage transcriptional response to inflammation, nor indication of its localization in the nucleus.

In order to validate the ChroP finding, we assessed Mpeg1 subcellular localization in RAW

264.7 cells. By western blot we observed that Mpeg1 is present both in cytosol and in the nucleus in a 60% to 40% ratio; moreover, the slightly lower molecular weight of the nucleosolic band of Mpeg1 compared to the cytosolic one would be compatible with the possible loss of the signal peptide, present at the N-terminus (Fig. 33A). Unexpectedly, we detected two distinct bands for Mpeg1: one at about 72kDa, in line with the predicted molecular weight, and a specific band around 50kDa which shows the same above-mentioned molecular shift between cytosolic and nucleosolic fraction. These two bands are specific, since both of them decrease upon Mpeg1 knock-down by shRNA (see below). We can exclude that the smaller band represents a splicing variant because *mpeg1* is constituted by only one exon; thus, we speculated that Mpeg1 might be subjected to cleavage by a still uncharacterized protease, which would generate the shorter isoform observed and a cleavage product of about 20kDa. By western blot we could not detect band around 20kDa reactive with the Mpeg1 antibody in use. Thus it is possible that either only the N-terminal region is stable upon cleavage, or that the 20kDa remnant is not recognized by the antibody. MS data did not help in understanding this mechanism, because we could only detected Mpeg1 peptides around the MACPF domain. To prove the protease cleavage hypothesis, it would be interesting to test a panel of protease inhibitors, to possibly identify the enzyme cleaving Mpeg1 and to characterize the functional differences between the full-length protein and the shorter isoform.

To confirm the interaction between Pu.1 and Mpeg1 and assess whether it occurs only on chromatin, we carried out a Pu.1 co-immunoprecipitation (co-IP) experiment by using RAW 264.7 nucleosolic fraction as input and IgG IP as negative control (Fig. 33B). We confirmed that the interaction exists and we observed that it is chromatin-independent: both bands recognized by the anti-Mpeg1 antibody were specifically immunoprecipitated in the nucleosolic co-IP using ant-Pu.1. This interaction was further assessed by label-free quantitative (LFQ) MS-proteomics analysis (Fig. 33C): the box-plot shows the specific



**Fig. 33: Characterization of Mpeg1 localization and dissection of chromatin-independent Pu.1 interaction.** **A)** WB analysis on cytosolic and nucleosolic localization of Mpeg1 (up), positive control for the subcellular fractionation (bottom). **B)** Chromatin-independent specific interaction between Pu.1 and Mpeg1 in Pu.1 nucleosolic IP; IgG IP and Ezh2 as negative controls for IP and WB, respectively. **C)** Coomassie-stained SDS-PAGE and MS-analysis of nucleosolic Pu.1 IP. Box-plots show the log<sub>2</sub> of the label-free quantitation (LFQ) value for Pu.1 and Mpeg1 proteins and for the general population of other proteins in Pu.1 IP, IgG IP and in the chromatin input; IgG IP is used as negative control.

enrichment of Mpeg1 in Pu.1 IP relative to the input, while the same enrichment is not observed in the IgG control IP.

All these experiments do not demonstrate a direct interaction between Pu.1 and Mpeg1; in order to test this hypothesis we are currently expressing a GST-tagged version of both the N-terminal and the C-terminal domains of Mpeg1.

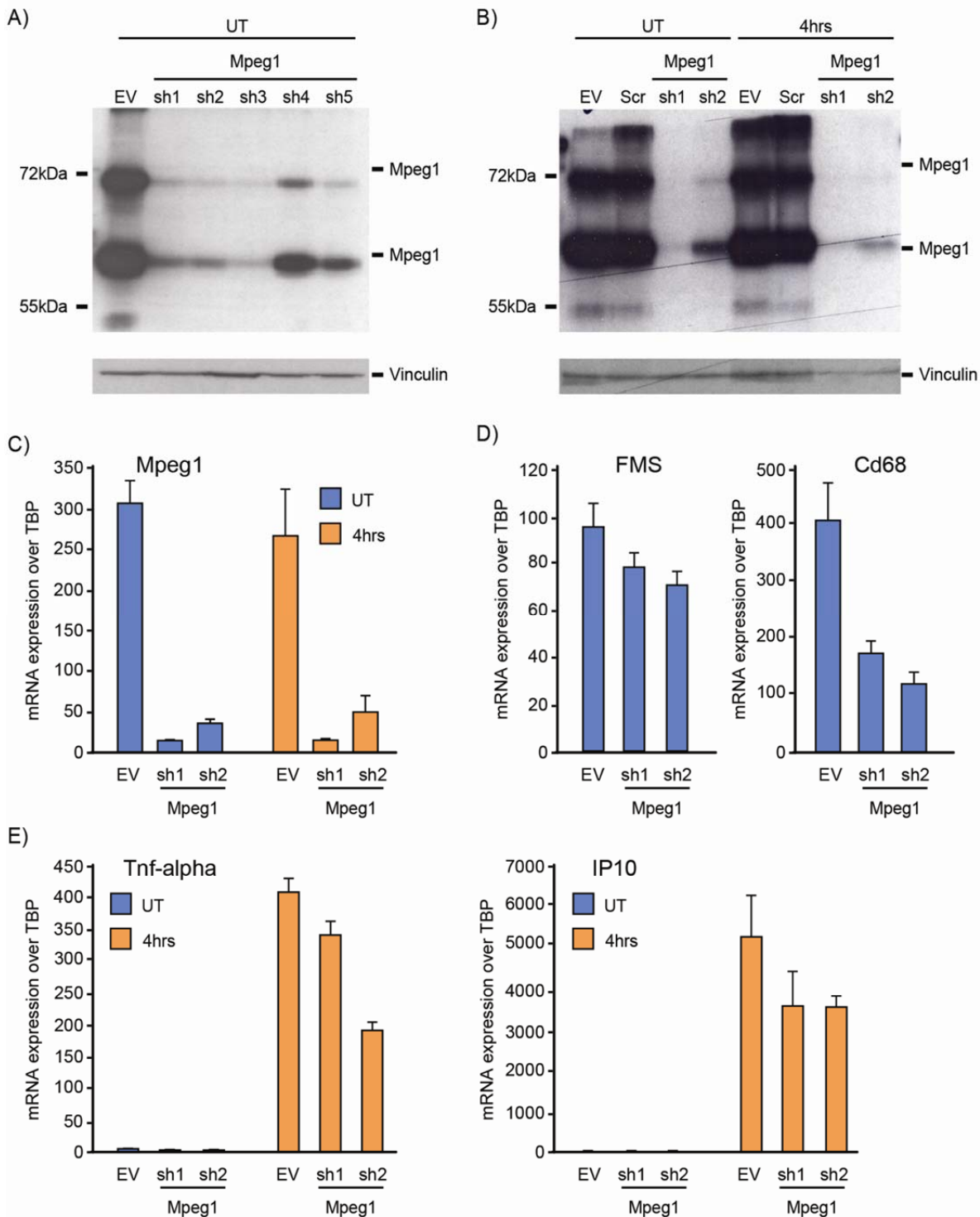
We already have available an expression vector containing Pu.1 cDNA that may be used

for the *in-vitro* translation and S35-methionine labeling of this transcription factor. The two recombinant products, namely N- or C-terminal GST-tagged Mpeg1 and S35-labelled Pu.1 will then be mixed and assayed in pull-down experiment to reveal the direct interaction between these two proteins.

We have also started the Mpeg1 knock-down experiments in primary bone marrow-derived murine macrophages (BMDMs), using short interfering RNAs (siRNAs) in order to assess the functional role of Mpeg1 in the expression of macrophage inflammatory genes, at basal state and upon inflammation. We designed and tested five different short hairpin RNAs (shRNAs), reported in Fig. 34A, in comparison with the empty vector and selected two shRNAs based on their increased knock-down efficiency assessed both at the mRNA and at protein level. These two shRNAs induce very efficient depletion of Mpeg1 after 48hrs, in both resting and activated BMDMs (Fig. 34B and C). We then screened a panel of 20 among macrophage-specific and inflammatory genes in both Mpeg1-depleted and control (empty vector transfected) BMDMs and observed a significantly decreased expression for FMS and Cd68 genes, two macrophage-specific genes coding for surface proteins important for macrophages recruitment and differentiation (Fig. 34D). This preliminary result could suggest a role of Mpeg1 in macrophage differentiation; we did not notice a significant delay in the differentiation of BMDMs in absence of Mpeg1 upon an initial assay; however this aspect deserves more in-depth investigation.

Remarkably, we also observed a decreased mRNA expression of some very interesting inflammatory genes, such as IP10 and Tnf- $\alpha$  (Fig. 34E), which are both involved in chemokine and cytokine secretion, respectively, upon the inflammatory burst.

Based on these encouraging but preliminary results, we have just started the RNA-seq analysis of BMDMs upon shRNA-depletion of Mpeg1 at both basal and inflamed state, to expand our investigation on the function of this protein in the regulation of development and inflammation.



**Fig. 34: Functional characterization of Mpeg1 depletion in macrophages.** **A)** Western blot analysis of the screening of shRNAs targeting Mpeg1 gene in bone marrow derived macrophages (BMDM) transfected with either empty vector (EV) or with shRNAs. **B)** WB analysis and **C)** quantitative PCR (qPCR) readout of BMDMs untreated (UT) or treated with LPS for 4hrs (4hrs) and transfected with either EV or the two shRNA constructs for Mpeg1 (sh1 and sh2). **D)** QPCR measurement of a set of macrophage-specific genes in UT BMDMs transfected with EV or sh1/sh2. **E)** Gene expression levels of LPS-responsive genes in BMDMs UT or LPS-treated and transfected with EV or sh1/sh2.

We are also setting ChIP-seq experiments in control and Mpeg1-depleted BMDMs to evaluate the genome-wide distribution and level of Pu.1, H3K4me1 and H3K27Ac in dependence of Mpeg1 in basal and stimulated condition.

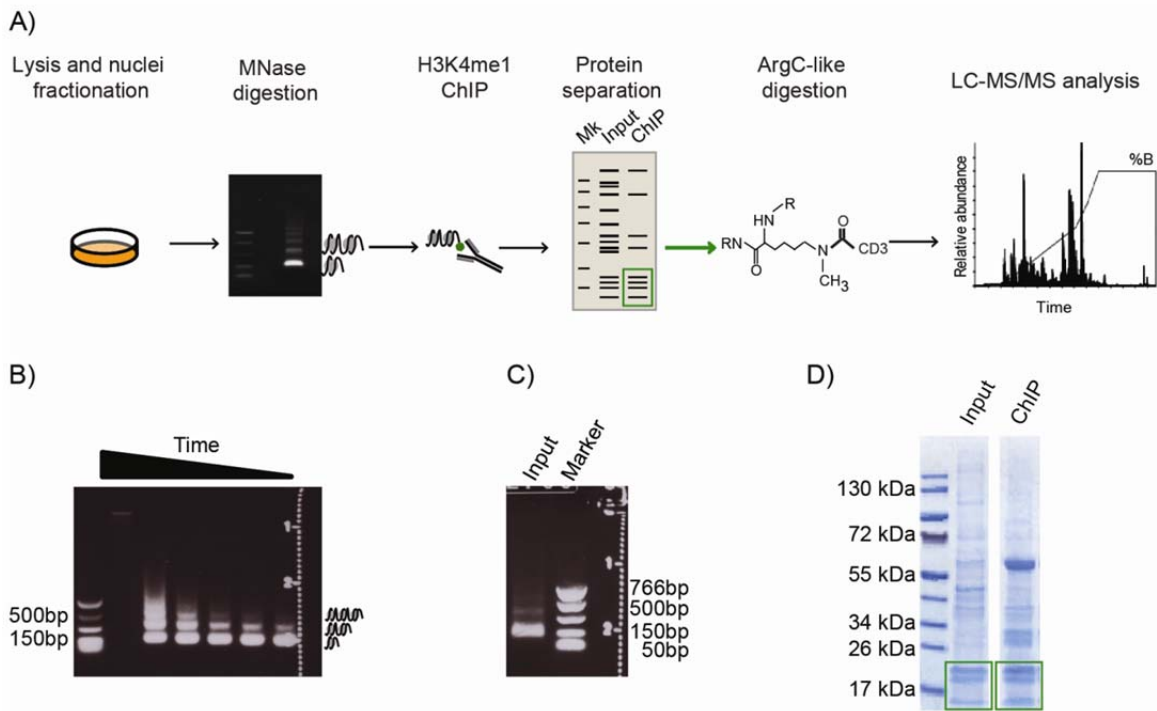
#### 5.4 Dissection of hPTMs associated with enhancers of inflammatory genes in macrophages.

In order to reconstruct the enhancers' chromatome in its completeness, we also investigated the histone post-translational modifications (hPTMs) associated with these *cis*-regulatory elements, the so-called "modificome". We employed the native-ChroP (N-ChroP) approach, recently established in our laboratory in HeLaS3 cells (Soldi and Bonaldi 2013). Similarly to X-ChroP, this strategy allows to comprehensively profile hPTMs co-enriching at distinct chromatin regions, biochemically immunopurified by ChIP using a defined mark as bait. The N-ChroP has the remarkable advantage of enabling the assessment of co-associating PTMs, not only at the intra-molecular level within the same histone, but also at the inter-molecular level, among the core-histones within the same intact mono-nucleosome, purified as a single entity.

##### 5.4.1 N-ChroP strategy to characterize the H3K4me1-modificome

We applied the standard N-ChroP strategy (Soldi and Bonaldi 2013) in RAW 264.7 cells, using H3K4me1 as bait.

The experimental approach is schematized in Fig. 35A: cells were first lysed and the nuclei were enriched by density centrifugation over a sucrose cushion. The purified nuclei were then digested using micrococcal nuclease S7 (MNase), monitoring the reaction over-time at fixed enzyme concentration (Fig. 35B) to obtain a chromatin fragment highly enriched in mono-nucleosomes (Fig. 35C). This fraction was used as input and incubated with anti-H3K4me1 antibody; the bound chromatin was then captured on magnetic beads and the



**Fig. 35: The N-ChroP strategy for the characterization of the modifichrome of H3K4me1-chromatin.** **A)** Schematic representation of the N-ChroP strategy. Unfixed cells are lysed and chromatin enzymatically fragmented prior to ChIP. Eluted proteins are separated via SDS and histones are ArgC-like digested and analysed by LC-MS/MS. **B)** Small scale MNase test digestion of purified nuclei. **C)** MNase-digested input prior to H3K4me1-ChIP. **D)** Coomassie stained SDS-PAGE of input and ChIP: green boxes indicate the histone bands excised for LC-MS/MS analysis.

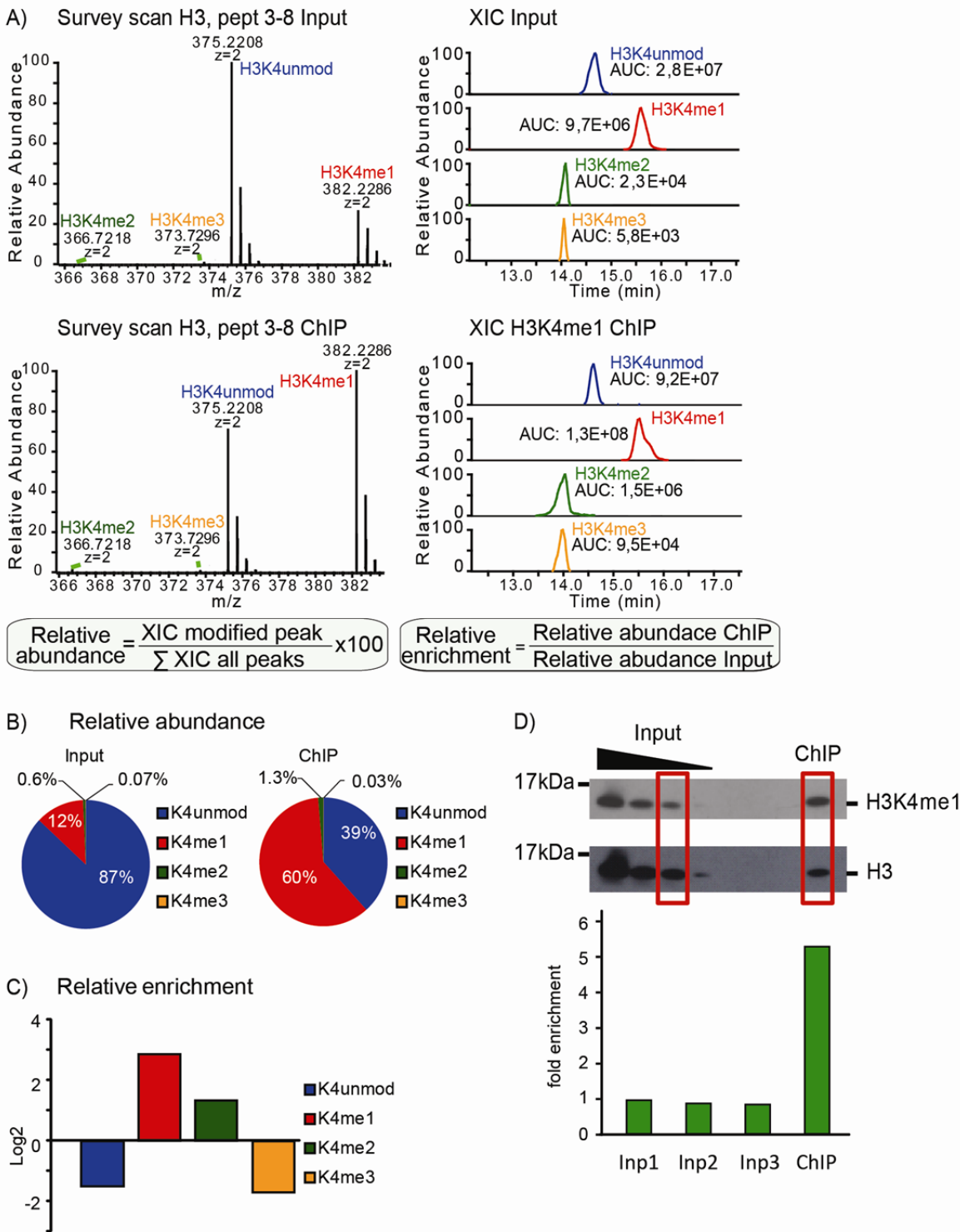
eluted proteins were separated *via* SDS-PAGE (Fig. 35D). Core histones from both the ChIP-ed material and the input were excised and subjected to ArgC-like digestion, to ensure the generation of peptides of optimal length for the subsequent LC-MS/MS analysis. Samples were analyzed on a high-resolution mass-spectrometer and the raw files were subjected to hPTM search using the MaxQuant software (Cox and Mann 2008; Cox, Neuhauser et al. 2011). Identified hPTMs were validated through manual inspection of the corresponding MS/MS fragmentation spectra, for exact modification site assignment.

In order to quantify the relative abundance of each distinct modified isoform in both ChIP and input, we used the intensity-based label-free strategy already described (Jung, Pasini et al. 2010; Soldi and Bonaldi 2013). Briefly, for each modified peptide isoform we reconstructed an extracted ion chromatogram (XIC), based on the theoretical exact mass (*See also Appendix 4*). Fig. 36A illustrates the survey scan (left) and the XIC

(right) for peptide 3 – 8 of histone H3 in the chromatin input (upper panels) and in the ChIP (bottom panels), when H3K4me1 is used as bait. The relative abundance of all the modified forms of a peptide is determined as the XIC of each modified peak over the sum of the XICs of all the modifications occurring on the same peptide, expressed as percentages (RA%). Then we determined the hPTM relative enrichment in the ChIP as the ratio between its RA% in the ChIP and the corresponding one in the input (Fig. 36A).

The relative abundances of the different modified forms of peptide H3 (3-8) were measured (Fig. 36B): in the input, 87% of K4 is unmodified, 12% is mono-methylated, while the di- and tri-methylated species account for 0.6% and 0.07% of the total, respectively. In the ChIP-ed material, 39% of K4 is unmodified, while 60% is mono-methylated and the di-/tri-methylated species still remain around the 1% in total (1.28% and 0.03%, respectively). The calculated relative enrichment indicates that mono-methylated K4 increases of about 5 fold (2.32 fold in log<sub>2</sub>) upon H3K4me1-N-ChroP. Importantly, both unmodified and tri-methylated K4 are depleted, suggesting that we could specifically enrich for enhancers (H3K4me1<sup>Hi</sup>/H3K4me3<sup>Low</sup>) (Heintzman, Stuart et al. 2007) and deplete the promoter regions, marked by H3K4me3. Di-methylation on lysine 4 is increased in the ChIP of about 2 fold (1 fold in log<sub>2</sub>); however its overall abundance remains around 1% (Fig. 36C).





**Fig.36: Intensity-based quantification of peptide 3 – 8 of histone H3 in H3K4me1 N-ChroP.** **A)** Survey scan (left) and XIC (right) for peptide 3 – 8 of H3 in the input (up) and in ChIP (bottom). This peptides contains Lys 4 that can be unmodified (blue), mono- (red), di- (green) or tri-methylated (orange); the relative abundance and the relative enrichment of a hPTM are calculated according to the shown formulas (AUC = area under the curve). **B)** Pie-charts representing the relative abundance of peptide H3 (3 – 8) forms in the input (left) and in ChIP (right). **C)** Log<sub>2</sub> relative enrichment of peptide 3 – 8 forms in H3K4me1 N-ChroP experiment. **D)** WB validation of H3K4me1 enrichment in the ChIP over the input; total H3 was used as normalizer for the fold enrichment calculation as  $\text{H3K4me1}(\text{ChIP}/\text{Inp})/\text{H3}(\text{ChIP}/\text{H3})$ .

We validated the enrichment of H3K4me1 in the ChIP relative to the input also by western blot (Fig. 36D), confirming about 5 fold enrichment of the bait, normalized on the total level of H3. These results suggest that the N-ChroP works efficiently also in RAW 264.7 cells using H3K4me1 as bait.

In addition to the expected enrichment of the bait and the depletion of the tri-methylated K4 (Fig. 37A) in H3K4me1 IPs, we observed a general decrease of H3K9me3, alone or in combination with acetylated K14 (K9me3 and K9me3/K14Ac, respectively), in accordance with the fact that this modification is a well-established mark of heterochromatin (Heard, Rougeulle et al. 2001; Peters, Kubicek et al. 2003). We also detected a depletion of H3K79me3, which has been associated with gene silencing (Barski, Cuddapah et al. 2007). Interestingly, the H3K79me2 was found enriched in our N-ChroP. This result confirms already published evidence suggesting that di-methylation of lysine 79 is mutually exclusive with the tri-methylation on the same residue and that H3K79me2 is typically enriched at actively transcribed genes (Sawado, Halow et al. 2008). As expected, we confirmed a global increase of the di-acetylated form of peptide (9 – 17) (K9Ac/K14Ac) and both the mono- and the di-acetylated forms of peptide (18 – 26) (K18Ac and K18/K23Ac). These results are in accordance with the general view that histone hyper-acetylation positively correlates with a relaxed chromatin conformation at actively transcribed regions (Marushige 1976). All these results confirm the robustness of our approach.

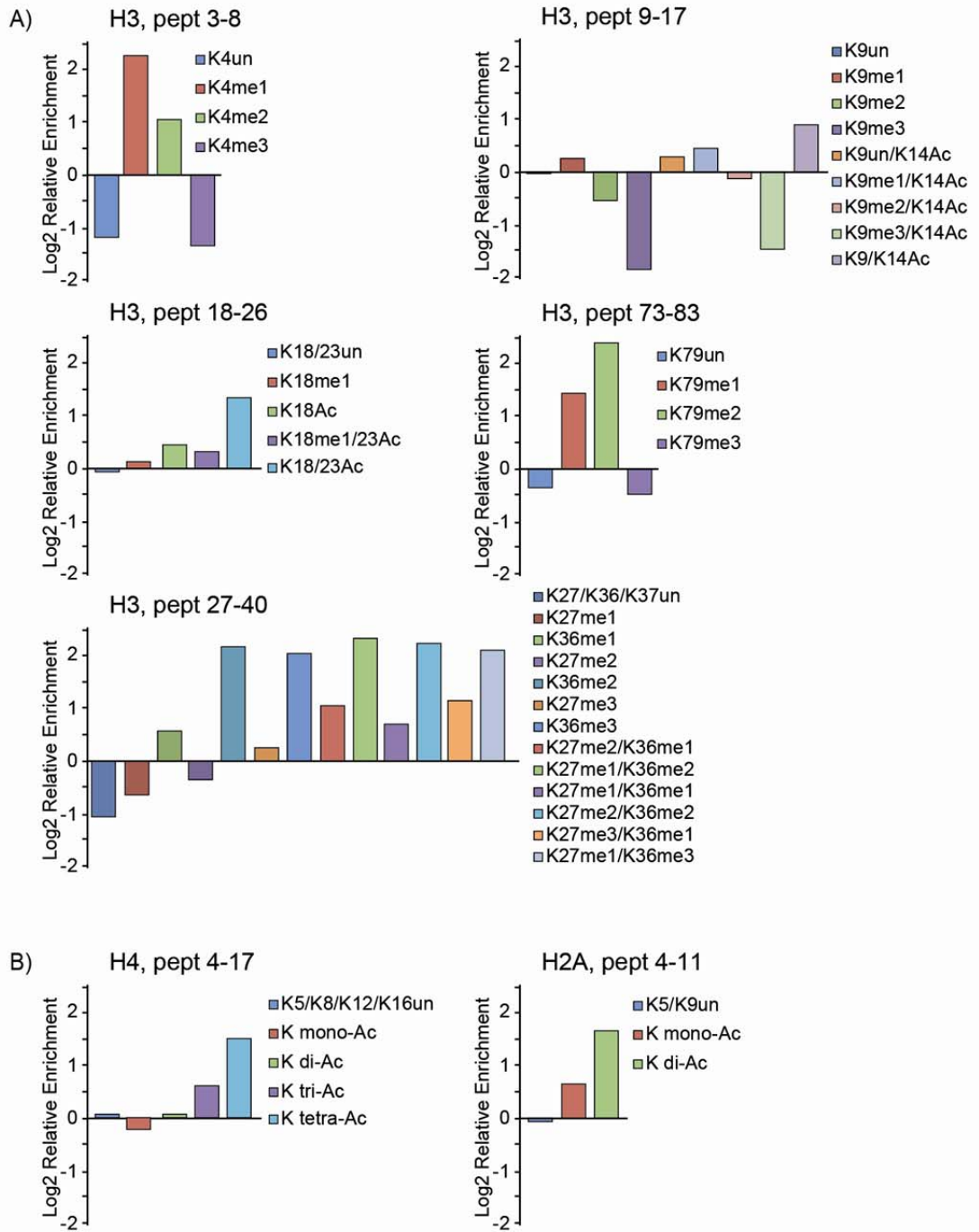
Due to their importance in enhancers' functional regulation, we extended the analysis to H3K27me3 and H3K27Ac, two markers that should allow distinguishing between poised and active enhancers, respectively (Creyghton, Cheng et al. 2010). Lysine 27 is at the N-terminus of peptide (27 – 40) produced upon ArgC-like digestion of histone H3; the unambiguous identification of modifications at this amino acid relies on the first *y-ion* (y1) and on all the *b-ions* produced upon fragmentation. Tri-methylation and acetylation

generate a very similar delta mass (+42.0469 Da and +42.0106 Da, respectively), that is however easily distinguishable by high-resolution mass-spectrometry. The use of Q Exactive allowed the correct modification assignment because of its “high-high” acquisition modality.

We obtained good fragmentation spectra that enabled the robust identification and quantification of tri-methylation of lysine 27. We did not observe any significant increase in ChIP, probably suggesting that we are enriching a pool of distinct H3K4me1-enhancers in different functional states, such as active, poised and inactive. Instead the mass spectra for H3K27Ac were not of a quality sufficient for a reliable identification and accurate quantification of this mark. This might be the consequence of the overall low abundance of the mark; to overcome this problem, we are setting a targeted-MS strategy to increase the signal-to-noise of the peak corresponding to this modification.

Surprisingly, within the same peptide we observed a strong increase of di- and tri-methylation at K36, alone or in combination with K27 mono- or di-methylation. Both K36me2 and K36me3 correlate positively with gene expression and with RNA polymerase II transcription elongation, together with H3K79me2 and with mono-ubiquitination of lysine 120 of histone H2B (H2BK120ub1) (Fuchs, Hollander et al. 2014). The enrichment of H3K36me2/3 levels at enhancers can be explained with twofold models: either we are enriching intragenic enhancers that are actively transcribed by the RNA polymerase II; or small non-coding RNAs transcription may occur at enhancers (eRNA), leading to deposition of these marks at these regions. In murine macrophages, about 30 – 40% of enhancers are intragenic (Ghisletti, Barozzi et al. 2010): a significant enrichment of this class of *enhancers* could indeed explain the observed increase of H3K36me3.

Interestingly, two new functions for H3K36me3 have been recently proposed: first, this mark could serve to distinguish, in combination with H3K4me1 and H3K27Ac, intragenic and extragenic active enhancers from inactive or poised ones (Zentner, Tesar et al. 2011).



**Fig. 37: Relative enrichment of hPTMs in H3K4me1 ChIP.** **A)** Relative enrichment of hPTMs in peptides 3-8, 9-17, 18-26, 27-40 and 73-83 of histone H3. **B)** Relative enrichment of hPTMs in peptide 4-17 of histone H4 and peptide 4-11 of histone H2A. Values of RE are expressed as log<sub>2</sub> values.

Second, H3K36me3 could work as good indicator for eRNA transcription in logistic regression model (Zhu, Sun et al. 2013). These considerations confirm that our result could be the consequence of the co-occurrence of both the two models. Additional ChIP-MS experiments could help to better dissect among these options. For instance, it could be

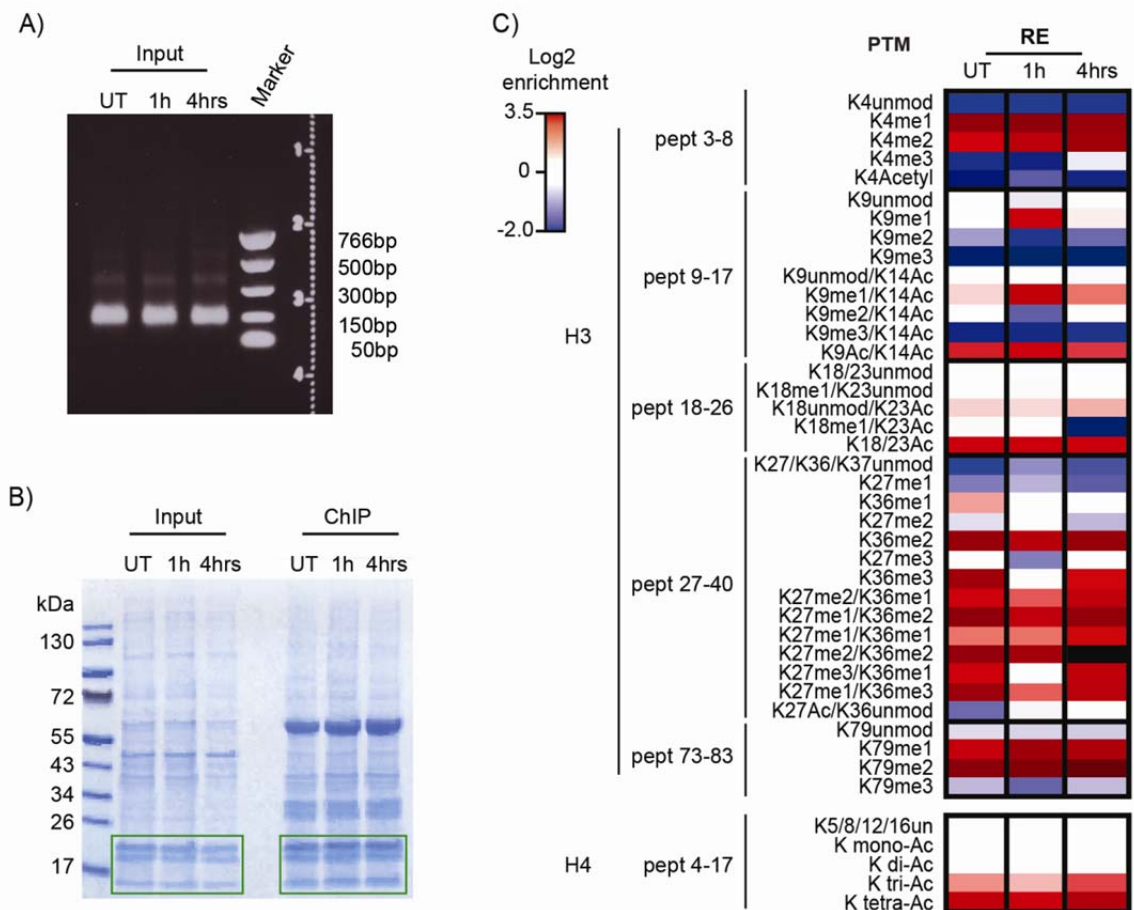
interesting to carry out N-ChroP experiments using H3K27Ac as bait to dissect active from poised enhancers. It would also be interesting to analyse the enhancers' modification upon RNase treatment to characterize eRNA role in H3K36me3 deposition.

As anticipated above, the N-ChroP approach allows also characterizing inter-molecular associations, namely hPTMs enriched on different histones within the same mono-nucleosome. By expanding our analysis we evaluated a general increase of multiply acetylated forms of the peptides 4 - 17 and 4 - 11 on histone H4 and H2A, respectively. This confirms the general enrichment of hyper-acetylated forms in H3K4me1-mono-nucleosomes (Fig. 37B).

To evaluate the dynamics of H3K4me1-modification upon an inflammatory stimulus, we carried out three N-ChroP experiments in parallel, using mono-nucleosomes from RAW 264.7 cells which were untreated or treated with LPS for 1 or 4hrs (Fig. 38A). Upon ChIPs, proteins were eluted and separated by SDS-PAGE (Fig. 38B) and core-histones were then analysed through gelLC-MS/MS, as above described. Histone PTMs relative enrichments in the three different conditions were calculated and are illustrated in the heat map in Fig. 38C.

In all the three experiments we confirmed a constant depletion of the heterochromatic marker H3K9me3, both alone and in combination with H3K14Ac. This effect is paralleled by a constant enrichment of the mono- and di-acetylated forms of the peptides (9 - 17) and (18 - 26) during the early inflammatory response. These trends were expected.

Preliminary analysis of the data revealed PTMs showing a dynamic behavior during the inflammatory response: H3K79 mono- and di-methylated are partially enriched, in line with the expected activation of gene transcription. The H3K36me2/me3-containing peptides showed different dynamics: H3K36me2 is stably enriched in all the conditions, while the association of this modification with H3K27me2 (K27me2/K36me2) shows a linear enrichment from untreated cells to 4hrs treated cells. We can explain this result



**Fig. 38: Histone PTMs within H3K4me1-regions profiled at basal conditions and upon inflammation.** **A)** Chromatin input from untreated (UT) or treated for 1h (1h) or 4hrs (4hrs) with LPS using the tc-N-ChroP approach. **B)** Coomassie-stained SDS-PAGE of the inputs and ChIPs from tc-N-ChroP using H3K4me1 as bait. **C)** Heat-map visualization of the relative hPTM enrichment for the marks identified in the H3K4me1-mononucleosomes and quantified through the intensity-based label-free approach. (Red = positive enrichment, blue = negative enrichment; results shown as log<sub>2</sub> of the relative enrichment).

either hypothesizing a higher number of accessible intragenic enhancers'

immunoprecipitated, or a higher rate of transcription occurring at these *cis*-regulatory regions upon LPS.

Unexpectedly, we also observed the enrichment of H3K9me1, alone and in combination with K14Ac (H3K9me1 or H3K9me1/H3K14Ac), at 1h and, to a lesser extent, at 4hrs upon LPS treatment. The presence of this mark has been described in association with H3K4me1 at the enhancers of genes activated during the differentiation of hematopoietic stem cells/progenitor cells into erythrocytes precursors (Cui, Zang et al. 2009). The increased

level of this mark in H3K4me1-regions could be explained by its specific deposition at the enhancer of very early-response genes, which are activated immediately after LPS stimulus, but already switched-off after 4hrs of LPS treatment. Such a model could also explain the observed depletion of H3K27me3, marker of poised enhancers, at the late time point.

We also observed a constant enrichment of the tri- and tetra-acetylated forms of peptide 4-17 of histone H4 during the inflammatory response; interestingly these enrichments seem to be slightly higher after 4hrs of LPS treatment, compatible with an accumulation of these marks in response to the further activation. Overall, additional analyses at shorter time-intervals are required to assess more in depth the dynamics of this model.

#### *5.4.2 Setup and initial assessment of a new strategy combining formaldehyde cross-linking with MNase digestion to analyse the modifecome of Pu.1-chromatin*

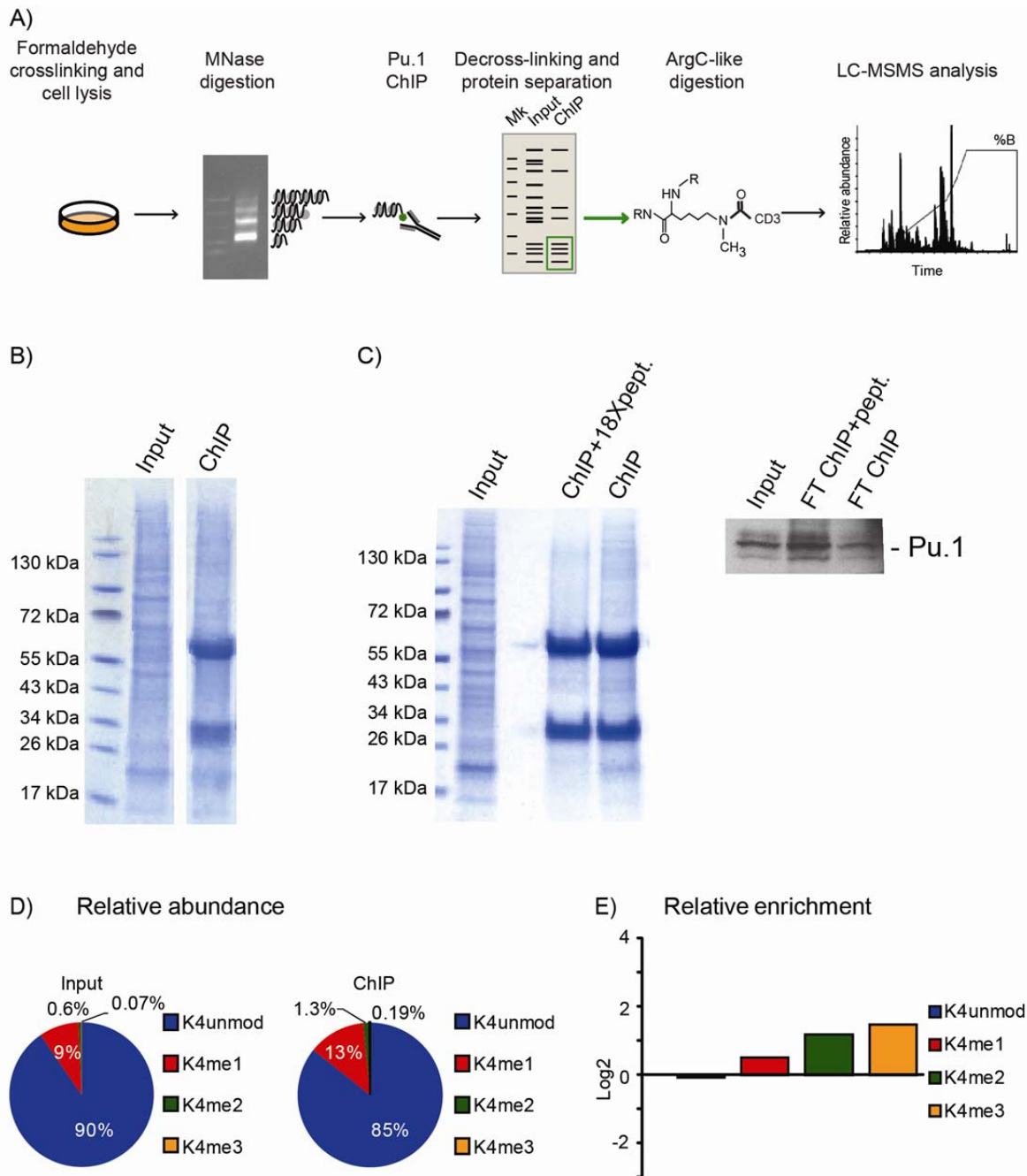
To dissect the Pu.1-modifecome, we modified the original N-ChroP protocol to use a transcription factor as bait, taking into account that TFs by definition bind in the nucleosome-free region (NFR). To achieve this goal we first stabilized with formaldehyde the interactions between the transcription factor and the underlying DNA, with the protocol already used in X-ChroP. Second, we combined the cross-linking with the MNase enzymatic digestion instead of sonication, to tightly regulate the DNA length and obtain a more homogeneous input for the Pu.1 ChIP.

Briefly, RAW 264.7 cells were cross-linked with formaldehyde and then purified nuclei were digested with a high concentration of MNase to cut the fixed chromatin. The reaction was monitored over time on an agarose gel at fixed MNase concentrations and the digestion was stopped when the input was mainly represented by mono-/di-/tri-nucleosomes. The input was incubated with antibody against Pu.1 and the eluted proteins

were separated by SDS-PAGE (Fig. 39A). Fig. 39B shows that we could successfully immunopurify an intact and stoichiometric histone octamer. Remarkably, this enrichment is specific, since it is abolished by competing the anti-Pu.1 antibody with the specific peptide used in X-ChroP experiment (*for details se 3.2.2 and material and methods*). The competition reflects the increase of the Pu.1 protein in the competed FT fraction relative to the non-competed one (Fig. 39C). All together these results indicate that we can enrich the chromatin fraction specifically associated with Pu.1.

We then focused on the analysis of hPTMs by MS. Firstly, we evaluated the relative abundance and enrichment of modifications on peptide H3 (3 – 8). H3K4me1 represents the positive control because of its known association with Pu.1. The RA% calculation of the input recapitulates what we already found in the N-ChroP input: 90% of K4 is unmodified, 9% is mono-methylated, while the remaining di- and tri-methylated K4 account for less than 1% (0.6% and 0.07%, respectively). Upon Pu.1 enrichment, K4me1 increases up to 13%, while the unmodified K4 slightly diminishes (85%) and di- and tri-methylated K4 account for 1.3 and 0.19% respectively (Fig. 39D). However, the relative enrichment of 1.4 fold for K4me1 is not enough to conclude that the X-ChIP worked successfully using the present experimental conditions. In fact, we also observed a relative enrichment of about 2 fold for H3K4me2 and 2.5 fold for H3K4me3 (1 fold and 1.3 fold in log<sub>2</sub>, respectively), even though their overall percentage account for maximum 1% of the total modifications at this residue (Fig. 39E).





**Fig. 39: X-ChroP/MNase combined strategy for the characterization of the modifcome of Pu.1-mononucleosomes.** **A)** Schematic representation of X-ChroP/MNase combined strategy. Formaldehyde-fixed cells are lysed and chromatin is enzymatically fragmented prior Pu.1 ChIP. Eluted proteins are de-crosslinked and separated via SDS-PAGE; histone bands are ArgC-like digested and analysed through LC-MSMS. **B)** Coomassie stained SDS-PAGE shows the Input and ChIP from Pu.1 X-ChroP/MNase experiment. **C)** Coomassie stained SDS-PAGE shows that the 18X fold peptide competition completely abolishes the specific nucleosome enrichment in the competed ChIP (left). This effect is mirrored in the WB (right) by the increase of Pu.1 levels in the corresponding flow-through (FT). **D)** Pie charts representing the relative abundance of peptide 3 – 8 of histone H3 in the input (left) and in the ChIP (right). **E)** Log<sub>2</sub> relative enrichment of peptide 3 – 8 in H3K4me1 X-ChroP/MNase combined experiment.

These observations suggest that the protocol needs further optimization to obtain a more efficient H3K4me1-enrichment and a corresponding H3K4me3 depletion, to achieve a more faithful representation of enhancers.



## **DISCUSSION**

In this thesis I describe how I carried out the in depth characterization of enhancers' protein composition in the macrophage-like cell line RAW 264.7 by using a combined approach based on biochemical methods for chromatin fractionation and enrichment and MS-based proteomics. We performed this analysis both in the basal and inflamed states, achieving a temporal analysis of the global transcriptional response elicited in macrophages by lipopolysaccharide.

We investigated the enhancers' chromatome based on the assumption that the protein composition and architecture of these regulatory regions strongly correlate with their functional state. Enhancers are well characterized functionally but information on their protein determinants is limited: they are defined as genomic sequences extensively bound by transcription factors thus leading to reduced nucleosome occupancy. Markers of these regions are hyper-sensitivity to DNase I and enrichment of mono-methylation of lysine 4 of histone H3 (H3K4me1), in the presence of low tri-methylation of the same residue (H3K4me1<sup>Hi</sup>/H3K4me3<sup>Low</sup>) [Heintzman et al 2007]. In macrophages, enhancers are further marked by the master regulator Pu.1, a pioneer TF that allows the recruitment of stimulus-specific transcription factors in response to distinct inflammatory stimuli, necessary for fine-tuning gene expression of different set of genes in time and space.

To study the enhancers' chromatome we employed the ChroP approach recently established by our laboratory. This strategy combines the ChIP technique and MS-based proteomics, thus allowing the characterization of both the hPTMs pattern (modificome) and the interactors (interactome) associated with a chromatin landscape marked by a specific hPTM. We enriched biochemically enhancers by using both H3K4me1 and Pu.1 as bait in the preparative ChIP.

The setup of ChroP in this model system required the optimization of different steps, such as: the SILAC labeling of RAW 264.7; the appropriate cross-linking conditions

to stabilize TF-DNA interaction for the subsequent MS analysis; the setup of a time-course experiment for the temporal analysis of enhancers during the early inflammatory response. Based on the SILAC ratio distribution we defined statistical cut-offs to discern specific binders from background proteins. One of the top scoring proteins in the H3K4me1-interactome was Pu.1 itself, thus demonstrating that in macrophages enhancer regions are truly marked by the co-occurrence of this hPTM and the master regulator. In genome-wide ChIP-seq analysis this information could only be addressed by carrying out sequential ChIPs, followed by sequencing and tracks' alignment.

We considered as genuine enhancers' determinants proteins enriched in both H3K4me1 and Pu.1 interactomes, identifying enhancer-binding proteins, transcription factors/co-activators and well-characterized markers of macrophage enhancers', such as the histone acetyl-transferase p300. Moreover, we observed an enrichment of the cohesin complex and of several SWI-SNF-associated complexes that may be involved in both positive transcription regulation (like the PBAF and the RNA polymerase II-associated complex) and negative transcriptional regulation (like the Brg1-Sin3a complex) or whose function may vary in dependence on the functional state (such as the LAR complex). Interestingly, we also identified two complexes related to RNA processing, namely Emerin 24 complex and p54nrb-PSF-matrin3 complex, and the spliceosome complex. The presence of these complexes as intrinsic part of the enhancers' architecture could be in accordance with the co-occurrence of gene transcription and mRNA processing. We also observed an unexpected enrichment of the MCM complex. In addition to an established function in DNA replication, some subunits of MCM have been recently described to associate with RNA polymerase II, thus suggesting a possible new role for this complex in the Stat1 target gene activation and RNA polymerase II-mediated transcription. In order to discern between enhancer-specific enrichment and chromatin-association of these unexpected complexes, it would be interesting to compare the data we have acquired with

experimental data coming from a completely unrelated landscape where we do not expect the presence of such complexes, such as the hetero-chromatic H3K9me3-marked chromatin.

In the time-course experiment, we profiled all these interactors during the early phases of the inflammatory response by treating RAW 264.7 cells for 1 or 4 hrs with lipopolysaccharide. These analyses revealed three dynamic scenarios: proteins that remain stably associated with these chromatin regions, proteins recruited to enhancers and proteins evicted upon the stimulus. Through our proteomics analysis we could conclude that enhancers are composed by a high-ordered structure which is quite stable and maintained throughout the inflammatory response. However, we also identified some dynamic changes in the enhancers' composition upon LPS treatment. Remarkably, Pbrm1 and Smarcc2 were both recruited during the inflammatory response; these proteins are subunits of the PBAF complex which is the only member of the SWI/SNF-related family that links chromatin remodeling activity with positive gene expression regulation. PcnA and Mki67 proteins are markers of DNA replication and were found to be dynamically displaced from enhancers during the inflammatory response, together with the DNA methyl-transferase Dnmt1, which is associated to gene silencing.

The time-course X-ChroP experiment also revealed a group of proteins newly recruited at enhancers upon LPS treatment. The induction of some of them during inflammation had been previously reported. This is for instance the case for the transcription factors Junb and Stat1, both of which promote activation of target inflammatory genes. Our finding suggests that the other newly recruited proteins may be involved in the activation of enhancers upon LPS treatment. Among them we found the enhancer binding protein Cebpz, the transcription elongation factor Supt6h and two PHYIN-domain-containing proteins (Mndal and Ifi204), involved in transcriptional regulation. To gain further insights into the role of these determinants in the regulation of

gene expression upon inflammation, we intend to carry out the ChIP-seq analysis of their genomic localization prior and during the inflammatory response; moreover, we will have available the genomic profiles of both the baits used in the tc-ChroP and of many other proteins with putative role as TFs from collaborators at IEO. We plan to compare the ChIP-seq profiles of these known and novel determinants, to confirm whether they localize at enhancers and to assess whether they synergize/cooperate with Pu.1 or other enhancers' markers.

Altogether our results suggest that macrophages respond to the inflammatory stimulus by stopping cell proliferation and committing towards a global gene transcriptional activation, which correspond to certain extent of chromatin remodeling and DNA de-methylation at enhancers. The finely-tuned gene expression regulation of inflammatory genes in time and space is achieved by the cell through the interplay between positive and negative transcriptional regulators. Moreover macrophages enhancers seem to be characterized by an overall stable higher-order structure, in line with a model that suggests the existence of a chromosomal-looping bridging together enhancer and promoter regions of corresponding genes. This higher-ordered structure seems to exist prior to the transcriptional activation induced by the inflammatory stimulus and is maintained throughout the inflammatory response, with only minor internal dynamics of specific factors. The plasticity of some chromatin determinants favors the activation of specific set of genes while the recruitment of distinct TFs and transcriptional co-activators serves to activate specific gene expression programs at the required time upon stimulus. Data from chromosome-conformation capturing (3C)-based sequencing technologies suggest a model where the majority of chromatin interactions, namely enhancers-promoter looping, are already formed within a cell and represent a stable architecture. Upon a specific stimulus, only some of these loops are reorganized while others maintain the previous structure; remarkably, this model is in perfect agreement with our findings.

The model of the chromosomal looping between enhancers and promoters as mechanism of transcription regulation is still under debate and our findings are not sufficient to prove or discard this model. To immunopurify enhancers, we used ChIP-grade antibodies against the two known markers that better define these *cis*-regulatory regions in macrophages, namely the H3K4me1 and the TF and master regulator of macrophages Pu.1. In the X-ChroP approach, as in many other techniques used to study chromatin composition, we utilized formaldehyde-fixed chromatin to stabilize DNA-/ protein-protein interactions, therefore it would be difficult to discriminate between determinants associated with promoters or enhancers in a scenario of chromosomal looping where these two *cis*-regulatory regions bridge together and appear as a unique and well-structured genomic entity that fine tunes gene transcription.

The X-ChroP strategy allows the unbiased characterization of a specific chromatin region and in our analysis we identified expected and already known determinants that have a central role in the enhancer regulation, thus confirming the robustness of the approach itself; moreover we characterized novel and potentially very interesting determinants both in basal condition and during the transcriptional activation response. Currently, we are working on the validation and follow-up of some of these novel determinants and therefore these findings may expand the knowledge on enhancer composition as well as on their mechanism of action during inflammatory response.

Mass-spectrometry is a well-established and powerful tool also for the analysis of PTMs without *a priori* knowledge. Histone post-translational modifications (hPTMs) represent an additional layer of epigenetic determinants that act in concert with TFs and chromatin remodelers to regulate the functional state of the underlying DNA. We achieved the dissection of the enhancers' modifecome through the ChroP approach using H3K4me1 and Pu.1 as baits. We applied to RAW 264.7 cells the already optimized N-ChroP protocol to purify of mono-nucleosomes enriched in H3K4me1 and analyzed by MS PTMs co-



associations, not only within the same histone molecule but also among distinct core-histones within the same mono-nucleosome.

Through the analysis of three parallel N-ChroP experiments, we performed the dynamic dissection of the H3K4me1-modificome starting from cells untreated or treated for 1h or 4hrs with LPS. Chromatin regions marked by H3K4me1 revealed an overall enrichment for markers of active chromatin, like the hyper-acetylated forms of H3 and H4 and H3K79me2; this enrichment was mirrored by a corresponding depletion of heterochromatic marks, such as H3K9me3. Unexpectedly, we observed the enrichment of H3K36me2 and K36me3, two markers typically associated with transcription elongation, which were further increased upon induction of the inflammatory response. Twofold models may explain this result: first, in macrophages around 30 – 40% of enhancers are intragenic, thus correspond to regions highly transcribed that are enriched with these markers. A second possibility may be occurrence of lncRNAs transcription at enhancers, the so-called eRNAs, which are proposed to promote deposition of these marks at transcribed *cis*-regulatory regions. Although the transcription of lncRNAs at enhancers is extensively described, their precise function is still under debate. At present, we cannot distinguish between these two hypotheses; one possibility could be to carry out ChroP experiment in presence of RNase to impair eRNA production and then assess the stability of the higher-order structure characterized and the presence of the corresponding modificome. We also detected a very rapid enrichment of H3K9me1 after 1hr of LPS, followed by a decrease to almost basal condition at 4 hours, which is mirrored by a quick depletion of H3K27me3 at the same time point. These two observations may be explained by the activation of a class of very early enhancers, which loose very rapidly the poised mark and acquire the H3K9me1 mark to activate target genes quickly, and then return with the same fast kinetic to a basal state.

A well-established marker of active enhancers in macrophages is represented by

H3K27Ac: in general, repressed genes lose this modification upon transcriptional activation, while active genes/enhancers acquire it. This mark is typically investigated *via* antibody-based approaches, indeed the correct site-specific assignment of modifications is particularly challenging by MS, since this modification is located at the N-terminal of the peptide H3 (27 – 40) generated by ArgC-like digestion, thus its identification only relies on the first *y*-ion and on the *b*-ion series. High-resolution mass spectrometry allows distinguishing with high confidence between tri-methylation and acetylation on this residue; however the very low abundance of H3K27Ac hampers its identification and quantification. Due to the functional relevance of this mark especially in the context of enhancers, we are currently establishing MS-targeted strategies to better isolate the peptide bearing this modification, to increase its signal-to-noise and obtain a better-quality fragmentation spectrum. To quantify this modification more reliably, we plan also to synthesize isotopically-labeled peptides, such as AQUA peptides, bearing H3K27Ac alone or in combination with other modifications. The use of AQUA peptides as spike-in will help building the optimal targeted method for accurate quantification by following the elution profile and fragment ions of the synthetic peptide.

Based on the data so far acquired, we envisage that an interesting next step will be to extend our analysis to other types of modifications co-occurring within the same enriched chromatin regions, such as arginine methylation and lysine ubiquitination, crotonylation and propionylation. The MS raw data already acquired in fact represent a valuable repository to gain information about more exotic PTMs and their possible combinations specifically enriched and functionally active at enhancers.

Interestingly, our temporal analysis of enhancers' determinants revealed the dynamic recruitment of the Polybromo protein 1 (Pbrm1). This protein contains six different bromo-domains with high specificity for the acetylation of lysine residues on histone H3, namely K4, K9, K14, K18 and K23. Remarkably, we identified the enrichment

of 4 out of 5 of these acetylated residues during the inflammatory stimulus, which function as docking sites for the recruitment of the PBAF complex. As such this represents an emblematic case, where the combined analysis of N- and X-ChroP enables observing the enrichment of a certain histone mark that serves to recruit the corresponding “reader”, which finally promotes gene activation through chromatin remodeling.

More in general, I foresee that the combination of ChroP and ChIP-seq can lead to the comprehensive characterization of a distinct chromatin region from two different but tightly complementary perspectives. While ChroP allows for a “blind” and unbiased characterization of known and unexpected chromatin determinants, their genome-wide localization can be then validated and investigated by a “targeted” ChIP-seq strategy.

**REFERENCES**

- Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*. 1981; 27(2 Pt 1): 299-308.
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007; 129(4): 823-37.
- Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell*. 2010; 143(3): 470-84.
- Beck HC. Mass spectrometry in epigenetic research. *Methods Mol Biol*. 2010; 593: 263-82.
- Benard EL, Racz PI, Rougeot J, Nezhinsky AE, Verbeek FJ, Spaink HP, et al. Macrophage-Expressed Perforins Mpeg1 and Mpeg1.2 Have an Anti-Bacterial Function in Zebrafish. *Journal of innate immunity*. 2014.
- Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M. A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nature biotechnology*. 2003; 21(3): 315-8.
- Blagoev B, Mann M. Quantitative proteomics to study mitogen-activated protein kinases. *Methods*. 2006; 40(3): 243-50.
- Bonaldi T, Imhof A, Regula JT. A combination of different mass spectroscopic techniques for the analysis of dynamic changes of histone modifications. *Proteomics*. 2004; 4(5): 1382-96.
- Bonaldi T, Regula JT, Imhof A. The use of mass spectrometry for the analysis of histone modifications. *Methods in enzymology*. 2004; 377: 111-30.
- Bremang M, Cuomo A, Agresta AM, Stugiewicz M, Spadotto V, Bonaldi T. Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. *Molecular bioSystems*. 2013; 9(9): 2231-47.
- Byrum SD, Raman A, Taverna SD, Tackett AJ. ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus. *Cell reports*. 2012; 2(1): 198-205.
- Byrum SD, Taverna SD, Tackett AJ. Purification of a specific native genomic locus for proteomic analysis. *Nucleic acids research*. 2013; 41(20): e195.
- Chandrasekaran R, Thompson M. Polybromo-1-bromodomains bind histone H3 at specific acetyl-lysine positions. *Biochemical and biophysical research communications*. 2007; 355(3): 661-6.
- Chang CS, Sun HL, Lii CK, Chen HW, Chen PY, Liu KL. Gamma-linolenic acid inhibits inflammatory responses by regulating NF-kappaB and AP-1 activation in lipopolysaccharide-induced RAW 264.7 macrophages. *Inflammation* 2010; 33(1): 46-57
- Cho H, Orphanides G, Sun X, Yang XJ, Ogryzko V, Lees E, et al. A human RNA polymerase II complex containing factors that modify chromatin structure. *Molecular and cellular biology*. 1998; 18(9): 5355-63.
- Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*. 2008; 26(12): 1367-72.
- Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research*. 2011; 10(4): 1794-805.

- Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(50): 21931-6.
- Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, et al. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell stem cell*. 2009; 4(1): 80-93.
- Dallas PB, Cheney IW, Liao DW, Bowrin V, Byam W, Pacchione S, et al. p300/CREB binding protein-related protein p270 is a component of mammalian SWI/SNF complexes. *Molecular and cellular biology*. 1998; 18(6): 3596-603.
- Dauffy J, Mouchiroud G, Bourette RP. The interferon-inducible gene, *Ifi204*, is transcriptionally activated in response to M-CSF, and its expression favors macrophage differentiation in myeloid progenitor cells. *Journal of leukocyte biology*. 2006; 79(1): 173-83.
- de Godoy LM, Olsen JV, de Souza GA, Li G, Mortensen P, Mann M. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome biology*. 2006; 7(6): R50.
- De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS biology*. 2010; 8(5): e1000384.
- Dejardin J, Kingston RE. Purification of proteins associated with specific genomic Loci. *Cell*. 2009; 136(1): 175-86.
- Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*. 2002; 295(5558): 1306-11.
- Dickins RA, Hemann MT, Zilfou JT, Simpson DR, Ibarra I, Hannon GJ, et al. Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nature genetics*. 2005; 37(11): 1289-95.
- Emmett MR, Caprioli RM. Micro-Electrospray Mass-Spectrometry - Ultra-High-Sensitivity Analysis of Peptides and Proteins. *Journal of the American Society for Mass Spectrometry*. 1994; 5(7): 605-13.
- Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science*. 1989; 246(4926): 64-71.
- Ficarro SB, McClelland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, et al. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nature biotechnology*. 2002; 20(3): 301-5.
- Frese CK, Nolting D, Altelaar AF, Griep-Raming J, Mohammed S, Heck AJ. Characterization of electron transfer dissociation in the Orbitrap Velos HCD cell. *Journal of the American Society for Mass Spectrometry*. 2013; 24(11): 1663-70.
- Fuchs G, Hollander D, Voichek Y, Ast G, Oren M. Cotranscriptional histone H2B monoubiquitylation is tightly coupled with RNA polymerase II elongation rate. *Genome research*. 2014; 24(10): 1572-83.
- Fujihara M, Muroi M, Muroi Y, Ito N, Suzuki T. Mechanism of lipopolysaccharide-triggered junB activation in a mouse macrophage-like cell line (J774). *The Journal of biological chemistry*. 1993; 268(20): 14898-905.
- Fuller-Pace FV. DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. *Nucleic acids research*. 2006; 34(15): 4206-15.

- Garcia BA, Mollah S, Ueberheide BM, Busby SA, Muratore TL, Shabanowitz J, et al. Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature protocols*. 2007; 2(4): 933-8.
- Garcia BA, Shabanowitz J, Hunt DF. Characterization of histones and their post-translational modifications by mass spectrometry. *Current opinion in chemical biology*. 2007; 11(1): 66-73.
- Geiger T, Cox J, Ostasiewicz P, Wisniewski JR, Mann M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature methods*. 2010; 7(5): 383-5.
- Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, Venturini E, et al. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity*. 2010; 32(3): 317-28.
- Giddings JC. Two-dimensional separations: concept and promise. *Analytical chemistry*. 1984; 56(12): 1258A-60A, 62A, 64A passim.
- Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & cellular proteomics : MCP*. 2012; 11(6): O111 016717.
- Gomard T, Michaud HA, Tempe D, Thiolon K, Pelegrin M, Piechaczyk M. An NF-kappaB-dependent role for JunB in the induction of proinflammatory cytokines in LPS-activated bone marrow-derived dendritic cells. *PloS one*. 2010; 5(3): e9585.
- Griffin PR, Coffman JA, Hood LE, Yates JR. Structural-Analysis of Proteins by Capillary Hplc Electrospray Tandem Mass-Spectrometry. *International Journal of Mass Spectrometry and Ion Processes*. 1991; 111: 131-49.
- Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology*. 1999; 17(10): 994-9.
- Heard E, Rougeulle C, Arnaud D, Avner P, Allis CD, Spector DL. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell*. 2001; 107(6): 727-38.
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics*. 2007; 39(3): 311-8.
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*. 2010; 38(4): 576-89.
- Hilger M, Mann M. Triple SILAC to determine stimulus specific interactions in the Wnt pathway. *Journal of proteome research*. 2012; 11(2): 982-94.
- Hillenkamp F, Karas M. Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. *Methods in enzymology*. 1990; 193: 280-95.
- Holaska JM, Wilson KL. An emerin "proteome": purification of distinct emerin-containing complexes from HeLa cells suggests molecular basis for diverse roles including gene regulation, mRNA splicing, signaling, mechanosensing, and nuclear architecture. *Biochemistry*. 2007; 46(30): 8897-908.
- Holland L, Gauthier L, Bell-Rogers P, Yankulov K. Distinct parts of minichromosome maintenance protein 2 associate with histone H3/H4 and RNA polymerase II holoenzyme. *European journal of biochemistry / FEBS*. 2002; 269(21): 5192-202.

- Icardi L, Mori R, Gesellehen V, Eyckerman S, De Cauwer L, Verhelst J, et al. The Sin3a repressor complex is a master regulator of STAT transcriptional activity. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109(30): 12058-63.
- Ishihama Y, Sato T, Tabata T, Miyamoto N, Sagane K, Nagasu T, et al. Quantitative mouse brain proteomics using culture-derived isotope tags as internal standards. *Nature biotechnology*. 2005; 23(5): 617-21.
- Ishimi Y, Ichinose S, Omori A, Sato K, Kimura H. Binding of human minichromosome maintenance proteins with histone H3. *The Journal of biological chemistry*. 1996; 271(39): 24115-22.
- Ju BG, Solum D, Song EJ, Lee KJ, Rose DW, Glass CK, et al. Activating the PARP-1 sensor component of the groucho/ TLE1 corepressor complex mediates a CaMKinase IIdelta-dependent neurogenic gene activation pathway. *Cell*. 2004; 119(6): 815-29.
- Jufvas A, Stralfors P, Vener AV. Histone variants and their post-translational modifications in primary human fat cells. *PloS one*. 2011; 6(1): e15960.
- Jung HR, Pasini D, Helin K, Jensen ON. Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. *Molecular & cellular proteomics : MCP*. 2010; 9(5): 838-50.
- Kelleher NL, Zubarev RA, Bush K, Furie B, Furie BC, McLafferty FW, et al. Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid. *Analytical chemistry*. 1999; 71(19): 4250-3.
- Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature*. 2010; 465(7295): 182-7.
- Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, et al. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature structural & molecular biology*. 2011; 18(8): 956-63.
- Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. *Science*. 1974; 184(4139): 868-71.
- Kouzarides T. SnapShot: Histone-modifying enzymes. *Cell*. 2007; 131(4): 822.
- Krivega I, Dale RK, Dean A. Role of LDB1 in the transition from chromatin looping to transcription activation. *Genes & development*. 2014; 28(12): 1278-90.
- Kuwahara S, Ikei A, Taguchi Y, Tabuchi Y, Fujimoto N, Obinata M, et al. PSPC1, NONO, and SFPQ are expressed in mouse Sertoli cells and may function as coregulators of androgen receptor-mediated transcription. *Biology of reproduction*. 2006; 75(3): 352-9.
- Kuzmanov A, Karina EI, Kirienko NV, Fay DS. The conserved PBAF nucleosome-remodeling complex mediates the response to stress in *Caenorhabditis elegans*. *Molecular and cellular biology*. 2014; 34(6): 1121-35.
- Lambert JP, Mitchell L, Rudner A, Baetz K, Figeys D. A novel proteomics approach for the discovery of chromatin-associated protein networks. *Molecular & cellular proteomics : MCP*. 2009; 8(4): 870-82.
- Lee CG, Jenkins NA, Gilbert DJ, Copeland NG, O'Brien WE. Cloning and analysis of gene regulation of a novel LPS-inducible cDNA. *Immunogenetics*. 1995; 41(5): 263-70.
- Lemon B, Inouye C, King DS, Tjian R. Selectivity of chromatin-remodelling cofactors for ligand-activated transcription. *Nature*. 2001; 414(6866): 924-8.

- Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human molecular genetics*. 2003; 12(14): 1725-35.
- Li X, Foley EA, Molloy KR, Li Y, Chait BT, Kapoor TM. Quantitative chemical proteomics approach to identify post-translational modification-mediated protein-protein interactions. *Journal of the American Chemical Society*. 2012; 134(4): 1982-5.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009; 326(5950): 289-93.
- Luu K, Greenhill CJ, Majoros A, Decker T, Jenkins BJ, Mansell A. STAT1 plays a role in TLR signal transduction and inflammatory responses. *Immunology and cell biology*. 2014; 92(9): 761-9.
- Mahajan MC, Narlikar GJ, Boyapaty G, Kingston RE, Weissman SM. Heterogeneous nuclear ribonucleoprotein C1/C2, MeCP1, and SWI/SNF form a chromatin remodeling complex at the beta-globin locus control region. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102(42): 15012-7.
- Mann M, Kelleher NL. Precision proteomics: the case for high resolution and high mass accuracy. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(47): 18132-8.
- Margueron R, Reinberg D. Chromatin structure and the inheritance of epigenetic information. *Nature reviews Genetics*. 2010; 11(4): 285-96.
- Marshall AG, Hendrickson CL. High-resolution mass spectrometers. *Annu Rev Anal Chem (Palo Alto Calif)*. 2008; 1: 579-99.
- Marushige K. Activation of chromatin by acetylation of histone side chains. *Proceedings of the National Academy of Sciences of the United States of America*. 1976; 73(11): 3937-41.
- Matthiesen R, Trelle MB, Hojrup P, Bunkenborg J, Jensen ON. VEMS 3.0: algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins. *Journal of proteome research*. 2005; 4(6): 2338-47.
- McCormack R, de Armas L, Shiratsuchi M, Podack ER. Killing machines: three pore-forming proteins of the immune system. *Immunologic research*. 2013; 57(1-3): 268-78.
- McNulty DE, Annan RS. Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Molecular & cellular proteomics : MCP*. 2008; 7(5): 971-80.
- Michalski A, Damoc E, Hauschild JP, Lange O, Wiegand A, Makarov A, et al. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Molecular & cellular proteomics : MCP*. 2011; 10(9): M111 011015.
- Min KJ, Cho KH, Kwon TK. The effect of oxidized low density lipoprotein (oxLDL)-induced heme oxygenase-1 on LPS-induced inflammation in RAW 264.7 macrophage cells. *Cell Signal*. 2012; 24(6): 1215-21.
- Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, et al. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular cell*. 2008; 30(6): 755-66.



- Nagaraj N, Kulak NA, Cox J, Neuhauser N, Mayr K, Hoerning O, et al. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Molecular & cellular proteomics : MCP*. 2012; 11(3): M111 013722.
- Nielsen ML, Vermeulen M, Bonaldi T, Cox J, Moroder L, Mann M. Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nature methods*. 2008; 5(6): 459-60.
- Nightingale KP, Gendreizig S, White DA, Bradbury C, Hollfelder F, Turner BM. Cross-talk between histone modifications in response to histone deacetylase inhibitors: MLL4 links histone H3 acetylation and histone H3K4 methylation. *The Journal of biological chemistry*. 2007; 282(7): 4408-16.
- Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods*. 2007; 4(9): 709-12.
- Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics : MCP*. 2002; 1(5): 376-86.
- Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, et al. Latent enhancers activated by stimulation in differentiated cells. *Cell*. 2013; 152(1-2): 157-71.
- Peters AH, Kubicek S, Mechtler K, O'Sullivan RJ, Derijck AA, Perez-Burgos L, et al. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Molecular cell*. 2003; 12(6): 1577-89.
- Pinkse MW, Uitto PM, Hilhorst MJ, Ooms B, Heck AJ. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Analytical chemistry*. 2004; 76(14): 3935-43.
- Poirier GG, de Murcia G, Jongstra-Bilen J, Niedergang C, Mandel P. Poly(ADP-ribosylation) of polynucleosomes causes relaxation of chromatin structure. *Proceedings of the National Academy of Sciences of the United States of America*. 1982; 79(11): 3423-7.
- Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. 2011; 470(7333): 279-83.
- Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols*. 2007; 2(8): 1896-906.
- Rappsilber J, Mann M. What does it mean to identify a protein in proteomics? *Trends in biochemical sciences*. 2002; 27(2): 74-8.
- Raschke WC, Baird S, Ralph P, Nakoinz I. Functional macrophage cell lines transformed by Abelson leukemia virus. *Cell* 1978; 15(1):261-7.
- Razin SV, Gavrillov AA, Pichugin A, Lipinski M, Iarovaia OV, Vassetzky YS. Transcription factories in the context of the nuclear and genome organization. *Nucleic acids research*. 2011; 39(21): 9085-92.
- Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic acids research*. 2007; 35(Web Server issue): W193-200.
- Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics: MCP*. 2004; 3(12): 1154-69.

- Ruepp A, Waegelé B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, et al. CORUM: the comprehensive resource of mammalian protein complexes--2009. *Nucleic acids research*. 2010; 38(Database issue): D497-501.
- Sandhu C, Hewel JA, Badis G, Talukder S, Liu J, Hughes TR, et al. Evaluation of data-dependent versus targeted shotgun proteomic approaches for monitoring transcription factor expression in breast cancer. *Journal of proteome research*. 2008; 7(4): 1529-41.
- Sawado T, Halow J, Im H, Ragoczy T, Bresnick EH, Bender MA, et al. H3 K79 dimethylation marks developmental activation of the beta-globin gene but is reduced upon LCR-mediated high-level transcription. *Blood*. 2008; 112(2): 406-14.
- Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature genetics*. 2010; 42(1): 53-61.
- Schwanhauser B, Gossen M, Dittmar G, Selbach M. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics*. 2009; 9(1): 205-9.
- Serandour AA, Avner S, Percevault F, Demay F, Bizot M, Lucchetti-Miganeh C, et al. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome research*. 2011; 21(4): 555-65.
- Shen Y, Zhang R, Moore RJ, Kim J, Metz TO, Hixson KK, et al. Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000-1500 and capabilities in proteomics and metabolomics. *Analytical chemistry*. 2005; 77(10): 3090-100.
- Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature protocols*. 2006; 1(6): 2856-60.
- Shih MF, Cheng YD, Shen CR, Cherng JY. A molecular pharmacology study into the anti-inflammatory actions of *Euphorbia hirta* L. on the LPS-induced RAW 264.7 cells through selective iNOS protein inhibition. *J Nat Med*. 2010; 64(3):330-5.
- Shukla AK, Futrell JH. Tandem mass spectrometry: dissociation of ions by collisional activation. *Journal of mass spectrometry : JMS*. 2000; 35(9): 1069-90.
- Sif S, Saurin AJ, Imbalzano AN, Kingston RE. Purification and characterization of mSin3A-containing Brg1 and hBrdm chromatin remodeling complexes. *Genes & development*. 2001; 15(5): 603-18.
- Skowronska-Krawczyk D, Ma Q, Schwartz M, Scully K, Li W, Liu Z, et al. Required enhancer-matrin-3 network interactions for a homeodomain transcription program. *Nature*. 2014; 514(7521): 257-61.
- Smith CM, Haimberger ZW, Johnson CO, Wolf AJ, Gafken PR, Zhang Z, et al. Heritable chromatin structure: mapping "memory" in histones H3 and H4. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99 Suppl 4: 16454-61.
- Smith JB, Herschman HR. The glucocorticoid attenuated response genes GARG-16, GARG-39, and GARG-49/IRG2 encode inducible proteins containing multiple tetratricopeptide repeat domains. *Archives of biochemistry and biophysics*. 1996; 330(2): 290-300.
- Snyder M, He W, Zhang JJ. The DNA replication factor MCM5 is essential for Stat1-mediated transcriptional activation. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102(41): 14539-44.

- Soldi M, Bonaldi T. The proteomic investigation of chromatin functional domains reveals novel synergisms among distinct heterochromatin components. *Molecular & cellular proteomics : MCP*. 2013; 12(3): 764-80.
- Soldi M, Cuomo A, Bonaldi T. Improved bottom-up strategy to efficiently separate hypermodified histone peptides through ultra-HPLC separation on a bench top Orbitrap instrument. *Proteomics*. 2014; 14(19): 2212-25.
- Song EJ, Lee KJ. [Identification of proteome molecules by proteomics using two-dimensional gel electrophoresis and MALDI-TOF MS]. *Experimental & molecular medicine*. 2001; 33(1 Suppl): 5-18.
- Spilsbury K, O'Mara MA, Wu WM, Rowe PB, Symonds G, Takayama Y. Isolation of a novel macrophage-specific gene by differential cDNA analysis. *Blood*. 1995; 85(6): 1620-9.
- Swaney DL, McAlister GC, Coon JJ. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nature methods*. 2008; 5(11): 959-64.
- Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(26): 9528-33.
- Tan M, Luo H, Lee S, Jin F, Yang JS, Montellier E, et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*. 2011; 146(6): 1016-28.
- Trelle MB, Salcedo-Amaya AM, Cohen AM, Stunnenberg HG, Jensen ON. Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite *Plasmodium falciparum*. *Journal of proteome research*. 2009; 8(7): 3439-50.
- Turner BM. Defining an epigenetic code. *Nature cell biology*. 2007; 9(1): 2-6.
- Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell*. 2010; 142(6): 967-80.
- Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, et al. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*. 2007; 131(1): 58-69.
- Waddington CH. The epigenotype. 1942. *International journal of epidemiology*. 1942; 41(1): 10-3.
- Wang CI, Alekseyenko AA, LeRoy G, Elia AE, Gorchakov AA, Britton LM, et al. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in *Drosophila*. *Nature structural & molecular biology*. 2013; 20(2): 202-9.
- Willmann KL, Milosevic S, Pauklin S, Schmitz KM, Rangam G, Simon MT, et al. A role for the RNA pol II-associated PAF complex in AID-induced immune diversification. *The Journal of experimental medicine*. 2012; 209(11): 2099-1111.
- Wu C and Morris J R. Genes, genetics and epigenetics: a correspondence. *Science*. 2001; 293(5532): 1103-1105.
- Wu TT, Chen TL, Chen RM. Lipopolysaccharide triggers macrophage activation of inflammatory cytokine expression, chemotaxis, phagocytosis, and oxidative ability via a toll-like receptor 4-dependent pathway: validated by RNA interference. *Toxicology letters*. 2009; 191(2-3): 195-202.

- Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, et al. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*. 2005; 121(6): 859-72.
- Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, et al. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*. 2006; 442(7098): 86-90.
- Yang J, Aittomaki S, Pesu M, Carter K, Saarinen J, Kalkkinen N, et al. Identification of p100 as a coactivator for STAT6 that bridges STAT6 with RNA polymerase II. *The EMBO journal*. 2002; 21(18): 4950-8.
- Zakrzewska A, Cui C, Stockhammer OW, Benard EL, Spaink HP, Meijer AH. Macrophage-specific gene functions in Spi1-directed innate immunity. *Blood*. 2010; 116(3): e1-11.
- Zentner GE, Tesar PJ, Scacheri PC. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome research*. 2011; 21(8): 1273-83.
- Zhang DE, Hetherington CJ, Meyers S, Rhoades KL, Larson CJ, Chen HM, et al. CCAAT enhancer-binding protein (C/EBP) and AML1 (CBF alpha2) synergistically activate the macrophage colony-stimulating factor receptor promoter. *Molecular and cellular biology*. 1996; 16(3): 1231-40.
- Zhang K, Kagan D, DuBois W, Robinson R, Bliskovsky V, Vass WC, et al. Mndal, a new interferon-inducible family member, is highly polymorphic, suppresses cell growth, and may modify plasmacytoma susceptibility. *Blood*. 2009; 114(14): 2952-60.
- Zhang Z, Carmichael GG. The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell*. 2001; 106(4): 465-75.
- Zhou Z, Licklider LJ, Gygi SP, Reed R. Comprehensive proteomic analysis of the human spliceosome. *Nature*. 2002; 419(6903): 182-5.
- Zhu Y, Sun L, Chen Z, Whitaker JW, Wang T, Wang W. Predicting enhancer transcription and activity from chromatin modifications. *Nucleic acids research*. 2013; 41(22): 10032-43.

**APPENDIX 1. The enhancers's basal chromatome: proteins identified upon the intersection of H3K4me1 and Pu.1 basal interactomes.** Protein determinants of enhancers where identified based on their SILAC ratio and ratio counts (RC) in four X-ChroP experimental replicas (H3K4me1 Forward, H3K4me1 Reverse, Pu.1 Forward, Pu.1 Reverse). (Fwd = Forward, Rev = Reverse). Fasta identifiers, gene names and protein names are given.

Fasta headers	Gene names	Protein names	H3K4me1_Fwd		H3K4me1_Rev		Pu.1_Fwd		Pu.1_Rev	
			Ratio H/L	H/L RC	Ratio L/H	H/L RC	Ratio H/L	H/L RC	Ratio L/H	H/L RC
P50247	Ahcy	Adenosylhomocysteinase	1.65	58	2.02	32	13.46	23	32.21	24
E9Q3L4	AI607873	Interferon-Inducible Protein X	2.27	8	2.78	7	12.28	4	18.23	4
Q9DBR0	Akap8	A-kinase anchor protein 8	2.39	21	3.58	15	11.28	3	20.59	2
O35381	Anp32a	Acidic leucine-rich nuclear phosphoprotein 32 family member A	1.98	31	2.84	25	14.16	18	47.36	19
Q9EST5	Anp32b	Acidic leucine-rich nuclear phosphoprotein 32 family member B	2.31	21	3.06	18	9.10	12	49.08	7
P28352	Apex1	DNA-(apurinic or apyrimidinic site) lyase;DNA-(apurinic or apyrimidinic site) lyase, mitochondrial	4.82	22	8.79	9	10.60	10	28.57	4
O35841	Api5	Apoptosis inhibitor 5	1.86	21	2.71	21	7.27	7	37.24	7
E9QAE3	Btaf1	RNA polymerase II, B-TFIID transcription factor-associated	1.92	10	2.01	12	12.12	13	14.14	12
Q8CH18	Ccar1	Cell division cycle and apoptosis regulator protein 1	2.57	20	3.95	11	10.13	5	21.71	5
Q8VDP4	Ccar2	DBIRD complex subunit KIAA1967 homolog	1.55	37	2.43	24	12.23	10	26.79	8
P24788	Cdk11b	Cyclin-dependent kinase 11B	2.00	18	2.74	16	10.24	2	18.74	7
Q9CXS4	Cenpv	Centromere protein V	5.33	6	7.06	8	12.53	8	38.24	2
Q8BHG9	Cggbp1	CGG triplet repeat-binding protein 1	2.55	4	3.31	5	11.09	3	29.90	2
E9QAS5	Chd4	Chromodomain-helicase-DNA-binding protein 4	2.39	80	2.99	61	13.50	40	33.32	38
G5E8I8	Cherp	Calcium homeostasis endoplasmic reticulum protein	1.61	8	2.41	4	9.47	5	19.96	5
O88712	Ctbp1	C-terminal-binding protein 1	2.15	4	3.08	3	14.06	6	38.23	5
P56546	Ctbp2	C-terminal-binding protein 2	1.83	17	2.65	14	14.07	9	28.55	11
Q9CWL8	Ctnnb1	Beta-catenin-like protein 1	2.05	16	2.41	9	12.80	2	40.31	4
Q3U741	Ddx17	Probable ATP-dependent RNA helicase DDX17	2.21	41	3.49	33	12.98	9	19.11	16
Q9JIK5	Ddx21	Nucleolar RNA helicase 2	2.89	91	4.15	76	9.82	21	17.37	27
Q8BTS0	Ddx5	Probable ATP-dependent RNA helicase DDX5	2.04	158	3.53	139	10.36	70	30.50	70
E9QNN1	Dhx9	ATP-dependent RNA helicase A	2.14	163	3.42	127	12.25	54	35.10	62

Q9CSH3	Dis3	Exosome complex exonuclease RRP44	1.95	29	2.39	23	8.34	12	16.10	13
Q6NZB0	Dnajc8	DnaJ homolog subfamily C member 8	2.71	7	3.81	4	9.68	6	42.77	3
P13864	Dnmt1	DNA (cytosine-5)-methyltransferase 1	2.70	46	3.38	38	12.30	34	20.35	32
P70372	Elavl1	ELAV-like protein 1	1.81	49	2.63	37	13.04	20	15.05	15
B2RWS6	Ep300	Histone acetyltransferase p300	1.98	6	3.33	4	11.00	8	17.47	6
P35550	Fbl	rRNA 2-O-methyltransferase fibrillarin	2.14	26	2.88	15	10.05	4	16.35	3
Q91Z50	Fen1	Flap endonuclease 1	2.50	33	3.60	15	12.60	15	35.92	10
P56959	Fus	RNA-binding protein FUS	2.08	29	3.60	22	7.45	8	15.27	8
Q8CHY6	Gatad2a	Transcriptional repressor p66 alpha	2.34	13	2.94	10	12.99	6	14.48	7
Q8VHR5	Gatad2b	Transcriptional repressor p66-beta	2.13	13	2.74	9	8.22	2	18.23	9
P62915	Gtf2b	Transcription initiation factor IIB	2.03	6	2.77	2	12.64	8	16.27	4
Q9D902	Gtf2e2	General transcription factor IIE subunit 2	2.42	4	2.88	5	11.75	3	26.49	2
Q9QZQ8	H2afy	Core histone macro-H2A.1	9.45	44	19.37	31	9.87	6	25.61	4
P0C0S6	h2afz	Histone H2A.Z;Histone H2A.V;Histone H2A	10.80	19	24.20	12	9.06	6	19.55	5
B1AUX2	Hcfc1	Host cell factor 1	2.09	26	2.66	20	12.95	17	22.46	22
P51859	Hdgf	Hepatoma-derived growth factor	10.73	21	19.71	19	11.40	3	16.03	8
P43275	Hist1h1a	Histone H1.1	8.68	22	18.29	19	9.12	5	19.40	6
P43276	Hist1h1b	Histone H1.5	9.43	62	18.57	45	8.94	14	18.57	13
P63158	Hmgb1	High mobility group protein B1	3.84	45	5.69	33	10.66	14	41.08	17
P30681	Hmgb2	High mobility group protein B2	3.77	57	6.92	50	11.70	21	24.17	19
Q9CX86	Hnrnpa0	Heterogeneous nuclear ribonucleoprotein A0	2.10	64	3.38	47	13.56	19	34.86	23
Q5EBP8	Hnrnpa1	Heterogeneous nuclear ribonucleoprotein A1	2.10	85	3.56	74	11.92	36	32.47	29
O88569	Hnrnpa2b1	Heterogeneous nuclear ribonucleoproteins A2/B1	2.13	181	3.31	162	12.18	87	24.70	70
Q8BG05	Hnrnpa3	Heterogeneous nuclear ribonucleoprotein A3	1.84	130	3.08	105	12.16	42	27.44	34
Q20BD0	Hnrnpab	Heterogeneous nuclear ribonucleoprotein A/B	2.42	47	3.39	48	12.82	24	35.82	27
Q9Z204	Hnrnpc	Heterogeneous nuclear ribonucleoproteins C1/C2	2.43	73	3.61	66	12.25	25	34.41	24
Q60668	Hnrnpd	Heterogeneous nuclear ribonucleoprotein D0	2.49	37	3.68	29	11.97	17	32.65	18
Q9Z2X1	Hnrnpf	Heterogeneous nuclear ribonucleoprotein F	2.03	66	3.21	46	12.66	30	27.57	21

Q8C2Q7	Hnrnp1	Heterogeneous nuclear ribonucleoprotein H	1.91	67	3.22	64	11.11	33	23.78	11
Q8R081	Hnrnp1	Heterogeneous nuclear ribonucleoprotein L	1.85	106	2.99	73	9.92	26	32.55	30
Q9D0E1	Hnrnpm	Heterogeneous nuclear ribonucleoprotein M	2.49	340	4.01	250	11.34	96	33.30	94
Q8VHM5	HnrnpR	Heterogeneous nuclear ribonucleoprotein R	2.01	68	3.05	63	11.10	12	24.87	16
Q8VEK3	HnrnpU	Heterogeneous nuclear ribonucleoprotein U	2.47	178	4.07	165	11.58	68	34.01	75
Q3TEA8	Hp1bp3	Heterochromatin protein 1-binding protein 3	10.71	61	24.69	45	7.81	7	21.14	5
Q5SWT9	Ikzf1	DNA-binding protein Ikaros	2.66	5	3.11	3	11.63	2	19.38	2
Q9CXY6	Ilf2	Interleukin enhancer-binding factor 2	2.32	15	3.44	15	13.66	6	39.30	5
Q60749	Khdrbs1	KH domain-containing, RNA-binding, signal transduction-associated protein 1	2.45	33	4.60	25	12.76	7	40.51	9
Q3U0V1	Khsrp	Far upstream element-binding protein 2	2.12	86	3.70	57	9.87	23	26.51	31
P48678	Lmna	Prelamin-A/C;Lamin-A/C	3.80	349	5.40	328	8.66	137	23.29	140
Q9CYI4	Luc71	Putative RNA-binding protein Luc7-like 1	2.05	7	2.83	4	7.51	2	14.64	1
Q7TNC4	Luc7l2	Putative RNA-binding protein Luc7-like 2	2.06	24	2.52	25	11.36	10	43.05	10
Q8K310	Matr3	Matrin-3	2.84	75	4.85	51	10.94	21	22.44	17
P97310	Mcm2	DNA replication licensing factor MCM2	2.12	58	2.74	44	11.72	27	15.92	31
P25206	Mcm3	DNA replication licensing factor MCM3	2.25	71	2.89	57	10.95	32	23.14	34
P49717	Mcm4	DNA replication licensing factor MCM4	2.25	52	2.86	42	11.57	23	15.68	27
Q52KC3	Mcm5	DNA replication licensing factor MCM5	2.19	60	2.94	48	10.42	21	34.83	18
P97311	Mcm6	DNA replication licensing factor MCM6	2.20	85	2.71	67	10.64	33	25.95	35
Q61881	Mcm7	DNA replication licensing factor MCM7	2.16	41	2.78	39	11.17	19	32.54	22
E9PVX6	Mki67	Marker of proliferation Ki-67	3.51	63	5.41	43	8.39	18	15.27	12
Q9JK91	Mlh1	DNA mismatch repair protein Mlh1	1.77	5	2.52	5	13.09	2	37.97	1
E9QN37	Mpeg1	Macrophage-expressed gene 1 protein	0.65	18	0.97	8	4.56	9	11.08	6
P54276	Msh6	DNA mismatch repair protein Msh6	1.94	25	2.68	19	8.83	17	14.65	14
Q9R190	Mta2	Metastasis-associated protein MTA2	2.20	45	2.75	37	13.09	18	19.90	31
Q7TPV4	Mybbp1a	Myb-binding protein 1A	2.67	70	3.11	42	7.65	11	15.60	18
Q3UYV9	Ncbp1	Nuclear cap-binding protein subunit 1	2.00	16	2.51	12	13.53	6	28.91	7
P09405	Ncl	Nucleolin	2.48	186	3.22	174	9.37	72	21.64	73



Q9D0T1	Nhp211	NHP2-like protein 1	1.90	7	2.24	7	8.92	3	19.12	3
Q99K48	Nono	Non-POU domain-containing octamer-binding protein	3.07	104	5.08	80	8.60	22	41.63	22
Q9D6Z1	Nop56	Nucleolar protein 56	2.36	26	2.77	13	7.89	1	14.12	5
Q9D6T0	Nosip	Nitric oxide synthase-interacting protein	1.94	3	2.49	5	8.83	5	43.58	3
Q61937	Npm1	Nucleophosmin	2.34	65	2.78	53	9.50	26	20.74	26
Q9CPP0	Npm3	Nucleoplasmin-3	2.08	2	2.47	2	10.93	3	20.92	2
Q1HFZ0	Nsun2	tRNA (cytosine(34)-C(5))-methyltransferase	2.77	44	3.79	35	11.67	9	24.96	14
Q9CQF3	Nudt21	Cleavage and polyadenylation specificity factor subunit 5	2.58	8	4.01	11	18.10	4	15.46	3
E9Q7G0	Numa1	Nuclear mitotic apparatus protein 1	1.94	96	2.14	76	10.46	58	15.71	53
Q9DCE5	Pak1ip1	p21-activated protein kinase-interacting protein 1	2.73	9	3.97	2	13.52	4	24.98	2
Q921K2	Parp1	Poly [ADP-ribose] polymerase 1	2.82	58	3.83	47	10.35	27	17.85	39
F8VQD1	Pbrm1	Protein polybromo-1	3.82	9	5.35	9	13.71	19	17.61	18
P17918	Pcna	Proliferating cell nuclear antigen	2.29	28	2.76	25	12.58	15	33.38	14
E9QPI5	Pds5a	Sister chromatid cohesion protein PDS5 homolog A	3.17	50	4.30	35	10.03	27	19.67	22
P52431	Pold1	DNA polymerase delta catalytic subunit;DNA polymerase	1.77	15	2.36	14	11.93	9	14.80	12
P08775	Polr2a	DNA-directed RNA polymerase II subunit RPB1	3.22	48	4.69	36	11.08	20	21.18	16
Q8CCF0	Prpf31	U4/U6 small nuclear ribonucleoprotein Prp31	1.72	14	2.29	8	13.54	5	16.61	9
Q99JF8	Psip1	PC4 and SFRS1-interacting protein	6.79	32	13.83	30	10.95	4	29.84	3
Q922I7	Ptbp1	Polypyrimidine tract-binding protein 1	1.67	70	2.64	71	12.31	43	32.08	22
Q64012	Raly	RNA-binding protein Raly	2.40	52	3.48	45	12.72	23	35.78	23
P62827	Ran	GTP-binding nuclear protein Ran	1.61	38	2.06	36	8.66	27	15.60	15
Q60972	Rbbp4	Histone-binding protein RBBP4	2.00	14	2.70	17	10.27	22	14.32	12
Q0VBL3	Rbm15	RNA binding motif protein 15	2.32	25	3.42	16	9.42	2	22.41	7
Q5SFM8	Rbm27	RNA-binding protein 27	1.65	7	2.15	7	8.83	7	17.86	7
Q9CWZ3	Rbm8a	RNA-binding protein 8A	1.58	2	2.36	4	11.06	3	17.73	2
Q91VM5	Rbmx11	RNA binding motif protein, X-linked-like-1	2.48	77	4.28	51	12.47	14	36.31	19

P31266	Rbpj	Recombining binding protein suppressor of hairless	2.81	20	3.83	23	8.70	4	23.91	9
Q6PFB2	Rcc1	Regulator of chromosome condensation	6.33	11	8.42	10	10.41	3	18.27	4
Q8BK67	Rcc2	Protein RCC2	4.66	68	6.62	73	8.57	22	28.55	15
Q9R1T2	Sae1	SUMO-activating enzyme subunit 1	2.44	15	3.39	8	12.00	5	41.46	2
D3YXK2	Safb	Scaffold attachment factor B1	2.95	14	3.51	9	12.71	2	37.08	3
Q9JLI8	Sart3	Squamous cell carcinoma antigen recognized by T-cells 3	1.51	19	2.00	19	10.96	6	14.30	9
Q9EQU5	Set	Protein SET	4.41	12	8.02	15	7.71	6	14.69	8
E9Q4Q2	Sf1	Splicing factor 1	2.42	22	3.73	17	21.45	2	15.66	7
Q8VIJ6	Sfpq	Splicing factor, proline- and glutamine-rich	3.42	134	5.41	105	11.68	36	23.47	48
Q60520	Sin3a	Paired amphipathic helix protein Sin3a	2.07	11	2.87	10	10.70	10	24.75	7
Q9CZU3	Skiv2l2	Superkiller viralicidic activity 2-like 2	1.63	27	2.06	26	11.17	13	15.08	14
Q3TKT4	Smarca4	Transcription activator BRG1	2.98	35	3.96	24	14.02	54	27.59	49
P97496	Smarcc1	SWI/SNF complex subunit SMARCC1	3.07	25	4.03	25	15.02	36	14.80	43
Q6PDG5	Smarcc2	SWI/SNF complex subunit SMARCC2	3.44	16	4.88	8	15.00	15	17.51	18
Q99JR8	Smarcd2	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 2	2.94	8	4.27	9	8.99	19	28.52	17
Q9CU62	Smc1a	Structural maintenance of chromosomes protein 1A	3.69	64	4.87	50	13.57	39	22.72	45
Q9CW03	Smc3	Structural maintenance of chromosomes protein 3	3.50	81	4.48	62	13.83	41	14.88	45
Q6P5D8	Smchd1	Structural maintenance of chromosomes flexible hinge domain-containing protein 1	3.75	36	5.25	24	11.95	29	21.48	28
Q62376	Snrnp70	U1 small nuclear ribonucleoprotein 70 kDa	2.08	30	2.98	28	12.25	2	26.56	8
Q62189	Snrpa	U1 small nuclear ribonucleoprotein A	1.87	28	2.99	23	12.49	14	38.94	14
P57784	Snrpa1	U2 small nuclear ribonucleoprotein A	1.53	13	2.07	16	10.85	6	37.08	7
Q9CQI7	Snrpb2	U2 small nuclear ribonucleoprotein B	1.58	6	2.05	4	11.44	1	27.37	2
P17433	Spi1	Transcription factor PU.1	6.49	3	10.06	2	12.02	36	39.73	27
Q99MR6	Srrt	Serrate RNA effector molecule homolog	1.95	52	2.73	41	16.90	32	25.71	36
H7BX95	Srsf1	Serine/arginine-rich splicing factor 1	1.86	35	2.80	30	12.33	13	29.20	14
Q62093	Srsf2	Serine/arginine-rich splicing factor 2	1.87	28	2.92	21	11.07	8	37.87	10

Appendix I

Q9D8S5	Srsf5	Serine/arginine-rich splicing factor 5	1.97	8	2.73	7	7.87	1	25.30	3
Q3TWW8	Srsf6	Serine/arginine-rich splicing factor 6	1.67	35	2.38	38	12.51	13	29.02	11
Q8BL97	Srsf7	Serine/arginine-rich splicing factor 7	1.89	31	3.07	20	10.17	13	23.86	18
P32067	Ssb	Lupus La protein homolog	2.30	46	3.46	39	12.79	23	27.28	15
Q08943	Ssrp1	FACT complex subunit SSRP1	3.11	25	4.10	22	10.27	11	36.82	12
P61957	Sumo2	Small ubiquitin-related modifier 2	2.92	16	3.73	11	8.18	9	40.94	8
O55201	Supt5h	Transcription elongation factor SPT5	3.06	34	3.97	30	10.46	13	15.45	13
Q93092	Taldo1	Transaldolase	3.28	31	5.15	21	9.77	11	19.93	14
Q921F2	Tardbp	TAR DNA-binding protein 43	1.87	44	2.66	37	12.41	17	43.66	18
Q8CGF7	Tcerg1	Transcription elongation regulator 1	1.99	21	3.09	21	11.97	11	25.30	10
Q99J36	Thumpd1	THUMP domain-containing protein 1	2.96	14	3.98	4	13.80	4	21.86	4
P40142	Tkt	Transketolase	1.74	91	2.52	89	10.20	29	26.95	41
Q61033	Tmpo	Lamina-associated polypeptide 2, isoforms alpha/zeta	2.65	26	4.46	26	10.67	8	38.35	6
Q62318	Trim28	Transcription intermediary factor 1-beta	1.88	90	2.42	75	13.26	46	37.77	61
Q9Z1F9	Uba2	SUMO-activating enzyme subunit 2	2.46	19	3.54	13	13.69	4	29.56	5
P62983	Ubc	Ubiquitin-40S ribosomal protein S27a	3.24	95	5.82	76	8.64	44	23.63	35
E9PXY8	Usp7	Ubiquitin carboxyl-terminal hydrolase;Ubiquitin carboxyl-terminal hydrolase 7	1.53	16	2.08	16	11.75	7	17.60	11
Q9R0G7	Zeb2	Zinc finger E-box-binding homeobox 2	2.51	7	3.17	4	13.82	9	16.26	8



**APPENDIX 2. Dynamic profiling of enhancers' determinants by *tc-ChroP*.** SILAC ratios and ratio counts for the 148 proteins previously identified as specific enhancers' determinants upon four time-course intriple-SILAC experiments (H3K4me1 Forward, H3K4me1 Reverse, Pu.1 Forward, Pu.1 Reverse). (1hr/UT correspond to M/L and H/L in the Forward and Reverse experiment, respectively. 4hr/UT correspond to H/L and M/L in the Forward and Reverse experiment, respectively). (Fwd = Forward, Rev = Reverse). (Red arrow = recruited, green arrow = evicted). Gene names are provided.

Gene names	H3K4me1 Fwd				H3K4me1 Rev				Pu.1 Fwd				Pu.1 Rev				Dynamics
	1h/UT	1h/UT RC	4h/UT	4h/UT RC	1h/UT	1h/UT RC	4h/UT	4h/UT RC	1h/UT	1h/UT RC	4h/UT	4h/UT RC	1h/UT	1h/UT RC	4h/UT	4h/UT RC	
Ahcy	1.39	14	1.65	14	0.73	27	0.86	27	0.62	16	7.00	16	0.82	12	1.00	12	
Akap8	1.06	1	2.30	1	0.81	8	0.94	8	2.00	2	2.15	2	1.13	4	1.06	4	
Anp32a	1.26	24	1.44	24	0.68	14	0.78	14	1.57	17	1.97	17	0.82	20	0.83	20	
Anp32b	1.27	5	1.43	5	0.64	12	0.81	12	1.63	8	2.05	8	0.76	10	0.86	10	
Apex1	0.88	19	0.90	19	0.83	21	0.80	21	1.53	10	1.69	10	0.81	6	0.78	6	
Api5	1.28	11	1.47	11	0.74	16	0.85	16	1.90	5	2.35	5	0.96	7	1.02	7	
Btaf1	1.32	2	1.56	2	0.60	3	0.79	3	1.87	2	2.80	2	0.84	7	1.02	7	
Ccar1	1.20	2	1.29	2	0.67	11	0.86	11	1.93	5	1.98	5	0.88	7	1.00	7	
Ccar2	1.39	4	1.85	4	0.83	26	0.90	26	2.14	4	2.64	4	1.03	18	1.01	18	
Cdk11b	1.23	2	1.35	2	0.84	13	0.90	13	NaN	0	NaN	0	1.10	6	0.92	6	
Cenpv	1.04	6	1.21	6	0.87	13	0.84	13	1.27	11	1.60	11	0.92	8	0.90	8	
Cggbp1	1.10	2	1.15	2	0.72	5	0.88	5	1.44	3	1.85	3	0.80	3	0.80	3	
Chd4	1.20	15	1.41	15	0.77	34	0.86	34	1.56	39	2.11	39	0.92	49	0.96	49	
Cherp	NaN	0	NaN	0	0.71	1	0.97	1	1.88	1	2.72	1	NaN	0	NaN	0	
Ctbp1	1.15	2	1.35	2	0.76	6	0.88	6	0.99	8	3.04	8	0.99	13	0.90	13	
Ctbp2	1.09	1	1.33	1	0.90	4	0.86	4	0.97	4	2.31	4	0.98	8	0.89	8	
Ctnnbl1	1.07	5	1.16	5	0.73	2	0.82	2	1.83	2	1.85	2	0.93	3	0.89	3	
Ddx17	1.16	18	1.53	18	0.79	47	0.87	47	1.73	32	2.12	32	1.07	25	0.96	25	
Ddx21	1.07	58	1.10	58	0.73	84	0.71	84	1.50	60	1.67	60	0.98	76	0.69	76	
Ddx5	1.11	58	1.32	58	0.75	178	0.95	178	1.71	103	2.02	103	1.00	107	1.02	107	
Dhx9	1.13	37	1.32	37	0.79	125	0.91	125	1.85	67	2.17	67	1.04	72	0.98	72	
Dis3	1.50	1	1.48	1	0.77	14	0.87	14	1.79	13	2.21	13	0.95	25	0.97	25	
Dnajc8	1.09	5	1.27	5	0.74	5	0.83	5	1.68	5	2.03	5	0.94	4	0.96	4	
Dnmt1	0.89	12	1.03	12	0.64	49	0.65	49	1.28	16	1.59	16	0.77	40	0.71	40	↓
Elavl1	1.02	7	1.11	7	0.76	26	0.82	26	1.53	20	1.68	20	1.00	16	0.95	16	
Ep300	NaN	0	NaN	0	1.25	2	0.76	2	1.41	5	1.49	5	0.96	9	0.92	9	

Appendix 2







Fbl	1.15	35	1.28	35	0.84	29	0.85	29	1.39	23	1.60	23	1.12	16	0.84	16	
Fen1	1.07	6	1.18	6	0.75	13	0.81	13	1.69	12	2.16	12	0.82	15	0.82	15	
Fus	1.16	3	1.40	3	0.79	19	0.98	19	1.85	3	2.40	3	0.99	14	1.17	14	↑
Gatad2a	1.11	1	1.78	1	0.79	9	0.89	9	NaN	0	NaN	0	0.92	20	0.99	20	
Gatad2b	NaN	0	NaN	0	0.65	8	0.72	8	3.57	2	2.77	2	0.86	11	0.85	11	
Gtf2b	1.00	1	1.47	1	0.62	7	0.84	7	NaN	0	NaN	0	0.83	10	0.94	10	↓
Gtf2e2	NaN	0	NaN	0	0.72	3	0.83	3	NaN	0	NaN	0	0.90	4	0.85	4	
H2afy	0.99	37	1.03	37	0.82	58	0.82	58	1.47	9	2.13	9	1.06	16	1.02	16	
h2afz	0.97	18	0.92	18	0.86	15	0.88	15	1.52	11	1.75	11	1.10	4	0.99	4	
Hcfc1	NaN	0	NaN	0	0.62	8	0.86	8	1.99	1	3.77	1	0.81	17	1.10	17	
Hdgf	0.83	24	0.88	24	0.85	26	0.86	26	1.50	10	2.27	8	0.94	6	1.23	6	
Hist1h1a	0.87	34	0.99	34	0.80	49	0.71	49	1.17	14	1.65	14	1.18	6	0.91	6	
Hist1h1b	0.94	53	0.95	53	0.81	70	0.82	70	1.39	18	1.72	18	1.10	23	1.01	23	
Hmgb1	1.07	30	1.09	30	0.81	26	0.84	26	1.67	32	2.01	32	0.90	17	0.88	17	
Hmgb2	1.16	45	1.06	45	0.84	49	0.91	49	1.72	61	1.97	61	0.90	28	0.92	28	
Hnrnpa0	1.16	22	1.35	22	0.81	36	0.90	36	1.74	42	1.98	42	1.09	26	0.99	26	
Hnrnpa1	1.22	16	1.36	16	0.79	39	0.94	39	1.76	38	1.70	36	1.08	14	1.10	14	
Hnrnpa2b	1.19	64	1.40	64	0.82	140	0.92	140	1.97	105	2.10	105	1.14	82	1.06	82	
Hnrnpa3	NaN	0	NaN	0	0.67	6	0.77	6	2.58	4	2.54	4	0.88	2	0.84	2	↓
Hnrnpab	1.24	14	1.24	14	0.82	46	0.87	46	1.83	26	2.06	26	0.99	34	0.94	34	
Hnrnpc	NaN	0	NaN	0	0.83	2	0.87	2	1.51	2	3.22	2	NaN	0	NaN	0	
Hnrnpd	1.28	15	1.38	15	0.81	46	0.92	46	1.81	36	2.18	35	1.03	26	0.95	26	
Hnrnpf	1.16	13	1.33	13	0.82	49	0.94	49	1.62	22	1.86	22	0.99	27	1.04	27	
Hnrnph1	1.02	8	1.08	8	0.79	44	0.89	44	1.54	24	1.97	24	1.06	21	1.00	21	
Hnrnpl	1.13	20	1.28	20	0.81	84	0.92	84	1.97	25	2.19	25	1.02	44	0.98	44	
Hnrnpm	1.11	68	1.29	68	0.80	241	0.91	241	2.04	138	2.50	138	1.11	130	1.05	130	
Hnrnpr	1.32	6	1.43	6	0.81	40	0.88	40	2.01	18	1.91	18	1.14	15	1.06	15	
Hnrnpu	1.11	98	1.11	98	0.83	189	0.87	188	1.98	115	2.15	115	1.10	93	1.01	93	
Hp1bp3	1.13	102	0.99	102	0.86	87	0.86	87	1.51	12	1.93	12	1.00	14	1.03	14	

Appendix 2

Ikzf1	NaN	0	NaN	0	0.64	8	0.73	8	1.63	6	2.11	6	0.91	22	1.00	22	
Ilf2	1.27	4	1.53	4	0.77	14	0.86	14	1.88	6	2.17	6	1.06	12	0.98	12	
Khdrbs1	1.22	4	1.41	4	0.75	24	0.93	24	1.31	3	1.23	3	1.11	14	1.08	14	
Khsrp	1.25	18	1.52	18	0.76	62	0.90	62	1.78	34	2.16	34	0.98	46	0.92	46	
Lmna	1.16	263	1.27	263	0.86	430	0.86	430	1.49	279	1.89	279	1.15	198	1.01	198	
Luc7l	1.25	1	1.11	1	0.85	4	0.86	4	1.63	2	1.50	2	1.09	2	1.13	2	
Luc7l2	1.03	11	1.04	11	0.88	20	0.94	20	1.70	10	1.99	10	0.99	9	1.22	9	
Matr3	1.06	6	1.37	6	0.77	37	0.92	37	1.87	21	2.21	20	1.10	24	1.10	24	
Mcm2	1.30	6	1.44	6	0.73	24	0.82	24	1.74	11	2.23	11	0.89	21	0.88	21	
Mcm3	1.30	19	1.37	19	0.70	31	0.76	31	1.68	21	1.94	21	0.90	26	0.85	26	
Mcm4	1.38	11	1.24	11	0.80	19	0.87	19	1.77	11	2.15	10	0.92	18	1.00	18	
Mcm5	1.15	17	1.22	17	0.71	47	0.76	47	1.57	21	1.95	21	0.92	37	0.86	37	
Mcm6	1.29	17	1.39	17	0.70	45	0.79	45	1.81	16	2.15	16	0.88	32	0.89	32	
Mcm7	1.26	20	1.28	20	0.74	34	0.81	34	1.63	23	1.98	23	0.91	39	0.92	39	
Mki67	NaN	0	NaN	0	0.66	79	0.76	79	NaN	0	NaN	0	0.67	27	0.77	27	↓
Mlh1	NaN	0	NaN	0	0.66	2	0.85	2	NaN	0	NaN	0	0.80	1	0.95	1	
Mpeg1	NaN	0	NaN	0	0.86	4	1.14	4	1.75	4	2.35	4	0.98	3	1.18	3	
Msh6	1.14	1	1.81	1	0.72	10	0.83	10	1.76	3	2.68	3	0.74	5	0.79	5	
Mta2	1.10	13	1.54	13	0.82	20	0.84	20	1.45	16	2.46	16	0.90	25	0.87	25	
Mybbp1a	1.06	61	1.26	61	0.70	41	0.73	41	1.64	41	2.15	41	1.04	27	0.78	27	
Ncbp1	1.08	4	1.20	4	0.75	9	0.87	9	1.99	4	2.43	4	0.91	6	1.01	6	
Ncl	1.14	115	1.20	115	0.74	174	0.76	174	1.72	145	1.91	145	1.07	122	0.90	122	
Nhp2l1	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	
Nono	1.04	31	1.09	31	0.84	66	1.02	66	1.81	32	2.27	32	1.12	30	1.13	30	↑
Nop56	1.16	48	1.31	48	0.75	31	0.83	31	1.46	33	1.82	33	1.13	15	0.86	15	
Nosip	1.12	1	1.04	1	0.66	7	0.74	7	1.57	3	1.94	3	0.85	7	0.75	7	
Npm1	1.13	59	1.32	59	0.74	44	0.83	44	1.66	51	2.09	51	1.07	32	0.93	32	
Npm3	NaN	0	NaN	0	0.77	3	0.75	3	NaN	0	NaN	0	1.00	4	0.85	4	
Nsun2	1.07	23	1.26	23	0.74	29	0.76	29	1.53	21	2.05	21	0.91	13	0.84	13	



Appendix 2

Nudt21	0,74	5	1,34	5	0,89	10	1,08	10	NaN	0	NaN	0	1,16	3	1,03	3	
Numa1	1,35	17	1,57	17	0,78	49	0,84	49	1,77	35	2,00	35	0,98	71	0,86	71	
Pak1ip1	1,18	3	1,05	3	0,75	5	0,77	5	1,16	1	1,50	1	0,95	3	0,85	3	
Parp1	1,19	26	1,40	26	0,79	45	0,82	45	1,53	45	2,14	44	0,86	34	0,84	34	
Pbrm1	NaN	0	NaN	0	0,92	9	0,91	9	1,94	9	1,87	9	0,98	24	0,93	24	
Pcna	1,12	27	1,18	27	0,70	24	0,71	24	1,25	15	1,56	15	0,74	16	0,77	16	
Pds5a	1,22	6	1,45	6	0,74	19	0,89	19	1,62	12	2,25	12	0,87	30	0,93	30	
Pold1	0,99	1	1,11	1	0,73	6	0,82	6	1,58	2	2,02	2	0,82	13	0,84	13	
Polr2a	1,18	8	1,32	8	0,79	26	0,80	26	1,37	4	1,52	4	0,89	14	0,91	14	
Prpf31	1,28	5	1,35	5	0,73	8	0,84	8	1,64	5	2,04	5	0,96	7	0,90	7	
Psip1	0,93	27	1,08	27	0,85	51	0,86	51	1,53	4	2,03	4	0,90	12	0,97	12	
Ptbp1	1,17	16	1,15	16	0,81	67	0,91	67	1,80	30	1,95	29	0,97	47	0,96	47	
Pyhin1	1,22	1	1,22	1	0,82	8	1,26	8	1,69	3	2,05	3	1,01	15	1,44	15	
Raly	1,17	14	1,28	14	0,78	46	0,88	46	1,97	19	2,13	19	1,05	32	0,99	32	
Ran	1,13	20	1,15	20	0,83	30	0,88	30	1,68	23	1,95	23	0,93	18	0,94	18	
Rbbp4	1,36	3	1,73	3	0,74	21	0,82	21	1,85	3	2,58	3	0,78	15	0,86	15	
Rbm15	0,95	2	1,35	2	1,04	58	1,13	58	1,67	4	2,21	4	1,12	9	1,04	9	
Rbm27	NaN	0	NaN	0	0,83	11	1,10	11	1,34	1	2,18	1	1,11	4	0,98	4	
Rbm8a	0,94	1	0,95	1	0,78	3	0,85	3	1,55	1	1,47	1	1,03	1	1,11	1	
Rbmx11	1,30	11	1,42	11	0,85	46	0,96	46	2,40	27	2,52	27	1,23	23	1,12	23	
Rbpj	1,02	11	1,11	11	0,75	25	0,81	25	1,16	4	1,51	4	0,86	12	0,86	12	
Rcc1	1,07	11	1,03	11	0,87	18	0,89	18	1,51	5	1,91	5	1,09	5	0,96	5	
Rcc2	1,00	58	0,95	58	0,83	76	0,89	77	1,63	24	1,84	24	0,98	30	0,96	30	
Rnmt	0,96	1	1,53	1	0,86	5	0,95	5	1,47	1	1,78	1	0,82	3	0,85	3	
Rpp30	1,31	2	1,14	2	NaN	0	NaN	0	1,72	2	2,00	2	1,08	2	1,02	2	
Sae1	1,26	3	1,51	3	0,63	7	0,72	7	1,40	6	1,76	6	0,77	6	0,80	6	
Safb	1,10	3	1,28	3	0,87	40	0,99	40	2,37	4	2,91	4	1,32	6	1,25	6	
Sart3	1,30	2	1,60	2	0,65	5	0,84	5	1,67	4	2,15	4	0,85	9	0,95	9	
Set	1,12	9	1,11	9	0,69	14	0,76	14	1,61	7	2,13	7	0,89	7	0,96	7	

Appendix 2

Sf1	NaN	0	NaN	0	0.72	18	0.88	18	2.17	2	2.39	2	0.88	5	0.93	5	
Sfpq	0.98	49	1.15	49	0.84	113	0.97	113	2.10	57	2.40	57	1.10	62	1.05	62	↑
Sin3a	NaN	0	NaN	0	0.60	4	0.69	4	1.36	2	1.74	2	0.87	5	0.79	5	
Skiv2l2	1.23	7	1.54	7	0.74	15	0.82	15	2.01	6	2.54	6	0.98	11	1.01	11	
Smarca4	1.09	4	1.35	4	0.84	25	0.84	25	1.65	22	1.97	22	0.89	52	0.85	52	
Smarcc1	1.16	1	1.90	1	0.74	11	0.81	11	1.56	13	2.53	13	0.86	47	0.80	47	
Smarcc2	0.85	1	49.56	1	0.92	1	0.79	1	1.81	7	2.19	7	0.90	12	0.82	12	↑
Smarcd2	NaN	0	NaN	0	0.70	2	0.79	2	1.09	5	1.33	5	0.90	24	0.83	24	
Smc1a	1.30	25	1.23	25	0.78	37	0.85	37	1.64	58	2.17	58	0.92	45	0.91	45	
Smc3	1.28	22	1.33	22	0.77	38	0.81	38	1.59	56	2.13	56	0.95	59	0.91	59	
Smchd1	NaN	0	NaN	0	NaN	0	NaN	0	1.60	1	1.78	1	0.90	2	0.93	2	
Snrnp70	1.23	8	1.38	8	0.83	25	0.97	25	1.97	5	2.09	5	1.08	10	1.09	10	↑
Snrpa	1.32	11	1.72	11	0.82	17	0.98	17	1.63	13	1.95	13	0.99	13	0.99	13	
Snrpa1	1.17	3	1.41	3	0.78	9	0.90	9	1.98	2	1.75	2	1.01	7	1.03	7	
Snrpb2	1.43	1	1.42	1	0.81	5	1.00	5	NaN	0	NaN	0	1.06	5	1.07	5	↑
Spi1	0.85	6	1.00	6	0.79	8	0.75	8	1.17	77	1.11	77	0.91	101	0.82	102	
Srrt	1.10	6	1.30	6	0.74	40	0.86	40	1.94	23	2.42	23	0.89	33	0.86	33	
Srsf1	1.12	15	1.25	15	0.92	31	1.01	31	1.77	16	2.06	16	1.16	16	1.06	16	↑
Srsf2	1.25	7	1.35	7	0.84	21	0.91	21	1.84	16	1.91	16	1.07	12	0.99	12	
Srsf5	1.05	3	1.31	3	0.89	6	1.02	6	1.73	6	2.04	6	1.02	3	1.12	3	↑
Srsf6	1.07	4	1.26	4	0.89	39	0.96	39	1.63	9	1.97	9	1.07	17	1.03	17	
Srsf7	1.12	13	1.25	13	0.88	37	1.00	37	1.65	17	1.81	16	1.14	17	1.08	17	
Ssb	1.18	14	1.13	14	0.77	28	0.81	28	1.50	20	1.94	20	0.99	15	1.00	15	
Ssrp1	1.29	14	1.29	14	0.88	32	0.86	32	1.69	10	2.22	10	0.98	17	0.97	17	
Sumo2	1.36	10	1.52	10	0.97	8	1.00	8	1.63	15	2.19	15	0.98	9	1.03	9	
Supt5h	1.21	7	1.45	7	0.82	20	0.92	20	1.72	11	1.49	11	0.97	7	1.09	7	
Taldo1	1.38	33	1.18	33	0.66	27	0.79	27	1.92	24	1.94	24	0.78	11	0.88	11	
Tardbp	1.15	8	1.29	8	0.80	30	0.87	30	1.81	18	2.26	18	0.93	20	0.97	20	
Tcerg1	1.31	2	1.96	2	0.51	8	0.84	8	1.90	8	2.18	8	0.92	14	0.98	14	

Appendix 2

Thumpd1	1.01	3	1.11	3	0.73	5	0.87	5	1.55	5	1.92	5	0.91	5	0.90	5	
Tkt	1.26	57	1.40	57	0.71	50	0.80	50	1.76	42	2.10	42	0.81	26	0.85	26	
Tmpo	0.99	9	1.21	9	0.80	62	0.85	62	0.99	6	2.30	6	1.07	16	1.01	16	
Trim28	1.05	20	1.36	20	0.75	49	0.80	49	1.73	65	2.23	65	0.98	72	0.94	72	
Uba2	1.29	4	1.46	4	0.68	7	0.85	7	1.78	2	2.17	2	0.83	7	0.89	7	
Ubc	0.90	69	0.86	69	0.79	101	0.74	101	1.15	45	1.35	45	0.93	50	0.81	50	
Usp7	1.35	2	1.62	2	0.79	6	0.83	6	1.49	4	2.38	4	0.89	11	0.86	11	
Zeb2	1.73	1	2.63	1	0.77	3	1.09	3	1.62	7	2.59	7	0.97	8	1.20	8	



**APPENDIX 3 List of proteins newly recruited at enhancers in *tc-ChroP*.** Proteins were identified based on their dynamic SILAC ratio in four X-ChroP experimental replicas (H3K4me1 Forward, H3K4me1 Reverse, Pu.1 Forward, Pu.1 Reverse). (1hr/UT correspond to M/L and H/L in the Forward and Reverse experiment, respectively. 4hr/UT correspond to H/L and M/L in the Forward and Reverse experiment, respectively). Fasta headers, Protein names and Gene names are provided together with number of peptides and unique peptides in all the four experiments

Appendix 3

Fasta headers	Protein names	Gene names	Peptides				Unique peptides				H3K4me1 Fwd				H3K4me1 Rev				Pu.1 Fwd				Pu.1 Rev			
			1h/UT	1h/UT RC	4h/UT	4h/UT RC	1h/UT	1h/UT RC	4h/UT	4h/UT RC	1h/UT	1h/UT RC	4h/UT	4h/UT RC	1h/UT	1h/UT RC	4h/UT	4h/UT RC	1h/UT	1h/UT RC	4h/UT	4h/UT RC				
Q9JIX8	Apoptotic chromatin condensation inducer in the nucleus	Acin1	4	13	5	5	4	13	5	5	1,14	4	1,44	4	0,87	31	0,93	31	2,01	5	2,65	5	1,09	9	0,99	9
Q9QUJ7	Long-chain-fatty-acid--CoA ligase 4	Acs14	0	6	2	4	0	5	2	3	NaN	0	NaN	0	0,94	6	0,92	6	1,20	2	1,60	2	1,00	4	1,12	4
Q9JII6	Alcohol dehydrogenase [NADP(+)]	Akr1a1	7	5	5	5	7	5	5	5	1,33	6	1,72	6	0,70	6	0,86	6	1,79	5	2,33	5	0,82	6	1,06	6
P84091	AP-2 complex subunit mu	Ap2m1	0	4	5	7	0	4	5	7	NaN	0	NaN	0	0,94	3	0,97	3	0,59	2	4,26	2	1,11	6	0,90	6
Q8K019	Bcl-2-associated transcription factor 1	Bclaf1	2	9	7	6	2	9	7	6	1,14	3	1,17	3	0,85	34	0,95	34	2,06	16	2,38	16	1,14	28	1,01	28
Q9DCA5	Ribosome biogenesis protein BRX1 homolog	Brix1	3	5	5	3	3	5	5	3	1,13	5	1,10	5	0,77	5	0,78	5	2,01	5	2,26	5	1,28	2	0,96	2
Q64152	Transcription factor BTF3	Btf3	0	4	2	3	0	4	2	3	NaN	0	NaN	0	1,02	7	0,93	7	NaN	0	NaN	0	1,19	3	1,22	3
P53569	CCAAT/enhancer-binding protein zeta	Cebpz	5	4	4	2	5	4	4	2	1,11	5	1,28	5	0,66	5	0,84	5	1,89	5	2,67	5	1,20	2	0,85	2
Q8BMK4	Cytoskeleton-associated protein 4	Ckap4	1	11	2	5	1	11	2	5	1,19	1	1,47	1	1,09	13	0,93	13	0,56	1	0,74	1	1,16	6	1,10	6
Q9EPU4	Cleavage and polyadenylation specificity factor subunit 1	Cpsf1	1	8	6	9	1	8	6	9	1,81	1	1,47	1	0,78	9	0,95	9	1,96	5	2,34	5	1,00	11	0,97	11
H3BJW3	Cleavage and polyadenylation specificity factor subunit 6	Cpsf6	2	11	3	7	2	11	3	7	1,43	2	2,41	2	0,84	23	1,04	23	1,60	4	3,12	4	1,13	11	1,07	11
Q9JKB3	DNA-binding protein A	Csda	0	3	5	2	0	1	1	1	NaN	0	NaN	0	0,99	1	0,79	1	NaN	0	NaN	0	1,00	1	0,92	1
D3Z0M9	DEAD (Asp-Glu-Ala-Asp) box polypeptide 23	Ddx23	6	6	10	8	6	6	10	8	1,28	5	1,41	5	0,81	8	0,98	8	1,85	12	2,23	12	1,08	11	1,03	11

Appendix 3

Q8VDW0	ATP-dependent RNA helicase DDX39A	Ddx39a	10	15	10	11	2	5	2	1	1,10	2	1,27	2	0,87	7	0,94	7	NaN	0	NaN	0	1,01	1	0,59	1
Q569Z5	Probable ATP-dependent RNA helicase DDX46	Ddx46	1	21	7	14	1	21	7	14	1,23	3	1,15	3	0,83	33	0,98	33	1,95	8	2,43	8	0,96	25	1,00	25
Q9CWX9	Probable ATP-dependent RNA helicase DDX47	Ddx47	0	2	4	2	0	2	4	2	NaN	0	NaN	0	0,97	2	0,82	2	0,61	3	10,92	3	0,94	1	1,11	1
A2AH85	116 kDa U5 small nuclear ribonucleoprotein component	Eftud2	9	19	14	16	8	18	13	15	1,12	10	1,40	10	0,77	26	0,94	26	1,98	15	2,35	15	1,04	17	1,04	17
P23116	Eukaryotic translation initiation factor 3 subunit A	Eif3a	0	16	2	6	0	16	2	6	NaN	0	NaN	0	1,15	26	1,02	26	0,87	1	9,73	1	1,20	9	1,16	9
P60229	Eukaryotic translation initiation factor 3 subunit E	Eif3e	1	6	8	3	1	6	8	3	NaN	0	NaN	0	0,94	6	0,93	6	0,69	8	7,42	8	1,01	4	1,08	4
Q9DBC3	Cap methyltransferase 1	Ftsjd2	2	2	3	1	2	2	3	1	1,13	2	1,21	2	0,95	3	0,88	3	1,83	2	2,16	2	0,71	1	0,76	1
Q61598	Rab GDP dissociation inhibitor beta	Gdi2	0	6	7	2	0	6	6	2	NaN	0	NaN	0	0,66	7	0,86	7	1,31	4	11,11	4	1,17	2	1,16	2
P68040	Guanine nucleotide-binding protein subunit beta-2-like 1	Gnb2l1	10	17	11	12	10	17	11	12	1,14	11	1,06	11	0,99	23	0,89	23	1,13	16	1,55	16	1,04	14	1,02	14
Q99ME9	Nucleolar GTP-binding protein 1	Gtpbp4	8	13	9	9	8	13	9	9	1,18	14	1,20	14	0,80	18	0,87	18	1,85	9	2,23	9	1,05	9	0,97	9
P10922	Histone H1.0	H1f0	7	9	6	6	7	9	6	6	0,95	29	0,98	29	0,89	70	0,89	70	1,24	10	1,45	10	1,26	14	1,15	14
Q6NVF4	DNA helicase B	Helb	1	5	2	3	1	5	2	3	1,59	1	1,60	1	1,20	5	1,12	5	2,05	2	2,72	2	1,31	3	1,14	3
P22752	HISTONE H2A type 1	Hist1h2ab	4	5	6	5	1	1	1	1	1,24	5	1,08	5	0,90	5	1,00	5	1,88	7	1,93	7	1,14	1	1,04	1
P15092	Interferon-activable protein 204	Ifi204	5	8	3	7	4	5	2	4	1,10	4	1,85	4	0,80	8	1,35	8	1,36	3	2,56	3	1,07	9	1,81	9
P09450	Transcription factor jun-B	Junb	1	2	3	2	1	2	2	2	NaN	0	NaN	0	2,05	2	2,80	2	2,70	3	5,32	3	2,60	5	2,90	5

Appendix 3

A2BE28	Ribosomal biogenesis protein LAS1L	Las1l	3	2	4	5	3	2	4	5	1,06	2	1,05	2	0,70	2	0,88	2	2,94	7	2,33	7	1,07	5	0,94	5
Q5SUF2	Luc7-like protein 3	Luc7l3	3	9	0	5	3	9	0	5	1,09	4	1,14	4	0,83	13	0,97	13	NaN	0	NaN	0	1,03	6	0,98	6
Q9WTL7	Acyl-protein thioesterase 2	Lypla2	0	3	2	4	0	3	2	4	NaN	0	NaN	0	0,81	3	0,72	3	1,90	4	1,74	4	1,04	7	0,93	7
D3Z3G6	Mitogen-activated protein kinase 3	Mapk3	2	4	4	6	1	2	3	5	1,69	2	1,95	2	1,00	5	1,02	5	2,08	5	2,61	5	0,98	6	1,13	6
DOQMC3	Myeloid cell nuclear differentiation antigen-like protein	Mndal	11	8	4	4	6	6	3	3	1,02	21	1,20	21	0,86	13	1,08	13	1,43	9	2,06	9	0,94	8	1,13	8
P70670	Nascent polypeptide-associated complex subunit alpha, muscle-specific form	Naca	3	4	2	2	3	4	2	2	1,17	5	1,04	5	1,01	12	0,88	12	1,03	5	1,32	5	1,04	2	0,97	2
Q91W39	Nuclear receptor coactivator 5	Ncoa5	0	10	1	6	0	3	0	2	NaN	0	NaN	0	0,81	27	0,99	27	1,93	1	3,23	1	1,32	9	1,25	9
Q3TZX8	Polynucleotide 5-hydroxyl-kinase NOL9	Nol9	2	5	6	3	2	5	6	3	0,84	3	1,13	3	0,91	5	0,63	5	2,53	5	2,14	5	1,13	3	0,92	3
Q99JX7	Nuclear RNA export factor 1	Nxf1	5	11	5	10	5	11	5	10	1,14	5	1,34	5	0,86	16	0,97	16	1,79	8	2,16	8	1,21	14	1,07	14
P50580	Proliferation-associated protein 2G4	Pa2g4	8	20	18	13	8	20	18	13	1,20	13	1,00	13	1,00	44	0,92	44	0,80	30	10,77	30	1,13	18	0,99	18
P29341	Polyadenylate-binding protein 1	Pabpc1	5	16	16	17	4	8	7	10	1,20	6	1,70	6	0,98	32	1,07	32	1,04	19	6,02	19	0,98	25	1,00	25
E9QN87	Plectin	Plec	6	42	12	27	6	41	12	25	2,01	5	0,96	5	0,91	44	0,91	44	0,93	8	1,75	8	1,19	33	1,19	33
Q8BVQ9	26S protease regulatory subunit 7	Psmc2	0	2	7	2	0	2	7	2	NaN	0	NaN	0	0,96	2	0,90	2	0,67	5	7,23	5	0,96	2	1,09	2
P70388	DNA repair protein RAD50	Rad50	1	6	5	11	1	6	5	11	1,02	1	1,23	1	0,92	7	0,93	7	1,80	5	2,44	5	1,00	12	1,01	12
Q8C2Q3	RNA-binding protein 14	Rbm14	10	23	11	15	7	18	8	12	1,14	13	1,55	13	0,84	69	0,99	69	2,31	15	3,13	15	1,26	28	1,25	28
Q9JJ80	Ribosome production factor 2 homolog	Rpf2	7	6	6	3	7	6	6	3	1,05	8	1,12	8	0,79	6	0,81	6	1,89	7	2,22	7	1,22	3	0,86	3



Appendix 3

Q6ZWW3	60S ribosomal protein L10	Rpl10a	4	9	6	6	4	9	6	6	1,13	10	1,18	10	0,91	29	0,87	29	0,99	11	1,35	11	1,10	11	0,98	11
Q5XJF6	60S ribosomal protein L10a	Rpl10a	7	10	3	9	7	10	3	9	1,34	10	1,35	10	1,02	32	0,97	32	1,00	6	1,17	6	1,13	17	1,03	17
P19253	60S ribosomal protein L13a	Rpl13a	5	12	1	8	5	12	1	8	1,30	5	1,15	5	1,01	25	0,90	25	0,75	3	1,03	3	1,21	11	1,02	11
Q9CR57	60S ribosomal protein L14	Rpl14	5	7	3	6	5	7	3	6	1,33	14	1,20	14	0,98	24	0,87	24	0,96	6	1,12	6	1,14	9	1,03	9
E9QAZ2	Ribosomal protein L15	Rpl15	8	12	5	7	8	12	5	7	1,19	15	1,17	15	1,01	26	0,94	26	1,03	17	1,23	17	1,18	10	1,09	10
Q6ZWZ7	60S ribosomal protein L17	Rpl17	2	10	5	4	2	10	5	4	1,24	4	0,98	4	1,05	29	0,95	29	0,93	5	1,06	5	1,11	8	1,03	8
P35980	60S ribosomal protein L18	Rpl18	6	10	4	5	6	10	4	5	0,88	12	1,19	12	1,01	28	0,93	28	0,87	5	1,30	5	1,03	8	1,06	8
P62717	60S ribosomal protein L18a	Rpl18a	5	12	8	8	5	12	8	8	1,08	12	1,14	12	1,00	22	0,92	22	1,23	19	1,55	19	1,17	14	1,02	14
Q8BP67	60S ribosomal protein L24	Rpl24	4	10	4	5	4	10	4	5	1,21	12	1,18	12	1,09	29	0,97	29	1,01	10	1,27	10	1,06	8	1,03	8
P61255	60S ribosomal protein L26	Rpl26	4	9	2	8	4	9	2	8	1,13	4	1,24	4	1,01	29	0,89	29	1,01	3	1,08	3	1,10	11	0,99	11
P14115	60S ribosomal protein L27a	Rpl27a	4	5	2	4	4	5	2	4	0,88	8	1,15	8	1,03	16	0,90	16	0,77	5	0,97	5	1,17	7	1,07	7
P41105	60S ribosomal protein L28	Rpl28	3	10	1	7	3	10	1	7	1,16	3	1,20	3	0,99	25	0,90	25	0,98	2	1,33	2	1,12	10	0,99	10
P27659	60S ribosomal protein L3	Rpl3	11	24	16	16	11	24	16	16	1,19	24	1,13	24	1,05	81	0,94	81	1,14	31	1,54	31	1,18	29	1,05	29
P62911	60S ribosomal protein L32	Rpl32	4	11	3	7	4	11	3	7	1,13	6	1,15	6	1,01	19	0,89	19	0,90	5	0,97	5	1,17	8	1,06	8
Q9D8E6	60S ribosomal protein L4	Rpl4	11	28	23	10	11	28	23	10	1,32	22	1,16	22	0,95	127	0,90	127	0,35	26	17,42	26	1,17	19	1,12	19
P47962	60S ribosomal protein L5	Rpl5	9	12	8	8	9	12	8	8	1,20	15	1,06	15	0,99	26	0,91	26	1,20	14	1,47	14	1,06	10	0,97	10
P47911	60S ribosomal protein L6	Rpl6	15	19	17	15	15	19	17	15	1,22	40	1,06	40	0,97	91	0,88	91	1,20	36	1,48	36	1,17	38	1,02	38
P14148	60S ribosomal protein L7	Rpl7	12	23	11	14	12	23	11	14	1,27	26	1,04	26	0,95	66	0,85	66	0,82	17	1,04	17	1,14	18	1,01	18
D3Z3R0	60S ribosomal protein L7a	Rpl7a	13	21	14	15	13	21	14	15	1,26	31	1,13	31	0,97	80	0,92	80	1,20	23	1,49	23	1,13	33	0,98	33
P62918	60S ribosomal protein L8	Rpl8	7	14	7	14	7	14	7	14	1,08	14	1,18	14	1,04	50	0,92	50	0,98	18	1,39	18	1,14	28	1,03	28
P62281	40S ribosomal protein S11	Rps11	8	14	10	8	8	14	10	8	1,16	19	1,07	19	1,10	36	0,93	36	1,18	24	1,49	24	1,12	13	0,94	13
P62301	40S ribosomal protein S13	Rps13	4	6	5	6	4	6	5	6	1,56	4	1,34	4	1,14	11	1,00	11	0,89	7	1,12	7	1,16	8	1,04	8

Appendix 3

P62849	40S ribosomal protein S24	Rps24	3	6	2	3	3	6	2	3	1,26	2	1,06	2	1,10	17	0,94	17	1,12	2	1,44	2	1,24	3	1,02	3
D3Z6C3	40S ribosomal protein S3a	Rps3a	12	22	17	12	12	22	17	12	1,26	33	1,21	33	1,11	69	0,91	69	0,97	26	1,30	26	1,16	29	0,98	29
P62702	40S ribosomal protein S4, X isoform	Rps4x	11	22	13	17	11	22	13	17	1,13	22	1,19	22	1,00	63	0,89	63	1,00	41	1,34	41	1,07	20	0,97	20
P62754	40S ribosomal protein S6	Rps6	10	11	10	7	10	11	10	7	1,17	22	1,18	22	1,03	38	0,90	38	1,33	11	1,81	11	1,12	13	1,06	13
P18654	Ribosomal protein S6 kinase alpha-3	Rps6ka3	0	8	1	5	0	8	1	5	NaN	0	NaN	0	0,78	8	1,06	8	1,56	2	2,24	2	0,97	8	1,16	8
P62082	40S ribosomal protein S7	Rps7	7	13	9	10	7	13	9	10	1,34	10	1,22	10	1,20	25	0,93	25	1,02	14	1,28	13	1,23	16	1,05	16
P62242	40S ribosomal protein S8	Rps8	9	12	9	11	9	12	9	11	1,13	25	1,13	25	1,04	54	0,93	54	1,00	18	1,17	17	1,12	23	1,00	23
Q80YR5	Scaffold attachment factor B2	Safb2	1	16	4	4	0	11	2	4	NaN	0	NaN	0	0,82	24	0,98	24	2,19	2	3,44	2	1,36	4	1,36	4
Q8CH25	SAFB-like transcription modulator	Sltm	2	22	6	10	2	21	6	10	1,18	3	1,40	3	0,96	75	1,08	75	2,43	9	3,10	9	1,20	20	1,21	20
Q78PY7	Staphylococcal nuclease domain-containing protein 1	Snd1	3	25	11	12	3	25	11	12	1,34	3	1,37	3	1,15	41	0,87	41	1,43	14	1,43	14	1,15	18	0,98	18
P84104	Serine/arginine-rich splicing factor 3	Srsf3	3	6	4	5	2	5	3	4	1,04	4	1,15	4	0,88	14	0,95	14	1,60	8	1,83	7	1,12	7	1,05	7
Q8C3V4	Signal transducer and activator of transcription 1	Stat1	2	7	4	5	2	7	4	5	1,15	2	1,59	2	0,88	7	0,95	7	1,44	4	2,16	4	0,97	5	1,33	5
Q62383	Transcription elongation factor SPT6	Supt6h	4	9	2	7	4	9	2	7	1,33	2	1,51	2	0,83	10	0,93	10	2,04	1	2,81	1	0,94	7	1,09	7
H3BL37	Treacle protein	Tcof1	7	9	5	2	7	9	5	2	0,76	35	0,90	35	0,66	13	1,15	13	1,24	8	1,24	8	1,06	3	1,15	3
P62996	Transformer-2 protein homolog beta	Tra2b	3	7	4	3	3	5	4	3	1,17	7	1,24	7	0,88	17	0,98	17	1,93	6	2,12	6	1,19	3	1,02	3
Q640M1	U3 small nucleolar RNA-associated protein 14 homolog A	Utp14a	2	5	3	3	2	5	3	3	1,11	2	1,40	2	0,77	5	0,98	5	1,77	3	2,62	3	1,29	3	1,10	3
P20152	Vimentin	Vim	22	29	32	29	19	26	29	27	1,02	72	1,06	72	1,01	102	0,83	102	1,14	127	1,57	127	1,22	86	1,01	86

**APPENDIX 4 hPTMs identified and quantified at enhancers.** Theoretical MH<sup>+</sup>, number of lysines contained in the sequence and alkylated in ArgC-like digestion, m/z used for XIC reconstruction and retention time normally observed are provided for all the peptides on histone H3, H2A and H4 reported together with amino acid sequence of the peptide and residue modified (in red).

Modification	Theoretical MH+	N° lysine (D3 mod.)	MH2+ D3	RT
<b>H3 (3-8) peptide = TK<sub>4</sub>QTAR</b>				
K <sub>4</sub> unmod	704.4050	1	375.2208	45.89
K <sub>4</sub> me1	718.4206	1	382.2286	54.21
K <sub>4</sub> me2	732.4363	0	366.7218	29.19
K <sub>4</sub> me3	746.4519	0	373.7296	28.22
<b>H3 (9-17) peptide = K<sub>9</sub>STGGK<sub>14</sub>APR</b>				
K <sub>9</sub> unmod/K <sub>14</sub> unmod	901.5214	2	496.2937	62.46
K <sub>9</sub> me1/K <sub>14</sub> unmod	915.5370	2	503.3016	68.84
K <sub>9</sub> me2/K <sub>14</sub> unmod	929.5527	1	487.7947	50.64
K <sub>9</sub> me3/K <sub>14</sub> unmod	943.5683	1	494.8025	50.35
K <sub>9</sub> unmod/K <sub>14</sub> Acetyl	943.5320	1	494.7843	62.67
K <sub>9</sub> me1/K <sub>14</sub> Acetyl	957.5476	1	501.7921	68.95
K <sub>9</sub> me2/K <sub>14</sub> Acetyl	917.5633	0	486.2853	51.04
K <sub>9</sub> me3/K <sub>14</sub> Acetyl	985.5789	0	493.2931	50.46
K <sub>9</sub> Acetyl/K <sub>14</sub> Acetyl	985.5425	0	493.2749	62.81
<b>H3 (18-26) peptide = K<sub>18</sub>QLATK<sub>23</sub>AAR</b>				
K <sub>18</sub> unmod/K <sub>23</sub> unmod	986.6105	2	538.8383	79.16
K <sub>18</sub> me1/K <sub>23</sub> unmod	1000.6262	2	545.8461	87.46
K <sub>18</sub> Acetyl/K <sub>23</sub> unmod	1028.6211	1	537.3289	79.77
K <sub>18</sub> Unmod/K <sub>23</sub> Acetyl	1028.6211	1	537.3289	79.77
K <sub>18</sub> me1/K <sub>23</sub> Acetyl	1042.6368	1	544.3367	86.11
K <sub>18</sub> Acetyl/K <sub>23</sub> Acetyl	1070.6317	0	535.8195	80.19
<b>H3 (73-83) peptide = EIAQDFK<sub>79</sub>TDLR</b>				
K <sub>79</sub> unmod	1335.6903	1	690.8635	118.49
K <sub>79</sub> me1	1349.7060	1	697.8713	125.68
K <sub>79</sub> me2	1363.7216	0	682.3644	105.25
K <sub>79</sub> me3	1377.7373	0	689.3723	107.55
<b>H3 (27-40) peptide = K<sub>27</sub>SAPATGGVK<sub>36</sub>K<sub>37</sub>PHR</b>				
K <sub>27</sub> /K <sub>36</sub> /K <sub>37</sub> unmod	1433.8336	3	784.9645	84.16
K <sub>27</sub> me1	1447.8492	3	791.9723	87.82
K <sub>36</sub> me1	1447.8492	3	791.9723	87.06
K <sub>27</sub> me2	1461.8649	2	776.4655	74.1
K <sub>36</sub> me2	1461.8649	2	776.4655	78
K <sub>36</sub> me3	1461.8649	2	776.4655	73.81
K <sub>27</sub> me3	1475.8805	2	783.4733	78
K <sub>27</sub> me2/K <sub>36</sub> me1	1475.8805	2	783.4733	76.89
K <sub>27</sub> me1/K <sub>36</sub> me2	1475.8805	2	783.4733	82.02
K <sub>27</sub> me1/K <sub>36</sub> me1	1461.8649	3	798.9802	90.65
K <sub>27</sub> me2/K <sub>36</sub> me2	1489.8962	1	767.9664	69.5
K <sub>27</sub> me3/K <sub>36</sub> me1	1489.8962	2	790.4811	76.73
K <sub>27</sub> me1/K <sub>36</sub> me3	1489.8962	2	790.4811	82.02

Modification	Theoretical MH+	N° lysine (D3 mod.)	MH2+ D3	RT
<b>H4 (4-17) peptide = GK<sub>5</sub>GGK<sub>8</sub>GLGK<sub>12</sub>GGAK<sub>16</sub>R</b>				
K <sub>5</sub> /K <sub>8</sub> /K <sub>12</sub> /K <sub>16</sub> unmod	1270.7702	4	725.9476	87.59
Mono-Acetyl K	1312.7808	3	724.4381	87.75
Di-Acetyl K	1354.7914	2	722.9287	87.96
Tri-Acetyl K	1396.8019	1	721.4193	88.17
Tetra-Acetyl K	1438.8125	0	719.9099	88.24
<b>H2A (4-11) peptide = GK<sub>5</sub>QGGK<sub>9</sub>AR</b>				
K <sub>5</sub> unmod/K <sub>9</sub> unmod	801.4690	2	446.2675	52.52
K <sub>5</sub> Acetyl/K <sub>9</sub> unmod	843.4795	1	444.7581	52.76
K <sub>5</sub> unmod/K <sub>9</sub> Acetyl	843.4795	1	444.7581	52.76
K <sub>5</sub> Acetyl/K <sub>9</sub> Acetyl	885.4901	0	443.2487	52.85