PhD degree in Molecular Medicine

(Curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM)

University of Milan and University of Naples "Federico II"

The role of arginine methylation in miRNA biogenesis investigated by MS-proteomics

Valeria Spadotto

European Institute of Oncology (IEO), Milan

Supervisor: Dr. Tiziana Bonaldi

European Institute of Oncology (IEO), Milan

My work is a game, a very serious game.

M. C. Escher

TABLE OF CONTENTS

FIGURE INDEX

LIST OF ABBREVIATIONS

3p	3 prime
5p	5 prime
ADMA	Asymmetric di-methylated arginine
Arg	Arginine
Co-IP	co-immunopurification
Da	Dalton
DDA	Data dependent acquisition
DSRM	Double strand RNA binding motif
ESI	Electrospray ionization
FDR	False discovery rate
Fwd	Forward experiment
geLC-MS/MS	In-gel digestion couple to tandem mass spectrometry
H4R3me2a	Histone H4 arginine 3 asymmetric di-methylation
HCD	High energy Collision Dissociation
HILIC	Hydrophilic interaction liquid chromatography
hmSILAC	Heavy methyl stable isotope labeling by amino acids in cell culture
HPLC	High pressure liquid chromatography
IP	Immunoaffinity purification
KD	knock down
LC	Liquid chromatography
LC MS/MS	Liquid chromatography and tandem MS
LC region	Low complexity region
LDC	Large Drosha Complex
LFQ	Label-free quantitation
Lys	Lysine
m/z	Mass-to-charge ratio
MALDI	Matrix assisted laser desorption ionization
Met	Methionine
miRISC	MicroRNA-induced silencing complex
miRNA	MicroRNA
MMA	Mono-methylated arginine
MS	Mass spectrometry
MS/MS	Tandem Mass spectrometry
nLC	Nano Liquid chromatography
Over	Overexpression
Pfam	Protein families database

PI	Post infection
PKMTs	Protein lysine Methyltransferases
Pre-miRNA	Precursor microRNA
Pri-miRNA	Primary microRNA
PRMTs	Protein Arginine Methyltransferases
PTM	Post translational modification
qPCR	Quantitative polymerase chain reaction
RBPs	RNA-binding proteins
Rev	Reverse experiment
RP	Reverse phase
RRM	RNA recognition motifs
SAM	S-Adenosyl-methionine
SCX	Strong cation exchange
SDMA	Symmetric di-methylated arginine
SDS	Sodium dodecyl sulfate
SDS-page	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Sh-RNA	Short hairpin RNA
SILAC	Stable isotope labeling by amino acids in cell culture
UTR	Un-translated region
WT	Wild type

1. ABSTRACT

MicroRNAs (miRNAs) are short, non-coding RNA molecules that fine tune gene expression at the post-transcriptional level. The efficient miRNA processing is fundamental to maintain their correct levels in the cell. However, although microRNA biogenesis is a tightly controlled pathway, alterations in this process occur in different pathological conditions, including cancer. Hence, dissecting the molecular mechanisms underlying miRNA biogenesis is crucial to understand how aberrant miRNA patterns are generated. The processing step operated by the Large Drosha Complex (LDC), which is crucial and rate limiting for the production of mature miRNAs, occurs in the nucleus and it consists in the cleavage of the primary miRNAs (pri-miRNAs) into the precursor miRNAs (pre-miRNAs). In our previous MS-based analysis of the cellular methylome, we showed that protein methylation frequently occurs on the majority of the LDC subunits [1], suggesting for the first time the possibility that this post translational modification (PTM) could contribute to the regulation of miRNA biogenesis.

To investigate the role of protein methylation in the regulation of the LDC, we thoroughly characterized the methylated sites on the subunits of the complex, thus generating the first comprehensive and high-confident methyl-proteome of the LDC.

By modulating the expression of PRMT1, the predominant Protein Arginine Methyltransferase in mammalian cells, we analyzed how methylations change at specific sites of distinct LDC subunits, thus identifying novel PRMT1 substrates. The alteration of the LDC-methylation state upon PRMT1 depletion correlated with an impairment of the pri- to- pre-miRNA processing step, which resulted in the massive deregulation of mature miRNA expression. Although the detailed molecular mechanism remains to be fully deciphered, the results described in this thesis uncover a key role of arginine methylation in the regulation of the LDC activity and, consequently, in miRNA biogenesis.

2. INTRODUCTION

2.1 MicroRNA biogenesis and the Large Drosha Complex

MicroRNAs (miRNAs) are short, non-coding RNA molecules that regulate gene expression by targeting specific mRNAs for degradation or translational repression, thus influencing diverse biological processes.

MicroRNAs are preferentially transcribed by RNA Polymerase II (RNA Pol II) into long primary transcripts (pri-miRNAs), which possess the characteristic 7-methyl-guanosine (m^7G) cap at the 5'-end, the poly-A tail at the 3'-end [2] and stem-loop structures where the mature miRNA sequences are embedded. MicroRNA genes are located in diverse regions of the genome: the majority of miRNAs derives from introns of non-coding or coding transcripts, while just some of them derive from exons [3, 4]. Often, miRNA loci, which are in close proximity, are transcribed as a unique transcript, giving rise to polycistronic transcription units [5]. In fact, many miRNAs are expressed as clusters that range from 2 to 19 miRNAs hairpins which are usually co-transcribed. Transcription of pri-miRNAs can be regulated positively or negatively by various transcription factors, such as p53 [6, 7], MYC [8] and ZEB1 and ZEB2 [9]. Moreover, epigenetic mechanisms, including both DNA methylation and histone modifications, also contribute to the regulation of miRNA expression [4].

In the nucleus, pri-miRNAs are cleaved by the type-III RNase enzyme Drosha to generate the precursor miRNAs (pre-miRNAs), which are ~60/100 nucleotides (nt) long [10]. Drosha cleaves the pri-miRNA hairpin 11bp far from the 'basal' junction and 22bp away from the 'apical' junction [11] **(Fig. 1)**. Pre-miRNAs are then exported by exportin-5 (XPO5) in the cytoplasm [12], where they are further processed by Dicer. The type-III RNase Dicer cleaves the pre-miRNAs close to the terminal loop sequence, producing small RNA duplexes about 22nt long [13]. Dicer interacts with the Argonaute proteins to form the miRNA-induced silencing complex (miRISC) [14]. One strand of the dsRNA is removed and degraded while the other (the guide strand) is retained by the miRISC and guides it to the recognition of complementary sequences within the 3' un-translated

2

regions (UTR) of target mRNAs, promoting their degradation or their translational inhibitions [15] **(Fig. 1)**.

In the multi-step event of miRNA biogenesis, the processing step operated by Drosha in the nucleus is crucial and rate limiting for the production of mature miRNA molecules. Although Drosha is the RNase effectively cleaving the RNA molecule, it requires binding to DiGeorge critical region 8 (DGCR8), forming the Microprocessor complex, to correctly interact with pri-miRNAs [16, 17]. DGCR8 binds Drosha through its C-terminus [11, 18] and provides additional RNA-binding recognizing the pri-miRNAs through its double strand RNA binding domains (dsRBDs). Although the Microprocessor is able to cleave primary transcripts, there is evidence suggesting that it requires accessory proteins to be fully efficient [19]. Different studies identified about 20 proteins as Microprocessor interactors which form a larger multi-protein complex called Large Drosha Complex (LDC) [17, 19]. These proteins are RNA-binding proteins (RBPs) that comprise the DEAD-box helicases DDX5 and DDX17 (also known as p68 and p72, respectively), a number of heterogeneous ribonucleoproteins (hnRNPs), the FET protein family EWSR, FUS, TAF15, and other factors.

The Microprocessor-associated proteins can modulate the specificity and/or activity of the Microprocessor and alterations in this regulation leads to aberrant miRNA biogenesis [20].
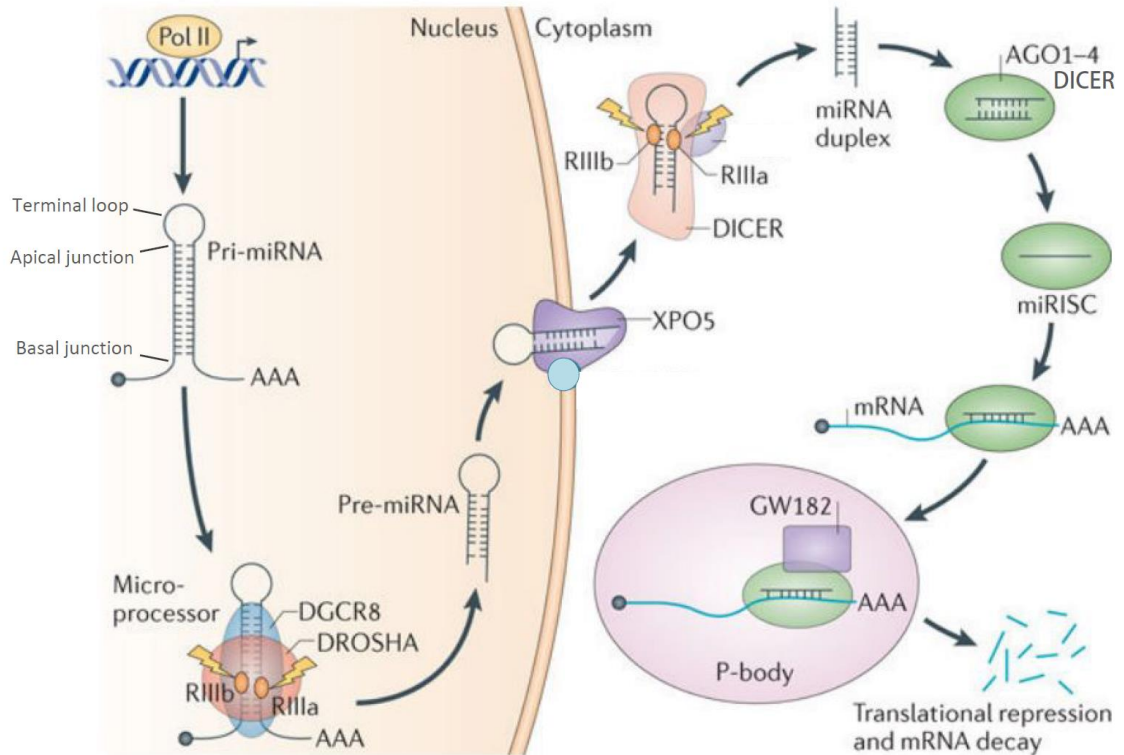
Fig. 1 Overview of microRNA biogenesis
MicroRNAs are transcribed as primary miRNAs (pri-miRNAs) by RNA polymerase II (Pol II) in the nucleus. The long pri-miRNAs are then cleaved by the Microprocessor, which is composed by DROSHA and DGCR8, to produce the pre-miRNAs. The pre-miRNAs are exported from the nucleus to the cytoplasm by exportin 5 (XPO5) and further processed by the type-III RNase DICER. The mature miRNA, also known as the guide strand from the miRNA duplex, is loaded into the miRNA-induced silencing complex (miRISC), which is composed from DICER and Argonaute (AGO) proteins. miRNA directs the RISC to target mRNAs and mediates gene suppression by mRNA degradation and/or translational repression. *Adapted from [21].*

2.2 Mechanisms of regulation of the Microprocessor

The efficiency of pri-miRNA processing is crucial for the maintenance of appropriate mature miRNAs level, thus several mechanisms of regulation of the Drosha complex exist.

One level of regulation is operated by accessory proteins within the LDC that regulate the efficiency of pri-miRNA processing by acting on the Microprocessor in different ways.

For instance, DDX5 and DDX17 are required for the efficient processing of a subset of miRNAs by proper recruiting the Microprocessor to their pri-miRNAs [22, 23]. Moreover, the interaction between DDX5 and/or DDX17 with the Drosha/DGCR8 complex can be further regulated by several signalling pathways [24, 25].

Some other accessory proteins, like TARDBP, increase Drosha stability and processing activity through a direct protein-protein interaction [26, 27], while ILF3 and ILF2 (also known as NF45 and NF90, respectively) are negative regulators of miRNA biogenesis. In fact, they inhibit pri-let-7a and pri-miR-21 processing into pre-miRNAs by sequestering them from the Microprocessor [28].

The protein levels of Drosha and DGCR8 in the cell are also tightly regulated since they may directly influence miRNA processing. A double-negative feedback loop ensures the accurate coupling of the Microprocessor proteins: on the one hand DGCR8 stabilizes Drosha protein level through direct interaction and, on the other, Drosha modulates the abundance of DGCR8 mRNA, cleaving two hairpin structure within its 5'UTR and leading its mRNA to degradation [29, 30].

Also post translational modifications (PTMs) can regulate several subunits of the complex. DGCR8 phosphorylation increases its stability, leading to an increase in pro-growth miRNA levels [31]; Drosha phosphorylation is required for its correct localization to the nucleus while acetylation prevents its ubiquitination and consequent degradation [32, 33]. In a previous study [1], we showed for the first time that protein methylation is frequently occurring on the LDC proteins. This observation was further confirmed by others [34] and points towards a role of this modification in the regulation of miRNA biogenesis.

2.2.1 MicroRNA biogenesis is altered in cancer

A number of large-scale miRNA expression profiling studies have recently shown that miRNAs are deregulated in different pathological conditions, including cancer [4]. Individual miRNAs can have either oncogenic or tumour-suppressive function, based on their inhibitory activity on tumour-suppressive or oncogenic targets, respectively. Moreover, miRNA signatures that distinguish between normal and tumoral samples have been identified [35, 36] and miRNA patterns of expression have shown to be useful tool to predict therapeutic outcomes in specific cancer types [37]. Interestingly, in the last years several studies showed that mature miRNAs are globally down-regulated in cancer, and

that this down-regulation is paralleled by the accumulation of pri-miRNAs, suggesting that the miRNA biogenesis might be impaired [36, 38].

Nowadays, it is well known that miRNA biogenesis plays a central role in cancer and that deregulation or deficiency of the components of the processing machinery often underlies cancer progression [21, 39, 40]. All the aforementioned regulatory mechanisms (paragraph 2.2) of miRNA processing pathway can be altered in cancer, as reviewed in [21]. Pri-miRNAs transcription can be affected by genetic alterations, altered epigenetic modifications or transcription factors, while pri-miRNA processing is deregulated by genetic mutations or altered expression levels of both the Microprocessor and the RBPs involved in its processing, such as DDX5 and DDX17. Whereas alterations of Drosha and DGCR8 globally affect mature miRNA levels, the deregulation of the accessory proteins impinges on specific subset of miRNAs.

Last, the export of the pre-miRNAs from the nucleus to the cytoplasm can also be affected, as well as the cytosolic processing of the precursors by Dicer and the Argonaute proteins.

For these reasons, there is increasing interest and necessity in dissecting the molecular mechanisms underlying miRNA biogenesis.

2.3 Protein methylation

Methylation is a post translational modification occurring on a variety of nuclear and cytoplasmic substrates. Initially, methylation was extensively studied as a key-determinant of the histone code in the context of the epigenetic regulation of gene transcription. However, the recent discovery that methylation occurs on an increasing number of non-histone proteins confirmed that it is implicated in every aspect of cellular biology, including transcription, RNA splicing, DNA repair, signal transduction and modulation of the immune response [34, 41-46]. Given its widespread involvement in different biological processes, deregulation of methylation correlates with diverse pathological conditions, such as neurodegenerative, cardiovascular diseases and cancer [47-50].

Methylation is a covalent modification that involves the addition of methyl-groups to the ε-

amine of lysine (K) and to the guanidino groups of arginine (R) residues. It is catalyzed by a family of enzymes called protein methyltransferases, specific for either lysines (PKMTs) or arginines (PRMTs). Lysine can be mono-, di- or tri-methylated by PKMTs, while arginine can be mono- or di-methylated by PRMTs. A distinct feature of R di-methylation is that it can be symmetric or asymmetric, depending on the number of nitrogen atoms of the guanidino group involved in the reaction [50] **(Fig. 2)**. Both PRMTs and PKMTs are S-Adenosyl-methionine (SAM or AdoMet) – dependent and catalyze the transfer of methyl-groups from the methyl-donor SAM to a plethora of biological targets in the cell. The addition of a methyl-group to an arginine or lysine residue within a protein reduces the hydrogen bonding potential of the K and R side chains and alters the bulkiness and hydrophobicity of the protein, but does not change the cationic charge of the residues. As a consequence, this can positively or negatively affect the interactions occurring between the substrate and other proteins [50, 51].
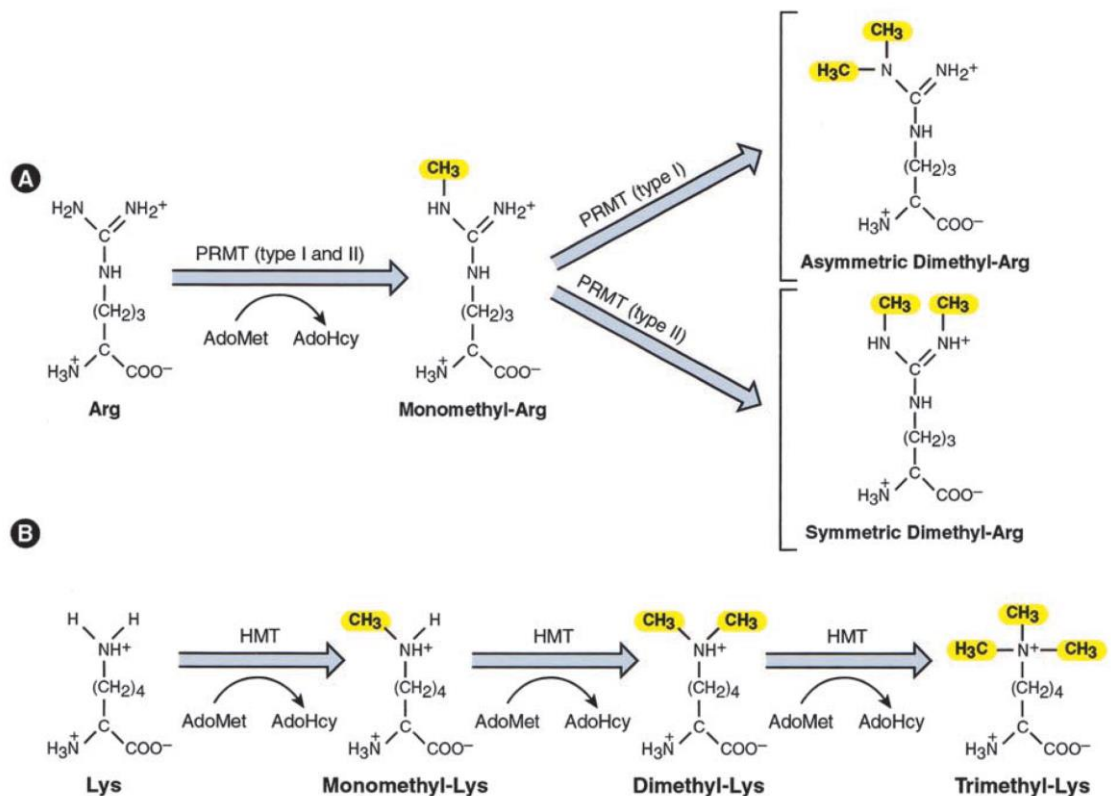
Fig. 2 Arginine and lysine methylation
A) Molecular structure of unmodified, mono- and di-methylated arginine. Type-I and II protein arginine methyltransferases catalyze asymmetric and symmetric di-methylation,

respectively. **B)** Molecular structure of unmodified, mono-, di-, and tri-methylated lysine. *Adapted from [52].*

In particular, arginine is structurally unique compared to other amino acids as its guanidino group is positively charged and can form extended hydrogen bonding networks with both proteins and nucleic acids [53]. The methylation of R residues alters both the shape of the R side-chain and the hydrogen bonding potential of the residue. Thus, cellular responses to mono- and symmetric/ asymmetric di-methylation, which are specific to the type of modification, all appear to be mediated by alteration in protein-protein, protein-RNA and protein-DNA interactions [41, 46].

For many years methylation was considered a stable and irreversible modification, but the reversal of lysine methylation is now well established thanks to the identification of two families of histone KDMs, which both act through the oxidation of the methyl-groups [54, 55]. In contrast to lysine methylation, the existence of arginine demethylases is still controversial. In fact, although the 2-oxoglutarate-dependent JMJD6 has been reported to catalyze methyl-arginine demethylation of histone and non-histone substrates [56] [57], the biochemical evidence for its activity is still not unequivocal [57, 58].

2.3.1 Protein Arginine methyltransferases

In mammals, arginine methylation is catalyzed by a family of nine members, called Protein Arginine Methyltransferases (PRMTs). All PRMTs have a highly conserved methyltransferase domain, while they contain a unique N-terminal region which varies in length and that can be responsible for substrates recognition and enzymatic activity regulation [46]. They can be classified according to the type of modification they catalyze: type-I enzymes catalyze the formation of mono-methylated arginine (MMA) and asymmetric di-methylated arginine (ADMA) (and include PRMT1, PRMT2, PRMT3, PRMT4, PRMT6 and PRMT8), type-II enzymes are responsible for MMA and symmetric di-methylated arginine (SDMA) (and comprise PRMT5 and PRMT9) and the type-III enzyme PRMT7 deposits only mono-methylation [41].

PRMT1 is the major type-I methyltransferases in mammalian cells and is responsible for

the 85% of the ADMA [59]. On histones, PRMT1 methylates histone H4 at arginine 3, generating the H4R3me2a mark [60], which is associated with transcriptional activation. In fact, this modification allows the recruitment of methyl-binding proteins and the deposition of other active PTM marks in the proximity [61]. Although several studies report that PRMT1methylates preferentially R located within arginine-glycine (RG) motifs, a recent evidence suggest that PRMT enzymes target arginine residues located in RG and non-RG sequence motifs with similar preference [34]. The PRMT1 gene can produce seven isoforms (PRMT1v1-v7) through complex alternative splicing in the 5'-end of its pre-mRNA. The isoforms differ in their N-terminal sequences and in their molecular weight; moreover, they have different enzymatic activity, substrate specificity, sub-cellular localization and interactions [62, 63]. Several studies have shown that PRMT1 is overexpressed in diverse types of cancer and that in human breast cancer cells and tumour tissues it is possible to observe an unbalance among distinct PRMT1 splicing variants, suggesting that these isoforms might be functionally different, especially in disease [64].

PRMT2 and PRMT3 are type-I PRMT that exhibits high sequence homology with the catalytic domain of PRMT1 [65]. The former has weak methyltransferase activity compared to PRMT1 [66] and, with the exception of H3R8 [67], specific substrates have not been identified yet. However, it has been shown to interact with estrogen and progesterone receptors in a ligand-dependent manner enhancing their activity [68]. The latter, methylates the ribosomal protein RPS2, thus regulating ribosome homeostasis, and p53 acting in a complex with ARF and VHL [69, 70],

PRMT4, also known as CARM1 (co-activator associated arginine methyltransferase 1) due to its role of transcriptional co-activator, is a type-I enzyme that methylates several histone and non-histone proteins. The recruitment of CARM1 to transcriptional promoters results in the methylation of histone H3 at R17 and R26 (H3R17me2a, H3R26me2a), marks associated to transcriptional activation [41, 71, 72].

PRMT5 is the major type-II PRMT in mammalian cells and exerts its activity both in the nucleus and in the cytoplasm. In the nucleus, it acts like a strong transcriptional repressor,

while in the cytoplasm it is involved in snRNP (Small Nuclear Ribonucleoproteins) biogenesis [73] and is a key regulator of the core splicing machinery [74].

PRMT6 is a nuclear type-I PRMT with a similar structure and preference for RG motifs compared to PRMT1 [75]. Its primary targets are histone H3, on which it produces the repressive mark H3R2me2a [76, 77], and DNA polymerase β [78]. Although its primary function is transcriptional repression, in some instances it has also been shown to act as a co-activator [79, 80].

PRMT7 activity is still controversial. In fact, it has been described to have type-III activity towards some substrates and type-II towards others, suggesting that distinct substrates might be methylated in different fashions by this enzyme [81, 82].

PRMT8 is a unique member of the PRMTs family since it is expressed only in the brain and is localized solely to the plasma membrane via myristoylation of its N-terminus [83]. Recently, it was reported that PRMT8 is involved in the pathogenesis of amyotrophic lateral sclerosis and that it plays critical roles in neural development [84, 85].

PRMT9, together with PRMT5, belongs to the type-II class of PRMTs. In fact it symmetrically methylates the spliceosome-associated protein SAP145, a component of the U2 snRNP that plays a critical role in the early stages of splicing [86].

2.4 Mass spectrometry and MS-based proteomics

In the last decade, technological improvements produced mass spectrometers characterized by progressively higher resolution, sensitivity and performance. Nowadays, the most widely used analytical setup for shotgun proteomic analysis is the liquid-chromatography coupled with tandem mass spectrometry (LC-MS/MS). The classical workflow for the identification and characterization of proteins from a complex mixture is the "bottom-up" approach, illustrated in **Fig. 3**, in which proteins are eventually separated and then digested with proteases into peptides prior to Mass Spectrometry (MS). The most common protease used in proteomic analyses is trypsin, which cleaves at the C-terminus of lysine and arginine residues unless they are followed by a proline residue.

Trypsin is commonly used in proteomics because its cleavage produces peptides about 10-12 amino acids long, which are easily detectable by MS.

In order to decrease the complexity of the mixture, peptides are then separated by nano-liquid chromatography -typically in reversed phase mode (RP-nLC)- which separates them according to their hydrophobicity. In RP-LC the polar and volatile mobile phase favors the ion transferring into the mass analyzer of the mass spectrometer, which measures the mass-to-charge ratio (m/z) of the peptide ions. In the mass spectrometer precursor peptides are isolated and scanned (full scan MS, MS or MS1) to identify the m/z that provides information on the elemental composition of the peptide. The most intense precursors are then fragmented into the constituent fragment ions through MS/MS events (tandem MS or MS2), to recover their amino acid sequence. Peptide identification is achieved by comparing the experimental MS/MS spectra derived from peptide fragmentation with theoretical MS/MS spectra generated by *in silico* digestion of a protein database. Protein identification is then inferred by assigning peptide sequences to proteins.
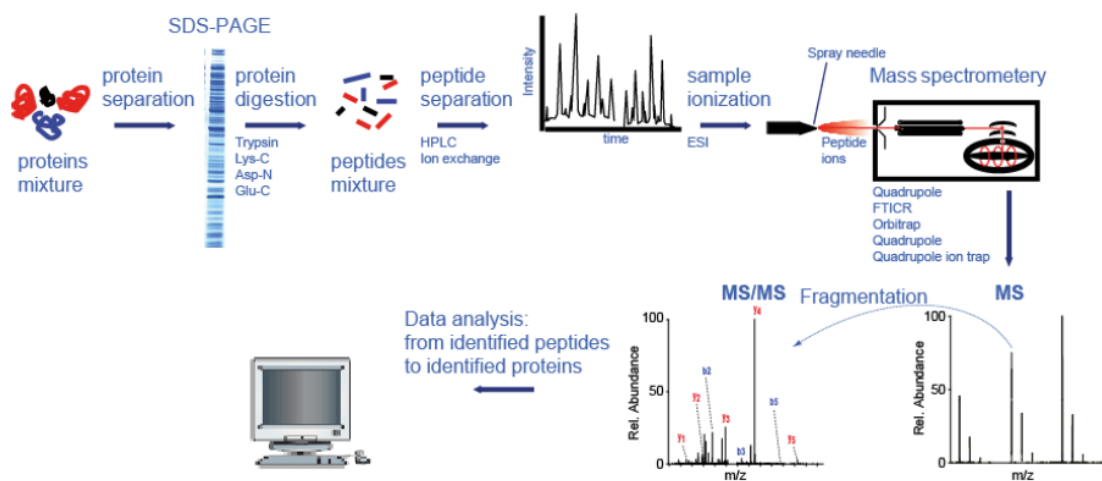
Fig. 3 Classical workflow for bottom-up MS analysis
A protein mixture from a biological sample is separated by SDS–PAGE and enzymatically digested into peptides, which are then separated by High-performance Liquid Chromatography (HPLC). Peptides are then ionized and analysed by tandem mass spectrometry, generating MS and MS/MS spectra, respectively. MS/MS spectra are searched against protein databases to obtain the protein identification. *Adapted from www.biochem.mpg.de*

A mass spectrometer consists of three main parts:

1. The ion source that converts the peptides into gas-phase ions.

2. The mass analyzer, which separates the ions according to their m/z.

3. The detector that records the number of ions at each m/z value.

1. Ion Source and ionization of peptides

Mass spectrometers can measure m/z values of ionized molecules in gas-phase. However, since proteins and peptides are non-volatile and polar molecules, they require an ionization method that transfers them into the gas-phase, for the subsequent analysis. One of the most important innovations in this type of instruments was the introduction of "soft-ionization" technologies, which allows the ionization of proteins and peptides without causing extensive degradation. Two techniques are commonly used: matrix-assisted laser desorption ionization (MALDI) [87, 88] and electrospray ionization (ESI) [89]. 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 such as α-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. 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 [90].

Unlike MALDI, ESI source produces ions from a solution. Peptides exist as ions in solution because they contain functional groups whose ionization is controlled by the pH of the solution. At acidic pH values, protonation of the amines confers overall positive charge to peptides, while at basic pH, de-protonation of the amines and carboxyl groups confers an overall negative charge. Fragmentation of peptide ions is favoured by positive charges on the peptide ions. For these reasons, ESI of peptides is most commonly done in the positive ion mode to analyse acidic samples. Briefly, the ESI process consists of the formation of an electrically charged spray driven by the application of a high voltage (2–6

kV) between the end of the LC-column and the inlet of the mass spectrometer. The electrically charged spray drives the desolvation of peptide-solvent droplets. The high temperature provided by a heated capillary and the sheath gas flow at the mass spectrometer inlet help this process **(Fig. 4)**. ESI sources are typically combined with reverse phase high-pressure liquid chromatography (RP-HPLC), which allows the separation of very complex peptide mixtures prior to MS analysis. An important implementation in ESI includes the micro and nano-ESI sources where the flow rates of the analytes are lowered to micro and nanoliter/min, respectively, improving the sensitivity [91, 92].
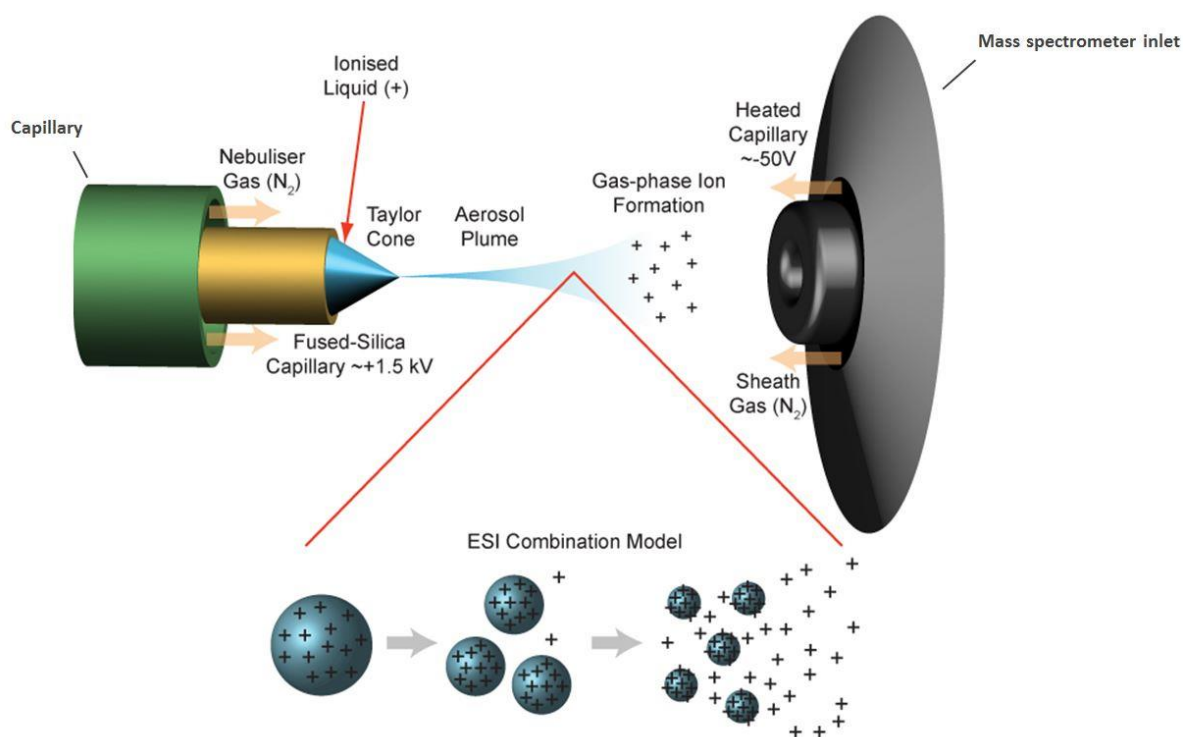
Fig. 4 Scheme of the electrospray ionization (ESI) process
Peptides eluted from the chromatographic column are ionized by a high voltage applied between the capillary and the mass spectrometer. Charged liquid forms a cone shape (known as the Taylor cone) and the analyte-solvent droplets burst away into a spray. *Adapted from http://www.lamondlab.com.*

2. Mass analyzer and dissociation methods

The mass analyzer is the core of the mass spectrometer and its key function is the storage and separation of ions based on their m/z. Common mass analyzers for peptide

analysis from complex mixture are: the linear ion trap (LIT), the Orbitrap, the FT-ICR, the quadrupole (Q) and the above-mentioned TOF. 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 producing the MS/MS spectrum. This enables extrapolating the primary sequence of the peptide of interest and inferring the presence and position of post translational modifications.

Various fragmentation techniques can be employed. In the collision-induced dissociation (CID) and higher energy collision dissociation (HCD) methods gas-phase protonated peptides/proteins are subjected to multiple collisions with rare gas atoms, resulting in the breakage of the peptide backbone at -CO-NH- bonds [93] and to the formation of the characteristic b- and y-ions (at N- and C-terminus, respectively) which are used for the univocal peptide identification. This type of fragmentation increases the frequency of relevant modification-specific fragmentation events in high resolution MS/MS spectra and is therefore suitable for the analysis of PTMs in bottom-up experiments [94]. These methods generate limited information for peptides longer than 15 amino acids, therefore Electron capture dissociation (ECD) and electron transfer dissociation (ETD) can be used. These approaches induce fragmentation of the peptide backbone based on gas-phase reactions using either thermal electrons or formation of radical ions, respectively, and are particularly suitable for sequencing longer peptides or whole proteins [95].

In a typical shotgun proteomics workflow, peptides are analysed in a data-dependent acquisition mode (DDA), in which the mass analyzer selects the most intense ions in a certain time window and subject them to fragmentation. Then the analyzer automatically switches back to the full scan MS mode and selects the next most intense ions to be fragmented. Hence, DDA is biased towards the most abundant species.

3. Detector

The detector is located at the end of mass spectrometer and records the number of ions at each m/z value. Most often, the detectors are electron multipliers or micro-channel plates, which emit a cascade of electrons when each ion hits the detector plate [96]. This cascade results in the amplification of each ion hit, to improve sensitivity. This process is

performed under vacuum to remove gas molecules and contaminating non-sample ions, which can collide with sample ions and produce non-specific reaction products.

2.4.1 Mass spectrometry analysis in Orbitrap instruments

In the last years mass spectrometers have continuously improved in terms of sensitivity, speed, mass accuracy and resolution. In particular, Orbitrap mass analyzer advancements have made the mass spectrometers based on this tool the instruments of choice for shotgun proteomics. The Orbitrap mass analyzer consists of a coaxial central spindle electrode surrounded by a barrel-like electrode [97]. The Orbitrap traps ions in an electric field and causes them to orbit and oscillate around a central electrode in ring shapes. The oscillation frequency along the electrode depends on the m/z. The detector picks up the axial motion of the ions and this measure is Fourier-transformed to yield high-resolution mass spectra [98].

The experiments reported in this thesis were performed on a Q Exactive HF Orbitrap MS operating in DDA mode (see previous paragraph) **(Fig. 5)**. Briefly, the ions generated in the ESI source are injected into the mass spectrometer and enter the S-lens. The S-lens focuses the ion beam before they pass through the Advanced Active Beam Guide (AABG). Here, the path is bent of 90° resulting in the removal of many uncharged/contaminants before entering the Quadrupole. In the Quadrupole the most abundant ions are selected and let into the C-trap to be then guided to the Orbitrap where MS1 are acquired. While they are transferred to the HCD cell to be fragmented, the following package of ions is already entered in to the C-trap, to start a new MS-MS/MS cycle. The fragment ions generated in to HCD cell are sent back into the Orbitrap analyzer where also the MS/MS spectra are detected at high resolution.
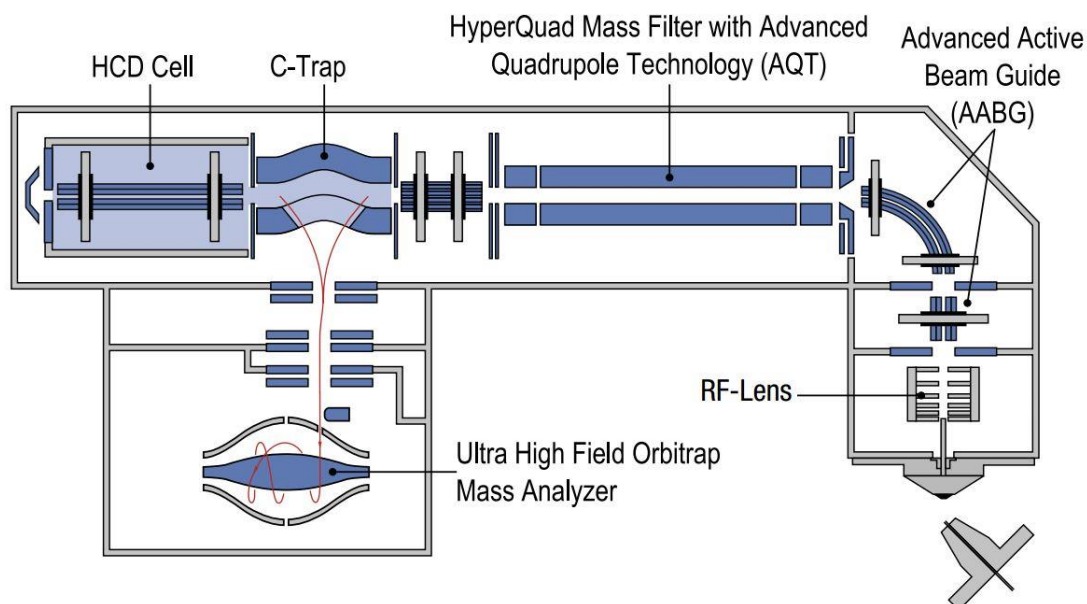
Fig. 5 Schematic description of a Q Exactive HF mass spectrometer
(*Adapted from www.thermofisher.com)*

2.4.2 Protein identification

After acquisition, data are interpreted by mapping the MS/MS spectra back to the peptide that was subjected to fragmentation. Several algorithms have been developed to analyze MS/MS spectra, but they all follow a similar rationale. Briefly, every MS/MS ion spectrum is searched against a database containing all the theoretical fragmentation patterns constructed for the peptides of the searched database. User-defined criteria, like the mass tolerance, the proteolytic enzyme used and the presence of fixed and variable modifications, restrict the pool of candidate peptides. The output of the search is a list of fragment ion spectra matched to the peptide sequence, ranked according to the search score. The score measures the similarity between the experimental spectrum and the theoretical one, and allows discriminating true and false identifications. To estimate the false discovery rate (FDR) one of the most common approaches is the target-decoy search [99], in which MS/MS spectra are searched against a target database of protein sequences augmented with the reversed sequences of the same database. Once established an FDR threshold, this is used to filter the data and remove false positive peptide identifications. Different analysis programs are available, such as Andromeda within the MaxQuant suite [100], MASCOT [101] and SEQUEST [102].

2.5 Quantitative proteomics

In the last few years, following the diffusion of mass spectrometry-based techniques, it has become evident that the qualitative identification of proteins is not sufficient for comparative analyses of functional states, and that quantitative data are much more informative to characterize a certain model system. However, mass spectrometry is not a quantitative technique *per se,* as a consequence of different facts. First, the ion intensity is proportional to its amount but it is also dependent on the chemical-physical properties of the corresponding peptide, such as the molecular composition, the charge and the hydrophobicity. Second, the comparative analysis between two different LC-MS/MS runs is influenced by external variations like the temperature and the chromatography reproducibility. Third, in data dependent acquisition (DDA) methods the choice of the precursor ion to be fragmented is stochastic and depends mostly on its abundance, thus limiting the possibility of profiling a sample completely and restraining the robust detection of low abundance peptides.

In order to overcome these limitations, different strategies have been developed and can be divided into two groups: label-free quantitation (LFQ) strategies, that are employed to compare the amount of proteins in different LC-MS runs without requiring particular sample handlings, and strategies based on isotope labeling, which require a specific preparation of the sample prior to LC-MS analysis **(Fig. 6)**.

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 strategies, 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.

In the former approaches, the ion chromatograms for every peptide are extracted from a LC-MS/MS run and their mass spectrometric peak areas are integrated over the chromatographic time scale (Extracted ion chromatogram, XIC or area under the curve,

AUC). These strategies rely on the concept that the XIC 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. Thus, 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 are based on the rationale that the abundance of a protein correlates linearly with the number of MS/MS spectra produced for its peptides. Thus, relative quantitation is achieved by comparing the number of such spectra between a set of experiments [103, 104]. Limits of spectral counting approaches are that physical and chemical properties of peptides influence this linear correlation and that accurate protein quantification requires the acquisition of a high number of spectra.

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.). However, LFQ requires highest technical and experimental reproducibility and leads to an overall lower accuracy in protein quantification in respect to isotope-based techniques.

Isotope-base quantitation methods are all based on the same rationale: creating a mass shift that distinguishes peptides deriving from different samples within a single MS analysis. Generally, samples from different biological conditions are isotopically light- (L) or heavy- (H) labeled and then mixed in equal amounts. The signals deriving from the two samples can be distinguished due to a known mass shift between L and H peptides. Thus, the H/L intensity ratio allows the accurate quantification of the relative amounts of peptides and proteins originally present in the two samples. The isotope can be introduced in the peptides at different stages during the sample processing, depending on the labeling approach **(Fig. 6)**, but the general rule is that the earlier this step is performed, the less are the variations introduced between the samples, thus the more accurate the quantification is. Depending on how proteins or peptides are labeled, isotope-based approaches can be divided into metabolic- and chemical- labeling strategies.

In the metabolic labeling, the isotope is added to growing cells as a metabolic precursor in order to be incorporated in the proteome during protein biosynthesis. These approaches are advantageous because they can be applied to *in vivo* studies and the samples mixing is carried out at the beginning of the workflow, thus minimizing the introduction of processing errors. For this reason, metabolic labeling approaches are compatible with complex purification procedures and allow accurate protein quantitation.

In chemical-labeling strategies, the tag is added covalently to the reactive side chains of amino acids either before or after the proteolytic cleavage through a chemical reaction. The most common methods for chemical labelling are: isotope coded affinity tagging (ICAT)[105] and isobaric tags for relative and absolute quantitation (iTRAQ) [106]. In ICAT, proteins are labeled with a chemical reagent that consists of a thio-reactive group that labels cysteines, an isotopic linker region displaying a known mass difference, and an affinity biotin tag. However, the main limit of this approach is that only cysteine-containing proteins can be labeled and quantified at MS level. In iTRAQ, 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 of multiplexing, by using up to 8 different isobaric tags. However, the efficiency of the labeling is variable and depends on sample complexity. Moreover, this method allows less accurate protein quantification due to the rather late step of labeling in the sample processing workflow.
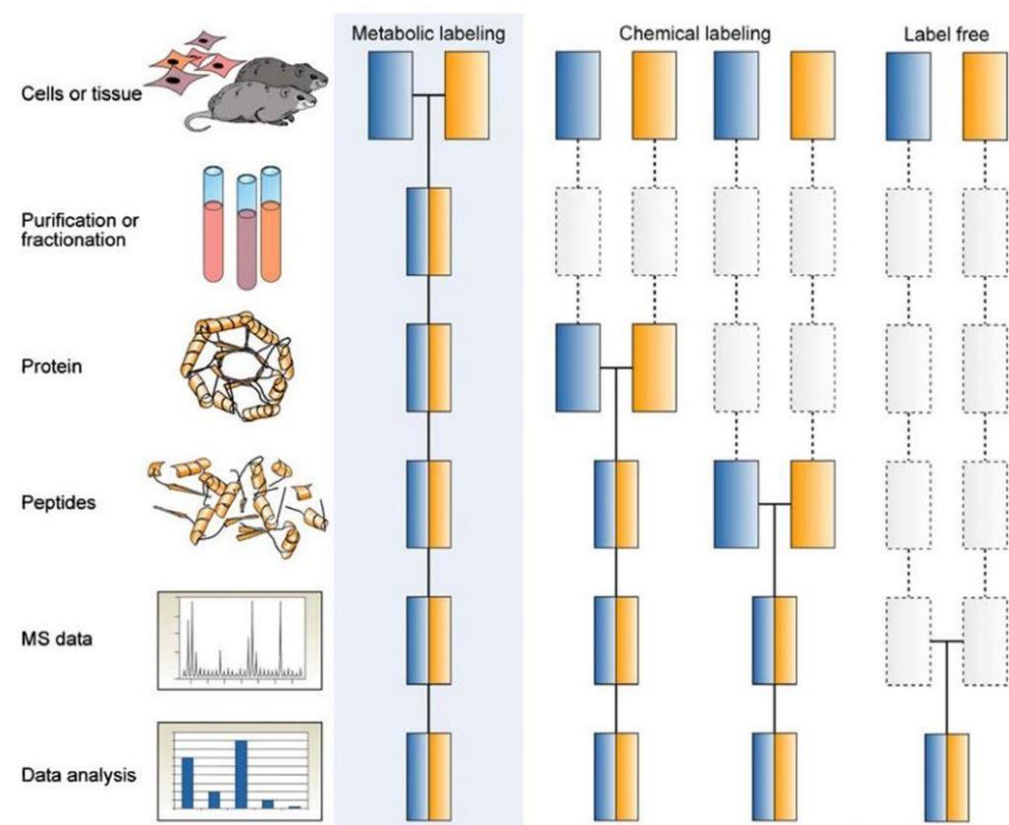
Fig. 6 Common quantitative mass spectrometry workflows
Boxes in blue and yellow represent two experimental conditions. Horizontal lines indicate when samples are combined. Dashed lines indicate points in which experimental variation and thus quantification error can occur. *Adapted from [103].*

2.5.1 Stable Isotope labeling by amino acids in cell culture (SILAC)

The most successful example of metabolic labeling strategies is the Stable isotope labeling by amino acids in cell culture (SILAC), introduced by Mann and co-workers in 2002 [107]. SILAC is a powerful and accurate procedure based on the use of essential amino acids where the naturally isotopes ^{12}C, ^{1}H and ^{14}N (or light, L) are substituted with heavy (or H) isotopes (^{13}C, ^{2}H and ^{15}N, respectively). In a classical SILAC experiment, two cell populations, representative of two different functional states, are grown in media containing either light or heavy isotope-coded essential amino acids for a sufficient number of cell doublings in order to ensure the full incorporation of the isotopes in the newly synthesized proteins. After complete labeling, cells grown in the two conditions will have two identical proteomes that only differ in protein isotope composition. Upon MS analysis, every peptide will be represented as a pair where the Δ mass is dependent on

both the number and the type of amino acids incorporated in the sequence [107] **(Fig. 7)**. In a standard SILAC experiment cells are usually labeled with H and L arginine and lysine, which in combination with trypsin digestion, ensure that all peptides of a protein (except the C-terminal one) carry one labeled amino acid.
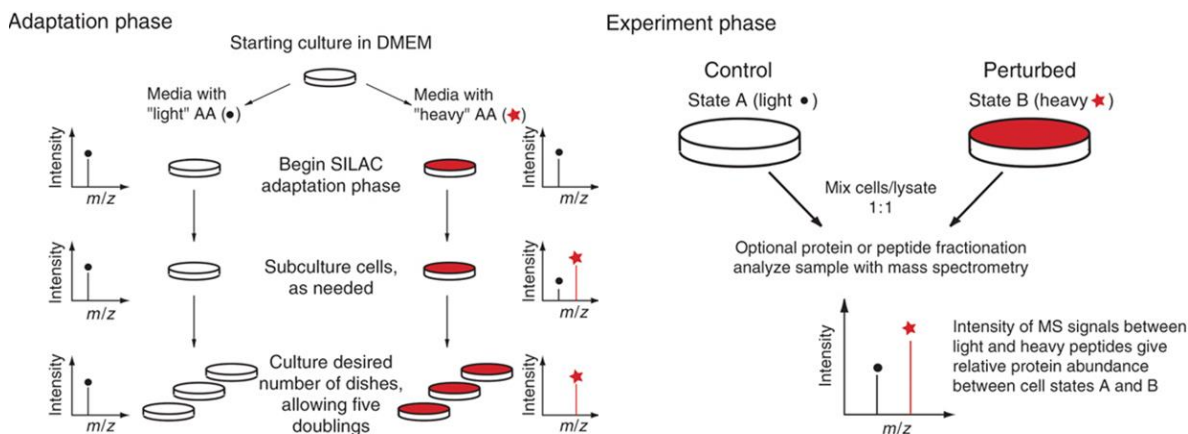
Fig. 7 Scheme of a standard SILAC experiment
SILAC experiment is composed of two different phases: in the adaptation phase (left panel), cells are grown in medium containing either light- or heavy-isotope coded amino acids until full incorporation (red asterisk). In the experimental phase (right panel) the two population are mixed, proteins are purified, digested and analysed by MS. *Adapted from [108].*

The SILAC labeling has been successfully applied to a wide variety of studies: from protein expression profiling to global PTM analysis, protein-protein and DNA-protein interaction analysis [109-112].

In the last years, the SILAC approach has been implemented and applied to a variety of proteomics workflows, as reviewed in [113]. It is now possible to quantify three and up to five cell states in a single run [114, 115] or to quantitatively profile protein translation dynamics with the pulse SILAC (pSILAC) [116]. Another development of SILAC is the super-SILAC, which allow quantifying proteins from animal or human tissues [117]. SILAC labeling can also be used to directly label PTMs, as in the Heavy Methyl SILAC (hmSILAC) approach, developed in the group of Mann in 2004 [118] (see paragraph 2.6.3).

2.6 Post translational modification analysis through MS-based proteomics

The identification of a post translational modification (PTM) by MS relies on the detection of a mass difference (Δ mass) introduced by the modification compared to the mass of the unmodified peptide. Generally, the workflow used for protein identification is also applicable to PTM studies, although this analysis is more complex due to three main features of modifications. First, they are substoichiometric and therefore their analysis requires enrichment steps and high sensitivity of detection. Second, the covalent bond between the PTM and the amino acid side chain of the peptide can be labile (as for phosphorylation), thus it is difficult to maintain the peptide in its modified form during sample preparation and ionization. Third, some PTMs are highly dynamic and thus the enzymes responsible for removing the PTM of interest need to be inactivated.

Given the sub-stoichiometric nature of PTMs, several strategies have been developed to enrich the modified peptides compared to the unmodified proteins or peptides. Examples of these approaches are immunopurification (IP), immobilized metal affinity chromatography (IMAC), titanium dioxide (TiO_2) metal-based chromatography, ion exchange (strong cation/anion exchange, SCX/SAX) chromatography and hydrophilic interaction liquid chromatography (HILIC) [119]. Immunopurification relies on the use of antibodies able to specifically recognize a certain modification. SCX and SAX are separation strategies orthogonal to the reversed phase and are used to decrease the complexity of highly complex samples [120]. IMAC, TiO_2 and HILIC are widely used in phosphoproteomics studies and are based on the principle that Fe^{3+} or TiO_2^+ chelate the phosphate group of phosphopeptides [121-123]. However, efficient enrichment strategies for many PTMs are still lacking [119]. Moreover, the starting amount of material required for modified peptides enrichment is sensibly larger than the one typically needed for proteome analysis.

In MS-based PTMs analysis, it is essential to generate enough peptide fragmentation information (MS/MS) for the high-confidence sequence identification and site localization of the modifications. Recently, the development of new instrumentation provided both high resolution and high mass measurement accuracies simultaneously at the MS and MS/MS

level, also thanks to the implementation of various fragmentation strategies (see paragraph 2.4). This has increased the reliability and efficiency of PTM identification and site-assignment [124]. In addition, to overcome the problems related with ambiguous PTM assignments, a number of algorithms, such as the PTM score [125], have been developed to statistically assess the location of a PTM on a peptide. However, despite their successful application in the context of phosphorylation analyses, such algorithms have not been optimized for the localization of other PTMs [126].

A final issue related to PTM identification by MS is the challenge in the database search, which implies the search of all potential combinations of mass shifts produced by the modifications for each identified peptide from a protein database, thus expanding exponentially the search space, with effects on site-assignment confidence and FDRs.

2.6.1 Classical technical approaches to study protein methylation

PRMTs are up-regulated in various types of cancer [127] and the consequent aberrant methylation of critical effectors and pathways involved in the development and progression of cancer has been described [64]. Nowadays it is well accepted that protein methylation plays a fundamental role in cellular biology, both in physiological and pathological conditions. Hence, the study of methylation and its targets has gained interested and a wide range of analytical approaches has been developed.

Classical approaches include *in vitro* radioactive labeling experiment using tritiated SAM as a methyl-donor. However, the limit of this approach is that it does not allow the specific assignment of the methylated sites within a protein. To address this issue, several antibodies that recognize modified residues and the different degree of modification have been developed, but the small size of the methyl-groups makes the production of such specific antibodies highly challenging.

However, these methods can be used only with single proteins, are often unspecific and require prior knowledge of the modifications [128].

Recently, mass spectrometry-based approaches have emerged as an extremely powerful analytical strategy for global investigation of protein PTMs, including methylation, due to

the intrinsic possibility of pinpointing different marks on distinct residues and of assessing quantitatively their abundances [129].

2.6.2 Challenges in the MS-based analysis of protein methylation

Although MS-based approaches have proven to be the most powerful strategy available for the global analysis of protein methylation, this approach remains very challenging, due to a number of issues.

1. As discussed in the previous paragraph, post translationally modified peptides are sub-stoichiometric and have to be enriched prior to MS analysis. However, in the case of methylated proteins/peptides the development of highly efficient pan-methyl antibodies has been difficult, due to the structural heterogeneity introduced by the addition of multiple methyl-groups [130]. Moreover, pan-methyl antibodies recognizing methylated motifs are likely not totally "unbiased" as they may recognize a subset of substrates in a context-specific manner [131]. Only recently, efficient antibodies against methylated arginine and lysine have been produced and successfully applied to study protein methylation [132, 133].

2. Several amino acids substitutions are isobaric to methylation. In fact, some amino acids have a chemical structure that differs by a mass difference of 14 Da and are therefore difficult to distinguish from mono-methylation. For instance, a peptide bearing glycine and methylated arginine would be identical to the same peptide with a glycine-to-alanine substitution [118].

3. *In vitro* artifacts introduced during sample preparation can be wrongly assigned as methylations. For instance, methanol induces the formation of mono-methylation on glutamic and aspartic acid residues, leading to the incorrect identification of arginine methylation sites during database searches [134, 135].

4. The large number and type of mass modifications resulting from the different degrees of this PTM (mono-, di- and tri-methylated lysine and mono- and di-methylated arginine) may lead to false positive assignments, due to the combinatorial increase of possible peptide matches during the database searching [118].

For all these reasons, the identification of methylated sites by mass spectrometry can produce a very high number of false positives. When studying PTMs the standard method to control false positive peptide identifications is the application of the target-decoy approach (see paragraph 2.4.2), which allow to estimate FDRs. Unlike for other modification types, recent works from the groups of Acuto and Wilkins showed that the use of this strategy for methyl-peptides identification is not efficient and need to be tailored for this specific PTM [44, 136].

2.6.3 Heavy methyl SILAC for the high-confidence identification of protein methylation

Heavy methyl SILAC (hmSILAC) is the first quantitative labeling approach that exploits metabolic labeling to investigate protein methylation. It is a variant of the classical SILAC where cells are grown in medium supplemented with either normal (light) or 'heavy' methionine. Heavy methionine contains a methyl-group with one carbon-13 and three deuterium atoms. Cells metabolically convert the $[^{13}C_3]$-methionine into $[^{13}C_3]$-AdoMet, which is the sole methyl donor in the cell. The heavy methionine is therefore incorporated into the backbone of newly synthesized proteins and, in addition, the heavy methyl-group is transferred by KMTs and PRMTs from SAM to their substrates, resulting in the heavy labeling of all methylations on proteins **(Fig. 8A)**. The workflow of an hmSILAC experiment is identical to a conventional SILAC experiment, allowing the identification and quantification of methylated peptides. For each methyl-group added, hmSILAC increases the mass of modified peptides by 4 Da in the case of mono-methylation, 8 Da for di-methylation and 12 Da for tri-methylation **(Fig. 8B)**.
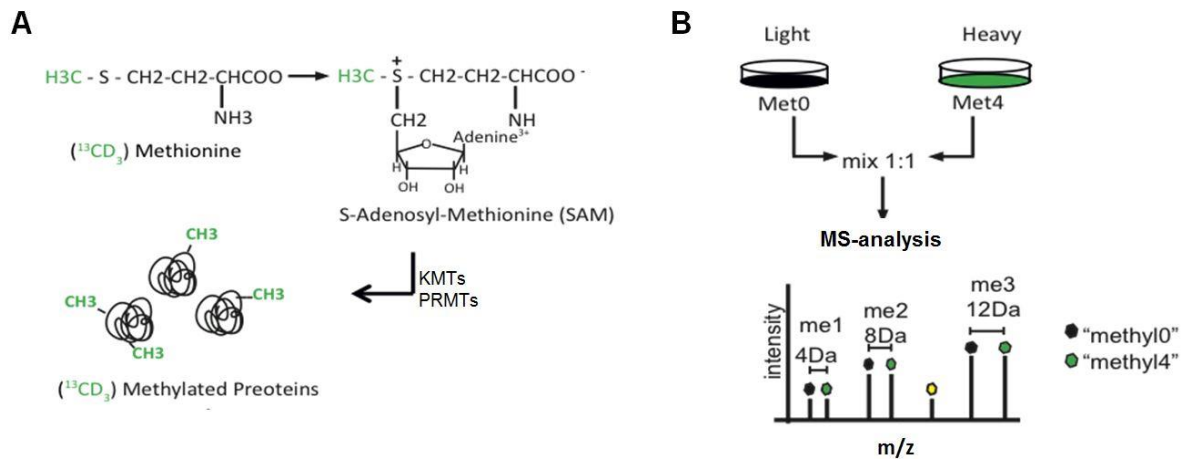
Fig. 8 Schematic description of the heavy methyl SILAC workflow
A) [$^{13}C_3$]-methionine (heavy) is metabolically converted into heavy S-adenosyl-methionine (SAM) in the cell. Through enzymatic reactions by KMTs and PRMTs, labeled methyl-groups are transferred onto methylated proteins. **B)** Upon mixing of the light (Met0) and heavy (Met4) labeled cells in equal amount (1:1) and MS analysis, the type of methylated residue can be identified on the basis of the mass separation of the peptide pair (4 Da = mono-methylation, 8 Da = di-methylation, 12 Da = tri-methylation). *Adapted from* [118].

Heavy methyl SILAC is the method of choice to study protein methylation because the tag is encoded in the methyl-groups itself, producing a modification-specific mass separation between light and heavy peptides [137]. Indeed, this increases the confidence in the MS-based identification of methyl-peptides, because the presence of the heavy and light pairs is used to distinguish *in vivo* methylation from potential false positives that are frequent in methylation analyses (as described in paragraph 2.6.2). Recent studies showed that the direct labeling of the PTM helps to reduce the methyl-peptides FDR [136] and strongly support the choice of hmSILAC to study protein methylation.

An issue of methylation analyses by MS is the multiplicity of methylation states. In addition, the number of modifications increases significantly with the hmSILAC, since the heavy version of all possible R and K methylation states is introduced in the search, together with the light counterpart, as variable modification. This substantially complicates the analysis of the data. At present, automated software for the analysis of hmSILAC data is not available. Software developed for the analysis of standard SILAC data, such as MaxQuant, can be adapted to search for heavy-methylation whereby light and heavy

labeled modifications are searched separately. However MaxQuant cannot be tailored to identify and match hmSILAC peptide pairs. For this reason, MaxQuant output files are integrated post-search using additional in house developed scripts to match corresponding light and heavy methylated peptides. Our group, as well as others, have contributed to the development of such analytical tools [1, 136], essential to handle large scale hmSILAC analysis of protein methylation.

Another potential drawback of hmSILAC is that it does not allow discriminating between methylated and methionine-containing peptides, since the heavy methionine is also incorporated into the backbone of the peptides. However, labeling strategies to bypass this issue are available, such as the isomethionine methyl SILAC (iMethyl-SILAC) [44]. In this approach the light methionine is substituted with the isomethionine (methionine-$^{13}C_4$). The two stable isotope-labeled methionines are nearly isobaric but differ in the distribution of the additional mass. The methyl-groups transferred during methylation are still light or heavy, but methionine incorporated during protein synthesis are nearly isobaric and as a result, i-methyl SILAC doublets will only arise from methylated peptides [44].

2.6.4 MS-based discrimination between symmetric and asymmetric arginine di-methylation

Symmetric and asymmetric di-methylation of arginine residues are isomeric modifications that have proven to have distinct biological effects [61]. Moreover, the evidence of the increase of symmetric di-methylation after loss of PRMT1 [138] or upon inhibition of type- I PRMTs [139] suggest the possibility that type-II enzymes may compensate or scavenge the loss of asymmetric di-methylation. This evidence highlights the importance of discriminating between the two types of di-methylated arginine residues. Although the majority of studies used selective antibodies capable of distinguishing between SDMA and ADMA, MS can also be employed for this purpose. Several studies indicate that, upon MS/MS fragmentation, differentially methylated peptides generate "diagnostic ions" that can be used to identify the distinct type of dimethyl-arginine decorating the peptide. In particular, the presence of the characteristic methyl-guanidine group ion was used to

increase the confidence of methyl-site assignment, in specific dissociation conditions and in a hmSILAC setup [118, 140, 141]. However, the systematic use of diagnostic fragment ions in global methylation analyses has not yet been explored systematically.

2.7 State-of-the-art on global methyl-proteomics studies

In the last decade several groups exploited MS-based approaches to perform global analysis of protein methylation with the aim of characterizing the cellular methyl-proteome. Different strategies of biochemical enrichment and fractionation methods prior to MS analysis have been used to increase the identification of methylated sites.

In 2003, Boisvert and colleagues conducted one of the first global analyses of methylation combining MS with antibody-based enrichment of asymmetrically and symmetrically R-methylated proteins. The authors reported over 200 proteins containing putative methylations but they were not able to specifically assign the methylation site [142]. Following this work, Ong and Mann introduced for the first time the hmSILAC strategy, which was used in combination with antibody-based enrichment to identify 59 high-confidence *in vivo* methylated sites [118]. Subsequent works from several groups, including ours, successfully applied the hmSILAC in global methylation studies. Heavy methyl SILAC was used in combination with different types of biochemical techniques to enrich methylated peptides based on their chemical properties, prior to the identification by LC-MS/MS [1, 143]. SCX, HILIC and isoelectric focusing were used, as well as SDS-page protein separation and antibody-based enrichments of methylated proteins, leading to the identification of up to 400 high-confidence *in vivo* methyl-sites.

In the last three years, the introduction of specific and efficient antibodies against mono- and asymmetrically di-methylated arginine set the basis for the development of protocols for the enrichment of methylated peptides instead of proteins, which have been successfully applied in combination with biochemical fractionation methods and MS analysis [132, 144]. These studies contributed to enlarge the knowledge on the methyl-proteome, with the identification of thousands of novel methyl-sites. The application of this new and successful experimental strategy, combined with SILAC, allows also studying R-

methylation dynamics following different stimulus. The group of Nielsen recently quantified methylation changes during transcriptional arrest [144] and after the depletion/inhibition of specific PRMTs [34], while Geoghegan and colleagues demonstrated that T cell activation induces global changes in R-methylation [44]. These works confirmed on a large scale that methylation is a dynamic PTM such as other better-characterized post translational modifications.

However, the majority of these studies, with the exception of the one from Geoghegan and co-workers, do not rely on strategies for the high-confidence validation of *in vivo* methylated sites, such as the hmSILAC and i-Methyl SILAC. This is a sever restrain, since it has recently demonstrated very clearly that the identification of methylated peptides with label-free approaches implicates quite high FDR in methyl-assignment. Indeed, minimizing the FDR is essential when detecting methyl-sites in global studies, because incorrect assignments can severely mislead follow-up functional analyses [44].

3. AIM OF THE PROJECT

Although microRNA biogenesis is a tightly controlled process, its deregulation, generally caused by alteration in the proteins of the pathway, has been observed in different pathological conditions, including cancer, and leads to aberrant miRNA expression [4, 130]. Hence, dissecting the molecular mechanisms underlying miRNA biogenesis is crucial to understand how this process is altered in pathological conditions, with the final aim of developing possible therapeutic strategies to revert aberrant miRNA levels.

During miRNA biogenesis, the Large Drosha Complex (LDC) catalyzes the cleavage of the primary-miRNAs to precursor-miRNAs, a rate-limiting step for the production of mature miRNA molecules, occurring in the nucleus. The enzymatic activity of the complex is finely tuned through multiple regulatory mechanisms, including post translational modifications (PTM) of its subunits. Our finding that protein methylation frequently occurs on most of the proteins of the complex [1] suggested for the first time that this PTM could have a role in regulating the complex, and hence in miRNA maturation.

The aim of my project is to investigate the role of protein methylation in the regulation of the LDC structure and function. To this aim, I have first characterized the methylated sites occurring on the different subunits of the complex and created a library of high-confident LDC methylations. I then focused on PRMT1, the major type I Protein Arginine Methyltransferase in mammalian cells, assessing the impact of its modulation (both depletion and induction) on the methylation state of LDC. The discovery of novel PRMT1 targets within the complex, achieved through a systematic and functional MS-based analysis, will help shedding light on the mechanisms through which protein R-methylation can regulate LDC structure and function, and ultimately impact on miRNA production.

4. MATERIALS AND METHODS

4.1 Cell culture

HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS, Invitrogen 10270-106), 1% glutamine, 100 U/ml penicillin, 100 mg/ml streptomycin and cultured at 37°C in a 5% CO_2 humidified atmosphere.

4.1.1 Heavy methyl SILAC labeling of cells

HeLa S3 cells were cultured in "Light" and "Heavy" SILAC media (PAA, custom) supplemented with L-Arginine (Sigma Aldrich, A6969) L-Lysine (Sigma Aldrich, L8662), and either L-[^{13}CD3]-Methionine (Met-4, heavy, Sigma Aldrich, 299154) or L-[^{12}CH3]-Methionine (Met-0, light, Sigma Aldrich M5308), respectively. Concentrations of amino acids were 84 mg/L, 146 mg/L and 30 mg/L, for L-Arginine, L-Lysine and L-Methionine, respectively. The SILAC medium was then supplemented with 10% dialyzed FBS (26400-044 Gibco, Life Technology), 1% glutamine, 100 U/ml Penicillin and 100 mg/ml Streptomycin and 10 mM HEPES pH 7.5.

Cells were grown in hmSILAC media for at least 9 replication cycles to ensure full incorporation, with a careful monitoring of growth rate, viability and overall morphology, which ensured that normal physiology, was preserved.

4.1.2 SILAC labeling of cells

HeLa cells were grown in "Light" and "Heavy" SILAC DMEM (Thermo Fisher Scientific 88420), supplemented with either normal L-Arginine and L-Lysine, or their heavy isotope-counterparts L-Arginine (Arg10, Sigma 608033) and L-Lysine (Lys 8, Sigma 608041). Arginine and Lysine were added at a concentration of 84 mg/L and 146 mg/L, respectively. The SILAC medium was then supplemented with 10% dialyzed FBS (26400-044 Gibco, Life Technology), 1% glutamine, 100 U/ml Penicillin and 100 mg/ml Streptomycin and 10 mM HEPES pH 7.5.

Cells were grown in SILAC media for at least 9 replication cycles to ensure full incorporation, with a careful monitoring of growth rate, viability and overall morphology.

4.2 Cell lysis and sub-cellular fractionation

For the preparation of whole cell extracts cells were collected, washed twice with cold PBS 1X and resuspended in 3 volumes of hot SDS Lysis buffer (0.1M Tris HCl pH 7.5, SDS 4%). Cells were sonicated 10 sec and then centrifuged 15 min at 13000 rpm. SDS total lysates were quantified using BCA (Pierce BCA Protein Assay kit, 23225).

For sub-cellular fractionation cells were harvested, washed twice with cold PBS 1X and re-suspended in 2 volumes of Lysis buffer A (10mM Hepes KOH pH 7.9, 1.5mM MgCl2, 10mM KCl, 0.2% NP40, 1X Roche Protease Inhibitors, 1U/μl NEB RNAse Inhibitors). After 20 strokes with a dounce homogenizer, cells were centrifuged 15 min at 3750 rpm. Supernatant (cytoplasmic extract) was collected and pellet (corresponding to crude nuclei) was washed twice with PBS 1X and re-suspended into 2 volumes of buffer C (420mM NaCl, 0.1% NP40, 0.5mM DTT, 20mM Hepes KOH pH 7.9, 2mM MgCl2, 0.2mM EDTA and 20% glycerol, 1X Roche Protease Inhibitors, 1U/μl NEB RNAse Inhibitors). The suspension was rotated 1h at 4°C and then ultracentrifuged at 33000 rpm for 1h.

For subsequent small RNA analysis, the chromatin pellet was washed with 10 volumes of buffer C, resuspended in Ripa buffer (10mM Tris HCl pH 8, 150mM NaCl, 0.1% SDS, 1% Triton, 1mM EDTA, 0.1% Na Deoxycholate) and centrifuged 5 min at 13000 rpm at 4°C. Protein extract were quantified with Bradford assay.

4.3 Protein co-immunoprecipitation (co-IP)

Nuclear protein extract was quantified with Bradford assay and 1-2 mg of extract was diluted with Ripa buffer lacking NaCl to decrease salt concentration to 150 mM NaCl. The extract was pre-cleared 3 times with 50 μl of Dynabeads-G beads, previously blocked with 0.5% BSA. After overnight IP, already blocked and washed beads were added to the extract and then incubated for 2.5 hours on a rotating wheel. Beads were then washed 4

times with Ripa buffer and then incubated with LSD Sample buffer supplemented with 100 mM DTT in order to elute the immunoprecipitated proteins. Proteins were resolved on 4-12% Bis-Tris acrylamide SDS-PAGE pre-cast gels (Novex Tris-Glycine Gels, Life technologies) and visualized on the gel using Colloidal Coomassie staining (Instant Blue, Expedeon), while one tenth was loaded on hand-made linear gel for western blot validation. The following antibodies were used for IP according to the manufacturer's–instruction: DGCR8 (Abcam ab90579, 4 μg/mg lysate), FUS (Bethyl A300-293A, 2 μg /ml lysate), DDX5 (Abcam ab126730, 1/50), Drosha (Santa Cruz sc-33778, 4 μg /ml lysate).

4.4 In-gel digestion of immunoprecipitated proteins

Processing of gel-separated proteins prior MS analysis was carried out as previously described, with minor modifications [145]. Briefly, slices were cut from gels and de-stained in 50% v/v acetonitrile (ACN)/50 mM NH_4HCO_3. Reduction was carried out with 10 mM DTT, followed by alkylation with 55 mM iodoacetamide. After each step samples were dehydrated with 100% ethanol and quickly "dried" in a centrifugal evaporator (SpeedVac). Subsequently, gel pieces were washed extensively with 50 mM NH_4HCO_3 and digested with 12.5 ng/μl trypsin (Promega V5113) overnight at 37°C. Digested peptides were extracted with extraction buffer (3% TFA, 30% ACN) and 100% ACN. Prior to MS, peptides are desalted and concentrated in a single step through reversed phase chromatography on micro-column called C18 Stage Tips [146]. Here, the separation depends on the hydrophobic interaction between the solute in the mobile phase and the immobilized stationary phase, which consist of hydrophobic C18 resin. The resin is composed of eighteen carbon atoms, and therefore is suitable for the binding and separation of small poly-peptides. Stage Tips are prepared by placing three disks of Empore™ C18 47 mm Extraction Disks (Model 2215, 3M) in an ordinary P200 pipette tip. Stage Tips were activated with 100% methanol (CH_3OH) and equilibrated twice with buffer A (0.1% TFA, 0.5% acetic acid). Tryptic digested samples were diluted in buffer A and loaded on the Stage Tips. Samples were then eluted from the tips with high organic

solvent (80% ACN, 0.5% acetic acid), lyophilized, re-suspended in 1% TFA and subjected to LC-MS/MS analysis.

4.5 Liquid Chromatography and Tandem Mass Spectrometry (LC-MS/MS)

Peptide mixtures were analyzed by online nano-flow liquid chromatography tandem mass spectrometry using an EASY-nLC™ 1000 (Thermo Fisher Scientic, Odense, Denmark) connected to a quadrupole/Orbitrap mass spectrometer (Q Exactive, Thermo Fisher Scientific) through a nanoelectrospray ion source. The nano-LC system was operated in one column set up with a 25 cm analytical column (75 µm inner diameter, 350 µm outer diameter) packed with C_{18} reversed-phase resin (ReproSil, Pur C18AQ 1.9 µm, Dr. Maisch, Germany) configuration. Solvent A was 0.1% formic acid (FA) and 5% ACN in ddH2O and solvent B was 80% ACN with 0.1% FA. Peptides were injected at a flow rate of 500 nl/min and separated with a gradient of 5–40% solvent B over 90 min, followed by a gradient of 40–60% for 10 min and 60–80% over 5 min at a flow rate of 250 nl/min. The Q Exactive was operated in the data-dependent mode (DDA). HCD-fragmentation method when acquiring MS/MS spectra consisted of an Orbitrap full MS scan followed by up to 10 MS/MS experiments (Top10) on the most abundant ions detected in the full MS scan. Mass spectrometer conditions for all experiments were as follows: full MS (AGC $3e^6$; resolution 70,000; m/z range 300-1650; maximum ion time 20 ms); MS/MS (AGC 17,500; maximum ion time 50 ms; isolation width 1.8 Da with a dynamic exclusion time of 20s). Singly charged ions and ions for which no charge state could be determined were excluded from the selection. Normalized collision energy was set to 28%; spray voltage was 2.2 kV; no sheath and auxiliary gas flow; heated capillary temperature was 275°C; S-lens RF level of 60%.

4.6 Assigning hmSILAC/SILAC peptide sequences using MaxQuant and data analysis

Acquired Raw data were analyzed using the integrated MaxQuant software v1.3.0.5 or v.1.5.2.8, using the Andromeda search engine [147] [100]. The January 2014 version of

the Uniprot sequence was used for peptide identification. Enzyme specificity was set to Trypsin/P, meaning that trypsin cleavage occurs also in the presence of proline after Lysine or Arginine residues. In MaxQuant, the estimated false discovery rate (FDR) of all peptide identifications was set to 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. Carbamidomethylation of cysteine was set as a fixed modification.

4.6.1 hmSILAC peptide assignment and data analysis

To assign hmSILAC peptide sequences, we defined new modifications in MaxQuant (v1.3.0.5) with the mass increment and residue specificities corresponding to the heavy versions of mono-methylated K/R, di-methylated K/R, and tri-methylated K. Additionally, we defined new modifications for heavy methionine and oxidized heavy methionine. To reduce the complexity of the search, and given the computational resources available, each experimental set of raw data was analyzed three times with three distinct sets of variable modifications, namely: (1) N-terminal Acetylation, Met4, Met4ox, oxidation, mono-methyl-K/R, mono-methyl4-K/R; (2) N-terminal Acetylation, Met4, Met4ox, oxidation, di-methyl-K/R, di-methyl4-K/R; (3) N-terminal Acetylation, Met4, Met4ox, oxidation, tri-methyl-K, tri-methyl4-K. Identification of high-confidence methylated sites was carried out as previously described [1]. Briefly, we developed a Perl analytical pipeline that filters and integrates the output tables from MQ, to identify hmSILAC doublets (pairs) of methylated peptides **(Fig. 9)**. The script pairs peptides containing the same MQ-assigned modification in both heavy and light form, applying the following criteria: (1) first the peptide pairs that were not present in the same RAW file or that were not assigned the same charge state were excluded; (2) the peptide sequences annotated as containing heavy and light forms of modifications simultaneously were excluded; (3) heavy and light peptides were required to have an Andromeda score > 62; (4) all remaining methyl-peptides were required to have the modification localized with a probability score > 0.75.
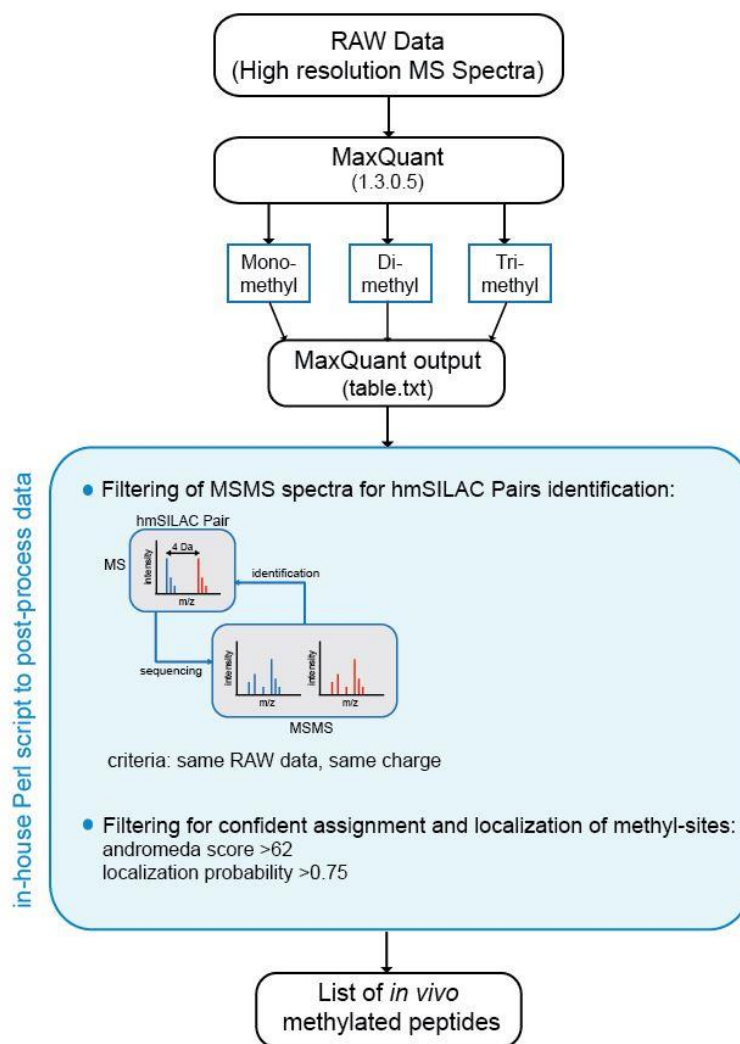
Fig. 9 Schematic description of hmSILAC data processing through an analysis pipeline developed in-house.

RAW data generated upon LC-MS/MS analysis, were processed using the MaxQuant software v1.3.0.5. Each experimental set of RAW data was analyzed three times in parallel, with distinct sets of variable modifications: (1) N-terminal Acetylation, Met4, Met4ox, oxidation, mono-methyl-K/R, mono-methyl4-K/R; (2) N-terminal Acetylation, Met4, Met4ox, oxidation, di-methyl-K/R, di-methyl4-K/R; (3) N-terminal Acetylation, Met4, Met4ox, oxidation, tri-methyl-K, tri-methyl4-K. The MaxQuant outputs were processed using an in-house Perl script (light blue boxes) that filters MS/MS spectra to identify hmSILAC pairs, following the criteria displayed in the picture. The Perl pipeline output consists in a .txt table listing high-confidence identified *in vivo* methylated peptides.

4.6.2 SILAC peptide assignment and data analysis

SILAC peptide and protein quantification was performed automatically with MaxQuant (v.1.5.2.8) using default settings for parameters. N-terminal acetylation of protein, methionine oxidation mono-methyl-K/R, di-methyl-K/R and tri-methyl-K were set as

variable modifications in MaxQuant. Outputs from MQ were manually filtered as fallow: proteins were accepted if identified with at least two peptides of which at least one unique; protein quantified were considered for further analysis only if they have ratio count (RC) equal or greater than 1.

A data analysis pipeline, written in Perl, was developed in-house to process MaxQuant output. In this pipeline, the evidence.txt file was first filtered: potential contaminants and reverse sequences were removed, and all accepted methyl-peptides were required to have all described methylations localised with a probability score >0.75. No Andromeda score cut-off was imposed. For quantitative analysis, methyl-peptides ratios were normalised on protein-level H/L information within each IP experiment and the median ratio of redundant methyl-peptides was calculated. To define significantly regulated methyl-peptides upon PRMT1 KD, for each IP experiment, the mixtools R-package (www.r-project.org) was used to define three Gaussian distribution related to three methyl-peptide populations: depleted, background and enriched. The mean and standard deviation (SD) of the background population were then used to define statistically significant- regulated peptides in each IP experiment.

Localization analyses of methylated peptides were performed using the Pfam database (http://pfam.xfam.org/), the Eukaryotic Linear Motif (ELM) resource (http://elm.eu.org/) or RCSB PDB (http://www.rcsb.org/pdb/home/home.do).

4.7 Cell transduction with lentiviral vectors

HeLa cells were transduced using second-generation lentiviral vectors, whose stocks were produced by the transient transfection of 293T cells with the packaging plasmid pCMV-DR8.74, the envelope plasmid pMD2G-VSVG and the respective transfer gene-carrying vector.

To knock down (KD) PRMTs, cells were infected with two different sh-RNAs targeting PRMT1, PRMT4, PRMT5, PRMT6 and PRMT7 or with plkO-Empty vector as a control (kindly provided by E. Guccione, IMCB Singapore). Sh-Scr and sh-Luciferase vectors were also tested as controls.

To overexpress PRMT1, the cDNA of the v2 isoform was amplified starting from cDNA of HeLa using the following primers:

FW –GGGGATCCCCCGGGCTGCAGATGGCGGCAGCCGAGGCCGCGAACTGCA-

REV –GAGGTTGATTGTCGACTCAGCGCATCCGGTAGTCGGTGG-

PRMT1 cDNA was cloned into pCCLsin.CPPT.PGK.GFP.WPRE lentiviral vector after plasmid linearization with PstI and SalI using the in-fusion HD EcoDry cloning plus according to the manufacturer's instructions (Clontech Laboratories, Inc., CA). The GFP cassette was removed from the vector after digestion with XhoI.

For virus preparation 293T packaging cells were plated (DMEM, 10% heat-inactivated FBS, 2mM glutamine, 100 U/ml penicillin and 100 mg/ml streptomycin) and co-transfected with 4 µg of pMD2G-VSVG, 6 µg of pCMV-DR8.74 and 10 µg of either the empty vector or the vector carrying the shRNA sequences. After 15 min 20 µM final of cloroquine was added to inhibit lysosomal degradation of exogenous DNA. After few hours the medium of 293T packaging cells was replaced with fresh medium. Supernatants from transfected 293T were collected at 48 hours post-transfection and used for HeLa infections. HeLa were infected at 50% of confluence and after 6-8 hours the virus-containing medium was gently removed and replaced by fresh medium. Puromycin selection (1 µg/ml) started 24 hours after the infection.

4.8 Western Blot analysis

Protein extracts from IP or cellular extracts were resolved by SDS-PAGE electrophoresis and blotted on PVDF membrane (Immobilon PVDF, Merck). Membrane blocking (10% BSA, TBS 0.1% Tween-20 for 1 hour) was followed by incubation with the appropriate primary antibodies and HRP-conjugated secondary antibodies (Cell Signaling Technology). Proteins were revealed by ECL system (Clarity ECL Western Substrate, Biorad) and chemo luminescence was measured with a ChemiDoc instrument (Biorad). The following antibodies were used, according to the manufacturer's instructions, at the indicated dilutions: DDX5 (1:100), DGCR8 (1:10000), FUS (1:1000), DROSHA (1:1000), VCL (Millipore 06-866, dilution 1:10000), PRMT1 (Abcam ab73246, 1:1000), PRMT4

(Bethyl A300-421A, 1:5000), PRMT5 (Abcam ab109451, 1:10000), PRMT6 (Abcam ab47244, 1:1000) , PRMT7 (sc-98882, 1:200), ASYM24 (Millipore, 1:1000), SYM10 (Millipore, 1:1000), MeR4-100 (Cell Signaling Technology 8015, 1:1000), LAMIN A/C (sc-6215, 1:1000), Lamin B (Abcam ab16048, 1:1000), GAPDH (Abcam ab9484, 1:1000), H3 (Abcam ab1791, 1:5000), H4 (Abcam ab7311, 1:5000), H4R3me2a (Active Motif 39705, 1:1000), ILF3 (Bethyl A303-615A, 1:2000), ILF2 (sc-271718, 1:1000), DDX17 (sc-130650, 1:1000), DDX3 (LS-C64573, 1:5000).

4.9 Total and small- RNA extraction

Total and small- RNA were prepared using mirVANA^{TM} miRNA Isolation kit (Ambion), according to the manufacturer's specification.

For the analysis of small-RNA in distinct cellular compartments, RNA was isolated from the chromatin fraction, the nucleosol and the cytosol by adding 1ml of Trizol (Trizol LS Reagent, Ambion) and 200 µl of chloroform to each fraction. The solution was mixed, incubated 5 min at RT and then centrifuged at 13000 rpm for 15 min at 4°C. The aqueous phase containing the RNA was collected and loaded on miRVANA column for RNA isolation.

DNAse (Zymo Research) treatment of RNA was performed before reverse transcription.

4.10 cDNA synthesis and real-time PCR

The cDNA was produced using the reverse-transcriptase miScript II RT Kit (Qiagen), starting from 1 µg of total RNA, according to the manufacturer's specifications for mRNA, pri-miRNA, pre-miRNA and miRNA profiling. One tenth of the reaction was used for Real-time PCR reactions in a 7,500HT Fast Real-Time PCR System. miScript SYBR Green PCR Kit (Qiagen) was used for miRNA and pre-miRNA analysis, following manufacturer's instruction, while mRNA and pri-miRNA were analyzed with FAST Sybr Green Master Mix (Life Technologies).

4.11 Primers used for quantitative PCR

For primary miRNA analysis we designed standard quantitative PCR (qPCR) primers that match a region just before the stem loops of each miRNA cluster under investigation **(Fig. 10A)**. Since the mature miRNAs of a cluster originates from the same primary transcript, miR-15a and miR-16-1 share the same pri-miRNA, likewise all the miRNAs of miR-17/92. MicroRNAs miR-15a, miR-16-1, miR-17, miR-18, miR-19a, miR-19b, miR-20, miR-92, miR-150 were amplified using forward primers from miScript primer assays (Qiagen) and reverse universal primers from miScript SYBR Green PCR Kit (Qiagen). Briefly, mature miRNAs are not polyadenylated, thus, when reverse transcription reactions are performed using miScript RT kit from Qiagen, mature miRNAs are polyadenylated by poly(A) polymerase and reverse transcribed in cDNA using oligo-dT primers. The oligo-dT primers have a 3' degenerate anchor and a universal tag sequence on the 5' end allowing amplification of mature miRNA in the real-time PCR step. In fact, the combination of the forward primer (miScript Primer Assay) with the miScript SYBR Green PCR Kit allows miRNA amplification.

Since precursor miRNAs are not polyadenylated, we adapted the miScript Qiagen strategy for miRNA analysis also to precursor amplification. We designed forward primers specific for the stem-loop sequences of each pre-miRNAs of interest and exploited the reverse universal primer from the miScript SYBR Green PCR Kit **(Fig. 10B, upper panel)**. Upon qPCR amplification, the strategy we used resulted specific for mature and precursor miRNAs detection **(Fig. 10B, lower panel)**.
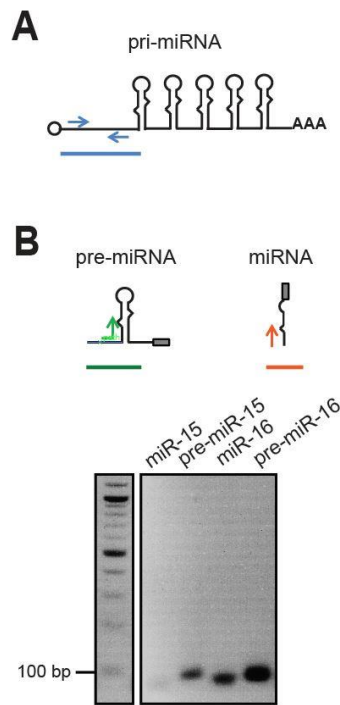
Fig. 10 qPCR primer designed for pri-, pre- and mature miRNAs analysis
A) Schematic representation of the strategy used to design specific qPCR primers for pri-miRNAs. **B)** Upper panel: schematic representation of the strategy used to design specific qPCR primers for pre-miRNAs amplification (in green) and comparison with Qiagen mature miRNAs primers (in red). Grey boxes represent the reverse universal primers. Lower panel: visualization of the qPCR products obtained using primers specific for mature and precursor miR-15a and miR-16-1 (representative of all miRNAs tested); mature miRNAs are shorter than the corresponding precursors.

Primer sequences are the following:

DGCR8: Fwd AAAACTTGCGAAGAATAAAGCTG,

      Rev TCTGTTTAACAAAGTCAGGGATGA

pri-miR-15a/16-1: Fwd GCCCTGTTAAGTTGGCATAGC,

      Rev ACTGAAGTCCATTCTGTGCCC

pri-miR-17/92: Fwd TGCCACGTGGATGTGAAGAT,

      Rev GGCCTCTCCCAAATGGATTGA

pre-miR-15: Fwd CCTTGGAGTAAAGTAGCAGCACA

pre-miR-16: Fwd CAGTGCCTTAGCAGCACGTA

pre-miR-17: Fwd AAAGTGCTTACAGTGCAGGTAGT

pre-miR-18: Fwd GGTGCATCTAGTGCAGATAGTGA

pre-miR-19a: Fwd GTCCTCTGTTAGTTTTGCATAGTTG

pre-miR-19b: Fwd GTTAGTTTTGCAGGTTTGCATCC

pre-miR-20: Fwd GTAGCACTAAAGTGCTTATAGTGCAGG

pre-miR-92: Fwd CTACACAGGTTGGGATCGGT

4.12 Taqman Array Human microRNAs

For the global analysis of miRNAs we used TaqMan Array Human microRNA A+B Cards (Applied Biosystems). In this case small RNA was prepared using mirVANA^{TM} miRNA Isolation kit (Ambion), and cDNA was prepared using Megaplex^{TM} RT reaction kit (Applied Biosystems).

Data were normalized on the geometric mean of a panel of housekeeping genes (MammU6, U6snRNA). MicroRNAs were considered significantly up/down-regulated when greater/lower than 1.5 fold changes.

4.13 Small-RNA sequencing: preparation of libraries, sequencing and data analysis

1 μg of total RNA was used as input for libraries preparation with the TruSeq Small RNA Library Prep Kit (Illumina), following manufacturer's instructions.

Briefly, RNA 3' and RNA 5' adapters were sequentially ligated to the RNA. Reverse transcription followed by PCR was used to create cDNA constructs based on the small RNA ligated with the adapters. The resulting cDNA constructs were gel-purified, eluted and concentrated by ethanol precipitation. The DNA fragment library was quantified with Bioanalyzer (Agilent Technologies) and sequencing was performed on an Illumina HiSeq2000 at 50bp single-read mode and 20 million read depth. Data analysis was performed with the IsomiRage workflow, as previously described [148]. Data normalization was performed after reads mapping, assignment and filtering. Normalization of the data was performed with a reads-per-million (RPM) normalization, using small nucleolar RNAs (snoRNA) or canonical miRNAs reads in each sample as normalizer.

5. RESULTS

5.1 Characterization of the Large Drosha Complex methylation pattern

In order to carry out a comprehensive analysis of the methylation pattern of the LDC, we set up a strategy combining the hmSILAC approach with affinity enrichment of the complex, followed by mass spectrometry analysis. As outlined in the Introduction (paragraph 2.6.3), the hmSILAC approach allows improving the confidence in methyl-site identification, overcoming several limitation of MS-based analysis of this type of modification in label-free mode, thanks to the specific isotope labeling of *in vivo* methylation. Moreover, the co-immunoprecipitation (co-IP) of the LDC facilitates the enrichment of all components of the complex, independently from their abundance in the cell. This aspect is crucial to increase the MS-coverage of all 23 subunits of the LDC and boost the comprehensive identification of methylations also on sub-stoichiometric proteins.

The strategy followed is outlined in **Fig. 11**. Briefly, HeLa S3 cells were metabolically labeled with the two isotopic variants (light, L and heavy, H) of methionine, up to full incorporation. Labeled cells were harvested, mixed in equal amount (1:1) and cell extracts fractionated into nuclear and cytosolic extracts. Nuclei were used as input for the immuno-affinity enrichment of the complex by co-IP and the immunoprecipitated material was separated on SDS-page and subjected to in-gel digestion with trypsin prior to LC-MS/MS. In parallel, the IP of the baits and the co-IP of some other subunits of the LDC, were validated through western blot (WB). Upon MS analysis, the high-confidence identification of *in vivo* mono-, di- and tri-methylated peptides was achieved taking advantage of the hmSILAC labeling: upon mixing H/L cells in 1:1 ratio, the *in vivo* methylated peptides result in the MS spectra as peptide-pairs separated by a specific mass difference corresponding to the number of methyl-groups present.
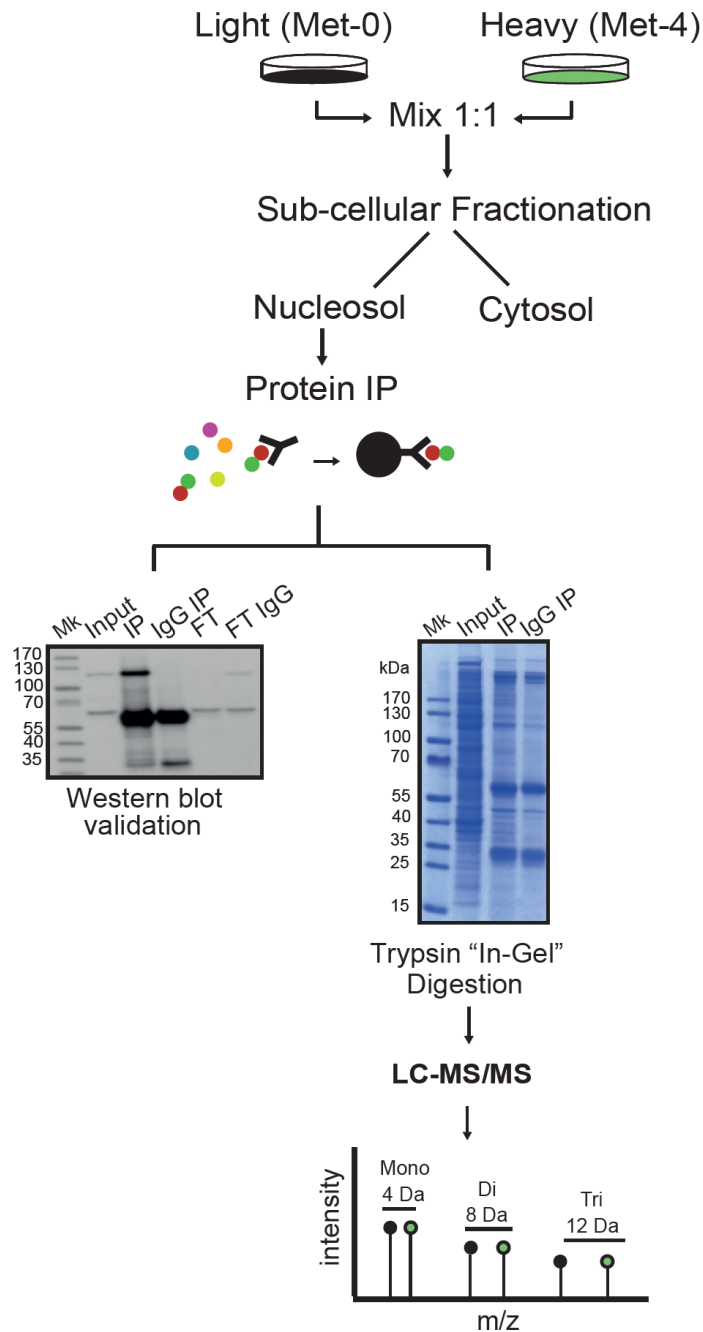
Fig. 11 Description of the hmSILAC/co-IP approach to characterize the LDC methylome

HeLa S3 were metabolically labeled with the light (Met-0, black dish) and heavy (Met-4, green dish). Cells were harvested, mixed 1:1 and subjected to fractionation into cytosolic and nuclear extracts. The nucleosol was used as input for the immuno-affinity enrichment of the complex, by co-IP, using different subunits of the complex as baits. The precipitated proteins were in-gel digested with trypsin, and analyzed with LC-MS/MS. Upon MS analysis, the type of methylated residue can be identified on the basis of the mass separation of the peptide pair (4 Da = mono-methylation, 8 Da = di-methylation, 12 Da = tri-methylation). In parallel, each co-IP was validated through WB analysis.

5.1.1 Heavy methyl SILAC labeling of HeLa S3

HeLa S3 cells were cultured in SILAC DMEM supplemented with light Arginine and Lysine and either heavy (Met-4) or light (Met-0) methionine, in parallel. Moreover, the SILAC medium contains heat-inactivated and dialyzed FBS, to ensure that the isotope-coded light (Arg-0, Lys-0, and Met-0) and heavy (Met-4) amino acids represent the only source of amino acids used for new protein synthesis. Growth curves of HeLa S3 cultured in light and heavy media are comparable **(Fig. 12A)**, indicating that cell viability is not influenced by the different isotope-coded amino acid composition of the media.

Cells were cultured in heavy hmSILAC medium for nine replications and the incorporation rate of heavy methionine was then assessed; upon cell lysis, the total extract was separated on SDS-PAGE and MS-analyzed upon in-gel trypsin digestion. The percentage of heavy methionine-containing peptides was evaluated by quantitative MS: out of 1031 peptides containing methionine identified, 795 (77%) were present only in the heavy form and only 8 in the light form, whereas 214 peptides were identified in both heavy-Met and light-Met forms. The distribution of heavy-Met/light-Met ratios associated to these peptides allowed to estimate an incorporation rate of 95% **(Fig. 12B)**. The almost complete incorporation of heavy methionine in HeLa S3 cell was therefore suitable for further experiments.
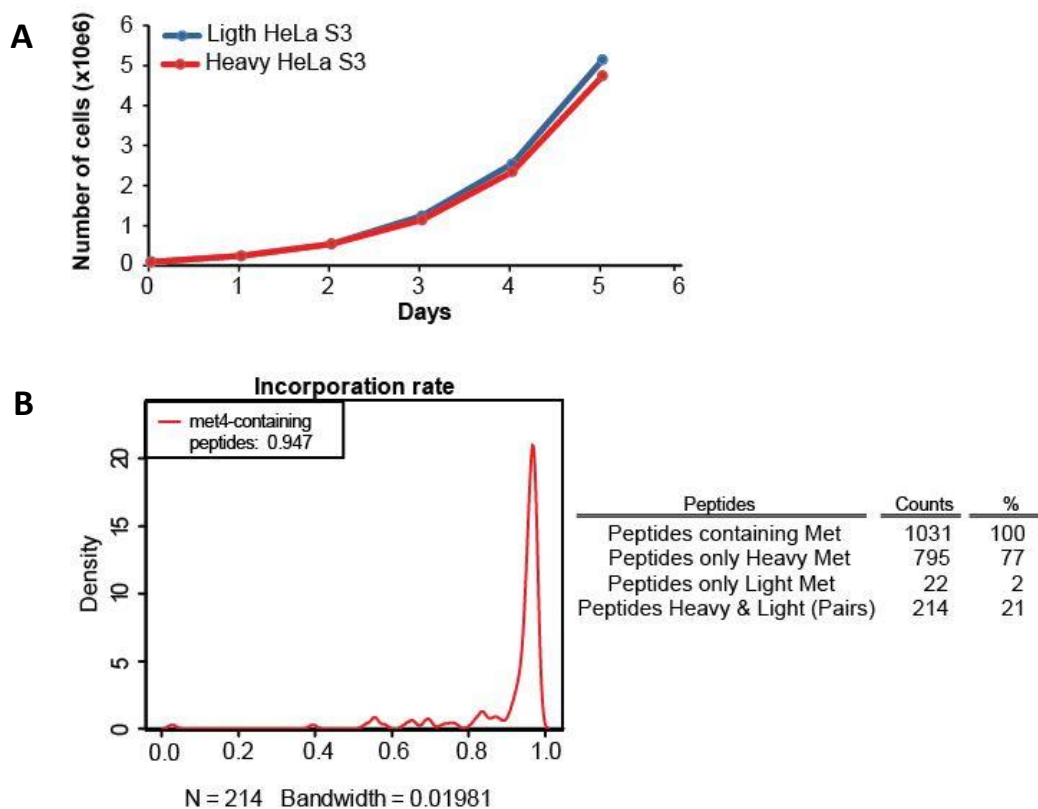
Fig. 12 Heavy methyl SILAC labeling of HeLa S3 cells
A) Comparison of the growth curves of heavy-labeled (red) and light-labeled (blue) cells.
B) Heavy methionine (Met-4) incorporation rate in HeLa S3. The Table summarizes the number and corresponding percentages of peptides identified in each form.

5.1.2 Nuclear and cytosolic sub-cellular fractionation of HeLa S3

The LDC is a nuclear complex and, even if many of its components shuttle between nucleus and cytosol, the majority of them are usually located in this cell compartment. For this reason, we decided to optimize the affinity enrichment of the complex using as input nuclear extracts.

 We optimized a cell fractionation protocol (described in details in Materials and Method, paragraph 4.2) that allows an efficient separation of cytosol and nucleus, as confirmed by the correct localization of cytoplasmic (Vinculin) and nuclear (Lamin A/C) markers profiled by Western blot (WB) **(Fig. 13A)**. Moreover, this protocol allows further separating the nucleosol from the chromatin fraction, which was excluded from the analysis since we want to favor the identification of non-histone methylation. As a matter of fact, histones

are highly methylated and very abundant in the nucleus and shotgun proteomic approaches, including MS-based analysis of PTMs, have a bias towards more abundant proteins. Hence, the choice of using nucleosolic extracts as input for the enrichment of the complex, excluding both cytosol and chromatin, was strategic to support in-depth annotation of the LDC methyl-proteome.

5.1.3 Co-immunoprecipitation of the Large Drosha Complex

The LDC is a large multi-protein complex with several subunits (like DDX5, DDX17 and DHX9) that are also engaged in other complexes, such as the spliceosome. Hence, to be able to enrich efficiently the complex in its completeness, we set up co-IP protocols using four different subunits as bait: DGCR8 and Drosha -which constitute the core components of the LDC- and the two associated proteins DDX5 and FUS. Using unlabeled nuclear extract, we first carried out an extensive optimization of the co-IP protocols, testing each antibody against the four different baits, in order to achieve the highest coverage of the LDC (both in terms of number of subunits identified and of their protein coverage by MS) and to reduce the cross-contamination with background proteins. The co-enrichment of other subunits of the complex with the baits was validated by WB **(Fig. 13B)** and the precipitated proteins were loaded on SDS-page and Coomassie-stained. Each IP and input was cut in 9 gel slices and in-gel digested (see Materials and Methods, paragraph 4.4) with trypsin, prior to MS analysis **(Fig. 13C)**.

We identified reproducibly all subunits of the complex in all experiments performed. However, DGCR8 was identified only when either the protein itself or Drosha were used as baits, which indicates that DGCR8 is a low abundant protein and becomes detectable only when enriched through direct protein-protein interaction with its partner in the core complex. The 23 subunits of the complex were identified in the different experiments with overall average sequence coverage of 40% **(Fig. 13D)**. To validate the efficiency of each IP in enriching the complex, the normalized Label Free Quantitation (LFQ) intensity values of the LDC components were compared to the intensities of the background proteins: the

shift in the median LFQ values indicated the efficient enrichment of the complex, compared to background **(Fig. 13E).**
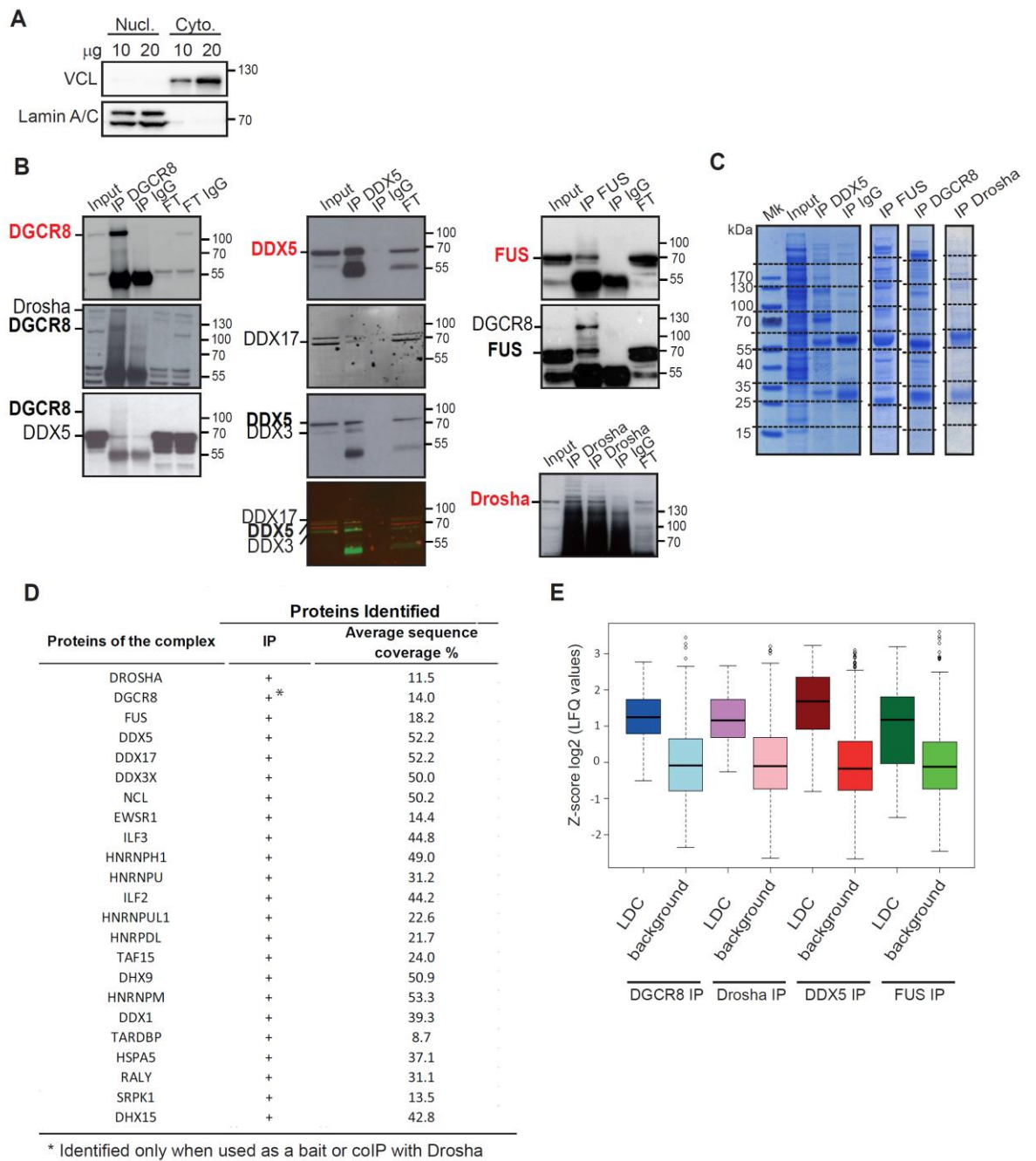
Fig. 13 Optimization of the co-IP protocols for the enrichment of the Large Drosha Complex

A) WB validation of HeLa S3 sub-cellular fractionation. Vinculin (VCL) and Lamin A/C were used as marker for cytoplasm and nucleus, respectively. Increasing amounts of extracts (10 and 20 μg) were used for the analysis. **B)** WB validation of the co-IP of some subunits of the complex together with the four baits used (labeled in red). Immunoprecipitation with Immunoglobulin G (IgG) was used as negative control. FT =

Flow-through. **C)** Coomassie-stained gels of the optimized co-IPs are illustrated. Dashed lines indicate the slices in which each lane was cut for protein in-gel digestion and peptide extraction, prior to LC-MS/MS. **D)** List of the subunits identified by MS using DGCR8, Drosha, DDX5 and FUS as baits in the co-IP. The mean sequence coverage of each protein from the four different IPs is indicated. **E)** Efficiency of each IP in enriching the complex was assessed comparing the normalized LFQ intensity values of the LDC subunits in the IPs, to those of the background: the shift in the median LFQ intensities represents the enrichment of the complex compared to background proteins.

5.1.4 Features of the Large Drosha Complex methyl-proteome

Upon the optimization of the affinity-enrichment protocol with unlabeled extracts, we performed the IP experiments using hmSILAC-labeled nuclear extract as input, to characterize the methylation pattern of the LDC. Upon five independent co-IPs followed by MS analysis and data processing (as described in details in Materials and Methods, paragraph 4.6 and 4.6.1) we identified 94 individual methylations on 68 distinct sites, on 14 out of 23 proteins of the complex. Combining these results with those obtained in our previous global analysis of methylation based on hmSILAC [1], we extended the annotation of high-confidence *in vivo* methylated sites for the LDC to 123 modifications on 90 sites, occurring on 17 of the 23 components of the complex **(Fig. 14A, appendix 1)**, confirming our previous observation that the Large Drosha Complex is highly methylated.

 The majority of modifications occurs on Arginine (R) residues (86%), which are equally mono- or di-methylated, with a significant proportion of residues identified in both forms, while the few Lysines (K) identified are mostly mono-methylated **(Fig. 14B)**. Interestingly, the analysis of the methyl-sites localization within protein sequences revealed that the majority of all identified sites (89%), as well as the one of R-sites (90%), were localized within Low Complexity (LC)/ disordered regions **(Fig. 14C)**. LC regions are sequences with very little variation in their amino acid composition, to which a role in mediating protein sub-cellular localization, protein-protein and protein-RNA interactions has been ascribed, especially for RNA-binding proteins [149-152].

Our combined strategy allowed not only to confirm many of the sites annotated in our previously published dataset [1], but also to identify novel modification sites also on proteins not described as methylated before **(Fig. 14D)**. Finally, we could also generate a

comprehensive library of high-confidence methylations on the LDC, to be used as a reference for all further functional experiments on methyl-dynamics.

Overall, these results confirmed that methylation, especially on R, massively modifies the LDC corroborating the hypothesis that this modification may play a crucial role in the regulation of its composition and/or function.
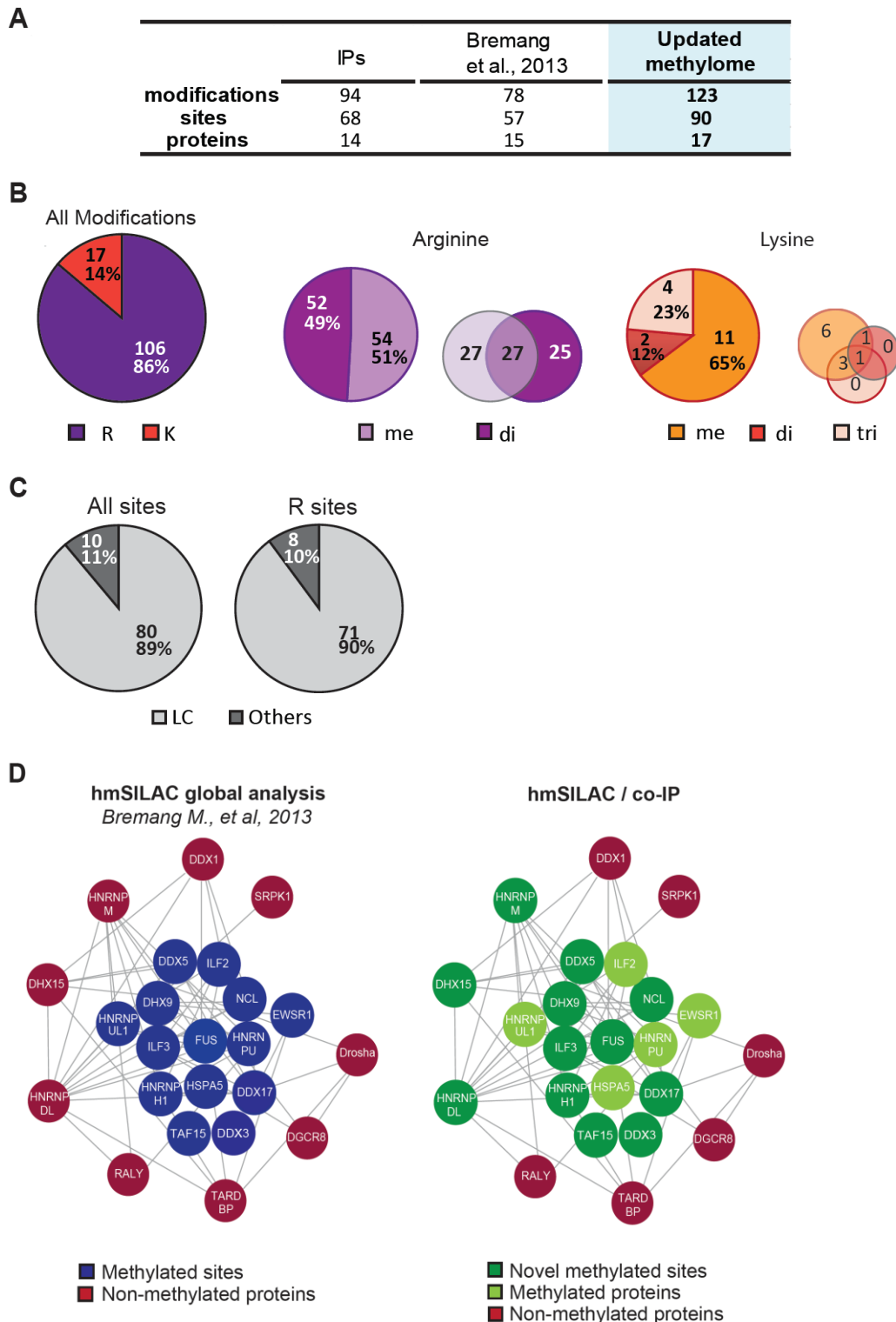
Fig. 14 Features of the *in vivo* methyl-proteome of the Large Drosha Complex
A) Summary of the LDC methyl-proteome. **B)** Distribution of R and K methylations identified on the complex (left panel) and R and K methyl-proteome composition (right panel). **C)** Pfam localization analysis of all the sites identified and of the R sites (LC= Low Complexity regions). **D)** Comparison between previously identified methyl-sites on the complex [1] and the novel identifications achieved through the combination of hmSILAC and co-IP/MS.

5.2 Set up of PRMT1 depletion and overexpression to modulate the Large Drosha Complex methylation state

Given the recognized role of protein methylation in regulating the function of other protein complexes, such as the spliceosome [42, 153] we explored the possibility that it could also regulate the Large Drosha Complex.

Arginines, which resulted as the most frequently methylated residue on the complex, can be modified by a set of arginine methyltransferases (PRMTs) in mammalian cells. Thus, we decided to focus on PRMT1, the most abundant and active PRMT [59], and to modulate its expression levels both through knock down (KD) and overexpression, with the aim of studying if and how methylation on the LDC change and to understand whether possible methylation changes may affect the complex function, thus impacting miRNA biogenesis.

The strategy we followed to deplete and overexpress PRMT1 is based on HeLa transduction with lentiviral particles carrying either different short hairpin (sh) RNAs specific for the enzyme or PRMT1 cDNA respectively, as schematized in **Fig. 15**. In both cases we used Empty vectors as a control. HeLa were harvested upon puromycin selection of positively infected cells and subjected to either protein extraction for expression and methylation analysis or to miRNA isolation to study the LDC processing activity.

Fig. 15 Scheme of the strategy used to deplete and overexpress PRMT1 in HeLa cells
Wild-type HeLa cells were transduced with lentiviral particles carrying Empty vector (green), different short hairpin RNAs (sh) (light blue) or PRMT1 cDNA (red). Upon puromycin selection of infected cells (dashed grey line) cells were harvested and subjected to different type of analyses.

5.2.1 Set up of PRMT1 knock down

For PRMT1 depletion, we initially tested three different sh-RNAs, called sh-B6, sh-B10 and sh-B8 respectively. Even if they all resulted efficient in achieving the depletion of the enzyme, we noticed that PRMT4 and PRMT6 expression was slightly decreased when using sh-B8 **(Fig. 16A)**. Hence, we discarded the sh-B8 construct and used sh-B6 and sh-B10 for the subsequent experiments.

Although there is evidence that the PRMTs might compensate each other [138], when we profiled the expression of other enzymes by WB, none of the sh-RNAs employed led to a detectable increase of their protein levels **(Fig. 16A)**. Yet, this does not exclude the possibility that the other PRMTs compensate the loss of PRMT1 through mechanisms that do not require changes in their expression, such as variation in their activity or substrate specificity.

Loss of PRMT1 expression is reported to be embryonic lethal in mice [154] and for this reason we assessed cell growth after its depletion. We performed a growth curve analysis starting from 24 hours (h) post cell infection comparing the growth of control and PRMT1 KD cells. Up to 96h after the infection, the growth of PRMT1-depleted cells is comparable

to control **(Fig. 16B, left panel)**. However, starting from 96h post infection, PRMT1 depletion starts to affect cell viability, ultimately being not compatible with cell growth at later time points **(Fig. 16B, right panel)**.

We thus profiled global methylation changes by WB after infection. Taking advantage of anti-pan-methyl antibodies that recognize different forms of R-methylation, we analysed asymmetric di-methylation and mono-methylation since they are the types of modifications catalyzed by PRMT1. PRMT1 is completely depleted at 48h post infection (PI), using both sh-RNAs **(Fig. 16C, left panel)**, while major changes in global methylation can be appreciated only at 72 and 96h PI. Indeed mono-methylation is already increasing at 72h while di-methylation shows an overall decrease at 96h **(Fig. 16C, left panel)**. Similar studies have shown that PRMT1 depletion causes a decrease in asymmetric di-methylation that corresponds to a strong increase on mono-methylations [138]. Still, it is not clear whether this effect is due to compensatory mechanisms or competition by other PRMTs or just as a consequence of the lack of di-methylation. Moreover, it has been reported that after an initial loss of di-methylation, four days upon PRMT1 depletion its levels starts to increase again [138]. We also observed a similar effect when profiling asymmetric di-methylation at later time points, up to 6 days PI **(Fig. 16C, right panel)**.

Based on these data, we chose to perform all following experiments 72 hours after PRMT1 KD, on the one hand because of the affected cell viability at later timing and on the other hand because of the changes in the global methylation occurring within 4 days PI.

Fig. 16 PRMT1 KD set up
A) WB analysis of PRMT1, PRMT4, PRMT5, PRMT6 and PRMT7 expression upon PRMT1 KD with three different sh-RNA sequences (sh-B6, sh-B10, sh-B8). Empty is the negative control. VCL was used as loading control. **B)** Growth curves of control and PRMT1-depleted cells between 24 and 96 hours post infection (PI) (left panel), and between 96 and 192 hours PI (right panel). Dashed grey line at 24h PI indicates puromycin selection. **C)** Time course profiling of PRMT1 depletion, asymmetric di-methylation (ASYM24) and mono-methylation (mono) between 24 and 96 hours PI (left panel). Time course profiling of PRMT1 depletion and of asymmetric di-methylation (ASYM24) between 4 and 6 days post infection (right panel). VCL was used as loading control. Red asterisks highlight changes in di- or mono-methylation.

5.2.2 PRMT1 overexpression set up

To achieve the enzyme overexpression in HeLa cells, PRMT1 cDNA was cloned into a lentiviral vector following the same strategy used for the knock down. The corresponding empty vector was used as a negative control. To evaluate the overexpression efficiency and the activity of the ectopic enzyme, we profiled PRMT1 and asymmetric di-methylation global changes in a time window between 24h and 96h after the infection. The expression

of the protein starts to increase already after 24-48h and correlates with a global increase of di-methylations **(Fig. 17A, upper panel)**. 48h PI we observed the increase of asymmetric di-methylation of R3 of histone H4 (H4R3me2a), a well-known PRMT1 histone target [60] **(Fig. 17A, lower panel)**. This result reassured that the gain of modification observed was due to the augmented PRMT1 level.

In HeLa PRMT1 is ubiquitously expressed both in the nucleus and in the cytoplasm. Thus, we analysed the localization of the ectopic enzyme, by fractionating control and PRMT1-overexpressing cells into nuclear and cytosolic extracts. We performed this experiment at 72h and 96h after the infection because at this time points PRMT1 was highly overexpressed. The enzyme expression level is increased in both cell compartments compared to the control **(Fig. 17B)**. Accordingly, upon PRMT1 overexpression, we observed increased asymmetric arginine di-methylations and decreased mono-methylations, in particular in the nucleus where the LDC is located and exerts its function **(Fig. 17B)**. Moreover, the expression of other PRMTs (PRMT4, 6 and 7) was not affected by PRMT1 overexpression **(Fig. 17C)**, suggesting that methylation changes may be due to ectopic PRMT1 activity rather than to the activity of other enzymes.
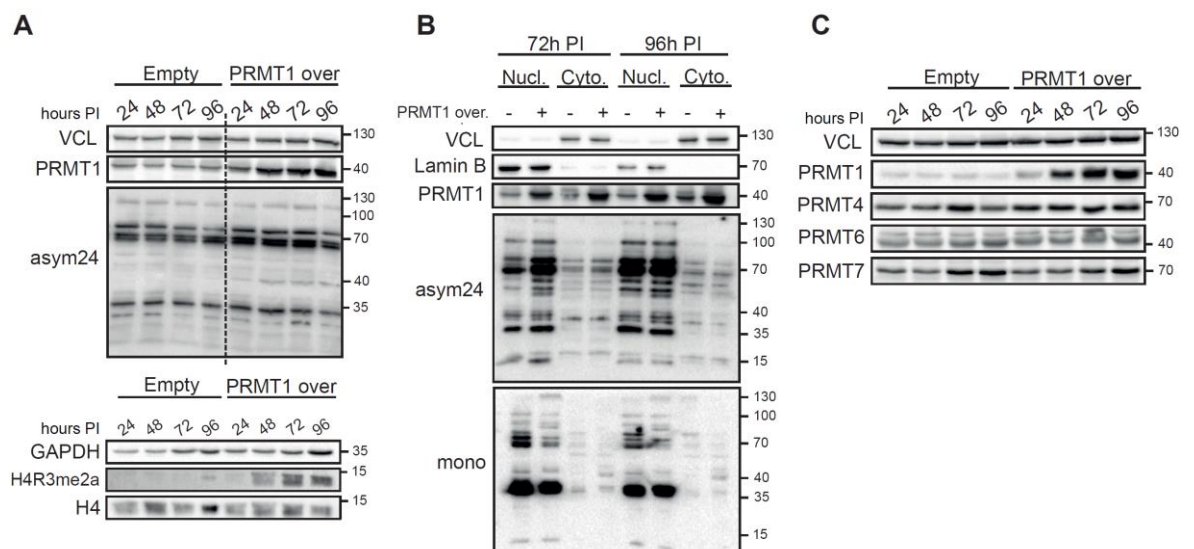
Fig. 17 PRMT1 overexpression set up
A) Western Blot profiling of PRMT1 overexpression and of asymmetric di-methylation (ASYM24) between 24 and 96 hours PI. VCL was used as loading control (upper panel).

Time course profiling of H4R3me2a on the same samples. GAPDH was used as loading control (lower panel). **B)** WB analysis of endogenous and ectopic PRMT1 localization at 72h and 96h PI. Cells were fractionated into nuclear (Nucl.) and cytoplasmic (Cyto.) extracts. Global asymmetric di-methylation and mono-methylation are also profiled. VCL was used as cytoplasmic marker. Lamin B was used as nuclear marker. **C)** WB analysis of PRMT1, PRMT4, PRMT6 and PRMT7 expression upon PRMT1 overexpression. VCL was used as loading control.

5.3 Identification of PRMT1-specific targets on the Large Drosha Complex

5.3.1 Analysis of the effect of PRMT1 depletion on the LDC at protein and methyl-peptides levels

The strategy we followed to identify PRMT1 targets on the LDC is outlined in **figure 18A**: we combined standard SILAC labeling of control and PRMT1-depleted cells with affinity enrichment of the complex by co-IP, followed by high resolution MS analysis and MaxQuant quantification of methyl-peptides in the two functional states. Heavy methyl SILAC is a useful tool for the high-confidence identification of *in vivo* methylation sites, but it has the limitation that only methionine-containing peptides can be quantified. For this reason, having built a reference library of confidently identified methylated sites on the LDC, we decided to move to a standard SILAC setup for this type of experiment. Moreover, SILAC allows identifying and quantifying peptides through a dedicated software package making the quantitation analysis more straightforward and robust [22, 23].

HeLa cells were labeled by growing them in media supplemented with light (Arg0, Lys0) or heavy (Arg10, Lys8) acids. Upon full incorporation, cells were infected with either the empty vector or with sh-B6 vector and they were harvested and mixed in equal amount (1:1) 72 hours after infection, based on the evidence that protein methylation changes appeared to peak at this time point. Cells were fractionated into nuclear and cytosolic fraction and the nuclear extract was partly analyzed directly by MS, partly used as input for the immuno-affinity enrichment of the complex. We performed three IPs in parallel using DGCR8, DDX5 and FUS as baits, in order to identify as many methyl-sites as possible. The three IPs were analyzed through geLC-MS/MS and raw data analyzed by MaxQuant software, as described in Materials and Methods (paragraph 4.6.2). Within this experimental setup, the ratio of the heavy versus the light peak intensity for the methyl-

peptide of a specific protein reflects potential changes in its methylation state, whereas the heavy-to-light ratio of the intensity of the unmodified peptides for the corresponding protein are used to normalize over possible changes in protein levels **(Fig. 18A)**. We performed both 'forward' and 'reverse' experiments by swapping the SILAC channel in which PRMT1 was depleted.

In the "forward" experiment, PRMT1 KD was carried out in the heavy-labeled cells. Collected cells were fractionated **(Fig. 18B, upper panel)** and nuclear extract was used as input for the immuno-affinity enrichment of the complex through DGCR8, DDX5 and FUS IPs **(Fig. 18B, lower panel)**. The immunoprecipitated material was separated by SDS-page; nine slices were excised for gel lane and subjected to trypsin digestion followed by MS analysis with a Q Exactive instrument. Upon data analysis, we evaluated the efficiency of the IPs in enriching the complex, comparing the iBAQ intensity values of the LDC with those of the background proteins. The shift in the median iBAQ intensity indicated the successful enrichment of the complex for all the IP experiments **(Fig. 18C)**.
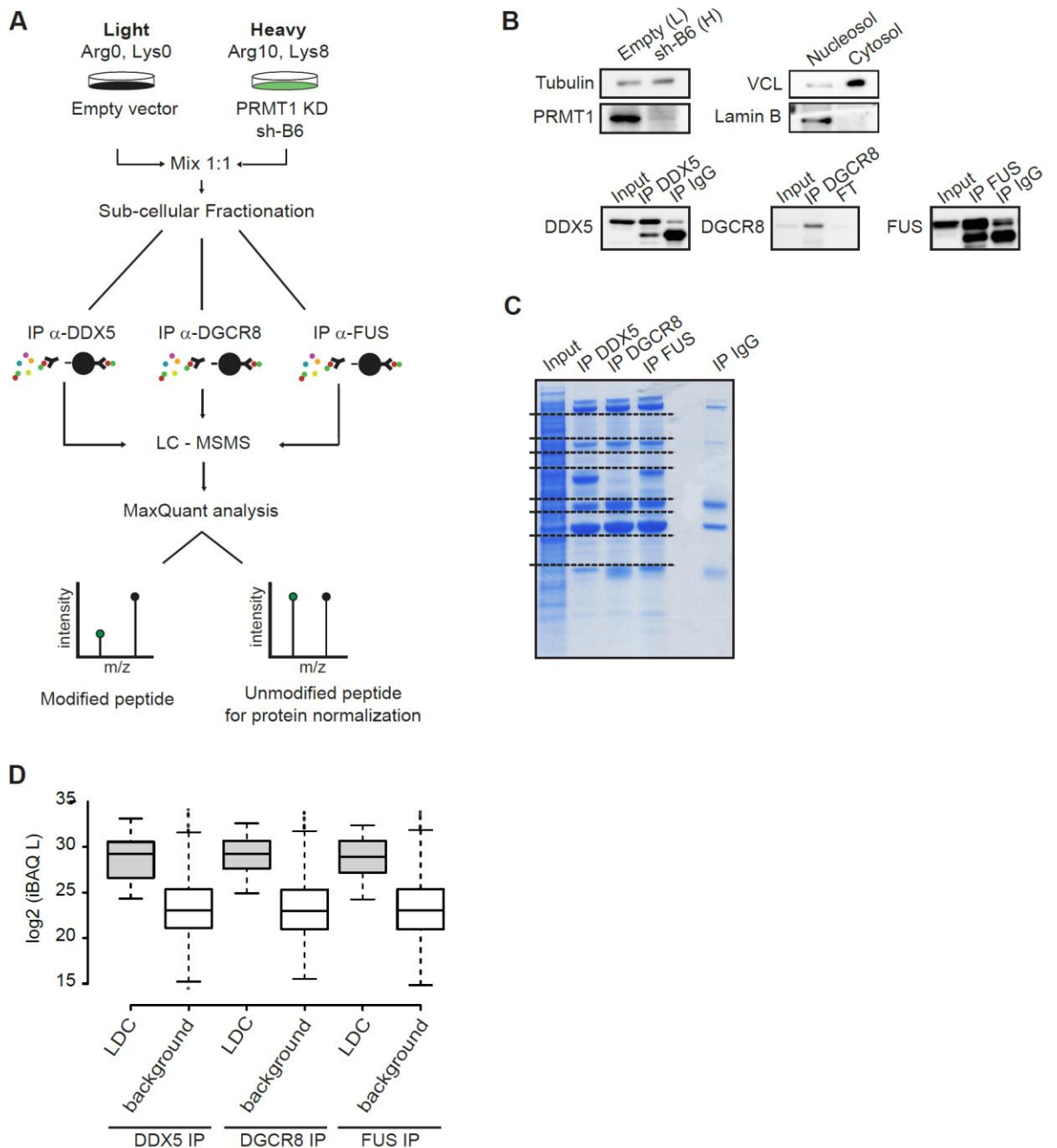
Fig. 18 Description of the SILAC/co-IP 'forward' experiment for the identification of PRMT1-specific targets on the Large Drosha Complex
A) Workflow of the experimental strategy: a"forward" experimental setup is schematized.
B) Upper panel: WB analysis of PRMT1 KD and evaluation of cell fractionation; tubulin was used as loading control, VCL and Lamin B were used as marker for cytoplasm and nucleus, respectively. Lower panel: validation of DDX5, DGCR8 and FUS IPs by WB. IgG was used as negative control for IP. FT = Flow Through. **C)** Coomassie-stained gel of DDX5, DGCR8, FUS and IgG IPs. Dashed lines indicate the slices in which each lane was cut for protein in-gel digestion and peptide extraction prior to LC-MS/MS. **D)** Evaluation of the efficiency of each IP in enriching the complex. For each IP experiment the iBAQ intensities of the proteins of the complex in the light (control) sample were compared to those of the background proteins.

We then performed a global analysis of the nuclear proteome upon PRMT1 depletion, with a specific focus on the LDC subunits, to assess possible changes in their protein expression. We examined SILAC ratios for all proteins identified and quantified confidently in the nuclear input of both 'forward' and 'reverse' experiments. We calculated the mean (μ) and standard deviation (σ) of the protein ratio distributions ($\log_2$ of H/L and L/H ratios in 'forward' and 'reverse' experiments, respectively) and defined as significantly up-/down-regulated those proteins that were greater or lower than $\mu \pm 2\sigma$ ('forward': $\log_2$ ratios greater than 1.01 and lower than -0.966; 'reverse': $\log_2$ ratios greater than 0.896 and lower than -1.174). Based on the fact that the H/L and L/H protein ratios of the LDC subunits lie within these cutoffs, we concluded that their expression was not significantly affected by PRMT1 depletion **(Fig. 19A, red dots)**, whereas we could confirm the substantial down-regulation of PRMT1 upon sh-B6 **(Fig. 19A, blue dot)**. We validated the MS results by WB, confirming that PRMT1 depletion does not affect the expression level of some key component of the complex, such as DGCR8 and Drosha **(Fig. 19B)**.
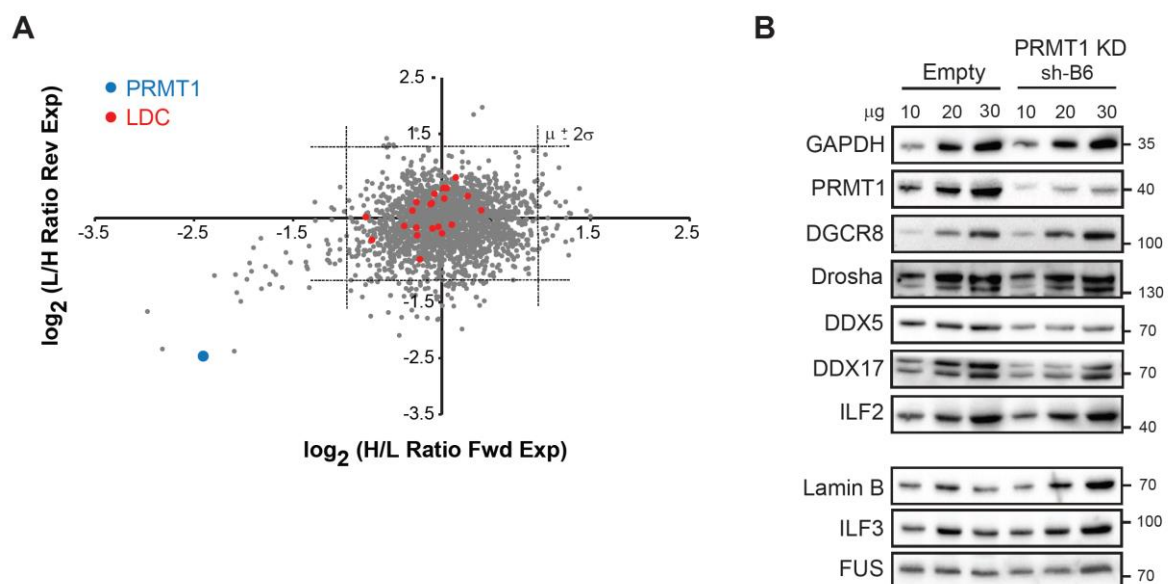
Fig. 19 Quantitative proteomics to profile changes in nuclear protein levels after PRMT1 depletion
A) The scatterplot shows the $\log_2$ SILAC distribution (H/L ratio and L/H ratio for the 'forward' and 'reverse' experiment, respectively) of the nuclear proteins identified in 'forward' and 'reverse' experiments, upon PRMT1 KD. PRMT1 is highlighted in blue; the components of the LDC are highlighted in red. Dashed lines represent cut off for

statistically significant regulated proteins. **B)** WB analysis of the expression of some key components of the complex in control and PRMT1 KD cells (sh-B6). Lamin B and GAPDH were used as loading control. Increasing amounts of nuclear extract (10, 20, 30 μg) were used for the analysis.

Upon MS data analysis of 'forward' and 'reverse' experiments by MaxQuant, we identified 142 methylated peptides within the LDC. Of these, 124 methyl-peptides were also SILAC-quantified **(Fig. 20A, appendix 2)** and the majority (118) contained methylated arginine. Overall, the 'reverse' experiment led to a lower rate of methyl-peptides identification, probably as a consequence of a less efficient enrichment of the proteins of the complex in the co-IPs. However, the quantitation of methylated peptides identified was reproducible between the 'forward' and the 'reverse' experiments. Moreover, the observation that the amount of methyl-peptides identified on the LDC in the IPs is sensibly higher than in the input, confirms that the immuno-affinity enrichment strategy does efficiently increase the methyl-sites identification within a set of proteins of interest **(Fig. 20A)**.

Out of the quantified 124 methyl-peptides, almost 70% (86) resulted significantly regulated upon PRMT1 depletion (after statistical analysis described in Materials and Methods, paragraph 6.2), with the majority (76) containing R methyl-sites. Regulated peptides occur on 13 of the 23 subunits of the complex. On these peptides, we identified 97 different modifications on 80 distinct sites **(Fig. 20B, upper panel)**; of note the number of peptides and sites does not coincide, since we identified more than one form of the same peptide bearing the same site. Among the 76 regulated R methyl-peptides, 46% were already annotated in the internal hmSILAC reference library **(Fig. 20B, lower panel).** Moreover, studying the sites identified on the regulated peptides, the 67% was reported in the library and therefore considered true and confident *in vivo* methyl-sites. The majority of them (46) occurred on arginine residues within 10 different subunits of the LDC, of which just FUS, TAF15, EWSR1 and ILF3 were already reported as PRMT1 targets, while DDX5, DDX17, ILF2, NCL and HNRNPH1 might represent newly identified substrates of the enzyme.

Overall, we cannot extrapolate a clear trend of regulation of the methyl-peptides as we observe equal proportion of down- and up-regulated peptides. However, it is interesting

that in various cases, the same peptide is down-regulated when di-methylated and up-regulated when mono-methylated (the peptides 465-GMPPPLR(me)/(di)GGPGGPGGPGGPMGR-486 of EWSR1 and 602-RAPVPVR(me)/(di)GGPK-613 of ILF3 are representative cases **(Fig. 20A)**. This trend is in line with the type-I enzymatic activity of PRMT1, which asymmetrically di-methylates either unmodified or mono-methylated arginines. Noteworthy, this result also confirms previous experimental evidence showing that loss of PRMT1 caused a global decrease of arginine di-methylation, with a corresponding increase of R-mono-methylation [139].

A localization analysis showed that the majority of the sites identified on regulated peptides occurred within LC regions of the proteins, with a subset (21%) located in close proximity to RNA recognition Motifs (RRM) or to double strand RNA-binding Motifs (DSRM) **(Fig. 20C)**. This suggests a potential regulatory effect of methylation in the binding of the complex to RNA, possibly newly processed microRNAs.
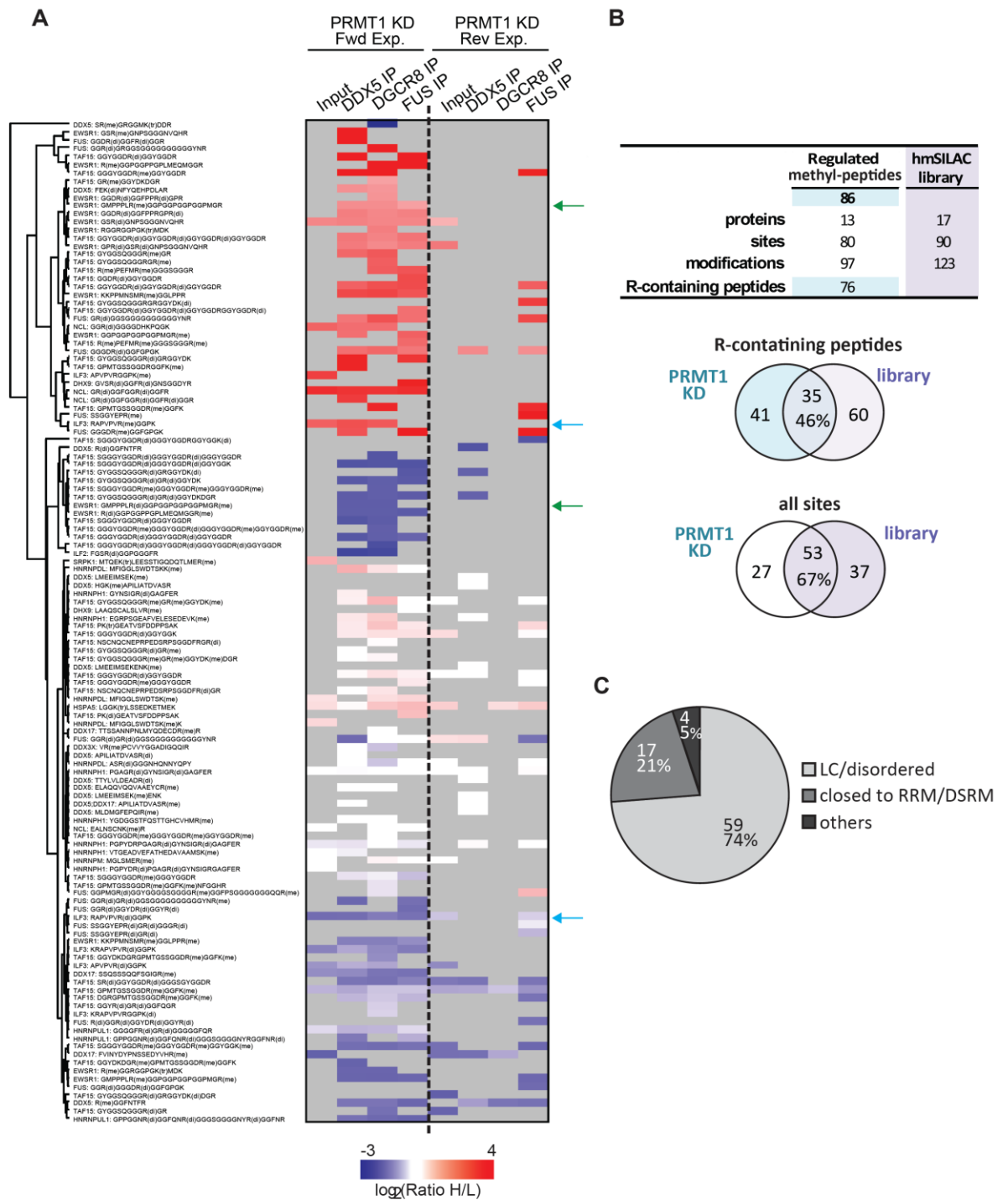
Fig. 20 Large Drosha Complex methyl-peptides changes upon PRMT1 depletion

A) Heatmap display of $\log_2$ H/L ratio of identified methyl-peptides in input and IP samples of the 'forward' and 'reverse' experiments (red = high ratio; blue = low ratio). Green arrows indicate the mono- and di-methylated forms of the peptide 465-GMPPPLRGGPGGPGGPGGPMGR-486 of EWSR1; light blue arrows indicate the mono- and di-methylated forms of the peptide 602-RAPVPVRGGPK-613 of ILF3. **B)** Summary of the regulated methyl-peptides identified on the LDC upon PRMT1 KD and summary of the hmSILAC library used as a reference (upper panel). Comparison of the results obtained with the hmSILAC library (lower panel). **C)** Pfam localization analysis of the methylated sites identified on regulated peptides (LC = Low Complexity regions; RRM = RNA Recognition Motifs; DSRM = Double Strand RNA-binding Motifs).

5.3.2 Validation of some PRMT1-specific targets on the complex through PRMT1 overexpression

Although the results obtained suggested that PRMT1 is responsible for the methylation of several components of the LDC, we could not exclude the possibility that other PRMTs could compensate to the loss of the enzyme. To validate the PRMT1-specific targets identified on the LDC, limiting the possible compensation by after overexpression of PRMT1.

The scheme of the experiment is outlined in **Figure 21A**: HeLa were grown in light and heavy SILAC media up to full incorporation of light and heavy amino acids, respectively. A 'forward' experiment was carried out, in which the light cells were infected with the empty vector (control) and the heavy-labeled ones with the construct for PRMT1 overexpression. Control and PRMT1-overexpressing cells were harvested, mixed 1:1 and separated into nuclear and cytoplasmic fractions **(Fig. 21B, upper panel)**. Nuclear extract was either directly MS analyzed, or used as input for the immuno-affinity enrichment of the complex through DGCR8, DDX5 and FUS IPs **(Fig. 21B, lower panel)**. The immunoprecipitated material in the three IPs and the nuclear input were separated by SDS-page; nine slices for gel lane were excised and subjected to trypsin digestion, followed by MS analysis with a Q Exactive instrument. MS data were processed with MaxQuant for protein identification and quantification of unmodified and methylated peptides. The IP experiments performed led to an overall efficient enrichment of the complex compared to background **(Fig. 21C)**. We defined proteins that were significantly up-/down-regulated by PRMT1 overexpression, based on the distribution of $\log_2$ SILAC H/L ratios for all proteins identified in the nuclear input considering $\mu \pm 2\sigma$ **(Fig. 21D)**: as expected, PRMT1 (marked in red) is significantly up-regulated, while the expression of the LDC subunits (green population centered on zero) is not significantly affected.
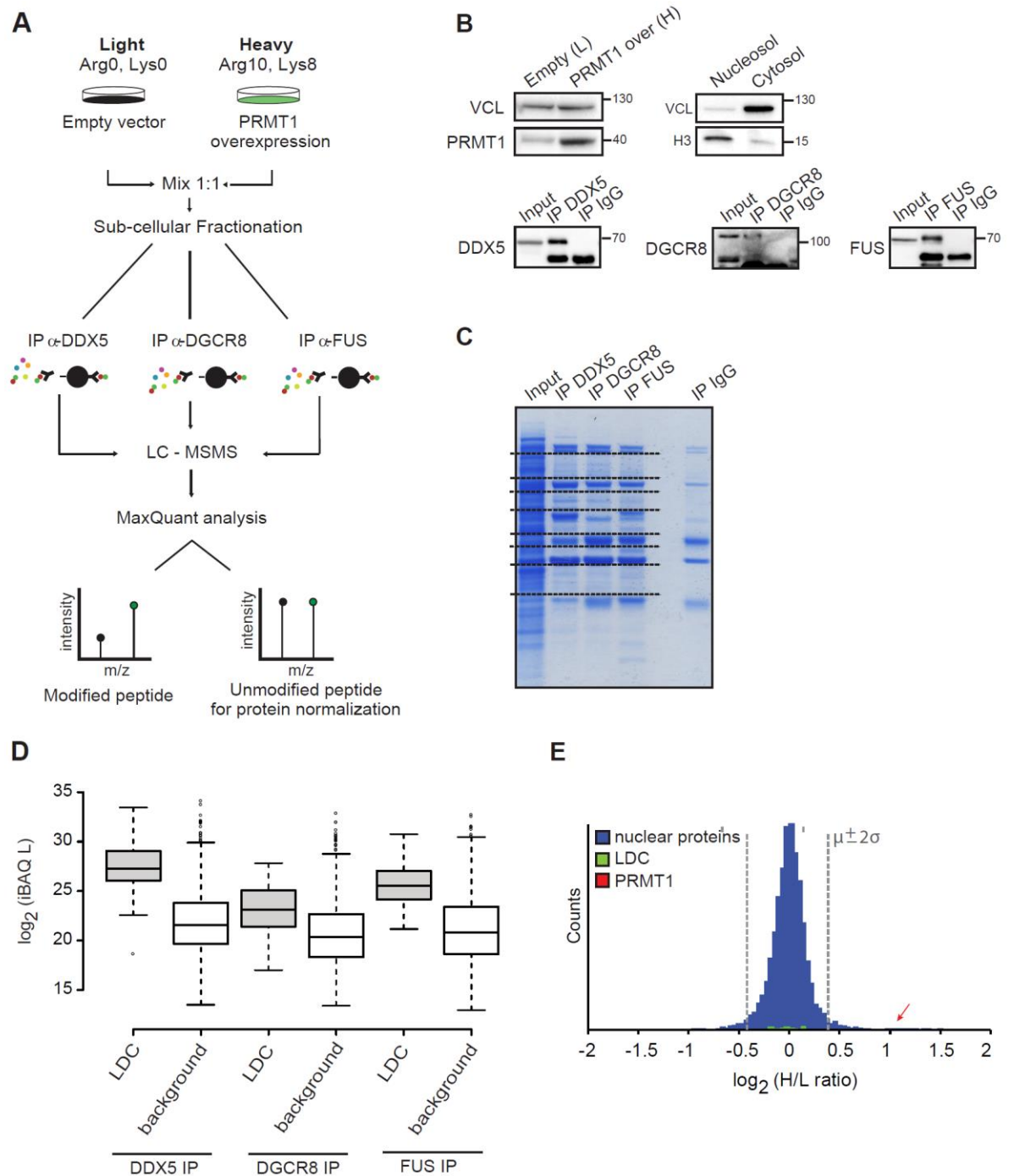
Fig. 21 SILAC/co-IP approach for the validation of PRMT1-specific targets on the Large Drosha Complex

A) Workflow of the experimental strategy: the 'forward' experiment is represented. **B)** Upper panel: WB analysis of PRMT1 overexpression and WB evaluation of cell fractionation; VCL was used as loading control; VCL and Histone H3 (H3) were used as marker for cytoplasm and nucleus respectively. Lower panel: WB validation of DDX5, DGCR8 and FUS IPs. IgG-IP was used as negative control. **C)** Coomassie-stained gel of DDX5, DGCR8, FUS and IgG IPs. Dashed lines indicate the slices in which each lane was cut for protein in-gel digestion and peptide extraction prior to LC-MS/MS. **D)** Evaluation of the efficiency of each IP in the enrichment of the complex. For each IP experiment the iBAQ intensities of the proteins of the complex in the light (control) sample were compared

to those of the background proteins. **E)** Log$_2$ H/L ratio distribution of nuclear proteins identified in the 'forward' experiment (in blue), upon PRMT1 overexpression. Dashed grey lines indicate the cut off to define significantly regulated proteins ($\mu \pm 2\sigma$ = mean $\pm$ 2 standard deviations). LDC components are highlighted in green; the red arrow highlights the position of PRMT1 within the distribution.

We identified and quantified a total of 80 methyl-peptides within the LDC, 67 of them (84%) containing R-methyl-sites. They occur on 14 proteins of the complex and within these peptides, we identified 91 methyl-sites bearing 103 different modifications **(Fig.22A, left panel)**. Comparing these results with the hmSILAC library, we noticed that just the 36% of R-containing methyl-peptides and the 49% of the methyl-sites identified was already annotated **(Fig. 22A, right panel)**. This limited overalap is somehow unexpected. A possible explanation could lie in the fact that in both the endogenous and ectopic PRMT1 are expressed and active in this experiment, leading to an increased cellular R-methylation on protein/sites that are not modified (and therefore not reported in the library) at the basal state.

The overexpression experiment was set up with the aim of validating PRMT1 targets identified in the previous depletion experiment, circumventing the compensation caused by other PRMTs when PRMT1 is down-regulated. Hence, we directly compared all the methyl-peptides identified in both experiments and found 42 methyl-peptides in common, occurring on 11 proteins of the complex **(Fig. 22B, appendix 3)**. Several methyl-peptides show a specular trend between PRMT1 depletion and overexpression: for instance, the mono-methylated peptide 465-GMPPPLR(me)GGPGGPGGPGGPMGR-486 of EWSR1 (indicated by the upper green arrow) is up-regulated upon PRMT1 KD and down-regulated when the enzyme is overexpressed. Interestingly, its di-methylated version (lower green arrow) behaves in the opposite way, being down-regulated upon the enzyme depletion and slightly up-regulated after the increase of PRMT1. These results are in line with the enzymatic activity of PRMT1 and with previous evidence [139], confirming that these specific methyl-peptides are real novel experimental PRMT1 targets.

However, we also observed cases where the peptides do not display opposite trend of methylation between the depletion and overexpression experiments: for instance some peptides are strongly down-regulated upon KD but not up-regulated when PRMT1 is induced (light blue line in **Fig. 22B**). This could be due to the fact that peptides are methylated at the basal level by the endogenous PRMT1 at saturation, thus cannot be further modified by the ectopic enzyme, thus producing "non-changing" ratios.

Considering these data (and taking into account that the 'reverse' experiment of PRMT1 overexpression is still on-going), we were able to confirm the identification of a number of PRMT1 targets of the LDC, which are listed in **Fig. 22C**. We excluded from the analysis those methyl-peptides that were non-regulated in both experiments and those that contained K methyl-sites. Only FUS, TAF15 and ILF3 were already reported as PRMT1 targets, even though for none the specific methyl-sites had been specified before, while all the remaining represents novel PRMT1 substrate annotations. The new PRMT1 target sites identified occur mainly in LC/disordered regions, corroborating the hypothesis that methylation may be functionally important to mediate protein-protein or protein-RNA interactions.

A

	methyl-peptides	hmSILAC library
	80	
proteins	14	17
sites	91	90
modifications	103	123
R-containing peptides	**67**	
in common with PRMT1 KD Exp.	**42**	
proteins	11	
sites	64	

R-containing peptides

PRMT1 over   43   24 36%   71   library

all sites

PRMT1 over   49   42 49%   48   library

B

PRMT1 KD | PRMT1 over.
Fwd Exp Rev Exp | Fwd Exp

FUS: GGR(di)GRGGSGGGGGGGGGGYNR / GGR(me)GR(me)GGSGGGGGGGGGGYNR
EWSR1: R(me)GGPGGPPGPLMEQMGGR
FUS: GGGDR(me)GGFGPGK
TAF15: GGYGGDR(di)GGYGGDR(di)GGYGGDR
EWSR1: GMPPPLR(me)GGPGGPGGPGGPMGR / GMPPPLRGGPGGPGGPGGPMGR(me)
EWSR1: GMPPPLR(me)GGPGGPGGPGGPMGR
EWSR1: GPR(di)GSR(di)GNPSGGGNVQHR
EWSR1: GMPPPLR(di)GGPGGPGGPGGPMGR(me)
TAF15: GYGGSQGGGR(di)GR(di)GGYDK
DDX5: TTYLVLDEADR(di)
TAF15: GGGYGGDR(di)GGGYGGDR(di)GGYGGDR
EWSR1: GMPPPLR(me)GGPGGPGGPGGPMGR(me)
FUS: GGR(di)GR(di)GGSGGGGGGGGGGYNR(me)
TAF15: SGGGYGGDR(di)GGGYGGDR(di)GGYGGK
TAF15: GGGYGGDR(di)GGYGGK / GGGYGGDR(me)GGYGGK(me)
TAF15: PK(tr)GEATVSFDDPPSAK
TAF15: GYGGSQGGGR(me)GR(me)GGYDK(me)
HNRNPH1: EGRPSGEAFVELESEDEVK(me)
HNRNPM: MGLSMER(me)
HNRNPDL: MFIGGLSWDTSK(me)
HNRNPDL: MFIGGLSWDTSKK(me)
DDX5: ELAQQVQQVAAEYCR(me)
DDX5: APILIATDVASR(me)
HNRNPH1: YGDGGSTFQSTTGHCVHMR(me)
HNRNPDL: ASR(di)GGGNHQNNYQPY
HNRNPH1: PGAGR(di)GYNSIGR(di)GAGFER
FUS: GGR(di)GR(di)GGSGGGGGGGGGGYNR
DDX17: TTSSANNPNLMYQDECDR(me)R
DDX5: APILIATDVASR(di)
NCL: EALNSCNK(me)R
FUS: GGR(di)GGGDR(di)GGFGPGK
FUS: GGPMGR(di)GGYGGGGSGGGGR(me)GGFPSGGGGGGGGGQQR(me) / GGPMGR(di)GGYGGGGSGGGGR(di)GGFPSGGGGGGGGGQQR
DDX5: R(me)GGFNTFR
ILF3: RAPVPVR(di)GGPK / RAPVPVR(me)GGPK(me)
ILF3: APVPVR(di)GGPK / APVPVR(me)GGPK(me)
FUS: SSGGYEPR(di)GR(di)
TAF15: GYGGSQGGGR(di)GR / GYGGSQGGGR(me)GR(me
TAF15: SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)
DDX17: SSQSSSQQFSGIGR(me)
TAF15: SR(di)GGYGGDR(di)GGGSGYGGDR
TAF15: DGRGPMTGSSGGDR(me)GGFK(me)
HNRNPUL1: GGGGFR(di)GR(di)GGGGGFQR

-1.5   4

log₂ (Ratio H/L)

C

PRMT1 target	Already reported in literature as PRMT1 target	Protein domains containing the identified methyl-peptides
FUS	+	LC/disorder
TAF15	+	LC
HNRNPUL1		LC
EWSR1		LC
DDX17		LC/disorder
ILF3	+	disorder closed to DSRM

Fig. 22 Profiling of the Large Drosha Complex methyl-peptides changes after PRMT1 overexpression
A) Summary of the regulated methyl-peptides identified on the LDC upon PRMT1 overexpression and summary of the hmSILAC library as a reference (left panel). Comparison of the results obtained with the hmSILAC library (right panel). **B)** Heatmap of $\log_2$ H/L ratio of methyl-peptides commonly identified in the PRMT1 depletion and overexpression experiments (red = high ratio; blue = low ratio). Ratios reported are the average ratios of input, DGCR8 IP, DDX5 IP and FUS IP for each experiment. Green arrows indicate the mono- and di-methylated forms of the peptide 465-GMPPPLRGGPGGPGGPGGPMGR-486 of EWSR1. Light Blue line indicates cases in which peptides were down-regulated in the KD experiment, while were not regulated in the overexpression one. In cases of ambiguously localized methylations on the same peptide, both possible forms of the methyl-peptide are reported. **C)** Summary table of PRMT1 targets identified in this study. Already known targets are reported. Protein domains in which identified methyl-peptides occur are reported for each target (LC = Low Complexity; DSRM = Double Strand RNA-binding Motifs)

5.4 Study of the impact of PRMT1-mediated methylation on miRNA biogenesis

The results obtained from the LDC methylation analysis confirmed that R-methylation frequently occurs on the complex and that several methyl-sites on specific subunits are modified by PRMT1. In fact, the modulation of its expression causes alteration in the methylation state of the LDC. Hence, we explored the possibility that methylation of the complex could regulate its function, with a direct impact on miRNA biogenesis.

5.4.1 The effect of PRMT1 depletion on the expression of miRNAs

We profiled by quantitative PCR (qPCR) analysis the effect of PRMT1 on the expression of a panel of miRNAs for which previous studies have reported both a crucial role in cancer [155] [156] and a deregulation upon the knock down of some LDC subunits [23, 157, 158]. Specifically, we studied the miRNAs from the miR-15a/16-1 cluster (which includes miR-15a and miR-16-1), the miR-17/92 cluster (that comprises miR-17, miR-18, miR-20, miR-19a, miR-19b and miR-92) and miR-150.

In the previous proteomic experiments, we compared the R-methylation changes between PRTM1 KD and control cells, the latter being generated by cell transduction with lentiviral particles carrying the pLKO.1 empty vector, which did not generate any sh-RNA. Importantly, the processing of exogenous sh-RNAs could saturate the RNAi cellular machinery regardless of their sequences, ultimately impinging on the synthesis of

endogenous miRNAs. We thus profiled the expression levels of a panel of miRNAs upon cell transduction with lentiviral particles carrying different sh-RNA sequences not targeting PRMT1. We tested the empty vector, a scramble sh-RNA (Scr) which sequence is not specific for any gene, a sh-RNA specific for the Luciferase gene (Luc) and, as further control, we tested also lentiviral particles not carrying any sh-RNA (empty lentiviral particles).

Cells were collected after 96h from infection and miRNAs were isolated. qPCR profiling showed that miRNA expression was comparable when cells were transduced with empty vector, Scr and with empty lentiviral particles, while significantly decreased when using sh-Luc **(Fig. 23A)**. Hence, we excluded this last construct from the analysis. In parallel the WB profiling of PRMT1 protein level indicated that PRMT1 is slightly but reproducibly increased when using the scramble construct **(Fig. 23B)**. On the basis of these results, we chose the empty vector as negative control for all subsequent analyses.

We also carefully optimized the infection protocol of HeLa cells paying specific attention to the cell confluence. miRNA expression is tightly regulated by cell to cell contacts and when cells reach confluence, their expression is enhanced in the confluent cells compared to the proliferating ones [25, 159]. As a matter of fact, HeLa growth was not affected by the PRMT1 KD within the 0-96h time window **(Fig. 16B, left panel)**, and when cells reach confluence, a significant difference of miRNA expression between confluent and non-confluent cells was observed **(Fig. 23C)**. Hence, we introduced in our protocol a cell dilution step at 24h post infection, in order to avoid cell confluence while profiling methylation changes and miRNA expression in the 72-96h post infection time window.
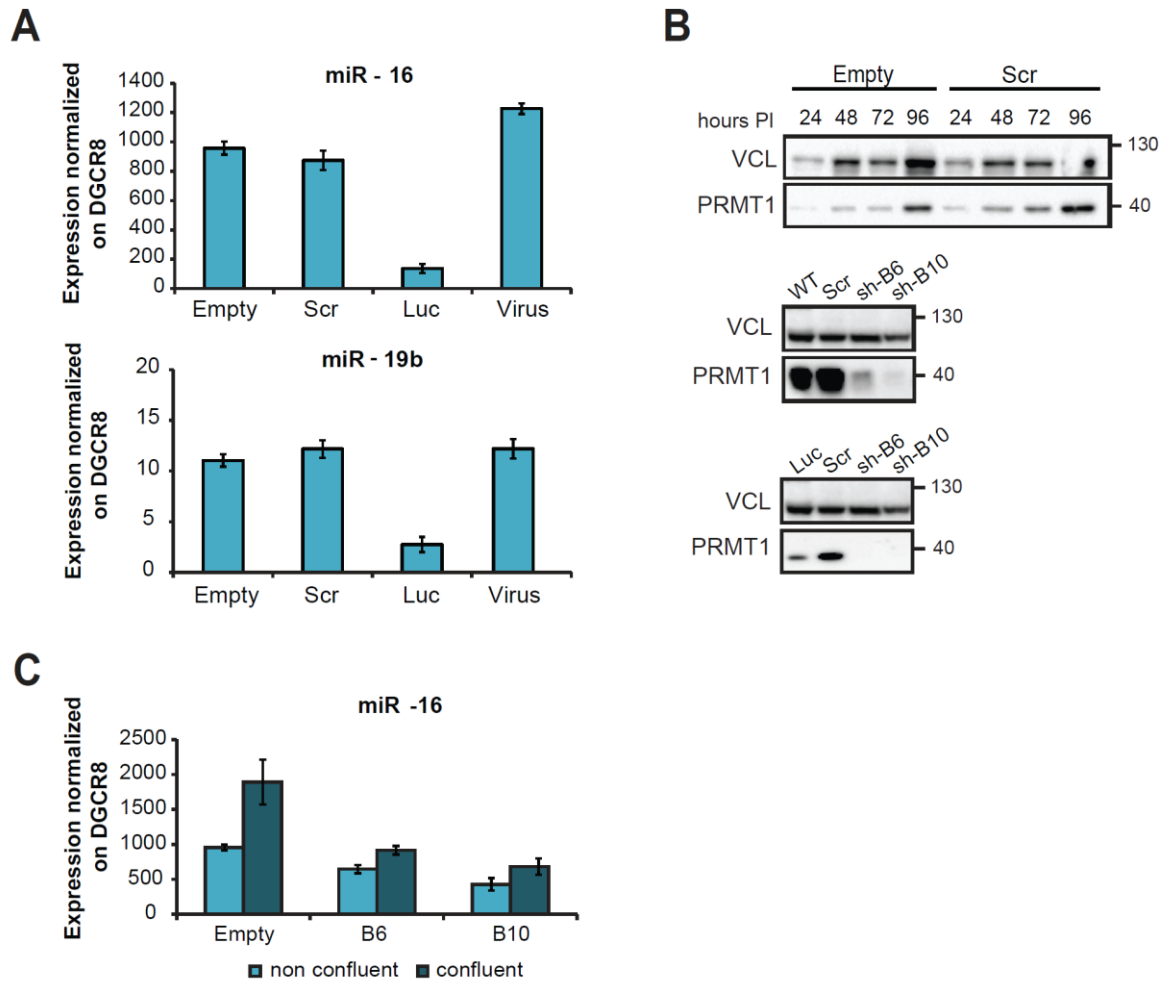
Fig. 23 Optimization of PRMT1 KD protocol for miRNA expression profiling
A) Selection of the best negative control: qPCR analysis of miR-16 and miR-19b (representative for miR-15a/16-1 and miR-17/92 clusters) upon lentiviral infection with different constructs (Empty = empty vector; Scr = sh-scramble; Luc = sh-Luciferase; Virus = lentiviral particles not carrying any sh-RNA). DGCR8 used for data normalization. **B)** WB analysis of PRMT1 level upon infection with sh-B6, sh-B10 and with the different negative control constructs (Empty, Scr, Luc); WT = HeLa wild-type. VCL was used as loading control. **C)** qPCR analysis of miR-16 expression in non-confluent and confluent cells upon PRMT1 depletion. DGCR8 was used for data normalization.

We then profiled the expression of mature miRNAs of the miR-15a/16-1 and miR-17/92

clusters upon PRMT1 KD, in a time window between 24 and 96 hours after the infection.

A reduction of miRNAs levels occurred between 72 and 96h post infection **(Fig. 24A)**, with

a striking decrease at 72h. Since we also observed changes in the LDC methylation state

in this time frame, we decide to focus on this time point for further analysis, which

confirmed that PRMT1 depletion induces significant down-regulation of all miRNAs

analyzed, with the exception of miR-150 **(Fig. 24B)**, which was selected as negative control in subsequent assays.
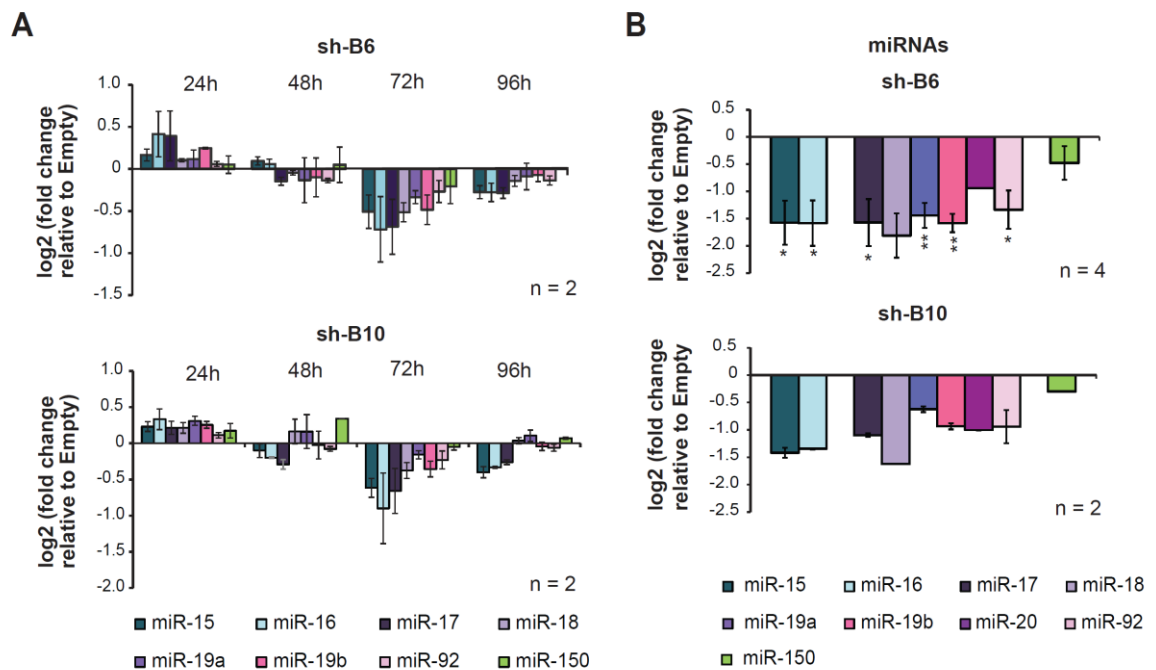
Fig. 24 PRMT1 depletion causes down-regulation of mature miRNA expression
A) Time-course expression profiling by qPCR of a subset of miRNAs upon PRMT1 KD with sh-B6 and sh-B10. Histograms represent mean ± SEM (n=2). DGCR8 was used for data normalization. **B)** qPCR profiling of miRNA upon PRMT1 depletion (72h PI) with sh-B6 and sh-B10. DGCR8 was used for data normalization. Histograms represent mean ± SEM (sh-B6: n=4, t-Test: equal variances, two-tail, * = p≤0.05, **=p≤0.01, ***≤0.001; sh-B10: n=2).

5.4.2 Global miRNA expression analysis

To assess the general effect of PRMT1 depletion on miRNA biogenesis, we initially performed a global study of miRNA expression using TaqMan Array Human microRNA cards. We profiled 600 miRNAs, 172 of which were robustly quantified in both control and PRMT1 KD cells **(Fig. 25A)**. The expression of 34 miRNAs was completely suppressed upon PRMT1 KD, while 72 were significantly down-regulated (fold change < 1.5). The results obtained for miR-17/92, miR-15a/16-1 and miR-150 expression were in line with the qPCR analysis, confirming the down-regulation of the miRNAs from the two clusters, while miR-150 was unaffected **(Fig. 25B)**. This analysis extends the result achieved by

qPCR and shows that PRMT1 is crucial for the correct miRNA biogenesis, as its depletion causes a general deregulation of miRNAs levels, with several significantly down-regulated miRNAs and a minor fraction of up-regulated miRNAs.

| miR	log2 (fold change relative to Empty)	Detected just in Empty
hsa-miR-15a	ND	+
hsa-miR-16	-0.83	
hsa-miR-17	-0.90	
hsa-miR-20a	-0.31	
hsa-miR-18a	-1.09	
hsa-miR-19a	-0.38	
hsa-miR-19b	-0.39	
hsa-miR-92a	-0.01	
hsa-miR-150	0.87	
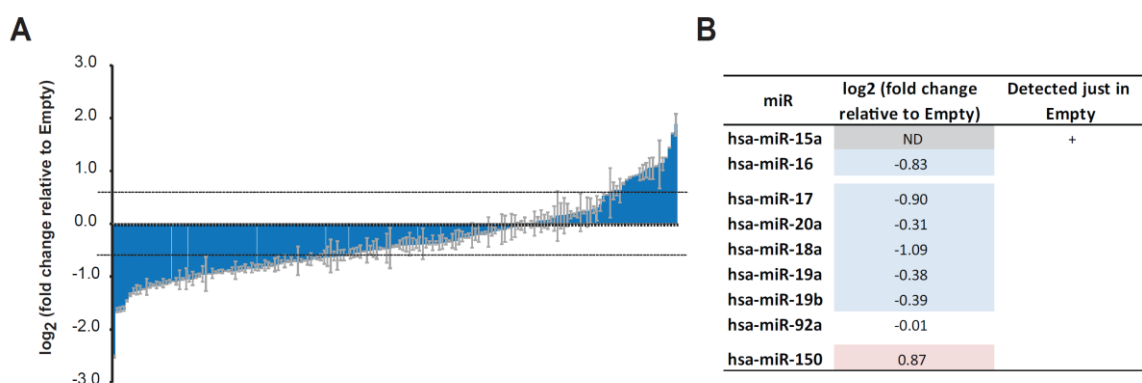
Fig. 25 PRMT1 depletion causes a general deregulation of mature miRNA expression A) Global expression analysis of miRNAs upon PRMT1 KD (sh-B6) performed with TaqMan Array Human miRNA Cards. Data were normalized on the geometric mean of a panel of housekeeping genes (mammU6, RNU44, RNU46, U6snRNA). Histograms represent mean ± SEM of miRNAs identified and quantified in two technical replicates. Significantly regulated up/down-regulated miRNAs considered when greater/lower than 1.5 fold changes. **B)** Table reporting the expression of miRNA of the miR-15a/16-1 and miR-17/92 clusters and miR-150 in the TaqMan card analysis. The average $\log_2$ ratio of two technical replicates is reported. ND = Not Detected.

We then took advantage of Small-RNA Sequencing to further extend and strengthen this global analysis, using the two sh-RNA constructs available for PRMT1 depletion, in two independent biological replicates.

We observed an overall reduction in the number of reads corresponding to canonical miRNAs in the samples with PRMT1 depletion, compared to controls **(Fig. 26A, upper panel)**, suggesting a global unbalance in miRNA expression. Interestingly, the analysis of small nucleolar RNA (snoRNA) showed a comparable amount of the reads among samples, confirming that the effect of PRMT1 depletion is specific for miRNA processing **(Fig. 26A, lower panel)**. In fact, snoRNAs are small non-coding RNAs that function as a guide for the chemical modification of ribosomal RNA and are independent from the

Drosha processing pathway [147]. Since snoRNA were not affected by the knock down of PRMT1, we used them as normalizers for the data.

The expression of the majority of the 225 miRNAs detected in control and PRMT1-depleted cells was reduced after the KD with both sh-RNAs, in the two biological replicates, with 41 miRNAs significantly down-regulated **(Fig. 26B, left panel)**. Only a small subset of miRNAs was not affected or slightly increased by the PRMT1 KD. Among the down-regulated miRNAs we identified several miRNAs from the cluster miR-15a/16-1 and miR-17/92, confirming the results from qPCR and miRNA array **(Fig. 26B, right panel)**. Moreover, the small-RNA sequencing allowed a more in-depth analysis, investigating the effect of PRMT1 on the processing of miRNAs of different origin. MicroRNAs are generally classified as "intergenic" or "intronic" based on their genomic location. Intronic miRNAs derive from the processing of introns and require spliceosomal components for their biogenesis [160] [161]. Interestingly, we did not observed any differential effect of PRMT1 KD between intronic and intergenic miRNAs **(Fig. 26C, upper panel)**, suggesting that they are all equally affected. In addition, we analyzed the effect of PRMT1 on Drosha cleavage. In fact, miRNAs can originate from the 5' (5p) or 3' (3p) arm of the hairpin, depending on Drosha activity, and an error in the cleavage could alter the relative stability of the two strands, resulting in the incorporation of the improper miRNA strand into the RISC complex [162]. However, the observation that 5p and 3p miRNAs are equally regulated suggests that PRMT1 depletion does not have a direct impact on Drosha cleavage **(Fig. 26C, lower panel)**.

Overall, these global studies of miRNA expression indicate that PRMT1 is fundamental for the correct biogenesis of miRNAs, as its depletion causes their massive down-regulation, even if it does not directly affect Drosha cleavage efficiency.
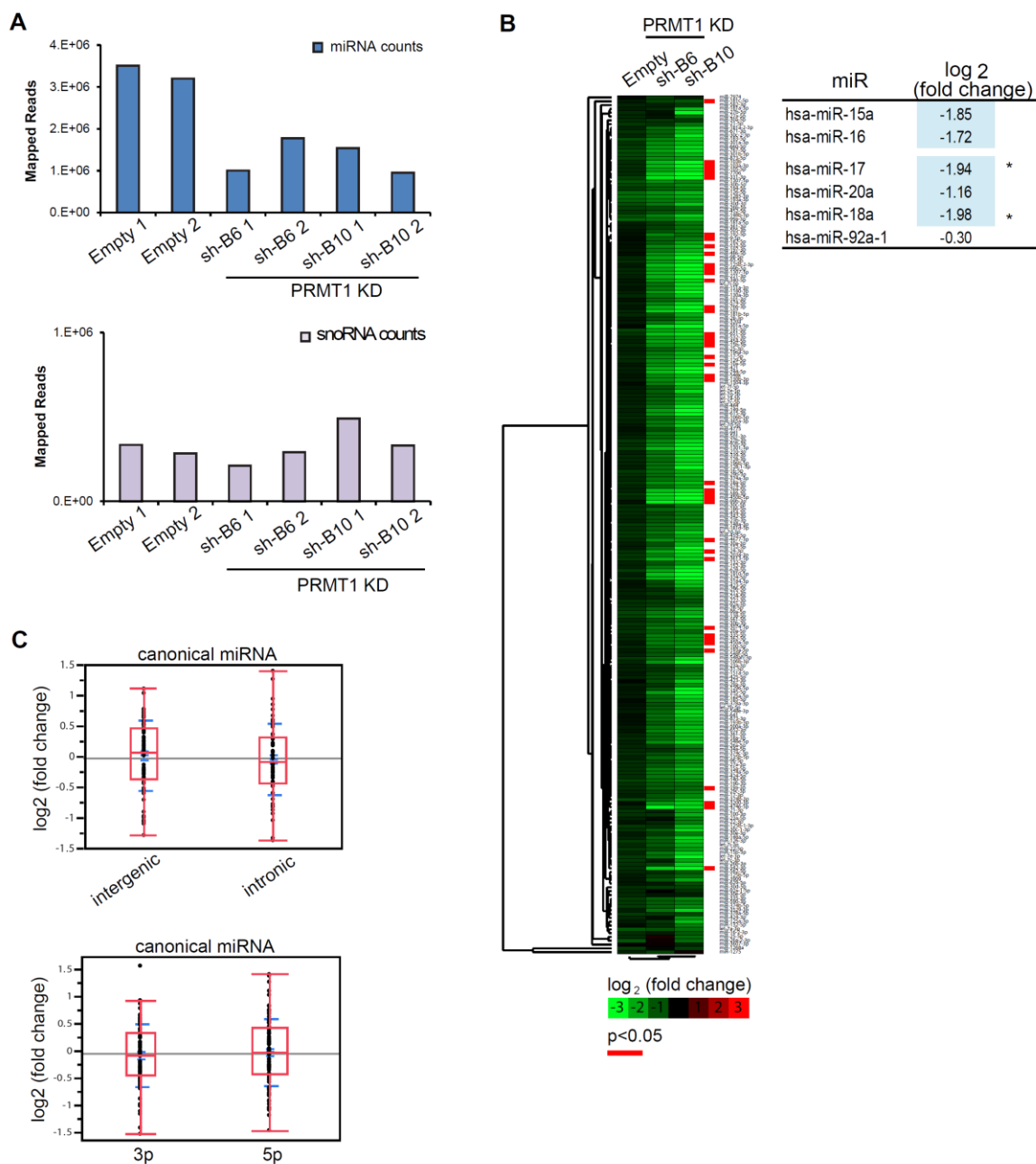
Fig. 26 PRMT1 depletion causes global down-regulation of mature miRNA expression

Global miRNA expression analysis by Small-RNA sequencing. **A)** Upper panel: histograms representing the number of reads corresponding to canonical miRNAs in control (Empty 1-2) and PRMT1 knock down (sh-B6 1-2, sh-B10 1-2) samples. Lower panel: histogram representing the number of reads corresponding to snoRNAs in control (Empty 1-2) and PRMT1 knock down (sh-B6 1-2, sh-B10 1-2) samples. The experiment was performed using two biological replicates (1-2); PRMT1 KD was obtained with two sh-RNAs (sh-B6, sh-B10). **B)** Left panel: the heatmap shows log$_2$ fold change of miRNAs identified in control (Empty) and PRMT1 KD (n = 225); (red = high fold change; green = low fold change). Data are the average of two independent biological replicates. Red lines indicate significantly down-regulated miRNAs (41); (t-Test: equal variances, two-tail, * = p-value < 0.05). Right panel: table reporting the expression of miRNA of the cluster miR-15a/16-1 and miR-17/92 detected by Small-RNA sequencing. Log$_2$ fold change of the

average of the two sh-RNAs is reported (* = p-value < 0.05). **C)** Upper panel: box plot show the distributions of intergenic and intronic miRNAs upon PRMT1 KD. Data are the average of the two independent biological replicates using both sh-B6 and sh-B10. Data were normalized on miRNAs RPM. Lower panel: box plot show the distributions of 5p (n = 79) and 3p (n = 60) miRNAs upon PRMT1 KD. Data are the average of the two independent biological replicates using both sh-B6 and sh-B10. Data were normalized on miRNAs RPM.

5.4.3 Primary and precursor miRNA analysis upon the depletion of PRMT1

Encouraged by the evidence that PRMT1 depletion affects miRNA expression, we hypothesized that the altered methylation state of the Large Drosha Complex caused by PRMT1 down-regulation may affect the processing step of primary- to precursor- miRNAs catalyzed by the complex, thus altering the relative ratio of pri- to pre- levels. The ultimate effect would be measured as a reduction of mature miRNAs **(Fig. 27A)**. Therefore, we looked more specifically at the effect of PRMT1 depletion on this processing step and designed specific qPCR primers that allow profiling the different miRNA intermediates of miR-15a/16-1 and miR-17/92 and that are described in the Materials and Methods section, paragraph 4.11.

We did not observe significant changes in the primary transcript levels upon PRMT1 KD with both sh-RNAs **(Fig. 27B).** However, interestingly, we found a significant and reproducible down-regulation of several pre-miRNAs **(Fig. 27C)**. These results confirm that the depletion of PRMT1 not only reduces mature miRNA expression but also affects the pri-to-pre miRNA processing step. In addition, the observation that the expression of the primary transcripts is not diminished but rather slightly increased excludes the possibility that the reduction of pre-miRNAs may be a consequence of the effect of PRMT1 depletion on pri-miRNAs transcription. The slight increase of pri-miRNAs could reflect the accumulation of the un-processed primary transcripts due to the impairment of the LDC-mediated processing **(Fig. 27A)**.
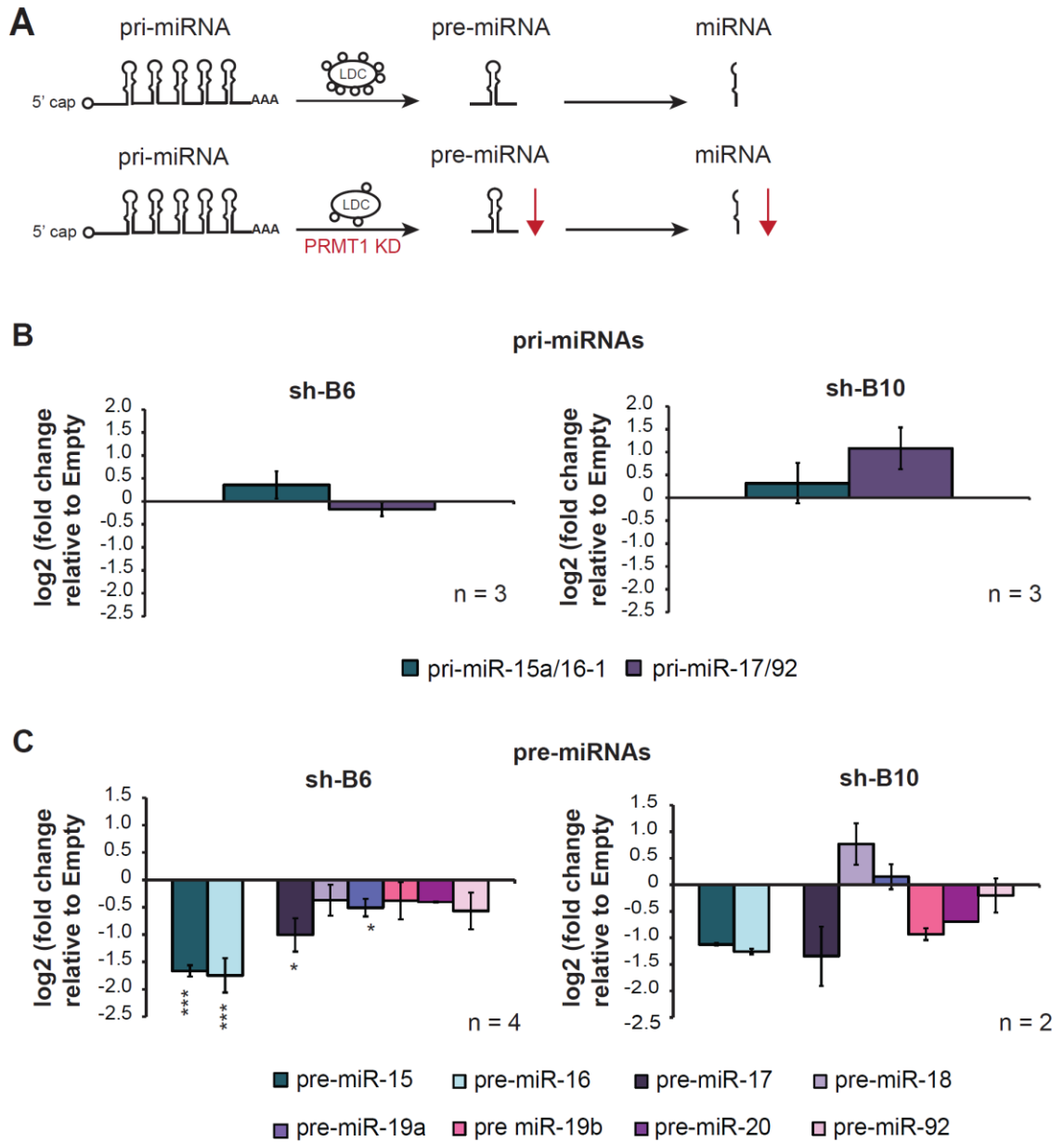
Fig. 27 The processing of primary to precursor miRNAs is impaired by PRMT1 depletion

A) Schematic representation of miRNA biogenesis and hypothesized PRMT1 role within this process. When the LDC-methylation is unaffected, pri-miRNAs are normally processed to pre-miRNAs (upper panel). When the LDC methylation pattern is altered, the pri- to pre-miRNA processing is impaired and pre- and mature miRNAs levels are reduced (lower panel). **B)** qPCR profiling of the pri-miRNAs of the clusters miR-15a/16-1 and miR-17/92 upon PRMT1 KD with sh-B6 and sh-B10. DGCR8 was used for the normalization of the data. Histograms represent mean ± SEM (n =3; t-Test, equal variances, two-tail, * = p≤0.05, **=p≤0.01, ***≤0.001). **C)** qPCR profiling of the pre-miRNAs of the clusters miR-15a/16-1 and miR-17/92 upon PRMT1 KD with sh-B6 and sh-B10. DGCR8 was used for the normalization of the data. Histograms represent mean ± SEM (sh-B6: n=4: t-Test, equal variances, two-tail, * = p≤0.05, **=p≤0.01, ***≤0.001. sh-B10: n=2).

We focused on the cell compartment where the processing step of the miRNA biogenesis catalyzed by the LDC occurs, to further confirm that the activity of the complex is impaired in the absence of PRMT1. Therefore, we analyzed each miRNA intermediate product in the different cellular compartments, namely cytosol, nucleosol and chromatin fraction. The same amount of control and PRMT1-depleted cells was subjected in parallel to total cell lysis or to sub-cellular fractionation **(Fig. 28A)**. Total RNA (including miRNAs) was then extracted from all samples. Pri-, pre- and mature miRNAs of the clusters miR-15a/16-1 and miR-17/92 were isolated from the total extract and profiled by qPCR, confirming the accumulation of the pri-miRNAs and the corresponding down-regulation of pre- and mature miRNAs upon PRMT1 depletion **(Fig. 28B, left panel)**. We then focused on the pri- to pre- processing step in the nuclear fraction, which is the compartment where the LDC is active. We profiled pri-miRNAs in the chromatin fraction, where they are firstly located being transcribed by RNA polymerase II. Subsequently the pri-miRNAs are processed by the LDC into pre-miRNAs, which are exported from the nucleosol into cytoplasm, where they undergo further processing. We observed a 10% increase of pri-miRNAs in the chromatin fraction upon PRMT1 KD **(Fig. 28B, right panel)**, corroborating the hypothesis that the depletion of the enzyme correlates with the impairment of the processing. In line with this observation, we observed decreased pre-miRNAs level in the nucleosol, followed by a slight increase of their levels in the cytoplasm **(Fig. 28B, right panel)**. This result led us to speculate that, when the processing is impaired, the export in the cytoplasm of residual pre-miRNAs might be favored.
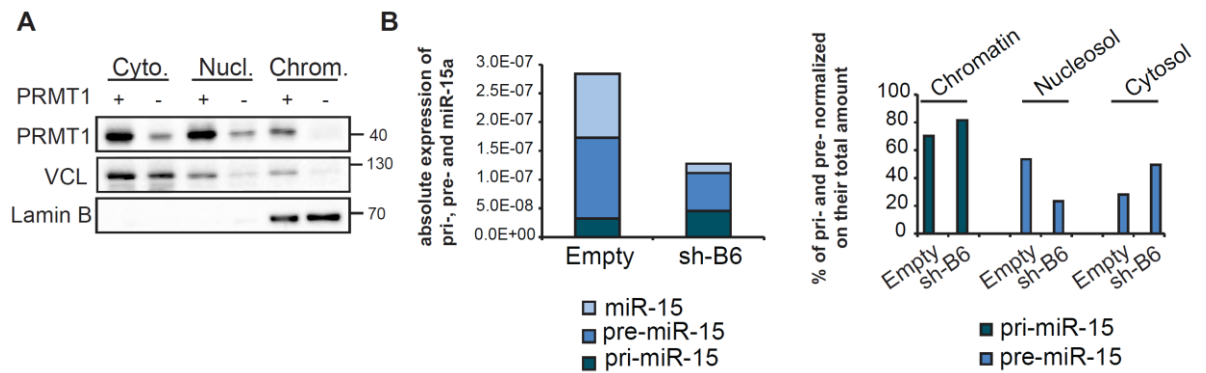
Fig. 28 Pri-, pre- and mature- miRNA analysis in specific cell compartments
A) WB validation of the efficiency of cell fractionation into cytosolic (Cyto.), nucleosolic (Nucl.) and chromatin (Chrom.) fractions in control and PRMT1 KD cells. VCL and Lamin B were used as cytosolic and chromatin markers, respectively. **B)** Left panel: qPCR analysis of pri-, pre- and mature miR-15a (representative of all miRNAs analyzed) in total cell extract from control (Empty) and PRMT1-depleted cells (sh-B6). Absolute expression of each miRNA form is reported. Right panel: qPCR analysis of pri- and pre-miR-15a in specific cell compartments in control (Empty) and PRMT1-depleted cells (sh-B6). The percentage of pri- and pre-miR-15a normalized on their total amount in all cell compartments is reported.

5.5 Investigation of the molecular mechanisms linking protein methylation to altered miRNA biogenesis

Protein methylation regulates several aspects of cellular homeostasis, essentially through three mechanisms: either altering protein sub-cellular localization, or modulating protein-protein and protein-nucleic acid interactions [46, 163]. We took into account all these mechanisms to try to dissect the molecular link between PRMT1-regulated arginine methylation on the LDC and altered miRNA biogenesis.

5.5.1 Analysis of the sub-cellular localization of Large Drosha Complex subunits in the absence of PRMT1

We performed a WB analysis of the sub-cellular localization of several LDC components upon PRMT1 depletion, to assess whether we could detect changes in their subcellular distribution. None of the subunits tested show significant changes in their localization in the KD sample, compared to control **(Fig. 29)**, which indicated that the absence of PRMT1

and the related changes in methylation do not affect the cellular distribution of the majority of the LDC components, at least the one profiled by WB.
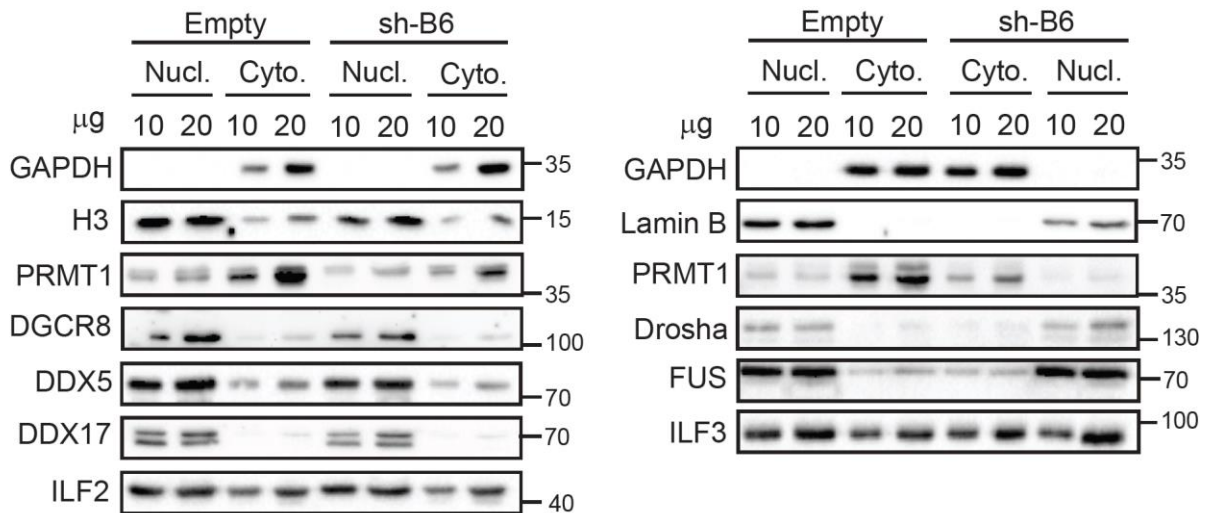
Fig. 29 The absence of PRMT1 does not affect the cellular distribution of several subunits of the LDC
WB analysis of the sub-cellular localization of the LDC proteins, in control (Empty) and PRMT1 KD (sh-B6) cells. VCL and GAPDH were used as cytosolic marker; Lamin B and H3 were used as nuclear markers. Increasing amounts of extracts (10 and 20 μg) were used.

5.5.2 The role of methylation in the regulation of protein-RNA or protein-protein interactions

The interaction among DGCR8 and Drosha, the core components of the Large Drosha Complex, occurs through a direct protein-protein contact [18], while the binding of other subunits of the complex with the Microprocessor can be mediated both by RNA [17] or by protein-protein interaction [164]. However, a comprehensive study of how accessory proteins associate to the Microprocessor within the LDC is still absent.

To understand whether RNA mediates interactions within the complex, we performed LDC immunoprecipitation using DCGR8 as bait, starting from nuclear extracts treated or not with RNase A. The binding of several components was profiled by WB analysis. The efficiency of the affinity-purification of DGCR8 was not affected by RNase A treatment **(Fig. 30A, left panel)** as well as the DGCR8/Drosha interaction, as expected. Moreover,

we did not observe relevant changes in the interaction of other subunits, such as DDX17, DDX5 and ILF2, with the bait; from this we can conclude that their interaction with the Microprocessor is not mediated by RNA **(Fig. 30A, right panel)**.

However, it is known that RNA mediates the interaction of DGCR8/Drosha with some LDC components such as Nucleolin [17]. Thus, to extend the analysis also to subunits which interaction was not assessed by WB, we are performing the proteomic experiment illustrated in **Fig. 30B**. We labeled wild-type (WT) HeLa in standard heavy and light SILAC media, up to full incorporation. Heavy and light cells were harvested, separated into nuclear and cytosolic fractions and the heavy-labeled nuclear extract was treated with RNase A, while the light-labeled one was left untreated, in a 'forward' experimental setup. We performed two DGCR8 IPs in parallel, using both RNase A-treated and untreated extracts as input. The immunoprecipitated samples were then mixed 1:1, analyzed by geLC-MS/MS and raw data analyzed by MaxQuant software, as described in Materials and Methods (paragraph 4.6). Within this experimental setup, proteins whose interaction with DGCR8 is RNA mediated will show a decreased H/L SILAC ratio, while interactions that are not affected by the treatment will produce H/L ratio equal 1. If we will discover that the interaction of some subunit is RNA-mediated, we plan to setup experiments, which will allow assessing whether the modulation of R-methylation on specific subunits might affect their recruitment on the complex.
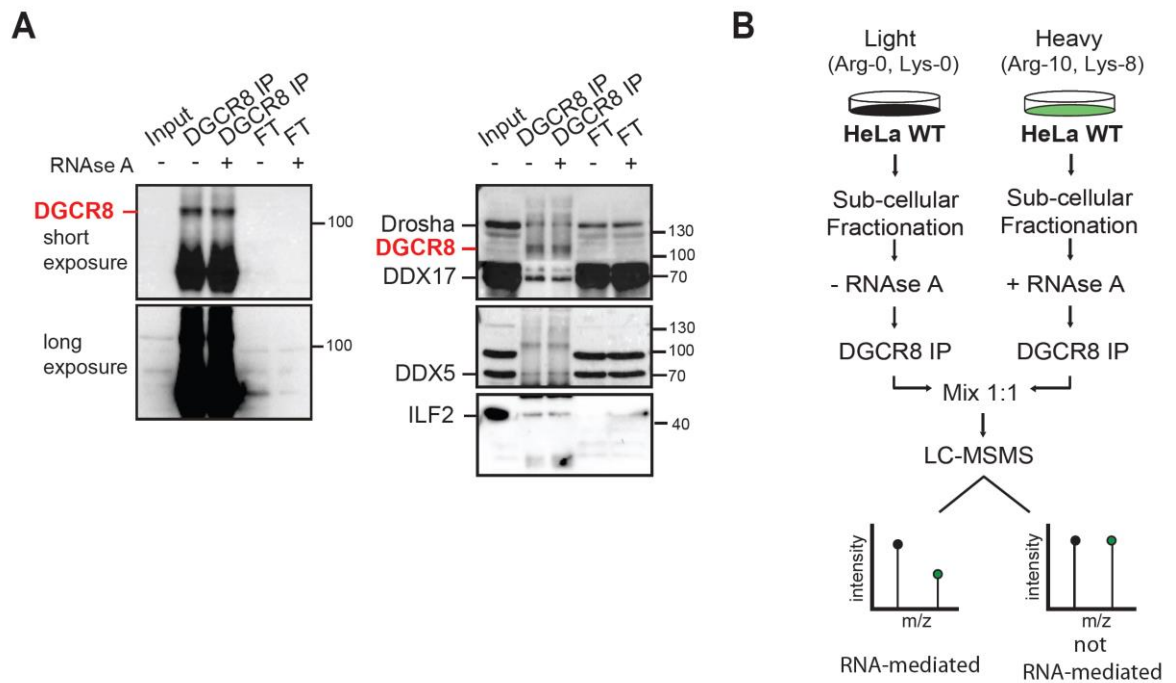
Fig. 30 Analysis of the role of RNA in modulating the interactions among subunits of the LDC

A) Immunoprecipitation of DGCR8 (indicated in red) from nuclear extracts treated or not with RNAse A, followed by WB analysis. FT = Flow-through. **B)** Workflow of the SILAC-based proteomic experiment: the 'forward' experiment is represented. Upon LC-MS/MS analysis, a decreased H/L ratio indicates a loss of the interaction (RNA-mediated); unchanging H/L ratios represent proteins whose interaction with DGCR8 is not affected by the loss of RNA (not RNA-mediated).

Based on the observation that methyl-sites that are regulated by PRMT1 occur mainly within LC/disordered regions of various LDC subunits **(Fig. 20C and Fig. 22C)**, we evaluated whether R-methylation affects protein-protein interactions within the complex. We performed DGCR8 IPs starting from nuclear extracts of PRMT1-depleted and control cells, followed by WB analysis. The amount of immunoprecipitated DCGR8 was comparable in the two functional states. However, the co-immunoprecipiation of other subunits known to directly interact with DGCR8, such as DDX17, DDX5 and ILF3 [19, 165], was altered upon PRMT1 knock down **(Fig. 31A)**, suggesting that the modulation of methylation exerted by the enzyme depletion may influence the interaction among proteins of the complex. We extended this analysis to the whole complex by performing a SILAC-based proteomic experiment, which is also ongoing **(Fig. 31B)**. We labeled HeLa

by growing them in standard SILAC media. Upon full incorporation, cells were infected with either the empty vector or with sh-B6, harvested and fractionated in nuclear and cytosolic fractions. We performed in parallel DGCR8 IPs starting either from control or PRMT1-depleted nuclear extracts. The immunoprecipitated materials were then mixed in equal amount (1:1), analyzed by geLC-MS/MS and raw data analyzed by MaxQuant software, as described in Materials and Methods (paragraph 4.6). Alterations in the H/L SILAC ratio of the proteins of the LDC will reflect changes in their interaction with DGCR8 in the absence of PRMT1.
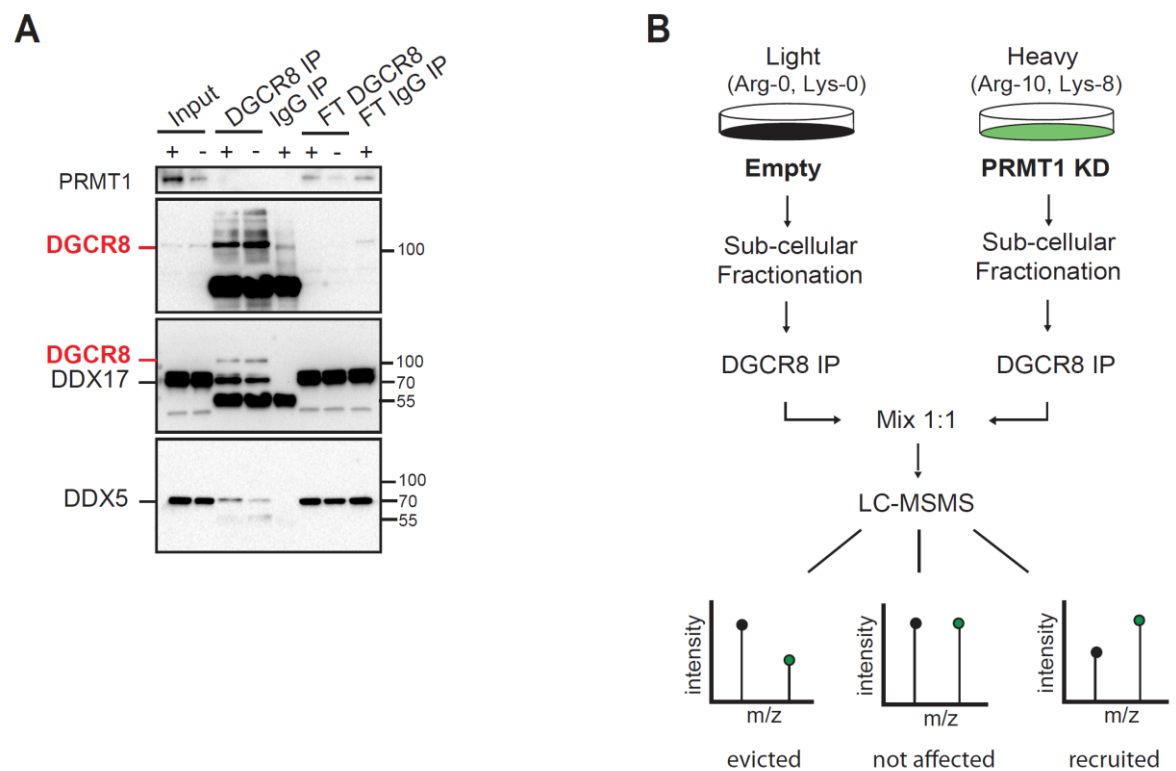
Fig. 31 Study of the role of methylation in the regulation of protein-protein interactions within the LDC
A) Immunoprecipitation of DGCR8 (indicated in red) from nuclear extracts of control (empty) and PRMT1-depleted (sh-B6) cells, followed by WB analysis. The co-IP of DDX17 and DDX5 was also assessed. FT = Flow-through; IgG IP was used as negative control.
B) Workflow of the SILAC-based proteomic experiment: the 'forward' experiment is represented. Upon LC-MS/MS analysis, an increased H/L ratio indicates proteins that are recruited to DGCR8 upon PRMT1 KD; decreased H/L ratios indicates a loss of the interaction (evicted); unchanging H/L ratios represent proteins whose interaction with DGCR8 is not affected by the loss of PRMT1.

Based on the results of the proteomic experiments we are performing, we will have insights on how the accessory proteins of the LDC interact with each other and/or with the Microprocessor, and how R-methylation may regulate the structure of the Large Drosha Complex, ultimately impacting on its activity.

5.6 Impact of other PRMTs on miRNA biogenesis

The PRMTs family comprises several enzymes besides PRMT1, thus we wanted to investigate if they might also play a role in the regulation of miRNA biogenesis.

In parallel to the set up of KD and overexpression protocols to modulate PRMT1 expression, we established protocols to deplete other important PRMTs, such as PRMT4, PRMT6, PRMT5 and PRMT7. For each enzyme we tested two different sh-RNA sequences, which were all efficient in depleting the respective target, except for PRMT5 sh-446 **(Fig. 32A)**. Given the crucial role of PRMTs in the regulation of cell cycle [50, 166] we studied cell growth after the depletion of each individual enzyme, in a time window between 4 and 8 days post infection. Similarly to the results obtained upon PRMT1 KD, we observed that the depletion of PRMT4, PRMT6 and PRMT7 strongly reduces cell viability starting from 4 days after the infection, while PRMT5 knock down results in a less prominent reduction **(Fig. 32B)**. We then profiled global methylation changes in the same time window, by WB analysis of total extract, using anti pan-methyl antibodies that recognize asymmetric (ASYM24) or symmetric (SYM10) di-methylation: in fact, PRMT4 and PRMT6 catalyze asymmetric di-methylation and PRMT5 and PRMT7 asymmetric di-methylation of their targets [50]. Four days upon infection, all enzymes are completely depleted (except with PRMT5 sh-446), and changes in di-methylations compared to control cells are clearly detectable **(Fig. 32C)**.
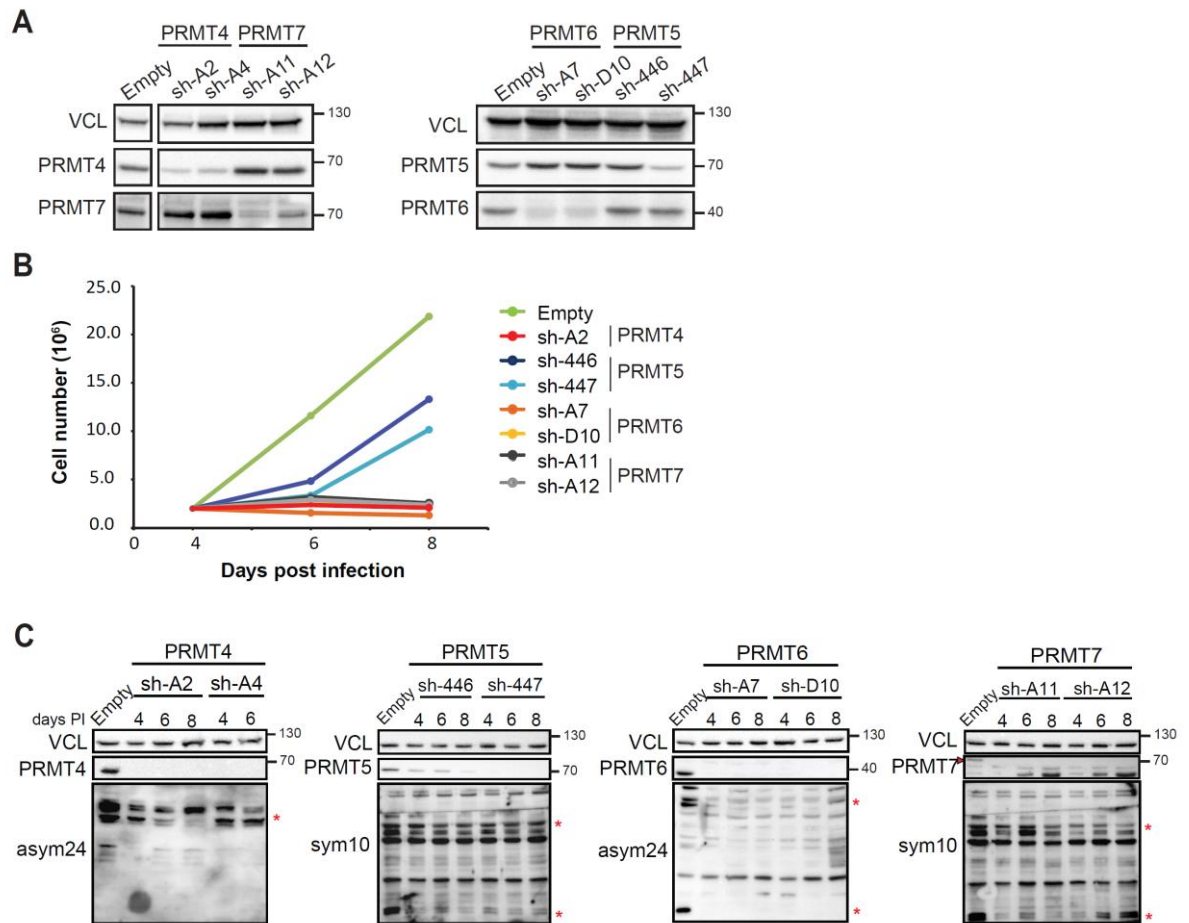
Fig. 32 Set up of the knock down of other PRMTs
A) WB analysis of PRMT4, PRMT5, PRMT6 and PRMT7 expression upon their depletion using two sh-RNA sequences per protein (PRMT4: sh-A2, sh-A4; PRMT5: sh-446, sh-447; PRMT6: sh-A7, sh-D10; PRMT7: sh-A11, sh-A12). Empty vector was used as negative control. VCL was used as loading control. **B)** Growth curves of control and PRMTs-depleted cells between 4 and 8 days PI. **C)** Time course profiling of PRMTs depletion and asymmetric (asym24) and symmetric (sym10) di-methylation between 4 and 8 days PI. VCL was used as loading control. Red asterisks highlight changes in di-methylation.

Starting from the evidence that changes in global protein methylation is observed when most PRMTs are depleted, we asked whether their depletion affects also miRNA biogenesis in a similar manner and to a similar extent as PRMT1. We profiled the expression of the miRNAs of miR-15a/16-1, miR-17/92 clusters and of miR-150 by qPCR and observed a reduction of mature miRNAs 72h after PRMT1 depletion, paralleled by the accumulation of their corresponding pri-miRNAs. However, the effect of PRMT4, PRMT5 and PRMT7 depletion on the expression of both primary and mature miRNAs is not similar to that of PRMT1, at least to the subset of miRNAs tested **(Fig. 33A)**, and the few

changes observed may be miRNA- or PRMT- specific. However, we observed a significant down-regulation of all mature miRNAs analyzed upon PRMT6 depletion using both sh-RNAs **(Fig. 33B)**. PRMT6 is an important transcriptional regulator, acting as either co-activator or repressor depending on the different R-sites that methylates (H3R42me2a and H3R2me2a, respectively) [76, 77, 80]. Therefore, we focused on the analysis of the expression levels of pri-miRNAs in order to evaluate whether the effect on mature miRNAs may be due to transcriptional inhibition rather than an impairment of LDC function. Pri-miR-15a/16-1 and pri-miR-17/92 were slightly down-regulated upon PRMT6 KD **(Fig. 33C)**, suggesting that this enzyme might have an impact on the primary miRNAs transcription. Yet, these results do not exclude that also PRMT6 might affect miRNA biogenesis altering the methylation state of the LDC.
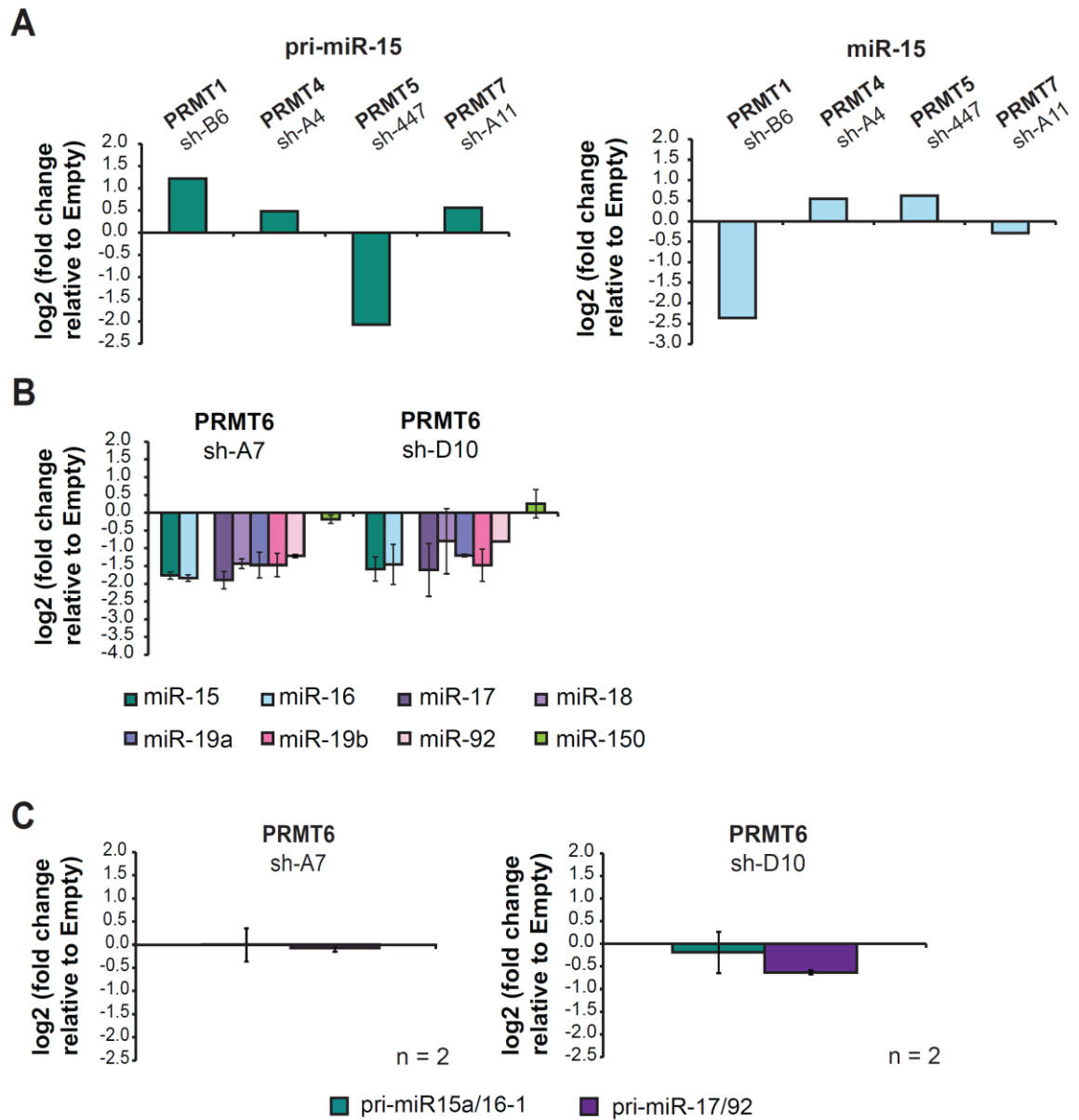
Fig. 33 qPCR analysis of miRNAs upon the KD of other PRMTs
A) qPCR analysis of pri- and mature miR-15a (representative of all the miRNAs analyzed) upon PRMT1, PRMT4, PRMT5 and PRMT7 KD. DGCR8 was used for data normalization. **B)** qPCR analysis of mature miRNAs of the clusters miR-15a/16-1 and miR-17/92 and miR-150, upon PRMT6 KD with sh-A7 and sh-D10. DGCR8 was used for data normalization. **C)** qPCR analysis of pri-miR-15a/16-1 and pri-miR-17/92, upon PRMT6 KD with sh-A7 and sh-D10. DGCR8 was used for data normalization.

6. DISCUSSION

In this study, we employed MS-based proteomics to investigate the role of arginine methylation in regulating the activity of the Large Drosha Complex and consequently the process of miRNA biogenesis.

We initially combined the hmSILAC approach with co-immunoprecipitations of the LDC to carry out a thorough characterization of un-ambiguously identified methylations occurring on the diverse components of the complex. We identified 90 different methylated sites on the complex, the majority of which occurs on arginine residues, confirming the highly methylated nature of the LDC.

Almost all the subunits of the complex are modified, some of them on multiple sites. However, interestingly, the core components, DGCR8 and Drosha, are not. It has been reported that they interact with each other through a direct protein-protein interaction, and with the primary transcript through specific RNA Binding Domains (RBDs) (see Introduction, paragraph 2.1). However, this result suggests that methylation may affect the binding of the accessory subunits to the Microprocessor rather than the structure/activity of the Microprocessor itself, since it can regulate both protein-protein and protein-RNA interactions.

In a very recent work from the Nielsen group, R-methylated sites have been identified on several protein complexes, including the LDC and other complexes involved in RNA metabolism [34]. On the Large Drosha complex, they identified methyl-sites on all the proteins of the complex, with the exception of DGCR8. However, in this study the authors did not use orthogonal validation approaches for high-confidence identification of methylated sites, such as hmSILAC or its variants. Thus, although the majority of the methyl-sites identified by Nielsen and colleagues has been characterized also by us, upon hmSILAC validation, the confidence of the remaining sites identified by them remains uncertain. Indeed, different groups have shown that label-free approaches are not ideal for global analysis of methylation, since they lead to high FDR in methyl-assignment [44, 136]. Since the generation of robust and reliable data is fundamental for the subsequent

biological and functional studies, and, at present, the most efficient strategy to decrease the FDR in large-scale MS-based methylation analyses is the hmSILAC, we decided to exploit this strategy in our study. Hence, we built a library of confidently annotated methyl-sites for the LDC and used it as a reference for the analysis of changes in the methylation state of the complex after modulation of PRMT1 expression.

For this purpose, we combined SILAC with the affinity enrichment of the LDC. The observation that the amount of methyl-peptides identified after the enrichment of the complex is higher than in the input confirmed that the immuno-affinity enrichment strategy adopted is an efficient method for the characterization of methylations of proteins and complexes of interest. This result also highlights the importance of overcoming the sub-stoichiometric nature of methylation by adopting efficient enrichment strategies prior to MS, as discussed in the Introduction (paragraph 2.6 and 2.6.2). Quantification analysis of asymmetric di-methylated arginine (ADMA), symmetric di-methylated arginine (SDMA) and mono-methylated arginine (MMA) at steady state, in MEFs, showed that ADMA is the predominant methylated arginine residue [138]. However, in the majority of the global proteomics studies mono-methylated arginine is the most represented modification. This result is likely due to the fact that these analyses are based on the specific enrichment of ADMA, SDMA and MMA peptides by antibodies that recognize the differently modified residues, and among them the MMA antibody has proven to be the most efficient [132]. In our study we enriched the proteins of the LDC instead of the methylated peptides, and this leads to the annotation of an equal amount of ADMA and MMA at steady state **(Fig. 14B)** and also to the identification of more ADMA compared to MMA on regulated methyl-peptides after PRMT1 depletion **(Fig.20, appendix 2)**. These observations suggest that the co-IP approach allows overcoming possible biases introduced by the use of pan-methyl antibodies.

Comparing the PRMT1-regulated methyl-peptides identified after the depletion of the enzyme with the hmSILAC library, we found that the 67% of the methyl-sites (53) were already annotated. The sites not reported in the hmSILAC reference might be either false positive identifications or true methyl-sites, missed by the initial screening. Indeed

drawbacks of shotgun proteomics are linked to its recognized bias towards high abundance peptides and to its stochastic nature that do not allow identifying all proteins/PTMs in a sample, in a single run. Considering that our library was built performing five different affinity enrichments of the complex, it is likely that some methyl-sites were not detected in these experiments. Hence, the manual inspection at the MS/MS level of these modified peptides will be essential to understand whether they are true sites, or artifacts.

The profiling of the methyl-proteome upon PRMT1 KD suggested that several subunits of the complex are PRMT1 targets and that the depletion of the enzyme causes changes in the methylation state of the LDC. However we cannot exclude the possibility that other PRMTs may compensate or compete for the loss of the major enzyme. Indeed, the depletion of PRMT1 seems to cause a complex response in terms of methylation changes on the LDC. Although on the one hand we observed the down-regulation of several mono- and di-methylated peptides after PRMT1 depletion, in line with the enzymatic activity of this type-I enzyme, on the other hand a subset of peptides was up-regulated, among them mono- and di-methylated were equally distributed. This finding could be partially explained with the fact that the global decrease of R di-methylation after the loss of PRMT1 is associated with an increase in mono-methylation, as previously described by Dhar and colleagues [138]. However, the observed increase of di-methylated peptides cannot be explained in light of this evidence and rather suggests the existence of compensatory/scavenging effects by other PRMTs. In fact, it has been shown that also symmetric di-methylation increases both upon PRMT1 loss [138] and after the inhibition of type-I PRMTs [139]. Since our MS-based analysis did not discriminate between asymmetric and symmetric di-methylation, we cannot exclude that type-II enzymes are compensating the decrease of ADMA, resulting in the up-regulation of some di-methylated methyl-peptides.

To help clarifying this issue, we carried out a SILAC-based proteomic experiment overexpressing PRMT1. The methylated peptides identified in this experiment were directly compared to those found after PRMT1 depletion, resulting in 42 common methyl-

peptides **(Fig. 12B)**. As expected, 10 methyl-peptides that were not regulated by depletion were not affected by overexpression. Five peptides showed the same trend in both KD and overexpression experiments, while the remaining were considered genuine PRMT1 targets; they either displayed an opposite trend between depletion and overexpression experiments or they were up-/down-regulated after the PRMT1 KD but not changing when the enzyme was overexpressed. A possible explanation for the latter observation is that they might be already methylated by the endogenous PRMT1 and thus cannot be further modified, resulting in 'non-changing' ratios. Here, we considered PRMT1 targets also methyl-peptides that were not reported in the hmSILAC library, since they were regulated consistently after both PRMT1 KD and overexpression.

When comparing the methylated peptides identified after PRMT1 overexpression with the reference library, we observed only a 50% overlap. In fact, 49 methylated-sites, occurring on 12 different subunits of the LDC, were not annotated in our library. This could be due not only to the fact that these sites might represent false positive identifications (and need to be further validated at the MS/MS level) but also to the fact that both endogenous and ectopic PRMT1 are present in the cell and they might methylate R residues that are not modified at the basal level, thus leading to the identification of new sites. PRMT1, as well as other PRMTs, is overexpressed in cancer [50], thus understanding whether the overexpression of a certain PRMT leads to an increase of the amount of methylated sites on its substrates might be particularly informative, also in a translational perspective.

The concomitant overexpression and depletion of one or more PRMTs at the same time could be particularly useful, for multiple purposes: first, to minimize the compensatory/compensation mechanisms among PRMTs; second to avoid the saturation effects due to the presence of the endogenous PRMT; third, to assess systematically the contribute of each enzyme to the cellular and LDC methyl-proteomes. Recently, the Structural Genomics Consortium of the University of Toronto has developed selective inhibitors of all type-I enzymes (including PRMT1, 3, 4, 6 and 8), of CARM1 and of all type-II enzymes (PRMT5 and PRMT7) [139, 167-169]. We will employ them, alone or in

combination, and in association with distinct PRMTs overexpression to dissect more systematically the PRMT1 targets in the LDC.

When we performed methylation analysis to identify PRMT1 targets on the complex, we observed ambiguous cases of modification assignments on some of the peptides. This is a common issue related to PTMs analysis by mass spectrometry (see Introduction, paragraph 2.6), which is particularly relevant in the case of methylation, since there are no tailored algorithms to statistically assess the localization of a methyl-site on a peptide, as in the case of other PTM such as phosphorylation. The most frequent cases are those of peptides for which MaxQuant software assigns a certain methylation type to a specific residue with the highest localization probability score, although the MS/MS information is not sufficient to confidently localize the detected Δmass **(Fig. 34)**.
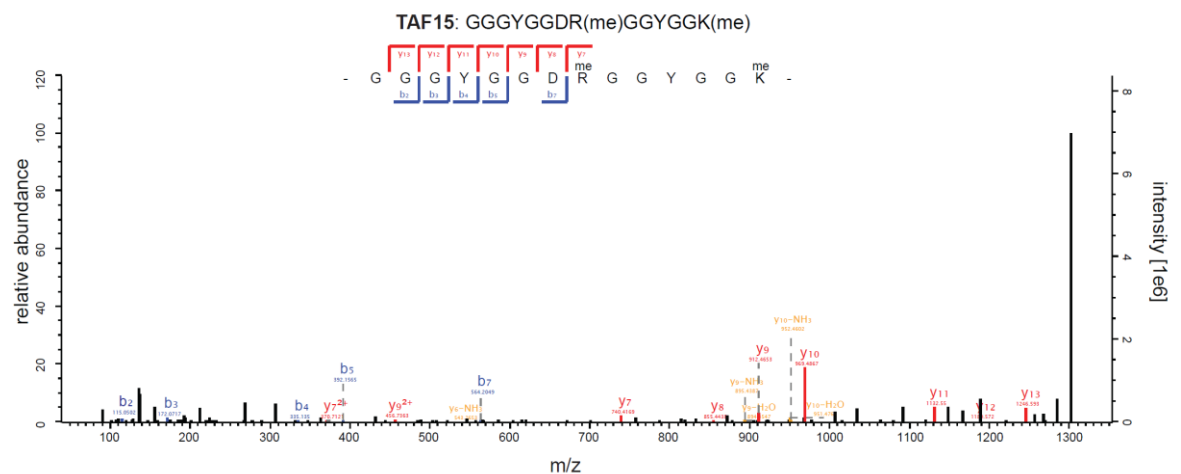
Fig. 34 Example of ambiguous methylation assignment
HCD MS/MS spectrum for the peptide 562-GGGYGGDRGGYGGK-576 of TAF15 from the proteomic experiment with PRMT1 overexpression **(Fig 22B)**. MaxQuant assigns the Δmass of 28 Da to two mono-methylations on R and K, even if the peptide is not entirely fragmented. (Andromeda score = 59.59; localization probability score = 1).

When methylated peptides with ambiguous methylation assignments resulted consistently regulated after PRMT1 depletion and overexpression, we decided to consider them as substrates of PRMT1 **(Fig. 22B)**. Taking into account these cases, we can define PRMT1 targets at a 'peptides level resolution', without precise site localization.

After establishing that the modulation of the expression of PRMT1 alters the methylation state of the Large Drosha Complex, we explored the possibility that methylation could regulate its function, impairing miRNA biogenesis. We initially profiled mature miRNA expression upon the depletion of PRMT1 and observed a global massive down-regulation of mature miRNAs. Even if the majority of miRNAs were down-regulated by the KD, a small subset of miRNAs was not affected or slightly up-regulated. This might reflect the fact that depletion of PRMT1 causes methylation changes on several subunits of the complex that are known to regulate the Microprocessor in opposite way. In fact, some accessory proteins are already recognized as positive regulators of miRNA biogenesis, such as DDX5 and DDX17, while others act as negative regulators, such as ILF3 and ILF2 (see paragraph 2.2). Therefore, the alteration of methylation induced by PRMT1 KD could lead to a subunit-specific and miRNA-specific effect.

Several components of the LDC are also involved in other cellular processes, such as splicing. A specific class of miRNAs, intronic miRNAs, derives from the processing of gene introns and requires spliceosomal components for their biogenesis [160]. However, we did not observe differences between the two classes when comparing intronic and intergenic miRNAs upon PRMT1 depletion, suggesting that they are equally affected. Furthermore, PRMT1 depletion does not have a direct impact on how the Microprocessor cleaves the pri-miRNAs. This is particularly important since the precise position of Drosha cleavage is critical for the generation of mature miRNAs and an error in the cleavage could invert the relative stability of the two miRNA strands (5p and 3p), resulting in the incorporation of the incorrect strand into the RISC complex [162].

However, the significant down-regulation of mature miRNAs observed could be the result of the impairment also of other protein complexes that are involved in miRNA pathway downstream the LDC, such as Dicer, the Ago proteins and the RISC complex. Thus, we focused on the biogenesis step catalyzed by the LDC, which is the processing of the pri- to pre-miRNAs, showing that the depletion of PRMT1 leads to a reduction of the precursor miRNAs, while the corresponding pri-miRNAs are slightly increased. The accumulation of the pri-miRNAs led us to exclude the hypothesis that the observed reduction of pre-

miRNAs may be a consequence of the effect of PRMT1 KD on the transcription of primary transcripts. On the contrary, the accumulation of the pri-miRNAs suggested that there is a general blockade of LDC-mediated miRNA processing activity.

When assessing the subcellular distribution of the pre-miRNAs of miR-15a/16-1 and 17/92 clusters, we observed that their decrease in the nucleus is followed by an increase in the amount of precursors in the cytoplasm. This result led us to speculate that, when the processing is affected, the export of the residual pre-miRNAs in the cytoplasm might be favored to sustain the production of the mature miRNAs.

The observation that the expression of Drosha and DGCR8 is not affected by PRMT1 depletion further suggests that the activity, rather than the amount, of the Microprocessor may explain the altered miRNA biogenesis. To specifically assess the enzymatic activity of the complex when PRMT1 is depleted, we plan to set up and carry out the Microprocessor reporter assay, developed in the group of Gregory [25]. In this assay, the pri-miRNA of interest is cloned into the 3'UTR of the Renilla luciferase gene. The cleavage by the Microprocessor is expected to destabilize the Renilla luciferase mRNA and lead to decreased Renilla luminescence allowing to measure directly the Microprocessor activity on targeted pri-miRNAs.

To further confirm that the enzymatic activity, rather than the expression, of PRMT1 is crucial for the correct processing of miRNAs, we will take advantage of different combination of PRMTs inhibitors and we will profile miRNA expression.

We started to investigate the molecular mechanism linking the PRMT1-regulated arginine methylation of the complex to its activity. Methylation is well known to regulate protein subcellular localization, protein-protein interactions or protein-RNA interactions. We did not observe major changes in the subcellular distribution of several components of the LDC in the absence of PRMT1, hence we focused on other mechanisms. The hypothesis that methylation might regulate the association of the accessory proteins to the Microprocessor is supported by the results on the localization analysis of the methyl-sites on the LDC. We found that the methyl-sites regulated by PRMT1 occur mainly in low complexity (LC) regions, that are typically involved in both protein-protein and protein-

RNA interactions [151], or in close proximity of RNA recognition domains **(Fig. 20C and Fig 22C)**. Evidence of a role of RNA in mediating interactions of specific subunits with the Microprocessor as well as the description of specific protein-protein interactions within the LDC were already reported in the literature [17], but at present, a comprehensive analysis of how accessory proteins associate to the Microprocessor is still missing. Thus, we set up proteomic experiments (described in the Results session, paragraph 5.5.2) that combine SILAC with co-IP of the complex to address these open issues.

If we will observe that the interaction of the accessory subunits with the Microprocessor is RNA-mediated, we will further investigate whether the modulation of R-methyl-sites on the complex affects their recruitment to the Microprocessor, ultimately impairing miRNA biogenesis. In addition, it has been recently reported that R-methylation is able to regulate the RNA-binding function of specific RNA-binding proteins [34], thus it will be interesting to assess also whether the binding ability of specific accessory proteins to distinct pri-miRNAs will be affected by PRMT1 depletion.

Our preliminary data suggest that the modulation of methylation caused by the depletion of PRMT1 can influence the interactions of some accessory proteins with the Microprocessor. Although the comprehensive proteomic experiment to further confirm this result is still on-going, the fact that specific subunits of the complex might be differentially recruited to the Microprocessor when their methylation state is altered is intriguing. Indeed, this could impair the processing of specific pri-miRNAs, similarly to what has been recently described for DDX17 [25]. The authors of this study showed that at high cell density DDX17 associates with the Microprocessor and pri-miRNAs, enhancing their processing. However, at low cell-density, DDX17 is sequestered from the Microprocessor causing a widespread miRNA down-regulation.

These experiments will provide insights on how the subunits of the LDC interact with the Microprocessor and on how PRMT1-mediated R-methylation may regulate the structure of the complex, ultimately affecting its activity.

Indeed, we cannot exclude that other PRMTs may also catalyze the methylation of the subunits of the LDC, modulating its function. In fact, while we observed that the KD of

PRMT4, PRMT5 and PRMT7 does not significantly impair miRNA processing, at least for the subset of miRNAs analyzed, we found that the depletion of PRMT6, which is also a type-I enzyme localized in the nucleus, affects the expression of the mature miRNAs tested. However, these analyses are still preliminary and further study on both global miRNA expression and the methylation state of the complex shall be performed to dissect the contribution of each enzyme in the LDC regulation.

Taken together, the results I obtained during my PhD show that the Large Drosha Complex contains a large number of high-confidence methylated sites, which are mostly catalyzed by PRMT1. Our results also show that the modulation of these R-methylated sites has a functional impact on miRNA processing. Hence, although the molecular mechanisms linking R-methylation to the altered miRNA biogenesis are still under investigation, the results described in this thesis suggest for the first time that R-methylation plays a key role in the regulation of the LDC activity and therefore in miRNA biogenesis.

Elucidating the molecular mechanisms that underlie altered miRNA biogenesis is essential to develop therapeutic strategies to treat aberrant miRNA expression, which is frequently associated with pathological states. Thus, studying post translational modification on miRNA processing factors might allow the identification of signalling pathways that regulate and fine-tune the miRNA machinery. Various inhibitors of protein methyltransferases and demethylases have been developed for the so-called epigenetic therapy, to revert aberrant methylation patterns on histones. However, these molecules may have effects also on non histonic targets, given the emerging awareness that PMTs and demethylases act of a much wider panel of substrates. In this scenario, inhibitors of PRMTs could be employed to correct the aberrant process influenced by methylation, such as miRNA biogenesis. In this perspective, a systematic assessment of the extent of protein methylation on the LDC and of its dynamic changes in dependence of distinct PRMTs inhibitors can provide crucial information for the implementation of these compounds as possible targeted drugs. This is an example of how MS-proteomics can

address fundamental questions on a specific cellular mechanism and at the same time, offer precious insights towards more translational applications.

REFERENCES

1.	Bremang, M., et al., *Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome.* Mol Biosyst, 2013. **9**(9): p. 2231-47.

2.	Lee, Y., et al., *MicroRNA genes are transcribed by RNA polymerase II.* EMBO J, 2004. **23**(20): p. 4051-60.

3.	Ha, M. and V.N. Kim, *Regulation of microRNA biogenesis.* Nat Rev Mol Cell Biol, 2014. **15**(8): p. 509-24.

4.	Davis-Dusenbery, B.N. and A. Hata, *MicroRNA in Cancer: The Involvement of Aberrant MicroRNA Biogenesis Regulatory Pathways.* Genes Cancer, 2010. **1**(11): p. 1100-14.

5.	Lee, Y., et al., *MicroRNA maturation: stepwise processing and subcellular localization.* EMBO J, 2002. **21**(17): p. 4663-70.

6.	Chang, T.C., et al., *Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis.* Mol Cell, 2007. **26**(5): p. 745-52.

7.	Bommer, G.T., et al., *p53-mediated activation of miRNA34 candidate tumor-suppressor genes.* Curr Biol, 2007. **17**(15): p. 1298-307.

8.	O'Donnell, K.A., et al., *c-Myc-regulated microRNAs modulate E2F1 expression.* Nature, 2005. **435**(7043): p. 839-43.

9.	Bracken, C.P., et al., *A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition.* Cancer Res, 2008. **68**(19): p. 7846-54.

10.	Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.

11.	Han, J., et al., *The Drosha-DGCR8 complex in primary microRNA processing.* Genes Dev, 2004. **18**(24): p. 3016-27.

12.	Lund, E., et al., *Nuclear export of microRNA precursors.* Science, 2004. **303**(5654): p. 95-8.

13.	Park, J.E., et al., *Dicer recognizes the 5' end of RNA for efficient and accurate processing.* Nature, 2011. **475**(7355): p. 201-5.

14.	Chendrimada, T.P., et al., *TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing.* Nature, 2005. **436**(7051): p. 740-4.

15. Gregory, R.I., et al., *Human RISC couples microRNA biogenesis and posttranscriptional gene silencing.* Cell, 2005. **123**(4): p. 631-40.

16. Denli, A.M., et al., *Processing of primary microRNAs by the Microprocessor complex.* Nature, 2004. **432**(7014): p. 231-235.

17. Shiohama, A., et al., *Nucleolar localization of DGCR8 and identification of eleven DGCR8-associated proteins.* Exp Cell Res, 2007. **313**(20): p. 4196-207.

18. Yeom, K.H., et al., *Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing.* Nucleic Acids Res, 2006. **34**(16): p. 4622-9.

19. Gregory, R.I., et al., *The Microprocessor complex mediates the genesis of microRNAs.* Nature, 2004. **432**(7014): p. 235-40.

20. van Kouwenhove, M., M. Kedde, and R. Agami, *MicroRNA regulation by RNA-binding proteins and its implications for cancer.* Nat Rev Cancer, 2011. **11**(9): p. 644-56.

21. Lin, S. and R.I. Gregory, *MicroRNA biogenesis pathways in cancer.* Nat Rev Cancer, 2015. **15**(6): p. 321-33.

22. Fukuda, T., et al., *DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs.* Nat Cell Biol, 2007. **9**(5): p. 604-11.

23. Wang, D., J. Huang, and Z. Hu, *RNA helicase DDX5 regulates microRNA expression and contributes to cytoskeletal reorganization in basal breast cancer cells.* Mol Cell Proteomics, 2012. **11**(2): p. M111 011932.

24. Newman, M.A. and S.M. Hammond, *Emerging paradigms of regulated microRNA processing.* Genes Dev, 2010. **24**(11): p. 1086-92.

25. Mori, M., et al., *Hippo signaling regulates microprocessor and links cell-density-dependent miRNA biogenesis to cancer.* Cell, 2014. **156**(5): p. 893-906.

26. Kawahara, Y. and A. Mieda-Sato, *TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes.* Proc Natl Acad Sci U S A, 2012. **109**(9): p. 3347-52.

27. Di Carlo, V., et al., *TDP-43 regulates the microprocessor complex activity during in vitro neuronal differentiation.* Mol Neurobiol, 2013. **48**(3): p. 952-63.

28. Sakamoto, S., et al., *The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway.* Mol Cell Biol, 2009. **29**(13): p. 3754-69.

29. Siomi, H. and M.C. Siomi, *Posttranscriptional regulation of microRNA biogenesis in animals.* Mol Cell, 2010. **38**(3): p. 323-32.

30. Han, J., et al., *Posttranscriptional crossregulation between Drosha and DGCR8.* Cell, 2009. **136**(1): p. 75-84.

31. Herbert, K.M., et al., *Phosphorylation of DGCR8 increases its intracellular stability and induces a progrowth miRNA profile.* Cell Rep, 2013. **5**(4): p. 1070-81.

32. Tang, X., et al., *Acetylation of drosha on the N-terminus inhibits its degradation by ubiquitination.* PLoS One, 2013. **8**(8): p. e72503.

33. Paroo, Z., et al., *Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling.* Cell, 2009. **139**(1): p. 112-22.

34. Larsen, S.C., et al., *Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells.* Sci Signal, 2016. **9**(443): p. rs9.

35. Volinia, S., et al., *A microRNA expression signature of human solid tumors defines cancer gene targets.* Proc Natl Acad Sci U S A, 2006. **103**(7): p. 2257-61.

36. Lu, J., et al., *MicroRNA expression profiles classify human cancers.* Nature, 2005. **435**(7043): p. 834-8.

37. Calin, G.A., et al., *A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia.* N Engl J Med, 2005. **353**(17): p. 1793-801.

38. Thomson, J.M., et al., *Extensive post-transcriptional regulation of microRNAs and its implications for cancer.* Genes Dev, 2006. **20**(16): p. 2202-7.

39. Kumar, M.S., et al., *Impaired microRNA processing enhances cellular transformation and tumorigenesis.* Nat Genet, 2007. **39**(5): p. 673-7.

40. Rakheja, D., et al., *Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours.* Nat Commun, 2014. **2**: p. 4802.

41. Bedford, M.T. and S.G. Clarke, *Protein arginine methylation in mammals: who, what, and why.* Mol Cell, 2009. **33**(1): p. 1-13.

42. Bezzi, M., et al., *Regulation of constitutive and alternative splicing by PRMT5 reveals a role for Mdm4 pre-mRNA in sensing defects in the spliceosomal machinery.* Genes Dev, 2013. **27**(17): p. 1903-16.

43. Lee, Y.H. and M.R. Stallcup, *Roles of protein arginine methylation in DNA damage signaling pathways is CARM1 a life-or-death decision point?* Cell Cycle, 2011. **10**(9): p. 1343-4.

44. Geoghegan, V., et al., *Comprehensive identification of arginine methylation in primary T cells reveals regulatory roles in cell signalling.* Nat Commun, 2015. **6**: p. 6758.

45.   Yang, Y., et al., *TDRD3 is an effector molecule for arginine-methylated histone marks.* Mol Cell, 2010. **40**(6): p. 1016-23.

46.   Morales, Y., et al., *Biochemistry and regulation of the protein arginine methyltransferases (PRMTs).* Arch Biochem Biophys, 2016. **590**: p. 138-52.

47.   Tradewell, M.L., et al., *Arginine methylation by PRMT1 regulates nuclear-cytoplasmic localization and toxicity of FUS/TLS harbouring ALS-linked mutations.* Hum Mol Genet, 2012. **21**(1): p. 136-49.

48.   Mastronardi, F.G. and M.A. Moscarello, *Molecules affecting myelin stability: a novel hypothesis regarding the pathogenesis of multiple sclerosis.* J Neurosci Res, 2005. **80**(3): p. 301-8.

49.   Lentz, S.R., R.N. Rodionov, and S. Dayal, *Hyperhomocysteinemia, endothelial dysfunction, and cardiovascular risk: the potential role of ADMA.* Atheroscler Suppl, 2003. **4**(4): p. 61-5.

50.   Yang, Y. and M.T. Bedford, *Protein arginine methyltransferases and cancer.* Nat Rev Cancer, 2013. **13**(1): p. 37-50.

51.   Hamamoto, R., V. Saloura, and Y. Nakamura, *Critical roles of non-histone protein lysine methylation in human tumorigenesis.* Nat Rev Cancer, 2015. **15**(2): p. 110-24.

52.   Zhang, Y. and D. Reinberg, *Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails.* Genes Dev, 2001. **15**(18): p. 2343-60.

53.   Slade, D.J., et al., *Chemical and biological methods to detect post-translational modifications of arginine.* Biopolymers, 2014. **101**(2): p. 133-43.

54.   Walport, L.J., et al., *Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases.* Nat Commun, 2016. **7**: p. 11974.

55.   Del Rizzo, P.A. and R.C. Trievel, *Molecular basis for substrate recognition by lysine methyltransferases and demethylases.* Biochim Biophys Acta, 2014. **1839**(12): p. 1404-15.

56.   Chang, B., et al., *JMJD6 is a histone arginine demethylase.* Science, 2007. **318**(5849): p. 444-7.

57.   Poulard, C., L. Corbo, and M. Le Romancer, *Protein arginine methylation/demethylation and cancer.* Oncotarget, 2016.

58.   Bottger, A., et al., *The oxygenase Jmjd6--a case study in conflicting assignments.* Biochem J, 2015. **468**(2): p. 191-202.

59.   Herrmann, F., et al., *Human protein arginine methyltransferases in vivo--distinct properties of eight canonical members of the PRMT family.* J Cell Sci, 2009. **122**(Pt 5): p. 667-77.

60.   Wang, H.B., et al., *Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor.* Science, 2001. **293**(5531): p. 853-857.

61.   Di Lorenzo, A. and M.T. Bedford, *Histone arginine methylation.* FEBS Lett, 2011. **585**(13): p. 2024-31.

62.   Goulet, I., et al., *Alternative splicing yields protein arginine methyltransferase 1 isoforms with distinct activity, substrate specificity, and subcellular localization.* J Biol Chem, 2007. **282**(45): p. 33009-21.

63.   Baldwin, R.M., et al., *Identification of the PRMT1v1 and PRMT1v2 specific interactomes by quantitative mass spectrometry in breast cancer cells.* Proteomics, 2015. **15**(13): p. 2187-97.

64.   Baldwin, R.M., A. Morettin, and J. Cote, *Role of PRMTs in cancer: Could minor isoforms be leaving a mark?* World J Biol Chem, 2014. **5**(2): p. 115-29.

65.   Morettin, A., R.M. Baldwin, and J. Cote, *Arginine methyltransferases as novel therapeutic targets for breast cancer.* Mutagenesis, 2015. **30**(2): p. 177-89.

66.   Lakowski, T.M. and A. Frankel, *Kinetic analysis of human protein arginine N-methyltransferase 2: formation of monomethyl- and asymmetric dimethyl-arginine residues on histone H4.* Biochem J, 2009. **421**(2): p. 253-61.

67.   Blythe, S.A., et al., *beta-Catenin primes organizer gene expression by recruiting a histone H3 arginine 8 methyltransferase, Prmt2.* Dev Cell, 2010. **19**(2): p. 220-31.

68.   Qi, C., et al., *Identification of protein arginine methyltransferase 2 as a coactivator for estrogen receptor alpha.* J Biol Chem, 2002. **277**(32): p. 28624-30.

69.   Choi, S., et al., *PRMT3 inhibits ubiquitination of ribosomal protein S2 and together forms an active enzyme complex.* Biochim Biophys Acta, 2008. **1780**(9): p. 1062-9.

70.   Lai, Y., et al., *Proteomic dissection of the von Hippel-Lindau (VHL) interactome.* J Proteome Res, 2011. **10**(11): p. 5175-82.

71.   Schurter, B.T., et al., *Methylation of histone H3 by coactivator-associated arginine methyltransferase 1.* Biochemistry, 2001. **40**(19): p. 5747-56.

72.   Bauer, U.M., et al., *Methylation at arginine 17 of histone H3 is linked to gene activation.* EMBO Rep, 2002. **3**(1): p. 39-44.

73.   Neuenkirchen, N., A. Chari, and U. Fischer, *Deciphering the assembly pathway of Sm-class U snRNPs.* FEBS Lett, 2008. **582**(14): p. 1997-2003.

74.   Koh, C.M., et al., *MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis.* Nature, 2015. **523**(7558): p. 96-100.

75.   Frankel, A., et al., *The novel human protein arginine N-methyltransferase PRMT6 is a nuclear enzyme displaying unique substrate specificity.* J Biol Chem, 2002. **277**(5): p. 3537-43.

76.   Hyllus, D., et al., *PRMT6-mediated methylation of R2 in histone H3 antagonizes H3 K4 trimethylation.* Genes Dev, 2007. **21**(24): p. 3369-80.

77.   Guccione, E., et al., *Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive.* Nature, 2007. **449**(7164): p. 933-7.

78.   El-Andaloussi, N., et al., *Arginine methylation regulates DNA polymerase beta.* Mol Cell, 2006. **22**(1): p. 51-62.

79.   Harrison, M.J., Y.H. Tang, and D.H. Dowhan, *Protein arginine methyltransferase 6 regulates multiple aspects of gene expression.* Nucleic Acids Res, 2010. **38**(7): p. 2201-16.

80.   Casadio, F., et al., *H3R42me2a is a histone modification with positive transcriptional effects.* Proc Natl Acad Sci U S A, 2013. **110**(37): p. 14894-9.

81.   Miranda, T.B., et al., *PRMT7 is a member of the protein arginine methyltransferase family with a distinct substrate specificity.* J Biol Chem, 2004. **279**(22): p. 22902-7.

82.   Lee, J.H., et al., *PRMT7, a new protein arginine methyltransferase that synthesizes symmetric dimethylarginine.* J Biol Chem, 2005. **280**(5): p. 3656-64.

83.   Lee, J., et al., *PRMT8, a new membrane-bound tissue-specific member of the protein arginine methyltransferase family.* J Biol Chem, 2005. **280**(38): p. 32890-6.

84.   Scaramuzzino, C., et al., *Protein arginine methyltransferase 1 and 8 interact with FUS to modify its sub-cellular distribution and toxicity in vitro and in vivo.* PLoS One, 2013. **8**(4): p. e61576.

85.   Simandi, Z., et al., *PRMT1 and PRMT8 regulate retinoic acid-dependent neuronal differentiation with implications to neuropathology.* Stem Cells, 2015. **33**(3): p. 726-41.

86.   Yang, Y., et al., *PRMT9 is a type II methyltransferase that methylates the splicing factor SAP145.* Nat Commun, 2015. **6**: p. 6428.

87.   Hillenkamp, F. and M. Karas, *Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization.* Methods Enzymol, 1990. **193**: p. 280-95.

88. Hillenkamp, F., et al., *Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers.* Anal Chem, 1991. **63**(24): p. 1193A-1203A.

89. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

90. Song, E.J. and K.J. Lee, *[Identification of proteome molecules by proteomics using two-dimensional gel electrophoresis and MALDI-TOF MS].* Exp Mol Med, 2001. **33**(1 Suppl): p. 5-18.

91. Griffin, P.R., et al., *Structural-Analysis of Proteins by Capillary Hplc Electrospray Tandem Mass-Spectrometry.* International Journal of Mass Spectrometry and Ion Processes, 1991. **111**: p. 131-149.

92. Emmett, M.R. and R.M. Caprioli, *Micro-Electrospray Mass-Spectrometry - Ultra-High-Sensitivity Analysis of Peptides and Proteins.* Journal of the American Society for Mass Spectrometry, 1994. **5**(7): p. 605-613.

93. Olsen, J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis.* Nat Methods, 2007. **4**(9): p. 709-12.

94. Nagaraj, N., et al., *Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation.* Journal of Proteome Research, 2010. **9**(12): p. 6786-94.

95. Zubarev, R.A., et al., *Electron capture dissociation for structural characterization of multiply charged protein cations.* Anal Chem, 2000. **72**(3): p. 563-73.

96. Finehout, E.J. and K.H. Lee, *An introduction to mass spectrometry applications in biological research.* Biochem Mol Biol Educ, 2004. **32**(2): p. 93-100.

97. Hu, Q., et al., *The Orbitrap: a new mass spectrometer.* J Mass Spectrom, 2005. **40**(4): p. 430-43.

98. Michalski, A., et al., *Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer.* Mol Cell Proteomics, 2011. **10**(9): p. M111 011015.

99. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.* Nat Methods, 2007. **4**(3): p. 207-14.

100. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nature Biotechnology, 2008. **26**(12): p. 1367-1372.

101. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.

102. Yates, J.R., 3rd, et al., *Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database.* Anal Chem, 1995. **67**(8): p. 1426-36.

103. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review.* Anal Bioanal Chem, 2007. **389**(4): p. 1017-31.

104. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present.* Anal Bioanal Chem, 2012. **404**(4): p. 939-65.

105. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nature Biotechnology, 1999. **17**(10): p. 994-9.

106. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.

107. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.

108. Ong, S.E. and M. Mann, *Stable isotope labeling by amino acids in cell culture for quantitative proteomics.* Methods Mol Biol, 2007. **359**: p. 37-52.

109. Blagoev, B., et al., *A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling.* Nature Biotechnology, 2003. **21**(3): p. 315-8.

110. de Godoy, L.M., et al., *Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system.* Genome Biol, 2006. **7**(6): p. R50.

111. Blagoev, B., et al., *Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics.* Nature Biotechnology, 2004. **22**(9): p. 1139-45.

112. Henriksen, P., et al., *Proteome-wide analysis of lysine acetylation suggests its broad regulatory scope in Saccharomyces cerevisiae.* Mol Cell Proteomics, 2012. **11**(11): p. 1510-22.

113. Ong, S.E., *The expanding field of SILAC.* Anal Bioanal Chem, 2012. **404**(4): p. 967-76.

114. Hilger, M. and M. Mann, *Triple SILAC to determine stimulus specific interactions in the Wnt pathway.* Journal of Proteome Research, 2012. **11**(2): p. 982-94.

115. Molina, H., et al., *Temporal profiling of the adipocyte proteome during differentiation using a five-plex SILAC based strategy.* Journal of Proteome Research, 2009. **8**(1): p. 48-58.

116. Schwanhausser, B., et al., *Global analysis of cellular protein translation by pulsed SILAC.* Proteomics, 2009. **9**(1): p. 205-9.

117. Geiger, T., et al., *Super-SILAC mix for quantitative proteomics of human tumor tissue.* Nat Methods, 2010. **7**(5): p. 383-5.

118. Ong, S.E., G. Mittler, and M. Mann, *Identifying and quantifying in vivo methylation sites by heavy methyl SILAC.* Nat Methods, 2004. **1**(2): p. 119-26.

119. Olsen, J.V. and M. Mann, *Status of large-scale analysis of post-translational modifications by mass spectrometry.* Mol Cell Proteomics, 2013. **12**(12): p. 3444-52.

120. Giddings, J.C., *Two-dimensional separations: concept and promise.* Anal Chem, 1984. **56**(12): p. 1258A-1260A, 1262A, 1264A passim.

121. Ficarro, S.B., et al., *Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae.* Nature Biotechnology, 2002. **20**(3): p. 301-5.

122. Pinkse, M.W., et al., *Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns.* Anal Chem, 2004. **76**(14): p. 3935-43.

123. McNulty, D.E. and R.S. Annan, *Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection.* Mol Cell Proteomics, 2008. **7**(5): p. 971-80.

124. Doll, S. and A.L. Burlingame, *Mass spectrometry-based detection and assignment of protein posttranslational modifications.* ACS Chem Biol, 2015. **10**(1): p. 63-71.

125. Mortensen, P., et al., *MSQuant, an open source platform for mass spectrometry-based quantitative proteomics.* Journal of Proteome Research, 2010. **9**(1): p. 393-403.

126. Kim, M.S., J. Zhong, and A. Pandey, *Common errors in mass spectrometry-based analysis of post-translational modifications.* Proteomics, 2016. **16**(5): p. 700-14.

127. Yoshimatsu, M., et al., *Dysregulation of PRMT1 and PRMT6, Type I arginine methyltransferases, is involved in various types of human cancers.* Int J Cancer, 2011. **128**(3): p. 562-73.

128. Afjehi-Sadat, L. and B.A. Garcia, *Comprehending dynamic protein methylation with mass spectrometry.* Curr Opin Chem Biol, 2013. **17**(1): p. 12-9.

129. Erce, M.A., et al., *The methylproteome and the intracellular methylation network.* Proteomics, 2012. **12**(4-5): p. 564-86.

130. Wang, K., et al., *An Antibody-free Approach for the Global Analysis of Protein Methylation.* Anal Chem, 2016.

131. Gayatri, S., et al., *Using oriented peptide array libraries to evaluate methylarginine-specific antibodies and arginine methyltransferase substrate motifs.* Sci Rep, 2016. **6**: p. 28718.

132. Guo, A., et al., *Immunoaffinity enrichment and mass spectrometry analysis of protein methylation.* Mol Cell Proteomics, 2014. **13**(1): p. 372-87.

133. Olsen, J.B., et al., *Quantitative Profiling of the Activity of Protein Lysine Methyltransferase SMYD2 Using SILAC-Based Proteomics.* Mol Cell Proteomics, 2016. **15**(3): p. 892-905.

134. Nielsen, M.L., M.M. Savitski, and R.A. Zubarev, *Extent of modifications in human proteome samples and their effect on dynamic range of analysis in shotgun proteomics.* Mol Cell Proteomics, 2006. **5**(12): p. 2384-91.

135. Sylvestersen, K.B. and M.L. Nielsen, *Large-Scale Identification of the Arginine Methylome by Mass Spectrometry.* Curr Protoc Protein Sci, 2015. **82**: p. 24 7 1-17.

136. Hart-Smith, G., et al., *Large Scale Mass Spectrometry-based Identifications of Enzyme-mediated Protein Methylation Are Subject to High False Discovery Rates.* Molecular & Cellular Proteomics, 2016. **15**(3): p. 989-1006.

137. Lau, H.T., K.A. Lewis, and S.E. Ong, *Quantifying in vivo, site-specific changes in protein methylation with SILAC.* Methods Mol Biol, 2014. **1188**: p. 161-75.

138. Dhar, S., et al., *Loss of the major Type I arginine methyltransferase PRMT1 causes substrate scavenging by other PRMTs.* Sci Rep, 2013. **3**: p. 1311.

139. Eram, M.S., et al., *A Potent, Selective, and Cell-Active Inhibitor of Human Type I Protein Arginine Methyltransferases.* ACS Chem Biol, 2016. **11**(3): p. 772-81.

140. Brame, C.J., M.F. Moran, and L.D. McBroom-Cerajewski, *A mass spectrometry based method for distinguishing between symmetrically and asymmetrically dimethylated arginine residues.* Rapid Commun Mass Spectrom, 2004. **18**(8): p. 877-81.

141. Gehrig, P.M., et al., *Fragmentation pathways of N(G)-methylated and unmodified arginine residues in peptides studied by ESI-MS/MS and MALDI-MS.* J Am Soc Mass Spectrom, 2004. **15**(2): p. 142-9.

142. Boisvert, F.M., et al., *A proteomic analysis of arginine-methylated protein complexes.* Mol Cell Proteomics, 2003. **2**(12): p. 1319-30.

143. Uhlmann, T., et al., *A method for large-scale identification of protein arginine methylation.* Mol Cell Proteomics, 2012. **11**(11): p. 1489-99.

144. Sylvestersen, K.B., et al., *Proteomic analysis of arginine methylation sites in human cells reveals dynamic regulation during transcriptional arrest.* Mol Cell Proteomics, 2014. **13**(8): p. 2072-88.

145. Shevchenko, A., et al., *In-gel digestion for mass spectrometric characterization of proteins and proteomes.* Nat Protoc, 2006. **1**(6): p. 2856-60.

146. Rappsilber, J., M. Mann, and Y. Ishihama, *Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips.* Nat Protoc, 2007. **2**(8): p. 1896-906.

147. Cox, J., et al., *Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment.* Journal of Proteome Research, 2011. **10**(4): p. 1794-1805.

148. Muller, H., M.J. Marzi, and F. Nicassio, *IsomiRage: From Functional Classification to Differential Expression of miRNA Isoforms.* Front Bioeng Biotechnol, 2014. **2**: p. 38.

149. Cumberworth, A., et al., *Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes.* Biochem J, 2013. **454**(3): p. 361-9.

150. Kato, M., et al., *Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels.* Cell, 2012. **149**(4): p. 753-67.

151. Calabretta, S. and S. Richard, *Emerging Roles of Disordered Sequences in RNA-Binding Proteins.* Trends Biochem Sci, 2015. **40**(11): p. 662-72.

152. Salichs, E., et al., *Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment.* PLoS Genet, 2009. **5**(3): p. e1000397.

153. Stopa, N., J.E. Krebs, and D. Shechter, *The PRMT5 arginine methyltransferase: many roles in development, cancer and beyond.* Cell Mol Life Sci, 2015. **72**(11): p. 2041-59.

154. Yu, Z., et al., *A mouse PRMT1 null allele defines an essential role for arginine methylation in genome maintenance and cell proliferation.* Mol Cell Biol, 2009. **29**(11): p. 2982-96.

155. Pekarsky, Y. and C.M. Croce, *Role of miR-15/16 in CLL.* Cell Death Differ, 2015. **22**(1): p. 6-11.

156. Olive, V., Q. Li, and L. He, *mir-17-92: a polycistronic oncomir with pleiotropic functions.* Immunol Rev, 2013. **253**(1): p. 158-66.

157. Allegra, D., et al., *Defective DROSHA processing contributes to downregulation of MiR-15/-16 in chronic lymphocytic leukemia.* Leukemia, 2014. **28**(1): p. 98-107.

158.   Pickering, B.F., D. Yu, and M.W. Van Dyke, *Nucleolin protein interacts with microprocessor complex to affect biogenesis of microRNAs 15a and 16.* J Biol Chem, 2011. **286**(51): p. 44095-103.

159.   Hwang, H.W., E.A. Wentzel, and J.T. Mendell, *Cell-cell contact globally activates microRNA biogenesis.* Proc Natl Acad Sci U S A, 2009. **106**(17): p. 7016-21.

160.   Lin, S.L., et al., *A novel RNA splicing-mediated gene silencing mechanism potential for genome evolution.* Biochem Biophys Res Commun, 2003. **310**(3): p. 754-60.

161.   Lin, S.L., J.D. Miller, and S.Y. Ying, *Intronic microRNA (miRNA).* J Biomed Biotechnol, 2006. **2006**(4): p. 26818.

162.   Davis, B.N. and A. Hata, *Regulation of MicroRNA Biogenesis: A miRiad of mechanisms.* Cell Commun Signal, 2009. **7**: p. 18.

163.   Yamaguchi, A. and K. Kitajo, *The effect of PRMT1-mediated arginine methylation on the subcellular localization, stress granules, and detergent-insoluble aggregates of FUS/TLS.* PLoS One, 2012. **7**(11): p. e49267.

164.   Morlando, M., et al., *FUS stimulates microRNA biogenesis by facilitating co-transcriptional Drosha recruitment.* EMBO J, 2012. **31**(24): p. 4502-10.

165.   Wolkowicz, U.M. and A.G. Cook, *NF45 dimerizes with NF90, Zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold.* Nucleic Acids Res, 2012. **40**(18): p. 9356-68.

166.   Kleinschmidt, M.A., et al., *Cell cycle regulation by the PRMT6 arginine methyltransferase through repression of cyclin-dependent kinase inhibitors.* PLoS One, 2012. **7**(8): p. e41446.

167.   Ferreira de Freitas, R., et al., *Discovery of a Potent and Selective Coactivator Associated Arginine Methyltransferase 1 (CARM1) Inhibitor by Virtual Screening.* J Med Chem, 2016. **59**(14): p. 6838-47.

168.   Shen, Y., et al., *Discovery of a Potent, Selective, and Cell-Active Dual Inhibitor of Protein Arginine Methyltransferase 4 and Protein Arginine Methyltransferase 6.* J Med Chem, 2016. **59**(19): p. 9124-9139.

169.   Smil, D., et al., *Discovery of a Dual PRMT5-PRMT7 Inhibitor.* ACS Med Chem Lett, 2015. **6**(4): p. 408-12.

Appendix 1. Heavy methyl SILAC-based library of the Large Drosha Complex methylated sites

Redundant list of all methylated sites identified through the hmSILAC/co-IP strategy. Uniprot protein identifiers (Uniprot ID), protein (Gene), type of methylated residue (Residue), site number (Position) and degree of methylation (Degree) and the sequence of the modified peptide identified by MS (Modified Sequence) are reported. The experiment in which the peptides were identified and the localization of the modified sites within each protein (Pfam based analysis) are also listed.

Uniprot ID	Gene	Residue	Position	Degree	Modified Sequence	Experiment	Protein Region LC/disordered	Others
P35637	FUS	R	216	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR	methylome	+	
P35637	FUS	R	218	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR	methylome	+	
P35637	FUS	R	491	di	GGDR(di)GGFR(di)GGR(di)GGGDR	methylome	+	
P35637	FUS	R	495	di	GGDR(di)GGFR(di)GGR(di)GGGDR	methylome	+	
P35637	FUS	R	498	di	GGDR(di)GGFR(di)GGR(di)GGGDR	methylome	+	
P35637	FUS	R	503	di	GGGDR(di)GGFGPGK	methylome	+	
P35637	FUS	R	503	me	GGGDR(me)GGFGPGK	methylome	+	
P35637	FUS	R	216	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR	DDX5 IP	+	
P35637	FUS	R	218	di	GR(di)GGSGGGGGGGGGGGYNR	DDX5 IP	+	
P35637	FUS	R	218	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR	DDX5 IP	+	
P35637	FUS	R	422	me	GGYGGGGSGGGGR(me)GGFPSGGGGGGGQQR(me)	FUS IP		Zf-Ran BP
P35637	FUS	R	371	me	VSFATR(me)	FUS IP	+	
P35637	FUS	R	216	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR	FUS IP	+	
P35637	FUS	R	407	me	GGYGGGGSGGGGR(me)GGFPSGGGGGGGQQR(me)	FUS IP	+	
P35637	FUS	K	510	me	GGGDR(me)GGFGPGK(me)	FUS IP		
P35637	FUS	R	242	di	SSGGYEPR(di)GR(di)	FUS IP	+	
P35637	FUS	R	244	di	SSGGYEPR(di)GR(di)	FUS IP	+	
P35637	FUS	R	503	di	GGGDR(di)GGFGPGK	FUS IP	+	
P35637	FUS	R	503	me	GGGDR(me)GGFGPGK(me)	FUS IP	+	
P35637	FUS	R	218	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR	FUS IP	+	
P35637	FUS	R	218	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR	DDX5 IP	+	
P35637	FUS	R	242	me	SSGGYEPR(me)	DDX5 IP	+	
P35637	FUS	R	216	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR	DDX5 IP	+	
P35637	FUS	R	503	di	GGGDR(di)GGFGPGK	DDX5 IP	+	
P35637	FUS	R	218	di	GR(di)GGSGGGGGGGGGGGYNR	DDX5 IP	+	
P35637	FUS	R	218	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR	Drosha IP	+	
P35637	FUS	R	216	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR	Drosha IP	+	
P35637	FUS	R	218	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR	DGCR8 IP	+	
P17844	DDX5	K	340	me	IMEEIMSEK(me)	methylome		coiled-coil
P17844	DDX5	R	502	di	R(di)GGFNTFR	methylome		P68 HR
P17844	DDX5	R	502	me	R(me)GGFNTFR	methylome		P68 HR
P17844	DDX5	R	20	me	GFGAPR(me)	DDX5 IP	+	
P17844	DDX5	R	484	di	SR(di)GR(di)GGMKDDR	DDX5 IP	+	
P17844	DDX5	R	484	me	SR(me)GR(me)GGMK(me)DDR	DDX5 IP	+	
P17844	DDX5	R	486	di	SR(di)GR(di)GGMKDDR	DDX5 IP	+	
P17844	DDX5	R	486	me	SR(me)GR(me)GGMK(me)DDR	DDX5 IP	+	
P17844	DDX5	K	490	me	SR(me)GR(me)GGMK(me)DDR	DDX5 IP	+	
P17844	DDX5	K	490	tr	SRGRGGMK(tr)DDR	DDX5 IP	+	
P17844	DDX5	R	502	di	R(di)GGFNTFR	DDX5 IP		P68 HR
P17844	DDX5	R	502	me	R(me)GGFNTFR	DDX5 IP		P68 HR
P17844	DDX5	R	484	me	SR(me)GR(me)GGMK(me)DDR	DDX5 IP	+	
P17844	DDX5	R	486	me	SR(me)GR(me)GGMK(me)DDR	DDX5 IP	+	
P17844	DDX5	K	490	tr	SRGRGGMK(tr)DDR	DDX5 IP	+	
P17844	DDX5	K	490	me	SR(me)GR(me)GGMK(me)DDR	DDX5 IP	+	
P17844	DDX5	R	502	di	R(di)GGFNTFR	DDX5 IP		P68 HR
P17844	DDX5	K	340	me	LMEEIMSEK(me)	DDX5 IP		coiled-coil
P17844	DDX5	R	486	di	SR(di)GR(di)GGMKDDR	DDX5 IP	+	
P17844	DDX5	R	484	di	SR(di)GR(di)GGMKDDR	DDX5 IP	+	
P17844	DDX5	R	502	me	R(me)GGFNTFR	DDX5 IP		P68 HR
P17844	DDX5	K	340	me	LMEEIMSEK(me)	FUS IP		coiled-coil
P17844	DDX5	R	502	me	R(me)GGFNTFR	FUS IP		P68 HR
P17844	DDX5	R	502	di	R(di)GGFNTFR	FUS IP		P68 HR
Q92841	DDX17	R	684	me	SSQSSSQQFSGIGR(me)	methylome	+	

Q92841	DDX17	R	684	me	SSQSSSQQFSGIGR(me)	DDX5 IP	+	
Q92841	DDX17	R	684	me	SSQSSSQQFSGIGR(me)	FUS IP	+	
Q92841	DDX17	R	684	me	SSQSSSQQFSGIGR(me)	DDX5 IP	+	
Q92841	DDX17	R	81	di	MR(di)GGGFGDR	DDX5 IP	+	
Q92841	DDX17	R	684	me	SSQSSSQQFSGIGR(me)	Drosha IP	+	
Q92841	DDX17	R	684	me	SSQSSSQQFSGIGR(me)	DGCR8 IP	+	
O00571	DDX3X	R	88	me	SSFFSDR(me)GSGSR	methylome	+	
O00571	DDX3X	R	592	me	SSR(me)FSGGFGAR	methylome	+	
O00571	DDX3X	R	333	me	LVDMMER(me)	FUS IP		DEAD box helic
P19338	NCL	R	694	di	GGR(di)GGGGDHKPQGK	methylome	+	
P19338	NCL	R	694	di	GGR(di)GGGGDHKPQGK	FUS IP	+	
P19338	NCL	R	656	me	GEGGFGGR(me)	FUS IP	+	
Q01844	EWSR1	R	267	di	R(di)GGPGGPPGPLMEQMGGR	methylome	+	
Q01844	EWSR1	R	455	di	KKPPMNSMR(di)GGLPPR	methylome	+	
Q01844	EWSR1	R	455	me	KKPPMNSMR(me)GGIPPR(me)	methylome	+	
Q01844	EWSR1	R	461	me	KKPPMNSMR(me)GGIPPR(me)	methylome	+	
Q01844	EWSR1	R	471	di	GMPPPLR(di)GGPGGPGGPGGPMGR	methylome	+	
Q01844	EWSR1	R	471	me	GMPPPLR(me)GGPGGPGGPGGPMGR	methylome	+	
Q01844	EWSR1	R	486	me	GMPPPLR(me)GGPGGPGGPGGPMGR(me)	methylome	+	
Q01844	EWSR1	R	503	di	GPR(di)GSR(di)GNPSGGGNVQHR	methylome	+	
Q01844	EWSR1	R	506	di	GSR(di)GNPSGGGNVQHR	methylome	+	
Q01844	EWSR1	R	471	di	GMPPPLR(di)GGPGGPGGPGGPMGR	DDX5 IP	+	
Q01844	EWSR1	R	615	di	R(di)GGPGGPPGPLMEQMGGR	DDX5 IP	+	
Q01844	EWSR1	R	615	di	R(di)GGPGGPPGPLMEQMGGR	FUS IP	+	
Q01844	EWSR1	R	471	di	GMPPPLR(di)GGPGGPGGPGGPMGR	FUS IP	+	
Q01844	EWSR1	R	471	me	GMPPPLR(me)GGPGGPGGPGGPMGR(me)	FUS IP	+	
Q01844	EWSR1	R	486	me	GMPPPLR(me)GGPGGPGGPGGPMGR(me)	FUS IP	+	
Q01844	EWSR1	R	471	me	GMPPPLR(me)GGPGGPGGPGGPMGR(me)	DDX5 IP	+	
Q01844	EWSR1	R	486	me	GMPPPLR(me)GGPGGPGGPGGPMGR(me)	DDX5 IP	+	
Q01844	EWSR1	R	471	di	GMPPPLR(di)GGPGGPGGPGGPMGR	DDX5 IP	+	
Q01844	EWSR1	R	615	di	R(di)GGPGGPPGPLMEQMGGR	DDX5 IP	+	
Q01844	EWSR1	R	486	me	GMPPPLR(me)GGPGGPGGPGGPMGR(me)	Drosha IP	+	
Q01844	EWSR1	R	471	me	GMPPPLR(me)GGPGGPGGPGGPMGR(me)	Drosha IP	+	
Q01844	EWSR1	R	471	me	GMPPPLR(me)GGPGGPGGPGGPMGR(me)	DGCR8 IP	+	
Q01844	EWSR1	R	615	di	R(di)GGPGGPPGPLMEQMGGR	DGCR8 IP	+	
Q01844	EWSR1	R	471	di	GMPPPLR(di)GGPGGPGGPGGPMGR	DGCR8 IP	+	
Q01844	EWSR1	R	486	me	GMPPPLR(me)GGPGGPGGPGGPMGR(me)	DGCR8 IP	+	
Q12906	ILF3	K	613	me	APVPVR(me)GGPK(me)	methylome	+	
Q12906	ILF3	R	609	di	RAPVPVR(di)GGPK	methylome	+	
Q12906	ILF3	R	609	me	APVPVR(me)GGPK(me)	methylome	+	
Q12906	ILF3	R	609	di	APVPVR(di)GGPK	FUS IP	+	
Q12906	ILF3	K	613	me	APVPVR(me)GGPK(me)	FUS IP	+	
Q12906	ILF3	R	609	me	APVPVR(me)GGPK(me)	FUS IP	+	
Q12906	ILF3	R	815	me	FNYSGSGGR(me)	DDX5 IP	+	
Q12906	ILF3	R	371	me	NENPVDYTVQIPPSTTYAITPMK(me)	DDX5 IP	+	
P31943	HNRNPH1	R	212	me	PGPYDR(me)PGAGR(me)GYNSIGR(me)GAGFER(me)	methylome	+	
P31943	HNRNPH1	R	217	di	PGPYDRPGAGR(di)GYNSIGR(di)GAGFER	methylome	+	
P31943	HNRNPH1	R	217	me	PGPYDR(me)PGAGR(me)GYNSIGR(me)GAGFER(me)	methylome	+	
P31943	HNRNPH1	R	224	di	PGPYDRPGAGR(di)GYNSIGR(di)GAGFER	methylome	+	
P31943	HNRNPH1	R	224	me	PGPYDR(me)PGAGR(me)GYNSIGR(me)GAGFER(me)	methylome	+	
P31943	HNRNPH1	R	230	me	PGPYDR(me)PGAGR(me)GYNSIGR(me)GAGFER(me)	methylome	+	
P31943	HNRNPH1	R	233	di	R(di)GAYGGGYGGYDDYNGYNDGYGFGSDR	methylome	+	
P31943	HNRNPH1	R	233	me	R(me)GAYGGGYGGYDDYNGYNDGYGFGSDR	methylome	+	
P31943	HNRNPH1	R	233	di	R(di)GAYGGGYGGYDDYNGYNDGYGFGSDR	DDX5 IP	+	
P31943	HNRNPH1	R	316	me	ATENDIYNFFSPLNPVR(me)	FUS IP		RRM 6

P31943	HNRNPH1	R	233	di	R(di)GAYGGGYGGYDDYNGYNDGYGFGSDR	FUS IP		RRM 6
P31943	HNRNPH1	R	233	me	R(me)GAYGGGYGGYDDYNGYNDGYGFGSDR	DDX5 IP		RRM 6
P31943	HNRNPH1	R	233	di	R(di)GAYGGGYGGYDDYNGYNDGYGFGSDR	DDX5 IP		RRM 6
P31943	HNRNPH1	R	217	me	PGPYDR(me)PGAGR(me)GYNSIGR(me)GAGFER(me)	DGCR8 IP	+	
P31943	HNRNPH1	R	224	me	PGPYDR(me)PGAGR(me)GYNSIGR(me)GAGFER(me)	DGCR8 IP	+	
P31943	HNRNPH1	R	217	di	PGPYDRPGAGR(di)GYNSIGR(di)GAGFER	DGCR8 IP	+	
P31943	HNRNPH1	R	217	di	PGAGR(di)GYNSIGR(di)GAGFER	DGCR8 IP	+	
P31943	HNRNPH1	R	230	me	PGPYDR(me)PGAGR(me)GYNSIGR(me)GAGFER(me)	DGCR8 IP	+	
P31943	HNRNPH1	R	224	di	PGAGR(di)GYNSIGR(di)GAGFER	DGCR8 IP	+	
P31943	HNRNPH1	R	212	me	PGPYDR(me)PGAGR(me)GYNSIGR(me)GAGFER(me)	DGCR8 IP	+	
P31943	HNRNPH1	R	224	di	PGPYDRPGAGR(di)GYNSIGR(di)GAGFER	DGCR8 IP	+	
Q00839	HNRNPU	R	733	di	R(di)GNMPQR(di)GGGGGGSGGIGYPYPR	methylome	+	
Q00839	HNRNPU	R	733	me	R(me)GNMPQR(me)GGGGGGSGGIGYPYPR	methylome	+	
Q00839	HNRNPU	R	739	di	R(di)GNMPQR(di)GGGGGGSGGIGYPYPR	methylome	+	
Q00839	HNRNPU	R	739	me	R(me)GNMPQR(me)GGGGGGSGGIGYPYPR	methylome	+	
Q00839	HNRNPU	R	733	di	R(di)GNMPQR(di)GGGGGGSGGIGYPYPR	DDX5 IP	+	
Q00839	HNRNPU	R	739	di	R(di)GNMPQR(di)GGGGGGSGGIGYPYPR	DDX5 IP	+	
Q00839	HNRNPU	R	739	di	R(di)GNMPQR(di)GGGGGGSGGIGYPYPR	DDX5 IP	+	
Q00839	HNRNPU	R	733	di	R(di)GNMPQR(di)GGGGGGSGGIGYPYPR	DDX5 IP	+	
Q00839	HNRNPU	R	739	di	R(di)GNMPQR(di)GGGGGGSGGIGYPYPR	Drosha IP	+	
Q00839	HNRNPU	R	733	di	R(di)GNMPQR(di)GGGGGGSGGIGYPYPR	Drosha IP	+	
P52272	HNRNPM	R	621	me	MGLAMGGGGGASFDR(me)	FUS IP	+	
P52272	HNRNPM	R	471	me	MGPLGLDHMASSIER(me)	FUS IP	+	
Q12905	ILF2	R	16	di	FGSR(di)GGPGGGFR	methylome	+	
O43143	DHX15	R	243	me	YMTDGMLLR(me)	FUS IP		DEAD box helic
B7Z4B8	HNRNPUL1	R	529	di	GGGGFR(di)GR(di)GGGGGFQR	methylome	+	
B7Z4B8	HNRNPUL1	R	531	di	GGGGFR(di)GR(di)GGGGGFQR	methylome	+	
O14979	HNRNPDL	R	408	di	ASR(di)GGGNHQNNYQPY	methylome	+	
O14979	HNRNPDL	R	408	di	ASR(di)GGGNHQNNYQPY	DDX5 IP	+	
Q92804	TAF15	K	192	me	GYGGSQGGGR(me)GR(me)GGYDK(me)	methylome	+	
Q92804	TAF15	K	192	tr	GYGGSQGGGRGRGGYDK(tr)DGR	methylome	+	
Q92804	TAF15	K	210	di	GPMTGSSGGDRGGFK(di)	methylome	+	
Q92804	TAF15	K	210	me	GPMTGSSGGDR(me)GGFK(me)	methylome	+	
Q92804	TAF15	K	576	me	SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)	methylome	+	
Q92804	TAF15	K	576	tr	SGGGYGGDRGGGYGGDRGGYGGK(tr)	methylome	+	
Q92804	TAF15	R	185	di	GYGGSQGGGR(di)GR(di)GGYDKDGR	methylome	+	
Q92804	TAF15	R	185	me	GYGGSQGGGR(me)GR(me)GGYDK(me)	methylome	+	
Q92804	TAF15	R	187	di	GYGGSQGGGR(di)GR(di)GGYDKDGR	methylome	+	
Q92804	TAF15	R	187	me	GYGGSQGGGR(me)GR(me)GGYDK(me)	methylome	+	
Q92804	TAF15	R	195	me	GYGGSQGGGR(me)GR(me)GGYDK(me)DGR(me)	methylome	+	
Q92804	TAF15	R	206	di	GPMTGSSGGDR(di)GGFK	methylome	+	
Q92804	TAF15	R	206	me	GPMTGSSGGDR(me)GGFK(me)	methylome	+	
Q92804	TAF15	R	326	di	RPEFMR(di)GGGSGGGR	methylome		RRM 1
Q92804	TAF15	R	385	me	NSCNQCNEPRPEDSR(me)PSGGDFR(me)GR(me)	methylome		Zf Ran BP
Q92804	TAF15	R	395	me	NSCNQCNEPRPEDSRPSGGDFR(me)GR(me)	methylome	+	
Q92804	TAF15	R	397	me	NSCNQCNEPRPEDSRPSGGDFR(me)GR(me)	methylome	+	
Q92804	TAF15	R	519	di	GGYGGDR(di)GGYGGDR(di)GGYGGDR(di)GGYGGDR	methylome	+	
Q92804	TAF15	R	519	me	GGGYGGDR(me)GGGYGGDR(me)GGYGGDR(me)	methylome	+	
Q92804	TAF15	R	526	di	SGGGYGGDR(di)GGGYGGDR(di)GGYGGK	methylome	+	
Q92804	TAF15	R	526	me	GGGYGGDR(me)GGGYGGDR(me)GGYGGDR(me)	methylome	+	
Q92804	TAF15	R	528	di	SR(di)GGYGGDR(di)GGGSGYGGDR	methylome	+	
Q92804	TAF15	R	535	di	SR(di)GGYGGDR(di)GGGSGYGGDR	methylome	+	
Q92804	TAF15	R	570	di	SGGGYGGDR(di)GGGYGGDR(di)GGYGGK	methylome	+	
Q92804	TAF15	R	570	me	SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)	methylome	+	
Q92804	TAF15	R	195	me	DGR(me)GPMTGSSGGDR(me)GGFK(me)	DDX5 IP	+	

Q92804	TAF15	R	206	di	GPMTGSSGGDR(di)GGFK	DDX5 IP	+
Q92804	TAF15	R	206	me	DGR(me)GPMTGSSGGDR(me)GGFK(me)	DDX5 IP	+
Q92804	TAF15	R	206	me	GPMTGSSGGDR(me)GGFK(me)	DDX5 IP	+
Q92804	TAF15	K	210	di	GPMTGSSGGDRGGFK(di)	DDX5 IP	+
Q92804	TAF15	K	210	me	GPMTGSSGGDR(me)GGFK(me)	DDX5 IP	+
Q92804	TAF15	K	210	me	DGR(me)GPMTGSSGGDR(me)GGFK(me)	DDX5 IP	+
Q92804	TAF15	R	459	me	SGGGYGGDR(me)GGGYGGDR	DDX5 IP	+
Q92804	TAF15	R	475	di	GGGYGGDR(di)GGGYGGDR	DDX5 IP	+
Q92804	TAF15	R	490	me	GGGYGGDR(me)GGYGGDR(me)	DDX5 IP	+
Q92804	TAF15	R	498	di	GGGYGGDR(di)GGYGGDR	DDX5 IP	+
Q92804	TAF15	R	505	me	GGGYGGDR(me)GGYGGDR(me)	DDX5 IP	+
Q92804	TAF15	R	528	di	SR(di)GGYGGDR(di)GGGSGYGGDR	DDX5 IP	+
Q92804	TAF15	R	535	di	SR(di)GGYGGDR(di)GGGSGYGGDR	DDX5 IP	+
Q92804	TAF15	R	562	di	SGGGYGGDR(di)GGGYGGDR	DDX5 IP	+
Q92804	TAF15	R	562	di	SGGGYGGDR(di)GGGYGGDR(di)GGYGGK	DDX5 IP	+
Q92804	TAF15	R	562	me	SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)	DDX5 IP	+
Q92804	TAF15	R	562	me	GGGYGGDR(me)GGGYGGDR	DDX5 IP	+
Q92804	TAF15	R	570	di	GGGYGGDR(di)GGYGGK	DDX5 IP	+
Q92804	TAF15	R	570	di	SGGGYGGDR(di)GGGYGGDR(di)GGYGGK	DDX5 IP	+
Q92804	TAF15	R	570	me	SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)	DDX5 IP	+
Q92804	TAF15	R	570	me	GGGYGGDR(me)GGYGGK(me)	DDX5 IP	+
Q92804	TAF15	K	576	me	SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)	DDX5 IP	+
Q92804	TAF15	K	576	me	GGGYGGDR(me)GGYGGK(me)	DDX5 IP	+
Q92804	TAF15	K	576	tr	SGGGYGGDRGGGYGGDRGGYGGK(tr)	DDX5 IP	+
Q92804	TAF15	R	206	me	GPMTGSSGGDR(me)GGFK(me)	FUS IP	+
Q92804	TAF15	K	576	me	GGGYGGDR(me)GGYGGK(me)	FUS IP	+
Q92804	TAF15	R	187	me	GYGGSQGGGR(me)GR(me)	FUS IP	+
Q92804	TAF15	R	528	di	SR(di)GGYGGDR(di)GGGSGYGGDR	FUS IP	+
Q92804	TAF15	R	570	me	SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)	FUS IP	+
Q92804	TAF15	R	570	di	GGGYGGDR(di)GGYGGK	FUS IP	+
Q92804	TAF15	R	483	di	GGGYGGDR(di)GGGYGGDR(di)GGYGGDR	FUS IP	+
Q92804	TAF15	R	397	di	PSGGDFR(di)GR(di)	FUS IP	+
Q92804	TAF15	R	206	me	GPMTGSSGGDR(me)GGFK(me)NFGGHR	FUS IP	+
Q92804	TAF15	R	206	di	GPMTGSSGGDR(di)GGFK	FUS IP	+
Q92804	TAF15	R	562	di	SGGGYGGDR(di)GGGYGGDR(di)GGYGGK	FUS IP	+
Q92804	TAF15	R	505	di	GGYGGDR(di)GGGYGGDR(di)GGYGGDR	FUS IP	+
Q92804	TAF15	K	576	me	SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)	FUS IP	+
Q92804	TAF15	R	535	di	SR(di)GGYGGDR(di)GGGSGYGGDR	FUS IP	+
Q92804	TAF15	K	210	me	DGR(me)GPMTGSSGGDR(me)GGFK(me)	FUS IP	+
Q92804	TAF15	R	187	di	GYGGSQGGGR(di)GR(di)GGYDK	FUS IP	+
Q92804	TAF15	R	570	me	GGGYGGDR(me)GGYGGK(me)	FUS IP	+
Q92804	TAF15	R	185	di	GYGGSQGGGR(di)GR(di)GGYDK	FUS IP	+
Q92804	TAF15	R	195	me	DGR(me)GPMTGSSGGDR(me)GGFK(me)	FUS IP	+
Q92804	TAF15	R	490	me	GGGYGGDR(me)GGYGGDR(me)	FUS IP	+
Q92804	TAF15	R	395	di	PSGGDFR(di)GR(di)	FUS IP	+
Q92804	TAF15	R	570	me	SGGGYGGDR(me)GGGYGGDR(me)	FUS IP	+
Q92804	TAF15	R	483	di	GGGYGGDR(di)GGYGGDR	FUS IP	+
Q92804	TAF15	R	397	me	PSGGDFR(me)GR(me)	FUS IP	+
Q92804	TAF15	R	505	me	GGGYGGDR(me)GGYGGDR(me)	FUS IP	+
Q92804	TAF15	K	210	me	GPMTGSSGGDR(me)GGFK(me)NFGGHR	FUS IP	+
Q92804	TAF15	R	206	me	DGR(me)GPMTGSSGGDR(me)GGFK(me)	FUS IP	+
Q92804	TAF15	R	562	di	SGGGYGGDR(di)GGGYGGDR	FUS IP	+
Q92804	TAF15	R	475	di	GGGYGGDR(di)GGGYGGDR(di)GGYGGDR	FUS IP	+
Q92804	TAF15	K	210	me	GPMTGSSGGDR(me)GGFK(me)	FUS IP	+
Q92804	TAF15	R	562	me	SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)	FUS IP	+

Q92804	TAF15	K	576	tr	SGGGYGGDRGGGYGGDRGGYGGK(tr)	FUS IP	+
Q92804	TAF15	R	320	me	VSFATR(me)	FUS IP	+
Q92804	TAF15	R	185	me	GYGGSQGGGR(me)GR(me)	FUS IP	+
Q92804	TAF15	R	562	me	SGGGYGGDR(me)GGGYGGDR(me)	FUS IP	+
Q92804	TAF15	R	395	me	PSGGDFR(me)GR(me)	FUS IP	+
Q92804	TAF15	R	570	di	SGGGYGGDR(di)GGGYGGDR(di)GGYGGK	FUS IP	+
Q92804	TAF15	K	210	me	DGR(me)GPMTGSSGGDR(me)GGFK(me)	DDX5 IP	+
Q92804	TAF15	R	498	di	GGGYGGDR(di)GGYGGDR	DDX5 IP	+
Q92804	TAF15	R	570	me	SGGGYGGDR(me)GGGYGGDR(me)	DDX5 IP	+
Q92804	TAF15	R	185	me	GYGGSQGGGR(me)GR(me)	DDX5 IP	+
Q92804	TAF15	R	570	di	SGGGYGGDR(di)GGGYGGDR(di)GGYGGK	DDX5 IP	+
Q92804	TAF15	K	210	me	GPMTGSSGGDR(me)GGFK(me)	DDX5 IP	+
Q92804	TAF15	R	321	me	R(me)PEFMR(me)GGGSGGGR	DDX5 IP	+
Q92804	TAF15	R	206	me	GPMTGSSGGDR(me)GGFK	DDX5 IP	+
Q92804	TAF15	R	185	me	GYGGSQGGGR(me)GR(me)GGYDK(me)DGR(me)	DDX5 IP	+
Q92804	TAF15	K	192	tr	GYGGSQGGGRGRGGYDK(tr)	DDX5 IP	+
Q92804	TAF15	R	562	di	SGGGYGGDR(di)GGGYGGDR	DDX5 IP	+
Q92804	TAF15	R	570	di	GGGYGGDR(di)GGGYGGDR(di)	DDX5 IP	+
Q92804	TAF15	R	395	di	PSGGDFR(di)GR	DDX5 IP	+
Q92804	TAF15	R	490	me	GGGYGGDR(me)GGGYGGDR(me)GGYGGDR(me)	DDX5 IP	+
Q92804	TAF15	R	528	di	SR(di)GGYGGDR(di)GGGSGYGGDR	DDX5 IP	+
Q92804	TAF15	R	195	me	GYGGSQGGGR(me)GR(me)GGYDK(me)DGR(me)	DDX5 IP	+
Q92804	TAF15	R	505	di	GGYGGDR(di)GGYGGDR(di)GGYGGDR	DDX5 IP	+
Q92804	TAF15	R	187	di	GYGGSQGGGR(di)GR(di)GGYDKDGR	DDX5 IP	+
Q92804	TAF15	R	562	di	SGGGYGGDR(di)GGGYGGDR(di)GGYGGK	DDX5 IP	+
Q92804	TAF15	R	562	me	SGGGYGGDR(me)GGGYGGDR(me)	DDX5 IP	+
Q92804	TAF15	R	187	me	GYGGSQGGGR(me)GR(me)GGYDK(me)DGR(me)	DDX5 IP	+
Q92804	TAF15	R	185	di	GYGGSQGGGR(di)GR(di)GGYDK	DDX5 IP	+
Q92804	TAF15	R	187	me	GYGGSQGGGR(me)GR(me)GGYDK(me)	DDX5 IP	+
Q92804	TAF15	R	570	di	GGGYGGDR(di)GGYGGK	DDX5 IP	+
Q92804	TAF15	R	562	me	SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)	DDX5 IP	+
Q92804	TAF15	K	192	me	GYGGSQGGGR(me)GR(me)GGYDK(me)	DDX5 IP	+
Q92804	TAF15	R	185	di	GYGGSQGGGR(di)GR(di)GGYDKDGR	DDX5 IP	+
Q92804	TAF15	R	326	di	RPEFMR(di)GGGSGGGR	DDX5 IP	+
Q92804	TAF15	R	206	me	DGR(me)GPMTGSSGGDR(me)GGFK(me)	DDX5 IP	+
Q92804	TAF15	R	505	me	GGGYGGDR(me)GGYGGDR(me)	DDX5 IP	+
Q92804	TAF15	R	185	di	GYGGSQGGGR(di)GR	DDX5 IP	+
Q92804	TAF15	R	545	me	GGYGGDR(me)GGGSGYGGDR(me)	DDX5 IP	+
Q92804	TAF15	R	490	me	GGGYGGDR(me)GGYGGDR(me)	DDX5 IP	+
Q92804	TAF15	R	483	me	GGGYGGDR(me)GGGYGGDR(me)GGYGGDR(me)	DDX5 IP	+
Q92804	TAF15	R	185	me	GYGGSQGGGR(me)GR(me)GGYDK(me)	DDX5 IP	+
Q92804	TAF15	R	187	di	GYGGSQGGGR(di)GR(di)GGYDK	DDX5 IP	+
Q92804	TAF15	K	576	me	GGGYGGDR(me)GGYGGK(me)	DDX5 IP	+
Q92804	TAF15	R	535	di	GGYGGDR(di)GGGSGYGGDR	DDX5 IP	+
Q92804	TAF15	R	187	me	GYGGSQGGGR(me)GR(me)	DDX5 IP	+
Q92804	TAF15	K	192	me	GYGGSQGGGR(me)GR(me)GGYDK(me)DGR(me)	DDX5 IP	+
Q92804	TAF15	K	576	me	SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)	DDX5 IP	+
Q92804	TAF15	R	467	me	SGGGYGGDR(me)GGGYGGDR(me)GGGYGGDR(me)	DDX5 IP	+
Q92804	TAF15	R	475	di	GGGYGGDR(di)GGGYGGDR(di)	DDX5 IP	+
Q92804	TAF15	R	570	me	GGGYGGDR(me)GGYGGK(me)	DDX5 IP	+
Q92804	TAF15	R	535	di	SR(di)GGYGGDR(di)GGGSGYGGDR	DDX5 IP	+
Q92804	TAF15	K	576	tr	SGGGYGGDRGGGYGGDRGGYGGK(tr)	DDX5 IP	+
Q92804	TAF15	R	195	me	DGR(me)GPMTGSSGGDR(me)GGFK(me)	DDX5 IP	+
Q92804	TAF15	R	483	di	GGGYGGDR(di)GGGYGGDR(di)	DDX5 IP	+
Q92804	TAF15	R	206	di	GPMTGSSGGDR(di)	DDX5 IP	+

Q92804	TAF15	R	535	me	GGYGGDR(me)GGGSGYGGDR(me)	DDX5 IP	+
Q92804	TAF15	R	570	me	SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)	DDX5 IP	+
Q92804	TAF15	R	326	me	R(me)PEFMR(me)GGGSGGGR	DDX5 IP	+
Q92804	TAF15	R	206	di	GPMTGSSGGDR(di)GGFK	DDX5 IP	+
Q92804	TAF15	R	185	di	GYGGSQGGGR(di)GR(di)GGYDK	Drosha IP	+
Q92804	TAF15	R	187	di	GYGGSQGGGR(di)GR(di)GGYDK	Drosha IP	+
Q92804	TAF15	R	570	me	GGGYGGDR(me)GGYGGK(me)	Drosha IP	+
Q92804	TAF15	K	576	me	GGGYGGDR(me)GGYGGK(me)	Drosha IP	+
Q92804	TAF15	R	535	di	SR(di)GGYGGDR(di)GGGSGYGGDR	Drosha IP	+
Q92804	TAF15	R	528	di	SR(di)GGYGGDR(di)GGGSGYGGDR	Drosha IP	+
Q92804	TAF15	K	210	me	GPMTGSSGGDR(me)GGFK(me)	Drosha IP	+
Q92804	TAF15	R	206	me	GPMTGSSGGDR(me)GGFK(me)	Drosha IP	+
Q92804	TAF15	R	528	di	SR(di)GGYGGDR(di)GGGSGYGGDR	DGCR8 IP	+
Q92804	TAF15	R	185	di	GYGGSQGGGR(di)GR(di)GGYDK	DGCR8 IP	+
Q92804	TAF15	R	206	me	GPMTGSSGGDR(me)GGFK(me)	DGCR8 IP	+
Q92804	TAF15	K	576	me	GGGYGGDR(me)GGYGGK(me)	DGCR8 IP	+
Q92804	TAF15	R	187	di	GYGGSQGGGR(di)GR(di)GGYDK	DGCR8 IP	+
Q92804	TAF15	R	570	me	GGGYGGDR(me)GGYGGK(me)	DGCR8 IP	+
Q92804	TAF15	R	570	di	GGGYGGDR(di)GGYGGK	DGCR8 IP	+
Q92804	TAF15	R	505	me	GGGYGGDR(me)GGYGGDR(me)	DGCR8 IP	+
Q92804	TAF15	K	210	me	GPMTGSSGGDR(me)GGFK(me)	DGCR8 IP	+
Q92804	TAF15	R	535	di	SR(di)GGYGGDR(di)GGGSGYGGDR	DGCR8 IP	
Q92804	TAF15	R	490	me	GGGYGGDR(me)GGYGGDR(me)	DGCR8 IP	
Q08211	DHX9	R	1160	me	YGDGPR(me)PPK	methylome	+
Q08211	DHX9	R	1175	di	R(di)GGSSYSGGGYGGGYSSGGYGSGGYGGSANSFR	methylome	+
Q08211	DHX9	R	1223	di	GVSR(di)GGFR(di)GNSGGDYR	methylome	
Q08211	DHX9	R	1227	di	GVSR(di)GGFR(di)GNSGGDYR	methylome	+
Q08211	DHX9	R	1249	di	GSGGFQR(di)GGGR(di)GAYGTGYFGQGR(di)GGGGY	methylome	+
Q08211	DHX9	R	1253	di	GSGGFQR(di)GGGR(di)GAYGTGYFGQGR(di)GGGGY	methylome	+
Q08211	DHX9	R	1265	di	GSGGFQR(di)GGGR(di)GAYGTGYFGQGR(di)GGGGY	methylome	+
Q08211	DHX9	R	1160	di	YGDGPR(di)PPK	DDX5 IP	+
Q08211	DHX9	R	1160	me	YGDGPR(me)PPK(me)	DDX5 IP	+
Q08211	DHX9	K	1163	me	YGDGPR(me)PPK(me)	DDX5 IP	+
Q08211	DHX9	R	1160	me	YGDGPR(me)PPK(me)	FUS IP	+
Q08211	DHX9	R	795	me	LETHMTPEMFR(me)	FUS IP	
Q08211	DHX9	R	1160	di	YGDGPR(di)PPK	FUS IP	+
Q08211	DHX9	K	1163	me	YGDGPR(me)PPK(me)	FUS IP	+
Q08211	DHX9	R	1253	di	GGGR(di)GAYGTGYFGQGR(di)GGGGY	DDX5 IP	+
Q08211	DHX9	R	1265	di	GGGR(di)GAYGTGYFGQGR(di)GGGGY	DDX5 IP	+
Q08211	DHX9	K	1163	me	YGDGPR(me)PPK(me)	Drosha IP	
Q08211	DHX9	R	1160	me	YGDGPR(me)PPK	Drosha IP	
Q08211	DHX9	R	1160	me	YGDGPR(me)PPK	DGCR8 IP	
Q08211	DHX9	R	1160	di	YGDGPR(di)PPK	DGCR8 IP	
Q08211	DHX9	K	1163	me	YGDGPR(me)PPK(me)	DGCR8 IP	
P11021	HSPA5	K	585	me	IGGK(me)ISSEDK	methylome	+
P11021	HSPA5	K	585	tr	LGGK(tr)LSSEDKETMEK	methylome	+
P11021	HSPA5	K	591	me	IGGK(me)ISSEDK(me)ETMEK	methylome	+
P11021	HSPA5	K	585	di	IGGK(di)ISSEDK	methylome	+

114

Appendix 2. Methylated peptides identified following the depletion of PRMT1

Redundant list of all methylated peptides identified through the SILAC/co-IP strategy. Uniprot protein identifiers (Uniprot ID), protein (Gene), type of methylated residue (Residue), site number (Position), degree of methylation (Degree) and the sequence of the modified peptide identified by MS (Modified Sequence) are reported. Log_2 H/L SILAC ratios of modified peptides in the three IPs (DDX5 IP, DGCR8 IP and FUS IP) are reported, both in the 'forward' and 'reverse' SILAC experiments. Significantly regulated peptides that were already reported in the hmSILAC-based library are indicated.

Uniprot ID	Gene	Residue	Position	Modification	Modified Sequence	Fwd Experiment log2 (Input)	log2 (DDX5 IP)	log2 (DGCR8 IP)	log2 (FUS IP)	Rev Experiment log2 (Input)	log2 (DDX5 IP)	log2 (DGCR8 IP)	log2 (FUS IP)	Sites regulated and present in the library
B7Z4B8	HNRNPUL1	R	567	di	GPPGGNR(di)GGFQNR(di)GGGSGGGGNYR(di)GGFNR		-0.566	-0.736	-0.904					
B7Z4B8	HNRNPUL1	R	550	di	GPPGGNR(di)GGFQNR(di)GGGSGGGGNYR(di)GGFNR		-0.566	-0.736	-0.904					
B7Z4B8	HNRNPUL1	R	556	di	GPPGGNR(di)GGFQNR(di)GGGSGGGGNYR(di)GGFNR		-0.566	-0.736	-0.904					
B7Z4B8	HNRNPUL1	R	550	di	GPPGGNR(di)GGFQNR(di)GGGSGGGGNYRGGFNR(di)		-0.476		-0.329					
B7Z4B8	HNRNPUL1	R	572	di	GPPGGNR(di)GGFQNR(di)GGGSGGGGNYRGGFNR(di)		-0.476		-0.329					
B7Z4B8	HNRNPUL1	R	556	di	GPPGGNR(di)GGFQNR(di)GGGSGGGGNYRGGFNR(di)		-0.476		-0.329					
Q12905	ILF2	R	16	di	FGSR(di)GGPGGGFR		-2.119	-2.019						+
P35637	FUS	R	244	di	SSGGYEPR(di)GR(di)								-0.314	+
P35637	FUS	R	242	di	SSGGYEPR(di)GR(di)								-0.314	+
P35637	FUS	R	495	di	GGDR(di)GGFR(di)GGR		4.221							+
P35637	FUS	R	491	di	GGDR(di)GGFR(di)GGR		4.221							+
P35637	FUS	R	216	di	GGR(di)GRGGSGGGGGGGGGGGYNR			4.012						+
P35637	FUS	R	503	di	GGGDR(di)GGFGPGK		1.468	1.598	1.355		1.114		1.061	+
P35637	FUS	R	394	di	GGPMGR(di)GGYGGGSGGGGR(me)GGFPSGGGGGGGQQR(me)			-0.198					0.522	
P35637	FUS	R	407	me	GGPMGR(di)GGYGGGSGGGGR(me)GGFPSGGGGGGGQQR(me)			-0.198					0.522	+
P35637	FUS	R	422	me	GGPMGR(di)GGYGGGSGGGGR(me)GGFPSGGGGGGGQQR(me)			-0.198					0.522	+
P35637	FUS	R	242	me	SSGGYEPR(me)								3.243	+
P35637	FUS	R	216	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR(me)		-0.833		-0.618					+
P35637	FUS	R	218	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR(me)		-0.833		-0.618					+
P35637	FUS	R	234	me	GGR(di)GR(di)GGSGGGGGGGGGGGYNR(me)		-0.833		-0.618					
P35637	FUS	R	503	di	GGR(di)GGGDR(di)GGFGPGK								-0.708	+
P35637	FUS	R	498	di	GGR(di)GGGDR(di)GGFGPGK								-0.708	+
P35637	FUS	R	216	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR		-0.741	0.115	0.101	0.322	0.333		-0.573	+
P35637	FUS	R	218	di	GGR(di)GR(di)GGSGGGGGGGGGGGYNR		-0.741	0.115	0.101	0.322	0.333		-0.573	+
P35637	FUS	R	503	me	GGGDR(me)GGFGPGK		1.895		3.820				3.748	+
P35637	FUS	R	481	di	R(di)GGR(di)GGYDR(di)GGYR(di)								-0.542	
P35637	FUS	R	473	di	R(di)GGR(di)GGYDR(di)GGYR(di)								-0.542	
P35637	FUS	R	485	di	R(di)GGR(di)GGYDR(di)GGYR(di)								-0.542	
P35637	FUS	R	476	di	R(di)GGR(di)GGYDR(di)GGYR(di)								-0.542	
P35637	FUS	R	218	di	GR(di)GGSGGGGGGGGGGYNR		1.378	1.916	1.462				2.048	+
O14979	HNRNPDL	K	162	me	MFIGGLSWDTSKK(me)		0.605	0.301	0.100					
O14979	HNRNPDL	K	161	me	MFIGGLSWDTSK(me)	0.298		0.327	0.366					
Q92804	TAF15	K	210	me	GPMTGSSGGDR(me)GGFK(me)	-0.322	-0.345	-0.268	-0.267	-0.355	-0.348	-0.270	-0.399	+
Q92804	TAF15	R	206	me	GPMTGSSGGDR(me)GGFK(me)	-0.322	-0.345	-0.268	-0.267	-0.355	-0.348	-0.270	-0.399	+
Q92804	TAF15	R	505	di	GGYGGDR(di)GGYGGDR(di)GGYGGDR		1.678	1.938	2.008				1.721	+
Q92804	TAF15	R	498	di	GGYGGDR(di)GGYGGDR(di)GGYGGDR		1.678	1.938	2.008				1.721	+
Q92804	TAF15	R	570	di	GGGYGGDR(di)GGYGGK		0.268	0.316	0.305	0.338			0.185	+
Q92804	TAF15	R	321	me	R(me)PEFMR(me)GGGSGGGR			1.753	1.858					+
Q92804	TAF15	R	326	me	R(me)PEFMR(me)GGGSGGGR			1.753	1.858					+
Q92804	TAF15	R	475	di	GGGYGGDR(di)GGGYGGDR(di)GGYGGDR		-1.463	-1.153	-0.942					+
Q92804	TAF15	R	483	di	GGGYGGDR(di)GGGYGGDR(di)GGYGGDR		-1.463	-1.153	-0.942					+
Q92804	TAF15	K	210	me	GPMTGSSGGDRGGFK(me)		2.714							+

Q92804	TAF15	R	185	di	GYGGSQGGGR(di)GRGGYDK	2.710		2.331				+
Q92804	TAF15	R	490	me	GGGYGGDR(me)GGGYGGDR(di)GGGYGGDR(me)GGYGGDR(me)		-1.136					+
Q92804	TAF15	R	467	me	GGGYGGDR(me)GGGYGGDR(di)GGGYGGDR(me)GGYGGDR(me)		-1.136					+
Q92804	TAF15	R	475	di	GGGYGGDR(me)GGGYGGDR(di)GGGYGGDR(me)GGYGGDR(me)		-1.136					+
Q92804	TAF15	R	483	me	GGGYGGDR(me)GGGYGGDR(di)GGGYGGDR(me)GGYGGDR(me)		-1.136					+
Q92804	TAF15	R	519	di	GGYGGDR(di)GGYGGDR(di)GGYGGDRGGYGGDR(di)			1.473				+
Q92804	TAF15	R	505	di	GGYGGDR(di)GGYGGDR(di)GGYGGDRGGYGGDR(di)			1.473				+
Q92804	TAF15	R	498	di	GGYGGDR(di)GGYGGDR(di)GGYGGDRGGYGGDR(di)			1.473				+
Q92804	TAF15	K	283	tr	PK(tr)GEATVSFDDPPSAK	0.266	0.463	0.279			0.352	
Q92804	TAF15	K	192	di	GYGGSQGGGR(di)GRGGYDK(di)DGR			-1.138				+
Q92804	TAF15	R	185	di	GYGGSQGGGR(di)GRGGYDK(di)DGR			-1.138				
Q92804	TAF15	R	187	me	GR(me)GGYDKDGR		0.798					+
Q92804	TAF15	R	475	di	GGGYGGDR(di)GGGYGGDR(di)GGGYGGDR(di)GGYGGDR		-1.805					+
Q92804	TAF15	R	467	di	GGGYGGDR(di)GGGYGGDR(di)GGGYGGDR(di)GGYGGDR		-1.805					+
Q92804	TAF15	R	483	di	GGGYGGDR(di)GGGYGGDR(di)GGGYGGDR(di)GGYGGDR		-1.805					+
Q92804	TAF15	K	210	me	DGRGPMTGSSGGDR(me)GGFK(me)	-0.312	-0.261	-0.308			-0.514	+
Q92804	TAF15	R	206	me	DGRGPMTGSSGGDR(me)GGFK(me)	-0.312	-0.261	-0.308			-0.514	+
Q92804	TAF15	K	576	di	SGGGYGGDR(di)GGGYGGDRGGYGGK(di)						-1.503	+
Q92804	TAF15	R	562	di	SGGGYGGDR(di)GGGYGGDRGGYGGK(di)						-1.503	+
Q92804	TAF15	R	187	di	GYGGSQGGGR(di)GR(di)GGYDKDGR	-1.252	-1.160	-1.308		-1.019		+
Q92804	TAF15	R	185	di	GYGGSQGGGR(di)GR(di)GGYDKDGR	-1.252	-1.160	-1.308		-1.019		+
Q92804	TAF15	K	192	di	GYGGSQGGGR(di)GRGGYDK(di)			-1.425		-1.064		+
Q92804	TAF15	R	185	di	GYGGSQGGGR(di)GRGGYDK(di)			-1.425		-1.064		+
Q92804	TAF15	R	535	di	SR(di)GGYGGDR(di)GGGSGYGGDR	-0.490	-0.564	-0.449	-0.478	-0.490	-0.485	+
Q92804	TAF15	R	528	di	SR(di)GGYGGDR(di)GGGSGYGGDR	-0.490	-0.564	-0.449	-0.478	-0.490	-0.485	+
Q92804	TAF15	R	505	di	GGYGGDR(di)GGYGGDR(di)GGYGGDR(di)GGYGGDR	1.363	1.198	1.400				+
Q92804	TAF15	R	512	di	GGYGGDR(di)GGYGGDR(di)GGYGGDR(di)GGYGGDR	1.363	1.198	1.400				+
Q92804	TAF15	R	498	di	GGYGGDR(di)GGYGGDR(di)GGYGGDR(di)GGYGGDR	1.363	1.198	1.400				
Q92804	TAF15	R	195	me	GGYDKDGR(me)GPMTGSSGGDR(me)GGFK	-0.984						+
Q92804	TAF15	R	206	me	GGYDKDGR(me)GPMTGSSGGDR(me)GGFK	-0.984						+
Q92804	TAF15	R	483	di	GGYGGDR(di)GGYGGDR	3.301		3.334				+
Q92804	TAF15	R	187	me	GYGGSQGGGRGR(me)		1.710					+
Q92804	TAF15	R	187	di	GYGGSQGGGR(di)GR(di)GGYDK	-1.378	-1.171	-1.345				+
Q92804	TAF15	R	185	di	GYGGSQGGGR(di)GR(di)GGYDK	-1.378	-1.171	-1.345				+
Q92804	TAF15	R	185	me	GYGGSQGGGR(me)GR	1.598	1.810					+
Q92804	TAF15	R	483	me	GGGYGGDR(me)GGYGGDR	3.081	3.351				2.729	+
Q92804	TAF15	R	206	me	GPMTGSSGGDR(me)GGFK		2.641				2.505	+
Q92804	TAF15	R	185	di	GYGGSQGGGR(di)GR		-0.593		-0.938			+
Q92804	TAF15	R	395	di	NSCNQCNEPRPEDSRPSGGDFR(di)GR	0.084	0.292					+
Q92804	TAF15	R	415	di	GGDR(di)GGYGGDR			2.019				
Q92804	TAF15	R	570	di	SGGGYGGDR(di)GGGYGGDR(di)GGYGGK	-1.596	-1.547	-1.619				+
Q92804	TAF15	R	562	di	SGGGYGGDR(di)GGGYGGDR(di)GGYGGK	-1.596	-1.547	-1.619				+
Q92804	TAF15	R	334	me	R(me)PEFMR(me)GGGSGGGR(me)			1.579				
Q92804	TAF15	R	321	me	R(me)PEFMR(me)GGGSGGGR(me)			1.579				+
Q92804	TAF15	R	326	me	R(me)PEFMR(me)GGGSGGGR(me)			1.579				+

Accession	Gene	AA	Pos	Mod	Sequence									+
Q92804	TAF15	K	576	me	SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)		-0.639	-0.677	-0.556	-0.578			-0.798	+
Q92804	TAF15	R	570	me	SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)		-0.639	-0.677	-0.556	-0.578			-0.798	+
Q92804	TAF15	R	467	di	SGGGYGGDR(di)GGGYGGDR(di)GGGYGGDR			-1.557						+
Q92804	TAF15	R	459	di	SGGGYGGDR(di)GGGYGGDR(di)GGGYGGDR			-1.557						+
Q92804	TAF15	R	187	me	GYGGSQGGGR(me)GR(me)GGYDK(me)		0.237	0.559	0.130	0.027			0.101	+
Q92804	TAF15	K	192	me	GYGGSQGGGR(me)GR(me)GGYDK(me)		0.237	0.559	0.130	0.027			0.101	+
Q92804	TAF15	R	185	me	GYGGSQGGGR(me)GR(me)GGYDK(me)		0.237	0.559	0.130	0.027			0.101	+
Q92804	TAF15	K	192	di	GYGGSQGGGRGRGGYDK(di)								2.207	+
Q92804	TAF15	R	459	di	SGGGYGGDR(di)GGGYGGDR		-1.547	-1.278						+
Q92804	TAF15	R	467	me	SGGGYGGDR(me)GGGYGGDR(me)GGGYGGDR(me)			-1.157						+
Q92804	TAF15	R	475	me	SGGGYGGDR(me)GGGYGGDR(me)GGGYGGDR(me)			-1.157						+
Q92804	TAF15	R	459	me	SGGGYGGDR(me)GGGYGGDR(me)GGGYGGDR(me)			-1.157						+
Q92804	TAF15	R	483	di	GGGYGGDR(di)GGYGGDR		0.188	0.311	0.239	0.071	0.098		-0.027	+
Q01844	EWSR1	R	486	me	GGPGGPGGPGGPMGR(me)		1.479		1.687					+
Q01844	EWSR1	R	506	di	GPR(di)GSR(di)GNPSGGGNVQHR		1.315	1.107	1.281	1.334				+
Q01844	EWSR1	R	503	di	GPR(di)GSR(di)GNPSGGGNVQHR		1.315	1.107	1.281	1.334				+
Q01844	EWSR1	R	503	di	GGDR(di)GGFPPRGPR(di)		1.267	1.148	1.229					+
Q01844	EWSR1	R	494	di	GGDR(di)GGFPPRGPR(di)		1.267	1.148	1.229					+
Q01844	EWSR1	R	471	me	GMPPPLR(me)GGPGGPGGPGGPMGR		1.067	1.071	0.814					+
Q01844	EWSR1	R	615	me	R(me)GGPGGPPGPLMEQMGGR			3.376	3.478					+
Q01844	EWSR1	R	494	di	GGDR(di)GGFPPR(di)GPR			0.959						+
Q01844	EWSR1	R	500	di	GGDR(di)GGFPPR(di)GPR			0.959						+
Q01844	EWSR1	R	471	me	GMPPPLR(me)GGPGGPGGPGGPMGR(me)		-0.899	-0.914	-0.868				-0.793	+
Q01844	EWSR1	R	486	me	GMPPPLR(me)GGPGGPGGPGGPMGR(me)		-0.899	-0.914	-0.868				-0.793	+
Q01844	EWSR1	R	461	me	KKPPMNSMR(me)GGLPPR(me)		-0.457	-0.452	-0.417					+
Q01844	EWSR1	R	455	me	KKPPMNSMR(me)GGLPPR(me)		-0.457	-0.452	-0.417					+
Q01844	EWSR1	R	633	me	R(di)GGPGGPPGPLMEQMGGR(me)		-1.202	-1.210	-1.321					+
Q01844	EWSR1	R	615	di	R(di)GGPGGPPGPLMEQMGGR(me)		-1.202	-1.210	-1.321					+
Q01844	EWSR1	R	486	me	GMPPPLR(di)GGPGGPGGPGGPMGR(me)		-1.226	-1.221						+
Q01844	EWSR1	R	471	di	GMPPPLR(di)GGPGGPGGPGGPMGR(me)		-1.226	-1.221						+
Q01844	EWSR1	R	506	me	GSR(me)GNPSGGGNVQHR		3.865							+
Q01844	EWSR1	K	641	tr	RGGRGGPGK(tr)MDK			1.267						
Q01844	EWSR1	R	633	me	R(me)GGRGGPGK(tr)MDK		-0.883	-0.881						
Q01844	EWSR1	K	641	tr	R(me)GGRGGPGK(tr)MDK		-0.883	-0.881						
Q01844	EWSR1	R	506	di	GSR(di)GNPSGGGNVQHR	1.155	1.234	1.162	1.161	0.564				+
Q01844	EWSR1	R	455	me	KKPPMNSMR(me)GGLPPR		2.060	2.117	1.872					+
P17844	DDX5	R	502	me	R(me)GGFNTFR		-0.479	-0.486	-0.855		-0.377	-0.502	-0.573	+
P17844	DDX5	R	502	di	R(di)GGFNTFR						-1.581			+
P17844	DDX5	K	391	me	HGK(me)APILIATDVASR						-0.024			
P17844	DDX5	K	490	tr	SR(me)GRGGMK(tr)DDR			-3.045						+
P17844	DDX5	R	484	me	SR(me)GRGGMK(tr)DDR			-3.045						+
P17844	DDX5	K	56	di	FEK(di)NFYQEHPDLAR		1.025	0.939						+
P17844	DDX5	K	340	me	LMEEIMSEK(me)						-0.068			+
P11021	HSPA5	K	585	tr	LGGK(tr)LSSEDKETMEK	0.326	0.197	0.449	0.511	0.408		0.331	0.424	

UniProt	Gene	Res	Pos	Type	Peptide								
Q92841	DDX17	R	505	me	FVINYDYPNSSEDYVHR(me)	-1.079				-0.606	-0.590	-0.350	
Q92841	DDX17	R	684	me	SSQSSSQQFSGIGR(me)	-0.418	-0.430	-0.507	-0.542				+
P19338	NCL	R	687	di	GR(di)GGFGGR(di)GGFR(di)GGR		2.248						
P19338	NCL	R	681	di	GR(di)GGFGGR(di)GGFR(di)GGR		2.248						
P19338	NCL	R	691	di	GR(di)GGFGGR(di)GGFR(di)GGR		2.248						
P19338	NCL	R	687	di	GR(di)GGFGGR(di)GGFR	2.509	2.383	2.377	2.489				
P19338	NCL	R	681	di	GR(di)GGFGGR(di)GGFR	2.509	2.383	2.377	2.489				
P19338	NCL	R	694	di	GGR(di)GGGGDHKPQGK	1.635	1.506	1.506					+
Q12906	ILF3	R	609	me	RAPVPVR(me)GGPK	1.699	1.952	1.803					+
Q12906	ILF3	K	613	me	APVPVRGGPK(me)	2.183							+
Q12906	ILF3	R	609	di	RAPVPVR(di)GGPK	-0.670	-0.661	-0.466	-0.552	-0.271		-0.246	+
P31943	HNRNPH1	R	217	di	PGPYDRPGAGR(di)GYNSIGR(di)GAGFER	-0.182	-0.194	-0.089	-0.188		0.184	0.119	+
P31943	HNRNPH1	R	224	di	PGPYDRPGAGR(di)GYNSIGR(di)GAGFER	-0.182	-0.194	-0.089	-0.188		0.184	0.119	+
P31943	HNRNPH1	K	68	me	EGRPSGEAFVELESEDEVK(me)		0.276	0.411			0.050		
Q08211	DHX9	R	1223	di	GVSR(di)GGFR(di)GNSGGDYR				2.759				+
Q08211	DHX9	R	1227	di	GVSR(di)GGFR(di)GNSGGDYR				2.759				+

Appendix 3. Methylated peptides commonly identified upon PRMT1 depletion and overexpression

List of methylated peptides commonly identified in PRMT1 KD and overexpression experiments through the SILAC/co-IP strategy. The sequence of the modified peptide identified by MS (Modified Sequence), the averaged $\log_2$ H/L SILAC ratios of modified peptides identified in DDX5 IP, DGCR8 IP and FUS IP upon KD and overexpression is reported (for PRMT1 depletion experiment, ratios from both 'forward' and 'reverse' experiment are reported). The sites identified within each peptide are indicated.

Modified Sequence	PRMT1 KD		PRMT1 Over	Modified sites
	average Fwd	average Rev	log2 ratio over	
HNRNPUL1: GGGGFR(di)GR(di)GGGGGFQR	-0.261		-0.119	R529, R531
FUS: SSGGYEPR(di)GR(di)		-0.314	-0.389	R242, R244
FUS: GGR(di)GR(di)GGSGGGGGGGGGGGYNR(me)	-0.726		0.104	R216, R218, R234
FUS: GGR(di)GGGDR(di)GGFGPGK		-0.708	-0.070	R216, R218
FUS: GGR(di)GR(di)GGSGGGGGGGGGGYNR	-0.175	0.027	-0.091	R498, R503
FUS: GGGDR(me)GGFGPGK	2.857	3.748	-1.143	R503
HNRNPDL: MFIGGLSWDTSKK(me)	0.335		0.160	K162
HNRNPDL: MFIGGLSWDTSK(me)	0.331		0.111	K161
HNRNPDL: ASR(di)GGGNHQNNYQPY	-0.073	0.016	0.020	R408
TAF15: GGYGGDR(di)GGYGGDR(di)GGYGGDR	1.875	1.721	-1.511	R505, R512
TAF15: GGGYGGDR(di)GGGYGGDR(di)GGYGGDR	-1.186		0.070	R475, R483
TAF15: PK(tr)GEATVSFDDPPSAK	0.336	0.352	-0.171	K283
TAF15: DGRGPMTGSSGGDR(me)GGFK(me)	-0.294	-0.514	-0.115	K210, R206
TAF15: SR(di)GGYGGDR(di)GGGSGYGGDR	-0.501	-0.484	-0.055	R528, R535
TAF15: GYGGSQGGGR(di)GR(di)GGYDK	-1.298		0.288	R185, R187
TAF15: SGGGYGGDR(di)GGGYGGDR(di)GGYGGK	-1.588		0.599	R562, R570
TAF15: SGGGYGGDR(me)GGGYGGDR(me)GGYGGK(me)	-0.624	-0.688	-0.110	R562, R570, K576
TAF15: GYGGSQGGGR(me)GR(me)GGYDK(me)	0.309	0.064	-0.204	R185, R187, K192
EWSR1: GPR(di)GSR(di)GNPSGGGNVQHR	1.234	1.334	-0.403	R503, R506
EWSR1: GMPPPLR(me)GGPGGPGGPGGPMGR	0.984		-0.770	R471
EWSR1: R(me)GGPGGPPGPLMEQMGGR	3.427		-0.116	R615
EWSR1: GMPPPLR(me)GGPGGPGGPGGPMGR(me)	-0.894	-0.793	0.082	R471, R486
EWSR1: GMPPPLR(di)GGPGGPGGPGGPMGR(me)	-1.223		0.191	R471, R486
DDX5: TTYLVLDEADR(di)		0.031	0.072	R252
DDX5: APILIATDVASR(me)	-0.033	0.064	0.012	R403
DDX5: R(me)GGFNTFR	-0.607	-0.484	-0.341	R501
DDX5: ELAQQVQQVAAEYCR(me)	0.008		0.012	R192
DDX5: APILIATDVASR(di)	0.057		0.016	R403
DDX17: TTSSANNPNLMYQDECDR(me)R	0.141		0.020	R586
DDX17: SSQSSSQQFSGIGR(me)	-0.474		-0.050	R684
NCL: EALNSCNK(me)R	0.008		-0.059	K545
HNRNPH1: EGRPSGEAFVELESEDEVK(me)	0.343	0.050	-0.024	K68
HNRNPH1: YGDGGSTFQSTTGHCVHMR(me)	-0.070		0.028	R294
HNRNPH1: PGAGR(di)GYNSIGR(di)GAGFER	-0.079	0.095	-0.070	R217, R224
HNRNPM: MGLSMER(me)	0.110	0.133	0.160	R531
ILF3: APVPVR(di)GGPK / APVPVR(me)GGPK(me)	-0.360		-0.421	R609, K613
TAF15: GGGYGGDR(di)GGYGGK / GGGYGGDR(me)GGYGGK(me)	0.296	0.261	-0.180	R570, K576
FUS: GGPMGR(di)GGYGGGGSGGGGR(me)GGFPSGGGGGGGGQQR(me) / GGPMGR(di)GGYGGGSGGGGR(di)GGFPSGGGGGGGQQR	-0.198	0.522	-0.327	R394, R407
FUS: GGR(di)GRGGSGGGGGGGGGGYNR / GGR(me)GR(me)GGSGGGGGGGGGGYNR	4.012		-0.550	R216, R218
EWSR1: GMPPPLR(me)GGPGGPGGPGGPMGR / GMPPPLRGGPGGPGGPGGPMGR(me	0.984		-0.955	R464
TAF15: GYGGSQGGGR(di)GR / GYGGSQGGGR(me)GR(me	-0.593	-0.938	-0.231	R185, R187
ILF3: RAPVPVR(di)GGPK / RAPVPVR(me)GGPK(me)	-0.587	-0.258	-0.425	R609, K613

121