



## PROJECT FINAL REPORT

**Grant Agreement number:** 277899

**Project acronym:** 4DCellFate

**Project title:** Developing a global understanding of the PRC and NuRD complexes in stem cell differentiation, in health and disease

**Funding Scheme:** Collaborative Project

**Period covered:** **from** 01.12.2011 **to** 30.11.2016

**Submission date:** 23.02.2017

**Name of coordinator:** CRG P1a, Luciano Di Croce  
**Tel:** +34 93 316 11 32  
**E-mail:** Luicano.DiCroce@crg.eu

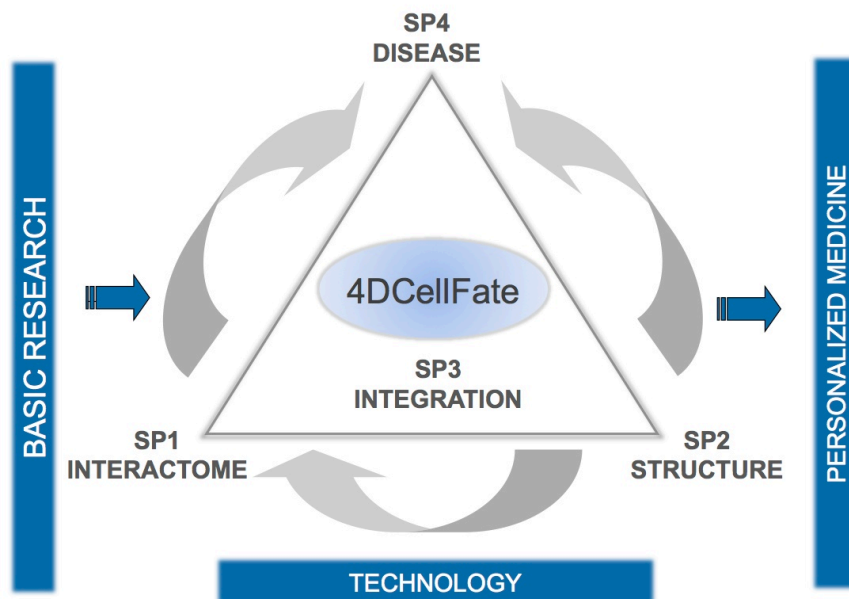
**Project website address:** <http://www.4dcellfate.eu>

## Table of Contents

|  |           |
|--|-----------|
| <b>Table of Contents</b> .....   | <b>2</b>  |
| <b>1. Executive summary</b> .....  | <b>3</b>  |
| <b>2. Summary description of project context and objectives.</b> .....   | <b>4</b>  |
| <b>3. Description of the project’s main S&amp;T results/foregrounds</b> .....  | <b>6</b>  |
| 3.1. SUBPROJECT 1: Polycomb repressive complex & NuRD Interactome .....  | 6         |
| 3.2. SUBPROJECT 2: Polycomb repressive complex & NuRD Structural Biology   | 15        |
| 3.3. SUBPROJECT 3: Polycomb repressive complex & NuRD Data Integration ...   | 25        |
| 3.4. SUBPROJECT 4: Polycomb repressive complex & NuRD In Disease .....   | 29        |
| <b>4. Potential impact (including the socio-economic impact and the wider societal implications of the project so far) and the main dissemination activities and exploitation of results</b> ..... | <b>33</b> |
| 4.1. Socio-economic impact and exploitation of results.....  | 33        |
| 4.2. Wider societal implications of the project.....   | 35        |
| 4.3. main dissemination activities .....   | 36        |
| <b>Annex</b> .....   | <b>39</b> |
| <b>Publications</b> .....  | <b>39</b> |

## 1. Executive summary

How does an undifferentiated cell "decide" which type of cell it will become? We are addressing a portion of this question by elucidating the role in cell fate taken by the epigenetic regulator complexes Nucleosome Remodelling and Deacetylation (NuRD) complex and the Polycomb Repressive Complexes (PRCs). These complexes have been previously shown to be critical in cell fate decisions made by embryonic stem (ES) cells. The basic mechanism of action for these complexes is well-established: they modulate the accessibility to the nucleosome by interacting directly with the histone proteins. This in turn dictates not only the level of chromatin compaction, but also the further interactions of transcription factors and other regulators with the DNA sequences. However, a highly interesting aspect of these complexes that is just now becoming clear is that multiple combinations of paralogous subunits generate considerable functional diversity among these complexes. This combinatorial assembly is also further complicated by their dynamic composition. Several publications (including some from this project) had shown that this dynamic composition can determine which gene networks these complexes regulate, and the type of regulation that they induce.



**Figure 1:** Concept and objectives of the 4DCellFate project.

Interestingly, a "reverse" type of differentiation can occur in cancer cells, in which a differentiated cell reverts to a "cancer" stem cell, as frequently occurs in leukemia. We would therefore also have intended to determine to what extent NuRD and PRC1/2 play a role in this reversion, and whether they could be manipulated to force such cancer stem cells to differentiate. This biochemical and structural work was also considered to be critical also for the molecular biologists in their decisions about how to approach analyzing the dynamics of PRC and NuRD functions and mechanisms *in vivo* and *in cell culture*. In Subproject 1 (WP1–3, Figure 1) we examined the roles of the specific subtypes of complexes, to distinguish

between their potential roles in cell differentiation (using ES cells). We also studied whether the distinct PTMs modulate these decisions. This could have occurred by changing the preferences of the complexes in their protein-protein, protein-nucleosome, or protein-nucleic acids interactions. Such interactions could have also been regulatory; for instance, interactions with long non-coding RNAs could regulate the complexes, and we investigated this aspect.

In Subproject 2 (WP4–6, Figure 1), we were analyzing the NuRD and PRC complexes by mass spectrometry to determine the exact composition of these. This allowed us to begin mass-producing the complexes in vitro to carry out detailed structural analyses, using X-ray crystallography, NMR, and electron microscopy. In parallel, we were determining the network of post-translational modifications (PTMs) that each component of the complexes is likely to undergo by mass spectrometry.

In Subproject 3 (WP8, Figure 1), data integration tools have been developed that are available to the consortium. Further, the different types of data generated by the project were being analyzed by a systems biology approach to elucidate details about how the expression, structure, interactions, and activity of the NuRD and PRC complexes varies with time during stem cell differentiation.

Finally, we were working on determining how alterations in PRC/NuRD complex activities contribute to disease and, importantly, whether it is possible to manipulate the function through specific small molecule inhibitors (Subproject 4, WP8–9, Figure 1). We aimed not to block the activity of all possible NuRD and PRC1/2 complex variations, but to interfere specifically with those that have been proven to be responsible for the cancer stem cell phenotype, or to modulate those important for a specific differentiation lineage. Such manipulation opens a plethora of opportunities for cancer treatments.

## 2. Summary description of project context and objectives.

### **Polycomb repressive complex & NuRD Interactome**

The aim of Sub-project 1 (SP1) was to understand the complex networks in which the NuRD/PRC proteins are involved, the way they interact with the genome, and the way they control changes in gene expression as ES cells differentiate. We use proteomics experiments to understand the complexes formed by NuRD/PRC proteins (WP1), and chromatin-immunoprecipitation experiments linked to high-throughput DNA sequencing (ChIP-seq) to understand the way they interact with the genome (WP2). This analysis was then correlated with genome-wide RNA-seq experiments to understand how these complexes control gene expression (WP2). Finally, we initiated several high-throughput genetic screens to identify new components and/or regulators of these complexes (WP3). The large amount of results generated during these 5 years indicate that variations in the assembly of both Polycomb and NuRD complexes are finely regulated to proper balance pluripotency and stem cell differentiation. Moreover, the chemical compounds we have identified will be of great use not only for further studies but also as potential drugs to modulated Polycomb and NuRD activities in vivo.

### **Polycomb repressive complex & NuRD Structural Biology**

The objective of Sub-project 2 (SP2) was to generate structural knowledge of the different complexes and their components. This has been a daunting challenge due to the flexibility,

size, and heterogeneity of the complexes involved. A highly collaborative and integrative approach, combining cross-linking/Mass Spectrometry (MS), Electron Microscopy (EM), Small Angle X-Ray Scattering (SAXS), X-ray crystallography, and other biochemical/biophysical approaches has been applied to generate a comprehensive 3D understanding of how PRC/NuRD complexes assemble. Our studies in SP2 have been greatly facilitated by the very fruitful collaboration with colleagues working in SP1 (mainly WP1 and WP2) to study both PRC and NuRD proteomics and PRC- and NuRD-genome interactions using ChIP-seq experiments. The structures we have determined, and are in the process of generating, will allow us to understand how PRC and NuRD complexes functions to alter chromatin structure. They are also providing the basis for the design of small molecules that can interfere with NuRD function.

### **Polycomb repressive complex & NuRD Data Integration**

Global changes in gene expression heterogeneity are associated with cell fate potential and particular functional gene classes tend to be enriched for high or low noise levels. However, each gene's underlying promoter architecture, local chromatin state and its broader regulatory environment are likely to be major determinants of these differences in the mode of transcription. The aim of this sub-project was to combine diverse data types generated in this project to provide an integrated view of how the PRC and NuRD complexes function to regulate cellular differentiation. The ultimate objective was to construct a predictive model that relates PRC and NuRD diversity, activity, and localisation to changes in gene expression and the differentiation process. We combined recently published single-cell RNA-seq data from mouse ESCs and cortex with matched publicly available chromatin state datasets to obtain an unbiased overview of features involved in the control of transcriptional consistency. Our results suggest that histone modifications directly influence transcriptional variability within tissues. However, an important caveat is the potential existence of cell-to-cell variability in chromatin status, which is invisible in data from the population-based ChIP-seq assays used in this analysis. This model was evaluated and refined in an iterative process of integration, prediction, and experimental validation.

Further, as part of the subproject the data generated by the consortium were made available to all consortium members and a user-friendly software tool was developed to facilitate data integration and analysis by wet-lab biologists. Generated datasets and software tools were made available to the wider scientific community.

### **Polycomb repressive complex & NuRD In Disease**

The aim of this subproject was to better understand the human NuRD and PRC complexes, and to use the information gleaned from the rest of the project to understand and manipulate NuRD- or PRC-dependent processes in human cells. Specifically, we sought to use our knowledge of PRC complexes to identify small molecules that would impact the properties of tumour cells, and do capitalise on our knowledge of NuRD biology to improve differentiation protocols for human pluripotent cells. The work in SP4 has benefited greatly from advances made in the other subprojects. SP1 provided important information about complex biochemistry and genome-wide localisations, SP2 helped us to understand the structure of these complexes, and SP3 was important in integrating all of the data together into a coherent picture. The inhibitors and protocols discovered through the work carried out in this subproject provide important ground work which will ultimately lead to improved therapeutic applications.

### 3. Description of the project's main S&T results/foregrounds

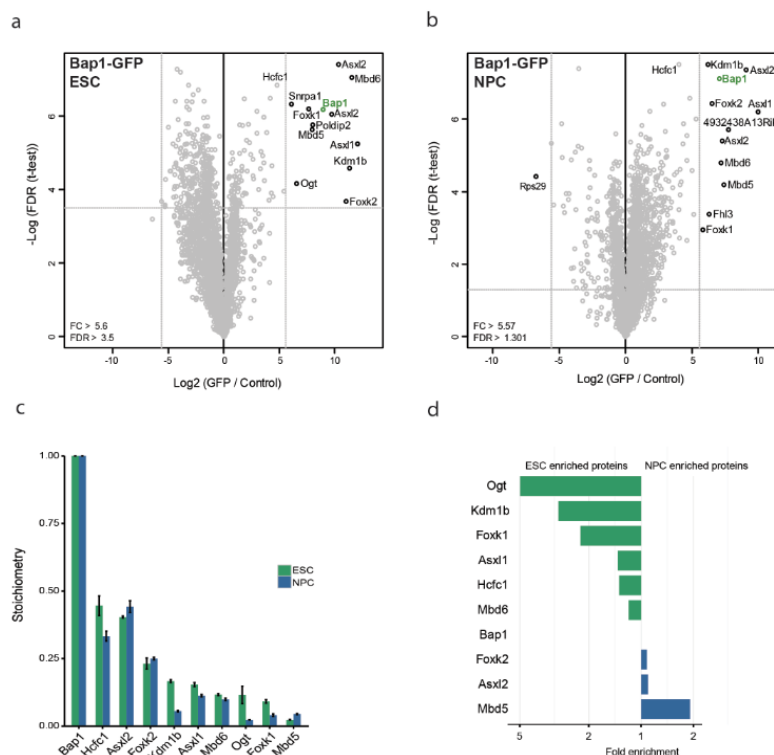
#### 3.1. SUBPROJECT 1: Polycomb repressive complex & NuRD Interactome

##### Proteomics over time

BAC-GFP constructs have been obtained for a number of NuRD and PRC complex subunits: For Polycomb: Asxl2, BAP1, EED, Ezh2, CBX8, CBX7, Ring1, C17orf96 (new PRC2 interactor)

For NuRD: MBD3, MBD2, CDKAP1 (DOC-1), MTA1, MTA2, MTA3, Zmynd8, Znf296, Znf219 (new NuRD interactors)

Many of these constructs have been transfected into mouse embryonic stem cells and stable cell lines have been made. GFP-based single-affinity purifications combined with label-free quantitative mass spectrometry-based proteomics (AP-MS) experiments are performed to characterize the protein complexes for all the tagged subunits, including their stoichiometry. To study the dynamics of Polycomb and NuRD protein complexes during stem cell differentiation, we make use of established protocols to differentiate mouse ES cells into neuronal precursor cells (NPCs) and finally into astrocytes or neurons. During the differentiation process, the BAC-GFP cell lines maintain nearly endogenous expression of the tagged Polycomb or NuRD complex subunits. Nuclear extracts are generated from these different cell types and AP-MS experiments are then performed to characterize the protein complexes. These experiments are yielding very interesting results. In the previous reporting period, we presented the purification and characterization of PRC1, PRC2 and NuRD in ES cells and NPCs, which revealed that the composition of these complexes is highly dynamic during cellular differentiation. During this past year, we also purified the PR-DUB complex from ESCs and NPCs (**Figure 2**).

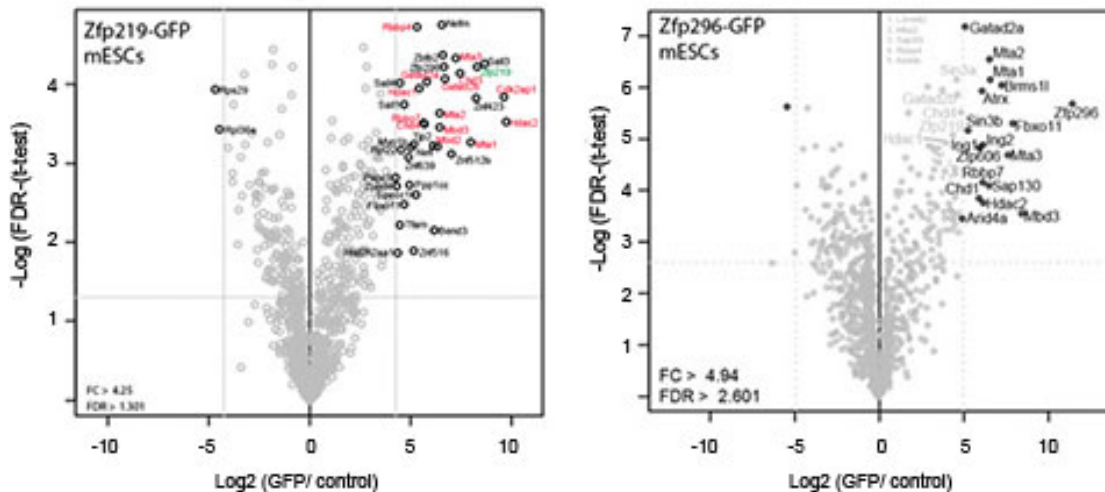


**Figure 2: AP-MS (A,B) and stoichiometry determination (C,D) of PR-DUB in mouse ESCs and neuronal precursor cells.** GFP-Bap1 pull-downs from ESCs (A) and NPCs (B) are plotted in a volcano plot (C). Stoichiometry determination indicates the core PR-DUB complex is very stable during differentiation, suggesting a conserved functional role. (D) The ratios of ESC/NPC stoichiometry values are plotted on a logarithmic scale.

This complex also shows a few dynamic subunits/interactors during differentiation, including Mbd5, which has a higher abundance in the NPC complex. However, the core of the PR-DUB complex remains stable during cellular differentiation.

We also performed ChIP-seq analyses for PRC1, PRC2, and NuRD to determine the differences in genome-wide binding sites for these complexes in mouse ES cells and neural progenitor cells (see below).

We have also generated tagged lines for both known and novel NuRD subunits/interactors. A transgenic BAC mouse ES line was created for the novel NuRD-interacting protein Zfp219. We also have a mouse ES line containing stably integrated GFP-Zfp296, another novel NuRD interacting protein. Label-free GFP pull-downs of these cell lines confirmed their interaction with NuRD. Interestingly, Zfp296 turns out to be a shared subunit between the NuRD and Sin3/Hdac complex in mouse ES cells (**Figure 3**). In parallel we also profile genome-wide binding targets for NuRD in ESCs and NPCs using ChIP-sequencing (see below).

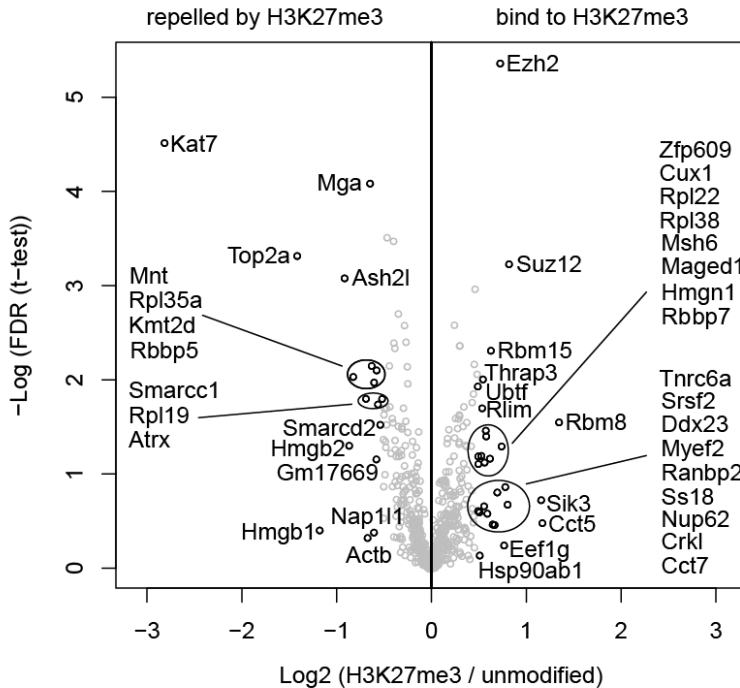


**Figure 3:** Label-free GFP affinity purifications in nuclear extracts from Zfp219-GFP (left) and Zfp296-GFP (right) mouse ES lines.

The Bartke lab (YIP3) has generated recombinant nucleosomes marked with H3K27me3. Together with the Bartke lab, we have performed pulldowns utilizing these recombinant nucleosomes together with lysates from mouse ES and NP cells.

As a positive control, we observed that the PRC2 complex binds to H3K27me3 in both the mESC and NPC pulldowns, specifically the Suz12 and Ezh2 subunits. Interestingly, subunits from many chromatin modifying complexes were repelled by the H3K27me3 modification in mESCs. These include SWI/SNF family (Smarcc1, Smarcd2), MLL family (Ash2l, Kmt2d), HBO1 (Kat7), and PRC1 (Mga).

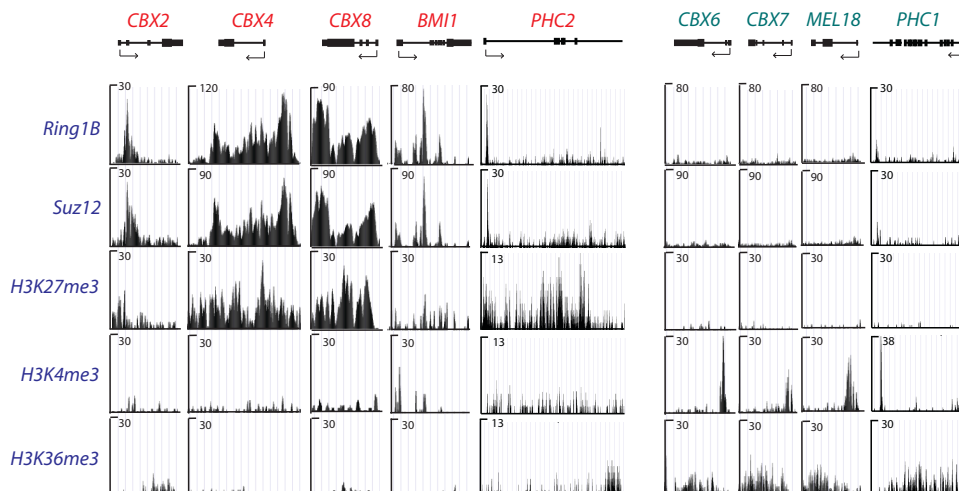
We also observed Top2a, a DNA topoisomerase, as a protein repelled by H3K27me3 in both mESCs and NPCs. In the NPC pulldowns, we identified the lysine demethylase Kdm5b as a prominent reader of the H3K27me3 modification (**Figure 4**).



**Figure 4:** Interaction plots of pull-downs using H3K27me3-modified dinucleosomes and mESC lysates. H3K27me3 interactors are on the right side of the plot, and proteins repelled by H3K27me3 are on the left.

### Integration of proteomics with genome-wide analysis and single molecule fluorescence microscopy

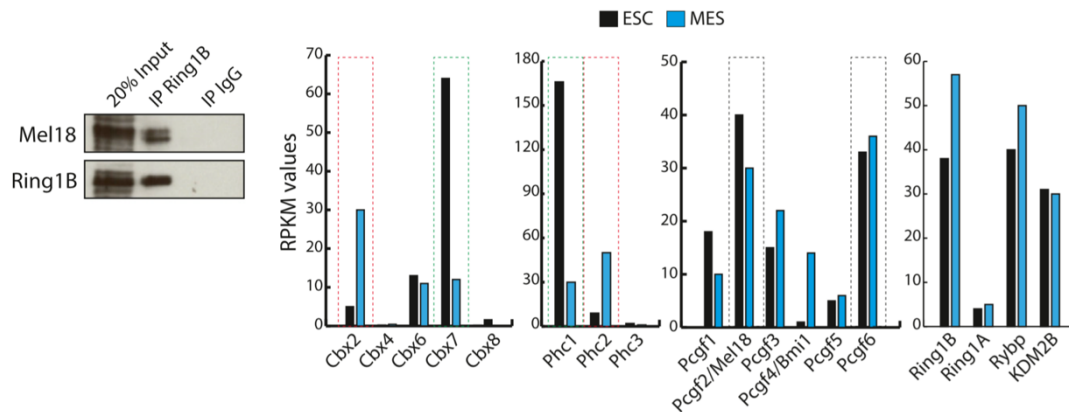
We have analysis of Polycomb subunits in proliferating and differentiating ES cells. Among those, we have characterized the genome-wide occupancy of Cbx2, Cbx6, PHC1, and Pcgf2/Mel18, which now can be integrated with our previous analyses of Suz12, Ring1b, Pcgf4/Bmi, Rybp, Cbx2, Cbx4, Cbx7, and Cbx8 (**Figure 5**). Additionally, we have also performed ChIP-seq analysis for several NuRD components, including MBD3, Mta1, Mta2, and Chd4.



**Figure 5.** Genome-wide distribution of several histone marks and Polycomb proteins in undifferentiated ES cells

We have also determined that Mel18 interacts with Ring1B in early cardiac-mesoderm precursors cells (MES) and measured the expression levels of all cPRC1 and ncPRC1 subunits (**Figure 6**). Cbx7 and Phc1 were strongly downregulated in MES cells, with Cbx2 and Phc2 the main Cbx and Phc family members expressed in these cells. Mel18 and Pcgf6 were the strongest Pcgf proteins expressed in MES cells. Strikingly, while global H2AK119ub

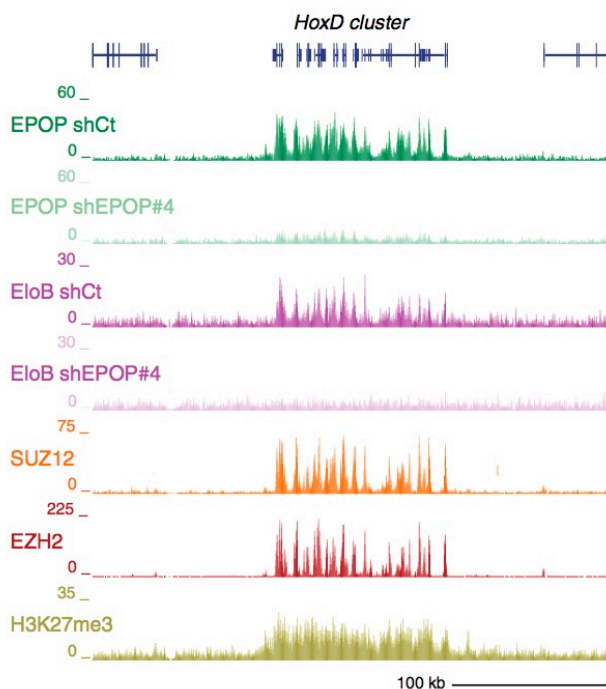
levels were constant in ESCs as compared to MES cells, H3K27me3 levels were reduced, concomitant with an accumulation of H3K27ac levels. Concordant with the effects of Mel18 depletion in ESCs, the levels of H2AK119ub and H3K27me3 also remained unaltered in shMel18 MES cells. Western blot analysis of several PRC1 subunits confirmed that Cbx2 and Phc2 were strongly upregulated, and that Mel18 was marginally downregulated in MES cells (**Figure 6**).



**Figure 6: PRC1 Composition, and expression in MES cells.** Endogenous coIP, using a Ring1B and IgG antibodies from MES cells (left). RNA-seq expression of PRC1 subunits in MES cells represented by FPKM values (right).

These data suggested that a new cPRC1 complex containing Cbx2 and Phc2 was assembled in MES cells, and that Mel18 was required to stabilize cPRC1 but not (as in ESCs) ncPRC1.

Furthermore, have identified C17orf96/EPOP as new interactor of PRC2 complex. This complex also contains Elongin B (EloB, or TCEB2) and Elongin C (EloC, or TCEB1).

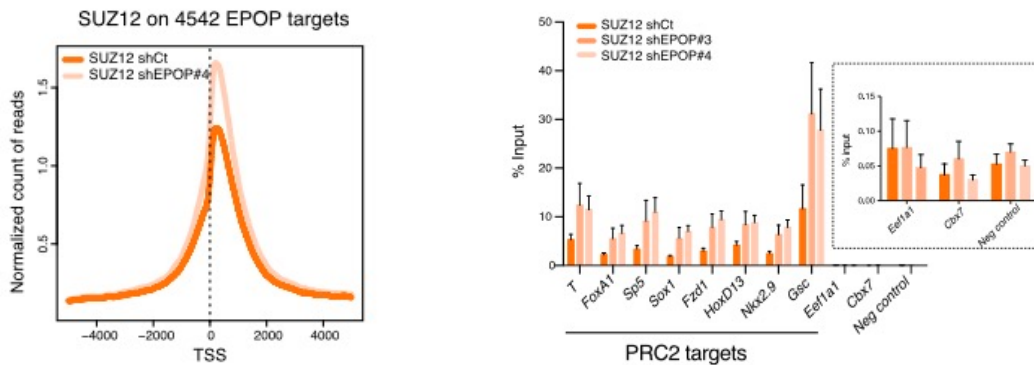


**Figure 7.** Genome-wide distribution of EPOP, EloB, H3K27me3 histone marks, and Polycomb proteins in undifferentiated ES cells.

The genome-wide analysis of the EPOP was reported in our previous annual report. This year we have fully analyzed the EloBC dimer in ES cells. ChIP-seq of endogenous EloB identified around 1000 significant peaks with highly enriched binding of EloB in WT mES cells (**Figure 7**).

The vast majority (752 out of 872 genes, 86%) of the EloB targets were also EPOP targets. Strikingly, the EloB ChIP signal at the EPOP-PRC2–occupied genes was drastically reduced in cells depleted of EPOP (Figure 5), in line with the biochemical observation that EPOP is required for efficient chromatin association of EloB.

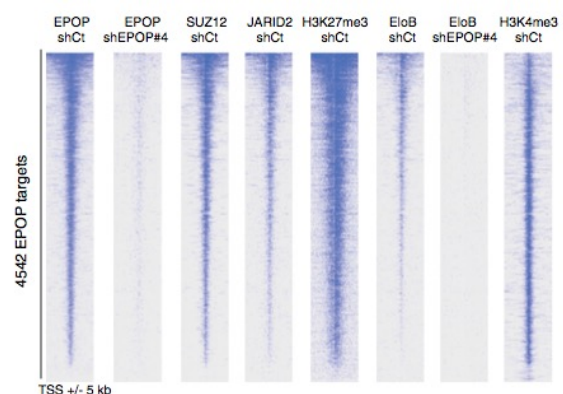
To understand the mechanistic basis for the requirement of EPOP and Elongin BC for low expression of PRC2 targets, we tested the effects of EPOP depletion on PRC2 binding to chromatin. In contrast to the function observed for other PRC2-associated factors like JARID2, EPOP is not required for PRC2 binding to PcG targets. Rather, we observed an increase in ChIP signal for both SUZ12 and the H3K27me3 mark on PRC2 targets after EPOP depletion (**Figure 8**).



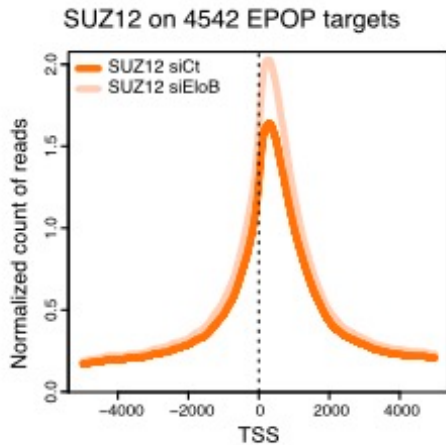
**Figure 8.** TSS (+/-5 kb) enrichment plots of SUZ12 ChIP-seq at 4,542 EPOP target sites from control (shCt) or EPOP-depleted (shEPOP#4) cells (left panel). Validation of effect of EPOP depletion on SUZ12 binding by ChIP-qPCR (right panel).

Because EPOP and JARID2 do not associate simultaneously with PRC2 yet occupy largely overlapping target sites on the genome (**Figure 9**), we also determined the ChIP signal of JARID2 after depletion of EPOP.

**Figure 9.** Heatmaps around the TSS (+/-5 kb) of 4,542 EPOP target genes. Genes are ranked by EPOP ChIP-seq intensity.



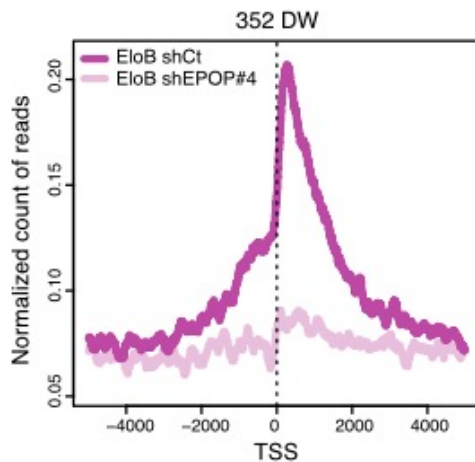
We also reported that Elongin BC regulates PRC2 binding to chromatin. Depleting Elongin BC by siRNA for 48 hr led to SUZ12 levels increased by more than 1.5-fold at 4542 targets, largely phenocopying the effect of EPOP depletion (**Figure 10**), showing that the function of EPOP and Elongin BC to impede excessive PRC2 binding shapes the transcriptional output.



**Figure 10.** TSS ( $\pm 5$  kb) enrichment plots of SUZ12 ChIP-seq at 4,542 EPOP target sites (left panel) or EPOP-negative control regions (right panel) after 48 hr treatment with control (siCt) or siEloB.

On the other hand, depletion of Elongin BC resulted in roughly equal numbers of genes up- or downregulated, indicating that the Elongin BC heterodimer has diverse roles in regulating transcriptional output.

Combination of expression and ChIP-seq results revealed that these genes exhibited a strongly decreased ChIP signal of EloB after EPOP depletion, whereas the 92 genes with an increased expression were independent of EloB (**Figure 11**). This indicates that the function of EPOP in bringing Elongin BC to PcG targets is indeed linked to the transcriptional status of these genes.



**Figure 11.** TSS ( $\pm 5$  kb) enrichment plots of EloB ChIP-seq from control (shCt) or EPOP-depleted (shEPOP#4) cells.

### **Single molecule fluorescence microscopy**

Our original aim was to develop two-colour single-molecule co-localisation imaging for the study of protein complexes. However, during the course of this project we have realised that, because many proteins co-localise to chromatin at sites of transcription, co-localisation does not necessarily help us understand the assembly and function of chromatin complexes. We have therefore found it more informative to carry out live-cell tracking of single proteins in the presence/absence of other components in order to understand the role of protein interactions within a complex. In this project, we have thus developed both fixed and live cell 2D and 3D single-molecule imaging (PALM and STORM) in mouse ES cells. This has allowed us to study either: 1) co-localisation and clustering of proteins within the nucleus; or 2) the role of

protein interactions within a complex by tracking single proteins in the presence/absence of other components.

### **Development of fixed and live cell 2D and 3D single-molecule imaging (PALM and STORM)**

For 2D single-molecule imaging, UCAM (Laue) in collaboration with Prof David Klenerman (Cambridge, UK) has built an nTIRF microscope. It incorporates 4 laser lines (405, 488, 561, 640 nm) and a dual view camera for multiple fluorophore imaging in either PALM (using mEos3.2) or STORM mode (using AF647). We have imaged a range of nuclear proteins such as the centromeric histone H3 variant CENP-A, the heterochromatin proteins HP1 $\beta$  and HP1 $\gamma$ , a nuclear periphery marker (lamin B2), and NuRD components (MBD3 and CHD4) (**Figure 12**). To determine co-localisation in ES cells on a 2D microscope, it is critical to ensure that localised molecules are within the detection plane (i.e. localised together) and not above or below this plane. We therefore developed and published an optical sectioning technique<sup>1</sup>. To study single ES cells, we also developed and published a novel microfluidic device capable of trapping cells for super-resolution imaging before releasing them for further processing<sup>2</sup>. To achieve 3D super-resolution imaging, the Klenerman group built a microscope with double-helix point spread function (DH-PSF) detection and adapted it for mammalian imaging (Carr *et al*, submitted). We have imaged single mEos3.2 and HaloTag®-JF549 dye-labelled CENP-A and CHD4 proteins with high x, y and z resolution over a 4  $\mu\text{m}$  range in z (**Figure 13**). We have used these microscopes to study the following:

#### **Clustering of NuRD proteins within the nucleus**

Taking advantage of the expertise of UCAM (Hendrich), knock-in cell lines were generated in which MBD3 and CHD4 were tagged at the C-terminus with mEos3 and the HaloTag®. Super-resolution imaging of the NuRD components MBD3 and CHD4 was demonstrated and published (**Figure 14**). CHD4 and MBD3 molecules form distinct foci in live ES cells suggesting that the NuRD complex forms locally enriched clusters. Analysis of these protein clusters has revealed that CHD4 binds to a considerably higher number of chromatin sites in the nucleus than MBD3, consistent with reports that it has significant non-NuRD functions (Stevens *et al*, in press).

#### **Single-particle tracking in the presence/absence of other components to understand the role of protein interactions within a complex**

If NuRD components co-localise/interact, the dynamics of specific proteins within the complex will be affected by the absence of other components. We investigated whether CHD4 dynamics is affected by the presence/absence of MBD3 using the knock-in cell lines described above and published this work (**Figure 12**)<sup>3</sup>. We measured CHD4 dynamics using PALM of mEos3.2-tagged CHD4 at 10 ms time resolution in both wild-type and MBD3-null cells. CHD4 exhibits two major diffusion coefficients ( $0.100 \pm 0.008$  and  $0.65 \pm 0.04 \mu\text{m}^2\text{s}^{-1}$ ).

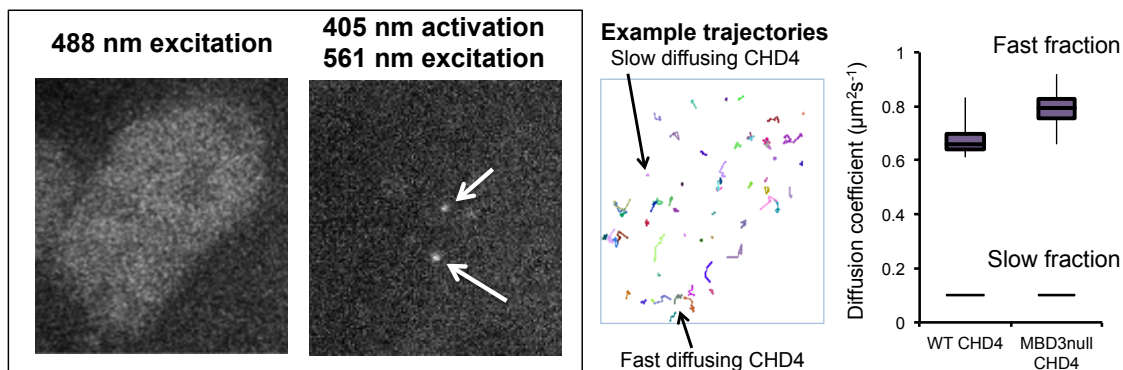
<sup>1</sup> Palayret, M. *et al*. Virtual-'light-sheet' single-molecule localisation microscopy enables quantitative optical sectioning for super-resolution imaging. *PLoS One* **10**, e0125438, doi:10.1371/journal.pone.0125438 (2015).

<sup>2</sup> Zhou, Y. *et al*. A microfluidic platform for trapping, releasing and super-resolution imaging of single cells. *Sens Actuators B Chem* **232**, 680-691, doi:10.1016/j.snb.2016.03.131 (2016).

<sup>3</sup> Zhang, W. *et al*. The Nucleosome Remodeling and Deacetylase Complex NuRD Is Built from Preformed Catalytically Active Sub-modules. *Journal of molecular biology* **428**, 2931-2942, doi:10.1016/j.jmb.2016.04.025 (2016)

In MBD3-null cells, CHD4 still exhibits the slowly diffusing fraction ( $0.100 \pm 0.008 \mu\text{m}^2\text{s}^{-1}$ ), but the fast fraction now moves significantly more quickly ( $0.80 \pm 0.08 \mu\text{m}^2\text{s}^{-1}$ ) ( $p < 0.004$ ). Since our published paper, we have carried out a similar analysis using our 3D microscope. Longer trajectories were achieved by combining the reduced photobleaching of HaloTag<sup>®</sup> dye-labelled CHD4 (compared to mEos3.2) and 3D tracking to prevent molecules diffusing out of the detection plane (**Figure 13**) (Carr *et al*, submitted). This study not only confirmed our previous 2D tracking results but allowed mean squared displacement analysis of the trajectories. The slowly diffusing fraction exhibits constrained diffusion over time and so appears to be stably bound to chromatin, whereas the fast fraction (**Figure 13D**) exhibits Brownian diffusion. Overall, the results suggest that a fraction of CHD4 associates with chromatin in either the presence or absence of MBD3, whereas the freely diffusing fraction of CHD4 molecules is significantly affected by the loss of MBD3 consistent with a lack of interaction with the NuRD complex. This demonstrates the feasibility of studying interactions within the NuRD complex using a combination of PALM and knockout experiments.

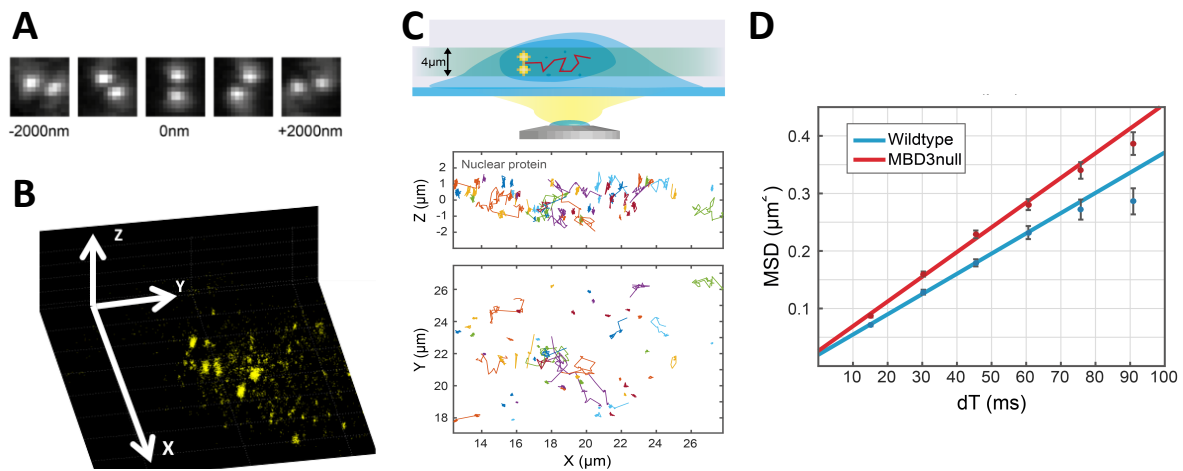
In collaboration with Prof David Klenerman (Cambridge, UK) and Anthony Carr (Sussex, UK), we have also used and published a live-cell motion blurring approach that can detect proteins bound to chromatin in yeast<sup>4</sup>. This approach uses long exposures to blur proteins that are diffusing so as to only detect chromatin bound proteins. We are now applying this to the detection of NuRD proteins in mouse ES cells to ask whether NuRD is required for stable binding of CHD4 (residence time changes) or simply for CHD4 association with chromatin (the number of non-specific binding events before a stable interaction will change) in a similar manner to which Sox2 has been shown to facilitate Oct4-DNA interactions<sup>5</sup>.



**Figure 12. 2D single molecule tracking of single mEos3-tagged CHD4 molecules in live mouse ES cells.** Representative images of the same cell are shown (left) using low power 488 nm excitation (green form of mEos3) and 405/561nm excitation (photo-activated red form of mEos3). A small number of the individual tracks from this cell are also shown indicating the fast and slow diffusing fractions of CHD4. The exact diffusion coefficients extracted from the data and the differences observed in the diffusion of a sub-population of CHD4 molecules in the presence (WT) and absence of MBD3 (MBD3-null) are shown in a box-and-whisker plot (right) for the two cell lines.

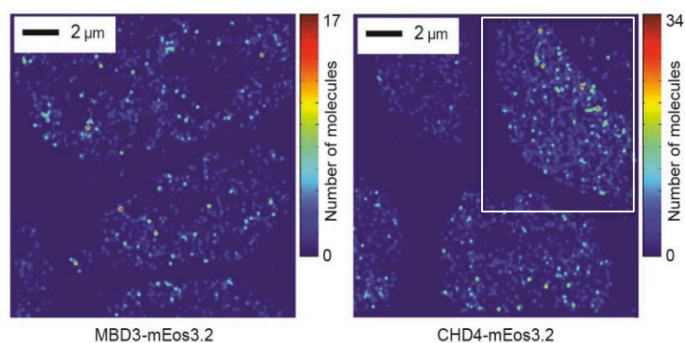
<sup>4</sup> Etheridge, T. J. *et al*. Quantification of DNA-associated proteins inside eukaryotic cells using single-molecule localization microscopy. *Nucleic Acids Res* **42**, e146, doi:10.1093/nar/gku726 (2014).

<sup>5</sup> Chen, J. *et al*. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274-1285, doi:10.1016/j.cell.2014.01.062 (2014).



**Figure 13. 3D single-molecule imaging in live mouse ES cells at 15 ms time resolution over a 4 µm range.** (A) Single protein molecules appear as 2 spots with the midpoint being its (x, y)-position and the angle indicating its z-position. (B) CENP-A molecules cluster at centromeres as expected. (C) Live-cell trajectories of single dye-labelled CHD4 molecules are shown. (D) Mean squared displacement analysis shows how deletion of MBD3 increases the speed of freely diffusing CHD4 molecules suggesting an interaction in ES cells.

100 nm density maps created from counting mEos3.2 localizations



**Figure 14. 2D super-resolution imaging shows clustering of both MBD3 and CHD4.** A 100nm density map is shown where the pixels have been colour-coded according to the number of molecules localised in that region.

**Fixed cell two-colour single-molecule co-localisation imaging approaches.** We have also developed multicolour super-resolution methods that allow us to co-localise different proteins and RNA molecules within one cell. A proof-of-principle experiment was carried out in *S. pombe* cells to demonstrate the feasibility of two-colour imaging. We labelled one protein with mEos2 and another with the Atto655 dye using a HaloTag® fusion and demonstrated co-localisation of mEos2-tagged CENP-A and Mis6-HaloTag®-Atto65. We have also carried out two-colour co-localisation using mEos3.2-tagged proteins in combination with Alexa Fluor 647-tagged antibodies. Finally, we have explored labelling proteins and RNA molecules using different activator-emitter dye combinations, in which illumination of an activator dye is used to stochastically activate an emitter dye<sup>6</sup>.

<sup>6</sup> Bates, M., Huang, B., Dempsey, G. T. & Zhuang, X. Multicolor super-resolution imaging with photo-switchable fluorescent probes. *Science* **317**, 1749-1753, doi:10.1126/science.1146598 (2007).

## **High-throughput genetic screening for components involved in PRC and NuRD-dependent transcription**

In order to identify new proteins that regulate PRC-mediated transcription, we have collaborated with Jeroen Krijgsveld's team of functional proteomics based at the EMBL of Heidelberg using a novel mass spectrometry-based approach called "SICAP", with the aim of unravelling the in vivo chromatin environment and composition of PRC2-bound CGI islands in mouse Embryonic Stem (ES) cells and differentiated cells.

SICAP method combines the x-linked Chromatin Immuno-Precipitation (ChIP) technique with protein analysis by mass-spectrometry. We combined SICAP analysis of Suz12 targets in mouse ES cells with Stable Isotope Labeling by Amino acids in Cell culture (SILAC). This allowed us to quantitatively compare SICAP analyses in Suz12 Wild-Type versus Suz12 Knock-Out mouse ES cells and in this way, to specifically enrich for PRC2-associated proteins over ChIP background. The data showed that this approach allowed us to identify all PRC2 core-components and known PRC2-associated factors among the most enriched and abundant proteins. In addition, this experiment allowed us to identify a recently identified PRC2-associated factor (Gm340), which the function remains unknown, and another uncharacterized protein, NP45104.

Currently we are currently developing molecular tools to investigate the putative role of these 2 factors in PcG function in mouse ES cells. Indeed, antibodies and CRISPR-mediated knockout cell lines are under development for both Gm340 and NP45104. This will allow us to investigate if they have a role in regulating PRC recruitment and catalytic activity.

### **3.2. SUBPROJECT 2: Polycomb repressive complex & NuRD Structural Biology**

#### **3.2.1 Structures of NuRD components and complexes**

##### **Structure of CHD4 using SAXS and chemical cross-linking/MS**

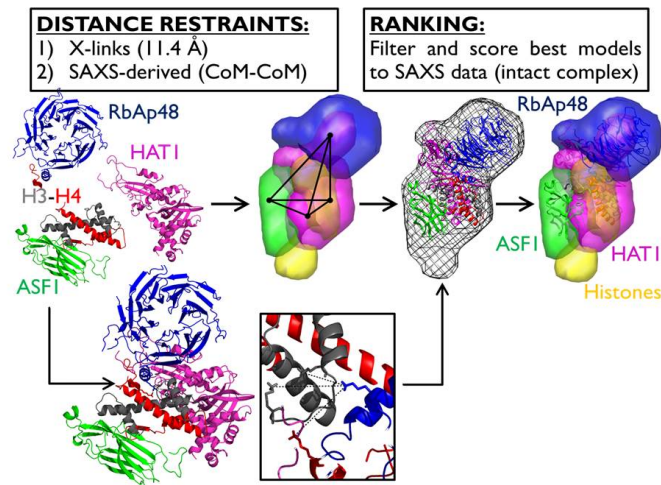
We expressed a large number of constructs of the CHD4 component of NuRD (containing different combinations of domains) in insect cells, and purified these proteins. We used SAXS, nucleosome binding, ATPase and remodelling assays, limited proteolysis, and cross-linking/MS, to generate a three-dimensional structural model describing the overall shape and domain interactions of CHD4<sup>7</sup>.

##### **Structure of RbAp48/Histone H3-H4 complex using SAXS and chemical cross-linking/MS**

SAXS and chemical cross-linking/MS were also employed to generate a three-dimensional structural model of RbAp48 in complex with Histones H3-H4 (**Figure 15**). The identified cross-links were used as distance restraints for rigid body docking of the 3D coordinates of the individual subunits. Each potential model of the complex was ranked based on its fit to the experimentally derived SAXS profiles of the complexes and components. We refined the method by calculating the centres of mass of the SAXS envelopes generated for each

<sup>7</sup> Watson, A. A. *et al.* The PHD and chromo domains regulate the ATPase activity of the human chromatin remodeler CHD4. *Journal of molecular biology* **422**, 3-17, doi:10.1016/j.jmb.2012.04.031 (2012).

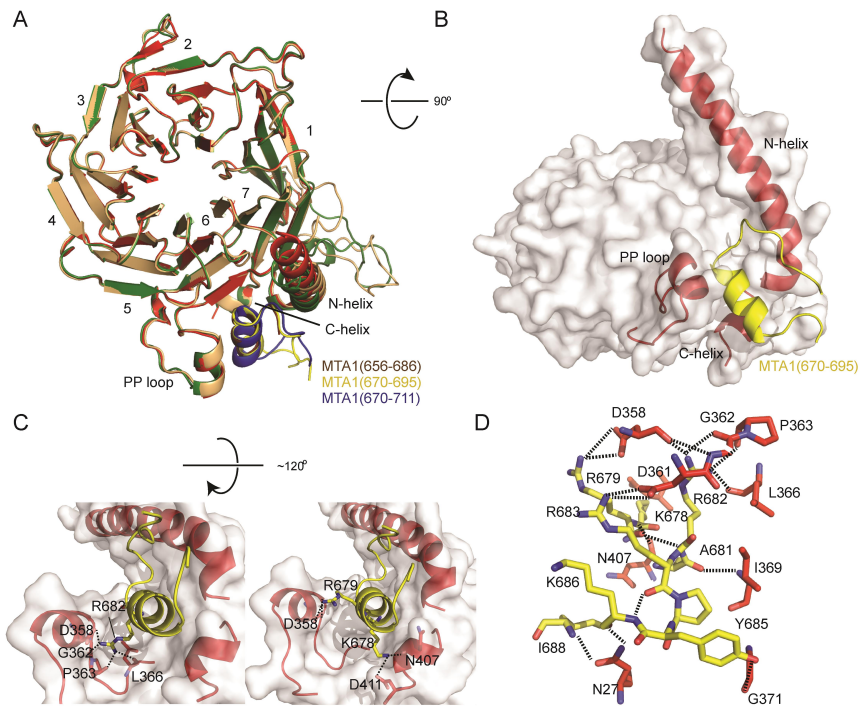
component and generating additional distance restraints between these points for use in rigid body docking. We carried out non-denaturing MS studies with the Sobott group, and FRET experiments to validate our model of the RbAp48+H3+H4+ASF1+HAT1 complex.



**Figure 15:** Structural model of the RbAp48 (blue) / HAT1 (magenta) / histones H3-H4 (grey, red) / ASF1 (green) complex determined using a combination of SAXS and chemical cross-linking/MS.

### Structural studies of the RbAp48/MTA-1 complex

In collaboration with Joel Mackay (Univ. of Sydney) we identified a C-terminal motif in MTA1 (resides 656–686), which is sufficient to bind full length RbAp48, and we subsequently solved the three-dimensional structures of three different complexes of RbAp48 bound to this region of MTA1 (Figure 2). Competition experiments using immobilized GST-H4(1-48) to pull down RbAp48 in the absence and presence of full-length MTA1 confirmed that MTA1 and H4 recognize the same site on RbAp48. These data support the conclusion that histone H4 and MTA1 contact overlapping surfaces on RbAp48 and indicate that their assembly into the NuRD complex modulates RbAp46/48 interactions with histones (Alqarni et al., 2014). Subsequently we showed that MTA1 contains a second similar RbAp46/48 binding site near the centre of the protein, supporting the finding from proteomics experiments that each MTA1 subunit binds two RbAp46/48 molecules (**Figure 16**).



**Figure 16:** X-ray crystal structures of RbAp48-MTA1 complexes. A. Cartoon superposition of RbAp48-MTA1(656–686), RbAp48-MTA1(670–695) and RbAp48-MTA1(670–711) over all C $\alpha$  atoms. RbAp48 is shown in green, red and gold and the MTA1 peptides are shown in brown, yellow and blue, respectively. The blades of RbAp48 are numbered. B. Surface/cartoon representation of RbAp48-MTA1(670–795), indicating the elements of the RbAp48 structure (red) that are contacted by MTA1 (yellow). C. Detail of the interactions of the RbAp48-MTA1(670–795) structure. D. Intermolecular hydrogen bonding network (dotted lines) formed between RbAp48 and MTA1.

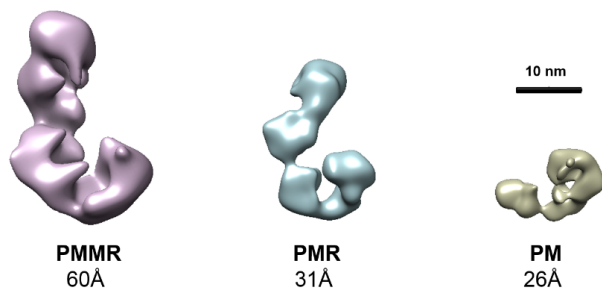
### Models of intact NuRD complexes

We have pursued two strategies to obtain structures of intact NuRD complexes. First, we analysed endogenous NuRD complexes affinity purified via GFP-tagged MBP-subunits from human and *Drosophila* cells. Secondly, by systematically producing recombinant combinations of the different subunits we identified different NuRD sub-complexes and we were able to reconstitute a close to complete *Drosophila* NuRD complex. The assembly pathway established here for intact NuRD complexes is likely to be conserved in humans and can thus be applied to generate homogenous human NuRD complexes with a defined subunit composition.

The endogenous *Drosophila* NuRD complex comprised the subunits p55, MTA-like, MBD-like and Rpd3 (PMMR), and contained only sub-stoichiometric amounts of Mi-2, Simjang and CG18292. Thus, PMMR is the core *Drosophila* NuRD complex. Proteomics analysis carried out in collaboration with the Vermeulen group suggested that the stoichiometry of the core NuRD (PMMR) complex is 4:2:1:2 with a single MBD-like subunit, which we confirmed using Multi-Angle Light Scattering of the recombinant complex. When we produced the recombinant NuRD complexes in insect cells, we found that stable sub-complexes comprising p55 and MTA-like (PM), MTA-like and Rpd3 (MR), as well as PMR exist. MBD-like appears to be added last to produce PMMR, in a process that likely mirrors the *in-vivo* assembly pathway. We were able to show that PMR, PMMR and endogenous NuRD all exist

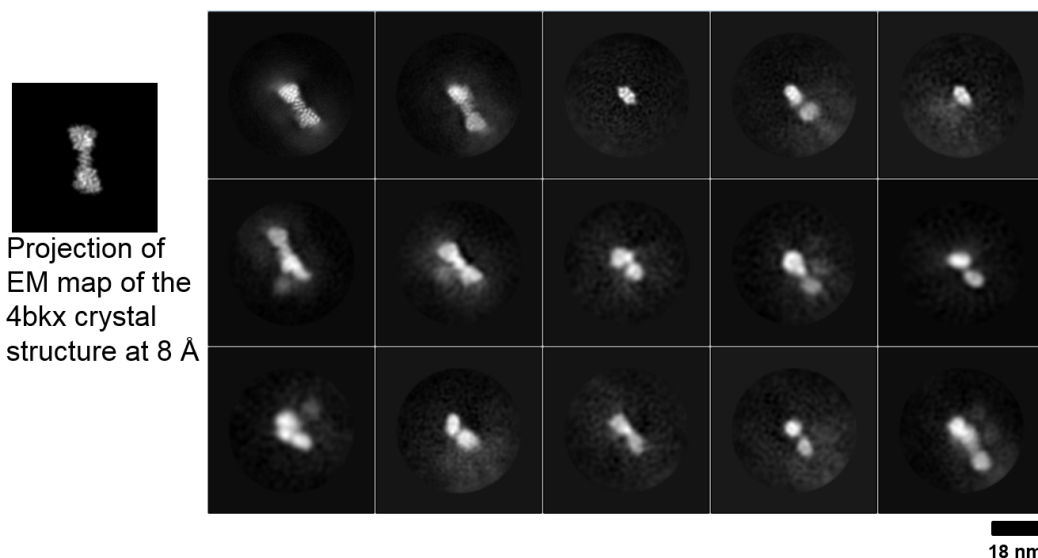
in both the cytoplasm and nucleus and have very similar deacetylase activities and specificity<sup>8</sup>. This has important implications for understanding NuRD complex function.

We calculated reference-free initial models for all three sub-complexes, which we subsequently refined (**Figure 17**). The PMMR and PMR 3D reconstructions resemble the endogenous *Drosophila* NuRD core volume in size and overall shape. PMR shows a structure that is clearly divided in two halves. When the two halves are superimposed, they show a correlation of 95% (as determined by the software Chimera). This is consistent with PMR being composed of two copies each of MTA-like and Rpd3, together with four copies of p55. The PM reconstruction can be fitted in either half of the PMR reconstruction with a correlation coefficient of 90%. The remaining density is attributed to Rpd3.



**Figure 17:** Refined volumes of PMMR (left), PMR (centre) and PM (right) at 60, 31 and 26 Å resolution, respectively. PMMR and PMR show similar size and shape, while the PM reconstruction shows a smaller

The recombinant PMMR was further analyzed by cryo-EM. Initial analysis of the dataset revealed 2D class averages showing high-resolution MTA features/secondary structure elements that clearly relate to the crystal structure of the MTA-HDAC/Rpd3 dimer (**Figure 18**) solved by the Schwabe group. However, many other 2D class averages showed only low-resolution features indicating that the dataset is still heterogeneous. We are currently attempting to overcome the heterogeneity of the NuRD complexes by biochemical stabilization strategies.



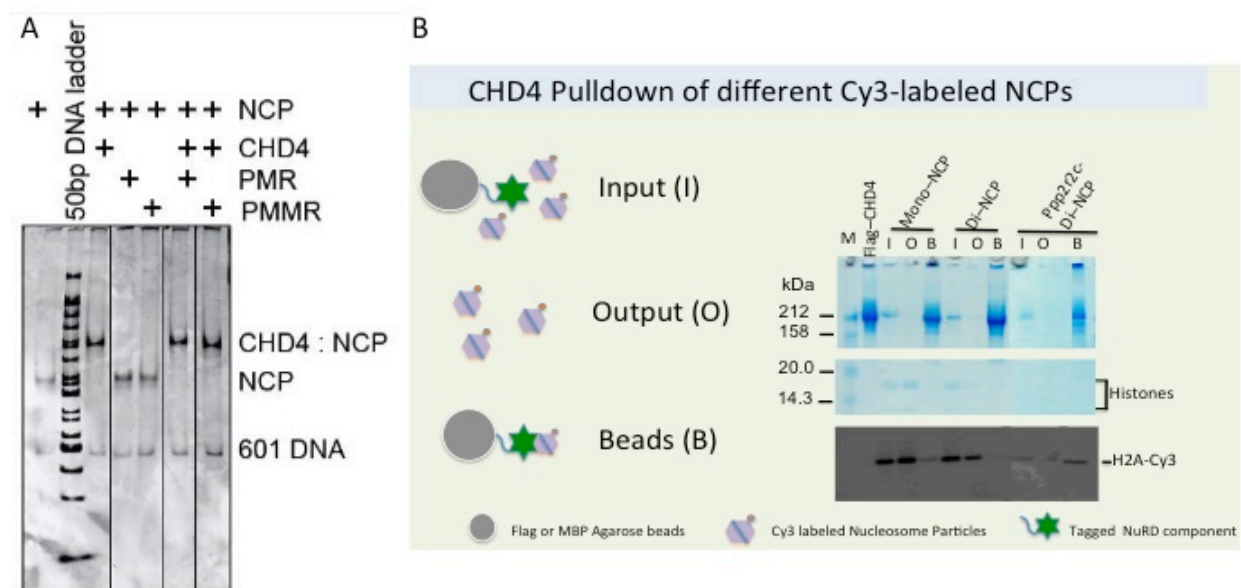
**Figure 18:** 2D cryo-EM class averages of the recombinant PMMR complex. Several 2D class averages show secondary structure elements corresponding to the MTA-RPD3 subunits, as evidenced by comparison to a projection of the MTA1-HDAC1 crystal structure filtered to 8 Å (left).

<sup>8</sup> Zhang, W. *et al.* The Nucleosome Remodeling and Deacetylase Complex NuRD Is Built from Preformed Catalytically Active Sub-modules. *Journal of molecular biology* **428**, 2931-2942, doi:10.1016/j.jmb.2016.04.025 (2016).

The other subunits of the complex – CHD4/Mi-2, GATA/Simjang and DOC1/CG-18292 – appear to loosely interact with the core PMMR complex. A sub-complex comprising CHD4 and GATA can be formed, defining two distinct functional entities – CG and core PMMR. We managed to reconstitute a close to complete NuRD complex by first binding GATA and then CHD4 to the PMMR sub-complex. These complexes are currently being analysed in detail by negative stain EM.

### Studies of NuRD complex-Nucleosome interactions

We have reconstituted mono- and di-nucleosomes with 601 DNA and DNAs from endogenous NuRD binding sites identified in CHIP-Seq experiments by the Hendrich group. We performed electrophoretic mobility shift assays (EMSA) to verify the interaction of NuRD with nucleosomes. Mono-nucleosomes interact with CHD4 as the bands shifted to a slower migration, but do not interact with PMR and PMMR (**Figure 19A**). We further investigated the binding preferences of CHD4 with different nucleosomes using pull-down assays (**Figure 19B**). We repeatedly see that CHD4 binds better to di-nucleosomes with a larger gap, such as the *Ppp2r2c* enhancer sequence, than it does to conventional nucleosomes reconstituted on to a Widom 601 DNA sequence.



**Figure 19:** Interactions of NuRD components with nucleosome particles. **A:** Mono-nucleosomes were incubated with CHD4, PMR or PMMR at room temperature for 30 min, then analysed by native PAGE using 0.2×TBE and 5% polyacrylamide gels. Nucleosomes were visualised by SYBR Gold staining. **B:** CHD4 was immobilised on  $\alpha$ -FLAG resin and incubated with mono-nucleosomes (Mono-NCP), di-nucleosomes (Di-NCP) and di-nucleosomes reconstituted on the *Ppp2r2c* enhancer sequence (*Ppp2r2c*-Di-NCP).

Currently we are purifying di-nucleosomes from cultured cells to perform an in vitro pull-down assay with assembled holo-NuRD complex, with the aim of identifying which specific genes NuRD binds to with the highest affinity. We then plan to reconstitute recombinant nucleosomes and make NuRD-Nucleosome complexes for both cryo-EM structural studies and single-molecule FRET assays to study remodelling function.

### **Screening NuRD complexes for small molecule inhibitors**

The objective of this part of the study was to assess the feasibility of identifying chemical lead molecules able to bind and perturb the function of isolated intact NuRD complexes.

We focussed on screening the core PMMR NuRD complex (which contains the histone deacetylase) and components of the NuRD complex that our structural studies had suggested might provide suitable druggable binding sites, in particular CHD4 (the chromatin remodeller) and RbAp48 (a histone chaperone). Two different CHD4 constructs (comprising the his-PP-CC and his-PP-CC-AH-D domains), and three different variants of RbAp48 (wild-type, histone H3-binding site mutant, and histone H4-binding site mutant) were produced. The Chung group at GSK using their Encoded Library Tag (ELT) technology and libraries to screen all these constructs for lead molecules. We have identified hit molecules that bind in repeated experiments, and our aim is now to further study selected hits using either fluorescence polarisation or HTRF/TR-FRET assays to confirm binding.

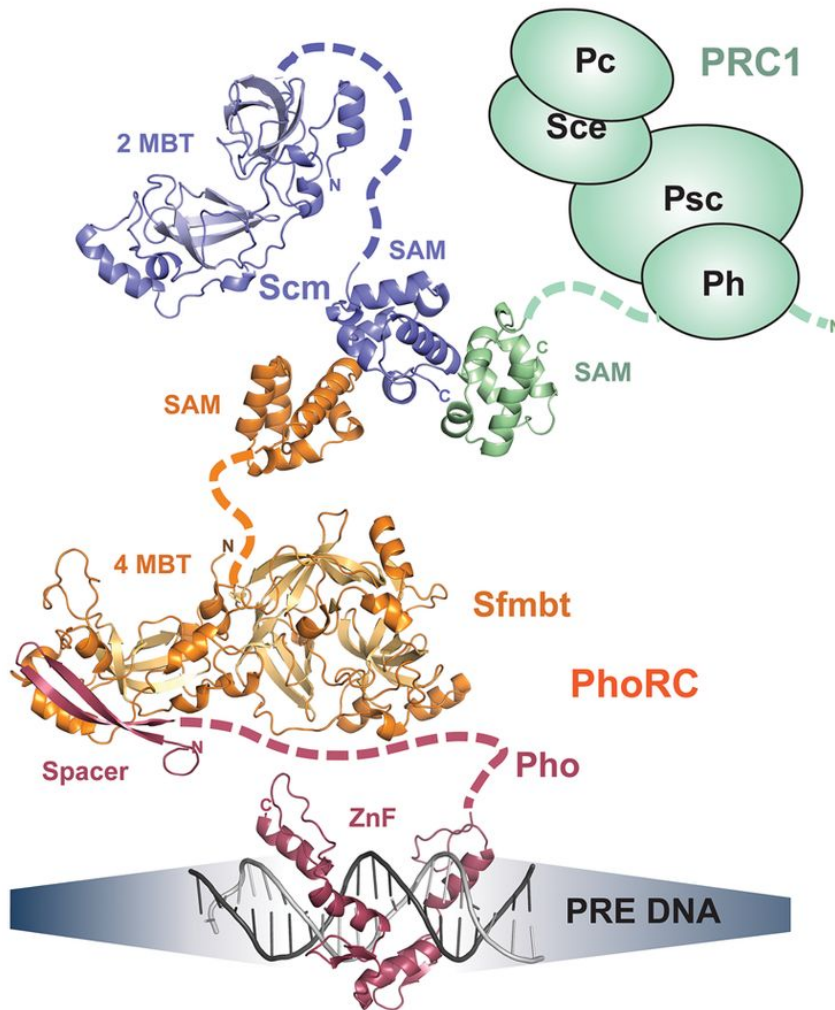
For CHD4 we have found that the affinity of the PHD domains for a N-terminal histone H3 peptide (1-20) increases in the presence of ATP (Watson et al., 2012). One possible explanation is that upon the binding of ATP to the ATP-binding site in the distal AH domain, conformational changes release the tandem PHD and chromo-domains, resulting in a higher binding affinity for H3. We therefore intend to screen the his-PP-CC-AH-D construct in the presence of a non-hydrolysable nucleotide (AMP-PNP) and/or the histone H3 binding peptide to favour identification of lead molecules binding to different conformational states of the protein.

Likewise, the RbAp48 histone chaperone has been crystallized both alone and with different binding peptides (FOG1, histones H3 and H4, MTA1) that result in conformational changes to the proteins overall structure. We noticed conformational changes in RbAp48 in which a flexible 25 residue loop in proximity of the N-terminal alpha helix becomes structured on binding histone H3. This suggests that ELT screening of RbAp48 in complex with the H3 binding peptide might in the future help to identify novel molecules.

### **3.2.2 Structures of Polycomb group complex components and Polycomb group complexes**

#### **Model of the *Drosophila* Polycomb Repressing Complex (PRC)**

The most complete structural description has been achieved for the *Drosophila melanogaster* **PRC** complex, where joint efforts of the Jürg Müller (MPI Munich) and the Christoph Müller (EMBL Heidelberg) groups have led to a model how PRC serves as bridging complex between *Drosophila* Polycomb Response Elements (PREs) and PRC1 (**Figure 20**).



**Figure 20 (adapted from Frey et al.<sup>9</sup>):** Molecular model of interactions with which PhoRC tethers canonical PRC1 to PREs. The Pho spacer region (purple) binds to the 4MBT domain of Sfmmt (orange) to form PhoRC (PDB ID 4C5I)<sup>10</sup>. The Sfmmt SAM domain binds to the ML surface of the Scm SAM domain (PDB ID 5J8Y)<sup>9</sup>. The EH surface of Scm SAM in turn binds to the ML surface of the Ph SAM domain (PDB ID 1PK1)<sup>11</sup>. Other earlier determined structure such as the ZnF domain of the Pho orthologue YY1 in complex with DNA (PDB ID 1ubd), 2MBT domain of Scm (PDB ID 2R57) are also shown.

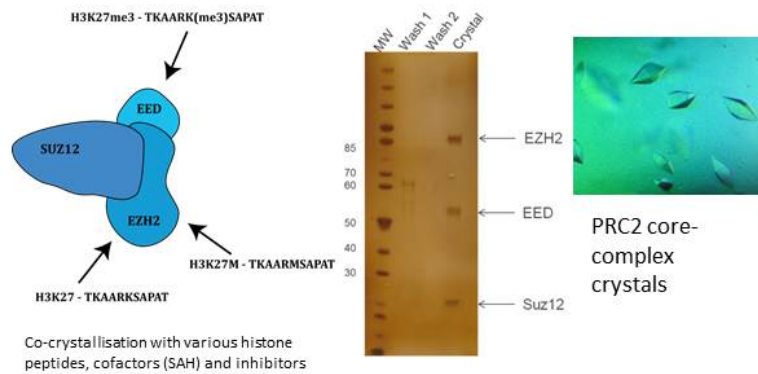
### Model of the PRC2 complex

For PRC2, efforts of the Müller group have been focusing on the co-expression in insect cells and crystallization of a ternary human subcomplex comprising EZHZ, the C-terminal moiety of SUZ12 and EED with different regulatory peptides. Crystals of this ternary complex were obtained that currently diffracted to about 4.0 Å (**Figure 21**).

<sup>9</sup> Frey, F. *et al.* Molecular basis of PRC1 targeting to Polycomb response elements by PhoRC. *Genes & development* **30**, 1116-1127, doi:10.1101/gad.279141.116 (2016).

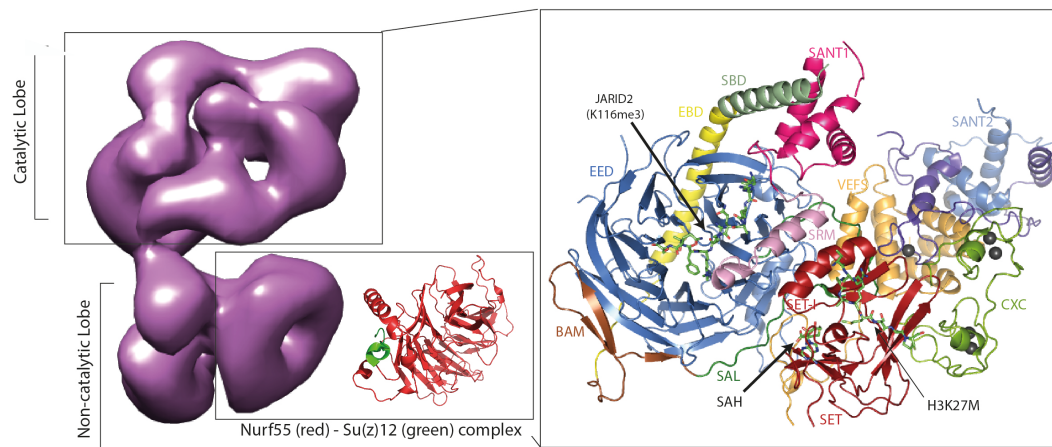
<sup>10</sup> Alfieri, C. *et al.* Structural basis for targeting the chromatin repressor Sfmmt to Polycomb response elements. *Genes & development* **27**, 2367-2379, doi:DOI 10.1101/gad.226621.113 (2013).

<sup>11</sup> Kim, C. A., Sawaya, M. R., Cascio, D., Kim, W. & Bowie, J. U. Structural organization of a Sex-comb-on-midleg/polyhomeotic copolymer. *The Journal of biological chemistry* **280**, 27769-27775, doi:10.1074/jbc.M503055200 (2005).



**Figure 21:** Schematic representation, co-expression and crystallization of a ternary PRC2 subcomplex.

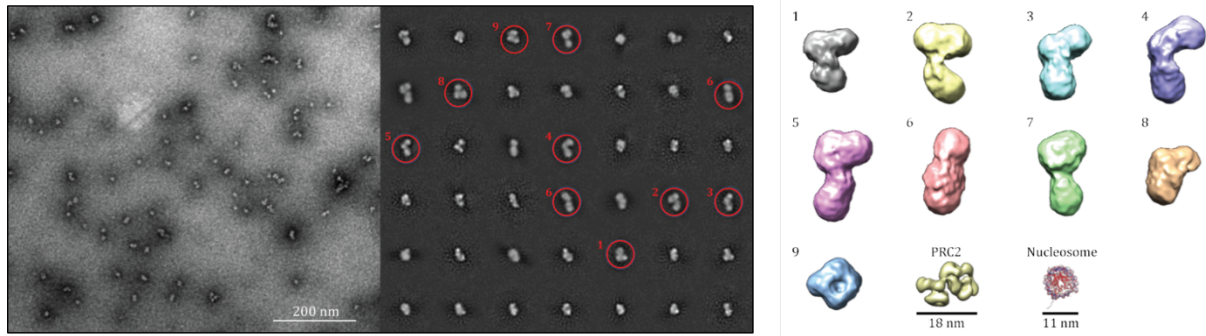
In spring 2016 two competing groups published crystal structures of a ternary PRC2 complex containing very similar constructs of EZHZ, the C-terminal moiety of SUZ12 and EED as identified by our group. The crystal structure of the ternary PRC2 complex is depicted in Figure 22. The ternary structure can be also fitted into a negative-stain electron microscopy reconstruction of a pentameric PRC2 complex that comprises EZHZ, SUZ12, EED, RbAp48 and AEBP2. Pentameric PRC2 complexes were analyzed by native MS (Frank Sobott). Both, monomeric (303,351.4 Da) and dimeric PRC2 are detected side by side. Collisional activation inside the mass spectrometer is used to confirm this assignment, with RbAp48 dissociating first indicating that it binds least tight with the other subunits.



**Figure 22:** Negative-stain EM structure of a pentameric PRC2 complex by Ciferri et al. eLife 2014. The crystal structure of the *Drosophila* Nurf55-Su(z)12 complex corresponding to human RbAp48-SUZ12 is shown in the inset (left panel). Crystal structure of the ternary PRC2 subcomplex (Justin et al. & Brooun et al. Nat Commun 2016) (right panel).

In parallel, the Müller group is also reconstituting PRC2 complexes bound to modified nucleosomes. Native MS in the Sobott group confirms that mono- and dimeric PRC2 pentameric complexes can bind to intact nucleosomes (1:1 complex: ca. 828 kDa). To preserve the integrity of PRC2-monomucleosome complex and increase its stability, the complex was cross-linked using the GraFix protocol. Subsequently, a single particle tomography dataset was collected on a Polara microscope and 2D classifications was performed (**Figure 23**, left). The selected 2D class averages were used for 3D model classification. The generated models appeared to have larger dimensions than PRC2 or nucleosome alone (Figure 23, right), indicative of PRC2-monomucleosome complex

formation. These models form the basis for the current efforts of the Müller group to obtain a high-resolution structure of the PRC2-monomucleosome complex using cryo-EM.



**Figure 23:** Negative stain EM micrographs and 2D class averages of the recombinant human PRC2-AEBP2 bound to K27M-nucleosomes (left) and its 3D models (right) generated from elected final 2D classifications (highlighted in red circles). The 3D models are shown in comparison to the published negative stain PRC2-AEBP2 model (EMD-2236) and the crystal structure of the nucleosome core particle (PDB: 1AOI). All models are drawn to scale.

In addition, the Jürg Müller group (MPI Munich) determined the crystal structure of a fragment of the PRC2 associated factor Polycomblike (Pcl) from *Drosophila* (**Figure 24**). Pcl, and its mammalian orthologues Pcl1/PCL1, Pcl2/PCL2 and Pcl3/PCL3 associate with PRC2 and are required for PRC2 anchoring at Polycomb target genes in the genome. The Jürg Müller lab showed that this part of the protein binds to DNA in a sequence non-specific manner with mid-micromolar affinity. Structure-guided mutations were then used to investigate how point mutations of residues that are critical for DNA binding by Pcl affect the nucleosome-binding activity and histone methyltransferase activity of Pcl-PRC2 assemblies *in vitro*. These studies established that the DNA-binding activity by Pcl/PHF1 protein plays a critical role for efficient H3K27 tri-methylation.

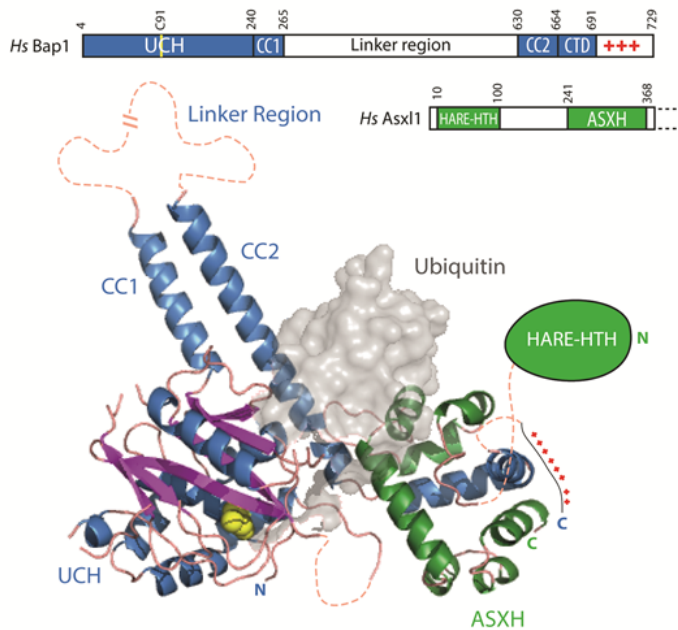


**Figure 24:** Domain architecture of a fragment of a Pcl orthologue. Resolution: 2.3Å, R/R<sub>free</sub>: 0.23/0.268.

### Model of the PR-DUB complex

The Müller group has continued to attempt to obtain a crystal structure of PR-DUB. Construct design was informed by biochemical characterisation of human PR-DUB constructs and a 3D structural model of human PR-DUB bound to ubiquitin based on the structure of Bap1-related protein UCH-L5 in complex with activating protein Rpn13 and with ubiquitin bound (**Figure 25**).

PR-DUB/nucleosome complexes have been visualized by negative stain EM. However, problems in processing the data arose due to the presence of nucleosomes without PR-DUB bound in the sample. An improved protocol has been developed for the purification of PR-DUB in complex with ubiquitinated nucleosomes to produce a homogenous solution of two PR-DUB molecules bound to a ubiquitinated nucleosome (**Figure 25**).



**Figure 25:** Domain architecture of human PR-DUB proteins Bap1 and Asx1 and a 3D structural model of the human PR-DUB complex with ubiquitin bound modelled on UCH-L5/Rpn13 complex with ubiquitin bound (PRD ID: 4WLR<sup>12</sup>).

### Native MS and ion mobility of intact nucleosomes

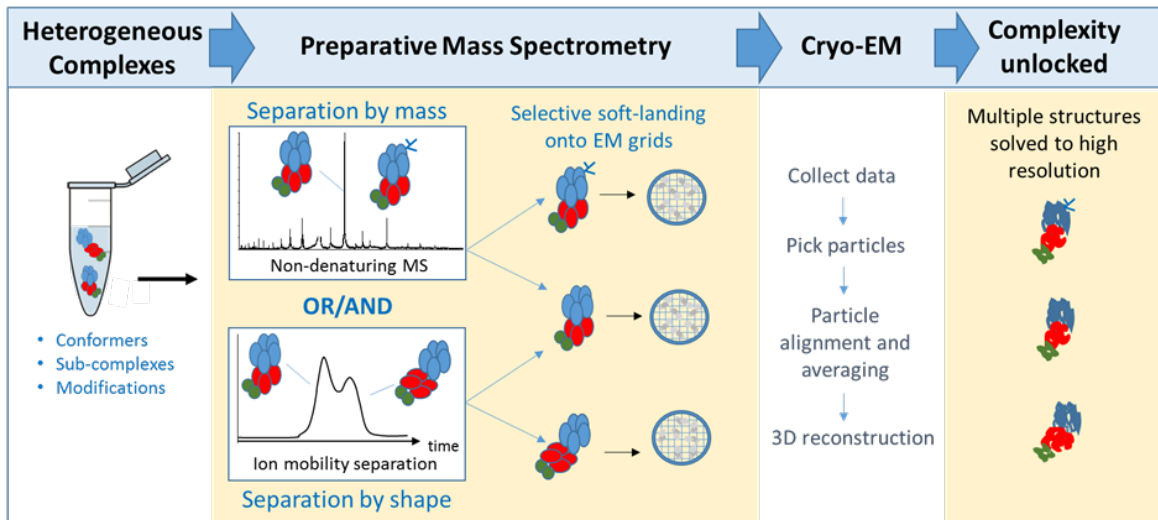
Native MS was used to assess homogeneity of reconstituted nucleosomes (from the Bartke and Christoph Müller groups), and to characterize their folding state using ion mobility. A typical mass spectrum of nucleosomes (222 kDa; measured at 10  $\mu$ M protein in 100 mM aqueous ammonium acetate buffer) shows complete assembly with a small excess of free 113 kDa DNA visible. The approach is suited to monitor assembly and partly sequence modified nucleosomes (with defined PTM state or PTM mimetics), and to provide a platform for rapid screening of possible interaction partners ( $K_d$  range low  $\mu$ M-nM).

### Preparative mass spectrometry

Native MS is able to separate lowly-populated ion species (e.g. partly assembled or heterogeneous complexes) in a narrow  $m/z$  window (**Figure 26**). When MS is combined with ion mobility separation, it also provides separation based on shape, resolving conformational or topological heterogeneity. Separated species can then be deposited onto a surface by preparative soft-landing, without affecting their composition. The Sobott group have coupled soft-landing MS with negative-stain EM, confirming the structural integrity of protein

<sup>12</sup> VanderLinden, R. T. *et al.* Structural basis for the activation and inhibition of the UCH37 deubiquitylase. *Molecular cell* **57**, 901-911, doi:10.1016/j.molcel.2015.01.016 (2015).

assemblies after deposition onto a target. This indicates the potential of the approach for high-resolution structure determination of heterogeneous assemblies. It has now also been shown that shape/size separation is possible using a fast-switching timed gate immediately after the ion mobility cell, by modifying a commercial instrument.



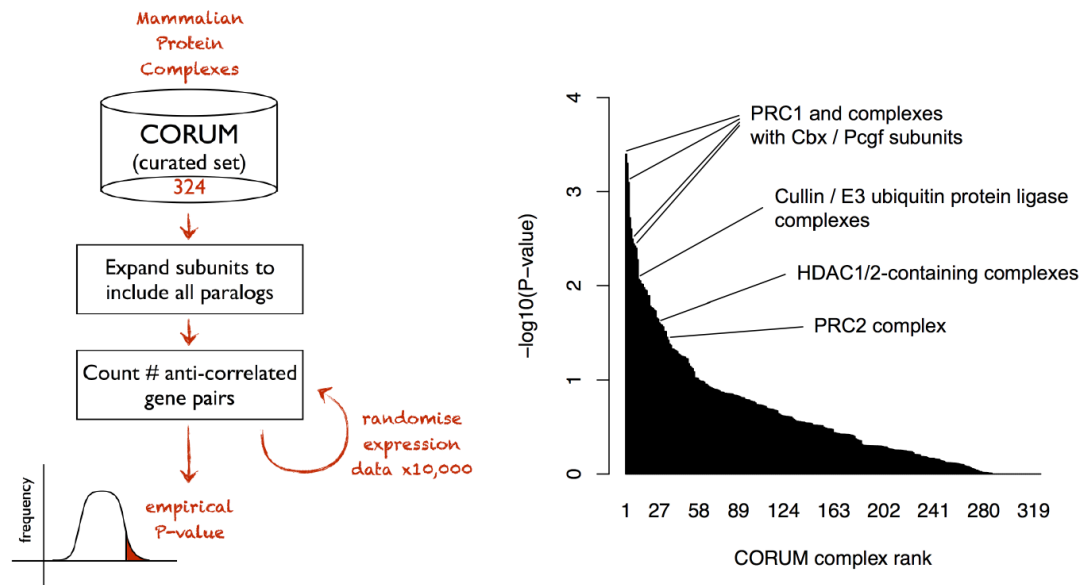
**Figure 26:** Coupling preparative MS with down-stream EM for high resolution structure determination.

As little as 1.5-2% global size difference (by collision cross section) can be detected. We tested, amongst others, shape/size selection of highly dynamic SMC dimers (Structural Maintenance of Chromosomes protein, *b. subtilis*). Corresponding EM and AFM data were collected in collaboration (Grenoble/Schaffitzel group, Bristol and Madrid) to define the corresponding structural variation of the protein. Coupling preparative MS with down-stream cryo-EM analysis will offer an elegant solution to unravelling biological heterogeneity complexity.

### 3.3. SUBPROJECT 3: Polycomb repressive complex & NuRD Data Integration

#### 3.3.1 Integrated overview of PRC and NuRD complex function during cellular differentiation.

Using a compendium of genome-wide expression data from differentiating mouse embryonic stem (ES) cells under various experimental perturbations (from P1a, Di Croce lab), we expanded this goal to encompass a global view of protein complex diversity and temporal dynamics. Results from our approach, which uses a staging technique based on principle component analysis (PCA) to efficiently merge expression data from heterogeneous genetic backgrounds, (i) confirmed the peculiar dynamics of PRC1 subunits during development and (ii) yielded a list of candidate complexes with unusually high levels of subunit switching for potential follow-up investigation (**Figure 27**).



**Figure 27.** Strategy to determine complexes with unusually high numbers of putative subunit switch events (left) and randomisation test results (right). Ranked nominal P values are shown although only significant complexes are indicated (FDR <0.25).

To identify and understand the influences on expression variation in mouse embryonic stem (mES) cells we integrated diverse sequence and functional genomic datasets. In our analysis we combined previously published RNA-seq data from single mouse embryonic stem cells (ESCs)<sup>13,14,15,16</sup> with definitions of promoter types, chromatin states and chromatin domains from the matched cell type.

Our aim was to determine whether sets of genes with similar properties tend to have unusually high or low expression noise at the transcriptional level. This revealed that the core promoter architecture of a gene is an important influence on the stability of expression across individual cells, with CpG island promoters and broad transcription initiation associated with low noise and a TATA box and sharp initiation associated with increased noise. However, the expression variability of each gene is also strongly linked to its chromatin architecture. In general, genes with high levels of activation-associated modifications have low expression variability across cells, with H3K36me3 in particular associated with low noise. Conversely, the presence of repressive modifications on expressed genes is associated with increased noise, suggesting a model where “conflicting” chromatin architectures – the absence of active modifications or the presence of repressive ones – increase noise (**Figure 28**).

In contrast, broad domains of active enhancer chromatin – super-enhancers – actually increase variability by sensitizing expression to fluctuations in the pluripotency network. Even

<sup>13</sup> Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome biology* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010)

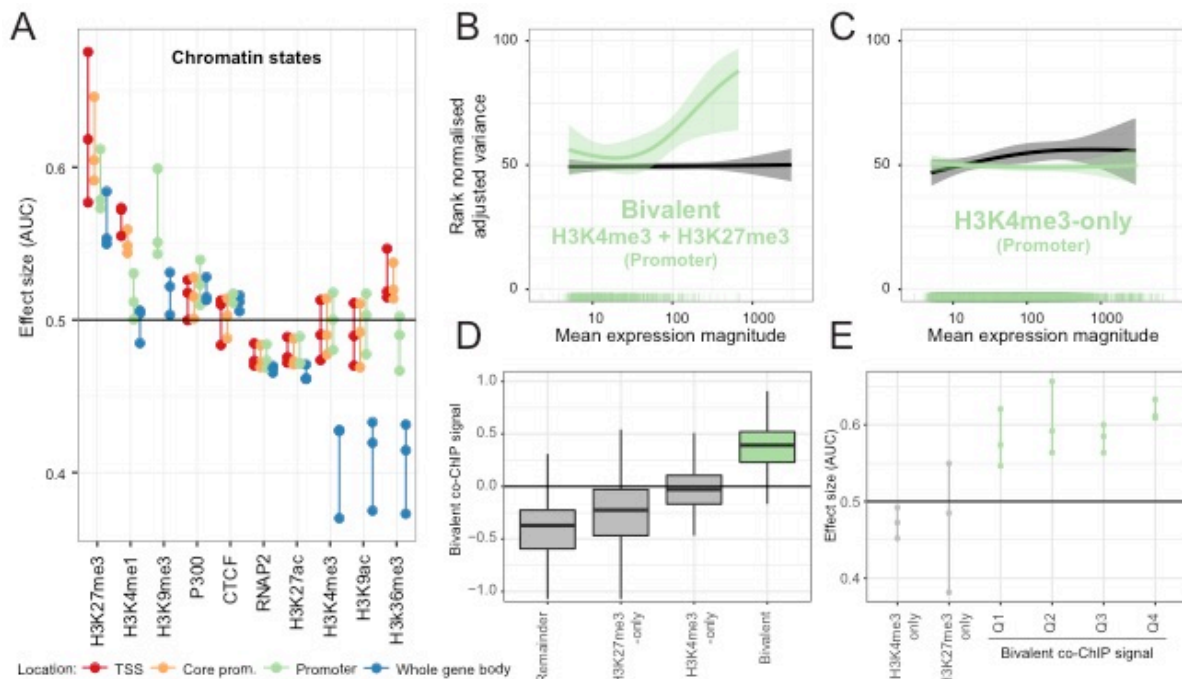
<sup>14</sup> Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature methods* **11**, 163-166, doi:10.1038/nmeth.2772 (2014)

<sup>15</sup> Grun, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nature methods* **11**, 637-640, doi:10.1038/nmeth.2930 (2014).

<sup>16</sup> Kolodziejczyk, A. A. *et al.* Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell stem cell* **17**, 471-485, doi:10.1016/j.stem.2015.09.011 (2015).

in 2i conditions fluctuations in pluripotency can be detected in the expression of a large number of genes.

Finally, we identified that unusually high mRNA stability of X-linked genes as a mechanism that compensates for the high noise associated with expression from the single copy of this chromosome. Our results provide an integrated view of how core promoters, chromatin and post-transcriptional regulation tune the stability of gene expression across individual stem cells.



**Figure 28.** Association of chromatin states with expression variation across single mouse embryonic stem cells. A. Association of individual states. B-C. Association of Polycomb targeted (H3K27me3-marked) bivalent and H3K4me3 promoters for genes with different expression levels. D-E. Evidence from co-ChIP that these noisy bivalent promoters carry both histone modifications at the single molecule level.

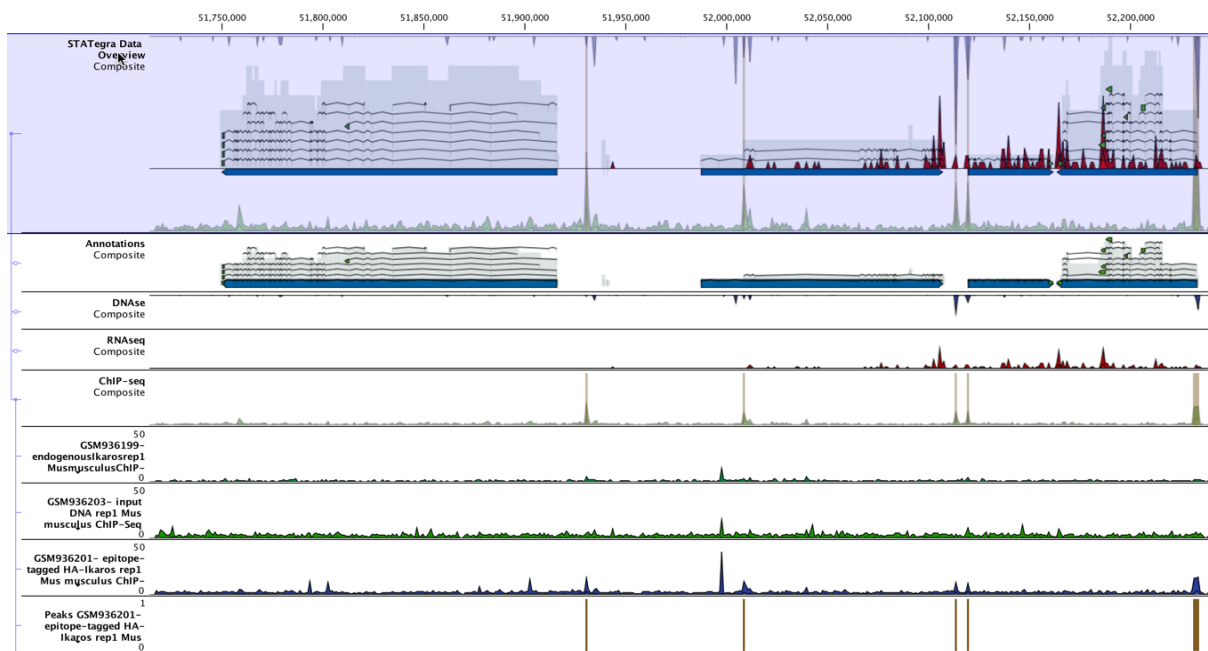
### 3.3.2 Open access to all datasets generated by the consortium and an analysis tool to facilitate the analysis and integration of these datasets by wet-lab biologists

To date, 87 publications have resulted from the 4DCellFate project and 50 of these publications are open access. In addition, all of the datasets generated by the consortium and submitted for publication have been deposited in the appropriate open access databases.

Based on the CLC bio Genomics Workbench (GxWB), we implemented extensions for integrative visual analytics of genomic track-data. The software integrates genomic datasets from both external sources and data generated within the consortium. With the new track-tools, computation and visualisation are handled in a seamless graphical framework that does not require any coding from the user.

Through discussions with consortium partners we recognised the identification of functional elements as peaks in genomic data as the central algorithmic challenge. This is addressed by a novel shape-based peak-detection algorithm that works in a robust manner while requiring only a minimum of parameterisation. The resulting combination of an interactive graphical approach with highly automated signal-detection algorithms makes the software ideally suited for data analysis by non-bioinformaticians. All results can be shared and exchanged between consortium members in the form of CLC-file objects.

A flexible back-end has been implemented for the CLC Genomics server that enables integration of systems-biology datasets and other bio-molecular networks. This way network data is made accessible for querying and interactive visualisation through the graphical user-interface (**Figure 29**).



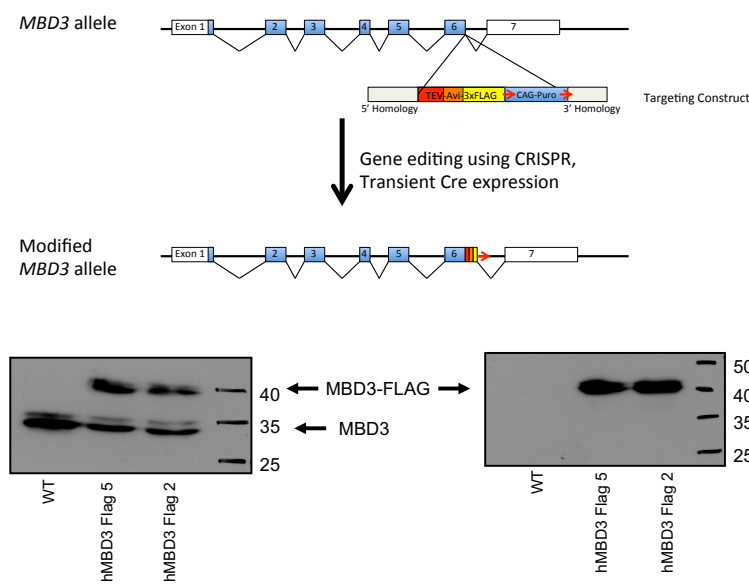
**Figure 29.** Screenshot of a composite Track Visualisation, comprising different annotation and data tracks together (blue background, top). The hierarchical organisation of the underlying tracks is indicated on the left (blue lines). Several annotation tracks (including gene annotations, exon-structure, mRNA and CDS tracks) are combined into a single composite track. Similarly, experimental data from DNase-seq and RNA-seq are combined into composite tracks themselves and collapsed. The RNA-seq data is visualized with red triangles in the overview on top of the gene annotations (blue arrows), while the DNase data (that indicates open chromatin regions) is drawn from the top down. The ChIP-seq track is expanded, showing the underlying data tracks consisting of 2 different conditions and one input track. The lowermost track shows the peak regions called from the data. All visual attributes such as color, offset, transparency, and thresholds can be freely adjusted for every track through the GUI.

### 3.4. SUBPROJECT 4: Polycomb repressive complex & NuRD In Disease

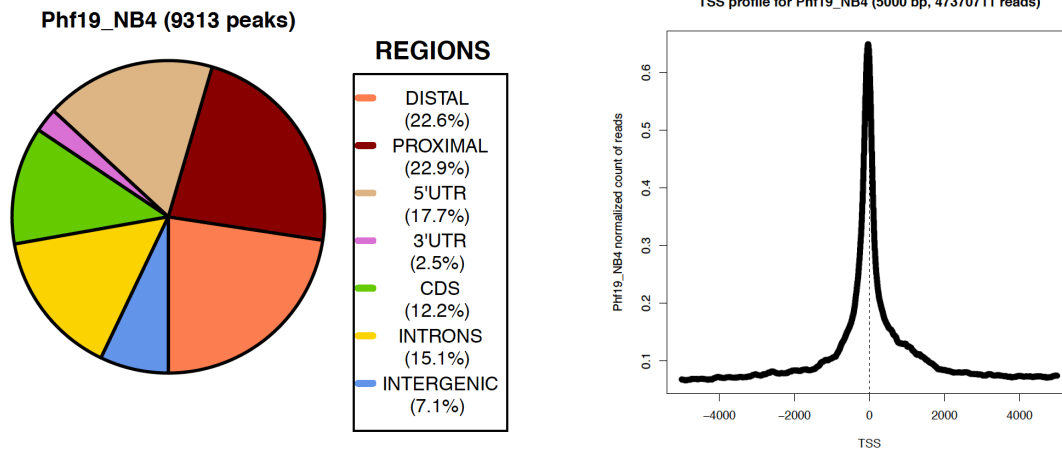
#### Characterisation of human NuRD and PRC complexes

Genome engineering was used to modify one *MBD3* allele in human iPS cells facilitate detection of the NuRD complex in human cells. Using CRISPR/Cas9, we modified one *MBD3* allele such that it now encodes for an MBD3 protein with a two-fold epitope tag at its C-terminus (**Figure 30**). After verification, this cell line was used to define components of the NuRD complex in human pluripotent cells. The epitope tag was used to purify native NuRD complexes and interacting proteins which were then subjected to quantitative mass spectrometry. This allowed us to both calculate the stoichiometry of NuRD in human cells and to identify potential interacting proteins.

PRC complexes have been well characterised in leukemic cells, so our focus was to identify key interacting proteins which can direct PRC action to specific sites. One such interactor is PHF19, identified by Partner 1a. Chromatin binding sites for PHF19 were determined in human leukemic cells (**Figure 31**), and this was compared with transcriptomics data to learn that PHF19 occupies promoters of both active and inactive genes in human leukemic cells, consistent with it playing a role in directing PRC2 activity in this cell type.



**Figure 30. Creation of an epitope-tagged *MBD3* allele in human iPS cells by CRISPR/Cas9 technology.**

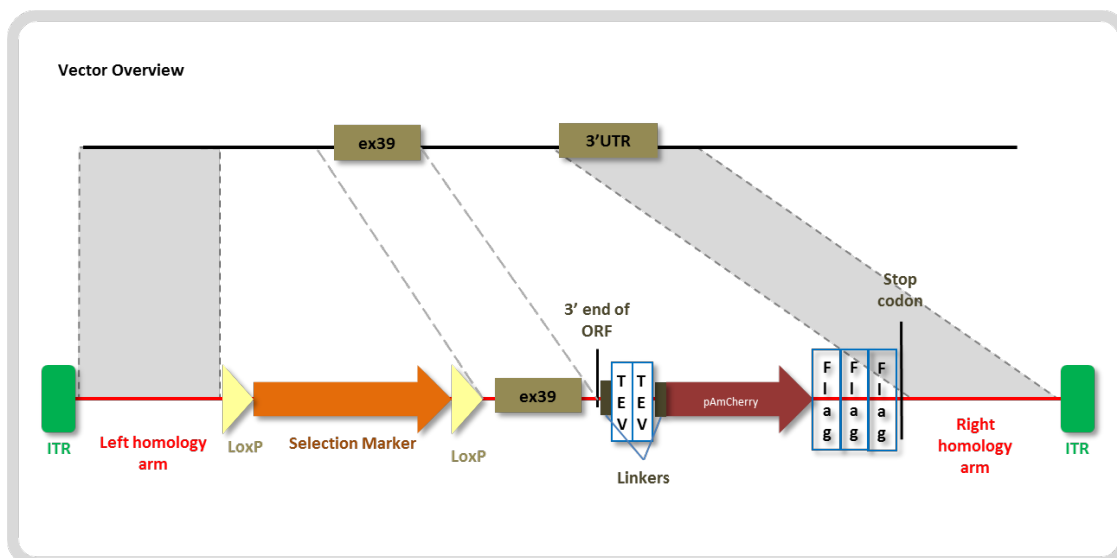


**Figure 31. Distribution of PHF19 in human leukemic cells**  
**Expression patterns of interacting proteins**

The expression patterns of two different isoforms of the PHF19 protein were determined in several human leukemic cell lines. Depletion of the long isoform triggered spontaneous cell differentiation. Subsequent RNAseq analysis in two different leukemic cell lines revealed widespread changes in gene expression, consistent with PHF19 being a subunit of the Polycomb repressor complex 2. Based on these results, we initiated a screening for the identification of small chemical compounds to inhibit PHF19. This screen is ongoing.

### Human reporter lines.

In addition to the MBD3-3xFLAG human iPS cell line described above, Partner 9 also created a tagged CHD4 allele in the same iPS cell parent line. This strategy combined a FLAG tag with fluorescent proteins to enable its use in microscopy experiments (**Figure. 32**). This line was constructed and delivered to Partner 2b, where it was independently verified. This resource will be of use in downstream applications.



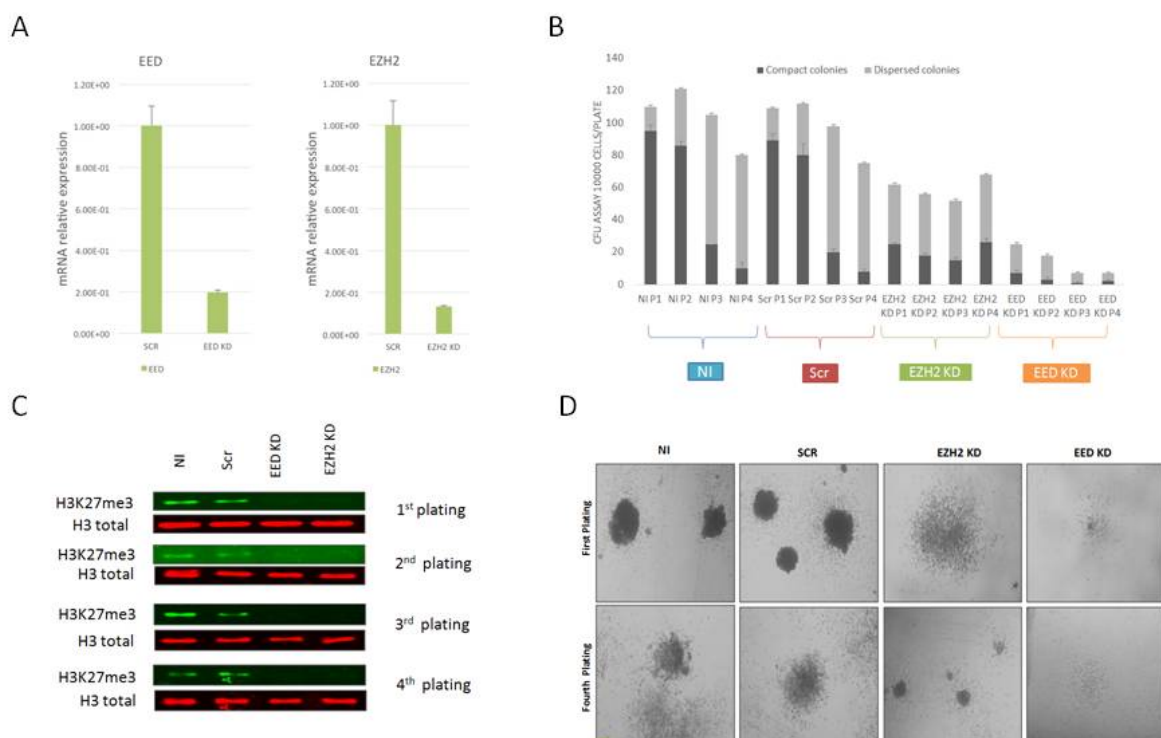
**Figure 32. Creation of a CHD4 fluorescent reporter human iPS cell line.**

### Small molecules to interfere with PRC function

We investigated the action of two different inhibitors of the PRC2 complex: GSK343 (selective for EZH2) and UNC1999 (targeting both EZH1 and EZH2. The inhibitors have been tested on human APL cell lines (NB4 and UF1, the latter resistant to RA induced differentiation), murine APL leukaemia, and Lin<sup>-</sup> cells from wild-type and PML-RAR knock-in mice (pre-leukemic phase of APL).

At the beginning of the treatment both molecules showed a strong inhibitory effect on cellular growth and colony formation which correlate with an induction of differentiation, further potentiated by retinoic acid co-treatment. Importantly, UNC1999 had a stronger effect compared to GSK343. Both molecules had a milder phenotype on Lin<sup>-</sup> cells derived from wild-type mice in comparison to the PML-RAR expressing cells, suggesting a possible therapeutic window.

At later time points (after several passages in methylcellulose) we observed the emergence of resistant clones in particular from cells expressing PML-RAR at all the stages of the disease (pre-leukaemia and fully established leukaemia), suggesting that continuous inhibition of EZH1/EZH2 may select for cells with the ability to sustain in vitro growth and therefore may have the potential of a leukemia initiating cell (LIC). This was verified using animal experiments. These observations may explain the rapid relapse of APL mice upon treatment discontinuation. These preclinical studies suggest that EZH2 inhibition improve survival but prolonged treatment enriches leukemia initiating cells. (**Figure 33**).

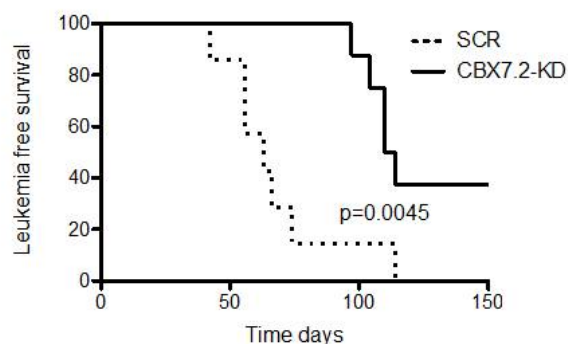


**Figure 33. EZH2 and EED knock down in murine APL cells.** Murine Pr1 cells were knocked down and analysed for: A) EZH2 and EED mRNA levels by quantitative RT-PCR (scrambled vector was used as control); B) serial re-plating ability in methylcellulose; C) H3K27Me3 levels analysis of the indicated samples; D) Morphology of the colonies from serial replating experiment showed in B.

Previous results on the role of CBX7 in AML maintenance suggested that this PRC1 component is fundamental for maintenance of leukemic cells. We therefore sought to further

validate CBX7 as a target in AML, and then to identify novel small molecules able to interfere with its function. Using two different AML models, knockdown of Cbx7 reduced both cell growth and clonogenicity, suggesting that CBX7 is indispensable for AML cells survival/proliferation. To further characterize the CBX7 domain required for the observed phenotype, we generated a CBX7 mutant in the chromodomain (CBX7<sup>K31A/W32A</sup> named CBX7<sup>AA</sup>) which is unable to bind H3. MLL-AF9 cells expressing the mutant CBX7<sup>AA</sup> showed a decreased ability to form colonies compared to cells expressing wild type CBX7. Importantly, colonies recovered 7 days after plating in methylcellulose were negative for the expression of CBX7<sup>AA</sup>, while wild type CBX7 was still expressed. These results suggest that CBX7<sup>AA</sup> acts in a dominant negative fashion and that CBX7 activity depends on its binding to histones through the chromodomain. To investigate the requirement of CBX7 in leukaemia maintenance, human primary AML cells were knocked-down for Cbx7 and transplanted into NSG mice. Leukaemia progression was significantly delayed in mice transplanted with Cbx7 knockdown leukemic cells (**Figure 34**). Importantly, in this experimental setting 3 mice out of 8 transplanted with leukemic cells knocked down for Cbx7 failed to develop leukaemia.

The results obtained so far identify CBX7 as a relevant target which, through its interaction with H3K27Me3 (mediated by its chromodomain), is required for leukaemia maintenance. In order to select molecules able to interfere with the association between Cbx7 and H3K27Me3, in collaboration with several groups (including Partner 1a) we exploited the new NanoBret technology to develop cell-based assays to identify small molecules able to interfere with the binding in cells and performed virtual screening to identify inhibitors in vitro of the interaction of the isolated Cbx7 chromodomain with a peptide derived from H3. Overall, the results of our small molecule activities have led to the identification of at least a dozen compounds (obtained through different approaches) that are at different stages of characterization in the drug discovery chain of activities (hit confirmation-hit to lead). Since there are essentially no validated small molecules able to interfere with the chromodomain so far, our studies represent a clear advancement in the demonstration of the drugability of this domain, that we are also showing to be critical for the oncogenic activity of CBX7 in AMLs.

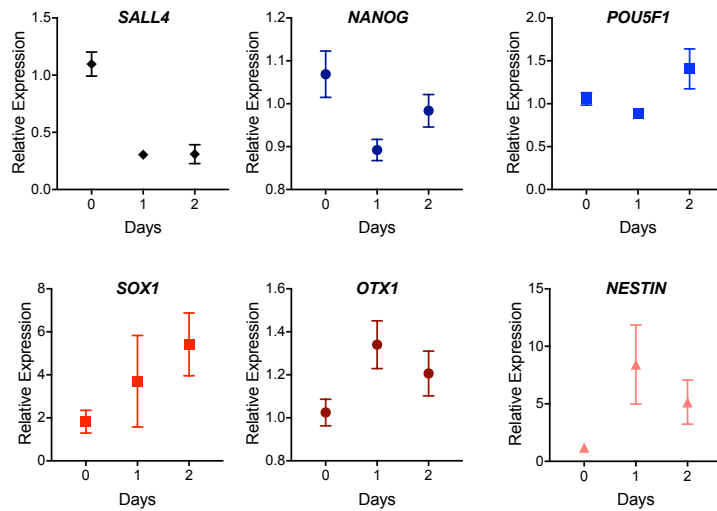


**Figure 34. CBX7 knock-down delays leukemia onset.** Kaplan-Meier curves showing leukaemia free survival following transplantation with control (SCR) or CBX7 knockdown AML cells.

### Improving differentiation protocols of human pluripotent cells

We identified SALL4 as a component of the NuRD complex in human pluripotent cells. Using our knowledge about SALL4 function in mouse cells, we designed a strategy to create an improved neural differentiation protocol in human pluripotent cells. By knocking down SALL4 transiently, we find we can prime cells to enter the neural lineage (**Figure 35**). This will

provide a more efficient neural differentiation protocol which should be amenable to the application of small molecule inhibitor screens on the way towards creating a more robust and generally applicable differentiation protocol.



**Figure 35. Expression of indicated genes upon knockdown of *SALL4* in human pluripotent cells.** Expression (Mean  $\pm$  SEM) is displayed relative to that in control cells at indicated days after *SALL4* knockdown.

## 4. Potential impact (including the socio-economic impact and the wider societal implications of the project so far) and the main dissemination activities and exploitation of results

### 4.1. Socio-economic impact and exploitation of results

#### Technical development:

##### Single molecule fluorescence microscopy

**Exploiting two-colour PALM and STORM methods to study PRC/NuRD co-localisation in mouse ES cells.** Our original aim was to develop two-colour single-molecule co-localisation imaging for the study of protein complexes. However, during the course of this project we have realised that, because many proteins co-localise to chromatin at sites of transcription, co-localisation does not necessarily help us understand the assembly and function of chromatin complexes. We have therefore found it more informative to carry out live-cell tracking of single proteins in the presence/absence of other components in order to understand the role of protein interactions within a complex. In this project, we have thus developed both fixed and live cell 2D and 3D single-molecule imaging (PALM and STORM) in mouse ES cells. This has allowed us to study either: 1) co-localisation and clustering of proteins within the nucleus; or 2) the role of protein interactions within a complex by tracking single proteins in the presence/absence of other components.

## **Development of fixed and live cell 2D and 3D single-molecule imaging (PALM and STORM)**

For 2D single-molecule imaging, UCAM (Laue) in collaboration with Prof David Klenerman (Cambridge, UK) has built an nTIRF microscope. It incorporates 4 laser lines (405, 488, 561, 640 nm) and a dual view camera for multiple fluorophore imaging in either PALM (using mEos3.2) or STORM mode (using AF647). We have imaged a range of nuclear proteins such as the centromeric histone H3 variant CENP-A, the heterochromatin proteins HP1 $\beta$  and HP1 $\gamma$ , a nuclear periphery marker (lamin B2), and NuRD components (MBD3 and CHD4). To determine co-localisation in ES cells on a 2D microscope, it is critical to ensure that localised molecules are within the detection plane (i.e. localised together) and not above or below this plane. We therefore developed and published an optical sectioning technique (Palayret *et al*, 2015). To study single ES cells, we also developed and published a novel microfluidic device capable of trapping cells for super-resolution imaging before releasing them for further processing (Zhou *et al*, 2016). To achieve 3D super-resolution imaging, the Klenerman group built a microscope with double-helix point spread function (DH-PSF) detection and adapted it for mammalian imaging (Carr *et al*, submitted). We have imaged single mEos3.2 and HaloTag®-JF549 dye-labelled CENP-A and CHD4 proteins with high x, y and z resolution over a 4  $\mu\text{m}$  range in z. We have used these microscopes to study the following:

**Clustering of NuRD proteins within the nucleus.** Taking advantage of the expertise of UCAM (Hendrich), knock-in cell lines were generated in which MBD3 and CHD4 were tagged at the C-terminus with mEos3 and the HaloTag®. Super-resolution imaging of the NuRD components MBD3 and CHD4 was demonstrated and published. CHD4 and MBD3 molecules form distinct foci in live ES cells suggesting that the NuRD complex forms locally enriched clusters. Analysis of these protein clusters has revealed that CHD4 binds to a considerably higher number of chromatin sites in the nucleus than MBD3, consistent with reports that it has significant non-NuRD functions (Stevens *et al*, in press).

**Single-particle tracking in the presence/absence of other components to understand the role of protein interactions within a complex.** If NuRD components co-localise/interact, the dynamics of specific proteins within the complex will be affected by the absence of other components. We investigated whether CHD4 dynamics is affected by the presence/absence of MBD3 using the knock-in cell lines described above and published this work (Zhang *et al*, 2016). We measured CHD4 dynamics using PALM of mEos3.2-tagged CHD4 at 10 ms time resolution in both wild-type and MBD3-null cells. CHD4 exhibits two major diffusion coefficients ( $0.100 \pm 0.008$  and  $0.65 \pm 0.04 \mu\text{m}^2\text{s}^{-1}$ ). In MBD3-null cells, CHD4 still exhibits the slowly diffusing fraction ( $0.100 \pm 0.008 \mu\text{m}^2\text{s}^{-1}$ ), but the fast fraction now moves significantly more quickly ( $0.80 \pm 0.08 \mu\text{m}^2\text{s}^{-1}$ ) ( $p < 0.004$ ). Since our published paper, we have carried out a similar analysis using our 3D microscope. Longer trajectories were achieved by combining the reduced photobleaching of HaloTag® dye-labelled CHD4 (compared to mEos3.2) and 3D tracking to prevent molecules diffusing out of the detection plane (Carr *et al*, submitted). This study not only confirmed our previous 2D tracking results but allowed mean squared displacement analysis of the trajectories. The slowly diffusing fraction exhibits constrained diffusion over time and so appears to be stably bound to chromatin, whereas the fast fraction exhibits Brownian diffusion. Overall, the results suggest that a fraction of CHD4 associates with chromatin in either the presence or absence of MBD3, whereas the freely diffusing fraction of CHD4 molecules is significantly affected by

the loss of MBD3 consistent with a lack of interaction with the NuRD complex. This demonstrates the feasibility of studying interactions within the NuRD complex using a combination of PALM and knockout experiments.

In collaboration with Prof David Klenerman (Cambridge, UK) and Anthony Carr (Sussex, UK), we have also used and published a live-cell motion blurring approach that can detect proteins bound to chromatin in yeast (Etheridge *et al*, 2014). This approach uses long exposures to blur proteins that are diffusing so as to only detect chromatin bound proteins. We are now applying this to the detection of NuRD proteins in mouse ES cells to ask whether NuRD is required for stable binding of CHD4 (residence time changes) or simply for CHD4 association with chromatin (the number of non-specific binding events before a stable interaction will change) in a similar manner to which Sox2 has been shown to facilitate Oct4-DNA interactions (Chen *et al*, 2014).

### **Development of (new) Screening Platforms**

During this project, we developed two different screening platforms to select small molecules able to interfere with CBX7-H3K27Me3 interaction.

- 1) Fluorescence polarization assay. This is an *in vitro* assay described in Simhadri *et al.*, (J Med Chem. 2014 Apr 10;57(7):2874-83) based on fluorescence polarization. The assay is carried out using a synthetic complex composed by either a recombinant truncated form of CBX7, containing the chromodomain (amino acids 8-62), or the full-length Cbx7, together with a modified peptide derived from H3 (5(6)-carboxyfluorescein-QLATKAAR-Lys(Me3)-SAATG).
- 2) Nano BRET Assay is a commercial available assay developed by Promega. The Nano BRET Assay is a protein-protein interaction assay that uses Nano-Luc Luciferase as the BRET energy donor and HaloTag Protein labeled with the HaloTag Nano BRET 618 Fluorophore as the energy acceptor to measure the interaction of two specific proteins. Importantly, Nano BRET Assays allow the measurement of protein interactions in live cells leading to the selection of molecules which are membrane permeable and able to interact with the full-length Cbx7 protein embedded into the whole physiological PRC1 complex.

### **4.2. Wider societal implications of the project**

Overall, the results from our small molecules screening of compounds (inhibiting CBX7-H3K27Me3 interaction) represent a clear advancement in the demonstration of the drugability of the chromodomain, that we are showing to be critical for the oncogenic activity of CBX7 in AMLs. Therefore, our studies mark the beginning of drug discovery activities against a novel class of targets in cancer.

Our studies using combinations of epigenetic drugs (LSD1, EZH2, HDAC inhibitors, retinoids) have demonstrated at the preclinical level that targeting both bulk leukemic cells and the leukaemia initiating cell compartment are both required for disease eradication, suggesting new clinical approaches for the treatment of AMLs.

### 4.3. main dissemination activities

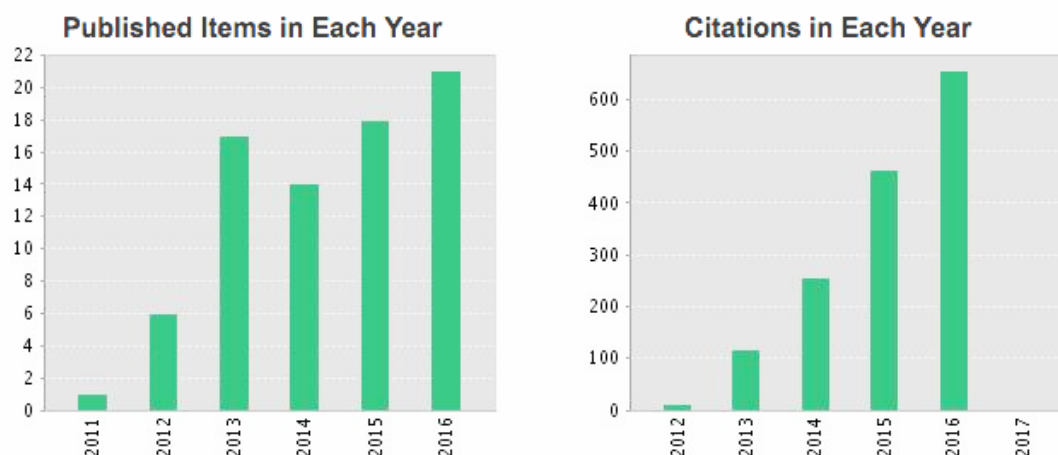
It was a central goal of 4DCellFate to promote dissemination activities within and outside the consortium, to scientists, to the industrial sector and to the general public, in order to raise awareness of the project, of its results and its future impact on society. In order to obtain support in dissemination actions from experts in science communication, the project created the 4DCellFate Communication Network, a network of all the Communication Offices at the partner institutions which actively participated in specific actions.

#### Dissemination within 4DCellFate

The Project Coordinator and the project management team ensured that all partners shared their data and results through e-mails, the website, annual meetings, phone and skype conferences as well as specific internal workshops. An intranet in form of dropbox professional was set up and maintained by the project manager to improve communication among consortium members and included a structured and searchable document repository (collecting official documents, presentations, relevant literature, etc.).

#### Dissemination to the scientific community

All partners were committed to communicate their (sub) project results to the scientific community and to share resources, protocols and reagents with other laboratories outside the project. Results were published in peer review journals and communicated at international and/or national meetings and conferences. To date, each of the four subprojects has generated large amount of novel data, which led to numerous publications in top impact journals (such as Nature, Nature Structural and Molecular Biology, Oncogene, etc.): 85 articles were published with support of the *4DCellFate* project, of which seven were collaborative within the project (**Figure 36**). These publications had an approximate average impact factor of 12.5 and were cited 1503 times in total (1389 times excluding self-citations). Eleven of the publications have been published in open access journals, in total 50 publications with open access (either green or golden way).



**Figure 36:** Number of publications of the 4DCellFate project (A) and numbers of their citations (B) in each year of project run time. Data extracted from Web of Science Citation Report.

4DCellFate worked actively to join forces with other European initiatives. Organization of joint meetings or workshops were explored to exchange results, technologies and foster communication with those initiatives. One example is the co-organization of a workshop with the 4DGenome initiative in Barcelona. Through the 4DCellFate Communication Network, joint press releases were distributed to local media in the partner countries, to the international press and to the EC press office (together with proper acknowledgement of the funding bodies). A public website was set up in the first few months of the 4DCellFate project ([www.4dcellfate.eu](http://www.4dcellfate.eu)) offering different types of information for different potential readers and users (scientists, funding bodies, companies, general public, etc). It was designed following the EC communication guidelines and with the support of the 4DCellFate Communication Network. The website includes scientific information on project goals, organization, partners, main achievements, and common resources.

### **Dissemination to the general public and outreach activities**

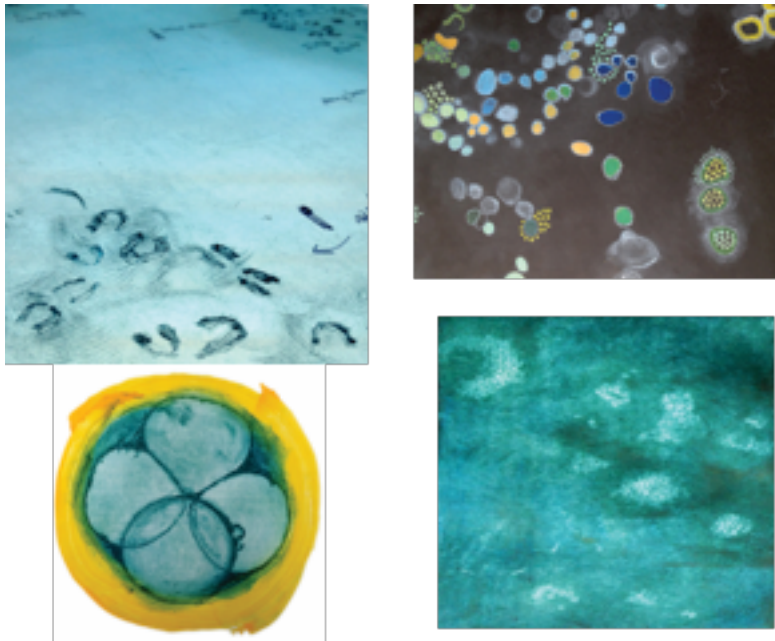
As a publicly funded project, 4DCellFate saw both the responsibility and the interest to freely communicate its innovative research and cutting-edge science to society, and to bring the questions, concerns and responses of the public back to the laboratory. This was achieved through different means and with the support of the 4DCellFate Communication Network. Most of the partner institutions in the project organized multiple outreach activities, such as open days, science & art activities, educational programmes for school teachers (see as examples ELLS at EMBL, and “BIO” at CRG), workshops for primary and secondary school kids, scientific cafes, and “easy” scientific lectures. The 4DCellFate Communication Network encouraged the partners (from principal investigators to PhD students) to get involved in such initiatives, and share their experience in the 4DCellFate network. P1a has hosted a series of experimental workshops for high school students at the CRG, reaching a total of 41 high school classes throughout Catalonia and more than 700 students for intense workshops in small groups, with a high student-to-scientist ratio. These experimental, 1-day workshops for late-stage high school students were held prior to the time when the students have to decide their “baccalaurate” topics. These workshops were intended to arouse the interest of the students in scientific research and biomedicine, with overviews about cancer and stem cells, at this critical time when they are deciding their future directions.

Furthermore, P1 organised on 16 September 2015 a Scientific Café dedicated to the topic “Epigenetics – beyond GATTACA”. On this public event with more than 80 participants, two researchers (one of them from P1 lab) presented their work and discussed towards the broad public the role of epigenetics on gene activity and potential applications to heal diseases like cancer.

Additionally, P1 has developed an experimental electrophoresis kit for introducing high school students to the basics of gene transcription, based on (and co-sponsored by) 4DCellFate, to be done on-site within the schools. These kits are highly economical (each for about 30 Euro) and can be distributed widely: in 2014, 50 kits were sent out to reach about 1500 high school students within Catalunya. P1 is considering starting a company for this dissemination. Numerous other outreach activities have also occurred through this reporting period, as listed in Table 3, that have used catchy methods for getting the public enthusiastic

about science as well as for explaining what they do. For instance, Science Speed Dating uses the popular social format to explain scientific careers to secondary school students.

The consortium worked furthermore together with the artist Ana Cid (Barcelona) on a project entitled “The Art of Science”, which will consist of artistic representations (etches) of images taken from the 4DCellFate project (for example, see **Figure 37**). These images were displayed on the 4DCellFate website with descriptions of the actual experiments, and an exhibit with A. Cid is being planned for the near future.



**Figure 37.** Examples of etches by A. Cid based on *4DCellFate* project images.

As one final joint outreach activity, the 4DCellFate consortium produced a short movie (10min), targeting students of Life Science and the interested broad public. 4DCellFate researchers themselves explained – through interviews and animations –the project, its activities and main results, their relevance and future challenges. The movie is uploaded on the youtube channel of P1 announced on the partner institutes websites and social media. To guarantee a high quality of the movie, a professional movie production company with experience in the scientific sector was subcontracted.

## Annex

### Publications

**Publications from the 4DCellFate project.** Collaborative publications are marked with an asterisk.

- 1\* Alfieri, C., Gambetta, M. C., Matos, R., Glatt, S., Sehr, P., Fraterman, S., Wilm, M., Müller, J. & Müller, C. W. Structural basis for targeting the chromatin repressor Sfmtb to Polycomb response elements. *Genes & development* **27**, 2367-2379, doi:DOI 10.1101/gad.226621.113 (2013).
- 2 Aloia, L., Demajo, S. & Di Croce, L. ZRF1: a novel epigenetic regulator of stem cell identity and cancer. *Cell cycle* **14**, 510-515, doi:10.4161/15384101.2014.988022 (2015).
- 3 Aloia, L., Di Stefano, B. & Di Croce, L. Polycomb complexes in stem cells and embryonic development. *Development* **140**, 2525-2534, doi:DOI 10.1242/dev.091553 (2013).
- 4 Aloia, L., Di Stefano, B., Sessa, A., Morey, L., Santanach, A., Gutierrez, A., Cozzuto, L., Benitah, S. A., Graf, T., Broccoli, V. & Di Croce, L. Zrf1 is required to establish and maintain neural progenitor identity. *Genes & development* **28**, 182-197, doi:10.1101/gad.228510.113 (2014).
- 5 Aloia, L., Gutierrez, A., Caballero, J. M. & Di Croce, L. Direct interaction between Id1 and Zrf1 controls neural differentiation of embryonic stem cells. *EMBO reports* **16**, 63-70, doi:10.15252/embr.201439560 (2015).
- 6\* Alqarni, S. S., Murthy, A., Zhang, W., Przewloka, M. R., Silva, A. P., Watson, A. A., Lejon, S., Pei, X. Y., Smits, A. H., Kloet, S. L., Wang, H., Shepherd, N. E., Stokes, P. H., Blobel, G. A., Vermeulen, M., Glover, D. M., Mackay, J. P. & Laue, E. D. Insight into the Architecture of the NuRD Complex: STRUCTURE OF THE RbAp48-MTA1 SUBCOMPLEX. *The Journal of biological chemistry* **289**, 21844-21855, doi:10.1074/jbc.M114.558940 (2014).
- 7 Aranda, S., Shi, Y. & Di Croce, L. Chromatin and Epigenetics at the Forefront: Finding Clues among Peaks. *Molecular and cellular biology* **36**, 2432-2439, doi:10.1128/MCB.00328-16 (2016).
- 8 Ballaré, C., Lange, M., Lapinaite, A., Martin, G. M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., Carlomagno, T., Benitah, S. A. & Di Croce, L. Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat Struct Mol Biol* **19**, 1257-1265, doi:DOI 10.1038/nsmb.2434 (2012).
- 9 Barberan-Soler, S., Fontrodona, L., Ribo, A., Lamm, A. T., Iannone, C., Ceron, J., Lehner, B. & Valcarcel, J. Co-option of the piRNA pathway for germline-specific alternative splicing of *C. elegans* TOR. *Cell reports* **8**, 1609-1616, doi:10.1016/j.celrep.2014.08.016 (2014).
- 10 Barford, D., Takagi, Y., Schultz, P. & Berger, I. Baculovirus expression: tackling the complexity challenge. *Curr Opin Struct Biol* **23**, 357-364, doi:DOI 10.1016/j.sbi.2013.03.009 (2013).
- 11 Baymaz, H. I., Fournier, A., Laget, S., Ji, Z., Jansen, P. W., Smits, A. H., Ferry, L., Mensinga, A., Poser, I., Sharrocks, A., Defossez, P. A. & Vermeulen, M. MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. *Proteomics* **14**, 2179-2189, doi:10.1002/pmic.201400013 (2014).
- 12 Baymaz, H. I., Karemaker, I. D. & Vermeulen, M. Perspective on unraveling the versatility of 'co-repressor' complexes. *Biochimica et biophysica acta* **1849**, 1051-1056, doi:10.1016/j.bbagr.2015.06.012 (2015).
- 13 Beringer, M., Pisano, P., Di Carlo, V., Blanco, E., Chammas, P., Vizan, P., Gutierrez, A., Aranda, S., Payer, B., Wierer, M. & Di Croce, L. EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells. *Molecular cell* **64**, 645-658, doi:10.1016/j.molcel.2016.10.018 (2016).
- 14 Bieniossek, C., Imasaki, T., Takagi, Y. & Berger, I. MultiBac: expanding the research toolbox for multiprotein complexes. *Trends in biochemical sciences* **37**, 49-57, doi:10.1016/j.tibs.2011.10.005 (2012).

- 15\* Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P. & Berger, I. The architecture of human general transcription factor TFIID core complex. *Nature* **493**, 699-702, doi:DOI 10.1038/nature11791 (2013).
- 16 Burga, A. & Lehner, B. Beyond genotype to phenotype: why the phenotype of an individual cannot always be predicted from their genome sequence and the environment that they experience. *The FEBS journal* **279**, 3765-3775, doi:10.1111/j.1742-4658.2012.08810.x (2012).
- 17 Chittock, E. C., Latwiel, S., Miller, T. C. & Muller, C. W. Molecular architecture of polycomb repressive complexes. *Biochem Soc Trans* **45**, 193-205, doi:10.1042/BST20160173 (2017).
- 18 Comet, I., Riising, E. M., Leblanc, B. & Helin, K. Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat Rev Cancer* **16**, 803-810, doi:10.1038/nrc.2016.83 (2016).
- 19 Corujo, D., Mas, G., Malinverni, R., Di Croce, L. & Buschbeck, M. Barcelona conference on epigenetics and cancer 2015: Coding and non-coding functions of the genome. *Epigenetics : official journal of the DNA Methylation Society* **11**, 95-100, doi:10.1080/15592294.2015.1131377 (2016).
- 20 Demajo, S., Uribealago, I., Gutierrez, A., Ballare, C., Capdevila, S., Roth, M., Zuber, J., Martin-Caballero, J. & Di Croce, L. ZRF1 controls the retinoic acid pathway and regulates leukemogenic potential in acute myeloid leukemia. *Oncogene*, 1-10, doi:10.1038/onc.2013.501 (2013).
- 21 Demajo, S., Uribealago, I., Gutierrez, A., Ballare, C., Capdevila, S., Roth, M., Zuber, J., Martin-Caballero, J. & Di Croce, L. ZRF1 controls the retinoic acid pathway and regulates leukemogenic potential in acute myeloid leukemia. *Oncogene* **33**, 5501-5510, doi:10.1038/onc.2013.501 (2014).
- 22\* Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol* **20**, 1147-1155, doi:10.1038/nsmb.2669 (2013).
- 23 Francesconi, M. & Lehner, B. The effects of genetic variation on gene expression dynamics during development. *Nature* **505**, 208-211, doi:DOI 10.1038/nature12772 (2014).
- 24 Francesconi, M. & Lehner, B. Reconstructing and analysing cellular states, space and time from gene expression profiles of many cells and single cells. *Mol Biosyst* **11**, 2690-2698, doi:10.1039/c5mb00339c (2015).
- 25 Frey, F., Sheahan, T., Finkl, K., Stoehr, G., Mann, M., Benda, C. & Muller, J. Molecular basis of PRC1 targeting to Polycomb response elements by PhoRC. *Genes & development* **30**, 1116-1127, doi:10.1101/gad.279141.116 (2016).
- 26 Gambetta, M. C. & Muller, J. O-GlcNAcylation Prevents Aggregation of the Polycomb Group Repressor Polyhomeotic. *Developmental cell*, doi:10.1016/j.devcel.2014.10.020 (2014).
- 27 Gambetta, M. C. & Muller, J. A critical perspective of the diverse roles of O-GlcNAc transferase in chromatin. *Chromosoma*, doi:10.1007/s00412-015-0513-1 (2015).
- 28 Jain, P. & Di Croce, L. Mutations and deletions of PRC2 in prostate cancer. *BioEssays : news and reviews in molecular, cellular and developmental biology* **38**, 446-454, doi:10.1002/bies.201500162 (2016).
- 29 Janich, P., Toufighi, K., Solanas, G., Luis, N. M., Minkwitz, S., Serrano, L., Lehner, B. & Benitah, S. A. Human Epidermal Stem Cell Function Is Regulated by Circadian Oscillations. *Cell stem cell*, doi:10.1016/j.stem.2013.09.004 (2013).
- 30 Julien, P., Minana, B., Baeza-Centurion, P., Valcarcel, J. & Lehner, B. The complete local genotype-phenotype landscape for the alternative splicing of a human exon. *Nat Commun* **7**, 11558, doi:10.1038/ncomms11558 (2016).
- 31\* Kalb, R., Latwiel, S., Baymaz, H. I., Jansen, P. W., Muller, C. W., Vermeulen, M. & Muller, J. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol* **21**, 569-571, doi:10.1038/nsmb.2833 (2014).
- 32 Kloet, S. L., Baymaz, H. I., Makowski, M., Groenewold, V., Jansen, P. W., Berendsen, M., Niazi, H., Kops, G. J. & Vermeulen, M. Towards elucidating the stability, dynamics and architecture

- of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. *The FEBS journal* **282**, 1774-1785, doi:10.1111/febs.12972 (2015).
- 33 Kloet, S. L., Irem Baymaz, H., Makowski, M., Groenewold, V., Jansen, P., Berendsen, M., Niazi, H., Kops, G. J. & Vermeulen, M. Towards elucidating the stability, dynamics and architecture of the NuRD complex using quantitative interaction proteomics. *The FEBS journal*, doi:10.1111/febs.12972 (2014).
- 34\* Kloet, S. L., Makowski, M. M., Baymaz, H. I., van Voorthuijsen, L., Karemaker, I. D., Santanach, A., Jansen, P. W., Di Croce, L. & Vermeulen, M. The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat Struct Mol Biol* **23**, 682-690, doi:10.1038/nsmb.3248 (2016).
- 35 Klosin, A. & Lehner, B. Mechanisms, timescales and principles of trans-generational epigenetic inheritance in animals. *Curr Opin Genet Dev* **36**, 41-49, doi:10.1016/j.gde.2016.04.001 (2016).
- 36 Koehler, C., Sauter, P. F., Wawryczyn, M., Girona, G. E., Gupta, K., Landry, J. J., Fritz, M. H., Radic, K., Hoffmann, J. E., Chen, Z. A., Zou, J., Tan, P. S., Galik, B., Junttila, S., Stolt-Bergner, P., Pruneri, G., Gyenesei, A., Schultz, C., Biskup, M. B., Besir, H., Benes, V., Rappsilber, J., Jechlinger, M., Korbelt, J. O., Berger, I., Braese, S. & Lemke, E. A. Genetic code expansion for multiprotein complex engineering. *Nature methods* **13**, 997-1000, doi:10.1038/nmeth.4032 (2016).
- 37 Kruger, A. V., Jelier, R., Dzyubachyk, O., Zimmerman, T., Meijering, E. & Lehner, B. Comprehensive single cell-resolution analysis of the role of chromatin regulators in early *C. elegans* embryogenesis. *Dev Biol*, doi:10.1016/j.ydbio.2014.10.014 (2014).
- 38 Kruger, A. V., Jelier, R., Dzyubachyk, O., Zimmerman, T., Meijering, E. & Lehner, B. Comprehensive single cell-resolution analysis of the role of chromatin regulators in early *C. elegans* embryogenesis. *Dev Biol* **398**, 153-162, doi:10.1016/j.ydbio.2014.10.014 (2015).
- 39 Lehner, B. Genotype to phenotype: lessons from model organisms for human genetics. *Nature reviews. Genetics* **14**, 168-178, doi:10.1038/nrg3404 (2013).
- 40 Lindeboom, R. G., Supek, F. & Lehner, B. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat Genet* **48**, 1112-1118, doi:10.1038/ng.3664 (2016).
- 41 Mas, G. & Di Croce, L. The role of Polycomb in stem cell genome architecture. *Curr Opin Cell Biol* **43**, 87-95, doi:10.1016/j.ceb.2016.09.006 (2016).
- 42 Matthews, G. M., Mehdipour, P., Cluse, L. A., Falkenberg, K. J., Wang, E., Roth, M., Santoro, F., Vidacs, E., Stanley, K., House, C. M., Rusche, J. R., Vakoc, C. R., Zuber, J., Minucci, S. & Johnstone, R. W. Functional-genetic dissection of HDAC dependencies in mouse lymphoid and myeloid malignancies. *Blood* **126**, 2392-2403, doi:10.1182/blood-2015-03-632984 (2015).
- 43 McManus, J., Cheng, Z. & Vogel, C. Next-generation analysis of gene expression regulation--comparing the roles of synthesis and degradation. *Mol Biosyst* **11**, 2680-2689, doi:10.1039/c5mb00310e (2015).
- 44 Mesa, P., Deniaud, A., Montoya, G. & Schaffitzel, C. Directly from the source: endogenous preparations of molecular machines. *Curr Opin Struct Biol* **23**, 319-325, doi:10.1016/j.sbi.2013.01.005 (2013).
- 45 Montes, M., Nielsen, M. M., Maglieri, G., Jacobsen, A., Hojfeldt, J., Agrawal-Singh, S., Hansen, K., Helin, K., van de Werken, H. J., Pedersen, J. S. & Lund, A. H. The lncRNA MIR31HG regulates p16(INK4A) expression to modulate senescence. *Nat Commun* **6**, 6967, doi:10.1038/ncomms7967 (2015).
- 46 Morey, L., Aloia, L., Cozzuto, L., Benitah, S. A. & Di Croce, L. RYBP and Cbx7 Define Specific Biological Functions of Polycomb Complexes in Mouse Embryonic Stem Cells. *Cell reports* **S2211-1247(12)00423-8**, doi:10.1016/j.celrep.2012.11.026 (2012).
- 47 Morey, L., Aloia, L., Cozzuto, L., Benitah, S. A. & Di Croce, L. RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. *Cell reports* **3**, 60-69, doi:10.1016/j.celrep.2012.11.026 (2013).

- 48 Morey, L., Santanach, A., Blanco, E., Aloia, L., Nora, E. P., Bruneau, B. G. & Di Croce, L. Polycomb Regulates Mesoderm Cell Fate-Specification in Embryonic Stem Cells through Activation and Repression Mechanisms. *Cell stem cell* **17**, 300-315, doi:10.1016/j.stem.2015.08.009 (2015).
- 49 Morey, L., Santanach, A. & Di Croce, L. Pluripotency and Epigenetic Factors in Mouse Embryonic Stem Cell Fate Regulation. *Molecular and cellular biology* **35**, 2716-2728, doi:10.1128/MCB.00266-15 (2015).
- 50 O'Shaughnessy, A. & Hendrich, B. CHD4 in the DNA-damage response and cell cycle progression: not so NuRDy now. *Biochem Soc Trans* **41**, 777-782, doi:DOI 10.1042/BST20130027 (2013).
- 51 Palayret, M., Armes, H., Basu, S., Watson, A. T., Herbert, A., Lando, D., Etheridge, T. J., Endesfelder, U., Heilemann, M., Laue, E., Carr, A. M., Klenerman, D. & Lee, S. F. Virtual-'light-sheet' single-molecule localisation microscopy enables quantitative optical sectioning for super-resolution imaging. *PLoS One* **10**, e0125438, doi:10.1371/journal.pone.0125438 (2015).
- 52 Park, S. & Lehner, B. Epigenetic epistatic interactions constrain the evolution of gene expression. *Mol Syst Biol* **9**, 645, doi:DOI 10.1038/msb.2013.2 (2013).
- 53 Park, S. & Lehner, B. Cancer type-dependent genetic interactions between cancer driver alterations indicate plasticity of epistasis across cell types. *Mol Syst Biol* **11**, 824, doi:10.15252/msb.20156102 (2015).
- 54 Pasini, D. & Di Croce, L. Emerging roles for Polycomb proteins in cancer. *Curr Opin Genet Dev* **36**, 50-58, doi:10.1016/j.gde.2016.03.013 (2016).
- 55 Pengelly, A. R., Kalb, R., Finkl, K. & Muller, J. Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes & development* **29**, 1487-1492, doi:10.1101/gad.265439.115 (2015).
- 56 Perez-Lluch, S., Blanco, E., Tilgner, H., Curado, J., Ruiz-Romero, M., Corominas, M. & Guigo, R. Absence of canonical marks of active chromatin in developmentally regulated genes. *Nat Genet* **47**, 1158-1167, doi:10.1038/ng.3381 (2015).
- 57 Reynolds, N., O'Shaughnessy, A. & Hendrich, B. Transcriptional repressors: multifaceted regulators of gene expression. *Development* **140**, 505-512, doi:DOI 10.1242/dev.083105 (2013).
- 58 Ribeiro, J. D., Morey, L., Mas, A., Gutierrez, A., Luis, N. M., Mejetta, S., Richly, H., Benitah, S. A., Keyes, W. M. & Di Croce, L. ZRF1 controls oncogene-induced senescence through the INK4-ARF locus. *Oncogene*, doi:10.1038/onc.2012.241 (2012).
- 59 Ribeiro, J. D., Morey, L., Mas, A., Gutierrez, A., Luis, N. M., Mejetta, S., Richly, H., Benitah, S. A., Keyes, W. M. & Di Croce, L. ZRF1 controls oncogene-induced senescence through the INK4-ARF locus. *Oncogene* **32**, 2161-2168, doi:10.1038/onc.2012.241 (2013).
- 60 Riising, E. M., Comet, I., Leblanc, B., Wu, X., Johansen, J. V. & Helin, K. Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Molecular cell* **55**, 347-360, doi:10.1016/j.molcel.2014.06.005 (2014).
- 61 Rinaldi, L., Datta, D., Serrat, J., Morey, L., Solanas, G., Avgustinova, A., Blanco, E., Pons, J. I., Matallanas, D., Von Kriegsheim, A., Di Croce, L. & Benitah, S. A. Dnmt3a and Dnmt3b Associate with Enhancers to Regulate Human Epidermal Stem Cell Homeostasis. *Cell stem cell* **19**, 491-501, doi:10.1016/j.stem.2016.06.020 (2016).
- 62 Rocha-Viegas, L., Villa, R., Gutierrez, A., Iriondo, O., Shiekhatar, R. & Di Croce, L. Role of UTX in RAR-mediated Gene Regulation in Leukemia. *Molecular and cellular biology*, doi:10.1128/MCB.00839-14 (2014).
- 63 Santoro, F., Botrugno, O. A., Dal Zuffo, R., Pallavicini, I., Matthews, G. M., Cluse, L., Barozzi, I., Senese, S., Fornasari, L., Moretti, S., Altucci, L., Pelicci, P. G., Chiocca, S., Johnstone, R. W. & Minucci, S. A dual role for Hdac1: oncosuppressor in tumorigenesis, oncogene in tumor maintenance. *Blood* **121**, 3459-3468, doi:DOI 10.1182/blood-2012-10-461988 (2013).

- 64 Schuster-Bockler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504-507, doi:10.1038/nature11273 (2012).
- 65 Sebe-Pedros, A., Ballare, C., Parra-Acero, H., Chiva, C., Tena, J. J., Sabido, E., Gomez-Skarmeta, J. L., Di Croce, L. & Ruiz-Trillo, I. The Dynamic Regulatory Genome of *Capsaspora* and the Origin of Animal Multicellularity. *Cell* **165**, 1224-1237, doi:10.1016/j.cell.2016.03.034 (2016).
- 66 Semple, J. I. & Lehner, B. Single and dual drug selection for transgenes following bombardment of *Caenorhabditis* species. *Methods* **68**, 409-416, doi:10.1016/j.ymeth.2014.05.004 (2014).
- 67 Simboeck, E., Gutierrez, A., Cozzuto, L., Beringer, M., Caizzi, L., Keyes, W. M. & Di Croce, L. DPY30 regulates pathways in cellular senescence through ID protein expression. *EMBO J* **32**, 2217-2230, doi:DOI 10.1038/emboj.2013.159 (2013).
- 68 Smits, A. H., Borrmann, A., Roosjen, M., van Hest, J. C. & Vermeulen, M. Click-MS: Tagless Protein Enrichment Using Bioorthogonal Chemistry for Quantitative Proteomics. *ACS chemical biology* **11**, 3245-3250, doi:10.1021/acschembio.6b00520 (2016).
- 69 Smits, A. H., Jansen, P. W., Poser, I., Hyman, A. A. & Vermeulen, M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res* **41**, e28, doi:DOI 10.1093/nar/gks941 (2013).
- 70 Smits, A. H. & Vermeulen, M. in *Systems Analysis of Chromatin-Related Protein Complexes in Cancer* (eds Andrew Emili, Jack Greenblatt, & Shoshana Wodak) 133-148 (Springer New York, 2014).
- 71 Smits, A. H. & Vermeulen, M. Characterizing Protein-Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. *Trends Biotechnol* **34**, 825-834, doi:10.1016/j.tibtech.2016.02.014 (2016).
- 72 Spruijt, C. G., Luijsterburg, M. S., Menafrá, R., Lindeboom, R. G., Jansen, P. W., Edupuganti, R. R., Baltissen, M. P., Wiegant, W. W., Voelker-Albert, M. C., Matarese, F., Mensinga, A., Poser, I., Vos, H. R., Stunnenberg, H. G., van Attikum, H. & Vermeulen, M. ZMYND8 Co-localizes with NuRD on Target Genes and Regulates Poly(ADP-Ribose)-Dependent Recruitment of GATAD2A/NuRD to Sites of DNA Damage. *Cell reports* **17**, 783-798, doi:10.1016/j.celrep.2016.09.037 (2016).
- 73 Supek, F. The Code of Silence: Widespread Associations Between Synonymous Codon Biases and Gene Function. *J Mol Evol* **82**, 65-73, doi:10.1007/s00239-015-9714-8 (2016).
- 74 Supek, F. & Lehner, B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature*, doi:10.1038/nature14173 (2015).
- 75 Supek, F., Lehner, B., Hajkova, P. & Warnecke, T. Hydroxymethylated cytosines are associated with elevated C to G transversion rates. *PLoS genetics* **10**, e1004585, doi:10.1371/journal.pgen.1004585 (2014).
- 76 Supek, F., Minana, B., Valcarcel, J., Gabaldon, T. & Lehner, B. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156**, 1324-1335, doi:10.1016/j.cell.2014.01.051 (2014).
- 77 Vavouri, T. & Lehner, B. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome biology* **13**, R110, doi:10.1186/gb-2012-13-11-r110 (2012).
- 78 Vizán, P., Beringer, M., Ballare, C. & Di Croce, L. Role of PRC2-associated factors in stem cells and disease. *The FEBS journal*, doi:10.1111/febs.13083 (2014).
- 79 Vizán, P., Beringer, M., Ballare, C. & Di Croce, L. Role of PRC2-associated factors in stem cells and disease. *The FEBS journal* **282**, 1723-1735, doi:10.1111/febs.13083 (2015).
- 80 Warnecke, T., Becker, E. A., Facciotti, M. T., Nislow, C. & Lehner, B. Conserved substitution patterns around nucleosome footprints in eukaryotes and Archaea derive from frequent

- nucleosome repositioning through evolution. *PLoS Comput Biol* **9**, e1003373, doi:10.1371/journal.pcbi.1003373 (2013).
- 81 Warnecke, T., Supek, F. & Lehner, B. Nucleoid-associated proteins affect mutation dynamics in *E. coli* in a growth phase-specific manner. *PLoS Comput Biol* **8**, e1002846, doi:10.1371/journal.pcbi.1002846 (2012).
- 82 Wu, X., Bekker-Jensen, I. H., Christensen, J., Rasmussen, K. D., Sidoli, S., Qi, Y., Kong, Y., Wang, X., Cui, Y., Xiao, Z., Xu, G., Williams, K., Rappsilber, J., Sonderby, C. K., Winther, O., Jensen, O. N. & Helin, K. Tumor suppressor ASXL1 is essential for the activation of INK4B expression in response to oncogene activity and anti-proliferative signals. *Cell Res* **25**, 1205-1218, doi:10.1038/cr.2015.121 (2015).
- 83\* Zhang, W., Aubert, A., Gomez de Segura, J. M., Karuppasamy, M., Basu, S., Murthy, A. S., Diamante, A., Drury, T. A., Balmer, J., Cramard, J., Watson, A. A., Lando, D., Lee, S. F., Palayret, M., Kloet, S. L., Smits, A. H., Deery, M. J., Vermeulen, M., Hendrich, B., Klenerman, D., Schaffitzel, C., Berger, I. & Laue, E. D. The Nucleosome Remodeling and Deacetylase Complex NuRD Is Built from Preformed Catalytically Active Sub-modules. *Journal of molecular biology* **428**, 2931-2942, doi:10.1016/j.jmb.2016.04.025 (2016).
- 84\* Zhang, W., Tyl, M., Ward, R., Sobott, F., Maman, J., Murthy, A. S., Watson, A. A., Fedorov, O., Bowman, A., Owen-Hughes, T., El Mkami, H., Murzina, N. V., Norman, D. G. & Laue, E. D. Structural plasticity of histones H3-H4 facilitates their allosteric exchange between RbAp48 and ASF1. *Nat Struct Mol Biol* **20**, 29-35, doi:DOI 10.1038/nsmb.2446 (2013).
- 85 Zhou, Y., Basu, S., Wohlfahrt, K. J., Lee, S. F., Klenerman, D., Laue, E. D. & Seshia, A. A. A microfluidic platform for trapping, releasing and super-resolution imaging of single cells. *Sens Actuators B Chem* **232**, 680-691, doi:10.1016/j.snb.2016.03.131 (2016).