EMBOpress

CROSS-JOURNAL FOCUS

Chromatin and epigenetics

EDITORS

Esther Schnapp
Senior Editor
esther.schnapp@embo.org | T +49 6221 8891 502
Esther joined EMBO reports in October 2008. She was awarded her PhD in 2005 at the Max Planck Institute for Molecular Cell Biology and Genetics in Dresden, Germany, where she studied tail regeneration in the axolotl. As a post-doc she worked on muscle development in zebrafish and on the characterisation of mesoangioblasts at the Stem Cell Research Institute of the San Raffaele Hospital in Milan, Italy.

Anne Nielsen
Editor
a.nielsen@embojournal.org | T +49 6221 8891 408
Anne received her PhD from Aarhus University in 2008 for work on miRNA processing in Joergen Kjems’ lab. As a postdoc she then went on to join Javier Martinez’ lab at IMBA in Vienna and focused on siRNA-binding proteins and non-conventional splicing in the unfolded protein response. Anne joined The EMBO Journal in 2012.

Maria Polychronidou
Editor
maria.polychronidou@embo.org | T +49 6221 8891 410
Maria received her PhD from the University of Heidelberg, where she studied the role of nuclear membrane proteins in development and aging. During her post-doctoral work, she focused on the analysis of tissue-specific regulatory functions of Hox transcription factors using a combination of computational and genome-wide methods.

Céline Carret
Editor
c.carret@embomolmed.org | T +49 6221 8891 310
Céline Carret completed her PhD at the University of Montpellier, France, characterising host immunodominant antigens to fight babesiosis, a parasitic disease caused by a unicellular Apicomplexan parasite closely related to the malaria agent Plasmodium. She further developed her post-doctoral career on malaria working at the Wellcome Trust Sanger Institute in Cambridge, UK and Instituto de Medicina Molecular in Lisbon, Portugal. Céline joined EMBO Molecular Medicine as a Scientific Editor in March 2011.
Epigenetic predisposition to reprogramming fates in somatic cells

Maayan Pour¹,², Inbar Pilzer¹,‡, Roni Rosner³, Zachary D Smith², Alexander Meissner² & Iftach Nachman¹,‡*

Abstract

Reprogramming to pluripotency is a low-efficiency process at the population level. Despite notable advances to molecularly characterize key steps, several fundamental aspects remain poorly understood, including when the potential to reprogram is first established. Here, we apply live-cell imaging combined with a novel statistical approach to infer when somatic cells become fated to generate downstream pluripotent progeny. By tracing cell lineages from several divisions before factor induction through to pluripotent colony formation, we find that pre-induction sister cells acquire similar outcomes. Namely, if one daughter cell contributes to a lineage that generates induced pluripotent stem cells (iPSCs), its paired sibling will as well. This result suggests that the potential to reprogram is predetermined within a select subpopulation of cells and heritable, at least over the short term.

We also find that expanding cells over several divisions prior to factor induction does not increase the per-lineage likelihood of successful reprogramming, nor is reprogramming fate correlated to neighboring cell identity or cell-specific reprogramming factor levels. By perturbing the epigenetic state of somatic populations with Ezh2 inhibitors prior to factor induction, we successfully modulate the fraction of iPSC-forming lineages. Our results therefore suggest that reprogramming potential may in part reflect preexisting epigenetic heterogeneity that can be tuned to alter the cellular response to factor induction.

Keywords cell fate decisions; live-cell imaging; reprogramming

Subject Category Stem Cells

DOI 10.15252/embr.201449264 | Received 3 July 2014 | Revised 11 December 2014 | Accepted 12 December 2014 | Published online 19 January 2015

EMBO Reports (2014) 16: 370–378

Introduction

Somatic cells can be reprogrammed to a pluripotent state by overexpression of defined transcription factors, Oct4, Sox2, Klf4, and c-Myc (OSKM) [1–4]. The reprogramming process is characterized by widespread epigenetic changes that generate induced pluripotent stem cells (iPSCs) with the functional and molecular characteristics of embryonic stem cells (ESCs) derived from the early embryo [3,5–9]. Generation of IPS cells is a robust and highly reproducible procedure, yet it is exceedingly inefficient at the per-cell level and requires an extended latency before autonomous pluripotency is acquired [10]. Different models have been suggested to explain these two notable attributes [11]. On one extreme, a fully stochastic model suggests that every cell division essentially constitutes a coin toss in which the cell ‘decides’ whether or not to reprogram. In this model, all cells are equally likely to reprogram at any time after factors have been induced, independent of their history prior to the time of induction. On the other extreme, a deterministic ‘elite’ model posits that the initial cell population contains a subpopulation that is predisposed to fated to successfully reprogram. Experimental work designed to test these models have offered different perspectives that vary between these two extremes. For instance, low-frequency stochastic reprogramming is inherent to any continuously proliferating lineage given enough time and cellular divisions [12]. We have previously described an early decision point after which the trajectory to successful reprogramming becomes defined, suggesting that the initial response to factor induction may determine the downstream trajectory [13]. By sampling single-cell transcription over the reprogramming timeline, high cell-to-cell variability in gene expression was found to describe early reprogramming, after which a deterministic hierarchical phase is acquired [14]. Finally, a recent study suggested that only select subpopulations of granulocyte-monocyte progenitors can reprogram, and do so with higher efficiency, while the majority of cells remain intrinsigent [15].

To investigate this in more detail, we used a live imaging approach to characterize key decision points and contributing factors during the reprogramming process. We find evidence suggesting that the potential to reprogram is largely pre-established within somatic cells before reprogramming factors are induced. We show that responding cells differ in their pre-induction properties from non-responding ones and that perturbing the epigenetic state of the somatic population prior to reprogramming can alter the potential of single cells to generate iPSC-forming lineages. Our
findings emphasize the relevance of preexisting cell-to-cell variability in reprogramming, expanding prior studies that pointed to the early stages following factor induction as critical to the final outcome. These observations will eventually lead to a better molecular definition of cellular state that includes a given cell’s potential to respond to transcriptional perturbation and has implications to other processes beyond reprogramming to pluripotency.

Results

We employed the NGFP2 MEF secondary reprogramming system, where all somatic cells contain identical integrations of OSKM factors under doxycycline (dox)-inducible promoters [13,16], and therefore, phenotypic variability must be non-genetic in origin. In all experiments, we focused on colonies generated within a 2-week reprogramming timeline to concentrate on the initial wave of iPSC colony formation, which is generally more defined than colonies that emerge later [12, 13]. To assist in cell lineage tracking, we transduced the MEFs with lentiviral vectors that constitutively express fluorescent proteins prior to reprogramming to create populations of uniquely labeled cells. To minimize efficiency calculation errors resulting from previously described satellite colonies that emanate from primary reprogramming lineages [13] and do not represent de novo acquisition of pluripotency, we used a colony-counting method which is estimated exclusively from colonies that can be traced back to the original fibroblast (see Materials and Methods).

Cells are predisposed to major cell fate decisions before factor induction

To determine when the potential to successfully generate iPSC colonies is established, we devised a strategy inspired by the Luria–Delbrück experiment. The original experiment demonstrated that acquisition of resistance through mutation precedes selection by employing a pre-growth period prior to screening for mutants [17]. In our version, we begin with a known number of MEFs and allow them to divide several times prior to factor induction, increasing the number of cells per well while holding the number of lineages constant (Fig 1A). If the potential to reprogram is largely predetermined, the fraction of iPSC containing wells will depend on the initial population size, assuming that the potential is inherited in daughter cells over the short term. In a post-determined model, reprogramming will depend only on the number of cells at the time of induction, increasing the fraction of iPSC containing wells as a function of population number.

We seeded cells at different low densities in 96-well plates (n = 17) and initiated reprogramming after 0 days (13 plates) or 5 days (4 plates), counting the exact number of cells in each well both at the day of plating and at the time of OSKM induction. After 2 weeks, we assessed the fraction of wells containing iPSC-marker-positive colonies and inferred the per-cell efficiency (Materials and Methods). By counting wells rather than colonies, we avoid inaccurate scoring of satellite colonies as unique reprogramming events when estimating efficiency [13]. Given an expected reprogramming efficiency of 1% for our system, we seeded only a small number of cells per well (10–100) to ensure that the fraction containing iPSCs will be within the dynamic range (i.e., lower than 96 wells) to precisely measure per-cell efficiency. As it has been reported that reprogramming potential in fibroblasts is diminished by progressive passaging [18], we separately tested reprogramming efficiency in cells that were expanded for 5 days before being replated and induced by OSKM (Supplementary Text S1, Supplementary Fig S1). We found no effect for the 5-day expansion period, ruling out a possible confounding effect of reduced efficiency due to a later generation.

We first calculated the reprogramming efficiency using plates in which dox was applied on day 0 (with no delay, Fig 1A, top). In these plates, the starting cell number and the number on the day of induction are trivially the same, so the estimated efficiency would be applicable to both models. By computing the fraction of positive wells for a certain starting number, the standard efficiency can be easily calculated (Fig 1B, blue). After verifying that reprogramming efficiency does not depend on the location of the well on the plate (Supplementary Fig S2), we determined the ‘standard failure rate’ parameter, which represents the probability that a cell does not produce any reprogrammed progeny, to be 0.998 ± 0.004 (maximum-likelihood estimator, see Supplementary Text S2), consistent with previous lineage normalized estimates using this system [13]. A similar efficiency was obtained using labeled subpopulations within a standard cell density (Supplementary Fig S3), verifying the low density in this experiment does not affect efficiency.

We next estimated the efficiency in wells where the addition of dox followed a 5-day expansion period of our restricted starting cell populations (Fig 1A, bottom). To increase the distinction between the two tested models, we only used wells that at least doubled their cell number between the two counts (n = 71 wells). We estimated the failure rate parameter separately according to starting cell count (0.990 ± 0.002) or count at time of dox (0.996 ± 0.001) using bootstrap sampling over a maximum-likelihood estimator (see Supplementary Text S2). The failure rate parameter according to starting cell count is closer to the standard failure rate parameter computed above, where the number of reprogramming lineages matches the number of cells at the time of induction. In fact, the efficiency as estimated from the number of cells at the time of induction (‘+ dox count’) is 2.5-fold lower than expected if the likelihood of reprogramming were stochastically distributed to all cells equally at that time (Fig 1C). To visualize the reprogramming efficiency according to each model, we divided the 71 wells to 6 groups of 10–13 wells, where each group has a similar initial cell count (Supplementary Table S1), and plotted the average efficiency for each group according to either initial count or dox-day count (Fig 1B). Efficiency as a function of starting cell number (Fig 1B, red) is closer to the day 0 ‘scaling points’ (Fig 1B, blue) than the efficiency as a function of cell number at time of induction (Fig 1B, green). This suggests that starting cell count (day 0) is a better predictor of reprogramming efficiency than the number of cells at the time of induction, favoring a predetermined per-lineage model over any post-determined model, including a fully stochastic one. To rule out the possibility that a higher local density after 5 days could reduce the per-cell exposure to dox, or deleteriously bias reprogramming in any way, we specifically disrupted cellular position within each well by brief trypsinization prior to factor induction, resulting in no effect on the apparent lineage dependence of reprogramming outcome (Supplementary Fig S4).
Epigenetic predisposition to reprogramming fates

Maayan Pour et al

EMBO reports

Predetermined potential is symmetrically maintained over the short term

The results above suggest the potential to reprogram is determined before OSKM induction, with limited acquired potential generated during ensuing divisions. However, we do not know how the potential is inherited within the lineage or how stable it is. For example, it could be restricted in sequential steps along the lineage, similar to fate restriction during early development. Alternatively, it could be equally inherited during each cell division. A third option is that cells within the potentiated lineage interconvert between ‘amenable’ and ‘recalcitrant’ states. To better understand how reprogramming potential is restricted and to validate the point of its appearance, we analyzed the fate statistics of lineage pairs of fluorescently labeled secondary MEFs. Cells were tracked from 2 days before induction of reprogramming. After the first division, prior to dox induction, two sister cells were tagged as paired lineages. We designated the fate of each cellular lineage into one of three categories: iPSC (Nanog-GFP+ forming), fast dividers (FD, Nanog-GFP-, indicative of transformation without reprogramming), and non-responder (NR, which do not acquire rapid proliferation or exhibit overt changes in fibroblast morphology, see Materials and Methods for full definitions). We then independently assigned these fates to both lineages within a pair to examine the possible combinations between them (Fig 2A). Notably, in each of the pairs examined, the two lineages adopted the same fate (Fig 2B, Supplementary Movie S1). We prospectively counted 58 pairs of FD-FD lineages, 6 pairs of iPS-iPS lineages, and 79 pairs of NR-NR lineages. We did not observe any pair of lineages...
acquiring mixed fates (such as iPSC-FD, where only one of the lineages contributes to iPSC and the other is transformed). Repeating the same experiment at a higher cell density resulted in similar within-pair correlations, verifying the effect is not density dependent (Supplementary Fig S5).

We used these lineage pair counts to rule out different models in which potential is acquired or lost after the first division. For example, by taking the observed counts of all 6 possible paired lineage combinations for the three fates (FD, NR, or iPSC), we can reject a model in which cells ‘decide’ their fate after the initial division (here, the term decision refers to gain or loss of a fate potential). It is possible, however, that multiple fate decisions may occur within discrete steps. For example, cells may or may not decide to proliferate in response to OSKM, and only as a second decision may proliferating cells acquire full reprogramming potential (Fig 2C). The time of acquiring each of these potentials would be reflected statistically within our lineage pair counts. A model in which cells acquire the potential to proliferate (shared between iPSC and FD fates) only after the first division can be ruled out by computing a P-value of iPSC and FD versus NR lineage pairs, which represents the probability of getting the observed count or higher of same-fate lineage pairs from our data compared to an alternate random pairing model. Given random pairing, if potential is acquired independently after the initial division, the observed combinations of pairs (in this case NR-NR, NR-FD/iPSC, or FD/iPSC-FD/iPSC) will follow a random distribution (see Supplementary Text S3). A model in which proliferative cells acquire or lose reprogramming potential after the initial division can also be ruled out by computing the P-value for which our FD versus iPSC lineage data reflect the random acquisition of FD-FD, FD-iPSC, and iPSC-iPSC pairs. Using our empirical paired lineage counts, we can reject all three of these models at high significance (Fig 2D, top). Repeating the same analysis for pairs resulting from the second observed division resulted in similar statistics (Fig 2D, bottom), suggesting fate potential is maintained (not gained or lost) over at least two divisions. Pairs from later divisions are harder to track, but some loss (or partial fulfillment) of fate potential in parts of the sub-lineages is observed (Fig 2B, Supplementary Movie S1).

The apparent predisposition toward different reprogramming fates suggests there are different internal states in the somatic cell population, which may be reflected by other cellular properties. We first tested whether the proliferation rate of the cells prior to induction correlates with their response to reprogramming by

Figure 2. Response potential is shared between daughter cells during early divisions.
A Schematic of the ‘paired lineage’ concept, depicting the last cell division prior to induction of OSKM (black arrow). The paired lineages can be classified into three categories with respect to when the potential to become an iPSC is obtained: potential may be acquired before the first division and both sub-lineages will include iPSC colony-forming events (left), potential may be acquired in one sub-lineage (or lost in its sibling lineage) after the first division, resulting in two different fates (center), or no potential is acquired over the lifetime (right). Green color denotes a sub-lineage that will form iPSCs.
B Snapshots from imaging a cell lineage originating from a single MEF where the first cell division occurred 8 h after imaging. From that division on, the paired lineages (marked red and green) were traced. After 18–40 h, the green lineage divided again and its sub-lineages were tracked separately. Dox was added at day 0. The final GFP+ colonies are composed of cells from different sub-lineages (see also Supplementary Movie S1). Scale bar, 500 μm.
C Possible paired lineage outcomes for two different decision points under a model assuming sequential acquisition of proliferation and reprogramming potentials. Left: possible pairs given the point of obtaining the potential for fast proliferation (marked by an arrow). Right: possible pairs given the point of obtaining the potential for reprogramming, assuming the cell already has fast proliferation potential.
D Counts of tracked lineage pairs from the 1st or 2nd division with each corresponding fate combination. No mixed-fate pairs (e.g., iPSC-FD) were observed. P-values are shown for each post-division fate decision model as described in text, rejecting all three post-division decision models.
E Division times during the 48 h that cells were tracked before initiating reprogramming, grouped according to their response as proliferating (either iPSC or FD, blue) and non-responding (NR, red) lineages. The ‘no division’ bar represents cells that did not divide prior to dox induction.
marking the number of divisions in each lineage during the 48 h prior to OSKM induction (Fig 2E). While proliferating cells of either the iPSC or FD fates could not be distinguished in this manner, both are likely to divide more times before induction than non-responding cells. Most responding cells divide at least twice during this period, while the majority of NR cells do not divide at all during the same period of time, and about 10% of them divide once. Thus, while division rate may distinguish between senescence-prone cells and proliferative cells, it cannot sufficiently predict whether responding cells will successfully navigate to pluripotency or simply acquire features associated with transformation. While we cannot rule out that the source of proliferation rate heterogeneity present in MEFs prior to factor induction could be affected by a mosaically represented genetic component, previous evidence converging non-responding to reprogramming cells by Mbd3 inhibition suggest these differences can be altered epigenetically [3].

Reprogramming potential is independent of local neighborhood and of early OSKM levels

What are the mechanisms that predetermine a cell’s response during the reprogramming process and what enables a predisposed lineage to realize this potential? One option is that environmental cues—such as reinforcing signals coming from neighboring cells—affect the future fate of the colonies. As cells within a lineage reside in close proximity to one another, they could respond similarly as a consequence of a shared local environment. Support from neighboring cells should be reflected by some preferential relative locations of future iPSC lineages to a specific type of lineage. To test this possibility, we examine the distribution of distances between starting cells in a reprogramming experiment (Fig 3A) and how it statistically depends on the final fates of their progeny. We observed no significant difference between the distance distributions of iPSC to FD, iPSC to NR, or FD to FD progenitors (Fig 3B), suggesting the relative location of starting cells does not affect their future fate. However, the reprogramming lineage itself could also supply a self-supportive local niche, which supports identical fates within each pair of sub-lineages. To temporarily remove the possible effect of lineage niche, we replated a CFP-labeled population of reprogramming MEFs at specific time points onto YFP-labeled cells reprogrammed in parallel (Fig 3C). In this system, reprogramming CFP cells are isolated from their original spatial niche, which includes both their lineage mates and neighboring lineages, and are randomly distributed among YFP lineages of different fates. After replating of the CFP cells, we followed the lineages for an additional 10–14 days and annotated the terminal fate of both CFP and YFP lineages. We then computed the distribution of distances between different lineage types (Fig 3D and E). At both early (days 2–6) or late (days 8–12) replating time points, we could not find any spatial effect—the distance distributions between all lineage types are similar. These results suggest that though signals from other colonies in the well may provide supportive and essential signals for successful reprogramming, the local proximity to specific neighboring lineages does not distinguish between different lineage fates.

It is possible that cell reprogramming rates during the course of reprogramming can disrupt the process. For example, it has been shown that a mesenchymal to epithelial transition (MET) occurs early during reprogramming [19, 20]. Replating during this phase may disturb MET mechanically and consequentially perturb downstream events. To study the effect of replating in isolation, we repeated the experiment differently by plating over empty or feeder-covered wells. We estimated final colony counts for different replating days (2–12) as well as time of colony appearance after replating (Supplementary Fig S6). Replating early in the process (days 2–4) resulted in both marked delay in appearance of colonies and lower number of final iPSC colonies compared to the non-replated case. Alternatively, replating during the later stages of the process (days 8–12) resulted in increasingly higher number of iPSC colonies that form with minimal delay after replating, suggesting that by this stage, iPSC-forming cells are more likely to maintain their route. For these later time points, replating may increase colony number as a trivial reflection of multiple iPSC-fated cells from each iPSC lineage being spatially distributed to different positions [14].

The per-cell expression level of the OSKM factors represents an ectopically induced cue that could also affect reprogramming potential. Despite the clonal origin of our secondary system, differential activation within single cells is possible. Different epigenetic states at sequence features of the lentiviral vectors could affect factor induction from their Tet-responsive promoters and lead to different fates in a simple way. For example, OSKM level could positively correlate to reprogramming outcome, and previous reports have shown that refractory reprogramming lineages with low factor expression can be rescued by elevating OSKM levels [21]. We found that OSKM levels are much higher in the NGFP2 inducible system than in the polycistronic OKSM cells used in Polo et al and that efficiency is not increased by additional supplementation (Supplementary Fig S7). To test whether the different behaviors are caused by different nuclear concentrations of the factors early in the reprogramming process, we examined the correlation between OSKM protein levels and the behavior of cells after induction. After 2 days of reprogramming, cells undergo consistent changes in morphology, usually resulting in a decrease in cell size [13] as well as nucleus size (Supplementary Fig S8). Using this behavior, we can distinguish cells that respond positively to factor induction (FD/iPSC) from those that do not. We stained reprogramming cells on days 0, 2, 4, and 6 days after induction using antibodies against OSKM. We indeed observe a variable level for each of the factors from day 2 onward, but found no negative correlation between nucleus size and the level of fluorescence (Fig 3F, Supplementary Fig S9). Together, these results suggest that the variable response to reprogramming is not due to obvious differences in OSKM factor levels at early stages.

Perturbing H3K27 or H3K4 methylation pre-induction alters future lineage fates

With exogenous explanations for these fated responses discounted, we hypothesized that differences in reprogramming potential may be epigenetic in origin and reflect innate differences in nuclear state. Discrete MEF responses may be a consequence of different chromatin states, either global or at the level of specific genes, that could permit constructive factor engagement at target sites upon their induction. Perturbation of chromatin modifiers has been extensively screened over the reprogramming process itself, some targets of
Figure 3. Reprogramming fate is not influenced by local signaling from neighbor lineages or by OSKM levels.

A, B All cells in a standard reprogramming experiment were annotated by final lineage fate, and distances between them at the time of induction were computed. Sample frame (A) at day 0 with 10 annotated cells. Scale bar, 100 μm. Histograms of cell-to-cell distances (B) between cells with different terminal fates show no significant relationship between proximity and outcome.

C, D Schematic of replating experiments. CFP- and YFP-labeled MEF cells were replated separately and in parallel. At specific time points within the 2nd to 12th day of reprogramming, CFP-labeled cells were replated onto stage-matched YFP cells. Scale bar, 200 μm.

E, F Histograms of distances between colonies of various fates replated after days 2–6 (B) or 8–12 (E). iPSC-C, iPSC-Y, CFP- or YFP-labeled iPSC colony, respectively. All-C, all-Y, CFP- or YFP-labeled colony of any type, respectively. The similarity between all distance distributions shows that there is no preference for the relative location of iPSC colonies.

G Correlation between the level of each factor on day 3 of reprogramming and morphological response. Plots show size of nuclear immunostaining signal for a given factor within induced cells against their median fluorescence intensity.

which contribute to population level effects in reprogramming efficiency [3,19,22,23]. Our findings suggest that such treatments could also be effective when limited to a period preceding OSKM induction, if they alter the cell’s epigenome in a manner that changes its predisposition to reprogram. To test this hypothesis, we subjected our secondary system to a panel of drugs that affect epigenetic modifications for either 24 or 48 h during the 2 days that precede factor induction (Fig 4A). We compared the NGFP+ colony count at day 14 to an untreated control. Of the drugs tested, both Ld1 and Ezh2 inhibitors showed the most significant increase in efficiency, increasing the number of Nanog-positive colonies by 3-fold compared to DMSO-treated and untreated controls (Fig 4B). Ezh2 is a histone methyltransferase that catalyzes repressive H3K27 methylation [24] (Supplementary Fig S10), while Ld1 is a histone demethylase removing H3K4 mono- and di-methylation [25]. Subsequently, inhibition of either may result in a permissive chromatin state that
could enable otherwise recalcitrant cells to switch to a reprogramming amenable state.

We further studied the effect of Ezh2 inhibitor pre-treatment on reprogramming. We hypothesized that Ezh2 inhibition could improve reprogramming by several different mechanisms: it could alter the number of reprogramming amenable MEFs, enable downstream stochastic fate switching, or simply amplify predisposed lineages, such that the final colony count is higher but the per-cell efficiency would be unchanged. To distinguish between these options, we repeated our ‘paired lineage’ experiment on cells treated with the Ezh2 inhibitor during the 2-day window where cell lineages are traced prior to dox induction. We reasoned that a delayed stochastic switch would enable the appearance of mixed pairs (e.g., FD-iPS), which would result in only one of the branches acquiring reprogramming potential. Alternatively, if all lineage pairs remain symmetrical in terms of their fates, as was observed in the untreated case, but with a higher fraction of pairs becoming iPSCs, then Ezh2 inhibition acts to increase the number of amenable cells within the population. Under the third scenario, the same number of lineages would reprogram, but would divide faster during the treatment period, creating more colonies of a secondary nature that increase the efficiency estimate artifactually. Supplementary Movie S2 shows a representative 16-day time lapse with lineage tracking for one such well. 27% of the wells (19/71) generated iPSC colonies, compared to 11% (11/96) in the no-treatment experiment (Fig 2), consistent with the global efficiency calculated during our screen (Fig 4B). Our data indicate the effect of Ezh2 inhibition is not on proliferation. Instead, the distribution of pair types (Fig 4C) suggests the majority of additional iPSC pairs may come from converted FD pairs. Though the majority of pairs we followed were still symmetrical, about 10% (10/98) were asymmetrical (4 NR-iPS pairs and 6 NR-FD pairs). These results suggest the Ezh2 inhibitor treatment increases the number of cells that generate iPSC-forming lineages, with some asymmetric lineage pairs between proliferative and non-proliferative fates possibly reflecting some cytotoxic effect. We compared the time of first division (that generates each pair) between Ezh2 inhibition and no pre-treatment conditions (Fig 4D).

The treatment appears to delay division time, causing a significant fraction of cells to divide for the first time only after OSKM induction. Interestingly, all mixed pairs divide early during the treatment period, while their following divisions occur mostly after dox induction and drug withdrawal (not shown). These pairs were exposed to Ezh2 inhibition for a longer period as individual cells than symmetrically fated pairs, allowing more time for the treatment to act differentially on isolated daughter cells, possibly allowing modulation between different responses.

Discussion

Our main results show that reprogramming potential is inherent to somatic cells prior to factor induction and that this potential is shared between a pair of lineages originating from the same pre-induced progenitor cell. These data suggest that reprogramming potential is set at least several divisions before induction and is heritable in the short term. Consistently, we do not find any effect of local signaling from neighboring lineages on the fate adopted by cells. Recently, an early stochastic phase was proposed to exist during reprogramming, based on high cell-to-cell variability in expression of specific genes in the same inducible system as the current study [14]. Our results suggest that this variability may stem from preexisting differences in cell states, rather than stochastic switching between different states after reprogramming is initiated [26]. Stochastic steps leading to potential loss or realization within each iPSC lineage may still occur at later divisions, but the likelihood of realization has to be high enough such that each pair member contributes to iPSC colonies within the allotted time window. The results imply that some predictive early marker could be identified and potentially used to isolate cells that will respond positively to factor induction. Our efforts to see whether Thy1 [21], a fibroblast specific marker heterogeneously present in the MEF population, can be used as such a marker indicate that Thy1 expression at day 0 is not predictive of lineage fate (Supplementary Fig S11). The stability of the transcriptional state may only be one marker for a cell’s response, while other relevant characteristics could include the presence of supportive (or the absence of deleterious) cofactors or the epigenetic configuration of target enhancer sequences. Genetic variants in the starting MEF population could also contribute to a cell’s potential to reprogram. However, efforts to characterize specific genetic variation within iPSC colonies generated from fibroblast pools have found only very rare instances of overrepresented polymorphisms [27], suggesting that genetic contributions to this process would fall below the overall frequency of ~1% that we observe for fibroblasts that successfully reprogram over our time course.

Our spatial dependence analysis shows no substantial contribution from the local niche, in terms of signaling from neighboring cells, to the final lineage fate. This suggests paired lineages do not adopt similar fates because of local ‘nurturing’ external effects but rather because of internal cell state. Additionally, the differential effect in colony formation between early and late replating during reprogramming suggests that from around day 6, cells are less prone to disturbance in their route to an iPSC fate, even though molecular markers associated with complete reprogramming have not yet been activated.

Effects of epigenetic perturbations on reprogramming efficiency have been demonstrated previously [3,22]. Here, we show that such a perturbation of H3K27 methylation through Ezh2 inhibition can change reprogramming potential by altering cellular state prior to factor induction. Furthermore, the shift in fates adopted by inhibitor treated cells suggests that the difference in potential may be chromatin-related, consistent with recent results where a major barrier to reprogramming is how transcription factors modify target chromatin once engaged [3]. Ezh2 inhibition increased the total number of iPSC-forming lineages at the expense of ‘fast dividers’, the other continuously proliferating response to factor induction, and not of senescence-prone ‘non-responding’ fates, contrary to other methods of improving reprogramming efficiency that act by altering the population of dividing cells only [12,18]. Paired lineages with mixed fates also arose solely under inhibitor treatment, but usually for cells that divided early during dosage, and as such spent more time under treatment as separate cells. This suggests an asymmetry either in the cumulative effect of Ezh2 inhibition on sister cells, or in the sisters’ internal states after division. All mixed-lineage pairs are between either iPSC- or FD-forming lineages and senescent non-responders, and never between alternate proliferating fates. As such, they may represent compounding effects between drug
toxicity and the oncogenic stress of OSKM induction. Finally, the effect of Ezh2 inhibition is likely not mediated through a change in OSKM levels, as these levels are unaltered by the inhibitor (Supplementary Fig S12).

Characterizing the molecular events that prescribe successful reprogramming is challenged by the low efficiency and extended latency of the process. As such, most studies have generally relied on inferences from static population sampling or via lineage tracing with a limited number of reporters. The methodology we present here seeks to address fundamental aspects of reprogramming lineages with minimal preconceived assumptions about the exact molecular mechanisms in play. This strategy may be used to track decision time points along other complex cellular lineages where little is known, such as in *in vitro* differentiation or cancer progression. Using live imaging and statistical analysis, we show that reprogramming potential in MEFs is preset and can be manipulated epigenetically. With a greater understanding of the key determinants through which reprogramming lineages are first established, future experiments may be designed to identify the underlying mechanisms that enable somatic cells to change fates in a directed fashion.

**Materials and Methods**

**Cell culture**

Secondary Nanog-GFP (NGFP2) MEFs derived from isolated E13.5 doxycycline-inducible murine fibroblasts as previously described [1] and cultured in ES cell medium, DMEM (Invitrogen) supplemented with 15% FBS, L-glutamine, penicillin-streptomycin, nonessential amino acids (Biological Industries), β-mercaptoethanol (Sigma), and 1,000 U/ml leukemia inhibitory factor (LIF, Millipore). All experiments were conducted after three passages from isolation. Collagen-OKSM-Oct4-EGFP MEFs [28] were grown in the same conditions as the NGFP2 MEFs.

**Reprogramming and image acquisition**

Nanog-GFP (YFP, or H2B-Cerulean labeled)-inducible MEFs were plated on gelatin-covered 24-well, 12-well, and 6-well plates (TC-treated polystyrene plate) at a density of 5,000, 10,000, and 20,000 cells per well, respectively. In 96-well plates, cells were seeded on feeder cells, either at low densities of 30–200 cells per well, or at high density, mixing 20–40 CFP-labeled cells with 1,000 YFP-labeled cells per well, as denoted in the text. Imaging started about 16 h after plating. Cells were cultured under serum starvation conditions (0.5% FBS) for ~16 h before switching into standard mouse ES medium supplemented with 2 μM doxycycline (Sigma) for all experiments to ensure all traced lineages began reprogramming from G1. Cells were kept on doxycycline for the duration of all imaging experiments. Growth medium (supplemented with dox) was replaced every 24–48 h. On day 12, the medium was switched to N2B27 + LIF + 2i + dox medium, containing neurobasal medium, DMEM/F12, B27, BSA (Invitrogen), Ndf (Millipore), 3 μM CHIR 99021 (Biovision), 1 μM PD0325901 (Santa Cruz). After 14 days, cells were fixed and immunostained against pluripotency markers. Inducible MEFs were imaged using a Nikon TiE epi-fluorescence microscope equipped with a motorized XY stage (Prior) and taken within a connected 6 × 6 or 7 × 7 spatial range at 10× magnification in up to three fluorescent wavelengths and phase contrast using NIS Elements software. Acquisitions were taken every 2–4 h for 14–18 days.

**Image analysis**

Tracking cell divisions during the first few generations, as well as tracking lineage dynamics at later generations, was done manually using ImageJ. Cell segmentation was done using CellProfiler [29]. Cell counting as well as final colonies counting was done automatically using CellProfiler and verified manually. In replating experiments, CFP and YFP colony identification was done using CellProfiler. iPSC-positive colonies were determined by coordinate comparison with the red channel, containing the Nanog staining data.

**Cell fate classification**

Cells were traced from 2 days before dox induction. All cells, at first division, if occurred during that 48 h period, were segmented as paired lineages. All lineages as well as cells that did not divide during that time were traced to their final fates. Final fate of a cell or a lineage was assigned to one of three categories: (1) non-responder (NR)—cells that did not divide at all after dox induction or divide slowly prior to death or senescence within 4 days of induction; (2) fast dividers (FD)—can also be referred to as partly reprogrammed cells)—cells that divide quickly after dox induction, creating a spread out lineage with morphological features of fibroblasts, and do not survive after switching into N2B27 medium (day 12) or do not stain out lineage with morphological features of fibroblasts, and do not survive after switching into N2B27 + LIF + 2i conditions at the end of the experimental time course. These colonies are confirmed by positive staining for E-cadherin and Nanog. A lineage is classified FD only if it yields no iPSC colonies; and (3) iPSC—cells that divide to form a condensed colony that resolves to a clear Nanog-GFP positive colony after switching into N2B27 + LIF + 2i conditions at the condition of the experimental time course. These colonies are confirmed by positive staining for E-cadherin and Nanog (for some experiments, alkaline phosphatase activity was also measured). A lineage is classified as iPSC if it yields any iPSC colony, even if some cells within the subsequent lineage are FD.

**Pre-treatment with chromatin modifiers inhibitors**

Ezh2 methyltransferase was inhibited with two different inhibitors: 3-deazaneplanocin A (DZNep, Sigma, 5 μM) which inhibits the expression of Ezh2, and JQ EZ005, an Ezh2 inhibitor that was kindly provided by J. Bradner. Each inhibitor was added to ES medium at concentration of 5 μM. Other chromatin modifiers tested included the LSD1 inhibitor RN-1 (Millipore, 0.01–1 μM), the DNA methyltransferase inhibitors RG108 (Cayman chemical, 10 μM), and 5-azacytidine (Sigma, 2 μM), the histone deacetylase inhibitor valproic acid (VPA, Sigma, 1 mM) and the G9a inhibitor BIX01294 (Stemgent, 1 μM). Effects of these inhibitors on reprogramming efficiency were measured against untreated and vehicle treated (0.1% DMSO, Sigma) controls. Cells were grown and plated for experiment as described above. During the 2 days before dox induction, cells were treated with either inhibitors for 2 different time periods (see Fig 4A). The concentration and duration of treatment was calibrated to identify conditions that maintain viability, enhance reprogramming, and aberrate their target epigenetic modification as determined by immunofluorescence or taken from the literature.
Experiments. Growth medium (supplemented with dox) was replaced for conditions (0.5% FBS) for about 16 h after plating. Cells were cultured under serum starvation well, or at high density, mixing 20 seeded on feeder cells, either at low densities of 30 plated on gelatin-coated 24-well, 12-well, and 6-well plates conditions as the NGFP2 MEFs.

Little is known, such as decision time points along other complex cellular lineages where molecular mechanisms in play. This strategy may be used to track lineages with minimal preconceived assumptions about the exact here seeks to address fundamental aspects of reprogramming on inferences from static population sampling or via lineage tracing.

Toxicity and the oncogenic stress of OSKM induction. Finally, the effect of Ezh2 inhibition is likely not mediated through a change in 5-azacytidine (Sigma, 2 l)

With a greater understanding of the key determinants through software. Acquisitions were taken every 2

Tracking cell divisions during the first few generations, as well as 3 spatial range at 10

Acquisitions were taken every 2

Image analysis

Additional experimental methods as well as description of colony counting are provided in the Supplementary Materials and Methods. Full description of the statistical methods employed is given in the Supplementary Information.

Acknowledgements

We thank J. Bradner for contributing an Ezh2 inhibitor and C. Sindhu for help with cell sorting. This study was supported in part by the Human Frontiers Scientific Program (HFSFP RG0080/2011), the Binational Science Foundation (BSF 2009403), and the Edmond J. Safra Center for Bioinformatics at Tel Aviv University.

Author contributions

MP, IP, ZDS, AM, and IN conceived or designed the experiments. MP and IP performed the experiments. MP and RR analyzed the data. MP, IP, ZDS, AM, and IN wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References


17. Luria SE, Delbrück M (1943) Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28: 491 – 511


Dynamics of genomic H3K27me3 domains and role of EZH2 during pancreatic endocrine specification

Cheng-Ran Xu1,2, Lin-Chen Li2, Greg Donahue1, Lei Ying3, Yu-Wei Zhang2, Paul Gadue3 & Kenneth S Zaret1,*

Abstract

Endoderm cells undergo sequential fate choices to generate insulin-secreting beta cells. Ezh2 of the PRC2 complex, which generates H3K27me3, modulates the transition from endoderm to pancreas progenitors, but the role of Ezh2 and H3K27me3 in the next transition to endocrine progenitors is unknown. We isolated endoderm cells, pancreas progenitors, and endocrine progenitors from different stages mouse embryos and analyzed H3K27me3 genome-wide. Unlike the decline in H3K27me3 domains reported during embryonic stem cell differentiation in vitro, we find that H3K27me3 domains increase in number during embryonic progenitor development in vivo. Genes that lose the H3K27me3 mark typically encode transcriptional regulators, including those for pro-endocrine fates, whereas genes that acquire the mark typically are involved in cell biology and morphogenesis. Deletion of Ezh2 at the pancreas progenitor stage enhanced the production of endoderm progenitors and beta cells. Inhibition of EZH2 in embryonic pancreas explants and in human embryonic stem cell cultures increased endocrine progenitors in vitro. Our studies reveal distinct dynamics in H3K27me3 targets in vivo and a means to modulate beta cell development from stem cells.

Keywords embryogenesis; endocrine; Ezh2; H3K27me3; pancreas

Subject Categories Chromatin, Epigenetics, Genomics & Functional Genomics; Development & Differentiation; Stem Cells

DOI 10.15252/embj.201488671 | Received 6 April 2014 | Revised 11 July 2014 | Accepted 16 July 2014 | Published online 8 August 2014

The EMBO Journal (2014) 33: 2157–2170

See also: HA Russ & M Hebrok (October 2014)

Introduction

The generation of insulin-secreting beta cells in the pancreas involves sequential cell fate choices. Pancreas progenitor cells, which initially express Pdx1, a pancreatic determination gene, arise from the definitive endoderm at the 7 somite pair (7S) stage of mouse embryogenesis, or embryonic day 8.5 (E8.5) (Jonsson et al., 1994; Offield et al., 1996) as an alternate fate from that of liver progenitors (Deutsch et al., 2001). The early differentiating PDX1+ cells are multipotent progenitors that can give rise to ductal cells, acinar cells, and endocrine progenitors (NGN3+) (Jorgensen et al., 2007). From about E13.5–E14.5, endocrine progenitors are specified and then differentiate into all five hormone-expressing endocrine lineages (α, β, δ, ε, and PP) which will comprise the pancreatic islets (Oliver-Krasinski & Stoffers, 2008; Zaret & Grompe, 2008; Pan & Wright, 2011). Due to technical limitations in working with the small cell numbers that can be harvested from embryos, progress in mapping chromatin transitions during beta cell development has been primarily from embryonic stem cell (ESC) cultures. The extent to which chromatin dynamics in vitro mirror those in vivo has yet to be explored. Since chromatin modifications are created by enzymes, and enzymes can be inhibited by small molecules, understanding chromatin dynamics can help control cell fates and thus enhance the generation of desired cell types, such as beta cells.

Progress in understanding chromatin states relevant to beta cell development includes the discovery that the H3K27me3 demethylases UTX (KDM6A) and JMJD3 (KDM6B) regulate endoderm differentiation from human ESCs by modulating the WNT signaling pathway (Jiang et al., 2013). Mouse ESC studies further showed that JMJD3 cooperates with the transcription factors Tbx3 and Eomes to promote endoderm induction (Kartikasari et al., 2013). We found a functional “prepattern” of chromatin states whereby regulatory elements of the Pdx1 gene, but not regulatory elements of liver genes, are marked by H3K27me3 in mouse embryonic endoderm, where all of these genes are silent and the cells are not yet committed to one fate or another (Xu et al., 2011). The histone H3K27 methyltransferase EZH2, which binds to the Pdx1 regulatory elements in endoderm, was found to modulate the pancreas versus liver fate choice by suppressing the pancreas lineage (Xu et al., 2011), consistent with a generally repressive role for PRC2 (Conely et al., 2011; Schwartz & Pirrotta, 2013).

In a human ESC model of pancreatic development, the global number of H3K27me3 peaks declined during in vitro differentiation to...
endoderm and pancreas progenitor stages [see Fig 3D of Xie et al (2013)], with transcriptional regulatory genes being among those losing the mark, over time. Whether a cumulative loss of H3K27me3 occurs globally in vivo is unknown. Another study of in vitro huESC differentiation to endoderm and posterior foregut progenitors, including pancreatic progenitors, observed a wide diversity of chromatin mark patterns that did not cohesively predict classes of enhancers as being prepatterned or common gene sets at each multipotent progenitor stage (Loh et al, 2014). Due to challenges in maturing huESC cultures to endoderm progenitors in vitro, this latter step was not explored.

A cross-tissue study found that genomic locations of the active H3K4me3 mark are similar between beta cells and neural tissue, whereas locations of H3K27me3 displayed a similar profile between beta cells and acinar cells, thereby better reflecting a shared developmental lineage (van Arensbergen et al, 2010). Notably, this study also showed that H3K4me3/H3K27me3 bivalency in pancreas progenitors at E10.5 did not predict competency of genes to be activated later in beta cells. Therefore, a primary focus on H3K27me3 dynamics seems most informative for endoderm development. A subsequent in vivo study showed that Ring1b, a PRC1 complex subunit, establishes repressed domains in pancreas progenitors but is not required to maintain them in insulin cells (van Arensbergen et al, 2013). Taken together, the above studies show that H3K27me3 dynamics are crucial to early pancreatic development, but roles at the endocrine induction step are unclear.

Prior studies of native embryonic cells compared ES chromatin profiles with that in pancreas progenitors and fully differentiated beta cells. Here, we present the first assessment of H3K27me3 dynamics in the step-by-step transitions between foregut endoderm cells, pancreas progenitor cells, and endocrine progenitor cells isolated from mouse embryos. Notably, the overall H3K27me3 peak dynamics in embryos differ from that observed with ESCs in vitro. With regard to gene networks that change in H3K27me3 coverage during the induction of pancreas progenitors into endocrine cells, H3K27me3 was acquired predominantly at genes involved in cell biological and morphogenetic changes. We genetically deleted Ezri2 during the pancreatic endocrine induction step in embryos and pharmacologically inhibited EZH2 in human ESC cultures in vitro and observed an increased yield of functional beta cell progenitors. These findings reveal gene networks specific to cells undergoing organogenesis in vivo and demonstrate how a detailed analysis of chromatin during native embryonic development provides insight that can be applied to stem cell differentiation.

Results

Net increase of H3K27me3 peaks during pancreas progenitor and endocrine progenitor specification in vivo

To obtain a genomic view of H3K27me3 locations during the transitions from endoderm to pancreas progenitors and from pancreas progenitors to endocrine progenitors, we used FACS to isolate cells and performed ChIP-Seq with an antibody for H3K27me3 (Supplementary Fig S1). For the pancreatic progenitor specification step, we isolated undifferentiated endoderm cells from E8.25 mouse embryos with the antibody ENDM1+ (Gadue et al, 2009; Xu et al, 2011) and Pdx1+/Liv2− pancreatic progenitor cells at E10.5 from Pdx1-GFP transgenic embryos (Supplementary Fig S2, Q3) (Gu et al, 2004). The selection of Liv2− cells eliminated about a fifth of the Pdx1−GFP− population that co-expresses the Liv2− hepatoblast surface antigen (Supplementary Fig S2, Q2) (Xu et al, 2011); these cells were not characterized further. For the endocrine specification step, we isolated Pdx1+/Liv2− cells at E10.5 and Ngn3−endocrine progenitor cells at E14.5 from Ngn3-GFP embryos (Lee et al, 2002; White et al, 2008) (Supplementary Fig S2). The experiments employed 5 × 104 to 105 cells per ChIP with an antibody against H3K27me3, pooled from about 100 E8.25, 50 E10.5 embryos, and 40 E14.5 pancreata (Fig 1A, Supplementary Fig S3, Supplementary Table S1). We mapped the H3K27me3 peaks in the genome at each stage (see Supplementary Methods, Supplementary

Figure 1. Increasing numbers of H3K27me3 peaks during pancreatic progenitor specification and endocrine specification.

A Scheme employed to use FACS to obtain ENDM1+endoderm progenitors (EN), Pdx1-GFP+/Liv2−pancreatic progenitors (PP), and Ngn3-GFP−endocrine progenitors (EP) for H3K27me3 ChIP-Seq.

B Top, schematic of mouse chromosome 9, red boxed region is expanded below, showing a local genomic view of 2 Mb of input-depleted H3K27me3 aligned sequence tags at the EN, PP, and EP stages. Note the clustering of tags over the gene-dense regions. Chromosomal coordinates (mm9) are shown above the tag display.

C Expanded view of the Hnf6/Onecut1 focus, showing a local diminution of sequence tags at the PP stage, when the Hnf6 gene is expressed (Jacquin et al, 2000).

D, E Venn diagrams depicting the unique and overlapping H3K27me3 aligned tag counts, illustrating a progressive increase in aligned tags at each of the pancreatic progenitor and endocrine progenitor transitions.
H3K27me3 gained and lost at distinct gene networks during pancreatic endocrine specification in vivo

To broadly assess the dynamics of the relevant gene networks marked by H3K27me3, we also mapped H3K27me3 targets for the alternate fates of ENDM1+ cells, that is, non-endoderm fate, and Liv2+ cells, that is, the hepatoblast (non-pancreatic) fate (Xu et al., 2011). This provided a collection of genes marked by H3K27me3 at either outcome of fate choices leading to endocrine cells. Of the H3K27me3 peaks from all of the above populations, 3,437 overlapping RefSeq genes in at least one tissue or stage examined (Fig 2A). We then created a heat map of the dynamics of H3K27me3 peaks at the genes during the EN, PP, and EP stages (Fig 2A). In agreement with the Venn diagrams in Fig 1D and E, the heat map in Fig 2A shows that while about 60% of the genes marked by H3K27me3 in endoderm remain marked into the endocrine progenitor stage, a large number of genes acquire H3K27me3 during the two underlying transitions. In addition, many genes positive in all three stages gained in H3K27me3 tag density by the endocrine progenitor stage (more red in Fig 2A, “EP”). This latter feature, that genes that retain H3K27me3 in development exhibit increased tag density over time, was also seen in the in vitro human ESC data [see Fig 3D of Xie et al (2013)].

We then examined the genes that lost H3K27me3 when pancreatic progenitors became Ngn3+ endocrine cells (115 genes, “++ – −”) or that gained H3K27me3 during the transition (598 genes, “− – +”), where the state of positive or negative for H3K27me3 had been stable for the previous endoderm to pancreas progenitor transition (Fig 2C). This focused the analysis on genes that underwent their first H3K27me3 transition at the endocrine step within the pancreatic endoderm lineage. Boxplots showing the distribution of fold-changes of gene expression scores for Ngn3+YFP+ versus Ngn3–YFP– pancreatic cells (E15.5) (Soyer et al, 2010) revealed that genes in the “++ – −” category showed a net increase in the overall expression level (Fig 2B, red box) while genes in the “− – +” category showed a net decrease in expression level (Fig 2B, green box), consistent with a repressing effect of the H3K27me3 mark. Genes that did not change their + or – status of H3K27me3 in these three populations exhibited no effect at the median (Fig 2B, “others”, gray). This analysis revealed an inverse correlation between our calls for H3K27me3 and gene expression during endocrine pancreas development, as expected for a repressive chromatin mark.

To understand the cellular networks that could explain the increase in H3K27me3 peaks seen in vivo, but not in vitro, we performed Gene Ontology (GO) analyses. Strikingly, the category of “− – +” typically encompassed genes related to cytoskeletal structure, membrane proteins, and cell adhesion (Fig 2C, “gained in EP”, Supplementary Table S2, Supplementary Dataset S1). It would therefore appear that the acquisition of H3K27me3 at the endocrine progenitor stage helps extinguish cell functions that are associated with cell biology and morphogenesis, features that are prominent during organogenesis in vivo and could be missing in cell culture.

By contrast, the category of “++ – −” predominantly encompassed genes related to transcriptional regulation (Fig 2D, “lost in EP”, Supplementary Table S3, Supplementary Dataset S1). This agrees with the data from human ESC differentiation in vitro, where loss of the H3K27me3 mark was seen extensively for transcriptional regulators (Xie et al., 2013). Notably, the “− – +” category, where the H3K27me3 mark is lost during endocrine cell induction, includes the transcription factor genes Ngn3, Nkx6.1, Nkx2.2, and Neurod1 (Supplementary Table S3; GO:0031018), which are necessary for the specification of endocrine progenitors and for establishing beta cell identity (Gradwohl et al, 2000; Gu et al, 2002; Schaffer et al, 2010, 2013; Papazian et al, 2011; Mastracci et al, 2013).

For example, inspection of the aligned sequence tags shows that H3K27me3 spans the Ngn3, Nkx2.2, and NeuroD1 genes at the endoderm and pancreatic progenitor stages, but is depleted during endocrine induction stage, when the genes are activated (Fig 3A). While Nkx6.1 is first activated in pancreatic progenitors and is needed for an endocrine versus acinar fate (Schaffer et al, 2010), recent studies indicate that Nkx6.1 also has a secondary role downstream of Ngn3 (Schaffer et al, 2013). In this context, we observe marked levels of H3K27me3 on Nkx6.1 in both endoderm and pancreatic progenitors, and then a loss of such in endocrine progenitors (Fig 3A). By contrast, the Hoxb5 gene is inactive and is blanketed by H3K27me3 at all stages tested and the Gapdh gene is expressed at all stages and lacks patches of H3K27me3. To validate these results, we performed ChIP-qPCR on the enhancer and promoter elements of Ngn3 and the promoters of Nkx2.2, NeuroD1, and Nkx6.1 (Fig 3; at location of red boxes in panels A, B), with the Hoxb5 promoter as a positive control and a Gapdh exon as a negative control (Fig 3). We found that in the endoderm and pancreatic progenitors, H3K27me3 was present at the Ngn3, Nkx2.2, NeuroD1, and...
genes overlapping/bound by H3K27me3 (red):

A  

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bound in EN only (+ - -)</td>
<td>54</td>
</tr>
<tr>
<td>Bound in EN and EP (+ - +)</td>
<td>278</td>
</tr>
<tr>
<td>Bound in PP only (- - -)</td>
<td>47</td>
</tr>
<tr>
<td>Bound in PP and EP (- - +)</td>
<td>634</td>
</tr>
<tr>
<td>Bound in EP only (- - -)</td>
<td>115</td>
</tr>
<tr>
<td>Bound in EN and PP (+ + -)</td>
<td>115</td>
</tr>
<tr>
<td>Bound in entire lineage (+ + +)</td>
<td>2,621</td>
</tr>
</tbody>
</table>

B  

<table>
<thead>
<tr>
<th>Gene Expression Fold Change</th>
<th>Ngn3+/Ngn3-</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4</td>
<td>+</td>
</tr>
<tr>
<td>-2</td>
<td>++</td>
</tr>
<tr>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>2</td>
<td>++</td>
</tr>
<tr>
<td>4</td>
<td>--</td>
</tr>
<tr>
<td>6</td>
<td>++</td>
</tr>
<tr>
<td>8</td>
<td>--</td>
</tr>
</tbody>
</table>

C  

<table>
<thead>
<tr>
<th>Gene Ontology Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0045941: pattern specification process</td>
<td>1.42E-03</td>
</tr>
<tr>
<td>GO:0007389: regulation of transcription</td>
<td>1.42E-03</td>
</tr>
<tr>
<td>GO:0006357: positive regulation macromolecule metabolism</td>
<td>3.20E-04</td>
</tr>
<tr>
<td>GO:0030528: regulation transcription DNA-dependent</td>
<td>3.79E-04</td>
</tr>
<tr>
<td>GO:0016055: Wnt receptor signaling</td>
<td>1.16E-04</td>
</tr>
<tr>
<td>GO:0032989: cell component morphology</td>
<td>1.16E-03</td>
</tr>
<tr>
<td>GO:0008289: lipid binding</td>
<td>1.16E-03</td>
</tr>
<tr>
<td>GO:0042995: cell projection</td>
<td>4.49E-04</td>
</tr>
<tr>
<td>GO:0043005: neuron projection</td>
<td>1.24E-04</td>
</tr>
<tr>
<td>GO:0005886: plasma membrane</td>
<td>3.05E-13</td>
</tr>
<tr>
<td>GO:0008092: cytoskeletal protein binding</td>
<td>9.49E-05</td>
</tr>
<tr>
<td>GO:0043005: neuron projection</td>
<td>1.24E-04</td>
</tr>
<tr>
<td>GO:0003779: actin binding</td>
<td>1.35E-04</td>
</tr>
<tr>
<td>GO:0044549: plasma membrane part</td>
<td>2.88E-04</td>
</tr>
<tr>
<td>GO:0045202: synapse</td>
<td>4.43E-04</td>
</tr>
<tr>
<td>GO:0042995: cell projection</td>
<td>4.49E-04</td>
</tr>
<tr>
<td>GO:0030054: cell junction</td>
<td>4.58E-04</td>
</tr>
<tr>
<td>GO:0008509: anion transport</td>
<td>8.87E-04</td>
</tr>
<tr>
<td>GO:0008289: lipid binding</td>
<td>1.16E-03</td>
</tr>
<tr>
<td>GO:0005516: calmodulin binding</td>
<td>1.16E-03</td>
</tr>
<tr>
<td>GO:0031225: anchored to membrane</td>
<td>1.16E-03</td>
</tr>
<tr>
<td>GO:0030695: GTPase regulator activity</td>
<td>1.16E-03</td>
</tr>
<tr>
<td>GO:0005085: guanyl-nuc. exchange</td>
<td>1.16E-03</td>
</tr>
<tr>
<td>GO:0060589: nucleoside-triphos. reg.</td>
<td>1.16E-03</td>
</tr>
<tr>
<td>GO:0042127: reg. cell proliferation</td>
<td>1.16E-03</td>
</tr>
<tr>
<td>GO:0006811: ion transport</td>
<td>1.16E-03</td>
</tr>
<tr>
<td>GO:0016055: Wnt receptor signaling</td>
<td>1.16E-03</td>
</tr>
<tr>
<td>GO:0005856: cytoskeleton</td>
<td>1.16E-03</td>
</tr>
<tr>
<td>GO:0044456: synapse part</td>
<td>1.16E-03</td>
</tr>
</tbody>
</table>

Figure 2. Dynamic patterns of H3K27me3 during pancreatic progenitor specification and endocrine specification.

A  Heat map indicating intensity of H3K27me3-bound genes (red, more tags per positive gene; black, called as negative) at the endoderm (EN), pancreas progenitor (PP), and endocrine progenitor (EP) stages. The number of genes in each sequential dynamic expression category is shown to the right of the heat map.

B  Boxplots with y-axis indicating gene expression fold change; that is, net increase in the overall RNA expression level of genes in the “++” category, that lose H3K27me3 at the endocrine induction step, while genes in the “−−” category, that gain H3K27me3, show a net decrease. Genes not changed in H3K27me3 + or − in these three populations exhibit no effect at the median.

C, D  Gene Ontology analysis of the category of “−−” (gain in H3K27me3 in Ngn3-GFP+ endocrine progenitor cells, panel C) and of the category of “++” (loss of H3K27me3 in Ngn3-GFP− endocrine progenitor cells, panel D). P-values are shown. See Supplementary Tables S2 and S3 for details.

Nkx6.1 regulatory sequences (Fig 3C, blue, red bars) and at the positive control site, but not at the negative control site. In Ngn3+ cells, H3K27me3 was depleted from the Ngn3, Nkx2.2, NeuroD1, and Nkx6.1 elements, consistent with the activation of the respective genes (Fig 3C, green bars). In summary, H3K27me3 is gained at differentiation and morphogenesis genes and is lost from key transcriptional regulatory genes in a stepwise manner during pancreatic endocrine development in vivo.
**Ezh2 suppresses the normal extent of endocrine cell induction**

Given the loss of H3K27me3 at key endocrine transcription factor genes during endocrine cell induction, we tested whether precocious loss of Ezh2, which catalyzes the H3K27me3 mark (Cao et al., 2002), would affect the timing or extent of endocrine induction. Since the Ezh2 null mouse is embryonic lethal during gastrulation (O’Carroll et al., 2001), we used an Ezh2 conditional allele (Ezh2^CA) (Su et al., 2003) and a Pax1-Cre transgene (Gu et al., 2002; Hingorani et al., 2003), the latter of which starts to express the Cre recombinase around E9.5, in an Ngn3-GFP background. This would cause precocious loss of Ezh2 prior to E12.5, when endocrine cell production begins. In sorted GFP^+ cells from E14.5 Ezh2^CA/Pax1-Cre; Ngn3-GFP pancreata, the expression level of Ezh2 is greatly diminished from that seen in wild-type or heterozygous floxed embryos (Fig 4A). In heterozygous and homozygous floxed Ezh2 embryos, the Ngn3 gene is modestly up-regulated in Ngn3-GFP^+ cells, possibly because precocious expression causes increased accumulation of the Ngn3 mRNA (Fig 4A). Importantly, H3K27me3 is greatly diminished in the Ezh2 knockout pancreas and undetectable in nascent endocrine progenitors expressing endogenous NGN3 (Fig 4B). In addition, Ngn3-GFP^+ cells displayed a higher fluorescence intensity in Ezh2 heterozygous and homozygous pancreas tissue (Fig 4C). Flow cytometry analysis provided a quantitative view of the three cell populations, Ngn3-GFP^{pro}, Ngn3-GFP^{pro} and Ngn3-GFP^{pro} in E14.5 pancreases. In Ezh2^{h/h} pancreases, Ngn3-GFP^{pro} cells were increased nearly threefold and Ezh2^{h/h} pancreases showed about a two-fold increase (Fig 4D and E). Together, these data show that Ezh2 depletion prior to the normal time of endocrine cell induction causes more endocrine progenitors to develop. Furthermore, the
The EMBO Journal

Dynamics of H3K27me3 in endocrine development

Cheng-Ran Xu et al

A

Ezh2
sorted Ngn3+ cells mRNA
Gapdh

B

Ezh2+/+; Pdx1-Cre Ezh2fl/fl; Pdx1-Cre

H3K27me3 NGN3

C

Ezh2+/+; Ngn3-GFP Ezh2fl/fl; Pdx1-Cre; Ngn3-GFP Ezh2fl/fl; Pdx1-Cre; Ngn3-GFP

D

83.2% 5.85% 8.04%
Ngn3-neg Ngn3-L Ngn3-H

74.8% 13.3% 8.46%
Ngn3-neg Ngn3-L Ngn3-H

73.0% 14.5% 6.75%
Ngn3-neg Ngn3-L Ngn3-H

E

Ngn3-neg Ngn3-L Ngn3-H

% cells/total Panc. cells

Ngn3 WT
rate of bromodeoxyuridine (BrdU) incorporation in the Ezh2fl/+ pancreas was indistinguishable from WT (Supplementary Fig S4). In prior studies (Jacquin et al., 2006), we showed that there is essentially no apoptosis in early pancreatic epithelial cells. Thus, the increase in Ngn3-GFP+ cells in Ezh2fl/+ pancreas was not due to increased cell proliferation or programmed cell death and therefore apparently by enhanced specification.

During pancreas development, multipotent progenitors undergo cell fate choices to differentiate into endocrine and exocrine pancreatic cells, the latter including acinar and duct cells. The exocrine compartment is highly branched, with the branches capped by acinar cells at the tips and connecting to duct trunks (Zhou et al., 2007). To address where the excess Ngn3-GFP+ endocrine cells arise in the Ezh2fl/+ pancreas, we performed immunohistochemistry on E14.5 pancreases for the acinar cell marker amylase or the trunk marker Sox9 (Seymour et al., 2007), combined with Ngn3 antibodies. In WT and Ezh2fl/+ pancreas, Ngn3+ cells were evident in the central trunk domains, as expected (Supplementary Fig S5, upper panels, blue cells). We did not observe any Ngn3+ cells in the peripheral, amylase+ tip areas (Supplementary Fig S5, upper panels, brown cells); Ngn3+ cells in Ezh2fl/+ were only evident in the Sox9+ duct trunk area (Supplementary Fig S5, lower panels). Taken together, the genetic data show that Ezh2 normally suppresses the extent of endocrine cell induction in the trunk cells of the developing pancreas. Ezh2 diminution at an earlier stage, due to Pdx1-Cre, increases the number of endocrine cells that develop.

Ezh2 loss at pancreas progenitor stage leads to increased beta cells

To assess the developmental consequence of an increase in Ngn3+ endocrine cells, we examined the mass of islets at postnatal day 9 (P9) by insulin immunohistochemistry. In the Pdx1-Cre;Ezh2fl/+ pancreas, we observed a 1.5-fold increase in the beta cell mass in the P9 animals, compared to wild type (Supplementary Fig S6A and B). These findings revealed that deleting Ezh2 at the pancreas progenitor stage enhances endocrine cell induction, allowing a greater number of beta cells to develop. Consistent with these data, in 2-month-old Pdx1-Cre;Ezh2fl/+ pancreases, the mass of islets shows a 1.34-fold increase compared to wild type (Fig 5A and B). However, surprisingly, in 2-month-old Pdx1-Cre;Ezh2fl/+ pancreases, the mass of islets was significantly reduced from WT (Fig 5A and B). Islet structure and beta cell morphologies appeared normal in the knockouts (data not shown).

Chen et al (2009) previously employed the Ezh2fl/+ model, but deleted Ezh2 later, at the beta cell stage, with a RIP-Cre. They found that the homozygous loss of Ezh2 after beta cells are formed induces Ink4a/Arf expression, which impairs beta cell proliferation and leads to mild diabetes (Chen et al., 2009). To test whether the loss of Ezh2 in the earlier stage, Pdx1-Cre model induces Ink4a/Arf in the islet, we isolated islets from about 2-month-old Pdx1-Cre;Ezh2fl/+ and Pdx1-Cre;Ezh2fl/+ male mice and their wild-type male littermates, and performed gene expression analysis. The expression of the Ezh2 gene could not be detected in Pdx1-Cre;Ezh2fl/+ islets, while the Ink4a/Arf gene, which was severely repressed in wild-type islets, was highly expressed in Pdx1-Cre;Ezh2fl/+ islets but not in Ezh2 heterozygotes (Fig 5C).  

Histone methyltransferase inhibitors modulate two stages of pancreas differentiation

Previously we showed that Pdx1 regulatory sequences contain H3K27me3 in undifferentiated ENDM1+ endoderm cells and that Ezh2 loss, via an endoderm Cre, causes an increase in the development of pancreatic progenitors at the expense of liver progenitors (Xu et al., 2011). To test small molecule inhibitors of EZH2 for their ability to mimic the Ezh2 conditional knockout at this stage, we set up half-embryo cultures that allow for gut tissue development in vitro (Wandzioch & Zaret, 2009). Briefly, “halves” of 2-75 embryos were taken anterior to the first somite and cultured for 48 h, allowing gut tube closure, the induction of early liver and pancreatic genes, and heart development (Wandzioch & Zaret, 2009). We cultured Pdx1-GFP half-embryos in the presence of 3-deazaneplanocin A (DZNep), which inhibits the methyl donor pathway for EZH2 (Fiskus et al., 2009), and the EZH2-specific inhibitor GSK-126 (McCabe et al., 2012). Notably, we observed a significant increase in the number of Pdx1-GFP+ cells in DZNep- and GSK-126-treated half-embryos (Supplementary Fig S7A and B). These data show that inhibitors of EZH2 can regulate a step of pancreatic differentiation similar to that seen in prior genetic studies with Ezh2fl/+ and an endoderm Cre (Xu et al., 2011).
A Immunohistochemistry for insulin (brown) shows beta cell area increases in different times and gave slightly different peak glucose levels for the different cohorts of WT animals, but the glucose levels were consistent within each experiment.

Pancreas, compared to WT counterparts. Sections were measured for their beta cell area over total pancreas area; (**Ezh expression) or (**Ink4a/Arf). Pdx1-Cre does not have a significant effect on the evA-Ezh expression) islets; nd means not detected.

B Ezh2 normal Ezh2fl/+; Pdx1-Cre Ezh2fl/fl; Pdx1-Cre

C Islet mRNA / Gapdh

<table>
<thead>
<tr>
<th></th>
<th>Ezh2</th>
<th>Insulin1</th>
<th>Ezh2</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>normal</td>
<td>fl/+</td>
<td>fl/fl</td>
<td>fl/fl</td>
</tr>
<tr>
<td>n=</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

D Glucose tolerance test in intermiate sets of WT and Pdx-Cre;Ezh2fl/fl male mice (n = 4 each) and WT and Pdx-Cre;Ezh2fl/fl mice (n = 5, 6 respectively) were analyzed at different times and gave slightly different peak glucose levels for the different cohorts of WT animals, but the glucose levels were consistent within each experiment (see error bars). Asterisks denote P < 0.05. Ezh2 heterozygous KO mice suppressed glucose levels more effectively than WT whereas homozygous KO mice were less effective than WT.

Figure 5. Beta cell mass expands in adult Ezh2 heterozygous knockout pancreas, but reduces in Ezh2 homozygous knockout pancreas.

A Immunohistochemistry for insulin (brown) shows beta cell area increases in 2-month-old male Ezh2 heterozygous KO pancreas, while decreases in homozygous KO pancreas, compared to WT counterparts.

B Statistics of the percentage of beta cell area to total pancreas area. Intermittent sections in 100 μM intervals were analyzed for each pancreas, and 2–3 middle sections were measured for their beta cell area over total pancreas area; n = number of animals. Asterisks denote P < 0.05.

C Ezh2 and Ink4a/Arf mRNA levels in isolated islet cells as analyzed by qR1-PCR, normalized to Gapdh within each sample and relative to levels in Ezh2 normal islets (Ezh2 expression) or Pdx-Cre;Ezh2fl/fl (Ink4a/Arf expression) islets; nd means not detected.

D Glucose tolerance test in intermediate sets of WT and Pdx-Cre;Ezh2fl/fl male mice (n = 4 each) and WT and Pdx-Cre;Ezh2fl/fl mice (n = 5, 6 respectively) were analyzed at different times and gave slightly different peak glucose levels for the different cohorts of WT animals, but the glucose levels were consistent within each experiment (see error bars). Asterisks denote P < 0.05. Ezh2 heterozygous KO mice suppressed glucose levels more effectively than WT whereas homozygous KO mice were less effective than WT.
Given this validation of the action of the EZH2 inhibitors in explants, we sought to determine whether treatment with DZNep can modulate the later step of endocrine progenitor induction. We dissected dorsal pancreatic tissue from E12.5 Ngn3-GFP embryos and treated cultured explants with DZNep for 4 days. The DZNep-treated Ngn3-GFP pancreas explants showed a higher GFP intensity compared to non-treated controls (Fig 6A). A 1.8-fold increase in the percentage of GFP+ cells in DZNep-treated explants, without a decrease in GFP− cells, was confirmed by flow cytometry analysis (Fig 6B). Hence, both the genetic and pharmacologic data indicate that EZH2 normally restrains the pancreatic progenitor and the endocrine progenitor induction steps of beta cell development.

**Histone methyltransferase inhibitors enhance the induction of beta-like cells in vitro**

We next tested EZH2 inhibitors with a human endodermal progenitor cell differentiation system in vitro. The endodermal progenitor cells are a recently described, pan-endodermal stem cell population derived from human ES or induced pluripotent stem (iPS) cells that can efficiently generate mono-hormonal pancreatic beta-like cells in vitro (Cheng et al., 2012). Differentiation of endodermal progenitor cells follows a stepwise developmental progression that mimics endocrine cell development in vivo, including pancreatic induction, endocrine specification, and subsequent expression of insulin (Fig 7A).

To determine whether methyltransferase inhibition could enhance endocrine differentiation of human cells in vitro, we treated endodermal progenitor cell cultures at distinct stages from differentiation initiation to endocrine specification. We used doses of DZNep and GSK-126 that could decrease H3K27me3 enrichment at the NEUROG3 promoter in undifferentiated human EP cells (Supplementary Fig S8). While treatment with DZNep early in the differentiation cultures had slight inhibitory effect on the later generation of C-peptide cells, treatment at days 8–10, during endocrine specification, led to an approximate doubling of the percentage of C-peptide cells in the end stage cultures without inducing the generation of poly-hormonal insulin/glucagon+ cells, as determined by intracellular flow cytometry (Fig 7B and D). The increase in C-peptide cells correlated with an increase in NGN3 mRNA expression at day 10 of the differentiation culture (Fig 7C). We found that treatment with GSK-126 had effects similar to DZNep both in NGN3 and in C-peptide induction (Fig 7C and D). Similar results were found with differentiation of an endoderm progenitor cell line derived from iPS cells (data not shown). These findings demonstrate that the positive effects of manipulating EZH2 function on pancreatic beta cell induction in mouse models can be applied to human stem cell differentiation cultures.

**Discussion**

The stepwise development of pancreatic beta cells is regulated by cell signaling, transcription factors, and chromatin modifiers. As noted in the Introduction, abundant evidence demonstrates a general role for the Polycomb complexes (PRCs) in lineage commitment (Lee et al., 2006; Surface et al., 2010) in general, and in endoderm and pancreatic progenitor development in particular. In addition, Ezh2 is necessary in the adult to maintain the physiological function of beta cells (Chen et al., 2009). Yet there was a gap in
performed H3K27me3 ChIP-Seq in cells isolated by FACS from fate transitions from endoderm progenitors to pancreatic progenitors, in human embryonic stem cell cultures. 

DZNep:

FACS analysis of cells expressing C-peptide versus glucagon on day 126 of differentiation in cultures treated at days 12-18 of differentiation when treated with DZNep at the times indicated by red bars in (A), expressed as mean ± SEM (n = 10) [***P-value < 0.001].

NGN3 mRNA

FACS analysis of cells expressing C-peptide versus glucagon on day 12 of differentiation in cultures treated at days 8-10 with DZNep or GSK-126.

NGN3+ 

INS+ 

PDX1+

D0-D2 D4-D6 D8-D10

DZNep:

control 

DZNep 

GSK-126

Figure 7. Histone methyltransferase inhibitors enhance pancreatic beta-like cell development from human endoderm progenitor cells.

A Experimental scheme of beta-like cell differentiation, showing the times of DZNep treatment (red bars).

B The % of C-peptide + endoderm cells at days 16-18 of differentiation when treated with DZNep at the times indicated by red bars in (A), expressed as mean ± SEM (n = 10) [***P-value < 0.001].

C Ngn3 mRNA was quantified by qRT-PCR at day 10 of differentiation, comparing control DZNep- and GSK-126-treated cultures at days 8-10, expressed as mean ± SEM (n = 4) [P-value < 0.05].

D FACS analysis of cells expressing C-peptide versus glucagon on day 18 of differentiation in cultures treated at days 8-10 with DZNep or GSK-126.

understanding PRC function in endocrine induction. Analogous to our earlier finding that Ezh2, encoding the enzyme for H3K27me3 modification, restrains the differentiation of endodermal cells into pancreatic lineage (Xu et al, 2011), we now find that Ezh2 restrains the induction of the later, endocrine commitment step. The existence of inhibitors of Ezh2 allowed us to employ such to enhance endocrine and beta cell induction in embryonic tissue explants and in human embryonic stem cell cultures.

To gain a dynamic view of H3K27me3 modification during the cell fate transitions from endoderm progenitors to pancreatic progenitors, and from pancreatic progenitors to endocrine progenitors, we performed H3K27me3 Chip-Seq in cells isolated by FACS from mouse embryos. Prior studies had mapped H3K27me3 specifically at promoters in pancreatic progenitors (van Arensbergen et al, 2010), but genome-wide assessments in such cells and in native endoderm and endocrine progenitors had not been performed. Three features of our dataset stand out, when compared to the in vitro ESC differentiation studies of (Xie et al, 2013). First, both in vitro and in vivo, genes that retain H3K27me3 throughout the developmental stages tested exhibit an increase in tag intensity over time. While the functionality of such changes remains to be determined, we suggest that it may reflect enhanced commitment to early-established lineage decisions. Second, both in vitro and in vivo, loss of H3K27me3 at each developmental stage frequently occurs at genes involved in transcriptional control. This would be consistent with a need to activate tissue-specific regulators that are silent in early development. The third feature reveals a disparity; we saw a clear increase in the number of H3K27me3 peaks, or patches, in the genome during the progression to the endocrine state, which contrasts markedly with a diminution of H3K27me3 peaks seen during the progression of embryonic stem cell cultures differentiated in vitro (Xie et al, 2013).

What could serve as the basis for the much greater number of genes gaining the H3K27me3 mark in embryonic development than in cell culture? It was striking that at the transition to endocrine progenitors, the genes gaining H3K27me3 largely fell into the categories of cell biology and morphogenesis. Since these functions are among the major determinants of embryonic organogenesis and are greatly minimized in cell culture, if they occur at all, we suggest that the deposition of H3K27me3 across the genome could be governed substantially by signaling that occurs during morphogenesis in vivo and that is largely absent in vitro.

The low cell number Chip-Seq method from sorted mouse embryo cells currently works best, in our laboratory, for chromatin marks such as H3K27me3 that exist in patches and thus are relatively easy to map to peaks. The disparity between our in vitro results and those reported from in vitro studies for H3K27me3 indicates that it is worth the effort to advance low cell number technologies further, in order to assess other chromatin modifications in vivo.

Releasing H3K27me3 from the key endocrine and beta cell transcriptional factor genes of Ngn3, Ink4a, NeuroD1, and neuroD1 is a prominent feature of the transition from pancreatic progenitors to endocrine progenitors. Concordantly, we found that precocious diminution of Ezh2 in pancreatic progenitor cells led to an increase in endocrine progenitors. Eviction of the Polycomb complex from key target genes was recently found to promote the developmental timing of gene expression in Arabidopsis (Sun et al, 2014). The mechanism of how the Ezh2 targets the development-specific genes remains elusive. The broad H3K27me3 patches we observed made it unproductive to screen for correlating DNA sequence motifs.

Upon precociously deleting the Ezh2 gene in pancreatic progenitors, we found that the expression of Ngn3 is elevated and the number of endocrine progenitors increased to a similar extent in E14.5 Ezh2/−/− and Ezh2H10/−/− pancreas. This suggests that the precise expression level of Ezh2 is critical to restrain endocrine pancreas development. Since the increased Ngn3+ endocrine progenitors occurred in the trunk domains, it appears that Ezh2 normally restrains endocrine induction in the natural trunk population, as opposed to a putative role in the exocrine compartment. Also, the lack of change in proliferation in the Ezh2H10/− pancreas is consistent with a cell fate control change instead of a putative non-specific increase in all cells of the organ.
At the adult stage, to our surprise, the mice with Ezh2<sup>fl/fl</sup> had enhanced glucose tolerance compared to their wild-type littermate, while the mice with Ezh2<sup>Wt/WT</sup> pancreas showed mild diabetes. We suggest that the gain in beta cell numbers, in the Ezh2<sup>Wt/WT</sup> pancreas, without Ink4a/Arf induction, is sufficient to enhance islet function, whereas in the Ezh2<sup>fl/fl</sup> pancreas, aberrant Ink4a/Arf induction results in impaired islet function. Thus our study is consistent with previous work (Chen et al, 2009) indicating that the proper amount of Ezh2 is necessary to maintain beta cell function in the adult pancreas. Based on these findings, we propose that during pancreatic development, the expression of the Ezh2 gene itself and/or Ezh2 enzymatic activity are under strict control.

Given the role Ezh2 plays at the endocrine induction step in vivo, we used small molecule inhibitors of Ezh2 on human embryonic stem cell cultures in vitro and discovered an enhanced yield of insulin-expressing cells. Since Ezh2 is also necessary to maintain beta cell function (Chen et al, 2009), Ezh2 inhibitors should be used in a short window prior to endocrine progenitor induction and then removed, to enhance the yield of beta cells. Our studies show that understanding mechanisms that underlie beta cell development in vivo can reveal processes that are not presently recapitulated in vitro, but yet can be used to enhance the in vitro approach.

Materials and Methods

Mouse strains and explant cultures

Pdx1-GFP (Gu et al, 2004), Ngn3-GFP (Lee et al, 2002; White et al, 2008), Ezh2<sup>Wt/WT</sup> (Su et al, 2003), and Pdx1-Cre (Hingorani et al, 2003) mice were used. Mice deleted for Ezh2 in pancreatic progenitor cells were generated by mating Pdx1-Cre<sup>+/−</sup>;Ezh2<sup>Wt/WT</sup> mice to one another.

For half-embryo cultures, ES.25 (2–7 somite pair) embryos were dissected, the posterior half was removed from the first somite site, and the anterior half was cultured at 37°C for 48 h in Dulbecco’s modified Eagle’s medium (DMEM) containing 10% calf serum (HyClone) (Wandzinski & Zaret, 2009). For pancreas explant cultures, E14.5 pancreata were dissected and cultured on Nucleopore Track-Etch Membranes (Whatman, 110414) floating on DMEM containing 10% FBS for 4 days (Metzger et al, 2012).

FACS analysis, cell sorting, RNA analysis

Endoderm cells were purified by FACS with the END1 and Liv2 antibodies as described (Xu et al, 2011) (Supplementary Fig S2). Embryos or pancreas tissue at later developmental stages was dissected from Pdx1-GFP or Ngn3-GFP strains, dissociated with trypsin for 5 min at 37°C then stopped with 0.4 volumes of FBS. The cells were filtered through a nylon mesh. GFP<sup>−</sup> cells were then analyzed or purified on a BD FACSDiva cell sorter.

Human beta-like cells were harvested at day 18 of differentiation and trypsinized into single cells. Cells were fixed with 1.6% paraformaldehyde and stained with an anti-C-peptide antibody (Cell Signaling) and anti-GG antibody (Sigma). Secondary antibodies were goat-anti-mouse-IgG1-PE (Jackson Immunoresearch) and goat-anti-rabbit-IgG-647 (Invitrogen). The stained cells were analyzed with the BD FACS Canto II Flow Cytometer (BD Biosciences), and the data were analyzed using Flowjo software (TreeStar).

Total RNA was isolated from cells isolated by FACS with the RNeasy Micro Kit (Qiagen) and was reverse transcribed with the iScript cDNA Synthesis kit (Bio-Rad). Real-time PCR was performed on the Power SYBR Green Supermix (Applied Biosystems) on the Applied Biosystems StepOnePlus. For RT-qPCR, primers were: Gapdh ATGTGGAAGCTGGTGTCGAC; GCTGTGACCTCACCCATAGATGAC; AAGAGCGCAAGTCACCTTCCCT; Ezh2 CTGCTTGCAACCTCTGA; GGTCCATCCACCAACAA. For human stem cell RT-qPCR total RNA was extracted with the RNAeasy Micro kit (Qiagen). RNA was reversely transcribed into cDNA using random hexamers with Superscript III Reverse Transcriptase (Invitrogen). qPCR was performed on the LightCycler 480 II (Roche) using LightCycler 480 SYBR Green I Master Mix (Roche). Expression levels were normalized to the housekeeping gene PPIG (Cyclophilin G). Human primers were: PPIG GAAAGTGCGGAT CAAGAACCATGAC; GTCTCTCTCCTCTCTCTCTCAT; TTTACT; NGN3 TGGATGCAAACACTCAAAC; AAGCCTGTGCGGGCAT.

Low cell number ChIP-qPCR and ChIP-Seq

H3K27me3 ChIP-qPCR using about 10,000 embryonic cells was as described (Xu et al, 2011). One μl of antibody against H3K27me3 (Millipore 07-449) was added per ChIP assay. The Western blot in Supplementary Fig S1 demonstrates the quality of the H3K27me3 antibody. All ChIP-qPCR data are from at least triplicate assays, normalized to input DNA. Primer sets: HoxB5 ACCGACTGGTCACAAAGAGCA; CCGATGCCACTCTACTCTGT; Gapdh exon TCACCCGCCACATCAAGCGC; TCACCAAAATTCACCGACC; Ngn3 enh TCCGCTTATGACCGAGCTGAT; TGTTGAAAGCGGGATGTTT; Ngn3 pro GAGAGTTGTGGCTGGTACGC; GGGACACGACTGCTTTGTC; Nkx6.1 pro GAGCCCGCTTAAACTGATAT; TCGCTTCTCTGCTTCTCTG; Nkx6.2 pro CACCTCTGCTCTAGCGCCCCTCAGGCT; ATTGTCAGATGTTAATTGTCG; NeuroD1 pro GTCCGGCGAGTCTCTAACAT; GAAACCTGTGACCTCGGAT.

For ChIP-Seq, we used chromatin from 5 × 104 to 1 × 105 sorted cells, pooled from many embryos harvested on multiple dates (see main text). Five μl of H3K27me3 antibody was used with the Magnify<sup>™</sup> Chromatin Immunoprecipitation System (Invitrogen 49-204). The preparation of multiplex libraries for sequencing (Supplementary Fig S3) was as per the “NENext ChIP-Seq Library Prep Set for SOLID” (NEB E6260S). Sequencing was performed on the SOLiD4.0 platform with input and H3K27me3-immunoprecipitated libraries. See Supplementary Methods for a detailed low cell number ChIP-Seq protocol.

ChIP-Seq data analysis

Alignment

Colorspace sequence tags were aligned to the mouse genome (assembly mm8, NCBI 36) using Bowtie v0.12.5. Up to three seed mismatches were allowed and 10 bp were trimmed from the 3′ end of each sequence tag prior to alignment. See Supplementary Table S1 for a summary of total and aligned reads for each sample.

Peak assessment

H3K27me3 peaks were called using the sliding window approach used to identify regions of H3K36me3 enrichment in (Guttman et al, 2009). Briefly, a 500-bp sliding window is scored for aligned tags
genome-wide; then, each window is assessed for its likelihood of enrichment given a Poisson model of the genomic background. A Bonferroni correction was applied to remove false discoveries (the effective alpha is 0.05 after correction), and overlapping peaks (at least 1 bp) were merged. Additionally, peaks were called in input using the same criteria. H3K27me3 peaks having 50% or more overlap to input peaks were discarded. The remaining peaks were then joined if the gap between them was less than 1 kb, and regions under 1 kb were discarded. A unified “super” peak set was constructed by pooling and merging overlapping peaks from each tissue (at least 1 bp overlap); then, an H3K27me3-input area under the curve (AUC) measurement was assessed for each unified peak in each tissue. In a given tissue, if a peak region had an AUC > 45, it was said to be a true enrichment. Each ReSeq gene-transcript was then assessed for overlap with true enrichments in each tissue; if any overlap could be detected the transcript was said to have H3K27me3 in that tissue. A detailed summary and analysis of the H3K27me3 target genes is presented in the Supplementary Dataset S1 in an Excel spreadsheet with multiple tabs.

Heat map
Each transcript therefore has an array of three values, one each for endoderm, pancreas progenitors, and endocrine cells, with “true” meaning “has H3K27me3 in this tissue” and “false” meaning “lacking H3K27me3 in this tissue”. To map H3K27me3 dynamics, transcripts with “true” or “false” in all tissues were filtered out. The remaining transcripts were recorded with their H3K27me3 profile and annotation information. The 90th percentile value for H3K27me3 AUC enrichment was set to the “maximum red” for the heat map.

Gene ontology
We assessed the genes that gained H3K27me3 upon the pancreas progenitor-to-endocrine transition and genes that lost the mark upon the transition, where the genes lacked or possessed H3K27me3, respectively, in the prior endoderm state. This focused the analysis on genes that underwent their first H3K27me3 transition at the endocrine step within the pancreatic endoderm lineage. GO analysis was performed on the two gene groups. Details of the GO analysis are presented in the Supplementary Dataset S1 in an Excel spreadsheet with multiple tabs.

Data sharing

Immunohistochemistry
We used a citrate-based antigen unmasking solution (Vector Labs) and performed immunohistochemistry with the ABC Kit (Vector Labs) (Xu et al., 2011). We also used antibodies against NGN3 (BCBC, ab2013), Sox9 (Millipore AB5535), amylase (Santa Cruz sc-46657), and biotinylated anti-mouse IgG or anti-rabbit IgG (Vector Labs).

Glucose tolerance analysis
Sixteen hours overnight fasted mice were injected intraperitoneally with glucose (2 mg/g body weight), and blood glucose levels were measured at 0, 15, 30, 60, 90 and 120 min after injection. The investigator was blinded as to the identity of the samples during the glucose assays. Littermate sets of WT and Pdx-Cre;Ezh2fl/fl mice (n = 4 each) and WT and Pdx-Cre;Ezh2flo/flo mice (n = 5, 6 respectively) were analyzed at different times and gave slightly different peak glucose levels for the different cohorts of WT animals, but the glucose levels were consistent within each experiment (see error bars in Fig 5D).

Human ES cell culture, differentiation, drug treatments, ChIP-qPCR
The human EP cell line used was from derived the H9 human ESC line (WiCell) as described previously (Cheng et al., 2012). EP cells maintenance and pancreatic differentiation were performed as described previously (Cheng et al., 2012). Differentiation cultures were treated with D2Nep (0.1 μM) or GSK-126 (2 μM) for 2-day periods as indicated. Five million EP cells were used for each ChIP assay with 5 μl of anti-H3K27me3 (Millipore 07-449). Primers used for the human NEUROG3 promoter are GTGAGAAGATGCACACATCACAAACAAAG and AGCAGTCTGCCACCATAGTGGA.

Supplementary information for this article is available online: http://embj.embopress.org

Acknowledgements
We thank Tianying Jiang for advice and assistance with histology, David Metzger and Angela Hines for advice on pancreas culture and the glucose tolerance test, Dario Nicetico for comments on the manuscript, members of the Zaret laboratory for advice and comments, and Eileen Huyle for help with the manuscript. We especially thank Bing Zhang at Beijing Genomics Institute for performing sequencing. This work was supported by NIH grants K01DK093886 to C-RX, R01DK092113 and U01DK072473 to PG, and R37GM03477 and U10DK072503 to KSZ

Author contributions
C-RX, GD and KSZ designed the research; C-RX, L-CL, LY and Y-WZ performed the research; C-RX, L-CL, GD, LY, PG and KSZ analyzed the data; and C-RX, PG and KSZ wrote the paper.

Conflict of interest
The authors declare that they have no conflict of interest.

References


Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes

Xu Li¹,f, Wenqi Wang¹, Jiadong Wang¹, Anna Malovannaya², Yuanxin Xi²,³, Wei Li²,³, Rudy Guerra⁴, David H Hawke⁵, Jun Qin² & Junjie Chen¹*

Abstract

The current knowledge on how transcription factors (TFs), the ultimate targets and executors of cellular signalling pathways, are regulated by protein–protein interactions remains limited. Here, we performed proteomics analyses of soluble and chromatin-associated complexes of 56 TFs, including the targets of many signalling pathways involved in development and cancer, and 37 members of the Forkhead box (FOX) TF family. Using tandem affinity purification followed by mass spectrometry (TAP/MS), we performed 214 purifications and identified 2,156 high-confident protein–protein interactions. We found that most TFs form very distinct protein complexes on and off chromatin. Using this data set, we categorized the transcription-related or unrelated regulators for general or specific TFs. Our study offers a valuable resource of protein–protein interaction networks for a large number of TFs and underscores the general principle that TFs form distinct location-specific protein complexes that are associated with the different regulation and diverse functions of these TFs.

Keywords: forkhead box; mass spectrometry; protein–protein interaction; transcriptional factor

Subject Categories: Network Biology; Transcription; Post-translational Modifications; Proteolysis & Proteomics

DOI: 10.3389/msb.2014.00504 | Received 18 June 2014 | Revised 25 November 2014 | Accepted 8 December 2014


See also: Z Ji & AD Sharrocks (January 2015)

Introduction

Over the years, considerable effort has been devoted to understand the signalling pathways, the basis of biological activities in all living organisms. Sophisticated signal transduction pathways are required for the development and survival of any organism, a minor disruption of which may cause developmental defects and diseases such as cancer (Fig 1A). The examples of these highly conserved signalling pathways include the Wnt (MacDonald et al, 2009), TGF-β (Massague, 1998) and NF-κB (Hayden & Ghosh, 2004) pathways. Many of these pathways function by ultimately regulating the activity of certain transcription factors (TFs), often by changing their localizations. Reports on individual proteins suggested that the chromatin association of TFs is tightly controlled by upstream signals. For example, NF-κB is known to translocate from the cytoplasm to the nucleus upon activation, which is a critical step coupling extracellular stimuli with transcriptional activation (Baldwin, 1996).

TFs are known to be rigorously regulated via their associations with other proteins (Blackwood & Eisenman, 1991). However, while the DNA-binding and the transcriptional activities of TFs on chromatin have been extensively studied, our knowledge of protein–protein interactions (PPIs) that may occur off the chromatin, which are important for the regulations and functions of these TFs, is very limited. Knowing what proteins TFs interact with and, especially, where they interact will greatly improve our understanding of how the activities of these TFs are controlled.

One example is the Forkhead box (FOX) family of TFs, which has been relatively well studied in their regulations of transcriptional activities, but little is known about protein–protein interactions involving these TFs. The term “Forkhead” was derived from a mutant Drosophila melanogaster that has a forklike head (Weigel et al, 1989). They have been classified into 19 subfamilies on the basis of the conservation of their DNA-binding domains (Kaestner et al, 2000). However, some have since been found to be variants of other family members. To date, there are 40 experimentally confirmed FOX family members and additional six or more FOX-like proteins in humans. Several in vitro studies have identified the consensus DNA sequences of a few FOX proteins, including those of FOXA, FOXD, FOXO, FOXP and FOXM, and their target genes using...
Molecular Systems Biology

Chromatin-associated and soluble human TF complexomes

A Disease correlation of 19 TFs and 4 well-studied FOX family members, based on their GO annotations. Each colour indicates one disease. The size of each coloured pie indicates the relative ratio of -log (P-value) of GO annotations in the corresponding disease.

B Pathway correlation and structural superfamilies of TFs. Each coloured area indicates one superfamily.

C Schematic diagram showing the major steps involved in TAP/MS screening and data analysis of human TFs and snapshot for each part of the data. Fifty-six transcription factors, together with their genomic localizations, were constructed into a vector harbouring a C-terminal SFB-tag through gateway technology. HEK293-T-Rex-TF-S-SBP-FLAG cells were transiently transfected with these vectors and puromycin selected. Proteomic analysis followed by puromycin selection. We picked 12 TFs: FOX1, FOX2, FOX3, FOX4, FOX5, FOX6, FOX7, FOX8, FOX9, FOX10, FOX11, and FOX12. These factors covered five structural TF superfamilies to exclude the potential bias caused by the initial proteomics study of the FOX family of TFs revealed that they may not be a total surprise, but it had never been systematically investigated. Indeed, our study confirmed that TFs indeed form distinct complexes on and off chromatin. This finding is not restricted to FOX family TFs, but it rather represents a general phenomenon of TFs, which is likely important for their differential regulations and diverse functions.

Figure 1. Proteomic analysis of human transcription factors.

- Disease correlation of 19 TFs and 4 well-studied FOX family members, based on their GO annotations. Each colour indicates one disease. The size of each coloured pie indicates the relative ratio of -log (P-value) of GO annotations in the corresponding disease.
- Pathway correlation and structural superfamilies of TFs. Each coloured area indicates one superfamily.
- Schematic diagram showing the major steps involved in TAP/MS screening and data analysis of human TFs and snapshot for each part of the data. Fifty-six transcription factors, together with their genomic localizations, were constructed into a vector harbouring a C-terminal SFB-tag through gateway technology. HEK293-T-Rex-TF-S-SBP-FLAG cells were transiently transfected with these vectors and puromycin selected. Proteomic analysis followed by puromycin selection. We picked 12 TFs: FOX1, FOX2, FOX3, FOX4, FOX5, FOX6, FOX7, FOX8, FOX9, FOX10, FOX11, and FOX12. These factors covered five structural TF superfamilies to exclude the potential bias caused by the initial proteomics study of the FOX family of TFs revealed that they may not be a total surprise, but it had never been systematically investigated. Indeed, our study confirmed that TFs indeed form distinct complexes on and off chromatin. This finding is not restricted to FOX family TFs, but it rather represents a general phenomenon of TFs, which is likely important for their differential regulations and diverse functions.
microarrays and ChIP-sequencings (Jolma et al., 2013). For example, the consensus sequence for FOXA1 is 5′-(G/A)(T/C)(A/C)AA(C/T)A-3′ (Georges et al., 2010), and FOXA1 plays a pivotal role in ER transcription activities (Hurtado et al., 2011).

FOX family TFs play important roles in regulating the expression of genes involved in a sundry of cellular processes, especially during development and tumorigenesis (Benayoun et al., 2011; Lam et al., 2013). Many FOX family members have been reported to be involved in cell, tissue and organ developments. For example, FOXA family members are required for normal development of liver, pancreas, lungs and prostate. Mutations or deletions of FOX genes often lead to developmental defects (Tuteja & Kaestner, 2007a,b), including severe organ and immune defects, premature ovarian failure, mental retardation, autism and speech disorders (Carlsson & Mahlapuu, 2002; Lehmann et al., 2003; Ariani et al., 2008; Hamdani et al., 2010). Recent studies indicated that the FOX family is involved in tumorigenesis as well (Lehtinen et al., 2006; Anders et al., 2011; Kress et al., 2011; Sykes et al., 2011; Ross-Innes et al., 2012). For example, Akt promotes cell survival by phosphorylating and inhibiting FOXO transcriptional activity (Brunet et al., 1999), which is important for the development of leukaemia and colorectal cancer (Tzivion et al., 2011). This regulation is achieved by changing the localization of FOXOs, since Akt-dependent phosphorylation of FOXOs promotes the relocation of FOXOs to the cytoplasm via enhancing the binding of FOXOs to 14-3-3 family proteins and thereby inhibiting their transcriptional activities (Brunet et al., 1999). FOXM1 is often involved in the oncogenesis of many different types of carcinoma (Koo et al., 2012). FOXM1 transcriptional activities are required for its oncogenic functions (Lam et al., 2013); however, FOXM1 could also promote β-catenin nuclear translocation independently of its DNA-binding activity, which may play a role in glioma formation (Zhang et al., 2011). These examples provide a rationale to further determine how these FOX proteins are regulated by protein–protein interaction network.

Thus, we started this project using the FOX TF family as a model to better understand how TFs in general are regulated on and off DNA. We used tandem affinity purification (TAP) followed by mass spectrometry (MS) analysis. As an unbiased approach, MS offers tremendous advantages over other methods in identifying PPIs under near-physiological conditions. Several large-scale MS-based studies have been conducted with yeast and human co-regulator protein complexes (Gavin et al., 2006; Malovannaya et al., 2011). In addition, several function-related large-scale studies have been conducted, which focus on specific signalling pathways (Behrends et al., 2010) or biological processes (Matsuoka et al., 2007; Bennett et al., 2010). While current methods are effective in identifying stable protein complexes, they are inadequate in recognizing regulated interactions, which are essential for understanding the complex signalling networks in the cell. This shortcoming is especially obvious when conducting large-scale proteomics analysis, since the appearance of abundant associated proteins in the interactomes drastically reduces the sensitivity for detecting small amounts of regulated but biologically significant interactions. Using a modified TAP/MS method, we have identified several relatively weak but regulated interactions for individual proteins and functionally validated these interactions (Liu et al., 2010; Wang et al., 2013). Thus, it is technically feasible to identify these functionally important interactions. The question is whether or not we can expand these studies to a relatively larger scale, with the ultimate goal of studying how PPIs change under different physiological conditions. We chose to start by revealing PPIs of TFs that are present on or off chromatin, since we believe that these location-specific PPIs are likely to be engaged in the differential regulations of these TFs.

We isolated soluble versus chromatin fractions based on our assumption that TFs on chromatin are likely to be involved in transcription-related functions, while they are not when in the soluble fractions. Most TFs localize constantly in the nucleus as determined by immunostaining experiments. It was assumed that these TFs would be chromatin bound all the time. This is not the case, since even for TFs that are always localized in the nucleus, they are often present in both soluble and chromatin fractions. Excitingly, our initial proteomics study of the FOX family of TFs revealed that they indeed form distinct complexes on and off chromatin. This finding may not be a total surprise, but it had never been systematically demonstrated. We wondered whether this is a general phenomenon that would apply to other TFs as well. Thus, we performed TAP/MS analyses for 19 non-FOX TFs involved in a variety of pathways associated with development and cancer. These non-FOX TFs are from five structural TF superfamilies to exclude the potential bias caused by the structural preference of their DNA-binding activities. With these additional TFs, we still observed distinct complexes of these TFs formed on and off chromatin. Altogether, our study provided location-specific (i.e. chromatin-associated and chromatin-free) complexes for 56 TFs involved in various signalling pathways and validated our working hypothesis that TFs are engaged in different PPIs on and off chromatin, which are likely important for their regulations and diverse functions.

Results

Proteomic analysis of transcription factors

We performed TAP/MS analyses for total 56 TFs (Supplementary Table S1), including 37 FOX family members and 19 non-FOX TFs involved in various human diseases (Fig 1A) and signalling pathways (Fig 1B, Supplementary Table S1) (i.e. MYC, MAX, TP53, NFKB1, JUN, FOS, SMAD4, TEAD2, RBPJ, TCF4, ATF6, CREB1, ETS1, GLI1, IRF3, MEF2A, NFATC1, PPARG and STAT3). These factors covered the five structural superfamilies of TFs (Matys et al., 2006), which include basic domain TFs (leucine zippers: ATF6, CREB1, FOS, JUN; helix-loop-helix factors: MYC, MAX, TCF4), zinc-coordinating DNA-binding domains TFs (GI1, PPARG), helix-turn-helix TFs (ETS1, IRF3, TEAD2, FOX), β-scaffold factors with minor groove contacts (MEF2A, NFATC1, NFKB1, STAT3, TP53) and others (RBPJ, SMAD4).

We established HEK293T-derivative cell lines stably expressing streptavidin-S-FLAG (SFB) triple-tagged TFs by transient transfections followed by puromycin selection. We picked 12–24 single clones for each bait, examined them by Western blotting and immunostaining and chose the ones with the correct subcellular localization and the lowest expression for affinity purifications. We compared the immunostaining results of our stable cell lines with those available in the literature. All of the tagged proteins were localized as previously reported (Supplementary Fig S1A and summarized in Supplementary Table S2). We also compared the expression levels of 12 tagged proteins with endogenous proteins in
our stable cell lines by Western blotting. Most of the tagged proteins were expressed similar to or slightly higher than that of endogenous proteins (Supplementary Fig S1B). To assess specific protein complexes of TFs on and off chromatin, we first isolated soluble fractions using a crude lysis step and then treated the insoluble pellets (i.e. the chromatin fraction) with TurboNuclease, which hydrolyses both single- and double-stranded DNA or RNA to oligonucleotides of 1–4 bases in length, to release chromatin-bound proteins. We detected very little histones, HMG proteins and other chromatin components in our purifications, suggesting that we were able to eliminated most of the non-specific interactions mediated by DNA. We performed TAP/MS using both soluble and chromatin fractions (for examples, see Supplementary Fig S1C; for fractionation specificity, see Supplementary Fig S1D) and then conducted data analysis to identify these location-specific interactions (i.e. interactions on or off chromatin; Fig 1C). We performed a total of 120 experiments and biological replicates for 24 of them (20%). We also included 70 control purifications, which comprise 61 unrelated protein purifications and nine vector only purifications. The vector-only purifications contain four in chromatin fraction, four in soluble fraction and one combined.

We identified a total of 29,919 interacting proteins from 120 TF and 70 control TAP/MS, which represents 3,751 unique preys (Supplementary Tables S3 and S4). To classify these preys, we adopted a significance analysis method for spectral count data (Choi et al, 2008, 2011) and assigned each prey appearing in the TF group (3,714 in total) with an abundance score and specificity score. The prey abundance score was a parameter estimated by the Poisson mixture model using the SAINT algorithm, which reflects the estimated protein abundance across all the experiments. The specificity score was another parameter, which represents the difference of the estimated prey abundance between the negative control group and the entire group (sample + control). We plotted these two scores to get the prey proteins specifically enriched in the sample group (Supplementary Fig S2A). We have removed the bait self-identifications to avoid any interference of data analysis due to bait overexpression. Using the same approach, we also obtained the prey specificity scores for proteins identified in chromatin (Supplementary Fig S2B) and soluble fractions (Supplementary Fig S2C). These analyses allowed us to easily rule out the generic preys (HSPs) and other non-specific binding proteins (actin, tubulin, RPLs), which appear as tails at the lower right corner of each atlas, while the proteins with high specificity and abundance appear in the upper right regions (Supplementary Fig S2). We then combined these three individual distributions into a bubble plot for prey categorization (Fig 2A, Supplementary Table S5). Using this prey fraction and specificity distribution map, we classified the preys into four coloured groups based on their positions: red: co-regulators of TFs may be involved in transcriptional regulation (e.g. MAX, TRRAP); purple: regulators with no fractional preference (MGA, Sin3A); blue: transcription-unrelated functions or negative regulators of TFs (CUL2, SKI); and green: potential regulators with less specificity (CBX3, COPSS, NTPCR) (Fig 2A). The grey indicates non-specific binding proteins with no preference (HSP, RPS, RPL), which locate at the bottom of this plot (Fig 2A). This prey distribution map may suggest how these preys act in regulating TFs in a generic or a specific manner.

To discover individual interactions with high confidence (HCIPs, high-confidence candidate interacting proteins), raw data from the MS analysis were subjected to a modified SAINT (Significance Analysis of INTeractome) algorithm (Choi et al, 2011; Wang et al, 2014) (Fig 2B). We used a two-pool analysis, and the spectra counts from TF group and control group were assembled as a matrix for all of the bait and prey proteins (Supplementary Table S3). In total, 29,919 protein matches were identified in 190 experiments, with 60 TF purifications from chromatin fractions, 60 TF purifications from soluble fractions and 70 control purifications (Fig 2C). We temporarily removed the bait self-identifications from the list to get a better estimation of bait abundance in cells and then added them back after the completion of the filtration. According to the SAINT methodology, the interactions with over 0.8 probability score were kept for further analysis as described below. 8,500 interactions passed this first filtration: 3,927 of which from chromatin fractions and 4,573 from soluble fractions (Fig 2C).

We then used the prey specificity information to further eliminate contaminants frequently shown in our purifications. The score \( \mu_i \) of individual preys as common contaminants were used to calculate the probability of abundant or non-specific preys that frequently showed up in these purifications. We filtered out preys with \( \mu_i \geq 0 \), which removed common contaminants and/or abundant non-specific binding proteins: 4,626 interactions passed this filtration. In total, 2,156 interactions passed both filtrations (Fig 2B) and were designated as HCIPs: 1,423 of which from chromatin fractions and 733 from soluble fractions (Table 1, Supplementary Table S6, Supplementary Dataset S1). We also performed biological replicates for 24 purifications, and the overlapped HCIPs were summarized in Supplementary Table S7. As predicted, the reproducibility of the HCIPs increases with the spectra counts (44% for 2 counts, 67% for 5 counts and 90% for 10 counts) (Fig 2C).

### Data validation and functional studies

To validate the reliability of our data set, we searched the HCIPs in various PPI databases, including BioGrid (Stark et al, 2006), STRING (von Mering et al, 2003), BIND (Badger et al, 2003), DIP ( Xenarios et al, 2000), HPRD (Prasad et al, 2009) and C. elegans TF data set (Reece-Hoyes et al, 2013). 15% interactions we identified have been confirmed by this combined knowledge database. We also compared the results with the CCI data set we published earlier (Malovannaya et al, 2011). 10% interactions we identified have been confirmed by the CCI database, which was created based on MS analysis of immunoprecipitates of endogenous protein complexes and therefore could be considered as an "endogenous co-IP validation". For example, proteins SIN3A, SAP130, ARID4B, MRC2, FOXK1 and FOXK2 from MAX purification were confirmed by the CCI database (Supplementary Table S6 “CCI-Confirmed” column). In total, 389 interactions of the total 2,156 HCIPs (~18%) were reported previously (Fig 2D and Supplementary Table S6). If only considering the 19 relatively well-studied non-FOX TFs, 33% (217 out of the 663 HCIPs) of the interactions were reported previously (Fig 2D and Supplementary Table S6), which confirmed the validity of our data set. We have also overlapping our results with the CRAPome, a collection of common contaminants in AP/MS data (Mellacheruvu et al, 2013). Using 20% frequency as the “non-specific interaction” cut-off, we found our HCIP set only generated 3.8% “potential false positives”.

---

**Molecular Systems Biology**

Chromatin-associated and soluble human TF complexomes **Xu Li et al**

© 2015 The Authors

4 *Molecular Systems Biology* 11, 775 | 2015

108306-CS-EMBO-ChromatinReprint-Text.indb 32

16/03/2015 15:23
To experimentally verify our proteomics data set (Supplementary Table S6), we performed reciprocal purifications using prey proteins and hoped to identify the corresponding bait proteins. We conducted reverse purifications for MAX-associated protein L3MBTL2, E2F6, FOXK2; NFATC1-associated protein JUN, HOXD13, CREB1, ATF1, ATF3; RBPJ-associated protein L3MBTL3, KDM1, FBXO42; and CREB1-associated protein ATF1, HMGAI, ZNF131, NFIX, NFATC2. In 14 out of 16 purifications, we identified the corresponding bait in the reverse purifications (Fig 2E, Table 2), 13 of which were in the corresponding fractions. This further validated that our purification results reflect endogenous PPIs under physiological conditions. We also performed co-IP experiments using tagged preys to pull down endogenous baits, to validate the HCIPs of MAX, FOXM1 and FOXO3. We have obtained 54% positive rate (82% if only counting the preys we actually tested since some prey constructs were not available) (Supplementary Fig S3A, B and D). shRNA screening of
Transcription factors form distinct functional complexes on and off chromatin, which could be functionally relevant

We listed the top HCIPs of each bait in each fraction (Fig 3A). It is clear that these HCIPs are very different between the fractions. The total spectra counts (TSC) of HCIPs for different baits were also compared between two fractions, which reflect the total amount of specific protein bindings of these TFs on or off chromatin (Fig 3B), which showed distinct preferences for these TFs to form complexes on or off the chromatin. RBPJ, a TF known to be chromatin bound before and after Notch activation, showed overwhelming protein binding in the chromatin fraction, while STAT3, which is predominantly present in the cytoplasm without stimuli, had very few interactions in the chromatin fraction (Fig 3B). These results indicate that our comparative interactomes are biologically meaningful.

Notably, only 14% of total HCIPs appeared in both soluble and chromatin fractions (Fig 4A), many of which are bait self-interactions. To make sure that this is not due to any artefact caused by data analysis, we used the CCI algorithm (Malovannaya et al., 2011) and an in-house written algorithm based on CompPASS (Sowa et al., 2009). The percentage of total HCIPs appearing in both fractions varies between 8 and 14% in these analyses, which confirmed our working hypothesis that TFs are differentially regulated on and off chromatin by distinct protein partners. We listed the protein families of these HCIPs (Supplementary Table S6) and summarized them in two different fractions (Fig 4B). As expected, proteins related with transcription are enriched in the chromatin fraction ($P = 4.07e-7$), while kinase ($P = 6.08e-5$), peptidase ($P = 7.16e-4$) and transmembrane proteins ($P = 0.037$) are enriched in the soluble fraction. This may reflect that the regulation of protein post-translational modifications and trafficking occurs preferentially in the soluble fraction. Since these functional indications based protein families may not be precise, we further annotated the HCIPs with ubiquitin-related function categories (Fig 4C). We found that degradation-related ubiquitin (Ub) E3 ligases/F-box proteins (FBXs), proteasome subunits and deubiquitinase (DUBs) only appeared in soluble fractions, while the chromatin remodelling ubiquitin E3s are highly enriched in chromatin fractions (Fig 4C). This more detailed analysis indicates that fraction preference is associated with the functions and/or the regulations of prey proteins.

### Table 1. Summary of step-by-step proteomics data analysis.

<table>
<thead>
<tr>
<th>Experiments</th>
<th>Total TFs</th>
<th>Chromatin</th>
<th>Soluble</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptides</td>
<td>185,394</td>
<td>113,654</td>
<td>48,077</td>
<td>65,577</td>
</tr>
<tr>
<td>Proteins</td>
<td>29,919</td>
<td>19,698</td>
<td>9,380</td>
<td>10,318</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Passed SAINT</th>
<th>8,500</th>
<th>8,500</th>
<th>1,927</th>
<th>4,573</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCIP</td>
<td>2,156</td>
<td>2,156</td>
<td>1,423</td>
<td>733</td>
</tr>
</tbody>
</table>

The identified proteins and peptide numbers of each group and step were presented.

### Table 2. Summary of reciprocal purifications results.

<table>
<thead>
<tr>
<th>Bait</th>
<th>Prey</th>
<th>In bait fraction</th>
<th>In prey fraction</th>
<th>Bait</th>
<th>Prey</th>
<th>In bait fraction</th>
<th>In prey fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAX</td>
<td>L3MBTL2</td>
<td>Chr</td>
<td>Sol</td>
<td>RBPJ</td>
<td>L3MBTL2</td>
<td>Chr</td>
<td>Sol</td>
</tr>
<tr>
<td>MAX</td>
<td>E2F6</td>
<td>Chr</td>
<td>Chr</td>
<td>RBPJ</td>
<td>KDM1</td>
<td>Chr</td>
<td>Chr</td>
</tr>
<tr>
<td>MAX</td>
<td>FOXO2</td>
<td>Chr</td>
<td>Chr</td>
<td>RBPJ</td>
<td>FBX042</td>
<td>Chr/Sol</td>
<td>Chr/Sol</td>
</tr>
<tr>
<td>NFATC1</td>
<td>JUN</td>
<td>Chr</td>
<td>Chr</td>
<td>CREB1</td>
<td>ATF3</td>
<td>Chr/Sol</td>
<td>Chr/Sol</td>
</tr>
<tr>
<td>NFATC1</td>
<td>HOXD13</td>
<td>Chr</td>
<td>Chr</td>
<td>CREB1</td>
<td>HMGA1</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>NFATC1</td>
<td>CREB1</td>
<td>Chr</td>
<td>Chr</td>
<td>CREB1</td>
<td>ZNF131</td>
<td>Chr</td>
<td>N</td>
</tr>
<tr>
<td>NFATC1</td>
<td>ATF3</td>
<td>Chr</td>
<td>Chr</td>
<td>CREB1</td>
<td>NF0X</td>
<td>Chr</td>
<td>N</td>
</tr>
<tr>
<td>NFATC1</td>
<td>ATF1</td>
<td>Chr</td>
<td>Chr</td>
<td>CREB1</td>
<td>NFATC2</td>
<td>Chr</td>
<td>N</td>
</tr>
</tbody>
</table>

Reciprocal purifications of 16 interactions identified from MAX, NFATC1, RBPJ and CREB1 purifications were performed with the same TAP/MS protocol. Chromatin and soluble fractions were separated and whether the corresponding baits appeared in the reciprocal purification was indicated by fraction name “Chr” or “Sol”.

MAX and FOXO3 HCIPs was conducted using, respectively, Ki67 staining and GADD45A mRNA level as read-outs. We have obtained eight positive potential regulators and six potential negative regulators of MAX, and six potential positive regulators of FOXO3 (Supplementary Fig S3C and E). From FOXO3 positive hits, we chose FOXK1 for further validation. We found that FOXK1 also regulates FOXO3 subcellular localization. Overexpression of FOXK1 translocated FOXO3 to the nucleus, while knocking down FOXK1 reduced the subcellular localization. Overexpression of FOXK1 translocated FOXO3 to the nucleus, while knocking down FOXK1 reduced the subcellular localization. FOXO3 nuclear translocation upon treatment with PI3K inhibitor LY294002 (Supplementary Fig S3F). Taken together, these data indicate that the interactomes are biologically highly reliable.

We have also searched post-translational modifications including phosphorylation and acetylation in all our MS results (Supplementary Table S8) and identified 8,043 peptides modified by phosphorylation and/or acetylation. In total, we identified 6,842 phosphorylation sites and 4,384 acetylation sites. Among these modified peptides, 47% of total and 36% of the bait peptides only exist in one fraction, which indicates that the PTMs of TF protein complexes are different in soluble and chromatin fractions.

Transcription factors form distinct functional complexes on and off chromatin, which could be functionally relevant

We listed the top HCIPs of each bait in each fraction (Fig 3A). It is clear that these HCIPs are very different between the fractions. The total spectra counts (TSC) of HCIPs for different baits were also compared between two fractions, which reflect the total amount of specific protein bindings of these TFs on or off chromatin (Fig 3B), which showed distinct preferences for these TFs to form complexes on or off the chromatin. RBPJ, a TF known to be chromatin bound before and after Notch activation, showed overwhelming protein binding in the chromatin fraction, while STAT3, which is predominantly present in the cytoplasm without stimuli, had very few interactions in the chromatin fraction (Fig 3B). These results indicate that our comparative interactomes are biologically meaningful.

Notably, only 14% of total HCIPs appeared in both soluble and chromatin fractions (Fig 4A), many of which are bait self-identification. To make sure that this is not due to any artefact caused by data analysis, we used the CCI algorithm (Malovannaya et al., 2011) and an in-house written algorithm based on CompPASS (Sowa et al., 2009). The percentage of total HCIPs appearing in both fractions varies between 8 and 14% in these analyses, which confirmed our working hypothesis that TFs are differentially regulated on and off chromatin by distinct protein partners. We listed the protein families of these HCIPs (Supplementary Table S6) and summarized them in two different fractions (Fig 4B). As expected, proteins related with transcription are enriched in the chromatin fraction ($P = 4.07e-7$), while kinase ($P = 6.08e-5$), peptidase ($P = 7.16e-4$) and transmembrane proteins ($P = 0.037$) are enriched in the soluble fraction. This may reflect that the regulation of protein post-translational modifications and trafficking occurs preferentially in the soluble fraction. Since these functional indications based protein families may not be precise, we further annotated the HCIPs with ubiquitin-related function categories (Fig 4C). We found that degradation-related ubiquitin (Ub) E3 ligases/F-box proteins (FBXs), proteasome subunits and deubiquitinase (DUBs) only appeared in soluble fractions, while the chromatin remodelling ubiquitin E3s are highly enriched in chromatin fractions (Fig 4C). This more detailed analysis indicates that fraction preference is associated with the functions and/or the regulations of prey proteins.
Different complexes formed with a given TF on and off chromatin may indicate how it is regulated differently and/or carry out distinct functions in these fractions. For instance, the DREAM complex, which is composed of LIN9, LIN37, LIN54, MYBL2 and S100A7, was identified only in the chromatin fraction of FOXM1 (Fig 4D). This finding agrees with the known function of the DREAM complex, which recruits FOXM1 to promote mitotic gene expression (Sadasivam et al., 2012; Chen et al., 2013; Grant et al., 2013) and thereby regulates cell cycle progression (Litovchick et al., 2007; Schmit et al., 2007). On the other hand, the tRNA splicing ligase complex, which is composed of IARS, DARS, EPRS, PARS2, QARS and C22orf28, was present only in the soluble fraction of FOXM1-associated proteins. This complex is known to be phosphorylated during mitosis and is required for tRNA splicing and mitotic functions (Dephoure et al., 2008; Olsen et al., 2010), implying that FOXM1 may have a previously unknown function in this process. The GO function annotation suggested that FOXM1 may have distinct functions on and off the chromatin, but still contribute to the overall functions of cell cycle, cellular development, cell growth and proliferation (Fig 4E). These data suggest that there might be a potential role of FOXM1 in tRNA splicing complex function or regulation. Our GO analysis of FOXM1 interactomes suggested similar biological functions of FOXM1 with different complex formations in these two fractions (Fig 4F).

In some other cases, the complexes in the two fractions may perform unrelated or even opposing regulatory roles. Many ubiquitin (Ub) E3 ligases/F-box proteins (FBXs) and proteasome subunits appeared in the left region of our prey specificity and abundance.

**Figure 3.** Transcription factors form distinct complexes on and off chromatin.

A HCPs with highest spectra counts were listed. The length of each box with the protein name on it indicates the protein size. Black fonts indicate new interactions identified by our purifications. Orange fonts indicate interactions defined by our purifications and the literature.

B Total spectra counts of TFs in different fractions. The y-axis indicates the total spectra counts (TSC) of HCPs in corresponding TF purifications. Red bar: TSC of HCPs in chromatin fractions; blue bar: TSC of HCPs in soluble fractions.
A FOXN2 cofactor that also helps other TFs to recognize target DNA (Gajiwala et al., 2000), in the chromatin fraction (Fig 5A). We confirmed this interaction via co-IP experiments (Fig 5B) and found 44% FOXN2 target genes overlapped with RFX1 target genes (Fig 5C). This indicates that FOXN2 and RFX1 form a transcriptional complex on chromatin and co-regulate a subset of gene transcription. On the other hand, FOXN2 binds to SKP1, βTrCP (also called BTRC/FBXW1), βTrCP2 (also called FBXW11) and some proteasome components such as PSMD8 in the soluble fraction. We have also identified CUL1 from our parallel virus-based FOXN2 TAP/MS.

Indeed, our pathway annotation analysis indicated that FOXN2 is tightly associated with the protein ubiquitination pathway in the soluble fraction, but not in the chromatin fraction. To further validate these results, we performed reciprocal purification in HEK293T cells using βTrCP2/FBXW11 as the bait, which is the adaptor protein that should associate with CUL1 substrates (Lyapina et al., 1998; Wu et al., 2000). Indeed, we identified many known βTrCP2-interacting proteins, which include CUL1, SKP1 (Lyapina et al., 1998; Wu et al., 2000), USP37 (Burrows et al., 2012), USP47 (Peschiaroli et al., 2010), EIF2AK (Meloche & Roux, 2012), CDC25A (Busino et al., 2003) and CTNNB1 (Hart et al., 1999) (Fig 5D). As expected, we also identified FOXN2 as a βTrCP2-binding protein (Fig 5D). We validated that FOXN2 binds to βTrCP and βTrCP2, but not to βTrCP substrate binding-defective mutant R474A (Inuzuka et al., 2010) (Fig 5E). βTrCP and βTrCP2, but not the mutant βTrCP (R474A), promoted FOXN2 ubiquitination in vivo (Fig 5F). While knocking down CUL1 significantly stabilized
**Figure 5. Functional validation of FOXN2 based on its interacting proteins in soluble and chromatin fractions.**

A FOXN2 HCIPs form distinct complexes in chromatin versus soluble fractions. The size of prey dots indicates the estimated abundance of preys. Lines indicate the interactions defined in the literature. CUL1 was identified in parallel virus-based FOXN2 purification.

B 293T cells were transfected with constructs encoding MYC-tagged RFX and SFB-tagged FOXN2 or its DNA binding-defective mutant FOXN2 (H162R) as indicated. Pull-down experiments were carried out with S-protein beads and immunoblotted with antibodies as indicated.

C Overlap Venn diagram of FOXN2 and RFX1 target genes identified by ChIP-sequencing. 293T cells stably expressing SFB-tagged FOXN2 or RFX1 were subjected to ChIP-sequencing using anti-FLAG antibody. Each experiment was performed with two biological replicates, and four control ChIP-sequencings were performed using 293T cells stably expressing other TFs.

D Reverse purification of FBXW11 (βTRCP2)-containing protein complex conducted using the same TAP/MS protocol recovered FOXN2 as FBXW11-binding protein. Prey names, peptide counts and whether or not the interactions have been reported are listed.

E 293T cells were transfected with constructs encoding SFB-tagged FOXN2 and MYC-tagged βTRCP, its substrate binding-defective mutant βTRCP (R474A), or βTRCP2 as indicated. Pull-down experiments were carried out with S-protein beads and immunoblotted with antibodies as indicated.

F In vivo ubiquitination assays were performed by co-transfecting constructs encoding FLAG-tagged FOXN2, His-tagged ubiquitin, MYC-tagged βTRCP, βTRCP (R474A) or βTRCP2 onto HEK293T cells as indicated. Cell lysates were denatured with 1% SDS and diluted 10-fold using PBS prior to the pull-down by Ni-NTA resin, followed by immunoblot with antibodies as indicated.

G 293T or 293T-βTRCP2, 293T-βTRCP2, 293T-βTRCP2, 293T-βTRCP2 cells were treated with 100 mM cycloheximide (CHX) for the indicated time. Immunoblotting was conducted with antibodies as indicated.

H A model showing on/off chromatin regulation of FOXN2 by transcriptional co-factors or E3 ligase complexes. All of the components indicated were identified from FOXN2 purifications.

Source data are available online for this figure.
FOXN2, knocking down βTRCP or βTRCP2 only partially stabilized FOXN2 (Fig 5G), which agrees with βTRCP and βTRCP2 being highly related proteins and having overlapping functions in the cell. On the basis of the data presented above, we propose that FOXN2 associates with RXF1 on chromatin to carry out its transcriptional functions, but that FOXN2 itself is regulated by proteasome-mediated degradation in the soluble fraction (Fig 5H). This example indicates that by investigating the different interacting proteins in soluble and chromatin fractions, we are able to gain further insights into the regulations and functions of transcription factors.

To further confirm that our newly established TF fraction-specific PPI network can be used for predicting novel protein functions or regulations, we expanded our studies on the JUN/CREB/ATF/NFATC subnetwork. Our TAP/MS and corresponding reciprocal TAP/MS analysis showed that CREB1 and NFATC1/2 bind to each other only in the chromatin fraction, which is also the case for ATF1/2/3/7 and NFATC1 (Table 2, Supplementary Table S4). However, CREB1 and ATF2s always bind to each other, regardless of whether they are on or off chromatin (Table 2, Supplementary Table S4). To confirm that these fraction-specific complex formations truly reflect the endogenous situation, we further conducted endogenous JUN purifications in chromatin and soluble fractions (Supplementary Table S9). 65% high-confidence interactions identified by our tagged TAP/MS have been confirmed by these AP/MS using antibodies against endogenous JUN. More importantly, we found that the fraction specificities are highly reproducible with endogenous AP. FOS, ATF2/3/7 and CREBS were found in both fractions, while NFATC1 still only appeared in chromatin fractions (Supplementary Table S9). Putting these TAP/MS data together, we built a JUN/CREB/ATF network (Fig 6A). Western analysis confirmed that NFATC1 binds to JUN, CREB1 and ATF1/2 predominantly in the chromatin fraction (Fig 6B), while ATF1 binds to JUN/CREB/ATF2 in both fractions (Fig 6C). ChIP-seq results from the ENCODE database suggested that 68% of CREB1 target genes overlapped with ATF2 target genes, while only 27% overlapped with NFATC1. 87% of the CREB1/NFATC1 interactions were confirmed by Western analysis of endogenous JUN.

**Figure 6.** Overview of JUN/CREB/ATF/NFATC1 subnetwork.

A. JUN/CREB/ATF/NFATC1 subnetwork map. Arrows indicate the identifications from TAP/MS. Bold arrows indicate the identifications from both TAP/MS and endogenous AP. Colours of arrows indicate the locations of interactions: red, in chromatin only; purple, in both fractions.

B. Pull-down: S-beads. Anti-MYC (NFATC1) and Anti-FLAG proteins were pulled down from chromatin and soluble fractions. Myc-NFATC1 and SFB-FLAG proteins were identified as interacting with CREB1, FOS and CREBS in both fractions.

C. Pull-down: S-beads. Anti-MYC (NFATC1) and Anti-FLAG proteins were pulled down from chromatin and soluble fractions. Myc-NFATC1 and SFB-FLAG proteins were identified as interacting with CREB1, FOS and CREBS in both fractions.

D. ChIP-seq gene overlaps. The ChIP-seq gene overlaps show that CREB1, ATF2 and NFATC1 target genes using ChIP-seq data sets generated by the ENCODE consortium in GM12878 cells.

Source data are available online for this figure.
co-target genes are also targeted by ATF2 (Fig 6D). These data indicate that JUN/CREB/ATFs are likely to form a stable complex, while they may only associate with NFATC when they are targeted to chromatin and act with NFATC to control gene transcription.

To further characterize these TFs and their associated proteins in the context of biological processes, we carried out pathway analysis to identify the new biological insights of baits indicated by their HCIPs (Supplementary Table S10) and the analysis of alteration of HCIPs of each bait in multiple cancer databases (Supplementary Table S11). These analyses link the TFs to a wide variety of cellular functions and disease correlations. For instance, we compared the GO annotation results of RBPI (Supplementary Fig S4A) and FOXO3 (Supplementary Fig S4B) based on their non-self-interact HCIPs (TF-HCIP) or with the known functions in the literature (TF literature). We have not only identified previously reported functions such as Notch signalling (Tangaki & Honjo, 2010) and PI3K/AKT signalling (Brunet et al., 1999; Tzivion et al., 2011), oestrogen receptor signalling (Guo & Sonenshein, 2004; Xia et al., 2006), but also novel functions or regulations, such as RBPI might be regulated by DNA methylation-mediated transcriptional repression (Supplementary Fig S4A), and FOXO3 is potentially involved in checkpoint signalling in the DNA damage response (Supplementary Fig S4B).

Discussion

In this study, we revealed chromatin-associated and soluble complexomes for each of the 56 TFs, which validated our hypothesis that TFs form unique protein complexes on and off chromatin. These results and other information presented in this study offer new insights into the regulation of TFs and their diverse in vivo functions.

Among our 214 TAP/MS results, there are several results that have few or no prey identifications. We believe this is due to several reasons. In many cases, such as in the cases of FOXC2 and FOXJ2, the bait proteins expressed well and could be found in both fractions. For example, we identified 38 and 64 peptides of FOXC2, 129 and 100 peptides of FOXJ2 in chromatin and soluble fractions, respectively. After the removal of non-specific interacting proteins such as chaperones, there is no HCIP left in the soluble fraction lists, which may indicate that these proteins only form functional complexes on chromatin. In some other cases, such as FOXA2 and ETS1, the bait proteins are highly enriched in one fraction. For example, we identified 20 peptides of FOXA2 in the chromatin fraction, but none in the soluble fraction. Similarly, we identified 59 peptides of ETS1 in the chromatin fraction and 402 peptides in the soluble fraction. In these cases when the bait protein was predominantly present in one fraction, the lack of HCIPs in the other fraction could reflect the nature of these bait TFs, which predominantly form functional complexes in one fraction. Of course, this could also be due to technical reasons, especially for chromatin fractions, since we may lose some of the associated proteins during extraction of chromatin-associated proteins. In the case of FOXO4, the failure to identify HCIPs could be just due to technical issues, since we only recovered a few peptides of FOXO4 in either fractions, which may indicate problems with protein expression or stability. It is known that AP/MS covers only a limited portion of the total peptide population (Liu et al., 2004), and some proteins are relatively difficult to recognize by MS because of their abundance or sequence/structural features (Ailieaar et al., 2013); the peptide readings per se may not reflect the real biological importance of these protein–protein interactions, especially if these interactions are regulated in a signalling pathway. For instance, we repeated our NF-kB1 purifications and were able to obtain RELB (Supplementary Table S12), a known NF-kB1-binding protein (Bouwmeester et al., 2004). Thus, more replicates may help to uncover additional HCIPs, but at a higher cost.

Affinity purification (AP) is commonly used in large-scale proteomics studies in mammalian systems, which led to several milestone discoveries. However, the method has its limitations, especially in detecting transient or regulated interacting proteins in the presence of highly abundant, non-specific associated proteins (Figeys et al., 2001; McHugh & Arthur, 2008). The major challenge to identify the relatively weak but regulated interactions is to eliminate the huge amount of common contaminants and abundant proteins frequently shown in the MS, since the real signals are often buried in the large amount of unspecific noise. These contaminants mainly come from two difference sources. One is from the pull-down assay per se, which usually generates a list of non-specific binding proteins that have affinity for the particular matrix (i.e. antibodies or other) one uses. This could be easily removed during our tandem affinity purification, since the non-specific binding proteins are unlikely to have affinity towards two different affinity matrices. The other source comes from the binding of overexpressed bait with in most cases abundant cellular proteins. For example, many tagged proteins would associate with various heat shock proteins, tubulins and RNA-binding proteins. In most cases, these commonly identified associated proteins that show up in multiple purifications are eliminated by our bioinformatics analysis using the modified SAINT method.

In this study, we employed a modified TAP method for isolating protein complexes and obtained results that could be further validated. Thus, we hope our TAP/MS-bioinformatics package could provide an easy and accurate way of studying protein–protein interactions, which can be expanded not only to protein families and signalling pathways, but also to the entire ORFeome in the future. Of course, this approach has its own drawbacks, including the use of tags that may interfere with certain PPis, overexpression of a given protein (which may lead to its mislocalization, misfolding or both) and very weak binding proteins that may be lost during the two-step purifications. In addition, our stringent criteria for selecting HCIPs could filter out some true but weakly or transiently interacting proteins. We may need to develop a new computational method specifically designed for this TAP/MS approach to obtain as many true interacting proteins as possible, but at the same time eliminate contaminants and abundant proteins that are often associated with the baits.

Several previous reports have suggested that the association between certain TFs and chromatin is tightly regulated and that these TFs could shuttle on and off chromatin. However, whether these applies to other TFs remains largely unknown. In other words, are there often “free” transcription factors that exist off chromatin? Based on the data presented above, the answer is yes. Indeed, the distribution of TFs and their associated proteins in these fractions differs dramatically (Supplementary Fig S1D; also see Fig 3). Most TFs form very different protein complexes on and off chromatin, which means that soluble and chromatin-bound fractions exist for
Molecular Systems Biology

Chromatin-associated and soluble human TF complexes

Xu Li et al

many TFs and the majority of the proteins in soluble fractions are not just the proteins released from chromatin during purification. The separation of soluble nuclear proteins versus soluble cytosolic proteins is irrelevant to our study, since these soluble TF-associated proteins should not be directly involved in transcriptional regulation, regardless of where they are (i.e. in the nuclei or in the cytosol).

Are these “location-specific” interactomes specific and relevant to the regulation of protein functions? We believe the answer is also a yes. The protein complexes in different fractions are highly specific and functionally relevant. In 14 reciprocal TAP/MS analyses that successfully identified bait proteins, 13 of which were uncovered in the corresponding fractions (Fig 2F and G). The reciprocal TAP/MS of βTRCP2/FBXW11, which was identified from FOXX2 soluble fractions, also captured FOXX2 only in the soluble fractions (Fig 5C). This means that complexes are not just randomly distributed in chromatin and soluble fractions, but they are highly organized and regulated. Another interesting example is that JUN/FOS, CREB1/S and ATF1/2/3/7 are likely to form a stable transcription complex, while they may only associate with NFATC when they are targeted to chromatin to regulate gene transcription. This hypothesis is, at least partially, supported by reports indicating that CREB1 is targeted to the same promoter sequence as NFATC under certain stimulations (Sato et al, 2006; Suehiro et al, 2010).

On the basis of our proteomics studies, we proposed that we could use this location-specific (i.e. chromatin-associated and soluble) interactomes to predict how TFs function or are regulated in the cell (Figs 4–6 and Supplementary Fig S4). For example, we found that FOXX1 formed two distinct complexes on and off chromatin, but both complexes are potentially involved in the same biological events: the regulation of mitosis and the promotion of cell cycle progression (Fig 4D–F), while the two FOXX2 complexes formed on and off chromatin had opposing roles in regulating FOXX2 functions (Fig 5). In addition, the functional annotation based on the HCIPs correlates well with some of the known functions in the literature. For example, RBP1/CFI, a TF that plays a central role in Notch signalling (Tanigaki & Honjo, 2010), is functionally annotated to be linked with the Notch signalling pathway (Supplementary Fig S4A). Similarly, on the basis of our analysis, the well-studied FOXX3 acts in PI3K/AKT signalling (Brunet et al, 1999; Tzivion et al, 2011), oestrogen receptor signalling (Guo & Sonenshein, 2004; Xia et al, 2006), ERK5 signalling (Finegan et al, 2009) and cell cycle/checkpoint regulation (Chung et al, 2012). These results agree well with the known diverse functions of FOXX3 in these processes (Supplementary Fig S4B). Moreover, both RBP1 and FOXX3 proteins could be functionally annotated to novel function or regulations based on our PPI studies: RBP1 might be regulated by DNA methylation-mediated transcriptional repression (Supplementary Fig S4A), and FOXX3 is potentially involved in BRCA1 signalling in DNA damage response (Supplementary Fig S4B). Therefore, we anticipate that these functional annotations defined by guilt by association will be beneficial, especially when studying proteins with unknown functions. However, since proteins often form distinct protein complexes in different environments or tissues to execute their tissue-specific functions, one needs to be cautious when drawing any conclusion solely based on PPI studies. In these cases, the functional validation and relevance should be the most important aspects and these leads should be pursued with a biological question in mind. Thus, our functional prediction and disease correlations based on the TAP/MS results performed in HEK293T cells only represent a fraction of the functions carried out by these bait proteins and should only be used as references.

In summary, our study offers a valuable resource of protein–protein interaction networks for transcription factors involved in many signalling pathways and human diseases. Although it may not come as a total surprise, our findings highlight that transcription factors form distinct complexes on and off chromatin. This location-based interactomes may be used to predict the molecular mechanisms underlying the regulations of these transcription factors and their associated biological functions.

Materials and Methods

Constructs and small hairpin RNAs (shRNAs)

FOXX1 and FOXX2 cDNAs were a generous gift from Dr. Andrew D. Sharrock (Li et al, 2012). βTRCP (R474A) cDNA was a generous gift from Dr. Wenyi Wei (Inuzuka et al, 2010). cDNAs encoding other known FOXX proteins and TFs, MYC, MAX, RBP1, TCF4, TEAD, JUN, FOS, NF-κB, SMAD4, ATF6, CREB1, ETFS1, GLI1, IRE1, MEF2A, NFATC1, PPAR, STAT3, L3MBTL2, L3MBTL3, E2F6, HOXD13, ATF1, ATF2, ATF3, ATF7, HMGAI, ZNF131, KDM1, FBXO42, NFIX, CBX3, EPC2, MTERF, XRCC1, TFAP2A, VRK3, MEC2P, TDFP1, EEF1D, LIG3, RPA2, H2AFY, RECQL, PAR52, DDX52, QARS, C1orf25, DARS, ORC2L, P53, IARS2, SMARCA1, SMARCA5, LARS, LRWD1, ZC3H11A, CDC27, RB1, RB1L1, PKN, TAF5L, ARHGEF2, RFIX, βTRCP and βTRCP2 were obtained from the hORFV5.1 library and Open Biosystems. cDNAs were subcloned into the pDONR201 vector (Invitrogen) as entry clones and subsequently transferred to gateway-compatible destination vectors for the expression of C-terminal SFB-, MYC-, GFP-tagged fusion proteins. Point or deletion mutants were generated using sequential PCR methods and verified by sequencing. FOXX2 DNA binding-defective mutant H162R was generated based on the conserved DNA-binding domain reported for FOXX3 (Harada et al, 2010). Four individual pCIP2 lentiviral shRNAs targeting βTRCP, βTRCP2, CUL1, ARHGEF2, TRRAP, TAF5L, CDC27, RB1, RB1L1, PKN, FOXX1, EPO300, CREBBP, MEC2P, LIG3, TFAP2A, TBP, FOXX2, L3MBTL2, SMARCA1, RPA2, EEF1D, WIZ, TDFP1, TWIST1, SAP130, H2AFY, XRCC1, RECQL, VRK3, MTERF, MORC2, BAZ2A, MXI1 and MXD4 were obtained from the shRNA and ORFeome core facility at The University of Texas MD Anderson Cancer Center. All lentiviral supernatants were generated by transient transfection of 293T cells with packaging plasmids pSPAX2 and pMD2G and harvested 48 h later. Supernatants were passed through a 0.45-μm filter and used to infect HEK293T and MCF10A cells with the addition of 8 μg/ml polybrene.

Cell culture, treatments and transfection

HEK293T cells were cultured in Dulbecco’s modified Eagle’s medium supplemented with 10% foetal bovine serum and 1% penicillin and streptomycin. MCF10A cells were cultured in Dulbecco’s modified Eagle’s medium/F12 supplemented with 5% horse serum, 10 μg/ml insulin, 20 μg/ml epidermal growth factor, 0.5 μg/ml hydrocortisone, 0.1 μg/ml chola toxin and 1% penicillin and streptomycin.
The activity of several TFs, such as SMAD4 and NF-xB1, can be greatly influenced by specific signalling events. For this reason, we also treated some stable cells with drugs to promote the nuclear translocation of some TFs. We performed purifications of SMAD4 with treatment of 5 ng/ml TGFβ or BMP4 for 16 h, TCF4 with treatment of 10 mM LiCl for 16 h or NF-xB1, JUN, FOS, with treatment of 25 ng/ml TNF-α for 8 h to promote chromatin association of the corresponding TFs. In many cases, we were able to isolate chromatin-associated proteins of these TFs even without any treatment.

Constructs encoding C-terminally SFB-tagged TFs were transfected into HEK293T cells using polyethyleneimine as previously described (Wang et al., 2013). Cells were selected with puromycin and 12–24 single clones were picked, examined by Western blotting and immunostaining. We chose the ones with the correct subcellular localizations and the lowest expression for the subsequent TAP/MS analysis.

**TAP of TF-associated protein complexes and MS analysis**

A total of \(1 \times 10^7\) HEK293T cells stably expressing tagged TFs were lysed with NETN buffer [20 mM Tris–HCl (pH 8.0), 100 mM NaCl, 1 mM EDTA and 0.5% Nonidet P-40, containing 1 μg/ml each of pepstatin A and apro tin] for 30 min. Crude lysates were saved as the soluble fraction by centrifugation at 16,000 g at 4°C for 30 min, and the pellet was digested with Western blotting (Acceleragen) for 10 min in digesting buffer [50 mM Tris (pH 8), 1 mM MgCl2 and protease inhibitor] to extract chromatin-bound proteins. The super nats were cleared at 16,000 g to remove debris from chromatin bound protein fractions. Both fractions were then incubated with streptavidin-conjugated beads (Amersham) for 2 h at 4°C. The beads were washed three times with NETN buffer, and the bead-bound proteins were eluted with NETN buffer containing 2 mg/ml biotin (Sigma). The elutes were incubated with S-protein beads (Novagen). The beads were again washed three times with NETN buffer and subjected to SDS-PAGE. Protein band containing the entire sample was excised, and MS analyses were performed by the TAPLN Biological Mass Spectrometry Facility of Harvard University.

For MS analysis, excised gel bands were cut into approximately 1-mm² pieces. Gel pieces were then subjected to in-gel trypsin digestion and dried. Samples were reconstituted in 5 μl of HPLC solvent A (2.5% acetonitrile, 0.1% formic acid). A nano-scale reverse-phase HPLC capillary column was created by packing 5-μm C18 spherical silica beads into a fused silica capillary (100 μm inner diameter × 20 cm length) with a flame-drawn tip. After equilibrating the column, each sample was loaded via a Famos autosampler (LC Packings, San Francisco, CA) onto the column. A gradient was formed and peptides were eluted with increasing concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid).

As peptides eluted, they were subjected to electrospray ionization and then entered into an LTQ Velos ion trap mass spectrometer (Thermo Fisher, San Jose, CA). Peptides were detected, isolated and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences (and hence protein identity) were determined by matching the acquired fragmentation pattern with protein databases by the software program, SEQUEST (ver. 28) (Thermo Fisher). Enzyme specificity was set to partially tryptic with two missed cleavages. Modifications included carboxyamidomethyl (cysteines, fixed) and oxidation (methionine, variable). Mass tolerance was set to 2.0 for precursor ions and 1.0 for fragment ions. The database searched was the Human IPI databases version 3.6. The number of entries in the database was 160,900, which includes both the target (forward) and the decoy (reversed) human sequences. Spectral matches were filtered to contain < 1% FDR at the peptide level based on the target-decoy method. Finally, only tryptic matches were reported and spectral matches were manually examined. When peptides matched to multiple proteins, the peptide was assigned so that only the most logical protein was included (Occam’s razor). This same principle was used for isoforms when present in the database. The longest isoform was reported as the match. Supplementary Tables S3 and S4 contain the lists of both the peptides identified and the proteins identified during these analyses.

The mass spectrometry data from this publication, including raw files and search results, have been deposited to ProteomeXchange (www.proteomexchange.org) with identifier PXD001383 and DOI 10.6019/PXD001383.

**Data analysis and bioinformatics analysis**

We downloaded protein sequence from the UniProt Consortium. The function annotations were generated through the use of Ingenuity Pathway Analysis software (www.ingenuity.com). The heatmap and clustering were generated using Multi Experiment Viewer version 4.8.1 (TM4) and Heatmap Builder (Dr. Euan Ashley, Stanford University) software.

For MS data filtration, protein results from MS sequencing were converted to NCBI gene identifiers and searched for protein length. We reorganized the data to the format compatible to the SAINT program and used two-pool analysis, which recognized control group as a separate pool. We did not remove outlier datapoints. However, during the data analysis, we temporally removed the bait self-identification in the identification list before applying the SAINT algorithms and added them back after the data filtration. The separation of positive and negative distributions was considered for the scoring of low-count interactions or for division of spectra counts by the total spectra counts of each purification. The statistics used to assess accuracy and significance of measurements was referred to the SAINT algorithms, where \(X_c > 0.80\) was taken as the threshold required for the data quantification.

For interactomes generated by Cytoscape, we analysed the network and created custom styles, then applied yFiles organic layout or unweighted force-directed distributions with minor adjustments when necessary. The GO annotations and disease correlations were generated using the literature or non-self HCPs identified in our studies, weighted by the spectra counts and searched in the Knowledge Base provided by Ingenuity pathway software (Ingenuity Systems, www.ingenuity.com), which contains findings and annotations from multiple sources including the Gene Ontology database, to estimate the significance of these correlations.

**ChIP-sequencing and genomic data analysis**

ChIP-sequencing was performed in HEK293T cells with two biological replicates. Samples were sequenced using the Illumina Miseq, raw reads were mapped to human reference genome (hg19), and...
peaks were selected using MACS in Galaxy and annotated with PAVIS (Zhang et al., 2008; Langmead et al., 2009; Huang et al., 2013). Data are available in the ArrayExpress database www.ebi.ac.uk/arrayexpress (Rustici et al., 2013) under accession number E-MTAB-3120.

Analysis of the NFATC1-, CREB1- and ATF2-bound regions was performed using ChIP-seq data sets generated by the ENCODE consortium in GM12878 cells (ENCODE-Project Consortium, 2012). Data sets were analysed in Galaxy and mapped to genes using PAVIS with −5000 to +1000 TSS windows.

Western blotting and immunostaining

Whole-cell lysates were prepared by lysing cells with NETN buffer (20 mM Tris–HCl (pH 8.0), 100 mM NaCl, 1 mM EDTA and 0.5% Nonidet P-40) on ice for 30 min and then boiling in 2× Laemmli buffer. Lysates were subjected to SDS–PAGE followed by immunoblotting with antibodies against various proteins, including TP53, c-MYC (the same one used for MYC-tag WB), GFP, GAPDH, FOXL2 (Santa Cruz Biotechnology), FOXK1, FOXM1, FOXO3, JUN, MAX, MCV (for endogenous MYC blotting), NF-κB1, STAT3, TCF4, βTRCP, CUL1 (Cell Signalling), β-actin, FLAG (Sigma), RBP1, βTRCP2 (Abcam) and histone H3 (Upstate). The rabbit polyclonal anti-FOXO3 antibody was generated by immunizing rabbits (Cocalico Biologicals) with GST-FOXO3 fusion protein and affinity purified.

For immunostaining assays, cells cultured on coverslips were washed with 0.5% (v/v) Triton X-100 solution for 5 min. Coverslips were washed with PBS and immunostained with primary antibodies in 5% goat serum for 60 min. Cells were then washed and incubated with rhodamine- or FITC-conjugated secondary antibodies for 60 min, and nuclei were stained with 1 μg/ml 4′,6-diamidino-2-phenylindole (DAPI). Slides were mounted and visualized using a Nikon ECLIPSE E800 fluorescence microscope with a Nikon Plan Fluor 40× oil objective lens (numerical aperture 1.30) at room temperature. Cells were photographed using a Spot camera (Diagnostic Instruments) and analysed using Photoshop software (Adobe).

Supplementary information for this article is available online:
http://msb.embopress.org

Acknowledgements

We thank the Chen laboratory for their advice and technical assistance, Kathryn Aziz for assistance on animal work and Elizabeth Hess for editing the manuscript. XL is a recipient of a Computational Cancer Biology Training Program Fellowship supported by the Cancer Prevention and Research Institute of Texas and a Jeffrey Lee Cousins Fellowship in Lung Cancer Research. This work was supported in part with a startup fund from The University of Texas MD Anderson Cancer Center (J.C.), Era of Hope Scholar Research Award W81XWH-09-1-0409 from the U.S. Department of Defense (J.C.) and National Institutes of Health Grant RO1HG007538 (W.L.). The ChIP-sequencing analysis was supported by Solexa allowance provided by the Center for Cancer Epigenetics and Sequencing and Microarray Facility (Core grant CA016672 SMF), MD Anderson Cancer Center. The University of Texas MD Anderson Cancer Center is supported by the National Institutes of Health Core Grant CA016672.

Author contributions

XL, WW, WL, JQ, DH, RG and JC designed the experiments. XL, WW and JW performed experiments. XL, AM, DH and YX analysed data. XL and JC wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

cycle-dependent gene expression through an atypical chromatin binding mechanism. Mol Cell Biol 33: 227 – 236
Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER, Bakalarski CE, Zhao Z, Solimini N, Lenthal Y, Shiloh Y, Gygi SP, Elledge SJ

Xu Li et al Chromatin-associated and soluble human TF complexomes

Molecular Systems Biology

© 2015 The Authors

Molecular Systems Biology 11: 775 | 2015

15

108306-CS-EMBO-ChromatinReprint-Text.indd 43
16/03/2015 15:23
Chromatin-associated and soluble human TF complexomes

Xi Li et al

Molecular Systems Biology


License: This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.