

# SCIENTIFIC REPORTS



OPEN

## FOXA and master transcription factors recruit Mediator and Cohesin to the core transcriptional regulatory circuitry of cancer cells

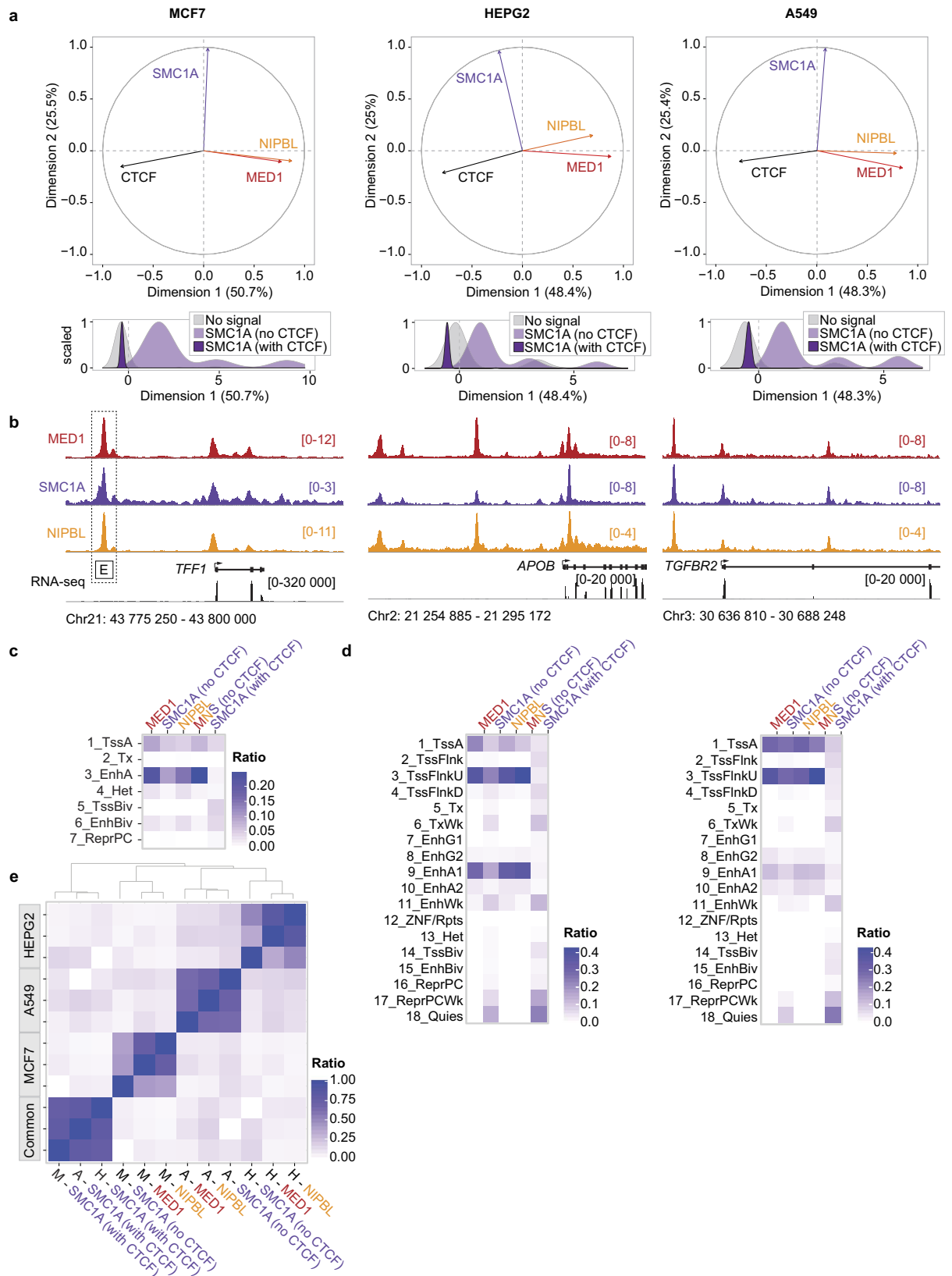
Received: 10 June 2016  
Accepted: 20 September 2016  
Published: 14 October 2016

Michèle Fournier<sup>1,2,\*</sup>, Gaëlle Bourriquen<sup>1,2,\*</sup>, Fabien C. Lamaze<sup>1,2,†</sup>, Maxime C. Côté<sup>1,2</sup>,  
Éric Fournier<sup>1,2</sup>, Charles Joly-Beauparlant<sup>2</sup>, Vicky Caron<sup>1,2</sup>, Stéphane Gobeil<sup>2,3</sup>, Arnaud Droit<sup>2,3</sup>  
& Steve Bilodeau<sup>1,2,4</sup>

Controlling the transcriptional program is essential to maintain the identity and the biological functions of a cell. The Mediator and Cohesin complexes have been established as central cofactors controlling the transcriptional program in normal cells. However, the distribution, recruitment and importance of these complexes in cancer cells have not been fully investigated. Here we show that FOXA and master transcription factors are part of the core transcriptional regulatory circuitry of cancer cells and are essential to recruit Mediator and Cohesin. Indeed, Mediator and Cohesin occupied the enhancer and promoter regions of actively transcribed genes and maintained the proliferation and colony forming potential. Through integration of publically available ChIP-Seq datasets, we predicted the core transcriptional regulatory circuitry of each cancer cell. Unexpectedly, for all cells investigated, the pioneer transcription factors FOXA1 and/or FOXA2 were identified in addition to cell-specific master transcription factors. Loss of both types of transcription factors phenocopied the loss of Mediator and Cohesin. Lastly, the master and pioneer transcription factors were essential to recruit Mediator and Cohesin to regulatory regions of actively transcribed genes. Our study proposes that maintenance of the cancer cell state is dependent on recruitment of Mediator and Cohesin through FOXA and master transcription factors.

Transcription factors enforce the cell-specific transcriptional program by binding regulatory elements dispersed throughout the genome<sup>1</sup>. Indeed, their unique ability to recognize specific DNA sequences makes them the central piece of cell state control. While hundreds of transcription factors are expressed in a cell, only a handful are essential to establish and maintain the transcriptional program<sup>2–4</sup>. A landmark demonstration established decades ago that expression of the transcription factor MYOD1 was alone sufficient to turn fibroblasts into myoblasts<sup>5</sup>. Since then, many transcription factors have been added to the list of the reprogramming transcription factors including the various combinations of OCT4, SOX2, KLF4 and MYC to force somatic cells into an induced pluripotent cell state<sup>6,7</sup>. Those dominant transcription factors enforcing a specific transcriptional program are typically part of the core transcriptional regulatory circuitry of a cell<sup>8</sup>. These core transcriptional regulatory circuitries include master transcription factors, like MYOD1 and OCT4, which are expressed at high levels relative to other transcription factors, are essential to cell state maintenance and positively regulate cell-type-specific genes in addition to their own expression<sup>8–10</sup>. Another class of transcription factors important for embryonic development and reprogramming are pioneer transcription factors<sup>11</sup>. These transcription factors, like FOXA1, possess the ability to bind DNA sequences wrapped around nucleosomes and positively or negatively influence the transcriptional program<sup>12–14</sup>. Overall, the intricate interplays between transcription factors define the transcriptional program and, therefore, the cellular state.

<sup>1</sup>Centre de Recherche sur le Cancer de l'Université Laval, Québec, Canada. <sup>2</sup>Centre de recherche du CHU de Québec – Université Laval, Québec, Canada. <sup>3</sup>Département de médecine moléculaire, Faculté de Médecine, Université Laval, Québec, Canada. <sup>4</sup>Département de biologie moléculaire, biochimie médicale et pathologie, Faculté de Médecine, Université Laval, Québec, Canada. <sup>†</sup>Present address: Ontario Institute for Cancer Research, Toronto, Canada. <sup>\*</sup>These authors contributed equally to this work. Correspondence and requests for materials should be addressed to S.B. (email: steve.bilodeau@crchuq.ulaval.ca)



**Figure 1. Mediator, Cohesin and NIPBL co-localize at regulatory regions of actively transcribed genes in cancer cells.** (a) Principal component analysis (PCA) of the distribution of MED1-, SMC1A-, NIPBL- and CTCF-occupied regions. Top – PCA variable factor map of the MED1, SMC1A, NIPBL and CTCF variables on the first two components showing a positive correlation between NIPBL and MED1. In addition, dimension 1 captures the negative correlation between the transcription factor CTCF and MED1/NIPBL-occupied regions. Bottom – Density plot of the genomic region coordinates on the dimension 1 showing a split distribution of

SMC1A. SMC1A is associated with CTCF (negative values) or MED1 and NIPBL (positive values). The y axes are normalized to 1 while the x axes represent arbitrary units. **(b)** Occupancy profiles of Mediator (MED1), Cohesin (SMC1A) and NIPBL showing colocalization at regulatory regions. The *TF1*, *APOP* and *TGFBR2* loci are displayed for MCF7, HEPG2 and A549 respectively. The characterized enhancer region of *TF1* (box labeled E) is indicated<sup>60</sup>. The scales of ChIP-Seq and RNA-Seq profiles are displayed in reads per million. **(c)** Overlaps between MED1, SMC1A (no CTCF), SMC1A (with CTCF), the regions co-occupied by MED1-SMC1A-NIPBL and the functional annotation of the MCF7 genome (see Supplementary Information). MED1, SMC1A (no CTCF) and NIPBL occupy active TSS (1\_TssA) and active enhancers (3\_EnhA), but not inactive regions. The color scale indicates the ratio of overlap. **(d)** Overlaps between MED1, SMC1A (no CTCF), SMC1A (with CTCF), the regions co-occupied by MED1-SMC1A-NIPBL and the functional annotation of the HEPG2 (left) and A549 (right) genomes using the ChromHMM 18-state model<sup>32,33</sup>. MED1, SMC1A (no CTCF) and NIPBL occupy active promoters (1\_TssA and 3\_TssFlnkU) and active enhancer regions (8\_EnhG2, 9\_EnhA1 and 10\_EnhA2), but not inactive regions in both HEPG2 and A549 cells (see Supplementary Information for abbreviations). The color scales indicate the ratio of overlap. **(e)** Heat map showing the colocalization frequencies of MED1, SMC1A and NIPBL enriched regions in MCF7, HEPG2 and A549 cells. The color scale reflects the colocalization of each pair of regulators. Regulators were clustered along both axes based on the similarity in their correlation with each other.

Transcription factors exert their control on gene expression through recruitment of a wide variety of cofactors, chromatin modifiers and regulators of the chromosome structure<sup>1,8,15</sup>. Ultimately, transcription factors are important to recruit the RNA Polymerase II (Pol II) machinery and to regulate elongation<sup>16,17</sup>. Bridging the transcription factors with Pol II is the role of the coactivator complex Mediator<sup>18</sup>. Indeed, Mediator serves as a central scaffold in the pre-initiation complex to regulate Pol II activity. Along with the Mediator complex, the Cohesin complex, which is essential to maintain sister chromatids cohesion during cell division<sup>19</sup>, and the cohesin loading factor NIPBL are essential to form cell-type-specific connections between enhancers and promoters<sup>20</sup>. Other transcription factors, like MYC, promotes Pol II pause-release by recruiting the P-TEFb complex at active promoters<sup>21</sup>. Therefore, a combination of transcription factors is required at enhancer and promoter regions to specify the transcriptional program of a cell. Identifying the combinations of transcription factors governing the transcriptional program of each cell provides the core information to control cell state.

The gene expression program controls normal and aberrant biological functions of a cell. Indeed, to reach a disease state, cells often modify their transcriptional program to acquire new identities and functions<sup>8</sup>. Interestingly, modulation of regulatory regions is emerging as a key feature of many human diseases<sup>22,23</sup>. For example, the vast majority of single nucleotide polymorphisms (SNPs) associated with cancer are found in the non-coding fraction of the genome with a large subset inside enhancer regions<sup>24</sup>. Furthermore, modulation of enhancer regions is a promising cancer treatment<sup>25,26</sup>. Additionally, mutations in transcriptional regulators occupying enhancer regions, including transcription factors and cofactors, are causative of many diseases<sup>27</sup>. Among them, mutations in Mediator, Cohesin and the Cohesin loader NIPBL have been associated to developmental syndromes as well as different forms of cancer<sup>28,29</sup>. While it is not clear how these mutations lead to a disease state, they point to the importance of proper control of regulatory regions to maintain a normal cell state.

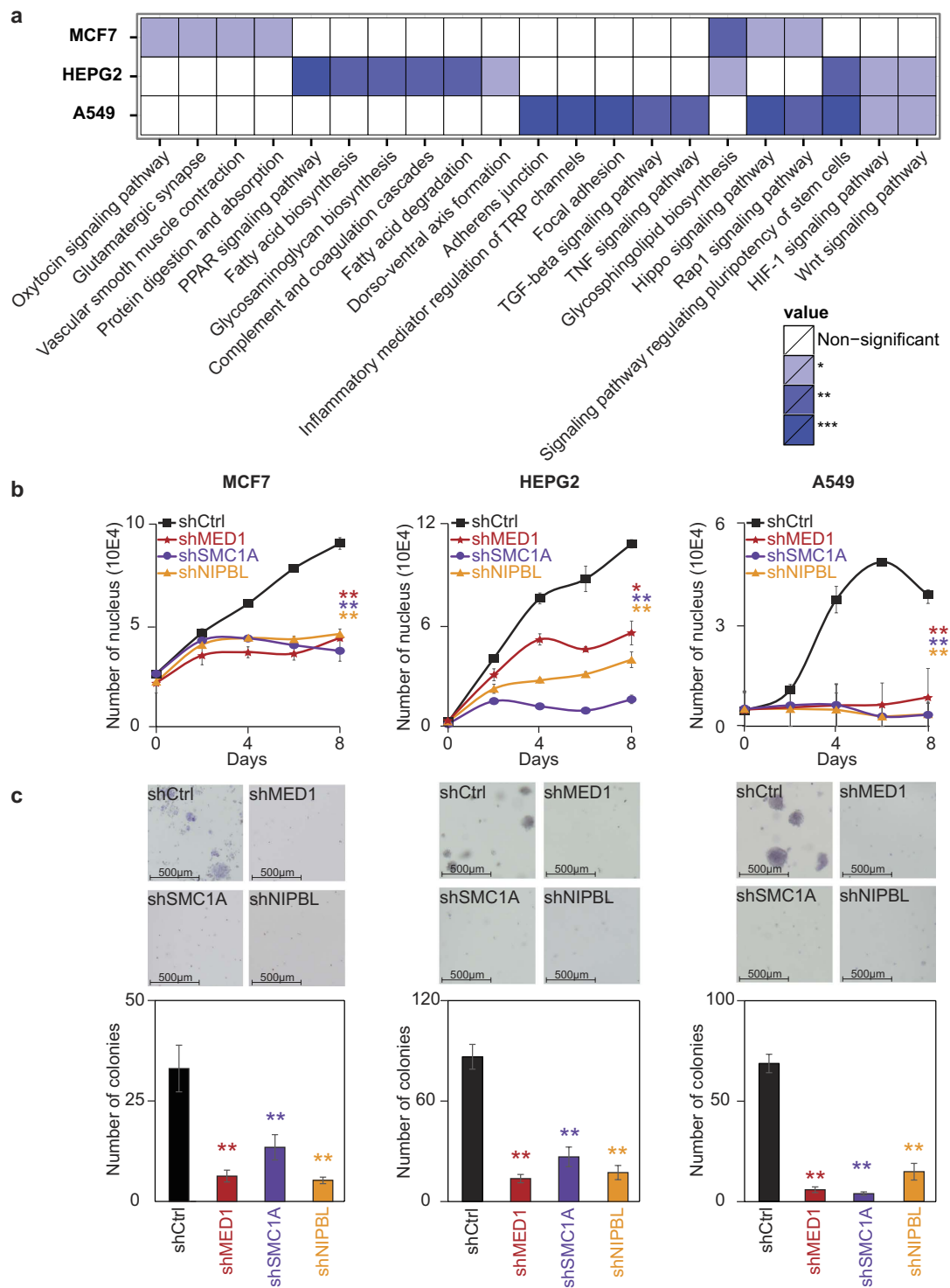
Here we show that FOXA pioneer transcription factors are, with master transcription factors, the core transcriptional regulatory circuitry of cancer cells. These transcription factors are essential to recruit the Mediator and Cohesin complexes to regulatory regions of actively transcribed genes. Taken together, our results suggest that Mediator and Cohesin are connected to the core transcriptional regulatory circuitry through FOXA and master transcription factors to maintain the cancer cell state.

## Results

### Mediator and Cohesin occupy regulatory regions of actively transcribed genes in cancer cells.

Maintenance of the transcriptional program depends on transcription factors and cofactors including Mediator and Cohesin. Indeed, Mediator and Cohesin occupy enhancer regions with master transcription factors to control the transcriptional program and identity of embryonic stem cells<sup>20</sup>. To define the contribution of Mediator and Cohesin to the regulation of the cancer transcriptional program, we established the genome-wide occupancy of MED1 (Mediator), SMC1A (Cohesin) and the cohesin loader NIPBL. Three cancer cell lines were profiled using chromatin immunoprecipitation coupled with massively parallel DNA sequencing (ChIP-Seq): MCF7 (breast), HEPG2 (liver) and A549 (lung). Principal component analyses (PCA) confirmed that the distribution profiles of NIPBL and MED1 were positively correlated ( $R = 0.53, 0.45$  and  $0.53$  for MCF7, HEPG2 and A549, respectively, all  $p$ -values  $< 10e-100$ ) across the genome of each cell type (Fig. 1a) in accordance with previous reports<sup>20</sup>. Furthermore, Cohesin, which is known to occupy regions with NIPBL and Mediator, but also the transcription factor CTCF<sup>30,31</sup>, showed a bimodal distribution in all three cell lines (Fig. 1a). Therefore, for subsequent analyses, SMC1A-occupied regions were divided into CTCF-occupied (with CTCF) and not (no CTCF). Detailed examination of the ChIP-Seq density profiles confirmed the presence of Mediator, Cohesin and NIPBL at the transcription start site (TSS) and well-characterized enhancers regions of genes associated with breast, liver and lung functions (Fig. 1b). These results suggest that similar to observations made in primary cells<sup>20</sup>, Mediator and Cohesin co-occupy the genome at regulatory regions of actively transcribed genes in cancer cells.

The mammalian genome is segmented into functional elements orchestrating transcriptional events and gene expression<sup>32</sup>. To extend the occupancy profiles of Mediator, Cohesin and NIPBL genome-wide, we compared the ChIP-Seq data with predicted functional elements. For MCF7 cells, histone marks were combined (see Supplementary Information) to define the different chromatin states while an 18-state model built using



**Figure 2. Mediator, Cohesin and NIPBL are essential for maintenance of cancer cells properties.** (a) Heat map representation of KEGG pathways enriched for genes occupied by MED1, SMC1A and NIPBL in MCF7, HEPG2 and A549 cells (Fisher's exact test, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ). (b) Growth curve of MCF7, HEPG2 and A549 following decreased expression of MED1, SMC1A and NIPBL showing a global negative effect on cell growth. Each line represents the average of a minimum of two biological replicates for two independent shRNA constructs. The p-values were calculated by comparing with the shRNA controls at day 8 (paired t-test, \* $p < 0.05$  and \*\* $p < 0.01$ ). (c) Soft agar colony formation assays with decreased levels of MED1, SMC1A and NIPBL showing a loss of colony formation potential. Colonies were quantified with the ImageJ software using 4 images per well. The bar graph represents the average of a minimum of two biological replicates for two independent shRNA constructs. The p-values were calculated by comparing with the shRNA controls (paired t-test, \* $p < 0.05$  and \*\* $p < 0.01$ ).

ChromHMM was available for A549 and HEPG2<sup>32,33</sup>. MED1, SMC1A and NIPBL occupied active enhancers and promoters in all three cell types (Fig. 1c,d). Lowly transcribed regions including heterochromatin and polycomb-occupied sites were not enriched. While the functional elements occupied by Mediator, Cohesin and NIPBL are similar in the three cancer cells investigated, the coordinates of these elements are different. Indeed, regions occupied by MED1, SMC1A (no CTCF) and NIPBL in MCF7 were different from those found in HEPG2 and A549 (Fig. 1e). However, regions occupied by CTCF and SMC1A (SMC1A with CTCF) were highly redundant between cell lines in accordance with previous reports<sup>34</sup>. Taken together, these results suggest a conserved function of Mediator, Cohesin and NIPBL in the control of regulatory elements throughout the genome of cancer cells.

**Proliferation of cancer cells depends on Mediator and Cohesin.** As in normal cells, the gene networks controlling the identity and functions of each cancer cell is unique. To determine the nature of the genes occupied by Mediator, Cohesin and NIPBL, we investigated the associated gene networks in MCF7, HEPG2 and A549. KEGG pathways analyses revealed the association of MED1, SMC1A and NIPBL with genes reflecting the normal and disease functions of the tissues (Fig. 2a and Supplementary Table S1). For example, in MCF7 (breast) cells, genes implicated in the oxytocin signaling pathway were enriched. For HEPG2 (liver) cells, genes associated with hepatic functions like PPAR signaling, complement and coagulation cascades as well as fatty acids metabolism were strongly enriched. In A549 (lung) cells, pathways associated with cell adhesion and TGF-beta signaling were identified. Interestingly, among the direct target genes, many known oncogenes including FGFR2, MAX and PIK3R1 in MCF7, PDGFB, PPARG and HIF1A in HEPG2 and MYC, MET and EGFR in A549 were identified. These observations suggest a role of Mediator and Cohesin in the maintenance of cancer cell properties through a cell-specific network of genes.

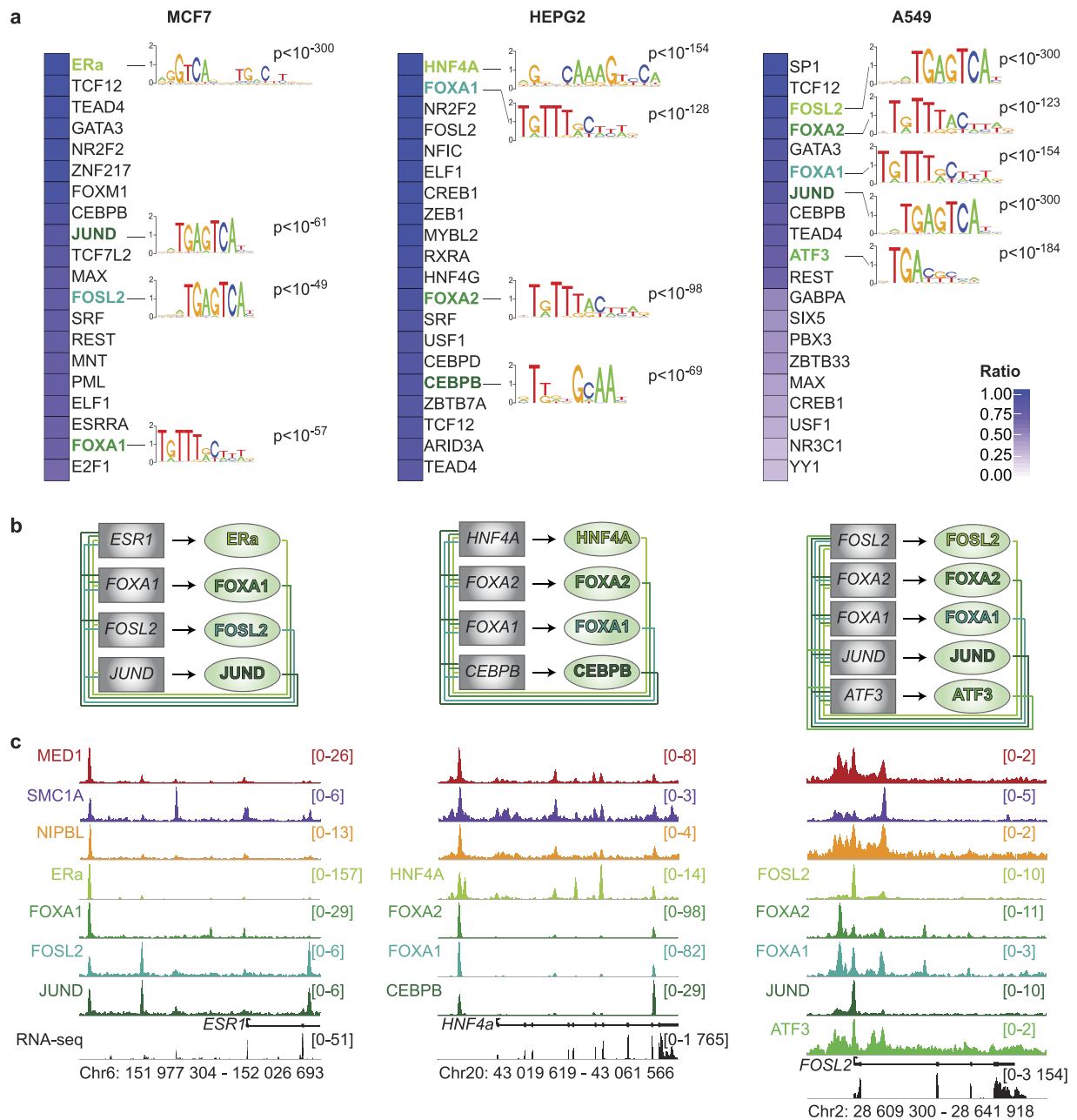
Whether or not cancer cells are dependent on Mediator, Cohesin and NIPBL to maintain their proliferative state is not known. Therefore, we quantified the cellular response of MCF7, HEPG2 and A549 with proliferation and soft-agar colony formation assays in MED1, SMC1A and NIPBL loss-of-function experiments. For each target, two independent shRNAs were validated for efficiency by Western blot and used in functional assays (Supplementary Fig. S1). In all three cell lines, similar results were obtained (Fig. 2b,c). Indeed, loss of MED1, SMC1A or NIPBL led to a global decrease in proliferation (Fig. 2b). During the 7-day experiment, cells were not dividing, but were maintained. To determine if these changes affected the colony formation potential of cancer cells, we used soft-agar colony formation assays. Loss of MED1, SMC1A and NIPBL reduced the colony formation potential of MCF7, HEPG2 and A549 (Fig. 2c). Indeed, all cells transduced with control shRNAs produced colonies while their number were significantly reduced with decreasing levels of MED1, SMC1A and NIPBL. Altogether, these results support a role of Mediator and Cohesin in the proliferation of cancer cells.

### Mediator and Cohesin connect with the core transcriptional regulatory circuitry of cancer cells.

Cofactors are recruited to regulatory elements by transcription factors. To determine the transcription factor(s) recruiting Mediator and Cohesin, we collected publicly available ChIP-Seq datasets for transcription factors (Supplementary Table S2) and computed pairwise genome-wide relationships for each cell line. Transcription factors were ranked based on their percentage of overlap with MED1, SMC1A and NIPBL across the genome of each cancer cell (Fig. 3a). The top 20 was selected for further filtering. Transcription factors are recruited to functional elements in *cis* through recognition of a DNA sequence but also in *trans* through protein-protein interactions<sup>35</sup>. To refine the list of candidates, we performed an *in silico* enrichment of the binding motifs found in the HOCOMOCO database in regions occupied by MED1, SMC1A and NIPBL. The 20 most frequent DNA motifs found in regions occupied by MED1, SMC1A and NIPBL were cross-referenced with the list of 20 co-occupying transcription factors (Fig. 3a). In each cell line, multiple transcription factors were identified. In MCF7, ERa, FOXA1, FOSL2 and JUND were found, while HNF4A, FOXA1, FOXA2 and CEBPB were identified in HEPG2 and FOSL2, FOXA1, FOXA2, JUND and ATF3 in A549. These results suggest that a combination of transcription factors is responsible for the recruitment of Mediator and Cohesin at regulatory regions.

Whether the identified transcription factors formed a core transcriptional regulatory circuitry in each cancer cell is unknown. To determine the connections between the transcription factors and their corresponding genes, we built the transcriptional network using the ChIP-Seq data for each transcription factor (Fig. 3b and Supplementary Fig. S2). Interestingly, the transcription factors were interconnected in all three cell types (Fig. 3b). In MCF7, ERa was found at the promoter of *ESR1*, *FOSL2*, *FOXA1* and *JUND*. In return, FOSL2, FOXA1 and JUND occupied the promoter of *ESR1* (Fig. 3c and Supplementary Fig. S2). The connections for MCF7, HEPG2 and A549 are illustrated in Fig. 3b,c. These results suggest that the transcription factors underlying the recruitment of Mediator and Cohesin are part of the core transcriptional regulatory circuitry of cancer cells.

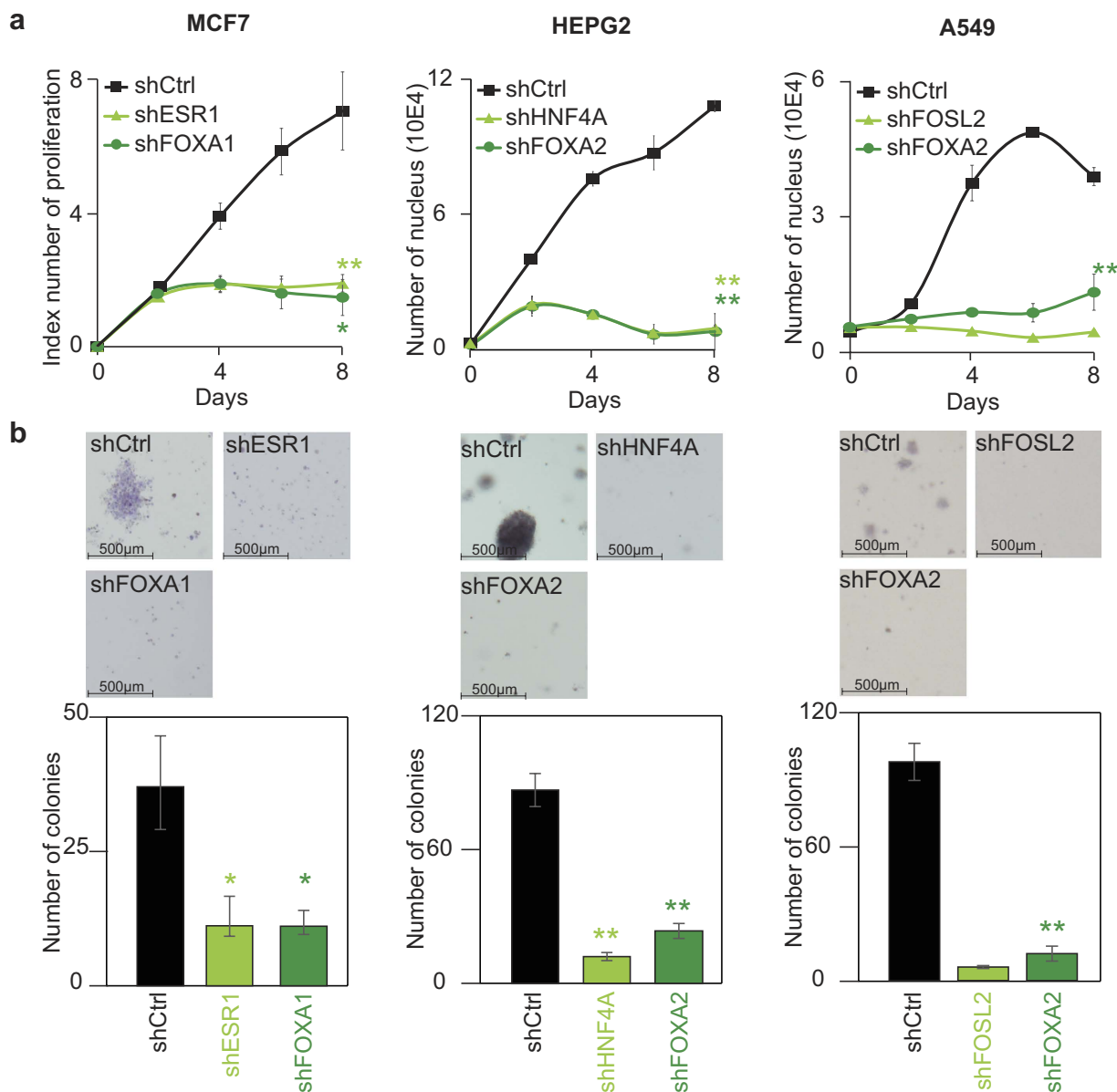
**FOXA and master transcription factors are essential to cancer cell maintenance.** To prioritize candidates for functional studies, we wanted to identify the central and most important transcription factors in each network. Interestingly, many of the identified transcription factors are or are presumed to be master and pioneer transcription factors. For example, ERa and HNF4A were implicated with the core transcriptional regulatory circuitry of breast and liver cells<sup>14,36</sup>. A shared characteristic of master transcription factors is that they form positive autoregulatory loops, regulating their own gene<sup>2</sup>. Therefore, we identified the transcription factors occupying their own regulatory regions with Mediator, Cohesin and NIPBL. A limited number of transcription factors fulfilled that characteristic and they became our leading candidates for controlling cancer cells. These transcription factors included ERa in MCF7; HNF4A and FOXA2 in HEPG2 and FOSL2 in A549 (Fig. 3b,c). Since, the FOXA family of pioneer transcription factors was identified in all cell lines, FOXA1 in MCF7 and FOXA2 in A549 were also included for further characterization. It is to be noted that many of these transcription factors are



**Figure 3. The core circuitry of cancer cells connect with Mediator and Cohesin.** (a) Percentage of overlapping regions occupied by MED1, SMC1A and NIPBL with transcription factors. The best 20 overlaps with transcription factors for each cell type are presented. The color scale indicates the percentage of overlap. The DNA binding motif logos found in the regions occupied by MED1, SMC1A and NIPBL are indicated next to their associated transcription factor. The motifs were found in the top 20 of the most enriched position-weight matrices. The associated p-values are indicated. (b) Core transcriptional regulatory circuitry of MCF7, HEPG2 and A549 cells. The transcription factors identified in (a) are interconnected. The box represents the transcription factor gene loci while the oval represents the protein identified in (a). Each line represents the interaction of a transcription factor with the indicated gene loci (−10 kb to the end of the gene). (c) Occupancy profiles of Mediator (MED1), Cohesin (SMC1A) and NIPBL in addition to the core transcription factors identified in (a) showing colocalization. The *ESR1*, *HNF4a* and *FOSL2* loci are displayed for MCF7, HEPG2 and A549 respectively. The scales of ChIP-Seq profiles are displayed in reads per million.

associated with Mediator and Cohesin in various cell types. Indeed, ERa and HNF4A were shown to interact with Mediator<sup>37–39</sup>. In addition, ERa was co-purified with Cohesin subunits<sup>40</sup>. Furthermore, FOXA1 and FOXA2 were found to recruit coactivators and interact with other transcription factors at functional enhancers<sup>35,41,42</sup>.

If the identified transcription factors are essential to the function of Mediator, Cohesin and NIPBL, they should phenocopy each other in functional loss-of-function assays. Indeed, we have shown that Mediator and Cohesin are essential to the proliferation of the cancer cells (Fig. 2). Therefore, we used the same cell proliferation and

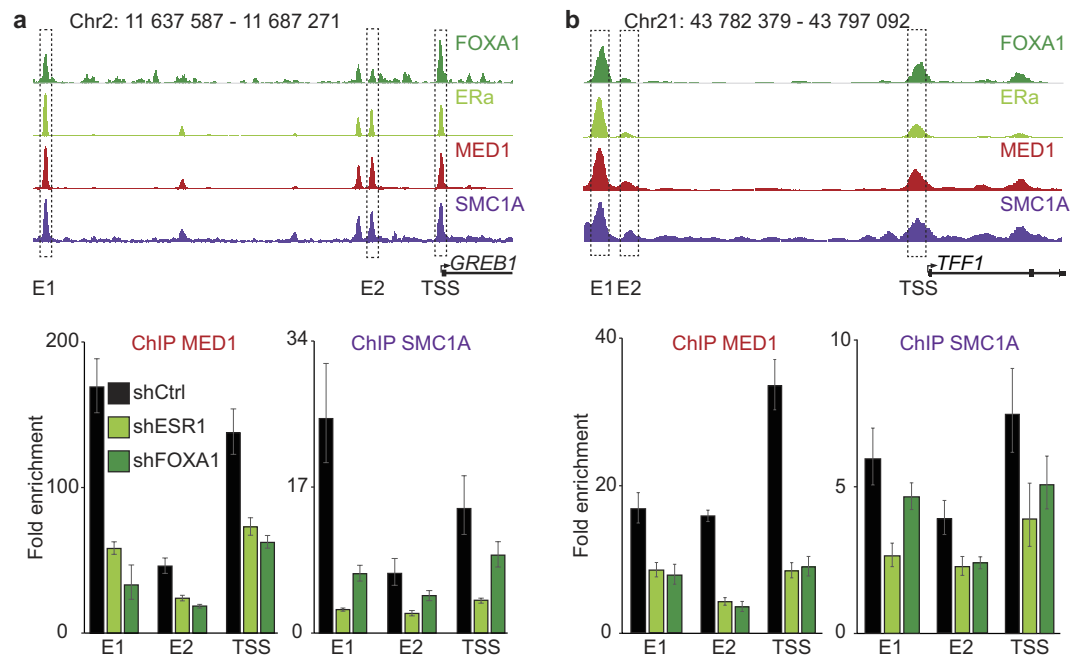


**Figure 4. Depletion of the core transcription factors phenocopies the loss of Mediator and Cohesin.**

(a) Growth curve of MCF7, HEPG2 and A549 following decreased expression of ER $\alpha$  and FOXA1 in MCF7, HNF4A and FOXA2 in HEPG2 and FOSL2 and FOXA2 in A549. The results show a global negative effect on cell growth. Each line represents the average of a minimum of two biological replicates for two independent shRNA constructs with the exception of FOSL2 where one shRNA construct was used. The p-values were calculated by comparing with the shRNA controls at day 8 (paired t-test, \* $p < 0.05$  and \*\* $p < 0.01$ ). (b) Soft agar colony formation assays showing a loss of colony formation potential with decreased expression of core transcription factors. Colonies were quantified with the ImageJ software using 4 images per well. The bar graph represents the average of a minimum of two biological replicates for two independent shRNA constructs with the exception of FOSL2 where one shRNA construct was used. The p-values were calculated by comparing with the shRNA controls (paired t-test, \* $p < 0.05$  and \*\* $p < 0.01$ ).

colony formation assays with validated shRNAs targeting each identified transcription factors (Supplementary Fig. S3). In all cases, loss of the transcription factor reduced the proliferation and the colony formation potential of cancer cells (Fig. 4a,b). Indeed, lower levels of ER $\alpha$  and FOXA1 in MCF7, HNF4A and FOXA2 in HEPG2 and FOSL2 and FOXA2 in A549 reduced the proliferation rate (Fig. 4a) and the colony formation potential (Fig. 4b). The results demonstrate that the identified core transcriptional regulatory circuitry in each cell is essential to maintain cancer cells.

**FOXA and master transcription factors recruit Mediator and Cohesin.** Integration of the ChIP-Seq results and the functional experiments suggest that Mediator and Cohesin are recruited to regulatory regions of



**Figure 5. FOXA and master transcription factors are essential to recruit Mediator and Cohesin.** Depletion of either the master transcription factor ERa or the pioneer transcription factor FOXA1 in MCF7 cells decreases MED1 and SMC1 recruitment at regulatory elements of the *TFF1* and *GREB1* genes in MCF7. Amplified enhancer (E1 and E2) and TSS regions are indicated by boxes. Enrichment fold were calculated relative to control regions without MED1 or SMC1 signal. The bar graphs display the enrichment fold of a representative experiment. Error bars show the standard error of the mean for technical qPCR triplicates.

actively transcribed genes by the core transcriptional regulatory circuitry. To functionally validate our results, we focused our analysis on MCF7 cells. Pioneer transcription factors like FOXA1 are known to bind the condensed chromatin prior to transcriptional activation<sup>12</sup>. On the other hand, master transcription factors like ERa tend to occupy nucleosome-free regions<sup>12</sup>. We entertained two models: (1) an all-or-nothing model where the presence of many transcription factors is required to stabilize large cofactors complexes like Mediator and Cohesin and (2) a sequential model where the transcription factors and cofactors are orderly recruited. To determine the best model, we proceeded with loss-of-function experiments for ERa and FOXA1 followed by ChIP of MED1 and SMC1A (Fig. 5). The same shRNA as previously described (Fig. 4 and Supplementary Fig. S3) were used. Molecularly, loss-of-function of the master transcription factor ERa and the pioneer transcription factor FOXA1 led to decreased recruitment of MED1 and SMC1A at enhancers and promoters of active genes *GREB1* (Fig. 5a and Supplementary Fig. S4) and *TFF1* (Fig. 5b and Supplementary Fig. S4). The total protein levels of the cofactors were not affected by the decreased expression of ERa and FOXA1 (Supplementary Fig. S4a). Therefore, we conclude that a combination of transcription factors is required to recruit and stabilize large protein complexes like Mediator and Cohesin to regulatory regions of actively transcribed regions.

## Discussion

While each cell presents unique biological characteristics, our results propose a shared common ground in transcriptional regulation. Indeed, Mediator, Cohesin and NIPBL occupy enhancer and promoter regions of actively transcribed genes in all primary<sup>20</sup> and cancer cells (Fig. 1) tested so far. In addition, cancer cells are no exception to the simple rule stating that transcription factors provide cell specificity to the transcriptional program. Our results establish a model where a combination of FOXA and master transcription factors are part of the core transcriptional regulatory circuitry and essential to maintain the cancer cell properties (Figs 3 and 4). Regulatory regions are occupied by dozens and potentially more transcription factors in each cell<sup>43</sup>. Surprisingly, loss of a single transcription factor affected the recruitment of Mediator and Cohesin (Fig. 5) pointing toward a combinatorial effect in the stabilization and function of Mediator and Cohesin. This observation is corroborated by most reprogramming experiments where a combination of transcription factors is required to reach a new cell state<sup>8</sup>. However, we cannot exclude the possibility that the similar phenotypes observed in Mediator, Cohesin and transcription factor loss-of-function experiments (Figs 2 and 4) could be attributable to independent mechanisms including cell cycle defects. Therefore, our results suggest a non-redundant role of transcription factors at regulatory regions and postulate that FOXA and master transcription factors are required to stabilize large cofactor complexes like Mediator and Cohesin.

The identification of FOXA1 or FOXA2 as essential transcription factors in cancer cells (Figs 3 and 4) proposes the existence of a conserved role of the FOXA family in the core transcriptional regulatory circuitry. As pioneer transcription factors, FOXA1 and FOXA2 recognize their DNA motif even when the DNA is wrapped around nucleosomes<sup>12</sup>. This property could lead the way for master transcription factors to activate gene transcription.



In hormone-dependent cancers, FOXA1 is essential to reprogram the genome-wide occupancy of the estrogen and androgen receptors and rewire the transcriptional program<sup>44,45</sup>. These observations suggest that FOXA transcription factors have the ability to direct binding of master transcription factors and to control the core transcriptional regulatory circuitry. Considering this central role of FOXA transcription factors, the successful development of specific inhibitors could lead to a new class of molecules to target cancer cells.

While key transcriptional regulators in breast adenocarcinoma (MCF7) and hepatocarcinoma (HEPG2) are known<sup>14,36,45</sup>, transcription factors maintaining lung adenocarcinoma (A549) are not well established. Integration of the genome-wide positioning of Mediator and Cohesin provide a valuable tool to predict key transcription factors in cell state maintenance. Indeed, our results suggest that the core transcriptional regulatory circuitry of A549 lung cancer cells is centered on FOSL2 and FOXA2 (Figs 3 and 4). Interestingly, both FOXA2 and FOSL2 are suspected to play major role in lung cancer progression. Indeed, FOSL2 is important to integrate the signaling through the TGF-beta/SMAD pathway and promote cellular migration of non-small cell lung adenocarcinomas<sup>46</sup>. Furthermore, decreased expression of FOXA2 following TGF-beta treatment leads to epithelial-to-mesenchymal transition<sup>47</sup>. Master transcription factors have been shown to determine the response to TGF-beta in various tissues through SMAD recruitment at regulatory regions<sup>48</sup>. One may speculate that decreased levels of FOXA2 allows the rewiring of the core transcriptional regulatory circuitry and free up FOSL2 for a role during the epithelial-to-mesenchymal transition. Therefore, the core transcriptional regulatory circuitry containing FOSL2 and FOXA2 could control the balance between maintenance of a lung epithelial *versus* mesenchymal phenotype.

The possibility of a general transcriptional mechanism centered on Mediator and Cohesin controlling the transcriptional program of normal and cancer cells raises questions regarding many human diseases. In addition to various types of cancers, genetic alterations in Mediator and Cohesin have been observed in developmental syndromes, neurological disorders and metabolic diseases<sup>29,49</sup>. For example, mutations of the cohesin subunit *STAG2* are frequent in various cancers<sup>50</sup>. Interestingly, evidence of aneuploidy are questioned and a transcription-dependent function is favored to explain the contribution of the Cohesin mutations toward cancer<sup>28,51</sup>. Similarly, *MED12* is mutated at high frequencies in uterine leiomyomas and prostate cancers<sup>52,53</sup>. However, in uterine leiomyomas, all *MED12* mutations are found in exon 2 emphasizing alteration of a specific function of the Mediator subunit<sup>52</sup>. Additionally, during embryonic development, complete loss of Mediator and Cohesin are severely detrimental while missense mutations are less deleterious, but with complex tissue-specific phenotypes<sup>29,49</sup>. Taken together, these observations suggest that Mediator and Cohesin are potentially involved in a tissue-specific rather than a general transcriptional mechanism. Furthermore, the absence of genomic abnormalities in patients with Mediator and Cohesin mutations minimizes the possibility of cell cycle defects as the primary cause of these diseases. Therefore, we propose that modulation of the activity of the core transcriptional program is the central theme of many human diseases.

If transcriptional regulation is the primary cellular function of Mediator and Cohesin, their importance for other diseases is likely underestimated. Indeed, modulating the activity of regulatory regions is suspected to be an important mechanism leading to human diseases as SNPs are common in enhancer regions<sup>24,25</sup>. Variations in the sequence of an enhancer or promoter regions affect the binding affinity of the residing transcription factors<sup>54</sup> which will result in differential recruitment of their associated cofactors including Mediator and Cohesin. Further experiments will demonstrate the importance of Mediator and Cohesin recruitment through FOXA and master transcription factors for many other human diseases.

## Methods

Full materials and methods are available in Supplementary Information.

**Cell Culture.** MCF7 (ATCC, HTB-22) and HEPG2 (ATCC, HB-8065) were grown in DMEM (Gibco, 11965–092). A549 (ATCC, CCL-185) were grown in F12K medium (Gibco, 21127022). Culture medium were supplemented with 10% fetal bovine serum (Invitrogen, qualified 12483020), 100  $\mu$ M MEM nonessential amino acids (Cellgro, 25–0250), 2 mM L-glutamine (Gibco, 25030–081), 100 U/ml penicillin, 100  $\mu$ g/ml streptomycin (Gibco, 15170–063).

For ChIP experiments, MCF7 were kept the last 3 days into DMEM w/o phenol-red (Gibco, 31053–028) supplemented with 5% of Charcoal/Dextran treated FBS (Hyclone, AVH78911), 100  $\mu$ M MEM nonessential amino acids (Cellgro, 25–0250), 2 mM L-glutamine (Gibco, 25030–081) and treated with 100 nM of Beta-Estradiol (SIGMA, E8875) for two hours.

**Chromatin Immunoprecipitation.** ChIP-Seq were performed in biological duplicates as previously described<sup>20,55</sup>. Briefly, cells were crosslinked with 1% formaldehyde. Sonicated DNA fragments were immunoprecipitated with antibodies directed against MED1 (Bethyl Laboratories, A300-793A), SMC1A (Bethyl Laboratories, A300-055A), NIPBL (Bethyl Laboratories, A301-779A) and ERa (Santa Cruz, sc-543x). ChIP-seq libraries and DNA sequencing were performed by the McGill University and G enome Qu ebec Innovation Centre.

For single gene analysis, primers were designed in the predicted enhancers and promoters regions of *TFF1* and *GREB1* loci. For the *TFF1* locus, primers GAGATGATGCCACCGTACAC & CCCTCACTCACTTTGAAGCA were used for enhancer 1 (E1), ATCCAGTCCAGGAAGGAGGT & GTCAAGAGAGGAGGCTGTGG for enhancer 2 (E2) and CCGAGTCAGGGATGAGAGG & GGCCTCCTTAGGCAAATGTT for the TSS. For the *GREB1* locus, primers GAGCTGACCTTGTGGTAGGC & CAGGGGCTGACAACTGAAAT were used for enhancer 1 (E1), GGGATATGGCTTGTCCATTGT & CATGACACCAGGACCGTAAAG for enhancer 2 (E2) and ACCCAGCAAAACACTTCAGG & ATTCAGCAGTAGCCCTTCCA for the TSS. Two negative control regions (ATGTCAGGCCCATGAACGAT & GCATTCATGGAGTCCAGGCTTT and AGGACCTGCAGCAAACAGAA & TGCTACATGGGCTAGTGTGCT) were used to calculate fold enrichments.

**Bioinformatic analyses.** All ChIP-Seq reads (Supplementary Table S2) were aligned to the hg19 build of the human genome using the ChIP-Seq pipeline developed at McGill University and Génome Québec Innovation Centre ([https://bitbucket.org/mugqic/mugqic\\_pipeline](https://bitbucket.org/mugqic/mugqic_pipeline)). BigWig files were generated using Samtools 1.2<sup>56</sup>, bedtools 2.17.0<sup>57</sup> and wigToBigWig (developed by the ENCODE team) with normalized count in millions of reads and a smoothing parameter of 225bp. Enriched regions were annotated using the UCSC database R package TxDb.Hsapiens.UCSC.hg19. KEGG pathways enrichments were assessed by hypergeometric tests. *In silico* motif searches were performed with the HOCOMOCO database using the PWMEnrich package. For overlaps with chromatin states and between ChIP-Seq datasets, the GenomicRanges R package was used.

**Lentiviral shRNAs.** Plasmid catalog numbers and shRNA sequences are provided in Supplementary Table S3. Lentiviruses were produced as previously described<sup>58</sup>. Cells were transduced with the lentiviruses for 24 h before puromycin selection for the indicated time. Validation of efficiency were performed by Western blot (Supplementary Figs 1 and 3) using MED1 (Bethyl A300-793A, 1:5000), SMC1A (Bethyl, A300-055A, 1:5000), NIPBL (Bethyl, A301-779A, 1:1000), ERa (Santa Cruz, sc-543x, 1:5000), FOXA1 (Santa Cruz, sc-101058 1:1000), FOXA2 (Abnova, 89-019-034, 1:1000), HNF4a (Santa Cruz, sc-8987x, 1:1000), FOSL2 (Santa Cruz, sc-604x, 1:1000), GAPDH (Pierce, PIMA515738, 1:25000) and Vinculin (SIGMA, V9131, 1:50000).

**Proliferation and colony formation assays.** For proliferation and soft agar colony formation assays, cells were transduced with two shRNA per target with the exception of FOSL2 for which only one shRNA was efficiently knocking down all isoforms (Supplementary Fig. S3). Briefly, cells infected with lentiviral particles were selected for 24 hours prior to seeding. For proliferation assays, cell counts were measured using the Cytation 5 cell Imaging Multi-Mode Reader (BioTek Instruments) or the xCELLigence Real-Time Cell Analyzer (RTCA) instrument (ESBE Scientific and ACEA Biosciences) up to 8 days following seeding. Soft agar colony formation assays were performed as previously described<sup>59</sup>. Cells were maintained for 13 days (A549 and HEPG2) or 28 days (MCF7). Colonies were quantified using the ImageJ 1.46r software (NIH).

## References

- Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
- Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567–582 (2008).
- Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587–594 (2009).
- Morris, S. A. & Daley, G. Q. A blueprint for engineering cell fate: current technologies to reprogram cell identity. *Cell Res.* **23**, 33–48 (2013).
- Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Hochedlinger, K. & Jaenisch, R. Induced Pluripotency and Epigenetic Reprogramming. *Cold Spring Harb. Perspect. Biol.* **7**, 1–25 (2015).
- Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251 (2013).
- Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
- Thayer, M. J. *et al.* Positive autoregulation of the myogenic determination gene MyoD1. *Cell* **58**, 241–248 (1989).
- Zaret, K. S. & Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* **37**, 76–81 (2016).
- Soufi, A. *et al.* Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. *Cell* **161**, 555–568 (2015).
- Wang, A. *et al.* Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386–399 (2015).
- Carroll, J. S. *et al.* Chromosome-Wide Mapping of Estrogen Receptor Binding Reveals Long-Range Regulation Requiring the Forkhead Protein FoxA1. *Cell* **122**, 33–43 (2005).
- Roeder, R. G. Role of General and Gene-specific Cofactors in the Regulation of Eukaryotic Transcription. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 201–218 (1998).
- Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
- Zhou, Q., Li, T. & Price, D. H. RNA Polymerase II Elongation Control. *Annu. Rev. Biochem.* **81**, 119–143 (2012).
- Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* **11**, 761–772 (2010).
- Nasmyth, K. & Haering, C. H. Cohesin: Its Roles and Mechanisms. *Annu. Rev. Genet.* **43**, 525–558 (2009).
- Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
- Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
- Smith, E. & Shilatifard, A. Enhancer biology and enhanceropathies. *Nat. Struct. Mol. Biol.* **21**, 210–219 (2014).
- Zhang, X. *et al.* Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat. Genet.* **48**, 176–182 (2016).
- Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
- Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell* **155**, 934–947 (2013).
- Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
- Gonda, T. J. & Ramsay, R. G. Directly targeting transcriptional dysregulation in cancer. *Nat. Rev. Cancer* **15**, 686–694 (2015).
- Losada, A. Cohesin in cancer: chromosome segregation and beyond. *Nat. Rev. Cancer* **14**, 389–393 (2014).
- Spaeth, J. M., Kim, N. H. & Boyer, T. G. Mediator and human disease. *Semin. Cell Dev. Biol.* **22**, 776–787 (2011).
- Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
- Parelho, V. *et al.* Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell* **132**, 422–433 (2008).
- Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
- Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680–1688 (2012).
- Liu, Z. *et al.* Enhancer Activation Requires trans-Recruitment of a Mega Transcription Factor Complex. *Cell* **159**, 358–373 (2014).
- Odom, D. T. *et al.* Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**, 1378–1381 (2004).

37. Burakov, D., Wong, C. W., Rachez, C., Cheskis, B. J. & Freedman, L. P. Functional interactions between the estrogen receptor and DRIP205, a subunit of the heteromeric DRIP coactivator complex. *J. Biol. Chem.* **275**, 20928–20934 (2000).
38. Maeda, Y. *et al.* Polyamines modulate the interaction between nuclear receptors and vitamin D receptor-interacting protein 205. *Mol. Endocrinol.* **16**, 1502–1510 (2002).
39. Malik, S., Wallberg, A. E., Kang, Y. K. & Roeder, R. G. TRAP / SMCC / Mediator-Dependent Transcriptional Activation from DNA and Chromatin Templates by Orphan Nuclear Receptor Hepatocyte Nuclear Factor 4. *Mol. Cell. Biol.* **22**, 5626–5637 (2002).
40. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–520 (2013).
41. Wallerman, O. *et al.* Molecular interactions between HNF4a, FOXA2 and GABP identified at regulatory DNA elements through ChIP-sequencing. *Nucleic Acids Res.* **37**, 7498–7508 (2009).
42. Alder, O. *et al.* Hippo signaling influences HNF4A and FOXA2 enhancer switching during hepatocyte differentiation. *Cell Rep.* **9**, 261–271 (2014).
43. Shu, J. *et al.* Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell* **153**, 963–975 (2013).
44. Pomerantz, M. M. *et al.* The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47**, 1346–1351 (2015).
45. Hurtado, A., Holmes, K. a., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet.* **43**, 27–33 (2011).
46. Wang, J. *et al.* FOSL2 Positively Regulates TGF- $\beta$ 1 Signalling in Non-Small Cell Lung Cancer. *PLoS One* **9**, e112150 (2014).
47. Tang, Y., Shu, G., Yuan, X., Jing, N. & Song, J. FOXA2 functions as a suppressor of tumor metastasis by inhibition of epithelial-to-mesenchymal transition in human lung cancers. *Cell Res.* **21**, 316–326 (2011).
48. Mullen, A. C. *et al.* Master transcription factors determine cell-type-specific responses to TGF- $\beta$  signaling. *Cell* **147**, 565–576 (2011).
49. Liu, J. & Krantz, I. D. Cohesin and human disease. *Annu. Rev. Genomics Hum. Genet.* **9**, 303–320 (2008).
50. Solomon, D. a. *et al.* Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science* **333**, 1039–1043 (2011).
51. Xu, H., Tomaszewski, J. M. & McKay, M. J. Can corruption of chromosome cohesion create a conduit to cancer? *Nat. Rev. Cancer* **11**, 199–210 (2011).
52. Mäkinen, N. *et al.* MED12, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. *Science* **334**, 252–255 (2011).
53. Barbieri, C. E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **44**, 685–689 (2012).
54. Maurano, M. T. *et al.* Large-scale identification of sequence variants influencing human transcription factor occupancy *in vivo*. *Nat. Genet.* **47**, 1393–1401 (2015).
55. Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. & Young, R. A. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* **23**, 2484–2489 (2009).
56. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
57. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
58. Tiscornia, G., Singer, O. & Verma, I. M. Production and purification of lentiviral vectors. *Nat. Protoc.* **1**, 241–245 (2006).
59. Oster, S. K., Mao, D. Y. L., Kennedy, J. & Penn, L. Z. Functional analysis of the N-terminal domain of the Myc oncoprotein. *Oncogene* **22**, 1998–2010 (2003).
60. Pan, Y. F. *et al.* Regulation of Estrogen Receptor-mediated Long Range Transcription via Evolutionarily Conserved Distal Response Elements. *J. Biol. Chem.* **283**, 32977–32988 (2008).

## Acknowledgements

We thank Samer Hussein and Anne-Marie Pulichino for critical review of the manuscript and the laboratories of Dimcho Bachvarov and Jean-Yves Masson for easing access to the xCELLigence Real-Time Cell Analyzer (RTCA) and the Cytation 5. In addition, we would like to thank Éric Paquet, Astride-Louise Deschênes and Audrey Lemaçon for insightful discussions regarding the statistical and bioinformatic approaches. This work was supported by funds to S.B from the Canada Research Chair in Transcriptional Genomics, the Louisa Gale Scholar of the Canadian Cancer Society (grant #2013-701963) and from the Natural Sciences and Engineering Research Council of Canada (grant #436266-2013).

## Author Contributions

M.F., G.B., M.C.C., V.C. and S.B. conceived and performed the experiments. F.C.L., E.F., C.J.-B., A.D. and S.B. conceived and performed the biocomputational analysis. S.G. provided shRNA plasmids. S.B. wrote the manuscript. All authors read and approved the final manuscript.

## Additional Information

**Accession codes:** The data sets supporting the results of this article are available in the NCBI's Gene Expression Omnibus repository and are accessible through GEO Series accession number GSE76893 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=qrspsmoohdytnt&acc=GSE76893>).

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Fournier, M. *et al.* FOXA and master transcription factors recruit Mediator and Cohesin to the core transcriptional regulatory circuitry of cancer cells. *Sci. Rep.* **6**, 34962; doi: 10.1038/srep34962 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016