Mitotic chromosome binding predicts transcription factor properties in interphase

Mahé Raccaud¹, Elias T. Friman¹, Andrea B. Alber¹, Harsha Agarwal², Cédric Deluz¹, Timo Kuhn², J. Christof M. Gebhardt² & David M. Suter¹

Mammalian transcription factors (TFs) differ broadly in their nuclear mobility and sequence-specific/non-specific DNA binding. How these properties affect their ability to occupy specific genomic sites and modify the epigenetic landscape is unclear. The association of TFs with mitotic chromosomes observed by fluorescence microscopy is largely mediated by non-specific DNA interactions and differs broadly between TFs. Here we combine quantitative measurements of mitotic chromosome binding (MCB) of 501 TFs, TF mobility measurements by fluorescence recovery after photobleaching, single molecule imaging of DNA binding, and mapping of TF binding and chromatin accessibility. TFs associating to mitotic chromosomes are enriched in DNA-rich compartments in interphase and display slower mobility in interphase and mitosis. Remarkably, MCB correlates with relative TF on-rates and genome-wide specific site occupancy, but not with TF residence times. This suggests that non-specific DNA binding properties of TFs regulate their search efficiency and occupancy of specific genomic sites.

¹Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. ²Institute of Biophysics, Ulm University, Albert-Einstein-Allee 11, 89081 Ulm, Germany. These authors contributed equally: Elias T. Friman, Andrea B. Alber. Correspondence and requests for materials should be addressed to D.M.S. (email: david.suter@epfl.ch)
Transcription factors (TFs) regulate gene expression by binding regulatory sequences of target genes. TF ability to occupy specific genomic sites depends on their nuclear concentration, their ability to search the genome, and the chromatin environment of their binding sites. How TFs maximize search efficiency for specific sites is incompletely understood. Pioneering theoretical work proposed that DNA-binding proteins display substantial non-specific DNA interactions, which modulate TF search efficiency. The length of the DNA sequence flanking the Lac operator was later shown to impact Lac Repressor on-rate, suggesting that local non-specific TF-DNA interactions increase search efficiency by one-dimensional diffusion along DNA. Experimental and computational modeling studies thus converge on a TF search model that combines 3D diffusion and facilitated diffusion, the latter resulting from local 1D search mediated by sliding along DNA, local jumps or hopping, and transfer between genomically-distant but physically close segments of DNA (intersegment transfer). Such local search mechanisms strongly modulate search efficiency and mainly depend on transient non-specific protein-DNA association mediated by electrostatic interactions. While gene arrays and more recently single molecule imaging have allowed monitoring specific DNA-binding events dynamics, non-specific DNA binding of most mammalian TFs remains uncharacterized, and thus to which extent this property impacts genome-wide occupancy of TFs is unknown.

A minority of TFs were shown to associate with mitotic chromosomes. These interactions can be identified by ChIP-seq on mitotic cells and TF-mitotic chromosome co-localization analysis by fluorescence microscopy. While ChIP-seq essentially identifies sequence-specific DNA binding, fluorescence microscopy allows quantifying mitotic chromosome association independently of enrichment on specific genomic sites. Importantly, immunofluorescence protocols involving chemical fixation cause the artificial eviction of chromatin-bound TFs. In contrast, live cell imaging of TFs fused to fluorescent proteins bypass this problem. Both non-specific and specific DNA binding of TFs to mitotic chromosomes have been described. However, the often small number of specifically-bound loci on mitotic chromosomes, the mild or null sensitivity to alterations of specific DNA binding properties, and the absence of quantitative relationship between mitotic ChIP-seq datasets and fluorescence microscopy suggest that co-localization of TFs with mitotic chromosomes as observed by microscopy is largely due to non-specific DNA interactions. Converging evidence from the literature further corroborates this view. SOX2 and FOXA1 strongly associate with mitotic chromosomes and display high non-specific affinity for DNA in vitro and high display non-specific affinity for DNA in vitro. In contrast, OCT4 displays less visible association with mitotic chromosomes and has low non-specific affinity for DNA in vitro. Finally, FOXA1 mutants with decreased non-specific DNA affinity but retaining their specificity for the FOXA1 motif also display reduced mitotic chromosome association.

Many TFs binding to mitotic chromosomes have pioneer properties, i.e., they can bind and open condensed chromatin regions. However, the existence of a common molecular mechanism underlying mitotic chromosome binding and pioneer activity remains uncertain.

Here we measure mitotic chromosome binding (MCB) of 501 mouse TFs in live mouse embryonic stem (ES) cells. We show that MCB correlates with interphase TF properties such as sub-nuclear localization, mobility, and with large differences in TF ability to occupy specific genomic sites. We propose that the co-localization of TFs with mitotic chromosomes is a proxy for TF non-specific DNA binding properties, which regulate TF search efficiency for their specific binding sites and thereby their impact on chromatin accessibility.

**Results**

Large-scale assessment of TF binding to mitotic chromosomes.

To measure MCB for a large number of TFs, we constructed a doxycycline (dox)-inducible lentiviral vector library of 757 mouse TFs fused to a yellow fluorescent protein (YPet) (Fig. 1a). This library was used to generate a corresponding library of mouse embryonic stem (ES) cell lines to quantify TF MCB by live cell fluorescence microscopy. Cells were seeded in 96-well plates and treated with dox to induce expression of TF-YPet fusion proteins. and the next day they were imaged by wide-field fluorescence microscopy. We used a semi-automated pipeline to detect cells in metaphase, allowing easy quantification of MCB since chromosomes are most spatially confined in this phase. Of note, we did not observe obvious differences in TF co-localization with mitotic chromosomes between prophase, metaphase and anaphase. We used the Mitotic Bound Fraction (MBF) as a metric for mitotic chromosome binding, defined as the averaged YPet fluorescence intensity on metaphase chromosomes multiplied by the fraction of cellular volume occupied by DNA (as measured by confocal microscopy, see Methods), divided by the total YPet signal (Fig. 1a). We then asked whether MCB values depend on over-expression levels, in at least 19 individual cells for each of 21 different TFs spanning a broad range of MCB. We found no consistent correlation between these two parameters (Supplementary Fig. 1a), suggesting that the MCB is largely independent of overexpression levels. We also measured the MCB of an endogenously expressed TF and the MBE of its overexpressed counterpart, using a homozygous SOX2-SNAP cell line and a dox-inducible SOX2-SNAP cell line. We only observed modest differences between their MBE values (Supplementary Fig. 1b).

We then compared wide-field fluorescence measurements with those performed by confocal microscopy, and found these to be in very good agreement (Supplementary Fig. 1c-d, Supplementary Table 1). In total 501 TFs yielded sufficiently strong fluorescent signals in metaphase to allow measuring their MBE, and for 94% of these we could measure the MBE in at least 10 cells (see Methods). We defined three bins of TFs based on visual inspection of the YPet signal on metaphase chromosomes: depleted (YPet signal lower than in the cytoplasm), intermediate (YPet signal equal to that of the cytoplasm) or enriched (YPet signal higher than in the cytoplasm), corresponding to MBEs <16.5%, 16.5–23% and >23%, respectively. Twenty-four percent of TFs fell in the depleted bin, 54% in the intermediate bin, and 22% in the enriched bin (Fig. 1b). Most TFs previously reported as enriched on mitotic chromosomes and for which we could obtain MBE values, such as FOXA1, GATA1, SOX2, ERBB3, HMG2, HMGN1, fell in the intermediate (ESRRB) or enriched (all five other TFs) category (Supplementary Table 2), suggesting that C-terminal YPet fusion does generally not perturb MCB. We also measured the MBE of a subset of TFs in NIH-3T3 cells, yielding consistent results (Supplementary Fig. 1e-f, Supplementary Table 3), suggesting that intrinsic TF properties are major determinants of mitotic chromosome association. We next compared our results with those obtained by a proteomics study investigating chromatin-bound proteins over the cell cycle. Importantly, that study used a different cell type (a glioblastoma cell line), species (human), and involved PFA cross-linking that can lead to artificial disruption of TF-mitotic chromosome interactions. Nevertheless, we found our results to globally agree with theirs (Supplementary Fig. 1g). Furthermore, while in contrast to the study by Ginno et al., our library contains TFs...
that are endogenously expressed in ES or NIH-3T3 cells but also many that are not, we did not find significant differences in MBF values between these two groups (Supplementary Fig. 1h). This suggests that endogenous TF expression status does not strongly impact its MBF.

Next, we clustered TFs according to their DNA binding domains (DBD) and found that members of some TF families (e.g., homeodomain or tryptophan cluster) were more likely to be enriched on mitotic chromosomes (Fig. 1c). In contrast, C2H2 Zinc-finger TFs were underrepresented in the enriched category.
in line with mitotic phosphorylation preventing their association with DNA during M-phase\textsuperscript{43,44}. However, the broad range of MBF within each family indicates that DNA binding domain type does not strictly govern the MBF (Fig. 1c), suggesting other TF characteristics involved in regulating mitotic chromosome association.

The MBF depends on electrostatic properties. We next used a machine learning algorithm (see Methods) based on a lasso regularized generalized linear model\textsuperscript{35} to uncover TF features that could explain differences in MBF (Fig. 1d and Supplementary Fig. 1). We collected a large number of features from the amino acid sequence of 401 different TFs (see Methods and Supplementary Table 4) to find those that are correlated with the MBF. We ran the algorithm 500 times on the data to obtain parameters used to predict the MBF of the remaining 100 TFs. The algorithm selects variables allowing predicting the MBF, while coefficients of non-predictive parameters are set to zero. Similarly, coefficients of covariates that are correlated among them are set to zero. We then kept only parameters with a coefficient $>0$ in at least 90% of the runs (Fig. 1d). As expected, certain types of DBD, such as the tryptophan cluster TF family and the C2HC Zinc fingers were moderately predictive of the MBF. The other parameters most strongly correlated with the MBF were the fractions of polar amino acids and disordered domains, and the absolute charge per DBD (number of basic minus number of acidic residues), while the fraction of aliphatic and hydrophobic amino acids was negatively correlated with the MBF (Fig. 1d). We then compared these parameters between the three different TF bins we defined above (Fig. 1b). Although broadly distributed in each bin, all of them were significantly different in TFs enriched on mitotic chromosomes (Fig. 1e). The absolute charge per DBD was the most distinctive parameter between TFs enriched on mitotic chromosomes (dark blue) versus those that are not (medium and light blue), suggesting that electrostatic interactions play an important role in mitotic chromosome association of TFs.

TFs usually harbor one or several nuclear localization signals (NLS), which often consist of a series of positively charged amino acids. During mitosis, Ran guanine nucleotide exchange factor (NLS), which often consist of a series of positively charged amino acids, plays an important role in mitotic chromosome association. We also added five positively charged amino acids to the N-terminus of four different TFs, and as expected these displayed an increased MBF (Supplementary Fig. 1k). Taken together, these results are in line with previous studies suggesting that (i) non-specific DNA binding is essentially mediated by electrostatic protein-DNA interactions, and (ii) non-specific DNA binding is largely responsible for the co-localization of TFs with mitotic chromosomes.

The MBF correlates with interphase TF-DNA co-localization. Since mitotic chromosome association reflects general non-specific DNA binding properties, we next asked whether the MBF scales with TF-DNA co-localization in interphase. To address this question, we used NIH-3T3 cells, which display easily identifiable regions of varying chromatin densities. We stained NIH-3T3 cells with Hoechst and classified the signal in three bins using automatic image thresholding: (i) very dense, H3K9me3-enriched heterochromatin regions (Fig. 2a); (ii) DNA-rich regions; (iii) DNA-poor regions (Fig. 2b, Methods). We then selected 38 NIH-3T3 cell lines allowing dox-inducible expression of selected TF-YPet fusions spanning a broad range of MBF. After overnight dox treatment and Hoechst staining, we performed two-color confocal microscopy to measure YPet-Hoechst signal co-localization (Fig. 2c). We observed a wide range of spatial repartition of TFs, with some of them (e.g., Duxbl, Bhlhb8, and Dlx6, Fig. 2c) highly enriched in heterochromatic regions. The MBF and the co-localization of TFs with DNA in interphase were positively correlated (Fig. 2d and Supplementary Table 5). We then determined this correlation within the different sub-nuclear regions we defined, and found the MBF to be positively correlated with enrichment in heterochromatin regions (Fig. 2e and Supplementary Table 5) and to a lesser extent with enrichment in DNA-rich regions (Fig. 2f and Supplementary Table 5). In contrast, the MBF was strongly, inversely correlated with enrichment in DNA-poor regions (Fig. 2g and Supplementary Table 5). Thus, TFs with a high MBF tend to be excluded from DNA-poor regions and are distributed mainly within heterochromatic and DNA-rich regions in interphase, suggesting that non-specific TF-DNA interactions also control interphase TF localization.

The MBF correlates with mitotic and interphase TF mobility. We reasoned that co-localization with DNA-rich regions resulting from transient TF-DNA interactions should decrease TF mobility in the cell. To test this hypothesis, we performed fluorescence recovery after photobleaching (FRAP) experiments in ES cells on 15 selected TFs with different MBF in mitotic and interphase cells, as well as FRAP on interphase cells only for 3 TFs excluded from mitotic chromosomes (Fig. 3a, b, Supplementary Table 6). We selected these 18 TFs based on their high signal to noise ratio required for reliable FRAP measurements,
Fig. 2 Mitotic chromosome binding is correlated with TF-DNA co-localization in interphase. **a** Immunofluorescence labeling of H3K9me3 (Abcam, ab8898) and Hoechst staining of a NIH-3T3 nucleus. Scale bar: 3 µm. **b** Automatic detection of regions displaying different densities of DNA: Heterochromatic (circled in red), DNA-rich (between red and yellow circles), or DNA-poor (circled in yellow). Scale bar: 3 µm. **c** Examples of TF-YPet interphase localization in NIH-3T3 cells, as compared to Hoechst staining and ranked by mitotic bound fraction (for n values see Supplementary Table 2). Bra Brachyury. Scale bar: 5 µm. **d** Correlation between the TF-YPet/Hoechst co-localization (MAFK: n = 8, NANOG and BHLHB8: n = 11, others: n = 10) and the mitotic bound fraction. **e-g** Correlation between the mitotic bound fraction and the fraction of YPet signal co-localized with heterochromatic regions (**e**), DNA-rich regions (**f**), and DNA-poor regions (**g**). n = 10. Error bars: SEM. r-values (r) and p-values (p) are based on Pearson correlation.
Fig. 3 TF mobility correlates with the mitotic bound fraction and interphase ψon-rate. a, b Top: Example of FRAP for CDX2-YPet (a) and HOXD10-YPet (b) for cells in interphase and mitosis (scale bar: 5 µm). Bottom: normalized interphase and mitosis FRAP recovery curves for CDX2-YPet (a) and HOXD10-YPet (b). c Correlation between FRAP $t_{1/2}$ recovery of TFs in interphase and mitosis. d Correlation between TF mitotic bound fraction and FRAP $t_{1/2}$ recovery in interphase. e Correlation between TF mitotic bound fraction and FRAP $t_{1/2}$ recovery in mitosis. f Correlation between TF mitotic bound fraction and FRAP immobile fraction in interphase. g Correlation between TF mitotic bound fraction and FRAP immobile fraction in mitosis. h Correlation between FRAP $t_{1/2}$ recovery in interphase and the ψon-rate. i Correlation between FRAP $t_{1/2}$ recovery in mitosis and the ψon-rate. j Correlation between FRAP $t_{1/2}$ recovery in interphase and TF relative residence times. k Correlation between FRAP $t_{1/2}$ recovery in mitosis and TF relative residence times. Error bars: SEM. r-values (r) and p-values (p) are based on Pearson correlation. For FRAP relative intensity, $t_{1/2}$ recovery time and immobile fraction: $n = 10$.
and their coverage of a broad range of MBF. We then used half-time ($t_{1/2}$) of fluorescence recovery as a metric of TF mobility. Mitotic and interphase $t_{1/2}$ of fluorescence recovery displayed a strong positive correlation (Fig. 3c and Supplementary Table 6), and TF mobility was strongly correlated with the MBF (Fig. 3d, e). These results suggest that intrinsic TF properties govern the MBF and TF mobility in both interphase and mitosis.

**TF size or off-rate do not explain differences in mobility.** FRAP recovery times depend on 3D diffusion, specific and non-specific DNA binding. Since mitotic chromosome association observed by fluorescence microscopy largely depends on non-specific TF-DNA interactions and correlates with $t_{1/2}$ of FRAP, we reasoned that differences in TF mobility might be mainly due to differences in non-specific DNA binding. However, since 3D diffusion and specific DNA binding also influence $t_{1/2}$ of fluorescence recovery, we first asked whether differences in these parameters could explain differences in TF mobility. According to the Stokes-Einstein equation, diffusion scales inversely with the size of molecules. Even though differences in TF radius are predicted to be very small since they scale with the third root of their mass, we assessed the correlation between TF-YPet molecular weight and FRAP $t_{1/2}$ recovery (Supplementary Fig. 2a, b and Supplementary Table 6). These two parameters were negatively correlated and thus do not explain differences in TF mobility. To test whether TFs quantitatively differ in their association with specific DNA sites, we determined the immobile fraction, which corresponds to the fraction of molecules that do not exchange over the time course of FRAP experiments. The immobile fractions in interphase and mitosis were well correlated with the mitotic bound fraction (Fig. 3f, g), suggesting that TFs with a high MBF tend to have a higher fraction of long binding events.

We then performed single molecule (SM) imaging of TFs in live cells to monitor specific DNA binding events. This approach allows comparing frequencies of specific DNA-binding events between TFs at a given TF nuclear concentration, thus providing a relative measure of TF on-rates. Furthermore, SM imaging allows determining differences in TF residence times on specific DNA sites. We generated 9 NIH-3T3 cell lines allowing dox-inducible expression of TFs C-terminally fused to a HaloTag and inducible expression with low doses of dox shortly before performing SM imaging in interphase cells, using highly inclined and laminated optical sheet (HILO) microscopy (see Methods). Single molecules were registered as bound to chromatin when they were confined for several frames inside a specified region (see Methods). We then associated these binding events with areas of bright, intermediate or dark Hoechst staining intensity (Supplementary Fig. 3a and Supplementary Table 7), similarly as for confocal measurements of TF nuclear distribution (see Fig. 2a, b). In order to compare different TFs, we calculated the ratio of DNA-binding events divided by the total number of detected molecules, the area of the corresponding intensity class and the recording time (Eq. 7 in Methods). This ratio provides a frequency of DNA binding events. We referred to this frequency for binding events lasting >1 s as the pseudo on-rate ($\psi$-on-rate hereafter). We approximated the residence time of a TF with the average time a molecule spent bound to DNA (Equation 8 in Methods). FRAP $t_{1/2}$ recovery times in interphase and mitosis were correlated with interphase TF $\psi$-on-rates (Fig. 3h, i, Supplementary Tables 6 and 7). In contrast, FRAP recovery times were not correlated with relative interphase residence times (Fig. 3j, k, Supplementary Tables 6 and 7). This suggests that while differences in TF $\psi$-on-rates could contribute to differences in TF mobility, differences in 3D diffusion or residence times on specific DNA sites do not. This is also in line with non-specific DNA binding contributing to differences in TF mobility and search efficiency, thereby impacting TF $\psi$-off-rates on specific sites.

**TF occupancy correlates with $\psi$-on-rates but not off-rates.** We next aimed to determine the genome-wide TF occupancy at specific sites for 21 TFs spanning a broad range of MBF, and we also included two FOXA1 mutants that have reduced electrostatic interactions with DNA, decreasing non-specific DNA binding activity without altering their specificity for the FOXA1 DNA motif. As expected, FOXA1 mutants displayed a lower MBF than wild-type FOXA1 (Supplementary Fig. 3b). For each TF, we generated an NIH-3T3 cell line allowing their dox-inducible expression with three HA tags fused to their C-terminus. We verified that different TFs were expressed at roughly comparable levels upon dox induction (Supplementary Fig. 3c, Supplementary Table 8), allowing ChIP-seq measurements to provide relative measurements of genome occupancy within a given TF concentration range. We then performed ChIP-seq using the same anti-HA antibody for all TFs and used peak calling to determine the number of specifically bound sites. Remarkably, for the 8 TFs for which we obtained both ChIP-seq and single molecule measurements, TF $\psi$-on-rates were strongly correlated with the number of ChIP-seq peaks (Fig. 4a, Supplementary Fig. 3d-f, Supplementary Tables 7 and 8). The offset of the linear fit suggests that a fraction of long-lived SM events is not detected by ChIP-seq. As previously reported, these can result from long-lived, non-specific DNA binding events or indirect TF-DNA interactions. In contrast, the number of ChIP-seq peaks was not correlated with TF relative residence times (Fig. 4b, Supplementary Tables 7 and 8). This suggests that the differences in genome-wide occupancy of TFs we observed are mainly due to differences in their association rate but not to longer residence times on specific DNA sites.

**The MBF correlates with genome-wide TF occupancy.** We used the ChIP-seq data to compare relative TF occupancy of specific sites in the genome with their MBF. We first determined the number of ChIP-seq peaks obtained for the 21 TFs, the two FOXA1 mutants, with two biological replicates for OCT4, SOX2, FOXA1, and BHLHB8 to estimate variability in recovering ChIP-seq peaks. Remarkably, the MBF and the number of ChIP-seq peaks were strongly correlated, with differences in ChIP-seq peak numbers ranging over three orders of magnitude (Fig. 4c, d and Supplementary Table 8). This correlation was robust to different q-value thresholds (Supplementary Fig. 4a), peak calling algorithms (Supplementary Fig. 4b), and downsampling to equalize the number of reads for each TF (Supplementary Fig. 4c). Very similar results were obtained when using the fraction of ChIP-seq reads in peaks as another metric for TF occupancy (Supplementary Fig. 4d and Supplementary Table 8), indicating that differences in ChIP-seq peak amplitude are small compared to differences in peak numbers. According to the law of mass action, the rate of formation of a TF-DNA complex scales linearly with its concentration, and therefore the six-fold range in TF expression level (Supplementary Fig. 3c) is unlikely to account for the very large differences in ChIP-seq peak numbers we observed. This relationship was also independent of the presence or absence of endogenous expression of these TFs in NIH-3T3 cells (Fig. 4d and Supplementary Table 8).

We then wondered why some TFs reported to display a high number of ChIP-seq peaks when expressed in their endogenous context displayed orders of magnitude fewer peaks in NIH-3T3 cells. While this may partly be due to differences in TF concentrations, antibodies used, and ChIP-seq protocols, we...
reasoned that in some cases differences in binding partners could also explain this discrepancy. In particular, OCT4 displayed two orders of magnitude fewer ChIP-seq peaks in NIH-3T3 as compared to ES cells. As SOX2 and OCT4 form heterodimers, and SOX2 can recruit OCT4 to DNA, OCT4 may have a low intrinsic ability to find its target sites in the absence of a heterodimeric partner. To test this hypothesis, we co-expressed a YPet-SOX2 fusion protein with OCT4-HA in NIH-3T3 and performed ChIP-seq against OCT4. While co-expression of YPet-SOX2 did not alter the expression level of OCT4-HA, the number of peaks strongly increased of OCT4-HA, suggesting that SOX2 drives OCT4 DNA binding in NIH-3T3 cells.

The MBF scales with TF search efficiency for specific sites. The number of sites bound by each TF may also depend on the number of specific sites available in the genome. We thus quantified the number of occurrences of the most highly enriched motif for the bound sites of each TF (see Supplementary Table 9 and Methods). We then quantified the fraction of each top motif that is occupied by its respective TF by dividing the number of ChIP-seq peaks containing the top motif (Supplementary Tables 8 and 9) by the total number of its occurrences in the genome (fraction of occupied motifs or FOM). In addition, we performed ATAC-seq on NIH-3T3 cells to map accessible chromatin regions, allowing to calculate the FOM in open chromatin regions. The FOM in both the whole genome and accessible chromatin strongly correlated with MBF. In this, the MBF measures the overall TF search efficiency, i.e., how likely a TF is to find its target sites given the genome-wide distribution of its recognition motif.
with the MBF (Supplementary Table 8), suggesting large differences in TF ability to find their specific sites, independently of the number of sites present in the whole genome or within accessible chromatin regions. Notably, some TFs were not enriched for their known consensus motif from the literature but rather for motifs annotated to other factors. This is not unexpected since TF binding profiles are often context-dependent, and motif enrichment can also be due to indirect binding or the presence of a motif in proximity of a true binding site. To exclude the possibility that our correlation is biased by indirect binding or incorrect motif identification, we analyzed the FOM at known consensus sites for all factors significantly enriched for a motif annotated to the factor itself or a closely related protein (e.g., the SOX3 motif for SOX2 and SOX15, see Supplementary Table 9 and Methods) and found this metric to also positively correlate with the MBF (Supplementary Fig. 4i, j and Supplementary Table 8). Therefore, our data suggests that the MBF is of strong predictive value for the ability of TFs to occupy their specific sites in the genome at a given TF concentration.

**MCB does not predict pioneer activity.** We then asked whether the MBF is predictive of TF ability to modify chromatin accessibility at their binding sites. Since pioneer TFs are suggested to be more capable of targeting closed chromatin regions than non-pioneer TFs, we first determined if the MBF was correlated with a higher propensity to bind closed chromatin regions. We found no correlation between the MBF and the fraction of TF ChIP-seq peaks in closed chromatin (Fig. 5a and Supplementary Table 8), and a weak positive correlation between the MBF and the FOM in closed chromatin (Supplementary Fig. 5a). We then selected 13 TFs that are not endogenously expressed in NIH-3T3 cells based on a published RNA-seq dataset to interrogate their ability to modify chromatin accessibility. For each TF, we performed two biological replicates of ATAC-seq after 48 h of dox induction in the respective NIH-3T3 cell line (Fig. 5b and Supplementary Fig. 5b). We counted the number of ATAC-seq reads in regions bound by each TF in the control cell line and each overexpression cell line and used edgeR and limma to call regions significantly (p < 0.05) either more or less accessible following overexpression of the TF (see Methods). Overall, the number of regions with changed accessibility coinciding with TF ChIP-seq peaks correlated with the number of TF ChIP-seq peaks (Fig. 5c and Supplementary Table 10). Of note, these regions displayed either significantly increased (Supplementary Fig. 5c) or decreased (Supplementary Fig. 5d) accessibility, and both correlated with the number of TF ChIP-seq peaks. As expected, a large number of genomic sites to which known pioneer TFs such as SOX2 and FOXA1 bound displayed changes in chromatin accessibility (Fig. 5c, Supplementary Fig. 5b, and Supplementary Table 10). The impact of TFs on both opening and closing chromatin can be explained by the fact that many of these TFs were reported to function as both activators and repressors. Importantly, these correlations were maintained when analyzing the same regions for all TFs (defined as all regions with an ATAC-seq peak in at least one sample, see Methods) (Supplementary Fig. 5e-g). To determine the intrinsic ability of each TF to modify chromatin accessibility, we quantified the fraction of sites bound by each TF that displayed significant increase or decrease in chromatin accessibility. Neither of them was correlated with the MBF (Fig. 5d-e and Supplementary Table 10), suggesting that TFs with a high MBF are not more potent on average in altering chromatin accessibility. Therefore, a high MBF is not a signature for an intrinsically higher propensity to alter chromatin accessibility.

**Discussion**

Recent studies have shown that TFs binding to mitotic chromosomes are rather common, and these were proposed to play a role in cell fate maintenance by controlling gene reactivation early during mitotic exit. Furthermore, the presence of SOX2 and OCT4 at the M-G1 transition was shown to be required for their role in regulating ES cell fate decisions. However, the molecular mechanisms underlying MCB are largely unknown and this property remains uncharacterized for the vast majority of TFs. Here we measured MCB of 501 TFs and found over 100 to be enriched on mitotic chromosomes. Our study thus provides a large database of TFs to be mined for their potential role in cell fate maintenance during cell division.

MCB as observed by fluorescence microscopy is distinct from the notion of mitotic bookmarking, which implies TF binding to specific sites during mitosis. However, all mitotic bookmarking TFs described so far are also decorating mitotic chromosomes as assessed by fluorescence microscopy. Here we show that TFs with a high MBF have a higher immobile fraction on mitotic chromosomes, raising the possibility that these also display a higher occupancy of specific sites during mitosis. However, mitotic ChIP-seq results obtained by different laboratories show little agreement, and the scarcity of data on TFs displaying poor co-localization with mitotic chromosomes by fluorescence microscopy does not allow to draw clear conclusions on the link between these two properties.

While our data indicate that the type of DNA binding domain and electrostatic characteristics of TFs impact MCB of TFs, the in silico determination of these properties based on TF amino acid sequence was not sufficient to accurately predict the MBF. Therefore, measuring other parameters such as post-translational modifications and three-dimensional structures of TF-DNA contact interfaces is arguably required to improve MBF prediction.

We found that quantitative measurements of MCB are well correlated to other TF properties. First, differences in the MBF scaled inversely with TF mobility in both interphase and mitosis. This can be explained by differences in non-specific DNA binding properties, which drive mitotic chromosome association and decrease TF mobility in the nucleus. Second, less mobile TFs displayed higher on-rates, suggesting that non-specific DNA association increases TF efficiency to search for target sites through facilitated diffusion. This is also corroborated by the high and low genome occupancy of TFs (SOX2 or FOXA1 versus OCT4) with high and low non-specific DNA binding activity in vitro, respectively. Third, TFs differed broadly in their ability to occupy specific sites in the genome during interphase, which scaled with their MCB, even though these two measurements were performed in different cell types. Since most TFs we studied are not endogenously expressed in NIH-3T3 cells, they are likely to depend essentially on their intrinsic ability to find their specific sites in the genome rather than on the cooperativity with other TFs. Importantly, here we aimed to determine the intrinsic ability of TFs to bind to their sites within a given concentration range. Thus, optimization of TF expression levels in their physiological context may further allow fine-tuning their genome occupancy. Interestingly, we found that OCT4 occupancy can be dramatically increased by co-expression of SOX2, and we had reported that SOX2 expression increases mitotic chromosome association of OCT4. Therefore, it is possible that some TFs depend on cooperativity with other DNA-binding proteins to mediate or enhance MCB and search for specific binding sites.

Finally, our data also sheds light on the proposed link between MCB and impact on chromatin accessibility. The MBF could predict neither preferential binding to closed chromatin nor the
inherent ability to modulate chromatin accessibility. However, the higher number of specific binding sites of TFs with a high MBF correlated with an increased absolute number of bound sites in closed chromatin regions, as well as a broader impact on chromatin accessibility. Therefore, the on-average higher impact on chromatin accessibility of TFs with a high MBF is mediated by their enhanced ability to occupy specific sites rather than an intrinsically higher ability to alter chromatin accessibility at target sites.

In summary, our findings converge on a model in which non-specific DNA binding properties play a central role in determining TF association to mitotic chromosomes and interphase DNA. We propose that non-specific DNA binding governs TF search efficiency for specific binding sites and thereby has a
modest impact on TF ability to alter the chromatin accessibility landscape (Fig. 6). Future studies should allow further understanding the molecular basis of differential non-specific DNA binding activity and how this allows to optimize chromatin scanning to find target sites in the genome.

**Methods**

**Cell lines and culture.** The E14 mouse embryonic stem (ES) cell line (provided by Didier Trono, EPFL) was used for all experiments involving ES cells. Cells were routinely cultured at 37 °C and 5% CO2 in GMEM (Sigma-Aldrich, #G5154), supplemented with 10% ES cell-qualified fetal bovine serum (ThermoFisher, #10270106), 1% non-concentrated lentiviral vector particles (infiltrated with 100 mM 2-mercaptoethanol (Sigma-Aldrich, #63689–01F00H), 2 mM L-glutamine (ThermoFisher, #25030–024), 100 µM 2-mercaptoethanol (Sigma-Aldrich, #63689–25ML–F), leukemias inhibitor factor (LIF, concentration not determined, produced by transient transfection of HEK 293T cells. For all cell lines, selection of transduced cells was achieved by removing the Oct4 coding sequence from the pLVTRE3G-Oct4-YPet plasmid22 using SacI and Ascl digestion. The coding sequences of Sox15 and Hoxd10 were then amplified with primers flanked with SacI and Ascl sites, digested with these enzymes, and ligated to SalI/AscI-cut pLVTRE3G-Oct4-YPet.

The two FoxA1 non-specific DNA-binding mutants36 were generated by site-directed mutagenesis before recombination in the pLVTRE3G-GW-HA/YPet, by PCR (primers in Supplementary Table 11) on the foxA1 pENTR vector.

The TFs fused to 5 arginine residues were generated by PCR amplification (primers in Supplementary Table 11) of the TOF ORF from the corresponding pENTR vectors.

**Live imaging of TFs and MBF quantification.** One day before imaging, cells were seeded in black-walled 96-well plates that were either uncoated (NIH-3T3 cells) or coated with 1 µg/ml of poly-L-lysine and 1 µg/ml of collagen. Puromycin and 1% Penicillin-Streptomycin. Didier Trono, EPFL) was used for all experiments involving ES cells. Cells were seeded in black-walled 96-well plates that were either uncoated (NIH-3T3 cells) or coated with 1 µg/ml of poly-L-lysine and 1 µg/ml of collagen.

The constructs for Halo-tagged TFs were generated using restriction cloning, by digesting the pLVTRE3G-GW-HA/YPet with Ndel and Ascl to replace the loxP-3xHA-2xSTOP-LoxP-YPet with a Ndel and Ascl-cut PCR product of the Halo-Tag coding sequence flanked by Ndel and Ascl restriction sites. The selected TFs were then shuttled from the pENTR library into the pLVTRE3G-GW-Halo by LR-Gateway recombination (ThermoFisher).

**DNA constructs.** A pENTR library containing the coding sequences of 750 mouse transcription factors (TFs) without STOP codon34 was recombined in the doxycycline-inducible expression vector described in ref. 32, (hereafter referred to as pLVTRE3G-GW-HA/YPet). This vector allows for dox-inducible expression of each TF fused either to YPet (a yellow fluorescent protein) or to three HA tags, depending on the presence or absence of Cre recombinase, respectively. Additionally, the coding sequences of Gbx2, Ki67, Klf4, Klf5, Nanog, Tcf3, and Sox17 were added to the library by PCR amplification from cDNA (primers in Supplementary Table 1) extracted from either ES cells maintained in the pluripotent state or ES cells differentiated for four days by removal of LIF and 2i. The PCR amplicons were then inserted in the pDONR221 by BP-Gateway recombination (ThermoFisher) and verified by Sanger sequencing. The pENTR library was then recombined into pLVTRE3G-GW-HA/YPet. The resulting lentiviral vector library was used to generate a corresponding ES library for the large-scale quantification of TF binding to mitotic chromosomes and for fluorescence recovery after photobleaching (FRAP) experiments. In the case of Sox15-YPet and Hoxd10-YPet, the corresponding E14 cell lines did not yield high enough fluorescence intensity signals to perform FRAP experiments. To circumvent this issue, the coding sequences of Sox15-YPet and Hoxd10-YPet were inserted into the pLVTRE3G-MCS backbone32, which allowed obtaining higher expression levels. This was achieved by removing the Oct4 coding sequence from the pLVTRE3G-Oct4–YPet.

**Figure 6 Mitotic chromosome binding predicts TF properties in interphase.** We propose that the MBF allows to predict TF non-specific DNA binding, which regulates TF mobility, on-rate and search efficiency, thus influencing genome occupancy. Modulation of chromatin accessibility at bound sites depends on genome-wide occupancy but even more so on pioneer activity.
FluoroBrite DMEM (ThermoFisher, #A18967-01) supplemented with 10% FBS. For NIH-3T3, to compare the MBF between wild type TFs and TFs tagged with 5 channel (Supplementary Fig. 1b). MBF were measured manually as described above for SNAP, the cell lines were labeled with 24 nM of the SNAP-Cell_647-SiR dye (New from67 (GSE77420).

Chromatome reporter signal in M-phase plus one, divided by the mean Proteome expression data from54 (GSE66243) and TFs endogenously expressed in ES cells (84%). For NIH-3T3, volumes measured for E14 cells were used as a proxy. The MBF was averaged over >10 cells per clone for 94% of the TFs and over 4 to 9 cells for 6% (n=29) of the TFs in E14 cell lines, and over 17 of the 20 NIH-3T3 cell lines, while the 3 others where averaged over 4 to 6 cells. For Supplementary Fig. 1g, Supplementary Data 2 from Ginn et al. was downloaded and data for genes present in that study and for which we measured a MBF was merged based on gene name. Mitotic enrichment data was calculated as the log of the mean Chromatome reporter signal in M-phase plus one, divided by the mean Proteome reporter signal in M-phase plus one. For Supplementary Fig. 1h, TFs endogenously expressed in NIH-3T3 were defined as in (average expression) >2 based on expression data from44 (GSE66243) and TFs endogenously expressed in ES cells were defined as In (average E14 2i 72 h expression) >0 based on expression data from42 (GSE77420).

To compare the MBF between endogenous and exogenous expression of SOX2- SNAP, the cell lines were labeled with 24 nM of the SNAP-Cell_647-SiR dye (New England BioLabs Cat# 9102S) and imaged on the Cell Analyzer with the Cy5 channel (Supplementary Fig. 1b). MBF were measured manually as described above for NIH-3T3. To compare the MBF between wild type TFs and TFs tagged with 5 arginine residues (Supplementary Fig. 1k), cells were imaged as described above for the large scale quantification. MBF were measured manually as described above for NIH-3T3 cells.

Immunofluorescence. Cells were fixed for 30 min with 2% PFA in PBS, permeabilized in 0.5% Triton in PBS, and blocked with PBS and 1% BSA for 30–60 min. Samples were incubated with primary antibodies, either anti-HA (anti-HA.11 IgG, Biolegend, #302301, 1:1500 dilution overnight at 4°C), or anti-mouse Alexa 488 antibody (Life Technologies, cat # A21202) or an anti-mouse Alexa 647 antibody (Life Technologies, cat # A-31571) for the OCT4-HA/YPet-Sox2 co-expression assay (Supplementary Fig. 1g). For the anti-HA immunostainings, samples were incubated with an anti-rabbit Alexa 647 antibody (Life Technologies, cat # A-31573). All secondary antibodies were used at 1:1000 dilution in PBS and 1% BSA, and left on samples for 45–60 min followed by three washes with 0.1% PBS-Tween, incubation with 2 ng ml−1 DAPI, three washes with PBS and 0.1% Tween, and two washes with PBS.

DNA-binding domain assignment and machine learning analysis. The number of DNA binding domains per TF and their family were extracted from the UniProt database. The DBD families (classified as described on http://www.edgar-wingender.de/mafT_classification-1.html) were divided into the following categories: 1) coherently present on mouse, not present on human, 2) coherently present on human, coherently absent on mouse, 3) coherently present on mouse and human, 4) coherently absent on mouse and human. In case more than one DBD type were included to each of the DBD families. Therefore, in Fig. 1c, some TFs are represented in several boxes. Amino acids were classified as described in Lee et al., 2009, into the following categories: positively charged, aromatic, polar, hydrophobic aliphatic, tiny, bulky, and small amino acids (see Supplementary Table 1). Additionally, patchy were calculated for the amino acid and average of hydropathy (GRAVY) score of the protein, the total number of consecutive positive amino acids, the total number of consecutive neutral amino acids and the total number of consecutive negative amino acids were calculated as previously described70. Most sequence-based parameters were extracted using a web tool71. The dispersion of positive charges was calculated using a sliding window of 5 amino acids to sum the number of arginine and lysine residues and quantified as the variance over the mean of these sums. All absolute counts of amino acids were normalized between 0 and 1. Absence or presence of DBD was annotated with 0 and 1, respectively, for each TF. TFs with missing variables were removed from the analysis. The coefficient for each parameter was calculated using a lasso regularized linear model from glmnet package on R50 on the log of the MBF, and averaged over all the runs. Significant parameters were described as parameters retained by the model in 90% of the runs (n = 500).

Imaging and co-localization analysis in NIH-3T3 cells. NIH-3T3 cells (kindly provided by the laboratory of Ueli Schibler, University of Geneva) were seeded in FluoroDishes (WPI, FD35-100) at densities of 36,000 cells cm−2 24 h before imaging and treated with 500 ng ml−1 of doxycycline. Shortly before imaging, cells were incubated with 1:62.5 µM of Hoechst 3342 (Invitrogen, #H35750) for 15 min and washed twice with PBS.

Cells were imaged using a confocal microscope (ZEISS LSM 700 INVERT) with a ×63 objective at 37 °C and 5% CO2. Channel settings were as following: EYFP 2.4% laser power, 700–900 gain, 41.1 µm pinhole; H342 2.6% laser power, 500–700 gain, 39 µm pinhole. Image dimensions: 50.8 × 50.8 µm, 0.05 µm pixel−1.

Whole cell signal co-localization was performed by pixel-pixel correlation of the Hoechst and YPet signal images using the R software. Brieﬂy, the Hoechst and YPet images were converted into text images and the correlation score between the two channels was calculated for each cell, using 10 cells per cell line.

The co-localization in different DNA regions was analyzed using an automated image segmentation pipeline in Fiji. Brieﬂy, nuclei were identiﬁed and segmented based on the Hoechst signal and 3 regions with high, medium and low Hoechst levels within each nucleus were deﬁned by k-means clustering. Subsequently, the corresponding YPet signal in each of the 3 regions was measured. Ten cells were analyzed per cell line.

Fluorescence recovery after photobleaching (FRAP). E14 cells (kindly provided by the laboratory of Didier Trono, EPFL) were seeded in FluoroDishes coated with 5 ng µl−1 recombinant mouse or rat E-cadherin Fc chimera protein (R&D systems, #748-EC-050) at a density of 120,000 cells cm−2 24 h before imaging and induced with 500 ng ml−1 of doxycycline.

Cells were imaged using a confocal microscope (ZEISS LSM 700 INVERT) with a ×63 objective at 37 °C and 5% CO2. Channel settings: EYFP 2.4% laser power, 700–900 gain, 41.1 µm pinhole; H342 2.6% laser power, 500–700 gain, 39 µm pinhole. Image dimensions: 50.8 × 50.8 µm, 0.05 µm pixel−1.

To image fluorescence recovery after photobleaching, a circular region of interest (ROI) with a diameter of 20 pixels on chromosomes of metaphase cells was selected for bleaching. In addition, two circular control ROIs of the same size were selected, one on the mitotic chromosomes to be used as a non-bleached control and one next to the cell to control for fluctuations of background fluorescence. Cells were first imaged five times with time intervals of 0.38 s to obtain pre-bleach intensity values, and subsequently the selected ROI was bleached for 0.6 s (five iterations) at 100% laser power. Fluorescence recovery was then imaged for 74 s at intervals of 0.38 s.

The same FRAP experiments were performed in interphase cells (one bleached ROI and one non-bleached control ROI within the nucleus, plus one background ROI next to the cell of interest).

To analyze the recovery time, the mean intensities of the bleached ROI and the two non-bleached control ROIs were measured in all time frames. As mitotic chromosomes tended to move throughout the acquisition, the ROIs on mitotic chromosomes were adjusted manually. The recovery curve of the bleached ROI was normalized based on the intensity values before bleaching and to the two control ROIs. The τ1/2 recovery time was calculated using easyFRAPzmac22 and averaged over 10 mitotic and 10 interphase cells. To analyze the FRAP data, the intensity was normalized as:

\[
I(t) = \frac{I(0) - \sum_{i=1}^{n} I(t_{ROI})}{n} \times \frac{I(t_{ROI})}{I(0)}
\]
Where \( n_{00} \) is the number of pre-blended frames, \( I(0,0) \) is the intensity of the bleached spot minus background and \( I(0,B) \) is the intensity of the non-bleached control minus background.

The fit is calculated for each cell as:

\[
I_{0j1} = I_0 - \alpha e^{-\alpha s + t}
\]

The immobile fraction \( I_{0m} \) and \( I_{11} \) are calculated for each cell as:

\[
I_{0m} = 1 - \frac{\alpha}{1 - (1 - a - \alpha)}
\]

\[
I_{11} = \frac{\ln(2)}{a}
\]

ChiP-seq.

E14 and NIH-3T3 cells were treated overnight with 500 ng/ml of Doxycycline one day after seeding. Briefly, at least 10^6 cells were fixed in 1% formaldehyde for 10 min at room temperature, quenched with 250 mM Tris-HCl pH 8.0, washed with PBS, spun down, and stored at −80 °C. The cell pellet was resuspended in 1.5 ml L1R (50 mM HEPEs-KOH pH 7.4, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10% Glycerol, 0.5% PNP, 0.25% Triton X-100), incubated 10 min at 4 °C, spun down, and resuspended in 1.5 ml L2R (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA), and incubated 10 min at 4 °C. The pellet was spun down and rinsed twice with SDS shearing buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.1% SDS), and finally resuspended in 0.9 ml SDS shearing buffer. All buffers contain 1:100 diluted Protease Inhibitor Cocktail in DMso (Sigma). The suspension was transferred to a milliTUBE 1 ml AFA fiber and sonicated on a E220 focused ultrasonicator (Covaris) using the following settings: 20 min, 200 cycles, 5% duty, 140 W, and input sample aliquots were taken.

Sonicated chromatin was incubated with 500 ng of the anti-HA.11 antibody (BioLegend, # 901501) per 10^6 cells at 4 °C overnight. Protein G Dynabeads (Thermo Fischer) were added to the chromatin and incubated for 3 h at 4 °C. The chromatin was washed several times at 4 °C with 5 min incubation between each wash and 2 min magneticization to collect beads; twice with Low Salt Wash Buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% EDTA), once with High Salt Wash Buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM PMSF), once with High Salt Wash Buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM PMSF, and finally with TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM PMSF). Beads were then resuspended in Elution buffer (TE buffer with 1% SDS and 150 mM NaCl), treated with 400 ng/ml Proteinase K and reverse crosslinked at 65 °C 1100 rpm overnight. Input samples were treated with 100 mg/ml RNase A and 0.4 ng/ml Proteinase K and reverse crosslinked at 65 °C 1100 rpm overnight. Samples were purified using QiaGen MinElute PCR purification kit. Libraries were prepared with NEB-Next ChiP-seq Library Prep Master Mix Set (NEB, #E6040) using insert size selection of 250 bp. Sequencing was performed using 37 nt paired-end reads on an Illumina NextSeq 500. Reads were aligned to the mouse reference genome mm10 (http://www.mousegenome.org) by the deepTools function bamCoverage (with setting ‘–normalizeUsingRPKM’). log2-ratio bigWig files were generated using the deepTools function bamCompare based on merged replicates for each TF expression sample over merged replicates for control (rtTA3G). Heatmaps were generated using deepTools computeMatrix (with setting ‘reference-point’ 1) and plothemap centered around the merged peak set for each TF.

Single molecule fluorescence microscopy.

Cells were seeded on glass-bottom dishes (Delta T culture dishes, Biotech, Pennsylvania, USA) one day before the measurement. To isolate the expression of Halo-tagged TFs, 10 ng ml−1 Doxycycline was added 4 h after transfection. Cells were then incubated with the anti-HA.11 antibody (BioLegend, # 901501) per 10^6 cells at 4 °C overnight. Protein G Dynabeads (Thermo Fischer) were added to the chromatin and incubated for 3 h at 4 °C. The chromatin was washed several times at 4 °C with 5 min incubation between each wash and 2 min magneticization to collect beads; twice with Low Salt Wash Buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% EDTA), once with High Salt Wash Buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM PMSF), once with High Salt Wash Buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM PMSF, and finally with TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM PMSF). Beads were then resuspended in Elution buffer (TE buffer with 1% SDS and 150 mM NaCl), treated with 400 ng/ml Proteinase K and reverse crosslinked at 65 °C 1100 rpm overnight. Input samples were treated with 100 mg/ml RNase A and 0.4 ng/ml Proteinase K and reverse crosslinked at 65 °C 1100 rpm overnight. Samples were purified using QiaGen MinElute PCR purification kit. Libraries were prepared with NEB-Next ChiP-seq Library Prep Master Mix Set (NEB, #E6040) using insert size selection of 250 bp. Sequencing was performed using 37 nt paired-end reads on an Illumina NextSeq 500. Reads were aligned to the mouse reference genome mm10 (http://www.mousegenome.org) by the deepTools function bamCoverage (with setting ‘–normalizeUsingRPKM’). log2-ratio bigWig files were generated using the deepTools function bamCompare based on merged replicates for each TF expression sample over merged replicates for control (rtTA3G). Heatmaps were generated using deepTools computeMatrix (with setting ‘reference-point’ 1) and plothemap centered around the merged peak set for each TF.

To investigate binding properties of TFs, we used two different illumination schemes: (i) continuous movies of cells illuminated with the 638 nm laser to excite the Halo-tagged TF and simultaneously observe the nucleus with a 633 nm laser. The nuclei were labeled with fluorescent SiR ligand (kindly provided by Kai Johnsson, EPFL, Switzerland, (Promega, #G82511)) according to the HaloTag protocol (Promega). DNA was labeled with 0.3 μg ml−1 Hoechst 33342 (ThermoFisher, #62249). Single molecule imaging was performed in phenol free Opti-MEM (ThermoFisher, #11031070) at 37 °C for up to 8 hours at 1 Hz.
This allowed accounting for cellular movements during long acquisition times in illumination scheme (ii).

We determined the fraction of bound TF-Halo molecules per area by dividing the number of molecules identified as bound in illumination scheme (i), \(N_b\), by the area \(A_s\) of the respective Hoechst intensity class, \(a = \text{bright}, \text{intermediate}, \text{or dark}\). We then divided the result by the total number of detected Halo-TF molecules, \(N_*\), and the video capture time \(T\). The result is a measure of binding frequency, which we referred to as pseudo on-rate (\(\psi\)-rate):

\[
\psi = \frac{N_b}{A_sN_*T}
\]

(6)

The \(\psi\)-rate of long bound Halo-TF molecules was determined analogously using the respective molecule counts of illumination scheme (ii):

\[
\psi = \frac{N_b}{A_sN_*T}
\]

(7)

To obtain a simple measure for the residence time \(t_{res}\) of specifically bound molecules, we calculated the average time that Halo-TF molecules spent bound in illumination scheme (ii):

\[
t_{res} = \frac{1}{N_*} \sum_{j=1}^{N_*} t_{b,j}
\]

(8)

where \(t_{b,j}\) is the binding time of molecule \(j\) in illumination scheme (ii). We then ordered the TFs with respect to this measure.

Quantification and statistical analysis. Statistical analyses of violin plots of TF distributions (Fig. 1e), on the MBF for NLS mutants (Supplementary Fig. 1f) and on the MBF for TFs tagged with five arginine residues (Supplementary Fig. 1k) were performed using Wilcoxon rank-sum test. For plots with linear scales, \(r\)-values and \(p\)-values are based on Pearson correlation (Figs. 2d-g, 3c-k, a-b, 5a-d, e, and Supplementary Figs. 1a, 1d, 1f, 2a-b, 3d-f). For plots displayed in log scale, Rho-values and \(p\)-values are based on Spearman’s rank correlation (Figs. 4d-e, 5c, Supplementary Figures 4a-d, 4h-j, 5a and 5c-g). Note that correlations were calculated using averages for replicates. Duplicate values were not averaged for the FOM of top motifs due to enrichment for different motifs.

Code availability. The code to analyze residence times and pseudo-on-rates from single molecule tracks is available as Matlab script on the Dryad Data Repository (https://doi.org/10.5061/dryad.9pc7458). Other codes used in this study are available from the corresponding author upon reasonable request.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Data supporting the findings of this manuscript are available from the corresponding author upon reasonable request. A reporting summary for this Article is available as a Supplementary Information file. ChIP-seq and ATAC-seq data that support the findings of this study have been deposited in GEO (Gene Expression Omnibus) with the accession code GSE119784.

Received: 18 September 2018 Accepted: 8 January 2019
Published online: 30 January 2019

References
1. von Hippel, P. H., Revzin, A., Gross, C. A. & Wang, A. C. Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. Proc. Natl Acad. Sci. USA 71, 4808–4812 (1974).
2. Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models. Theory Biochem. 20, 6929–6948 (1981).
3. Winter, R. B., Berg, O. G. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor-operator interaction: kinetic measurements and conclusions. Biochemistry 20, 6961–6977 (1981).
4. Dör, L., Robs, R. & Mandel-Gutfreund, Y. How motif environment influences transcription factor binding dynamics: finding a needle in a haystack. BioEssays News Rev. Mol. Cell. Dev. Biol. 38, 605–612 (2016).
5. Hu, L., Grosberg, A. Y. & Bruinsma, R. Are DNA transcription factor proteins with multiple DNA binding domains a thousand in one? Biophys. J. 95, 1151–1156 (2008).
6. Marklund, E. G. et al. Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models. Proc. Natl Acad. Sci. USA 110, 19796–19801 (2013).
7. Subekti, D. R. G. et al. The disordered linker in p53 participates in nonspecific binding to and one-dimensional sliding along DNA revealed by single-molecule fluorescence microscopy. Biochemistry 56, 4134–4144 (2017).
8. Takayama, Y. & Clere, G. M. Intra-and intermolecular translocation of the bi-domain transcription factor OctI characterized by liquid crystal and paramagnetic NMR. Proc. Natl Acad. Sci. USA 108, E169–E176 (2011).
9. Vuzman, D., Polonsky, M. & Levy, Y. Facilitated DNA search by multidomain transcription factors: cross talk via a flexible linker. Biophys. J. 99, 1202–1211 (2010).
10. Hettich, J. & Gehrdt, J. C. M. Transcription factor target site search and gene regulation in a background of unspecific binding sites. J. Theor. Biol. 454, 91–101 (2018).
11. von Hippel, P. H. & Berg, O. G. Facilitated target location in biological systems. J. Biol. Chem. 264, 675–678 (1989).
12. Matthew, J. B. & Ohlen, K. D. H. Electrostatic deformation of DNA by a DNA-binding protein. J. Biol. Chem. 260, 5860–5862 (1985).
13. Takeda, Y. et al. Different interactions used by Cro repressor in specific and nonspecific DNA binding. J. Biol. Chem. 261, 8608–8616 (1986).
14. Kenan, K. T., Garcia-Moreno, B. & Freire, E. A calorimetric characterization of the salt dependence of the GCG4 leucine zipper. Protein. Pept. Sci. 4, 1934–1938 (1995).
15. Kalodimos, C. G. et al. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science 305, 386–389 (2004).
16. Barbì, M. & Paillusson, F. Protein-DNA electrostatics: toward a new paradigm for protein sliding. Adv. Protein Chem. Struct. Biol. 92, 253–297 (2013).
17. Voz, T., Wang, S., Poen, G. M. K. & Wilson, W. D. Electrostatic control of DNA intersegmental translocation by the ETS transcription factor ETV6. J. Biol. Chem. 292, 13187–13196 (2017).
18. McNally, J. G., Muller, W. G., Walker, D., Wolford, R. & Hager, G. L. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. J. Biol. Chem. 278, 21004–21009 (2010).
19. Desjardins, G., Okon, M., Graves, B. J. & McIntosh, L. P. Conformational dynamics and the binding of specific and nonspecific DNA by the autoinhibited transcription factor Ets-1. Biochemistry 55, 4105–4118 (2016).
20. Vo, T., Wang, S., Poen, G. M. K. & Wilson, W. D. Electrostatic control of DNA intersegmental translocation by the ETS transcription factor ETV6. J. Biol. Chem. 292, 13187–13196 (2017).
21. Caravaca, J. M. et al. Bookmarking by specific and nonspecific DNA-binding proteins in mitotic chromosomes: regulatory mechanisms and impact on cell fate decisions. FEBS Lett. (2017). https://doi.org/10.1002/1873-4608.12828.
22. Teves, S. S. et al. A dynamic mode of mitotic bookmarking by transcription factors. eLife 5, e22280 (2016).
23. Lerner, J. et al. Human mutations affect the epigenetic/booking function of H3F1B. Nucleic Acids Res. 44, 8097–8111 (2016).
24. Palier, C. et al. Association of chromatin proteins high mobility group box (HMGBox) with mitotic chromosomes. Mol. Biol. Cell. 14, 3414–3426 (2003).
25. Kumar, S., Chaturvedi, N. K., Kumar, S. & Tyagi, R. K. Agonist-mediated down-regulation of HNF1B. J. Biol. Chem. 280, 251–260 (2013).
26. Caravaca, J. M. et al. Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. Genes Dev. 27, 251–260 (2013).
27. Deluz, C. et al. A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. Genes Dev. 30, 2538–2550 (2016).
28. Festuccia, N. et al. Transcription factor activity and nucleosome organisation in mitosis. (2015). https://doi.org/10.1101/043241.
29. Kadowaki, S. et al. Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. Cell 150, 725–737 (2012).
35. Festuccia, N. et al. Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. *Nat. Cell Biol.* 18, 1139–1148 (2016).
36. Seligy, T., Mathur, U. M., Luger, K., Tulin, A. V. & Zaret, K. S. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev.* 23, 804–809 (2009).
37. Soufi, A. et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell 161*, 555–568 (2015).
38. Soufi, A., Donahue, K. & Zaret, K. S. Facilitations and impediments of the pluripotency reprogramming factors’ initial engagement with the genome. *Cell 151*, 994–1004 (2012).
39. Zaret, K. S. Genome reactivation after the silence in mitosis: recapitulating the events of embryonic development. In *Mammalian Gene Regulation*. (Cold Spring Harbor Laboratory Press, 2006), pp. 139–142.
40. Strebing, D. et al. Endogenous fluctuations of OCT4 and SOX2 bias embryonic cell decisions. (2018). https://doi.org/10.1101/299073.
41. Alber, A. B., Paquet, E. R., Biseri, M., Naef, F. & Suter, D. M. Single live cell monitoring of protein turnover reveals intercellular variability and cell-cycle dependence of degradation rates. *Mol. Cell* (2018). https://doi.org/10.1016/j.molcel.2018.07.023.
42. Ginno, P. A., Burger, L., Seebacher, J., Isemanntavic, V. & Schübel, D. Cell cycle-stage specific histone modifications reveal extensive interdependence of the genomic regulatory landscape. *Nat. Commun.* 9, 4048 (2018).
43. Dovat, S. et al. A common mechanism for mitotic inactivation of C2H2 zinc finger DNA-binding domains. *Genes Dev.* 16, 2985–2990 (2002).
44. Rizkallah, R., Alexander, K. E. & Hurt, M. M. Global mitotic phosphorylation and its crosstalk with transcription, replication, and genome integrity. *Proteome Sci.* 2012, 10 (2012).
45. Liu, Y. et al. Widespread mitotic bookmarking by histone marks and transcriptionally active proteins. *Nature* (2018). https://doi.org/10.1038/s41586-018-0161-1.
46. Li, H.-Y. & Zheng, Y. Phosphorylation of RCC1 in mitosis is essential for the regulation of mitotic chromosome dynamics. *Genes Dev.* 22, 777–782 (2008).
47. Iwafuchi-Doi, M. The mechanistic basis for chromatin regulation by pioneer transcription factors. *J. Stat. Softw.* 74, 161–179 (2017).
48. Faial, T. et al. Brachyury and SMAD signalling collaboratively orchestrate embryonic patterning. *Nature* (2018). https://doi.org/10.1038/s41586-018-0161-1.
49. Schick, S. et al. Identifying novel transcriptional regulators with circadian expression profiles. *PLoS Genet.* 10, e1004267 (2014).
50. Mistri, T. K. et al. Selective in vivo chromatin modifications reveal the role of pioneer transcription factors in embryonic stem cells. *Mol. Cell. Biol.* (2018). https://doi.org/10.1128/MCB.00137-18.
51. Anderson, A. M., Weasner, B. M., Weasner, B. P. & Kumar, J. P. Dual roles for the RBPJ transcription factor in embryonic stem cells. *Mol. Cell. Biol.* 32, 1978–2078 (2012).
52. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137 (2008).
53. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature 489*, 57–72 (2012).
54. Iwafuchi-Doi, M. The mechanistic basis for chromatin regulation by pioneer transcription factors. *J. Stat. Softw.* 74, 161–179 (2017).
55. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Cell 158*, 356–358 (2014).
56. Sandelin, A., Alkema, W., Engstroem, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* 32, D91–D94 (2004).
57. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218 (2013).
58. Chen, W., Gardeux, V., Meireles-Filho, A. & Deplancke, B. Profiling of single-cell transcriptomes. *Curr. Protoc. Mouse Biol.* 7, 145–175 (2017).
59. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Oxf. Engl.* 26, 841–842 (2010).
60. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl.* 26, 56–58 (2010).
61. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47 (2015).
62. Clauss, K. et al. DNA residence time is a regulatory factor of transcription repression. *Nucleic Acids Res.* 45, 11221–11230 (2017).

**Acknowledgements**

This work was supported by the Teofo Rossi di Montelera and di Premuda Foundation advised by CARGEST SA, and an anonymous donor advised by CARGEST SA (D.M.S.), the German Research Foundation (DFG 6261/1-1 to J.C.M.G.); the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme (637987 ChromArch to J.C.M.G.) and the DFG Graduate School of Molecular Medicine at Ulm University (to H.A.). We thank Bart Deplancke for providing the pENTR transcription factor library, Rubina Davtyan and Katharina Soukop for their assistance in analyzing parts of the single molecule data, Vincent Gardeux for his assistance in imaging, and Philipp Bucher for critical reading of the manuscript.

**Author contributions**

Conceptualization, M.R. and D.M.S.; Methodology, M.R., D.M.S., H.A., and J.C.M.G.; Formal analysis, M.R., B.A., E.T.F., and T.K.; Investigation, M.R., B.A., H.A., and C.D.; Resources, D.M.S. and J.C.M.G.; Writing—Original draft, M.R., D.M.S., and J.C.M.G.; Reviewed manuscript, M.R., E.T.F., D.M.S., and J.C.M.G.; Funding Acquisition, D.M.S., and J.C.M.G.; Supervision, D.M.S. and J.C.M.G.

**Conflict of interest**

The authors declare no conflict of interest.

**Data availability**

All data generated in this study are available as a CeRCAD database (https://cercad.epfl.ch) and in the Dryad Digital Repository (https://doi.org/10.5061/dryad.2w025).
Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-08417-5.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: Nature Communications thanks Jeffrey Nickerson and the other anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019