GFI1 tethers the NuRD complex to open and transcriptionally active chromatin in myeloid progenitors

Anne Helness1, Jennifer Fraszczak1, Charles Joly-Beauparlant2, Halil Bagci1,7, Christian Trahan1, Kaifee Arman1, Peiman Shoshtarizadeh1, Rían Chen1, Marina Ayoub1,8, Jean-François Côté1,3,4,5, Marlene Oeffinger1,4, Arnaud Droit2 & Tarik Möröy1,5,6

Growth factor independent 1 (GFI1) is a SNAG-domain, DNA binding transcriptional repressor which controls myeloid differentiation through molecular mechanisms and cofactors that still remain to be clearly identified. Here we show that GFI1 associates with the chromodomain helicase DNA binding protein 4 (CHD4) and other components of the Nucleosome remodeling and deacetylase (NuRD) complex. In granulo-monocytic precursors, GFI1, CHD4 or GFI1/CHD4 complexes occupy sites enriched for histone marks associated with active transcription suggesting that GFI1 recruits the NuRD complex to target genes regulated by active or bivalent promoters and enhancers. GFI1 and GFI1/CHD4 complexes occupy promoters that are either enriched for IRF1 or SPI1 consensus binding sites, respectively. During neutrophil differentiation, chromatin closure and depletion of H3K4me2 occurs at different degrees depending on whether GFI1, CHD4 or both are present, indicating that GFI1 is more efficient in depleting of H3K4me2 and -me1 marks when associated with CHD4. Our data suggest that GFI1/CHD4 complexes regulate histone modifications differentially to enable regulation of target genes affecting immune response, nucleosome organization or cellular metabolic processes and that both the target gene specificity and the activity of GFI1 during myeloid differentiation depends on the presence of chromatin remodeling complexes.
The DNA-binding zinc finger proteins GFI1 and GFI1B act as transcriptional repressors by recruiting the complex containing the lysine-specific histone deethylase 1A (LSD1, official name KDM1A) and its cofactor CoREST (official name RCOR1) together with HDACs to sites of specific target genes that harbor a GFI1/GFI1B consensus DNA binding motif1 for a review see refs. 2–5. GFI1 is critical for the differentiation of myeloid cells into neutrophils, which is highlighted for example by GFI1 deficient mice that entirely lack neutrophil granulocytes and, as a consequence, show major defects in their innate immune response6–8. GFI1 and its shorter paralog GFI1B share a 20 aa N-terminal SNAG domain that shows sequence similarity to the N-terminal tail of histone H3. It has been suggested that the GFI1/B SNAG domains and the H3 N-terminus can compete for binding to the same pocket in the LSD1 protein8. Transcriptional repression is achieved by the enzymatic action of LSD1 and HDACs leading to the demethylation of histone H3 Lysine 4 (H3K4) and the deacetylation of histone H3 Lysine 9 (H3K9)9–11. The general applicability of this model has been challenged by recent observations indicating that H3K4 demethylation states do not change in cells upon treatment with an LSD1 inhibitor that not only blocks its enzymatic activity but also leads to the eviction of GFI1 and LSD1 from promoter sites12,13. Moreover, this report finds that LSD1’s demethylase function is not critical for GFI1 function but rather suggests that LSD1’s physical interaction with GFI1’s SNAG-domain is crucial and that LSD1 rather serves as a scaffold for other histone-modifying enzymes such as HDAC612,13. However, further investigation on this matter is necessary since this study was based on a presumed inactive LSD1 mutant for which another study showed some catalytic activity14.

To understand the precise molecular function of GFI1 as a transcriptional regulator, it is necessary to identify the epigenetic modifier complexes that are recruited by this GFI1/LSD1 scaffold and act on chromatin structure and more specifically on histone modifications. Indeed, it was recently shown that GFI1B can recruit members of the so-called BRAF-histone deacetylase (HDAC) (BHC) chromatin-remodeling complex that contains LSD1, members of the CoREST complex (RCOR1, -2 and -3), HDACs as well as a number of the high mobility group of proteins (HMG20A and -B) and other associated proteins15. The nucleosome remodeling and deacetylase (NuRD) complex would be an excellent candidate as well, since similar to the BHC complex it also facilitates histone deacetylase mediated chromatin condensation and some studies have reported that it also contains LSD116–18.

The NuRD complex differs however from the BHC complex in that it contains seven different proteins divided into two sub-complexes, one which comprises an ATP-dependent nucleosome remodeling activity and another which harbors HDACs targeting H3K9 or histone H3 Lysine 27 (H3K27)19–21. Characteristic for the NuRD complex are its major components, the closely related proteins CHD4 (chromodomain helicase DNA binding protein 4 or Mi-2β), CHD3, and CHD522. CHD4 contains an SNF helicase domain and PhD/Chromo domains that mediate its interaction between nucleosomes and methylated histones23. The methyl-CpG-binding domain proteins MBD2 and MBD3 represent its non-enzymatic components and link the ATPase remodeling activities to the HDAC1 and -2 containing subcomplexes. The metastasis-associated proteins MTA1, MTA2, and MTA3 are also part of this subcomplex and mediate binding to DNA, to HDAC1, and to other transcription factors that can recruit NuRD to specific loci in the genome24–26. The proteins RBBP7 and RBBP4 bind histones and most likely have roles as scaffolds27,28. The GATA zinc-finger domain-containing proteins GATAD2A and -B interact with MBD2/3 and are also canonical members of the NuRD complex, although their precise function remains to be determined29.

The NuRD complex can mediate both transcriptional repression or activation6,30 and can be recruited to sites of bivalent or poised targets in chromatin that are primed to be efficiently activated or repressed by modifying the histone marks during progenitor self-renewal or differentiation31,32. Bivalent target promoters or genomic loci with enhancers show both repression marks such as H3K27 trimethylation (H3K27me3) and activation marks such as H3K4 dimethylation (H3K4me2) or H3K27 acetylation (H3K27ac) at the same time and also feature modifications such as H3K4 monomethylation (H3K4me1), which identify so-called primed or induced enhancers31. To exert its function in a tissue and differentiation stage-specific manner, NuRD associates for instance during lymphoid development with lineage-specific transcription factors and co-regulators30, such as IKZF1 (IKAROS), BCL6, or BLIMP121–35.

Here, we show that GFI1 and GFI1B interact with members of the NuRD complex, notably with the chromodomain helicase DNA binding protein 4 (CHD4). In myeloid progenitors, GFI1 occupies chromatin together with CHD4 at specific target regions that are different from those regions occupied by GFI1 or CHD4 alone. These target regions bear characteristics of open chromatin and histone modifications that are associated with active transcription or poised enhancers. GFI1 occupies promoters and genomic sites that are different from those occupied by GFI1/CHD4 complexes and are upstream of different groups of genes that can be distinguished by the enrichment of either IRF1 or SPI1 binding consensus sequences, respectively. During neutrophil differentiation, different levels of chromatin closure and a reduction of H3K4me levels are seen depending on whether sites are occupied by GFI1, CHD4, or GFI1/CHD4 complexes. Lastly, GFI1, CHD4, or GFI1/CHD4 occupy promoters of genes that fall into three distinct groups termed immune system, chromatin/ transcription or poised enhancers. GFI1 occupies promoters and genomic sites that are different from those occupied by GFI1/CHD4 complexes and are upstream of different groups of genes that can be distinguished by the enrichment of either IRF1 or SPI1 binding consensus sequences, respectively. During neutrophil differentiation, different levels of chromatin closure and a reduction of H3K4me levels are seen depending on whether sites are occupied by GFI1, CHD4, or GFI1/CHD4 complexes. Lastly, GFI1, CHD4, or GFI1/CHD4 occupy promoters of genes that fall into three distinct groups termed immune system, chromatin/ nucleosome assembly, and metabolic process, respectively, that can be both upregulate and downregulate during neutrophil differentiation.

Results

GFI1 associates with the nucleosome-remodeling, and histone deacetylase (NuRD) complex. We had used AP–MS (affinity purification and mass spectrometry) to identify proteins that co-purified with Flag-tagged versions of GFI1 and GFI1B in HEK293T cells36,37. This approach revealed the presence of members of the NuRD complex such as chromodomain helicase DNA binding proteins 3 and -4 (CHD3 and -4), metastasis-associated 1, and -2 (MTA1 and -2), and the chromatin remodeling factor retinoblastoma binding protein 4 (RBBP4, also called chromatin assembly factor 1 or CAF-1) in both isolated GFI1 and GFI1B complexes (Fig. 1A, Suppl. Fig. 1A). The peptide coverage for these factors was in a similar range as for those proteins that are known to associate with GFI1 and GFI1B such as HDAC1 and members of the CoREST complex such as RCOR1, 2, or -3 (Fig. 1A). To validate these findings, we used a BioID approach in HEK293T cells for GFI1 associated proteins, compiled the data with known interactions from the IntAct and BioGrid databases38,39, and observed that GFI1 has the potential to associate with four major complexes, notably with the NuRD complex, in agreement with our findings from the AP–MS experiment, but also with the SWI/SNF, CtBP, and cohesin complexes (Fig. 1B, Suppl. Data 1).

Next, we compared the data with our previously reported BioID experiment for GFI1B37 and found again members of the NuRD complex as high ranking candidates for binding partners of both GFI1 and GFI1B (Fig. 1B, Suppl. Table 1), and proteins
already known to bind to these two factors such as HDAC1 and -2, LSD1 and the CoREST proteins RCOR1, -2, and -3 (Fig. 1C, Suppl. Data 1). Quantification of the mass spectrometry results showed high peptide coverage of CHD3 and CHD4, MTA1, -2, -3, and GATA Zinc Finger Domain Containing 2A and 2B (GATAD2A and GATAD2B); the abundance of recovered peptides for GFI1B being higher than for GFI1 (Fig. 1C). Also detectable were peptides for other NuRD complex members such as the methyl-CpG binding domain protein 3 (MBD3) and RBBP4 and -7 (Fig. 1C, Suppl. Data 1).

NuRD components CHD4 and MTA2 associate with GFI1 C- and N-termini. Next, we expressed Flag-tagged versions of GFI1 and GFI1B in HEK293T cells, precipitated NEs with anti-Flag antibodies, and analyzed them using mass spectrometry. The results showed that CHD4 and MTA2 are highly enriched in the precipitated NEs with GFI1 compared to GFI1B, indicating a stronger association of GFI1 with these components. This association is likely to play a role in the function of the NuRD complex, as CHD4 and MTA2 are known to be important for the activity of the NuRD complex.
Fig. 1 GFI1 is associated with protein components of the NuRD complex. a Result from AP-MS (affinity purification and mass spectrometry) experiments with GFI1- or GFI1B-Flag-tagged fusion proteins expressed in HEK293T cells for NuRD and LSD1/CoREST complexes. *: previously known GFI1 interacting proteins. b Interaction map of the GFI1 proteome obtained through BioID data. The interactions between GFI1 and the prey are “weighted” in color according to AvgSpec (e.g., GFI1-KDM1A: 104 AvgSpec, darker color; GFI1-HDAC2: 44 AvgSpec less dark color). The known Prey-Prey interactions (e.g., KDM1A-HDAC1) were imported from databases IntAct and BioGrid, and are represented by fewer dark connections since they cannot be normalized to the GFI1-BioID screen. GFI1-prey interactions remain visible. *: previously known GFI1 interacting proteins. c Dot Plot showing BioID interactions of GFI1-BirA4-Flag or GFI1B-BirA4-Flag with the indicated members of the NuRD or LSD1/CoREST complexes. The node color depicts the average spectral counts. The relative abundance of prey vs. the bait is shown by the circle size. The edge color represents the confidence score of the BioID/MS interaction (5% < BFDR as low confidence score, 1% < BFDR ≤ 5% as medium confidence or BFDR ≤ 1% as high confidence). BFDR Bayesian false discovery rate. *: previously known GFI1-interacting proteins.

GFI1 and CHD4 co-occupy sites of open chromatin and active transcription. We chose primary murine GMPs (lin−/Cd34+Scal−/Cd16/32−/Cd34+ cells) FACS-sorted from bone marrow, Suppl. Fig. 11), in which GFI1 is expressed to further investigate the association between NuRD and GFI1. Chromatin Immunoprecipitation and sequencing (ChIP-seq) experiments with antibodies against murine GFI1 and CHD4 in wild type (WT) GMPs showed 3188 peaks for GFI1 and 4236 peaks for CHD4 and revealed that at 1128 sites occupation of GFI1 and CHD4 overlapped (Fig. 3A). The sites that are co-occupied by GFI1 and CHD4 are primarily at promoter (~30%) and intergenic regions (~38%) (Fig. 3B). When we compared the overall binding of CHD4 to chromatin in sorted GMPs from WT with gene-deficient (Gfi1−/−) mice, we observed that of the 1128 sites that were co-occupied by both GFI1 and CHD4 in WT cells, 841 lost enrichment of CHD4 when GFI1 was absent (Fig. 3C, D), most of them in regions <3 kb from the TSS. At the remaining 287 sites, CHD4 was still enriched regardless of GFI1’s presence or absence (Fig. 3C, D). Except for 9 CHD4 enriched sites, no significant gain in CHD4 bound sites was observed in Gfi1−/− GMPs (Fig. 3C). The 3108 sites that were occupied only by CHD4 in WT cells, remained largely unaffected by Gfi1 deficiency, i.e., CHD4 remained at these sites (Fig. 3C).

Examples for loci occupied by GFI1/CHD4 are the genes Cd34, Csf1r (encoding M-CSFR), or Csf1 (encoding M-CSF) (Fig. 3E, Suppl. Fig. 2A). Gfi1 deletion is associated with a loss of CHD4 at these sites and RNA-seq data showed that this leads to the upregulation of Cd34 mRNA (Fig. 3E) but does not affect Csf1r and Csf1 mRNA levels (Fig. 3F and Suppl. Fig. 2A, B, Suppl. Table 3). For further validation, we chose five other genes selected from the group of 841 loci where CHD4 binding was lost in the absence of GFI1 according to the ChIP-seq data. ChIP-qPCR on these gene loci for CHD4 in primary WT and Gfi1−/− GMPs showed loss of CHD4 enrichment confirming the results of the CHD4 ChIP-seq experiment (Suppl. Fig. 2C, Suppl. Table 3). RNA-seq and flow cytometric analysis demonstrated that Cd34 mRNA and CHD4 protein levels were not altered in Gfi1−/− GMPs compared to WT GMPs (Suppl. Fig. 2D, E). Genes that lost CHD4 binding at their promoters in the absence of GFI1 or maintained CHD4 did not show a significantly different upregulation or downregulation of gene expression in Gfi1−/− vs. wt GMPs (Fig. 3G). Also, upregulation or downregulation of GFI1 target genes in Gfi1−/− vs. wt GMPs did not correlate with the presence or absence of CHD4 (Suppl. Fig. 3A, B, Suppl. Table 3).
Table 3). This suggested that the deletion of GFI1 affects gene expression independently of the presence of CHD4.

To test whether GFI1/CHD4 complexes affect chromatin remodeling, we performed ATAC-seq and ChIP-seq analyses of GMPs with antibodies against methylated and acetylated histone H3. The obtained data were filtered for promoter regions defined here as regions located between less than 2 kb upstream and less than 500 bp downstream of the transcription start sites (TSS), and separately also for enhancer regions selected based on the criteria defined in the Fantom5 enhancer atlas. The data were ordered according to CHD4 occupation and were separated into three groups: occupation by CHD4 alone, by GFI1/CHD4 together, or by GFI1 alone (Fig. 4A). We observed that promoters occupied by GFI1 show higher levels of markers associated with active
transcription such as H3K4me3 and H3K27ac than the promoters occupied by CHD4 alone. Promoters co-occupied by both GFI1 and CHD4 showed intermediate levels of H3K4me3 and H3K27ac (Fig. 4B). In contrast, promoters occupied by CHD4 showed higher levels of H3K27me3 H3K9me3 and H3K4me2, all associated with transcriptional repression, than promoters occupied by GFI1 or by both GFI1 and CHD4 (Fig. 4B), with a similar pattern for the levels of H3K4me2 and H3K4me3, which are markers of transcriptional activation (Fig. 4B). This situation is exemplified by the promoter regions of the genes encoding the GFI1 targets Csf1 (Fig. 4C) and Csf1r (suppl. Fig. 4), which are both expressed in GMPs (Suppl. Fig. 2A, B, Suppl. Table 3 and Fig. 3F).

H3K9 and H3K27 methylation and acetylation patterns at enhancers correlated similarly with GFI1 and CHD4 occupation to those at promoters (Fig. 4D). Also, we found relatively higher levels of H3K4me1 and lower levels of H3K4me3 at sites occupied by GFI1 only and gain of H3K4me3 at sites occupied by GFI1 and values for GFI1/CHD4 occupation in between (Fig. 4D). Analysis of RNA-seq data from GMPs of genes that are next to the promoters or enhancers defined here and were occupied by GFI1 showed higher expression levels than the genes next to sites bound by CHD4 or both CHD4 and GFI1 (Fig. 4E, F, respectively). These findings indicate that GFI1/CHD4 complexes occupy active or bivalent promoters and also active enhancers. These sites have a more closed chromatin configuration when CHD4 is present and have a more open chromatin conformation when GFI1 is present.

De novo motif analysis highlights myeloid-specific gene network. Next, we compared our ChIP-seq data for GFI1 and CHD4 binding and H3K4 methylation with a data set for the transcription factor CEBPα, also done in GMP cells, which has been shown to co-occupy promoter sites together with GFI1 in myeloid cells.42 The data were again analyzed for promoter regions as defined above, they were ordered according to GFI1 binding and were separated into four groups: occupation by GFI1 alone, by GFI1/CEBPα, by GFI1/CEBPα/CHD4, and by GFI1/CHD4 (Fig. 5A). A de novo motif analysis43 of consensus DNA binding sites at promoter regions revealed that the loci bound by GFI1 or by GFI1/CEBPα were very highly enriched for the GFI1/1B DNA binding motif, as expected, but the enrichment was even higher for the consensus DNA binding motif for the transcription factor IRF1, which contains the 5′-AANNAAA-3′ core sequence for all IRF factors (Fig. 5B). In addition, ETS2, STAT3, and SOX3 binding motifs were highly enriched at sites occupied by GFI1, but in contrast, KLF, SP1, RUNX, and ATF3 binding motifs were highly enriched at sites occupied by GFI1/CEBPα (Fig. 5B). At sites where GFI1 and CHD4 are both present, neither Irf nor Gfi1 consensus sites were enriched. However, the binding motifs for SPI1 (PU.1) and the related factor SPI-C were found to be enriched with the highest $E$ values (Fig. 5B). Sites occupied by GFI1/CHD4/CEBPα showed again enrichment for the PU.1 and SPI-C motifs, but in addition also the binding sequences for CEBPα and RUNX1 (Fig. 5B). This differential enrichment of binding motifs suggests that GFI1 or GFI1/CEBPα complexes bind to other genomic loci that those occupied by GFI1/CHD4 or GFI1/CEBPα/CHD4 complexes. Aggregation plots showed that H3K4me2 and H3K4me1 methylation levels at promoter regions occupied by GFI1 or GFI1/CEBPα complexes are differentially depleted or enriched at the TSS or 3′ of the TSS, respectively, according to the presence of CHD4 (Fig. 5C). The data suggest that GFI1 is more efficient in the depletion of H3K4me2 and -me1 marks in the presence of CHD4 (Fig. 5C). This also suggests that GFI1 and CEBPα are associated with promoters with a higher transcriptional activity than those where CHD4 is also present either together with GFI1 or GFI1/CEBPα complexes, and, in addition, suggesting again that the presence of CHD4 is associated with repressive histone marks.

Chromatin remodeling by GFI1, CHD4, or both CHD4 and GFI1 during neutrophil differentiation. The developmental steps from GMPs to mature neutrophils have been clarified using surface markers to define pre-neutrophils (preNeu), immature neutrophils (immNeu), and mature neutrophils (matNeu) stages.44 To explore how sites that are occupied by GFI1, CHD4, or both in GMPs are altered during neutrophil differentiation with regard to chromatin openness or H3K4 dimethylation levels, we sorted GMP, preNeu and matNeu populations from bone marrow following a published strategy (Suppl. Fig. 5A, Suppl. Fig. 12, Suppl. Table 5). Monitoring GFP expression in these same subsets isolated from Gfi1-GFP knockin mice45 showed that the Gfi1 gene is expressed and the promoter is active (Suppl. Fig. 5B, Suppl. Table 5). However, analysis by western blot of the sorted cells showed that GFI1 protein expression levels are higher in matNeus than in preNeu or immatNeu cells (Suppl. Fig. 5C, Suppl. Fig. 13, Suppl. Table 5). RNA-seq reads of groups of genes specific for chemotaxis or phagocytosis were obtained and were congruent with the published pattern (Suppl. Fig. 6A, B, Suppl. Table 5), similar to other genes regulated during myeloid differentiation indicating that the sorted populations represent indeed
Fig. 3 GFI1 and NuRD complex member CHD4 co-occupy sites at chromatin from granulocyte myeloid progenitors (GMPs). 

- **a**: Venn diagram indicating the number of sites occupied by GFI1, CHD4, or both in GMPs according to ChIP-seq experiments done with GMPs. 
- **b**: Relative distribution of sites occupied by GFI1, CHD4, or both in percent of all sites in promoter regions, 3′ and 5′ untranslated regions (UTRs), introns, exons, or other distal regions. 
- **c**: Venn diagram indicating the number of sites occupied by GFI1 in WT GMPs, CHD4 in WT GMPs and CHD4 in GFI1 deficient GMPs. 
- **d**: Heatmap of ChIP-seq results for sites occupied by GFI1, or GFI1 and CHD4 in WT GMPs or in GMPs sorted from GFI1 deficient mice (Gfi1 KO). Genes are sorted according to GFI1 occupation. Red: genes occupied by GFI1, but not by CHD4, dark blue: \( N = 841 \) sites occupied by both CHD4 and GFI1 in wt GMPs which lose CHD4 occupation when Gfi1 is deleted (i.e., in Gfi1 KO GMPs), pale blue: \( N = 287 \) sites occupied by both CHD4 and GFI1 in wt GMPs which maintain CHD4 occupation when Gfi1 is deleted (i.e., in Gfi1 KO GMPs). 
- **e**: Schematic depiction of the locus encoding murine CD34 and the Csf1r locus encoding M-CSFR. 
- **f**: The promoters of the genes targeted by GFI1 and by CHD4 were separated into two groups according to the presence of CHD4 in Gfi1 KO cells (w_CHD4 and wo_CHD4). The groups were then separated into two subgroups according to the direction of the fold change in expression after Gfi1 deletion (up or down). For the comparison w_CHD4_up vs. wo_CHD4_up, the p-value is 0.80. For the comparison w_CHD4_down vs. wo_CHD4_down, the p-value is 0.0095. The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles). The whiskers extend up to 1.5 times the inter-quartile range and data beyond this point were plotted individually.
preNeu and matNeu cells as previously reported (Suppl. Fig. 7, Suppl. Table 5).

Next, we performed both ATAC-Seq and ChIP-Seq experiments to determine whether and how chromatin openness and H3K4 di-methylation levels change at the loci that are occupied by GFI1 and CHD4 or both in GMPs during neutrophil differentiation. Data were filtered as described above for promoter and enhancer regions, were ordered according to CHD4 occupation, and were separated into three groups: occupation in GMPs by CHD4 alone, by GFI1/CHD4, and by GFI1 alone (Fig. 6A, B, Suppl. Table 5). We compiled the data for the locus of a typical myeloid-specific enhancer localized in the 3′ region of the PLBD1 gene and several sites downstream (Fig. 6C, Suppl. Table 5). Sites 3′ of the Plbd1 gene that are occupied by
**Fig. 4 GFI1 and GFI1/CHD4 occupy actively transcribed regions.** a Heatmap of ATAC-seq and ChIP-seq analyses were obtained with antibodies against GFI1, CHD4, and methylated and acetylated histone H3 from GMPs. Shown are reads for promoter regions with 5 kb 5′ and 3′ of TSS ordered according to CHD4 occupation and were separated into three groups: occupation by CHD4 alone, by GFI1 and CHD4, and by GFI1 alone. (TSS transcription start site). b Aggregation plots for promoters occupied by GFI1, CHD4, and methylated and acetylated histone H3 with the data shown in (a). Shown is the fold enrichment over a region of 2 kb 3′ and 5′ of the TSS. (TSS transcription start site). The differences in epigenetic marks for regions bound by CHD4 and or GFI1 in (b) and (d) (see below) are due to background signals since in some cases reads are counted which do not necessarily represent sites because cut off parameters were kept identical throughout the experiment. The signal in (b) and (d) represent the intensity of the reads that were mapped in the selected regions. To ensure all regions are compared between samples, the same parameters were used for the peak calling step in the same category. c Exemplary depiction of the ATAC-Seq and ChIP-Seq data at the GFI1 target gene Csf1. Indicated are the tracks corresponding to the individual experiments, the gene, and the TSS. Annotations represent locations on the mouse genome version GRCm38 (mm10). d Aggregation plots for enhancers occupied by GFI1, CHD4, and methylated and acetylated histone H3. Shown is the fold enrichment over a region of 2 kb 3′ and 5′ of the TSS (transcription start site). e Analysis of RNA-seq data from GMPs of genes that are next to the promoters of f enhancers and were occupied by GFI1, CHD4, or both CHD4 and GFI1. The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles). The whiskers extend up to 1.5 times the interquartile range and data beyond this point were plotted individually.

Fig. 5 De novo motif analysis and histone methylation aggregation plots for sites occupied by GFI1, CHD4, and CEBPα in GMPs. a Comparison of ChIP-seq data for the occupation of GFI1, CHD4, and H3K4 methylation at promoter regions in GMPs with an analogous ChIP-seq data set for the transcription factor CEBPα. Data are ordered according to GFI1 occupation and are separated into four groups: occupation by GFI1 alone, by GFI1/CEBPα, by GFI1/CHD4, and by GFI1/CHD4. b De novo motif analysis of consensus DNA binding sites at promoter regions by GFI1 alone, by GFI1/CEBPα, by GFI1/CEBPα/CHD4, and by GFI1/CHD4. c Aggregation plots of histone H3K4 methylation at promoters as defined in (a) for regions 2 kb 5′ or 3′ of the TSS (transcription start site).

GFI1/CHD4 in GMPs showed a decrease in ATAC-seq reads and low-level H3K4me2 marks compared to the site within the Plbd1 gene where CHD4 is present without GFI1. At this site, ATAC-seq levels increased and H3K4me2 levels accumulated in preNeu and matNeu cells (Fig. 6C, Suppl. Table 5).

To better quantify and facilitate the integration of signals from GMPs, preNeu, und matNeu cells of regions that are occupied by GFI1, CHD4, or both GFI1 and CHD4 in GMPs, we performed a Metagene analysis. This allowed us to directly compare the enrichment profiles of H3K4me2 and ATAC-seq signals between cell types and experiments and to visualize the result. We used the same definition of promoter and enhancer regions as in the previous analyses to extract the enrichment signal, which was then normalized using the NCIS algorithm. This permitted us to compare the mean coverage values between cell types (GMP, preNeu, matNeu), since the normalization integrates both background noise and the size of the library (RPM). Bins cover 1000 base pairs 5′ and 3′ of the TSS for the promoter region and of the defined center region (Ctr) of the enhancer (at 50 bins).

Using the ChIP-seq data from GMPs, we compiled values from 2190 promoter sites occupied by GFI1, 1779 sites occupied by CHD4, and 341 sites occupied by both CHD4 and GFI1. We observed that mean coverage of ATAC-seq RPM values for these promoters was around 15-fold lower in preNeu and matNeu cells than in GMPs (Fig. 6D, Suppl. Table 1), indicating that chromatin regions close during neutrophil differentiation and that sites occupied in GMPs by CHD4 are more contracted than sites occupied in GMPs by GFI1 or GFI1/CHD4. Similarly, at the same sites, the mean coverage of H3K4me2 levels were lower overall in preNeu and matNeu cells compared to GMPs (Fig. 6D, Suppl. Table 5). They were also lower at the promoter sites
occupied by GFI1 compared to sites occupied by CHD4 or GFI1/CHD4 complexes at the TSS (blue line Fig. 6D, Suppl. Table 2, Suppl. Table 5), suggesting that GFI1 may be active in removing methyl groups from H3K4 in GMPs and preNeu cells and that this is modiﬁed when CHD4 is present together with GFI1.

For enhancers, we included values from 382 sites occupied by GFI1, 2375 sites occupied by CHD4, and 173 sites occupied by both CHD4 and GFI1 in GMPs. The RPM values of ATAC-seq reads indicated a highly signiﬁcant chromatin contraction for these sites in preNeu and matNeu cells versus GMPs (Fig. 6E, Suppl. Data 1, Suppl. Table 5), but no differences were seen in each cell type between enhancer centers that were occupied by GFI1, CHD4, or GFI1/CHD4 in GMPs, only at regions 5′ and 3′ of the center when occupied by GFI1 (Fig. 6E, Suppl. Data 2, Suppl. Table 5). H3K4me2 values at these enhancers were again lower in preNeu and matNeu cells than in GMPs (Fig. 6E, Suppl. Data 2, Suppl. Table 5). However, sites at the enhancer centers occupied in GMPs by CHD4 had signiﬁcantly lower H3K4me2

Fig. 6 H3K4 methylation and chromatin status at sites occupied by GFI1, CHD4, or both CHD4 and GFI1 during neutrophil differentiation. a Heatmap of data from ChIP-Seq and ATAC-Seq experiments and to determine the occupation of GFI1, CHD4, and H3K4 di-methylation patterns during neutrophil differentiation from GMPs via preNeu (preneutrophils) to matNeu (mature neutrophils). Data were ﬁltered for promoter regions, ordered according to CHD4 occupation, and were separated into three groups: occupation by CHD4 alone, by GFI1/CHD4 and by GFI1 alone. b Heatmap of data from ChIP-Seq and ATAC-Seq experiments as in (a), but for enhancer regions. c ATAC-Seq and ChIP-seq read on loci covering the myeloid-speciﬁc enhancers at the 3′ end of the Plbd1 gene and at downstream regions for sites occupied by GFI1, CHD4, or both, CEBPa and carrying the indicated histone marks. Annotations represent locations on the mouse genome version GRCm38 (mm10). d Metagene analysis for ATAC-seq reads and H3K4me2 levels (e) in RPM values for GMPs, preNeu, and matNeu cells at promoters occupied by CHD4 alone, by GFI1/CHD4, and by GFI1 alone. Shown are regions 2 kb 5′ and 3′ of the TSS. e Metagene analysis for ATAC-seq reads and for H3K4me2 levels in mean RKPB for GMPs, preNeu and matNeu cells at enhancers occupied by CHD4 alone, by GFI1/CHD4, and by GFI1 alone. Shown are regions 2 kb 5′ and 3′ of the site of the enhancer (Ctr).
levels than sites occupied by GFI1 (Fig. 6E, red lines and blue lines, respectively, Suppl. Table 5), whereas sites occupied by GFI1/CHD4 had intermediate levels (Fig. 6E, green line, Suppl. Data 2, Suppl. Table 5). This indicates that in GMPs GFI1 and CHD4 act differently on H3K4 methylation at enhancer sites than at promoters and that this differential pattern is retained in preNeu and matNeus cells.

Genes occupied by CHD4, GFI1, or both belong to the different categories. Next, we reordered the Chip-Seq and ATAC-seq data once again according to CHD4 occupation at promoter and enhancer sites but now separated them into 6 groups: occupation by CHD4 alone, by GFI1/CHD4, or by GFI1 alone and according whether these groups of genes were up or down-regulated during the differentiation from GMPs to preNeu and matNeus (Fig. 7A, B, Suppl. Table 5). A GO classification showed that genes occupied by CHD4 alone are found in pathways typical for the immune and inflammatory response regardless of whether they were upregulated or downregulated during differentiation from GMPs to matNeus cells (Fig. 7B, Suppl. Table 5). Up and downregulated genes co-occupied by both CHD4 and GFI1 encode regulators of chromatin assembly and nucleosome organization, while genes occupied by GFI1 alone are involved in metabolic processes (Fig. 7C, Suppl. Table 2, Suppl. Table 5), suggesting that GFI1, CHD4, and the GFI1/CHD4 complex regulate separate defined groups of genes during neutrophil differentiation.

Discussion
In the present study, we describe results of BioID experiments indicating that the SNAG domain and zinc-finger transcriptional repressor GFI1 associates with several chromatin remodeling complexes such as the NuRD complex, which also contains LSD1 and members of the LSD1/CoREST complex such as Rcor1, -2 and -3, but also the CtBP and SWI/SNF chromatin remodeling and repressor complexes. We chose to elucidate the biological meaning of the interaction with the NuRD complex since we find that GFI1 can associate with almost all its components. We have focused on the chromodomain helicase CHD4, one of the NuRD complex components, and have used primary murine cells representing stages of neutrophil differentiation as a model system given the particularly important role of GFI1 in this process. We demonstrate that GFI1 can recruit CHD4 to specific sets of target genes that regulate processes such as nucleosome organization and chromatin assembly. While GFI1 occupies target gene promoters containing its own cognate DNA consensus sequence and those for IRF1, GFI1/CHD4 complexes target genes at consensus sites for ETS-related factors such as SPI1 (PU.1) and SPIC. Analysis of histone modifications and chromatin structure indicates that GFI1 and GFI1/CHD4 complexes occupy active or bivalent promoters and active enhancers, both up and downregulated during neutrophil differentiation.

Previous studies showed that GFI1 binds to LSD1 and HDACs, but these histone-modifying enzymes are also members of the NuRD complex. It is thus possible that GFI1 enters into association with the NuRD complex either via LSD1 or HDACs or through CHD4 or MTA2, the proteins identified here to bind to GFI1. Immunoprecipitation and mass spectrometry data, BioID experiments with full-length GFI1 and a GFI1 mutant that lacks the SNAG domain, biochemical analyses with mutated and truncated GFI1 proteins support a model in which GFI1 associates with the CHD4 and MTA2 components of the NuRD complex through several regions including the SNAG and zinc finger domains. Given that the SNAG domain specifically and directly interacts with LSD1 and at the same time with components of the NuRD complex such as CHD4 and MTA2, SEC fractionation data support this view, since the GFI1 complexes that elute around 400–500 kDa contain LSD1 and Rcor1, but also CHD4 and MTA2. Data from the SEC however also indicate that GFI1 is a member of other, high molecular weight complexes eluting at around 2 MDa, in which NuRD components or members of the LSD1/CoREST complex are much less abundant. This would be in agreement with the notion that GFI1 acts as a member of several different chromatin remodeling complexes, which is also supported by our BioID results. The precise nature of the 2 MDa GFI1 complex, however, remains to be determined.

Since we can demonstrate that GFI1 and CHD4 co-occupy promoter and enhancer sites in GMPs, it is conceivable that GFI1 recruits the NuRD complex to these regions in the chromatin. For a considerable fraction of these sites, the recruitment is likely to occur directly through GFI1, since the deletion of GFI1 abrogates CHD4 occupation. The deletion of GFI1 did not lead to a large redistribution of CHD4 to new sites as it was observed in similar experiments with the transcription factor IKAROS in B lymphocytes, which underlines an important difference to our study with myeloid cells. According to the existing model, sites occupied by GFI1 should be depleted of H3K4me2, -me1, and H3K9acetyl marks and be transcriptionally silent. However, our ChIP-seq data suggest that GFI1 and CHD4 and GFI1/CHD4 complexes are located at transcriptionally active promoters in GMPs. The enrichment of markers for active transcription such as acetylation at H3K27 and di- and tri-methylation at H3K4 and the depletion of markers indicating transcriptional repression such as tri-methylation of H3K27 and H3K9 at sites where GFI1 binds support this notion. However, while H3K4me1 that indicates active transcription is strongly depleted at the TSS of GFI1 occupied genes, it is possible that these are rather bivalent promoters, which are characterized by the presence of both active and repressive histone marks.

Similarly, enhancers occupied by GFI1, CHD4, and GFI1/CHD4 complexes show H3K4me1 and H3K27 acetylation and absence or depletion of H3K27me3 or H3K9me3 marks, which is consistent with active enhancers as opposed to inactive or poised enhancers, which would be characterized by the presence of both H3K4me1 and H3K27me3. Of interest here is that sites occupied by CHD4 show relatively lower levels of active histone marks and relatively higher levels of histone marks associated with inactive promoters and enhancers compared to sites occupied by GFI1 alone. Sites occupied by both GFI1 and CHD4 consistently show levels in between suggesting that GFI1 can modulate the effect of the NuRD complex on chromatin remodeling. RNA-seq data show that genes associated with promoters or enhancers occupied only by GFI1 are expressed at a higher level than genes occupied only by CHD4, while genes associated with promoters or enhancers occupied by both GFI1 and CHD4 have expression levels that fall in between these extremes. This supports the view that GFI1 is associated with active promoters and enhancers and can modulate the effect of the NuRD complex. This suggests also that the presence of the NuRD complex at sites occupied by GFI1 reduces transcription relative to sites occupied by GFI1 alone.

The active chromatin found here at sites occupied by GFI1 and GFI1/CHD4 complexes in progenitor cells such as GMPs seems at first sight to contradict the established function of both GFI1 and the NuRD complex as transcriptional repressors. In particular, H3K4me2 levels were expected to be depleted at GFI1-occupied sites given its association with LSD1, but this was not always observed when comparing values from aggregation plots. However, it is conceivable that GFI1 or GFI1/NuRD complexes are in a poised state in GMPs and become active only upon differentiation signals that enable them to repress the transcription of genes specific for neutrophil.
differentiation. Aggregation plots from ATAC-seq analyses suggest more accessibility at promoter sites at a narrow region around the TSS compared to the enhancer sites. However, the pattern of the ATAC-seq plot is different at enhancers where it is broader compared to promoters, where it is narrower. This does not allow to conclude that promoters are more accessible than enhancers but rather indicates that different mechanisms are probably in place controlling accessibility at enhancer versus promoter sites occupied by GFI1.

Our Metagene analyses, which allowed a quantification and direct comparison between cell types and experiments, provided some clarification and indicated that chromatin openness and H3K4me2...
levels at sites occupied by GF11, CHD4, and GF11/CHD4 complexes are strongly reduced when GMPs differentiate into preNeu and matNeu cells. In addition, chromatin was more compacted at enhancers bound by CHD4 than at those occupied by GF11 or GF11/CHD4. Interestingly, however, H3K4me2 patterns were lowest at the TSS of promoters occupied by GF11 alone relative to sites occupied by CHD4 or GF11/CHD4 complexes, which would be in agreement with the presumed function of GF11 to enable the removal of methyl groups from H3K4 via LSD1. At enhancers, however, H3K4me2 levels are highest at sites occupied by GF11 compared to sites bound by CHD4 or both GF11 and CHD4, indicating that the role of GF11 as a facilitator of chromatin remodeling can be different at promoters and enhancers.

Previous studies have shown that a regulatory network exists between the transcription factors PU.1, IRF8, C/EBPα, and GF11 in myeloid differentiation and that GF11 can bind to regions in chromatin together with PU.1 or C/EBPα or IRF8. Evidence from our motif analysis suggests that CHD4 affects the role of GF11 in this regulatory network. While GF11 and GF11/C/EBPα complexes bind to sites similarly enriched for IRF and GF11 consensus binding sites, complexes that contain CHD4 are found at sites enriched for PU.1 binding sites but lack a GF11 consensus site. This suggests that the presence of CHD4 redirects GF11 to a different set of promoters, or at least to a different region of a promoter. This also points to the possibility that when in a complex with CHD4, GF11 may not be required to directly contact DNA or does so through a component of the NuRD complex such as MTA2, which has a DNA binding domain. The analysis of genes up and downregulated during neutrophil differentiation from GMPs via preNeu to matNeu cells supported this view since it showed that genes occupied by CHD4, GF11, or GF11/CHD4 complexes belong to different groups. We propose that in myeloid progenitors GF11 tethers the NuRD complex through the binding to CHD4 and other components to a specific set of target genes with active or bivalent promoters and active enhancers but remains at a poised state. During neutrophil differentiation, CHD4, GF11, and GF11/CHD4 complexes enable the transcriptional regulation of different sets of target genes affecting the immune response, cellular metabolic processes, or nucleosome organization through chromatin remodeling. It has been reported that the deletion of CHD4 leads in mice to the reduction of neutrophil granulocytes and B-cell precursors, which are the same cells that require GF11, and to an increase in erythroid precursors, which are known to require GF11B. Although the biological role of the interaction of GF11 and GF11B with members of the NuRD complex remains to be investigated in more detail, and the absence of an in vivo model such as the use of Chd4 gene-deficient mice represent a limitation of the present study, our findings suggest that the association of GF11 and GF11B with the NuRD complex is important for all early hematopoietic lineages, notably for the development of the myeloid lineage.

**Methods**

**Mouse strains.** The mice used in this study have been bred on C57BL/6 genetic background and were maintained in a specific-pathogen-free plus environment at the Institut de recherches cliniques de Montreal (IRCM). The Institutional Animal Ethics Review Board of the IRCM approved all animal protocols and experimental procedures were performed in compliance with IRCM and CCAC (Canadian Council of Animal Care) guidelines. The work was performed under the IRCM identification number 2020-09 TM.

**BioloD and MS data analysis.** BioloD experiments were carried out similarly to those in previous reports, with modifications. Briefly, Flp-In T-Rex HEK293T cell lines from 15-cm plates expressing GF11WT-BirA*-Flag, GF11ASNAG-BirA*-Flag in a tetracycline-inducible manner were treated with 50 μM biotin and 1 μg/ml tetracycline for 24 h. The following day, the medium was removed and cells were washed three times with cold phosphate-buffered saline (PBS). Cell pellets were lysed in 1.5 ml radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris- HCl (pH 7.4), 150 mM NaCl, 1% NP-40, 1 ml EDTA, 0.1% SDS, 0.5% sodium deoxycholate, Sigma protease inhibitor 1:500 (P8340 5ml, Sigma), 1 ml DTT, 1 mM PMSF, supplemented with 250 U benzamidine, 5 mM DTT, 1 mM TCEP, and iodoacetamide, and then washed twice with 1 ml RIPA buffer. Beads were then washed three times in 1 ml of 50 mM ammonium bicarbonate (ABC) (AB0032, Biobasic). Samples were then subjected to reduction/alkylation with TCEP and iodoacetamide, and then trypsin digestion by the IRCM proteomics core facility. Trypsin-digested samples were desorbed in 2% acetonitrile prior to mass spectrometry analysis. Samples were analyzed by Orbitrap Fusion Mass Spectrometer (Thermo Fisher) at the IRCM proteomics core facility. Peptide and protein identification analyses for GF11 WT-BirA*-Flag or GF11ΔSNAG-BirA*-Flag samples were performed similarly to those in previous reports. Bira*-Flag controls and GF11BWT-BirA*-Flag samples were generated in our previous work. BioloD-MS data were processed using the ProHits software. The Protein wizard tool was used to convert RAW files to mzXML files. MS/MS spectra were searched by using Human ReSeq version 57. Peptide search was carried out through Mascot and X!Tandem. The Trans-Proteomic Pipeline (TPP) suite and iProphet integrated into ProHits were used for peptide validation and quantitation.

**Dot plot analysis.** Significance Analysis of INTeractome (SAINT) implemented in ProHits was used to score protein interactions based on the average number of spectral counts and for data filtering. In brief, SAINT files of the wild-type GF11 (GF11WT-BirA*-Flag) and GF11ΔSNAG (GF11ΔSNAG-BirA*-Flag) and the wild-type GF11B (GF11BWT-BirA*-Flag) were filtered using BirA*-Flag controls. Homogenate (B type pestel) and GF11WT-GFP (Fig. 1C), and GF11WT-GSNAG (Fig. 2D) generated in ProHits were processed by using the ProHits-viz tools to carry out dot plot analyses. Dot plots displaying the protein interaction data included the average number of spectral counts for each prey (AvgSpec), the relative prey abundance towards baits (relative abundance), and the Bayesian false discovery rate (BFDR) as confidence score for the indicated interacting proteins.

**Protein network analysis.** Protein network analyses were performed by using Cytoscape. The SAINT output file of GF11-BirA*-Flag BioID/MS data was imported to Cytoscape. The existing protein–protein interaction network between prey identified in GF11-BirA*-Flag BioID/MS screen was imported by using the Biogest network analysis tool and merged with the GF11IS-BirA*-Flag BioID/MS data. The merged protein network was then subjected to MOCADE clustering to visualize protein complexes that are connected with the GF11 bait.

**Preparation of nuclear extracts (NE) for SEC.** Kasumi cells (2.5 billion) were harvested from cell culture media by centrifugation (at RT) for 10 min at 2000 rpm. Divided into two 50 mL falcons. Pelleted cells were suspended in five volumes of 4 °C PBS and collected by centrifugation (at RT) for 10 min at 2000 rpm. Divided into two 50 mL falcons. Pelleted cells were suspended in five volumes of 4 °C PBS and collected by centrifugation; subsequent steps were performed at 4 °C. Nuclear extracts were subjected to second low-speed centrifugation as before and suspended in two packed cell volume (volume prior to centrifugation) and then washed three times in 1 ml of 50 mM ammonium bicarbonate (ABC) (AB0032, Biobasic). Samples were then subjected to reduction/alkylation with TCEP and iodoacetamide, and then trypsin digestion by the IRCM proteomics core facility. Trypsin-digested samples were desorbed in 2% acetonitrile prior to mass spectrometry analysis. Samples were analyzed by Orbitrap Fusion Mass Spectrometer (Thermo Fisher) at the IRCM proteomics core facility. Peptide and protein identification analyses for GF11 WT-BirA*-Flag or GF11ΔSNAG-BirA*-Flag samples were performed similarly to those in previous reports. Bira*-Flag controls and GF11BWT-BirA*-Flag samples were generated in our previous work. BioloD-MS data were processed using the ProHits software. The Protein wizard tool was used to convert RAW files to mzXML files. MS/MS spectra were searched by using Human ReSeq version 57. Peptide search was carried out through Mascot and X!Tandem. The Trans-Proteomic Pipeline (TPP) suite and iProphet integrated into ProHits were used for peptide validation and quantitation.

**Dot plot analysis.** Significance Analysis of INTeractome (SAINT) implemented in ProHits was used to score protein interactions based on the average number of spectral counts and for data filtering. In brief, SAINT files of the wild-type GF11 (GF11WT-BirA*-Flag) and GF11ΔSNAG (GF11ΔSNAG-BirA*-Flag) and the wild-type GF11B (GF11BWT-BirA*-Flag) were filtered using BirA*-Flag controls. Homogenate (B type pestel) and GF11WT-GFP (Fig. 1C), and GF11WT-GSNAG (Fig. 2D) generated in ProHits were processed by using the ProHits-viz tools to carry out dot plot analyses. Dot plots displaying the protein interaction data included the average number of spectral counts for each prey (AvgSpec), the relative prey abundance towards baits (relative abundance), and the Bayesian false discovery rate (BFDR) as confidence score for the indicated interacting proteins.

**Protein network analysis.** Protein network analyses were performed by using Cytoscape. The SAINT output file of GF11-BirA*-Flag BioID/MS data was imported to Cytoscape. The existing protein–protein interaction network between prey identified in GF11-BirA*-Flag BioID/MS screen was imported by using the Biogest network analysis tool and merged with the GF11IS-BirA*-Flag BioID/MS data. The merged protein network was then subjected to MOCADE clustering to visualize protein complexes that are connected with the GF11 bait.

**Preparation of nuclear extracts (NE) for SEC.** Kasumi cells (2.5 billion) were harvested from cell culture media by centrifugation (at RT) for 10 min at 2000 rpm. Divided into two 50 mL falcons. Pelleted cells were suspended in five volumes of 4 °C PBS and collected by centrifugation; subsequent steps were performed at 4 °C. Nuclear extracts were subjected to second low-speed centrifugation as before and suspended in two packed cell volumes (volume prior to the initial wash with buffer A) of buffer A (10 mM HEPES (pH 7.9), 1.5 mM MgCl2, 10 mM KCl, 0.5 mM DTT, 0.5 mM PMSF) and allowed to stand for 10 min. The cells were collected by centrifugation as before and suspended in two packed cell volumes (volume prior to the initial wash with buffer A) of buffer A and lysed by 15 strokes of a Dounce homogenizer B type pestel. The homogenate was then cleared microscopically for cell lysis and centrifuged for 10 min at 2000 rpm to pellet nuclei. The pellet obtained from the low-speed centrifugation of the homogenate was subjected to second
centrifugation for 15 min at 25,000g (Beckman Coulter JA-25.15 Fixed Angle Rotor), to remove residual cytoplasmic material and this pellet was designated as nuclei fraction. The nuclear pellet was resuspended in 500 µl of NE (50 mM Tris-HCl, pH 7.5) containing protease inhibitors (PI; 40 µg/ml and PMSF; 0.5 mM) and vortexed every 2 min. Protein-DNA complex was sonicated 3 times for 10 s at 50% output using the Branson digital sonifier, once for 15 s at 50% output then put on ice for 10 min and centrifuged at 13,000 rpm for 10 min at 4 °C. Totally, 3.25 ml of supernatant (the equivalent of 2.5 × 10^9 cells) and aliquoted into 250 µl (12 tubes) and 50 µl (5 tubes) at a concentration of 2.6 µg/ml.

**Flow cytometry analysis, sorting of GMPs**

Flow cytometry analysis was performed using BD FACs Diva software (BD Biosciences) or FlowJo (for histogram overlays; Tree Star). For cell sorting, lineage negative BM cells were first depleted using a mouse lineage cell depletion kit (Miltenyi Biotec) then applied to a five-laser FACSaria II sorter (BD Biosciences) (Suppl. Figs. 11 and 12).

**Cell culture.**

THP-1 (ATCC TIB-202), KASUMI (ATCC CRL-1573) cells were maintained in DMEM media (Multiscell) supplemented with 10% bovine growth serum (RMBio Fetalgro) and 100 IU Penicillin and 100 µg/ml Streptomycin (Multiscell). HEK293T (ATCC CRL-1573) cells were maintained DMEM media (Multiscell) with the above-mentioned supplements. We verified that none of the cell lines used in this study were found in the Register of Misidentified Cell Lines maintained by the International Cell Line Authentication Committee. All cell lines used were tested and shown to be negative for mycoplasma contamination using PCR amplification using a mix of primers (https://bitsexbio.com/23682).

**Gene expression profiling by RNA-seq analysis.**

Bone marrow from 2 tibiae, 2 femora, and 2 humeri was harvested in PBS/2.5% FBS and pooled before lineage negative depletion using autoMACS Pro separator (Miltenyi Biotec). Cells were incubated with a lineage antibody cocktail (B220, CD3, CD4, CD8, Gr1, CD11b, NK1.1, B220, CD19, Suppl. Table 4) and were sorted on FACSaria II sorter (BD Biosciences) to recover GMPs. RNA was extracted using MagMax-96 Total RNA Isolation Kit (Ambion) and quality-checked with RNA 6000 Pico kit (Agilent). RNA-seq libraries were prepared from the RNA extracts using the Illumina TruSeq Stranded mRNA Kit according to the manufacturer’s instructions and sequenced using the TruSeq PE Clusterkit v3-biot-HS on an Illumina Hiseq 2000 system. Sequencing reads were aligned to the mm10 genome using TopHat v2.0.19. Reads were processed with Samtools and then mapped to Ensembl transcripts using HTSeq. Differential expression was tested using the DESeq R package (R Core Team 2015, http://www.r-project.org/). A genome coverage file was generated and scaled to RPM using Bedtools. RNA-seq data produced for this study are available under accession number GSE173533. RNA-Seq samples are in triplicate.

**Chromatin immunoprecipitation (ChIP)**

ChIPs were performed on 1–20 × 10^6 sorted cells. The cells were cross-linked with 1.5 mM EGS for 20 min before sonication with a Covaris E220 to generate 200–600 bp fragments. Samples were immuno-precipitated with 2–5 µg of either anti-GFI1 (AF3540, R&D systems), anti-H3K4me1 (ab8895, Abcam), anti-H3K4me2 (ab11946; Abcam), anti-H3K4me3 (ab8850; Abcam) or anti-H3K9me3 (ab8898; Abcam). For antibodies see Suppl. Table 4. Libraries were generated according to Illumina’s instructions. Libraries were sequenced on the Illumina Hi-seq 2000 following the manufacturer’s protocols to obtain 50 bp paired-end reads. External datasets were obtained in the form of BED files of peaks and wig visualization tracks, aligned to the mm9 build, except for L5D1, which only included the bed peak file. There is one replicate per sample for ChipSeq.

**Gene expression analysis**

Bone marrow and thymic samples from 10–20 µg of total RNA from each cell type were sequenced on the Illumina Hi-seq 2000 platform at the Canadian Genomics Heartland (https://www.cgh.ca) using cell type-specific primers for gene expression analysis.

**Data availability.**

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE28 partner repository with the dataset identifier PXD028945 and 10.6019/PXD028945. The raw ChIP-seq and RNA-seq data have been uploaded to the GEO Datasets repository (https://www.ncbi.nlm.nih.gov/geo/) and are available under the following accession number: GSE173533. Previously published ChIP-seq and ATAC-Seq data, which are presented in Fig. 4 and suppl Figure 4, are available under the following accession numbers: H3K27ac (GSM1441173), H3K27me3 (DRR023959), H3K9me3 (DRR023962), H3K4me1 (GSM1441289), and ATAC-Seq (DRR023962).

Received: 1 June 2021; Accepted: 11 November 2021; Published online: 02 December 2021
60. Keller, A., Eng, J., Zhang, N., Li, X. J. & Aebersold, R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol. Syst. Biol. 1, 2005 0017 (2005).

61. Choi, H. et al. SAINT: probabilistic scoring of affinity purification-mass spectrometry data. Nat. Methods 8, 70–73 (2011).

62. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).

63. Martin, A. et al. BioGenie, a new tool for gene network building, visualization and analysis. BMC Bioinforma. 11, 91 (2010).

64. Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinforma. 4, 2 (2003).

65. Bensadoun, A. & Weinstein, D. Assay of proteins in the presence of interfering materials. Anal. Biochem. 70, 241–250 (1976).

66. Reimand, J. et al. gProfiler—a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 44, W83–W89 (2016).

67. Giurgiu, M. et al. CORUM: the comprehensive resource of mammalian protein complexes-2019. Nucleic Acids Res. 47, D559–D563 (2019).

68. Mellacheruvu, D. et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nat. Methods 10, 730–736 (2013).

69. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

70. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).

71. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

72. Anders, S. & Huber, W. Differential expression analysis for sequence count data. Genome Biol. 11, R106 (2010).

73. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).

74. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl Acad. Sci. USA 102, 15545–15550 (2005).

75. Lee, T. I., Johnstone, S. E. & Young, R. A. Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. 1, 729–748 (2006).

76. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 47, D442–D450 (2019).

77. Barrett, T. et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 41, D991–D995 (2013).

Acknowledgements

We are indebted to Mathieu Lapointe for technical assistance, Marie-Claude Lavallée and Jo-Anny Bissen for excellent animal care, Eric Massicotte and Julie Lord for FACS, and cell sorting. We thank Genome Quebec for performing HT sequencing. Tarik Möröy was supported by Canada Research Chair (Tier 1) and grants from the CIHR (MOP-84238, MOP-94846, PJDN-148372) and the Cancer Research Society. Charles Vadasz was supported by fellowships from the Fonds de recherche Québec—Santé (FRQS) and from the CIHR. Hall Bagci was supported by a doctoral training award from FRQS (#33603) and NSERC Discovery Grant (RGPIN-2016-04088 to Jean-François Coté). Jean-François Coté holds the TRANSAT chair in breast cancer research. Christian Trahan was supported by a CIHR Project grant (PTT-153313 to Marlene Oelfinger).

Author contributions

A.H.: Concept and design, collection of data, data analysis, paper writing; J.H.: Collection of data, data analysis, paper writing; C.T.: Collection of data, data analysis, paper writing; C.J:B. Analysis of genomic data, P.S: setting up BioID experiment, concept, and design, collection of data, data analysis; R.C.: protein data and western blot analyses; M.A.: setting up BioID experiment, collection of data, data analysis, K.A.: Preparation of samples, collection of data, data analysis; H.B.: Proteomic data analysis, paper writing, J.F.C.: Concept and paper writing, reagents; A.D.: Concept and analysis of genomic data; Marlene Oelfinger: Concept and paper writing, T.M.: Concept and design, data analysis, and supervision, paper writing, final approval, provision of funds.

Competing interests

The authors declare no competing interests. The work described here was conducted in compliance with Nature Springer’s ethics and biosecurity guidelines.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-02889-2.

Correspondence and requests for materials should be addressed to Tarik Möröy.

Peer review information Communications Biology thanks John Strooboulis, Jirko Kühnisch, and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editors: Eirini Trompouki and George Inglis. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

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

© The Author(s) 2021, corrected publication 2022