ARTICLES
https://doi.org/10.1038/s41588-018-0298-2

nature genetics

Corrected: Publisher Correction

Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments

Stephanie Fanucchi[1,2], Ezio T. Fok[1,2], Emiliano Dalla[1,3], Youtaro Shibayama[1,4], Kathleen Börner[5,6], Erin Y. Chang[7], Stoyan Stoychev[8], Maxim Imakaev[9], Dirk Grimm[5,6,10], Kevin C. Wang[7], Guoliang Li[11], Wing-Kin Sung[12,13] and Musa M. Mhlanga[1,14]*

Accumulation of trimethylation of histone H3 at lysine 4 (H3K4me3) on immune-related gene promoters underlies robust transcription during trained immunity. However, the molecular basis for this remains unknown. Here we show three-dimensional chromatin topology enables immune genes to engage in chromosomal contacts with a subset of long noncoding RNAs (lncRNAs) we have defined as immune gene–priming lncRNAs (IPLs). We show that the prototypical IPL, UMLILO, acts in cis to direct the WD repeat-containing protein 5 (WDR5)–mixed lineage leukemia protein 1 (MLL1) complex across the chemokine promoters, facilitating their H3K4me3 epigenetic priming. This mechanism is shared amongst several trained immune genes. Training mediated by β-glucan epigenetically reprograms immune genes by upregulating IPLs in manner dependent on nuclear factor of activated T cells. The murine chemokine topologically associating domain lacks an IPL, and the *Cxcl* genes are not trained. Strikingly, the insertion of UMLILO into the chemokine topologically associating domain in mouse macrophages resulted in training of *Cxcl* genes. This provides strong evidence that lncRNA-mediated regulation is central to the establishment of trained immunity.

Uncontrolled cytokine expression is the etiology of many inflammatory-based diseases, such as cancer[1]. When challenged, innate immune cells (monocytes) and non-immune cells (endothelium) elicit a pro-inflammatory response by rapidly expressing genes involved in processes such as cytokine signaling. Stochasticity in immune gene expression may undermine a successful immune response. Therefore, many immune gene promoters exist in an epigenetically primed, or accessible state prior to activation. This permits robust gene transcription upon activation by signal-dependent transcription factors[2]. Trimethylation of histone H3 at lysine 4 (H3K4me3) is associated with the promoters of primed genes. Importantly, the level of H3K4me3 on gene promoters reflects the strength of transcription, with high H3K4me3 levels associated with robust gene expression[3]. However, the upstream molecular events that coordinate H3K4me3 accumulation at inflammatory gene promoters remain poorly understood.

Monocytes pre-exposed to β-glucan prior to exposure to certain infectious agents exhibit an enhanced innate immune response, or 'trained immunity', and display improved clearance of the infectious agent[4–6]. This enhanced immunity is preceded by the accumulation of H3K4me3 marks on immune genes[4]. As a consequence of this, 'trained' genes are more robustly transcribed. Yet the molecular mechanisms describing how immunological signals alter H3K4me3 levels at discrete loci remain unknown[4–8].

Within the eukaryotic nucleus, chromosomal looping contributes to transcription by bringing co-regulated genes and enhancer–promoter elements together in cis and *trans*[9,10]. These chromosomal contacts are usually identified by chromosome conformation capture (3C)-based techniques (Hi-C, chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) etc.)[11]. Analyses of Hi-C data show that chromatin is divided into domains enriched in chromosomal loops, named topologically associating domains (TADs)[11].

Recently, Hi-C studies have shown that several co-regulated immune genes are segregated into TADs[12]. Due to the insulating role of the transcriptional activator/repressor CTCF, loci within TADs are more likely to participate in intra-TAD contacts[13–15]. However, a modest increase in intra-TAD chromosomal interactions exists compared with the frequency of inter-TAD interactions[15]. Thus, in addition to being located within the same TAD, other cooperative mechanisms may regulate immune gene transcription.

By acting as intermediates that link nuclear organization to gene regulation, long noncoding RNAs (lncRNAs) are emerging as key modulators of gene activity. lncRNAs range from >200 nucleotides to <10 kilobases (kb) in length[16]. Although not all are functional, many have been shown to tether protein-interacting partners, such as chromatin remodelers, near target genes to regulate their transcription[16–22].

[1]Gene Expression and Biophysics Group, Division of Chemical, Systems and Synthetic Biology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa. [2]BTRI, CSIR Biosciences, Pretoria, South Africa. [3]Department of Medicine, Università degli Studi di Udine, Udine, Italy. [4]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [5]Department of Infectious Diseases/Virology, BioQuant Center, Heidelberg University Hospital, Heidelberg, Germany. [6]Heidelberg Partner Site, German Center for Infection Research (DZIF), Heidelberg, Germany. [7]Department of Dermatology, Stanford University, Stanford, CA, USA. [8]Biomedical Technologies Group, CSIR Biosciences, Pretoria, South Africa. [9]Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Boston, MA, USA. [10]Cluster of Excellence CellNetworks, Heidelberg, Germany. [11]College of Informatics, Huazhong Agricultural University, Wuhan, China. [12]School of Computing, National University of Singapore, Singapore, Singapore. [13]Genome Institute of Singapore, Singapore, Singapore. [14]Gene Expression and Biophysics Unit, Instituto de Medicina Molecular, Faculdade de Medicina Universidade de Lisboa, Lisbon, Portugal. *e-mail: arhat@mhlangalab.org

lncRNA–protein complexes act in concert with DNA elements (for example, enhancers) to mediate the assembly of TADs[13,19]. Enhancer loci can be transcribed into noncoding RNAs, termed enhancer RNAs (eRNAs), which act with the Mediator complex to regulate enhancer–gene looping[20]. A second subtype of lncRNAs acts as a scaffold for chromatin-remodeling complexes that catalyze H3K4me3 of target gene promoters[17,21,22]. For example, NeST is a WD repeat-containing protein 5 (WDR5)-interacting lncRNA that regulates the H3K4me3 epigenetic activation of the *IFNG* locus[17]. However, it remains unknown how three-dimensional (3D) topology and lncRNA-related regulation affect the H3K4me3-primed state of trained immune gene promoters.

In this study we show how chromatin 3D structure enables a subset of lncRNAs, termed immune gene–priming lncRNAs (IP-lncRNAs or IPLs) to be brought proximal to immune genes prior to their activation. We show that the prototypical IPL, UMLILO, acts in cis to direct the WDR5–mixed lineage leukemia protein 1 (MLL1) complex across the chemokine promoters (*IL-8*, *CXCL1*, *CXCL2* and *CXCL3*) facilitating their H3K4me3 epigenetic priming. This mechanism is shared amongst dozens of trained immune genes (for example, *IL6* and *CSF1*). We demonstrate that β-glucan, a classical inducer of trained immunity, triggers a nuclear factor of activated T cells (NFAT)-mediated increase in the transcription of several IPLs. Corresponding to the absence of UMLILO in the murine chemokine TAD, trained immune responses are absent at the *Cxcl* locus. Strikingly, the insertion of UMLILO into the mouse chemokine TAD resulted in training of the *Cxcl* genes. This provides strong evidence that IPL-mediated changes to the epigenetic status of immune gene promoters permit trained immune responses.

Results

Identification of IPLs. Tumor-necrosis factor (TNF) induces the transcription of hundreds of innate immune genes, many of which are upregulated during trained immune responses[4,23]. In primary human endothelial cells (HUVECs) and other immune cells, many genes are highly expressed after TNF induction, including those involved in cytokine signaling (*TNFAIP3*, *IL1A*, *IL1B* and *IL6*) and chemotaxis (*CCL2*, *CXCL1*, *CXCL2*, *CXCL3* and *IL-8*)[24]. Importantly, the promoters of all of these robustly transcribed genes are primed with H3K4me3 prior to activation. We hypothesized that nuclear topology was correctly positioning lncRNAs near immune gene promoters to coordinate the deposition of H3K4me3. We devised a discovery pipeline to identify these lncRNAs (Fig. 1a).

Recently, Hi-C studies showed that several pro-inflammatory genes are segregated into TADs[12]. We investigated whether this organization extended to other TNF-responsive genes (Fig. 1a(i)). Using previously published Hi-C data we observed that many TNF-inducible genes were arranged into TADs that were: (1) unaccompanied by other pro-inflammatory genes (*IL6*) or (2) accompanied by other functionally related cytokines (*CXCL* chemokines) (Fig. 1b,c).

Our analysis identified annotated lncRNAs in both class I and class II TADs. Cap analysis gene expression (CAGE) sequencing presents an ideal approach to detect lncRNAs with low abundance. The FANTOM5 panel of CAGE-sequenced samples includes both primary immune samples (monocytes) and non-immune samples (endothelium)[19]. Therefore, we analyzed FANTOM5 CAGE data to determine which annotated lncRNA(s) were being expressed in each TAD (Fig. 1a(ii) and Supplementary Table 1). In some immune gene–containing TADs (*IL6* TAD), CAGE data showed that more than one lncRNA is transcribed). As these cytokines are expressed across many different innate immune cell types we selected the lncRNA variant that was also expressed across multiple tissue types.

To determine whether these lncRNAs act in cis, we analyzed RNA polymerase II (Pol II) (Ser2/Ser5) ChIA-PET data from a library constructed in HUVECs[25]. Upon TNF stimulation, we observed an increase in chromosomal contacts between TNF-responsive genes and lncRNAs for both class I and class II immune TADs (Fig. 1b and Supplementary Tables 2 and 3). Furthermore, these Pol II-associated chromosomal contacts were confined within TAD boundaries (Fig. 1b).

We combined the above with H3K4me3 and Pol II chromatin immunoprecipitation (ChIP) analysis to confirm that the promoters of the TNF-responsive genes were in a primed state (Fig. 1b,c). In summary, all of the identified lncRNAs were located in the same TAD, were expressed and were engaged in Pol II-associated contact with H3K4me3-primed TNF-responsive genes. We refer to these lncRNAs as IPLs.

In general, we found IPLs to be evolutionarily conserved in mice, multi-exonic and low-abundance lncRNAs. We observed that corresponding to their target genes, IPLs are expressed across a wide range of unstimulated cell types. Although no sequence conservation exists between the exonic portion of IPLs, the majority of human IPLs possess an Alu element in their first intron. Most of the promoters of IPLs share common transcription factor-binding motifs, including NFAT1/2, RELA (nuclear factor NF-kappa-B p65 subunit) and GATA3 (GATA binding protein 3) sites (Fig. 1d). IPLs located in the *IL6*, *CXCL* and *CSF1* TADs were the most enriched in primary monocytes infected with pathogens such as *Candida* or when treated with stimuli such as β-glucan (Supplementary Table 1). These data hinted that IPLs could be important regulators of innate immune transcription.

UMLILO is transcribed within the chemokine TAD. One candidate IPL, which we named upstream master lncRNA of the inflammatory chemokine locus (UMLILO), formed chromosomal contacts with the ELR$^+$ *CXCL* chemokine genes (*IL-8*, *CXCL1*, *CXCL2* and *CXCL3*; hereafter referred to as *CXCL* chemokines). These immune genes induce neutrophil chemotaxis and represent the first line of the mammalian innate immune defense[26]. Analysis of Hi-C data showed that the chemokines are located within the same TAD (Fig. 2a and Supplementary Fig. 1). ChIA-PET analysis confirmed that upon gene activation with TNF for 30 min there were numerous PETs/contacts that occurred between the super-enhancer region and the *CXCL* genes, which were confined exclusively to the region encompassing the chemokine TAD (Fig. 2b).

eRNAs can regulate chromosomal looping in TADs[20]. The majority of eRNAs have been shown to be unspliced, non-polyadenylated transcripts and to display bidirectional transcription and H3K4me1 and H3K27ac marks[19]. TADs containing super-enhancers may include co-regulated genes that are highly sensitive to external stimuli[27]. ChIP sequencing (ChIP-seq) analysis across the chemokine TAD highlighted a previously identified super-enhancer that is atypically large (~80 kb) (Fig. 2c and Supplementary Fig. 2)[24]. Analysis of FANTOM5 CAGE data identified several previously annotated eRNAs that were bidirectionally transcribed from regions corresponding to H3K4me1 and H3K27ac peaks (eRNA-01, eRNA-02 and eRNA-3; Fig. 2c).

We queried the FANTOM5 data for CAGE peaks (transcripts) not associated with any coding genes or eRNA chromatin marks. This analysis showed that UMLILO, a multi-exonic 575–base pair (bp) lncRNA, is transcribed from within the chemokine super-enhancer. CAGE data showed that UMLILO is unidirectionally transcribed and expressed in innate immune/non-immune cells (monocytes, endothelium; Fig. 2d). The exonic portion of UMLILO does not possess eRNA H3K4me1 and H3K27ac marks (Fig. 2c and Supplementary Fig. 3). Corresponding to UMLILO's higher levels of expression, its promoter region has H3K4me1 and H3K27ac marks in monocytes (Supplementary Fig. 3). Collectively, this suggests that while UMLILO might be brought in close proximity to the chemokine genes by chromosomal looping (Fig. 2b,c), it does not exhibit eRNA-like characteristics.

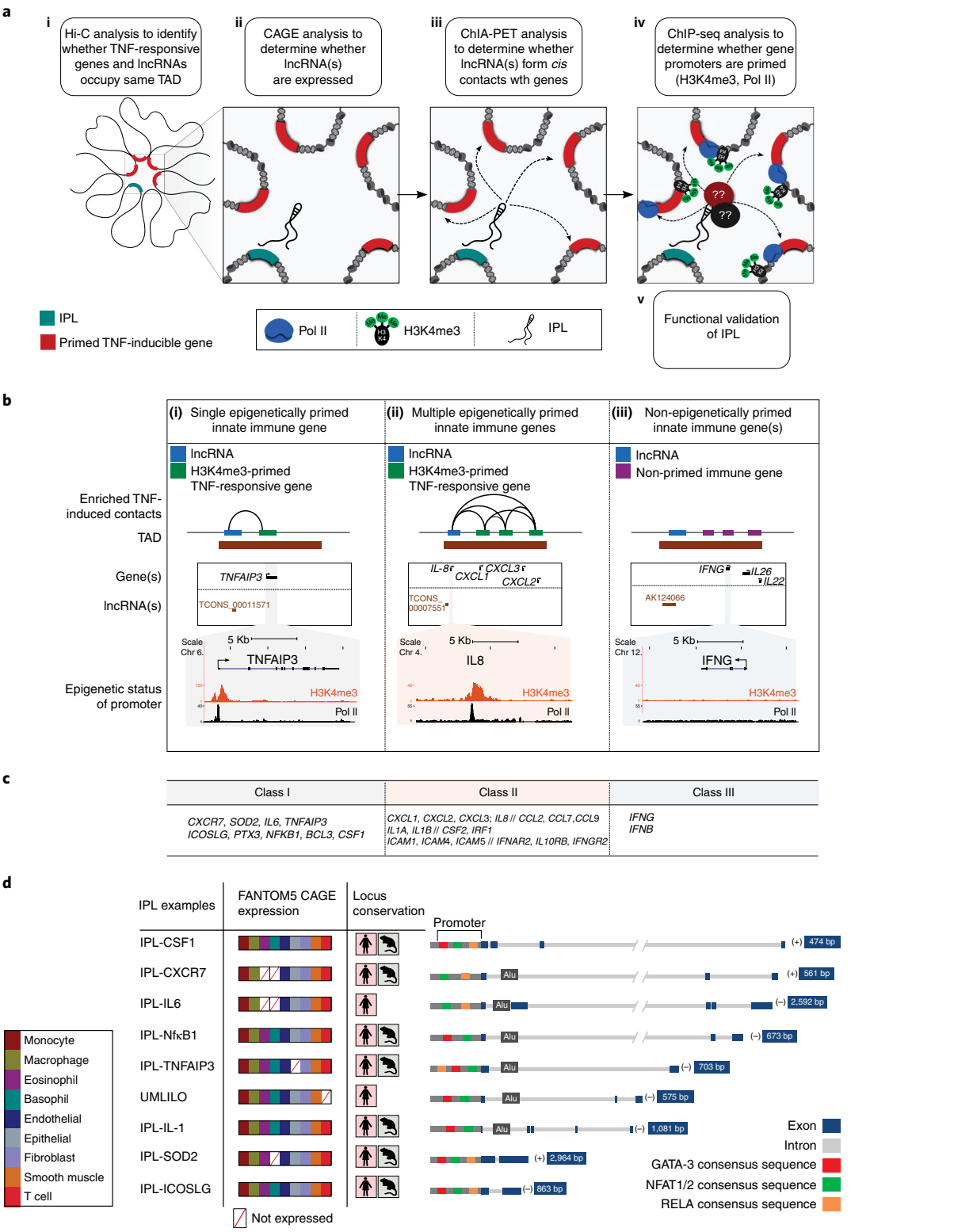

Fig. 1 | Chromatin 3D structure brings H3K4me3-primed TNF-responsive genes proximal to IPLs. a, The IPL discovery pipeline. **b**, Schematic to demonstrate class I and II immune TADs. H3K4me3-primed TNF-inducible genes are arranged into TADs that either are unaccompanied by another innate immune/non-immune genes (class I) or are accompanied by other functionally related cytokines, chemokines and signaling molecules (class II). TADs of both classes of TNF-responsive genes were co-occupied with IPLs. Upon TNF stimulation, chromosomal contacts were enriched between TNF-responsive genes and IPLs for both class I and class II innate immune TADs. No chromosomal contacts were detected between IPLs and non-TNF-responsive genes in HUVECs. **c**, Many H3K4me3-primed TNF-responsive genes fall into either class I or class II TADs. A double slash denotes the separation between genes that occupy different class II TADs. **d**, IPLs are multi-exonic, low-abundance lncRNAs expressed across a wide range of innate immune cell types. Sequence analysis showed that the majority of IPLs possess an Alu element in their first intron and that many of the promoters of IPLs share common putative transcription factor–binding motifs.

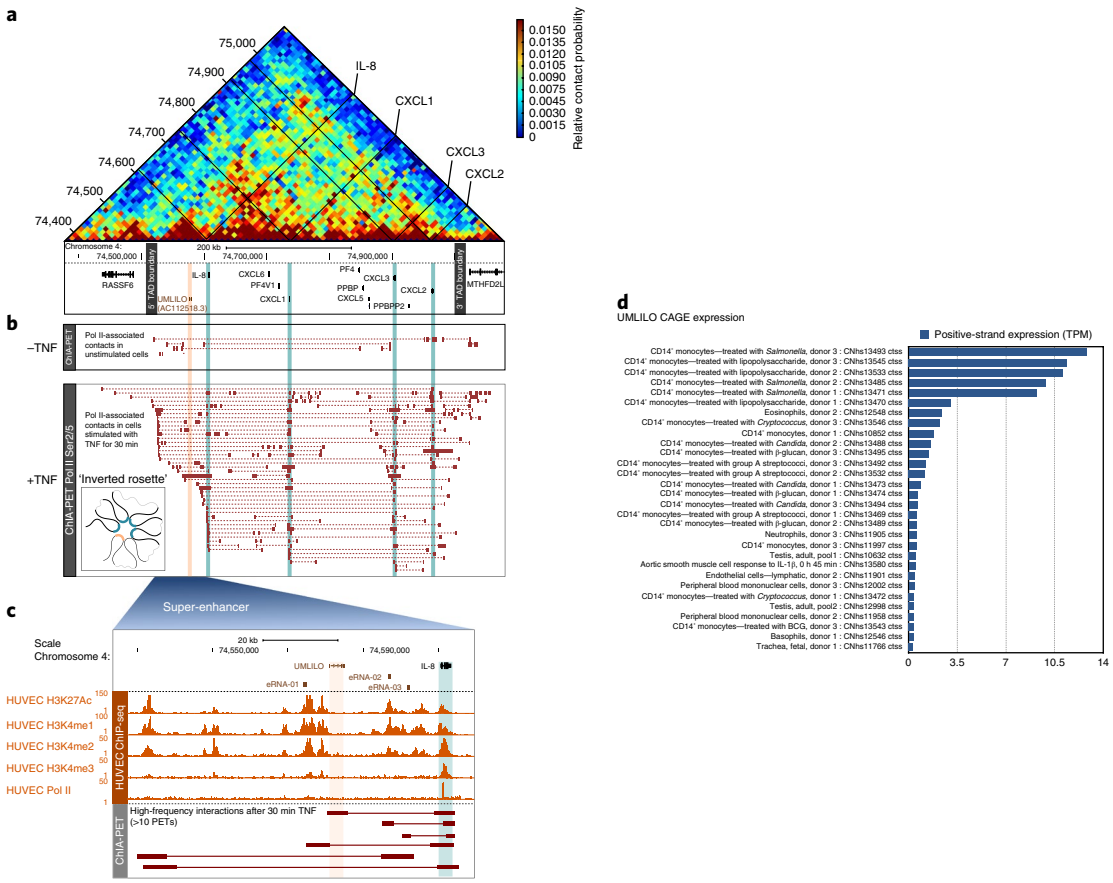

Fig. 2 | UMLILO is a new super-enhancer-resident lncRNA that is transcribed within the ELR + CXC chemokine TAD. a, Hi-C analysis across the *CXCL* chemokine locus showed that the chemokine TAD spans a region of ~500 kb. **b**, ChIA-PET data from a library constructed in HUVECs with an antibody that enriches for actively transcribing Pol II. Unstimulated HUVECs displayed relatively few PETs/chromosomal contacts. Upon chemokine gene induction with TNF for 30 min, numerous PETs/contacts were observed between the super-enhancer region and the chemokine genes, resulting in the formation of a putative 'inverted rosette' structure. Contacts detected by ChIA-PET between interacting chemokine genes were confined exclusively to the region encompassing the chemokine TAD. **c**, ChIP-seq analysis across the chemokine TAD showed that the super-enhancer is atypically large (~80 kb) and highly enriched for H3K4me1 and H3K27ac marks. UMLILO does not possess the typical eRNA histone modifications, H3K4me1 and H3K27ac. **d**, FANTOM5 CAGE data from the top 30 samples with the highest expression of UMLILO. UMLILO is expressed in multiple primary innate immune and nonimmune cells (such as $CD14^+$ monocytes and endothelium).

Depleting UMLILO alters H3K4me3 levels at *CXCL* promoters. As ChIA-PET data showed that UMLILO forms chromosomal contacts with the chemokine genes, but not with genes in neighboring TADs (Fig. 3a), we hypothesized that UMLILO might be important to chemokine transcriptional regulation. TNF strongly induces the chemokines (Fig. 3b) concomitant with an increase in UMLILO expression from baseline levels (Fig. 3b). We used either small interfering RNA (siRNA) or locked nucleic acid (LNA) GapmeRs to knock down UMLILO RNA and measured chemokine expression after TNF induction. Both siRNA and LNA knockdown of UMLILO in HUVECs (Fig. 3c), HeLa and THP-1 monocytic-like cells (data not shown) significantly abrogated chemokine expression (Fig. 3c). This demonstrated that UMLILO RNA influenced the transcription of the *CXCL* chemokines.

Overexpression of UMLILO did not alter chemokine transcription (Supplementary Fig. 4), suggesting that UMLILO may exert its effect in cis. To confirm that the effect of UMLILO was limited to the *CXCL* TAD, we constructed adeno-associated viral (AAV) vectors that contain clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9) and guides targeting UMLILO to delete the full-length UMLILO transcript. We performed RNA sequencing (RNA-seq) on a transduced THP-1 population to verify genome-wide effects of UMLILO depletion. This showed that *IL8*, *CXCL1*, *CXCL2* and *CXCL3* transcription was abrogated (Supplementary Figs. 4 and 5), but a similar effect was not seen for genes located outside of the *CXCL* TAD boundary (Supplementary Figs. 6 and 7). This strongly suggested that UMLILO regulates only the activation of *IL8*, *CXCL1*, *CXCL2* and *CXCL3* with its activity constrained to within the boundaries of the *CXCL* TAD.

The chemokine promoters are enriched for H3K4 mono-, di- and trimethylation, which are catalyzed by the MLL family of methyltransferases. MLL1 has been found to selectively regulate the activation of TNF-induced gene activation[28]. MLL1 associates with WDR5, an RNA-binding adapter protein that recognizes H3K4 methylation. Other lncRNAs, such as NeST[17] and HOTTIP[21], recruit the WDR5–MLL1 complex to regulate H3K4me3 levels on target genes. As all IPLs engage in chromosomal contact with H3K4me3-primed immune genes (Fig. 1b,c), we investigated whether UMLILO was regulating *CXCL* H3K4me3 promoter levels.

We performed siRNA knockdown of UMLILO followed by WDR5, H3K4me3, Pol II (Ser5) and MED12 ChIP analysis of the chemokine promoters (Fig. 3d). A high knockdown efficiency of UMLILO was observed (Fig. 3c). Consistent with UMLILO not being an eRNA, siRNA targeting UMLILO (siUMLILO) did not alter MED12 occupancy on the chemokine promoters (Fig. 3d).

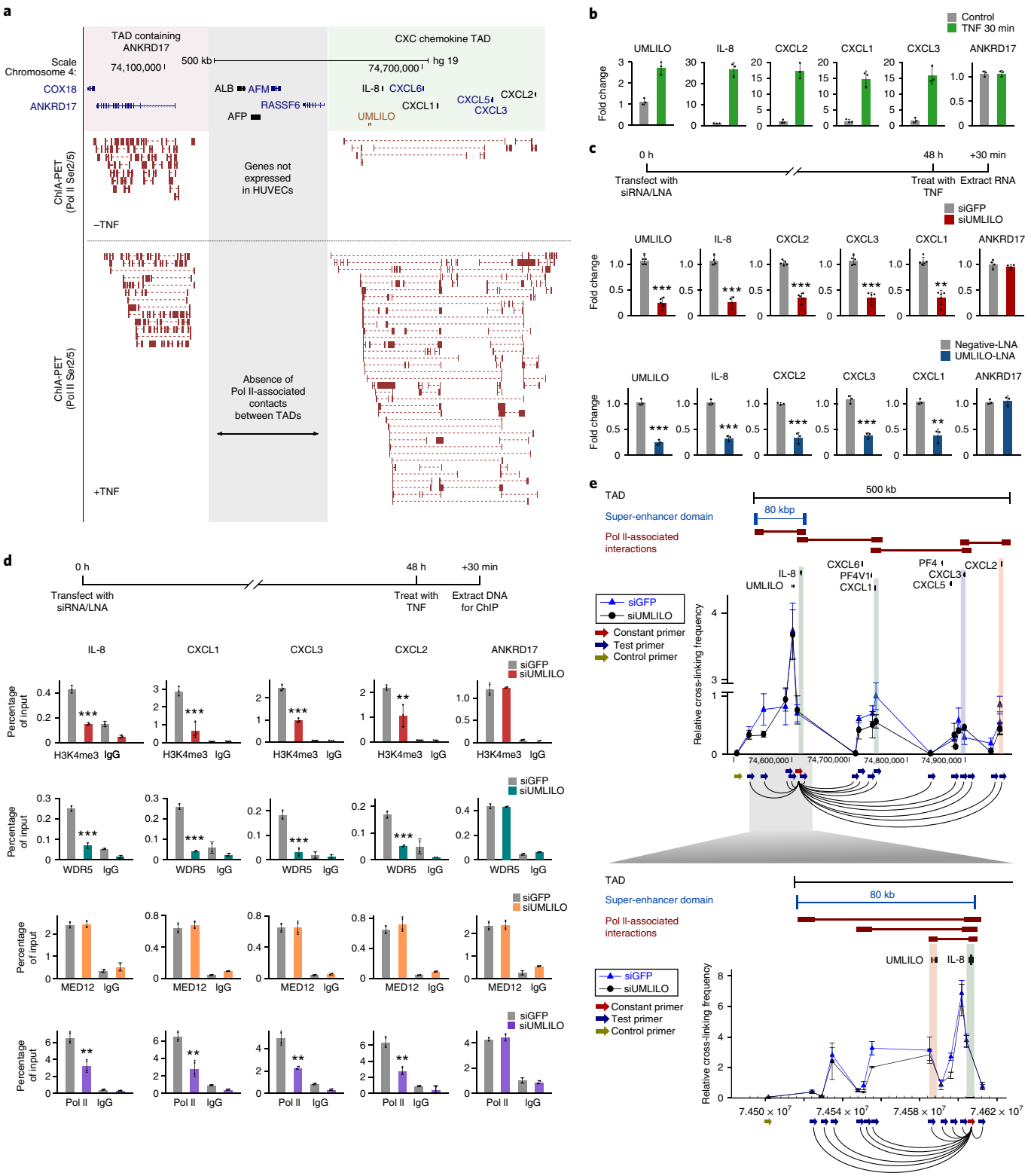

Fig. 3 | The UMLILO lncRNA regulates H3K4me3 across the *CXCL* chemokine promoters. a, Pol II (Ser2/5) ChIA-PET data show that UMLILO forms chromosomal contacts with the chemokines, but not with genes expressed in HUVECs that occupy a 5′-neighboring TAD. Genes located between these two TADs (such as *ALB* and *AFM*) are not expressed in HUVECs and do not engage in Pol II-associated chromosomal contacts with the *CXCL* chemokines. **b**, UMLILO and the chemokines are induced by TNF in HUVECs. **c**, UMLILO is necessary for chemokine expression. Knockdown of UMLILO with either siRNA or LNA GapmeRs, prior to chemokine gene induction using TNF, was sufficient to significantly abrogate chemokine expression but not an expressed gene in a TAD located 5′ of the *CXCL* TAD boundary. **d**, siRNA knockdown of UMLILO, prior to chemokine gene induction using TNF, modulates WDR5 binding as well as Pol II (Ser5) and H3K4me3 occupancy on the chemokine promoters. **e**, Depleting UMLILO using siRNA does not significantly alter chromosomal contact across the chemokine TAD. Mean ± s.d.; $n = 3$ independent experiments; $^{**}P < 0.01$, $^{***}P < 0.001$; two-tailed Student's *t*-test.

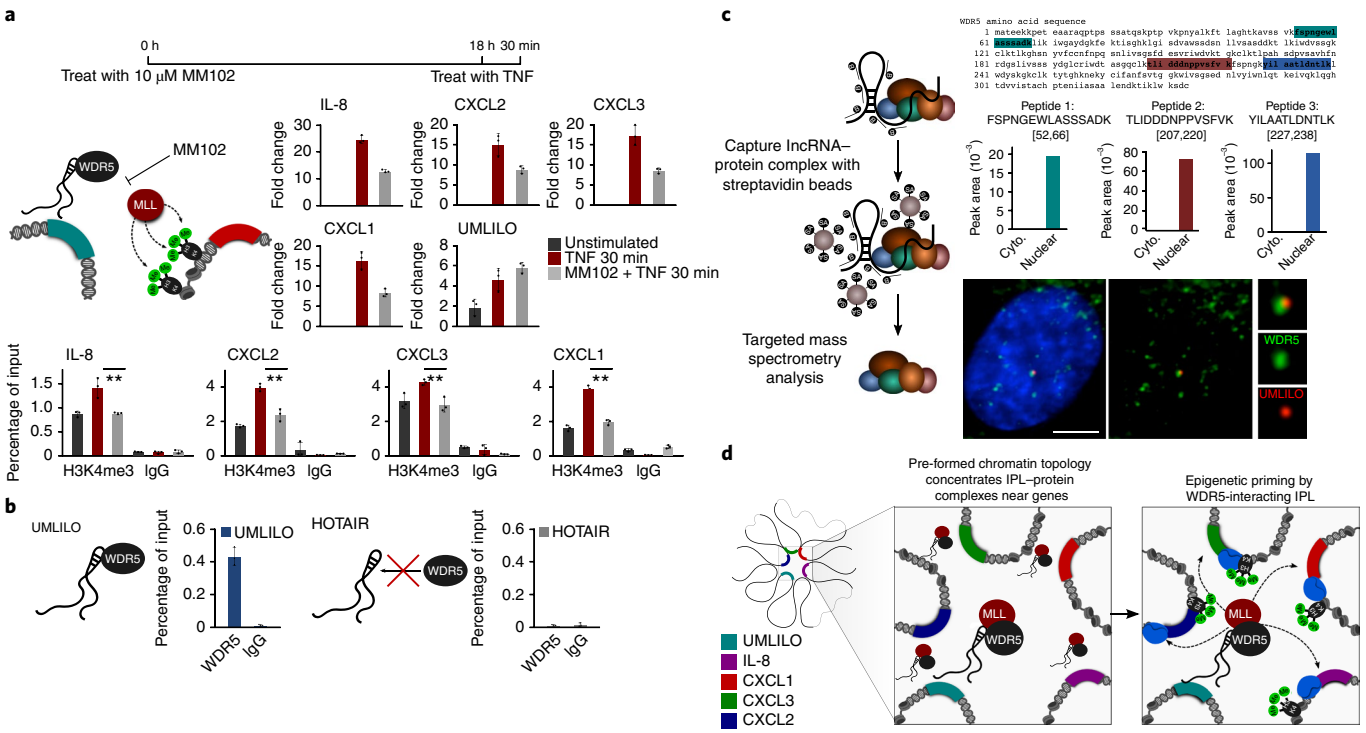

Fig. 4 | UMLILO interacts with WDR5. a, MM102, a small molecule inhibitor that prevents the WDR5–MLL1 interaction, led to a significant reduction in chemokine gene transcription and H3K4me3 epigenetic marks on the chemokine promoters. IgG, immunoglobulin G. **b**, Immunoprecipitation of WDR5 in formaldehyde-cross-linked extracts specifically retrieved UMLILO, but not HOTAIR. **c**, UMLILO was recovered from the nuclear fraction of cell lysates using biotinylated single-stranded DNA probes tiling the exonic portion of the lncRNA. Using the MRM^{HR} mass spectrophotometry approach, we identified WDR5 in the nuclear fraction of the UMLILO pulldowns. Immunofluorescence showed that WDR5 is a low-abundance protein located in the nucleus. Immuno–RNA FISH showed that distinct foci of WDR5 associate with UMLILO. Cyto., cytoplasmic fraction. **d**, lncRNAs are directed in close proximity to immune genes by 3D topology to concentrate the WDR5–MLL1 complex near the promoters of immune genes. This facilitates the deposition of the H3K4me3 active mark on immune gene promoters. As a consequence, immune genes are robustly transcribed upon the arrival of signal-dependent transcription factors. Mean ± s.d.; $n = 3$, independent experiments; $^{*}P < 0.05$, $^{**}P < 0.01$; two-tailed Student's t-test; cells were counterstained with DAPI; scale bar, 5 μm.

However, siUMLILO significantly reduced WDR5 binding and resulted in the loss of Pol II and H3K4me3 marks at chemokine promoters (Fig. 3d). Therefore, the loss of chemokine gene expression observed upon UMLILO knockdown (Fig. 3c) is likely caused by the loss of H3K4me3 deposition by a putative UMLILO–WDR5–MLL1 complex.

Corresponding to the study by Jin et al.[12], we observed that the chemokine TAD structure is formed prior to gene activation (Supplementary Fig. 8). Silencing of UMLILO by siRNA followed by 3C analysis showed that chromosomal contact across the chemokine TAD remained unaffected by loss of the RNA (Fig. 3e). These data suggest that the UMLILO transcript does not instruct chromosomal looping.

To provide support for a chromatin localization of UMLILO, we performed chromatin isolation by RNA purification (ChIRP) followed by locus-specific quantitative PCR (qPCR). ChIRP–qPCR showed that the UMLILO transcript is in close proximity to the *CXCL* promoters but not control genes (Supplementary Fig. 8). This provided evidence that UMLILO is a cis-acting, chromatin-associated transcript.

UMLILO interacts with WDR5. MLL1 activity is mediated via interactions with the WAR complex (WDR5, Ash2L and RbBP5) and other regulators. Treating HUVECs with a small molecule inhibitor, MM102, that blocks WDR5–MLL1 interactions[29] resulted in significant reductions in chemokine gene transcription and H3K4me3 marks (Fig. 4a). Other H3K4me3-primed TNF-responsive genes were also abrogated (data not shown). Although this does not directly implicate the IPLs, these data highlight the necessity of MLL1-dependent H3K4me3 deposition for immune gene transcription.

To directly test whether UMLILO interacts with WDR5, the epitope-tagged protein was co-expressed with UMLILO or HOTAIR (a control lncRNA). Immunoprecipitation of WDR5 in formaldehyde-cross-linked extracts retrieved UMLILO but not HOTAIR (Fig. 4b). We observed that a portion of UMLILO (part of exon 1 and exon 3) was conserved in the pig. We used qPCR and H3K4me3 ChIP to verify that porcine UMLILO acts in a similar manner to human UMLILO. Using RNA immunoprecipitation analysis we observed that exon 3 of UMLILO had an enriched interaction with WDR5 relative to exon 1 and 2, suggesting that exon 3 of UMLILO is important for its interaction with WDR5 (Supplementary Fig. 9).

Due to the low expression of UMLILO, we used a targeted multiple reaction monitoring with high-resolution mass spectrometry (MRM^{HR}) method to quantify and identify members of the WAR complex (Fig. 4c). We subsequently confirmed the presence of WDR5 in nuclear extracts of UMLILO pull-downs (Fig. 4c and Supplementary Fig. 10). Immunofluorescence labeling showed that WDR5 present at low abundance in HUVECs was located predominantly in the nucleus in discrete punctate spots. Using immuno-RNA FISH we observed distinct foci of WDR5 co-localized with UMLILO RNA (Fig. 4c). This provided additional evidence for the nuclear location of the UMLILO–WDR5 interaction (Fig. 4d).

Deletion of UMLILO abrogates chemokine transcription. As introns briefly persist after transcription, probes targeting introns

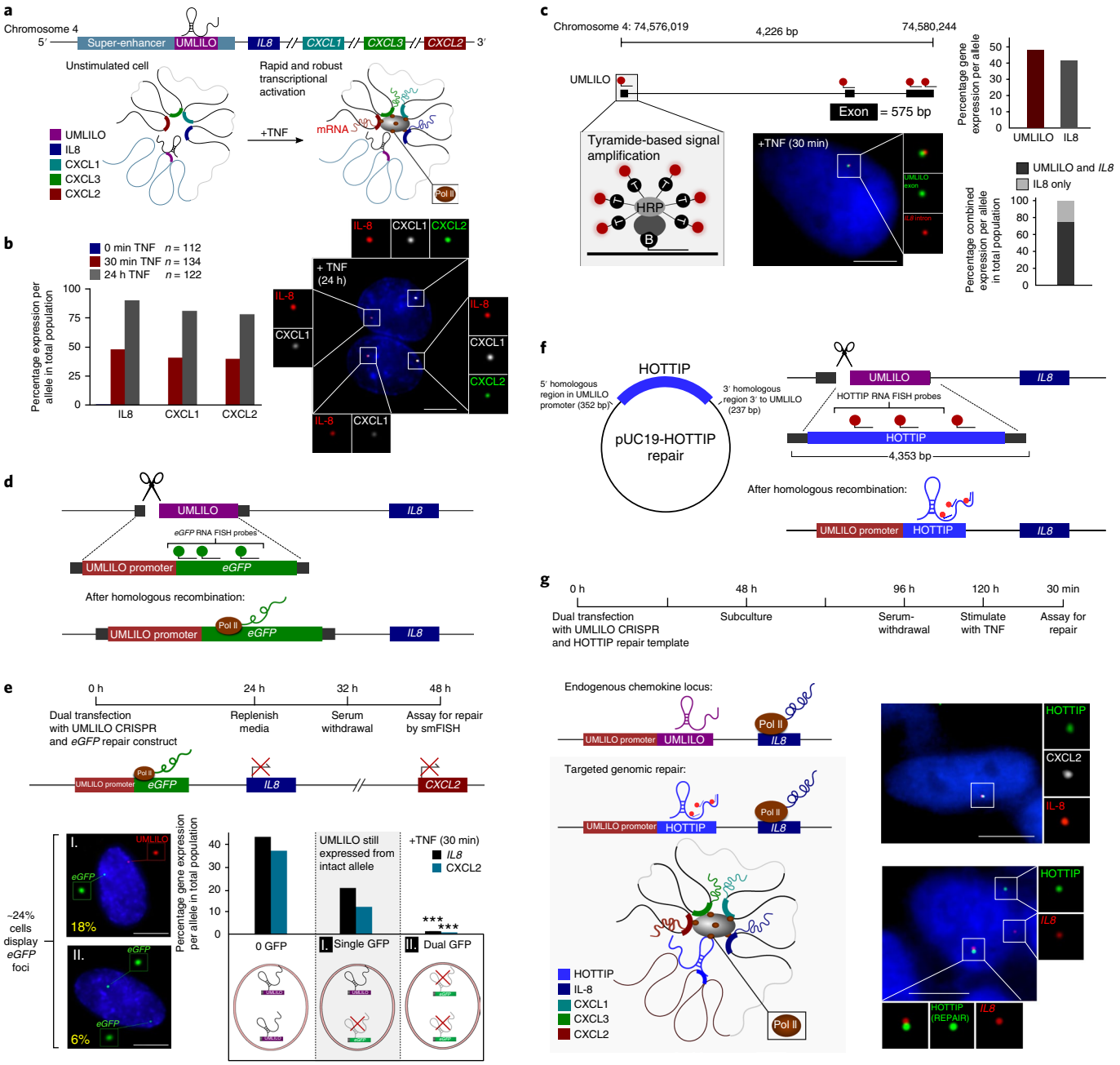

Fig. 5 | UMLILO acts in cis to regulate chemokine transcription. a, Schematic demonstrating how the chemokine TAD structure is formed prior to gene activation by TNF. **b**, Intronic RNA FISH foci of co-expressed CXCL chemokine genes always co-localize. **c**, Simultaneous exonic TSA RNA FISH on UMLILO and IL8 showed that co-localization between exonic UMLILO and the intronic portion of IL8 was frequently observed. **d**, The CRISPR–Cas9 system was used to delete the genomic sequence encoding UMLILO and replace it with an eGFP reporter sequence in primary HUVEC cells. Successful repair was detected using RNA FISH probes targeting eGFP RNA, driven by the endogenous UMLILO promoter. **e** Successful repair was observed in ~26% of cells, with ~19% of cells displaying a single focus and ~7% of cells displaying dual eGFP foci. Cells displaying a single eGFP focus were still able to express UMLILO from the opposite allele. There was a significant reduction in chemokine expression in cells displaying dual eGFP foci. $n = 362$; $^{***}P < 0.001$; two-tailed Fisher's exact test. **f**, The HOTTIP repair construct consists of the sequence of HOTTIP flanked by homologous arms that flank UMLILO. The CRISPR–Cas9 system was used to delete the genomic sequence encoding UMLILO and replace it with the sequence of HOTTIP in HeLa cells. **g**, Replacing endogenous UMLILO with HOTTIP restores chemokine transcription. Overlapping HOTTIP and chemokine foci were observed in ~5% of HeLa cells dually transfected with the CRISPR and HOTTIP repair construct for 120 h. $n = 646$; cells were counterstained with DAPI; scale bar, 5 μm.

label the site of transcription[30]. Using intronic single-molecule RNA FISH (smFISH), we observed that TNF induced robust chemokine transcription (*CXCL3* lacks an intron and was excluded; Fig. 5a,b). RNA FISH foci of co-expressed chemokine genes always co-localized (Fig. 5b). To reliably detect the short UMLILO transcript, we developed a tyramide-based signal amplification (TSA) approach to visualize exonic UMLILO (Fig. 5c)[31]. We observed a punctate focus of UMLILO, which frequently co-localized with IL8; (Fig. 5c), providing additional evidence that UMLILO is a cis-acting transcript.

However, not all lncRNAs produce functional RNA transcripts[32]. To establish a clear causal relationship between UMLILO RNA and CXCL transcription we replaced the UMLILO transcript with an

enhanced green fluorescent protein (eGFP) DNA sequence using homology-directed repair (HDR) directed by CRISPR–Cas9 (Fig. 5d). We designed the repair so that the endogenous promoter elements regulating UMLILO would be preserved and eGFP transcription would be regulated by the same endogenous regulatory elements regulating UMLILO (Supplementary Fig. 11). Following successful repair in HUVECs, we induced expression of the chemokines with TNF for 30 min. We observed a distinct focus of transcribed *eGFP* RNA by smFISH in the nucleus of ~24% of cells, with ~18% of cells displaying a single focus and ~6% of cells displaying dual *eGFP* foci, indicative of one at each allele (Fig. 5e). Co-localization between a focus of eGFP RNA and IL8 and/or CXCL2 RNA was never observed (Supplementary Fig. 11). We compared cells displaying a single eGFP focus but still expressing UMLILO from the intact allele with cells displaying dual eGFP foci. In cells displaying no eGFP foci, expression of IL8 and CXCL2 was comparable to the mock-transfected cells (Fig. 5e). However, in cells displaying a single eGFP focus, the expression of IL8 and CXCL2 occurred only at the allele where UMLILO was not deleted (Fig. 5e). Importantly, this was observed despite the fact that these cells are still able to express UMLILO from the opposite allele. Interestingly, in cells displaying two eGFP foci, we were unable to detect expression of any chemokine genes (Fig. 5e). This reinforced prior data indicating UMLILO to be a cis-acting lncRNA, playing a key role in the expression of chemokine genes.

HOTTIP can substitute for UMLILO activity. While the knockdown data and CRISPR–Cas9/eGFP exchange experiments strongly implicate the UMLILO transcript as a crucial regulator of CXCL transcription (Figs. 3c and 5e), the region encoding UMLILO may also include enhancer motifs that regulate chemokine expression. As UMLILO interacts with WDR5 (Fig. 4b,c), we investigated whether replacing UMLILO with a well-characterized WDR5-interacting lncRNA would restore CXCL transcription.

HOTTIP is a lncRNA that binds to WDR5 to facilitate the H3K4me3-mediated activation of the *HoxA* genes[21,22]. We designed a repair template with the HOTTIP exonic sequence and homologous arms that flanked UMLILO. We used this repair template and a CRISPR–Cas9 vector to exchange the UMLILO DNA sequence for the HOTTIP sequence, preserving regulation via the endogenous UMLILO promoter (Fig. 5f). TNF was then used to induce chemokine expression. Thereafter, we detected transcription of HOTTIP by RNA FISH and related the position of HOTTIP to the chemokine genes (also detected by RNA FISH).

In unstimulated HeLa cells, single allelic expression of HOTTIP was observed at ~25% of alleles (Supplementary Fig. 12). Upon successful HDR-mediated insertion of HOTTIP, strikingly, we observed HOTTIP overlapping with IL8 and/or CXCL2 foci in ~5% of cells (Fig. 5g), indicating successful functional substitution of HOTTIP for UMLILO. Remarkably, this demonstrates that substituting a WDR5-interacting lncRNA into the UMLILO locus restores chemokine transcription.

β-Glucan epigenetically reprograms immune genes in an IPL-dependent manner. Next we investigated whether our discovery pipeline could reliably identify other lncRNAs with similar properties. To validate these IPLs in independent experiments we used LNAs targeting IPL-CSF1, IPL-IL6 and IPL-ICOSLG (Fig. 6 and Supplementary Fig. 13) followed by WDR5 and H3K4me3 ChIP. Knocking down each of the IPLs led to the modulation of expression of both their target genes and WDR5 and H3K4me3 levels (Fig. 6), validating the IPL discovery pipeline as a reliable approach to identify other IPLs.

The increased responsiveness of trained monocytes is driven by epigenetic reprogramming (Fig. 7a). H3K4me3 ChIP-seq data generated by the Blueprint Consortium showed that β-glucan-mediated training of human monocytes increased transcription and H3K4me3 levels on IPL-regulated immune gene promoters (Fig. 7b)[4]. siUMLILO prior to β-glucan treatment significantly decreased chemokine transcription, indicating that UMLILO regulates training of the *CXCL* genes (Fig. 7c,d).

Training of monocytes is mediated by the β-glucan receptor, dectin-1 (ref. [4]). β-Glucan/dectin-1 signaling triggers activation of calcium-dependent NFAT[33]. As all IPL promoters contain NFAT motifs (recognized by NFAT1 and NFAT2; Fig. 1d), we hypothesized that IPL upregulation in trained monocytes may be triggered via NFAT (Fig. 7e). Calcium influxes activate calcineurin, which dephosphorylates NFAT, permitting its translocation into the nucleus, where it activates transcription[34]. Tacrolimus (FK506) inhibits calcineurin activation and, as a consequence, NFAT nuclear translocation (Fig. 7e).

Exposing monocytes to β-glucan increased expression of UMLILO and other IPLs (Fig. 7f and Supplementary Fig. 14), while pre-treatment of monocytes with tacrolimus prevented the β-glucan-induced upregulation of IPLs from baseline levels, as well as target gene transcription (Fig. 7f and Supplementary Fig. 15). These data suggest baseline expression of UMLILO is not regulated by NFAT, but by another transcription factor that binds to its promoter. Notwithstanding, β-glucan induced an increase in NFAT levels on the UMLILO promoter (Fig. 7g), while pre-treatment with tacrolimus abrogated this increase (Fig. 7g). Thus, β-glucan-mediated NFAT signaling increases IPL transcription, which increases WDR5–MLL1-mediated H3K4me3 accumulation on immune gene promoters (Fig. 7h).

Adding UMLILO to the mouse chemokine TAD restored training of the *Cxcl* chemokines. Mice are reported to be more resistant to inflammatory stimuli compared with humans[34,35]. β-Glucan-induced training has been shown to increase H3K4me3 on several mouse immune gene promoters (Supplementary Fig. 16)[4]. Significantly, the murine *Cxcl* TAD lacks UMLILO (Fig. 8a and Supplementary Fig. 17), and no increase in H3K4me3 at the *Cxcl1*, *Cxcl2* or *Cxcl3* promoters is observed after β-glucan-induced training (Fig. 8b). We hypothesized that this may be due to the lack of IPL-mediated regulation in the mouse *Cxcl* TAD.

To test whether UMLILO was able to restore robust chemokine transcription, we designed a knock-in experiment in the mouse RAW 264.7 macrophage-like cell line. From 3C analysis we selected a region within the murine *Cxcl* super-enhancer that engages in preformed chromosomal contact with *Cxcl2* (Fig. 8c). We designed a repair template that included UMLILO with its human endogenous promoter (Fig. 8c) to insert into the aforementioned region via HDR mediated by CRISPR–Cas9. We performed qPCR and ChIP in UMLILO knock-in cells and observed a significant increase in chemokine transcription (Fig. 8d). LNA knockdown of UMLILO in knock-in cells confirmed that this was due to UMLILO RNA and not to UMLILO genomic DNA. Subsequent β-glucan-induced training of the UMLILO knock-in cells resulted in a significant accumulation of H3K4me3 at *Cxcl* promoters (Fig. 8e). These data provided strong evidence that UMLILO is central to the deposition of H3K4me3 on chemokine promoters during trained immune responses (Fig. 8f) and more broadly that IPLs are the bona fide mechanism by which essential trained immune genes (for example *IL6*, *IL1B*) acquire enrichment of the H3K4me3 promoter mark.

Discussion

In this study we identify a general mechanism of immune gene regulation whereby 3D chromatin looping topology correctly positions UMLILO, and several other IPLs, in the appropriate genomic location proximal to innate immune genes. This permits IPLs to direct the WDR5–MLL1 complex to target gene promoters, facilitating their H3K4me3 epigenetic priming (Supplementary Fig. 18).

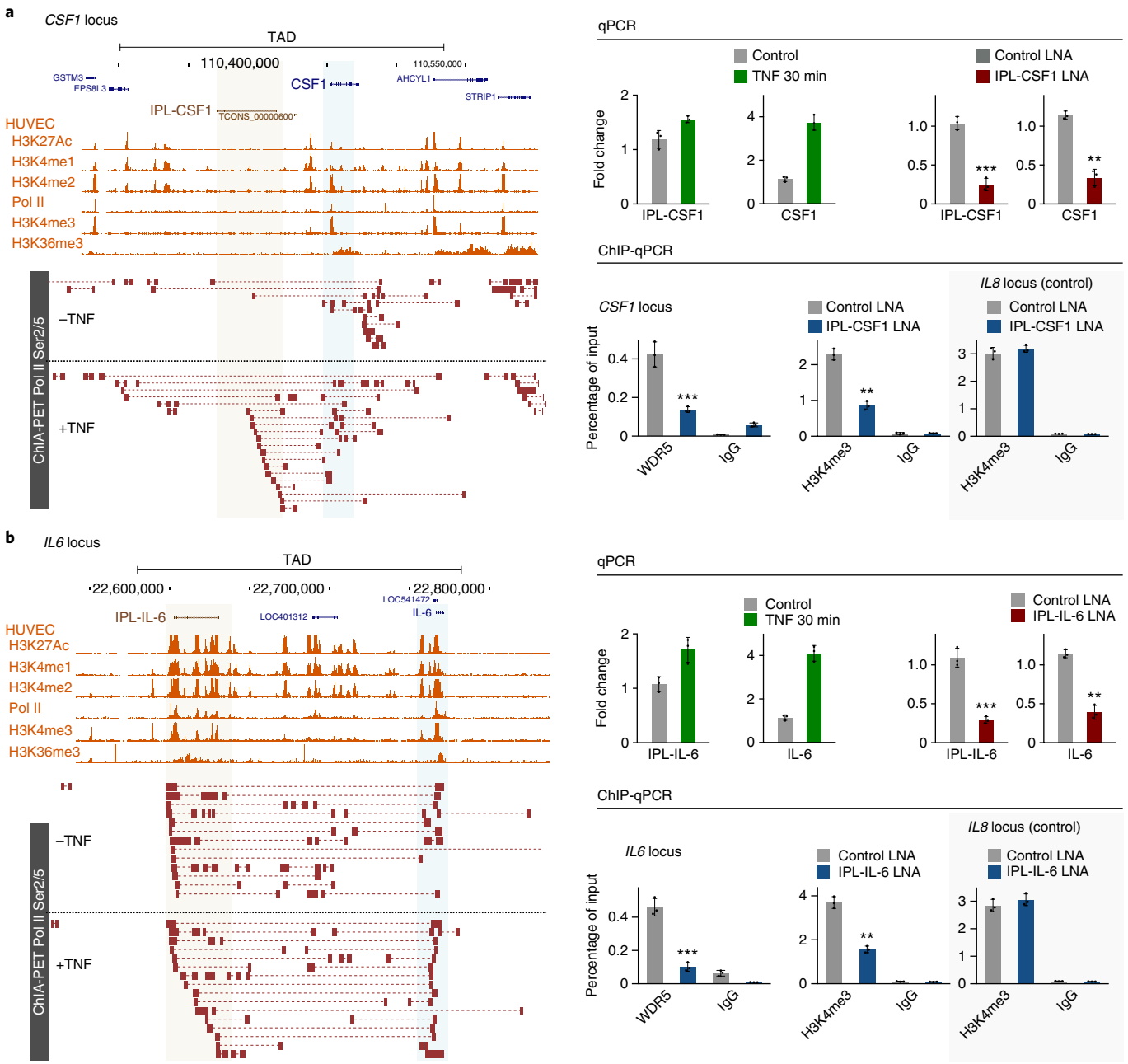

Fig. 6 | WDR5–lncRNA regulation is a general mechanism of H3K4me3-primed TNF-responsive genes. a, The *CSF1* TAD spans a region of ~250 kb. Pol II (Ser2/5)-associated ChIA-PET contacts are observed between *CSF1* and IPL-CSF1. IPL-CSF1 is a 476-bp annotated transcript that is composed of four exons and is transcribed from the positive strand. LNAs targeting IPL-CSF1 abrogated IPL-CSF1 and CSF1 transcription, as well as WDR5 and H3K4me3 levels on the *CSF1* promoter. **b**, The *IL6* TAD spans a region of ~190 kb. Pol II (Ser2/5)-associated ChIA-PET contacts are observed between *IL6* and IPL-IL-6. IPL-IL-6 is a 621-bp annotated transcript that is composed of four exons and is transcribed from the positive strand. LNAs targeting IPL-IL-6 abrogated IPL-IL-6 and IL-6 transcription, as well as WDR5 and H3K4me3 levels on the *IL6* promoter. Mean ± s.d.; $n = 3$ independent experiments; $^{**}P < 0.01$, $^{***}P < 0.001$; two-tailed Student's *t*-test.

This study addresses the long-standing question of what underlies the mechanism of trained immunity. We show that β-glucan epigenetically reprograms immune genes by upregulating IPLs in an NFAT-dependent manner. Pre-treating monocytes with tacrolimus, a first-line drug used for inflammatory conditions[36], prior to β-glucan treatment resulted in a significant reduction of IPL levels, cytokine expression and H3K4me3 levels (Fig. 7f,g). This suggests that β-glucan-induced NFAT signaling plays an important role in immune gene training and that tacrolimus blocks the establishment of trained immunity.

The lack of UMLILO in the murine *Cxcl* locus raised the question of whether inserting UMLILO into the appropriate genomic context would rescue the training of *Cxcl* genes (Fig. 8a). MLL1 contains a CXXC domain, which recognizes and binds to unmethylated CpG sites in promoters and deposits H3K4me3 on nearby histone tails. Over evolutionary time, cytosines may be converted into thymines due to spontaneous deamination[37]. In contrast to human *CXCL* promoters, mouse *Cxcl* promoters have maintained only a few CpG sites.

Remarkably, the addition of UMLILO to the chemokine TAD in mouse macrophages significantly increased chemokine expression

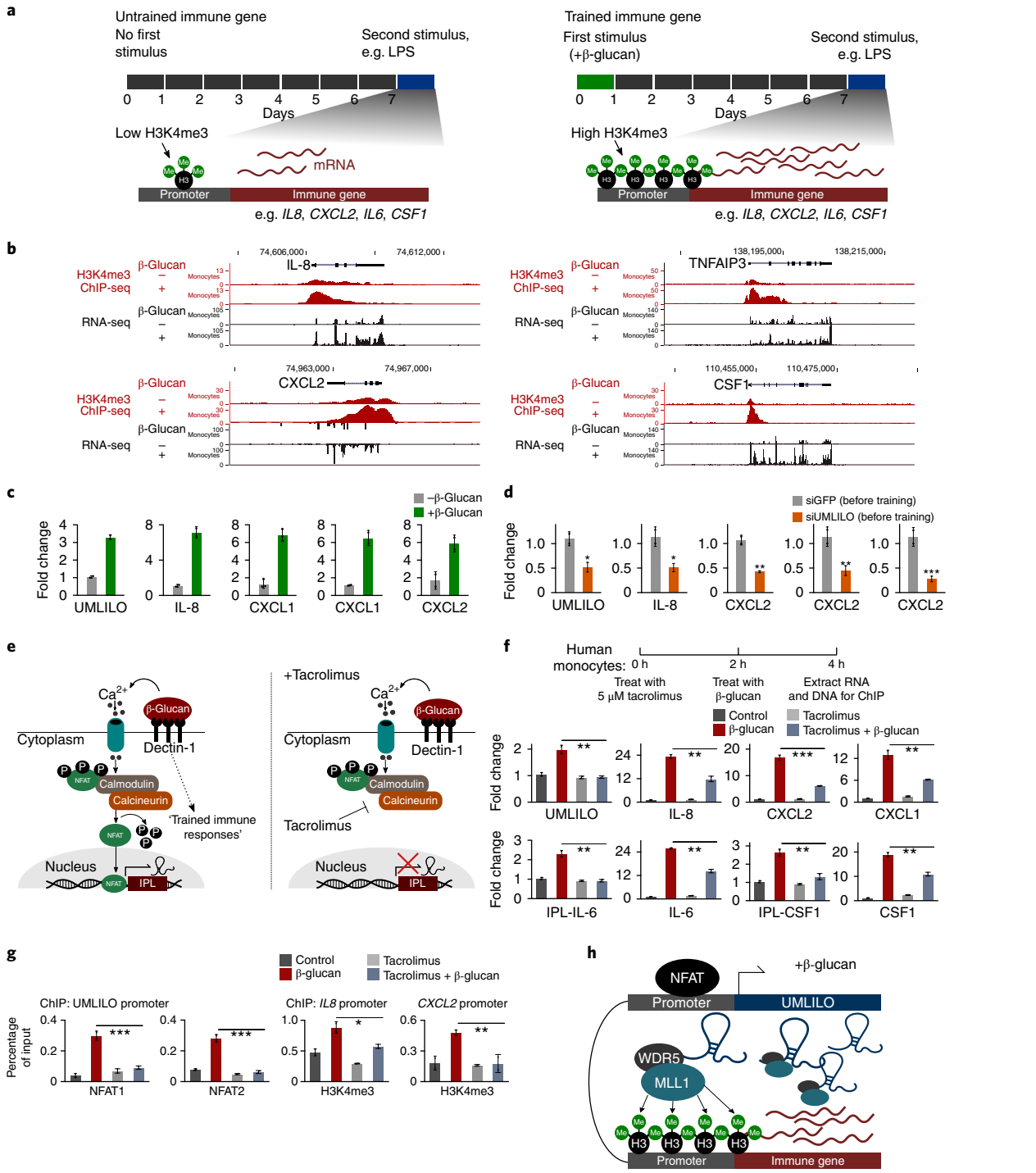

Fig. 7 | β-Glucan epigenetically reprograms immune genes by upregulating IPLs in a NFAT-dependent manner. a, Training of monocytes leads to an increase in immune gene transcription and H3K4me3 levels on dozens of IPL-regulated innate immune responsive gene promoters. **b,** Several immune genes are trained by β-glucan including *IL-8*, *CXCL2*, *CSF1* and *TNFAIP3*. **c**, β-Glucan-induced training increases UMLILO and CXCL transcription. **d**, siRNAs targeting UMLILO prior to β-glucan-induced training lead to a significant reduction in expression of both UMLILO and CXCL. **e**, β-Glucan/dectin-1 signaling has been shown to trigger activation of NFAT, a calcium-dependent transcription factor[33]. Calcium influx leads to the activation of calcineurin; this dephosphorylates NFAT, allowing its translocation into the nucleus, where it may bind to DNA and activate gene transcription. **f**, Pre-treatment of monocytes with tacrolimus prevented the β-glucan-induced upregulation of IPLs (UMLILO, IPL-IL-6 and IPL-CSF1) and immune gene transcription (CXCL genes, IL-6 and CSF1). **g**, In unstimulated monocytes, NFAT occupied the UMLILO promoter. Pre-treatment of monocytes with tacrolimus prevented this β-glucan-induced increase of NFAT on the UMLILO promoter. **h**, Upon exposure to β-glucan, NFAT signaling leads to an upregulation of IPL transcription. This in turn increases the concentration of the WDR5–MLL1 complex proximal to the innate immune gene promoters, facilitating the specific H3K4me3 priming of the innate immune response genes. Mean ± s.d.; $n = 3$ independent experiments; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$; two-tailed Student's *t*-test.

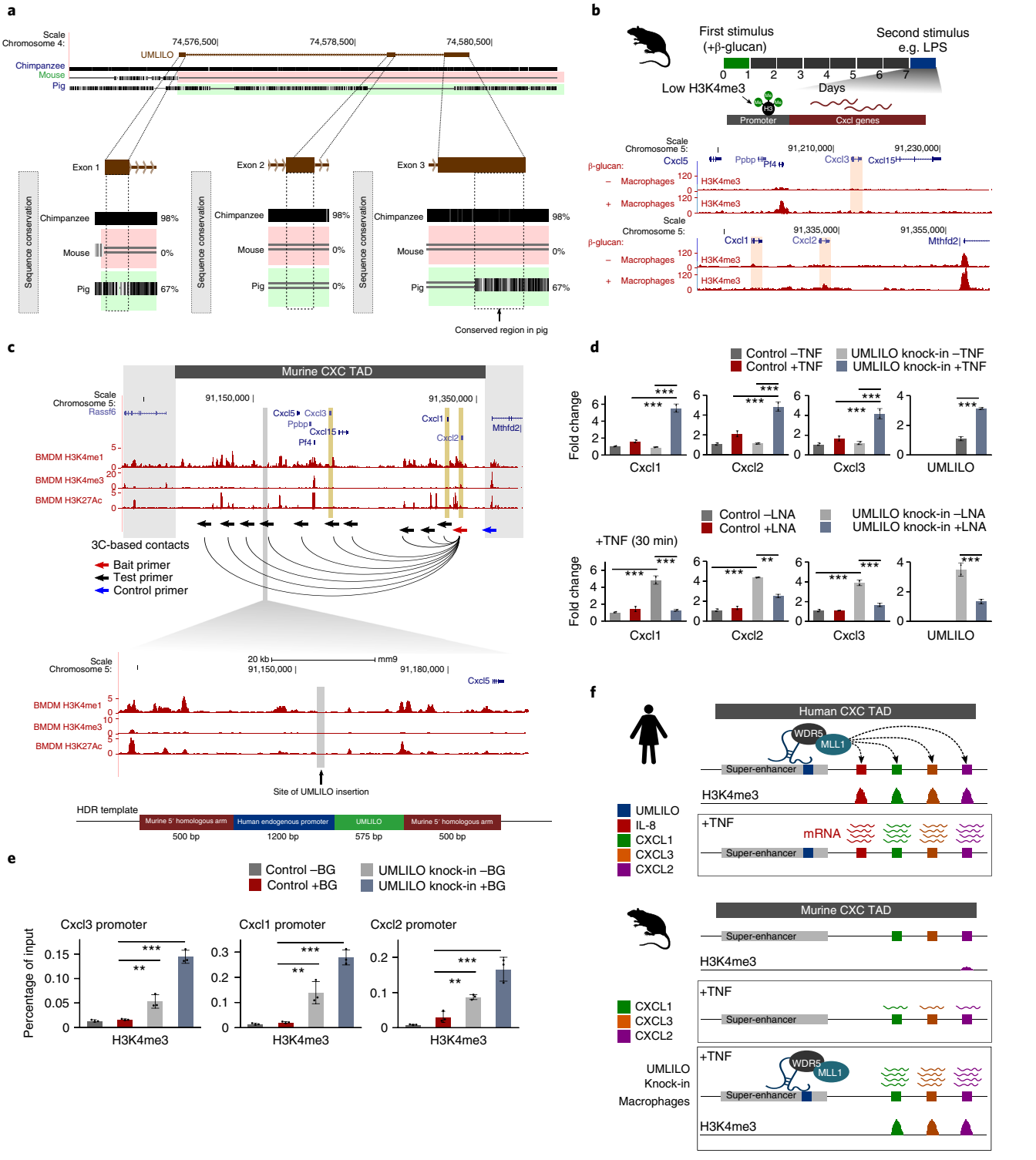

Fig. 8 | Inserting UMLILO within the mouse chemokine TAD restored training of the *Cxcl* chemokines. a, No homolog of UMLILO exists in mice. **b**, No stable accumulation of H3K4me3 was observed on the promoters of *Cxcl1*, *Cxcl2* or *Cxcl3* before or after β-glucan-induced training in mouse macrophages. **c**, Using 3C analysis in murine cells, a region was selected within the mouse super-enhancer that engages in chromosomal contact with *Cxcl2*. For HDR, a repair template was designed that included UMLILO with its own human promoter. UMLILO was successfully edited into the mouse *Cxcl* chemokine TAD in the RAW macrophage-like cell line to generate a knock-in clone. BMDM, bone marrow-derived macrophages. **d**, qPCR showed a significant increase in chemokine transcription in the UMLILO knock-in RAW cells. LNA knockdown of UMLILO in these cells confirmed that this increase in transcription was a result of the UMLILO transcript. **e**, A significant increase in baseline levels of H3K4me3 was observed in the UMLILO knock-in RAW cells compared with control cells. Further, β-glucan-induced training resulted in a significant increase in H3K4me3 levels on the *Cxcl* promoters. **f**, Schematic to demonstrate how the addition of UMLILO to the murine *Cxcl* TAD increases both transcription and training of the *Cxcl* genes. Collectively this shows how a cis-acting WDR5-interacting lncRNA insulates a key transcriptional pathway from gene-intrinsic noise, to achieve robust chemokine expression. Mean ± s.d.; $n = 3$ independent experiments; $^{**}P < 0.01$, $^{***}P < 0.001$; two-tailed Student's *t*-test.

and resulted in training of the chemokine genes (Fig. 8d,e). Though highly elevated, these levels did not fully recapitulate those observed at human *CXCL* genes. We speculate that this could be attributed to the reduced number of CpGs and the absence of other regulatory elements that could be missing through lack of evolutionary maintenance in the mouse TAD.

The mouse data suggest that IPLs are not an absolute requirement for gene activation. Rather, we propose that IPLs increase H3K4me3 levels on target genes and in this way aid robust immune gene transcription. Eukaryotic organisms have evolved several mechanisms to reduce gene-intrinsic noise[38]. We speculate that genes displaying non-stochastic gene expression may exhibit a distinct nuclear architecture and may also be assisted in transcriptional regulation by 'IPL-like' lncRNAs, such as UMLILO. This is exemplified during trained immune responses, when β-glucan/dectin-1 signaling increases IPL expression and, as a result, H3K4me3 levels, leading to robust transcription of trained immune genes.

Trained monocytes display a shift in metabolism towards an increase in glycolysis, which is also associated with the accumulation of certain metabolites, including succinate[39]. As many of these metabolites can alter the activity of histone-modifying enzymes, it is most likely the altered metabolic state observed in trained monocytes that is responsible for maintaining the IPL-mediated deposition of H3K4me3 that occurs after the first β-glucan stimulus.

Monocytes exposed to lipopolysaccharide (LPS) exhibit a tolerized phenotype, which is characterized by a reduction of epigenetic marks at immune gene promoters[6]. As a consequence, tolerized monocytes express lower levels of cytokines upon exposure to a secondary, unrelated infection. Recently, it was shown that β-glucan can partly reverse this phenotype and reactivate unresponsive genes, by reprogramming distal histone modifications[8]. Here we have shown that β-glucan/NFAT signaling upregulates IPLs and, as a result, H3K4me3 on target immune gene promoters, providing a mechanistic explanation for these prior observations[8]. This suggests that assaying IPL transcription levels may be a useful biomarker for assessing effective innate immune activation by various methods, including vaccination[40].

Historically, upstream steps that regulate cytokine signal transduction, such as antibodies targeting the CXCL receptors[41], have been the preferred therapeutic approach. Unfortunately, broad off-target effects make these approaches at times unsuccessful. There is a great need to develop inhibitors that are able to more discretely influence immune gene transcription. As aberrant expression of innate immune genes underlies many diseases, adjustment of immune gene levels by directly altering the activity of IPLs may represent a valuable therapeutic strategy to achieve tailored immunomodulation.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0298-2.

Received: 2 November 2017; Accepted: 30 October 2018;
Published online: 10 December 2018

References

1. Rogatsky, I. & Adelman, K. Preparing the first responders: building the inflammatory transcriptome from the ground up. *Mol. Cell* **54**, 245–254 (2014).
2. Bhatt, D. M. et al. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279–290 (2012).
3. Lauberth, S. M. et al. H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152**, 1021–1036 (2013).
4. Quintin, J. et al. *Candida albicans* infection affords protection against reinfection via functional reprogramming of monocytes. *Cell Host Microbe* **12**, 223–232 (2012).
5. Saeed, S. et al. Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science* **345**, 1251086 (2014).
6. Netea, M. G. et al. Trained immunity: a program of innate immune memory in health and disease. *Science* **352**, 6284 (2016).
7. Arts, R. J. et al. Glutaminolysis and fumarate accumulation integrate immunometabolic and epigenetic programs in trained immunity. *Cell Metab.* **24**, 807–819 (2016).
8. Novakovic, B. et al. β-Glucan reverses the epigenetic state of LPS-induced immunological tolerance. *Cell* **167**, 1354–1368 (2016).
9. Li, G. et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 (2012).
10. Fanucchi, S. et al. Chromosomal contact permits transcription between coregulated genes. *Cell* **155**, 606–620 (2013).
11. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
12. Jin, F. et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).
13. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
14. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
15. Dekker, J. & Mirny, L. The 3D genome as moderator of chromosomal communication. *Cell* **164**, 1110–1121 (2016).
16. Guttman, M. et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295–300 (2011).
17. Gomez, J. A. et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-γ locus. *Cell* **152**, 743–754 (2013).
18. Quinodoz, S. & Guttman, M. Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends Cell Biol.* **24**, 651–663 (2014).
19. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
20. Lai, F. et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497–501 (2013).
21. Wang, K. C. et al. A long non-coding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
22. Yang, Y. W. et al. Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *eLife* **3**, e02046 (2014).
23. Paulsen, M. T. et al. Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. *Proc. Natl Acad. Sci. USA* **110**, 2240–2245 (2013).
24. Brown, J. D. et al. NF-κB directs dynamic super enhancer formation in inflammation and atherogenesis. *Mol. Cell* **56**, 219–231 (2014).
25. Papantonis, A. et al. TNFα signals through specialized factories where responsive coding and miRNA genes are transcribed. *EMBO J.* **31**, 4404–4414 (2012).
26. Matsushima, K. & Morishita, K. Molecular cloning of a human monocyte-derived neutrophil chemotactic factor (MDNCF) and the induction of MDNCF mRNA by interleukin 1 and tumor necrosis factor. *J. Exp. Med.* **167**, 1883–1893 (1988).
27. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
28. Wang, X. et al. MLL1, a H3K4 methyltransferase, regulates the TNFα -stimulated activation of genes downstream of NF-κB. *J. Cell Sci.* **125**, 4058–4066 (2012).
29. Cao, F. et al. Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Mol. Cell* **53**, 247–261 (2014).
30. Raj, A. et al. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
31. Shibayama, Y., Fanucchi, S. & Mhlanga, M. M. Visualization of enhancer-derived noncoding RNA. *Methods Mol. Biol.* **1468**, 19–32 (2017).
32. Engreitz, J. M. et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2017).
33. Goodridge, H. S. et al. Dectin-1 stimulation by *Candida albicans* yeast or zymosan triggers NFAT activation in macrophages and dendritic cells. *J. Immunol.* **178**, 3107–3115 (2007).
34. Asfaha, S. et al. Mice that express human interleukin-8 have increased mobilization of immature myeloid cells, which exacerbates inflammation and accelerates colon carcinogenesis. *Gastroenterology* **144**, 155–166 (2013).
35. Sauter, C. & Wolfensberger, C. Interferon in human serum after injection of endotoxin. *Lancet* **2**, 852–853 (1980).
36. Thin, L. W. et al. Oral tacrolimus for the treatment of refractory inflammatory bowel disease in the biologic era. *Inflamm. Bowel Dis.* **19**, 1490–1498 (2013).
37. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916–919 (2010).

38. Kærn, M. et al. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).
39. Mills, E. L. et al. Succinate dehydrogenase supports metabolic repurposing of mitochondria to drive inflammatory macrophages. *Cell* **167**, 457–470 (2016).
40. Arts, R. J. W. et al. BCG vaccination protects against experimental viral infection in humans through the induction of cytokines associated with trained immunity. *Cell Host Microbe* **10**, 89–100 (2018).
41. Campbell, L. M. et al. Rationale and means to target pro-inflammatory interleukin-8 (CXCL8). *Signal. CancerPharmaceut.* **6**, 929–959 (2013).

Acknowledgements

We thank all members of the Gene Expression and Biophysics Laboratory (the M.M.M. laboratory). We thank M. Lusic, A Gontijo, F. Brombacher, Y. Negishi, L. Davignon and INTRIM consortium members for comments on the manuscript. The authors also thank S. Consalvi, M. Charpentier, A. Boucharlat and the Chemogenomic and Biological screening core facility at the Institut Pasteur in Paris for support during the course of this work. This research is supported by a Department of Science and Technology Centre of Competence Grant, an SA Medical Research Council SHIP grant, and a CSIR Parliamentary Grant, all to M.M.M., and M.M.M. is a Chan Zuckerberg Investigator of the Chan Zuckerberg Initiative. A full list of the investigators who contributed to the generation of the Blueprint Consortium data used in the ChIP-seq project is available from http://www.blueprint-epigenome.eu. Funding for that project was provided by the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement number 282510–BLUEPRINT.

Author contributions

S.F. and M.M.M. designed the study. S.F. performed most experiments and collected and analyzed data. E.T.F. carried out 3C experiments and ChIP and analyzed data. E.D. analyzed CAGE, ChIP and RNA-seq data. Y.S. designed 3C experiments and performed RNA FISH experiments. K.B. and D.G. designed and produced the AAV vectors. E.Y.C. and K.C.W. helped design and perform the UMLILO knock-in experiment. S.S. carried out mass spectrometry experiments and analyzed data. M.I. analyzed Hi-C data. G.L. and W.-K.S. analyzed ChIP and ChIA-PET data. S.F., Y.S., M.I., E.T.F. and M.M.M. discussed and edited the paper. S.F. and M.M.M. co-wrote the paper. M.M.M. designed experiments, analyzed data and supervised the study.

Competing interests

CSIR (Pretoria) has filed a provisional patent application on behalf of S.F., Y.S., E.D. and M.M.M. claiming some of the concepts described in this publication and licensed the patent to Immunolincs Genomics (Seattle, WA).

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0298-2.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.M.M.

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

Methods

Cell culture. Early passage HUVECs from pooled donors (Lonza) were grown to ~80% confluence in EGM-2 with supplements (Lonza), serum-starved (18 h) in EGM-2 + 0.5% FBS, and treated with TNF (10 ng ml^{-1}; Sigma-Aldrich) for up to 24 h. Prior to transfection, cells were grown in antibiotic-free EGM-2. The HeLa, THP-1 and RAW cell lines (Sigma-Aldrich) were maintained in DMEM + 10% FBS. Porcine corneal epithelial (PCE) cells were maintained in corneal epithelial cell medium with supplements (ScienCell). HUVEC, PCE, HeLa, THP-1 and RAW cells were not cultured simultaneously to prevent cross-contamination of cultures. All cell types were routinely checked for the appropriate cellular morphology as well as mycoplasma contamination. All single-cell experiments were performed in primary HUVECs (except for the HOTTIP repair experiment). Due to the large cell numbers required, 3C experiments were performed in the HeLa cell line.

Peripheral blood mononuclear cell and monocyte isolation. Buffy coats from healthy donors were obtained after written informed consent (Western Province Blood Transfusion Services). Peripheral blood mononuclear cell isolation was performed by dilution of blood in pyrogen-free PBS and differential density centrifugation over Ficoll-Paque (GE Healthcare). Cells were washed twice in PBS. Percoll isolation of monocytes was performed as previously described[42]. Briefly, 150×10^6 to 200×10^6 peripheral blood mononuclear cells were layered on top of a hyperosmotic Percoll solution (48.5% Percoll (Sigma-Aldrich), 41.5% sterile H_2O and 0.16 M filter-sterilized NaCl) and centrifuged for 10 min at 580 *g*. The interphase layer was isolated, and cells were washed with cold PBS. Cells were resuspended in RPMI culture medium supplemented with 10 μg ml^{-1} gentamicin, 10 mM Glutamax and 10 mM pyruvate and then counted.

Monocyte β-glucan treatment and inhibition experiment. Monocytes were added to flat-bottom 12-well plates. Monocytes were incubated with culture medium only as a negative control or with 5 μg ml^{-1} β-glucan (β-1,3-D-glucan), kindly provided by D. Williams (James H. Quillen College of Medicine, East Tennessee State University), and/or 5 μM tacrolimus (Sigma-Aldrich) each for up to 24 h.

Hi-C analysis. Hi-C sequencing data were preprocessed (iterative alignment and outlier removal) using the pipeline described by Imakaev et al. [43]. The heat map in the chemokine locus shows the Hi-C interactions as paired-read counts between pairs of sliding windows 50 kb in length.

ChIA-PET analysis. The two Pol II (Ser2/Ser5) ChIA-PET libraries for HUVECs were obtained from the National Center for Biotechnology Information (accession number GSE41553)[25]. One library was obtained 0 min after TNF treatment, while another was obtained 30 min after TNF treatment. Each library yielded about 35 million paired-end reads. Both libraries were processed by the ChIA-PET tool[44]. Briefly, the head and tail ChIA-PET tags are first extracted from the sequences. The ChIA-PET tags are then mapped to the human reference genome (hg19) by BatMis[45], and 10.8×10^6 and 8.8×10^6 are successfully aligned to hg19 for 0 min and 30 min samples, respectively. From the uniquely mapped PETs, we extract all intrachromosomal PETs. Precisely, intrachromosomal PETs are defined as the head and tail of the PETs mapped onto the same chromosome with a genomic distance at least 5 kb. The intrachromosomal PETs are then clustered to form ChIA-PET interaction clusters.

Identification of enhancers in the chemokine TAD. We used the web-based Promoter Enhancer Slider Selector Tool (PrESSTo; http://pressto.binf.ku.dk/) to query the atlas of active, in vivo bidirectionally transcribed human enhancers identified by analysis of the FANTOM5 panel of tissue and primary cell samples[19]. Across the chemokine locus (chr4: positions 74,500,000–75,000,000), we searched for enhancers significantly expressed in the "blood vessel endothelial cell" and the "macrophage" facets (percentage expression ≥ 1, P value < 0.001, Benjamini–Hochberg adjusted false discovery rate < 0.05). Every identified enhancer was associated to its overall percentage of expression and number of tags per million in the facets, as well as its relative tags per million in every single sample belonging to the analyzed facets.

Identification of lncRNAs in the chemokine TAD. We used the ENSEMBL mart human database (release GRCh37.p13) to identify all the lncRNAs mapping to the chemokine locus (chr4: positions 74,500,000–75,000,000). We then queried the FANTOM5 human expression atlas to evaluate the promoter activity associated with the transcription start site of every identified lncRNA. The ZENBU omics interactive visualization system (human hg19, February 2015 updated release) was used to derive the expression levels of each lncRNA in 1,829 human samples (primary cells, cancer cell lines and post-mortem tissues) profiled either in steady state or in time-course experiments.

ChIP-seq analysis. This study makes use of ChIP-seq datasets from Quintin et al.[4] (GSE34260) and ENCODE (accession numbers GSM733673, GSM733690, GSM733691, GSM733683, GSM822279, GSM822306, GSM733757, GSM1000074, GSM1000065 and GSM1000066). This study also makes use of data generated by the Blueprint Consortium.

CRISPR synthesis. Software developed by the F. Zhang laboratory at the Massachusetts Institute of Technology (http://www.genome-engineering.org/) was used to identify CRISPR candidate guide sequences[46]. CRISPRs were generated using the protocol by Cong et al.[46]. pX330 (1 μg) was digested with BbsI for 30 min at 37 °C, gel purified using the QIAquick Gel Extraction Kit (Qiagen) and eluted in elution buffer. Oligos were then phosphorylated using T4 PNK (New England Biolabs) and annealed using the following parameters: 37 °C for 30 min followed by 95 °C for 5 min, then cooling to 25 °C at 5 °C min^{-1}. Annealed oligos were then ligated into the BbsI-digested pX330 vector using Quickligase (New England Biolabs). The ligation reaction was then treated with PlasmidSafe exonuclease (Epicentre) at 37 °C for 30 min. Colony PCR was performed on *Escherichia coli* transformants to identify successful CRISPR clones. HUVECs and THP-1 cells were then microporated with the respective CRISPR using the Neon Transfection System (Life Technologies) according to the manufacturer's instructions. HeLa cells were transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Nuclease activity was assessed by the surveyor assay.

RNA interference. HUVECs, HeLa cells and THP-1 cells were transfected as per manufacturer's instructions with siRNAs (Darmacon) or GapmeRs (Exiqon) targeting UMLILO and other IPLs using the Neon Transfection System (Life Technologies) or Lipofectamine 2000 (Invitrogen). siRNA and LNA sequences are provided in Supplementary Table 4.

qPCR with reverse transcription (RT–qPCR). For all RT–qPCR analyses, whole RNA was extracted using TRIzol (Invitrogen) according to the manufacturer's instructions. Complementary DNA was generated using the SuperScript III cDNA Kit (Invitrogen). All qPCR primers were verified to produce specific products and to perform efficiently. qPCR reactions were performed with technical triplicates on a CFX Real-Time PCR Detection System (Bio-Rad) with SsoFast qPCR Supermix (Bio-Rad). Relative quantification was performed using the delta-delta Ct method with *HPRT* as a reference gene. qPCR primers are provided in Supplementary Table 4.

RNA FISH probes. RNA FISH was performed according to the protocol by Raj et al.[30] using 48 20-mer probes (Biosearch) targeting *IL-8* intron 1, *CXCL1* intron 1, *CXCL2* intron 1, and *HOTTIP* and 28 20-mer probes targeting *eGFP*. Each 20-mer bears a 3′-amino-modifier C6-dT. The amino group was subsequently conjugated to the following *N*-hydroxysuccinimide ester dyes: ATTO-488, ATTO-565, ATTO-647N (ATTO-TEC) or Alexa Fluor 647 (Invitrogen). For the TSA FISH and ChIRP experiment, *UMLILO* exonic sense, *UMLILO* exonic antisense and *HOTTIP* probes were conjugated to biotin. Briefly, oligonucleotide probes were ethanol precipitated and resuspended in 0.1 M sodium tetraborate (Sigma-Aldrich). Approximately 0.3 mg of the *N*-hydroxysuccinimide ester dye (ATTO-TEC) was dissolved in dimethyl sulfoxide (Sigma-Aldrich). The dye solution was added to the probe solution and incubated overnight in the dark at 37 °C. Following conjugation reaction, the probes were ethanol precipitated overnight and resuspended in 0.1 M triethyl ammonium (Sigma-Aldrich). Conjugated probes were separated and purified to enrich for dye-conjugated probes by reverse-phase HPLC on a C18 column.

Immuno-RNA FISH. For each experiment, early-passage HUVECs on coverslips were grown to ~80% confluence, treated with TNF, fixed in 3.7% formaldehyde for 10 min at room temperature, then washed three times in PBS. Cells were permeabilized in ice-cold 90% methanol for 10 min then washed twice with PBS and incubated in blocking buffer (1% BSA/PBS) for 30 min at room temperature on an orbital shaker. Cells were then incubated in primary antibody solution (diluted in 1% BSA/PBS) for 1 h. Goat polyclonal anti-WDR5 (H-35, sc-135245; Santa Cruz Biotechnology) was used to detect WDR5 protein. Coverslips were then washed five times with wash buffer (0.05% Tween-20/PBS) following incubation with secondary antibodies conjugated to either ATTO-488 or ATTO-565 for 1 h. Coverslips were then washed five times with wash buffer (0.05% Tween-20/PBS) and post-fixed with 3.7% formaldehyde/PBS for 10 min at room temperature, followed by further permeabilization in 70% ethanol overnight. For RNA FISH detection, coverslips were washed twice in PBS and incubated in wash buffer (10% formamide, 2X SCC–1X SCC in 0.15 M NaCl plus 0.015 M sodium citrate) for 5 min. Cells were then hybridized overnight in a humidified chamber at 37 °C in 50 μl of Hyb buffer (10% dextran sulfate, 1 μg μl^{-1} *E. coli* transfer RNA, 2 mM vanadyl ribonucleoside complex, 0.02% ribonuclease (RNase)-free BSA, 10% formamide) combined with 50 ng of single-molecule FISH probes. Coverslips were then washed three times (30 min each on the orbital shaker) in wash buffer (10% formamide, 2X SCC). Cells were then incubated in equilibration buffer (0.4% glucose, 2X SCC) for 5 min and counterstained with 1 μg ml^{-1} DAPI (Life Technologies). Coverslips were mounted in glox buffer (3.7 μg μl^{-1} glucose oxidase, 1 unit catalase) and imaged.

TSA FISH. TSA FISH was performed with the Tyramide SuperBoost Kit (Invitrogen) with modifications[31]. Early-passage HUVECs on coverslips were grown to ~80% confluence, treated with TNF, fixed in 3.7% formaldehyde for 10 min at room temperature, then washed three times in PBS. Cells were

permeabilized in ice-cold 90% methanol for 10 min, followed by additional permeabilization in 70% ethanol overnight. Cells were then hybridized overnight in a humidified chamber at 37 °C in 50 µl Hyb buffer combined with 50 ng biotinylated single-molecule RNA FISH probes. The cells were then incubated with 3% hydrogen peroxide solution for 60 min at room temperature and then rinsed three times with 1X PBS. Next, cells were washed twice with PBS and incubated in blocking buffer (1% BSA/PBS) for 30 min at room temperature on an orbital shaker, incubated with HRP-conjugated streptavidin for 60 min at room temperature or overnight at 4 °C and washed for 10 min with PBS at room temperature. This step was repeated three times. A tyramide working solution was prepared as per the manufacturer's recommendation and incubated for 10 min at room temperature. Cells were then washed three times with PBS and then incubated in equilibration buffer (0.4% glucose, 2X SCC) for 5 min and counterstained with 1 µg ml^{-1} DAPI (Life Technologies). Coverslips were mounted in glox buffer and imaged.

Image acquisition and processing. Cells were imaged on a custom-built Nikon Ti Eclipse wide-field total internal reflection fluorescence microscope using a 100X N.A. 1.49 Nikon Apochromat total internal reflection fluorescence oil immersion objective. Imaging was done using mercury lamp illumination through the appropriate filter sets at low camera gain in each of the fluorescent channels using an Andor iXion897 EMCCD camera cooled to −80 °C. The microscope was controlled using µmanager (1.0) open-source microscope management software (National Institutes of Health and University of California, San Francisco). A 20 ms exposure time was used for DAPI. Exposure times ranged from 200 to 500 ms for other dyes. Each field of view was captured as a series of images acquired on multiple focal planes through the samples, across a range of 2–10 µm in the axial plane. A 0.2 µm piezo step size was used for these *z*-stacks. Chromatic aberration was verified before image capture by alignment of Focal Check Fluorescent Microspheres (Molecular Probes). Signal intensities were measured using Fiji (1.0)[47]. The contrast of pictures shown was adjusted to fit a 16 bit gray scale. To facilitate the comparison between different fields of view on the same coverslip, signal intensity values were normalized relative to the intensity of fluorescent beads. No cells were purposely excluded from the single-cell analysis. Cells were imaged at random across each coverslip, for each single-cell experiment.

ChIRP–qPCR. ChIRP–qPCR was performed as per Chu & Chang[48], with modifications. For each ChIRP experiment, 20×10^6 cells were grown to ~80% confluence, fixed in fresh 1% glutaraldehyde for 10 min at room temperature, then washed three times in PBS. Then glycine was added to the medium a final concentration of 125 mM and incubated with shaking for 5 min at room temperature. Cells were rinsed two times with 10 ml cold PBS, scraped into 5 ml cold PBS and centrifuged for 5 min at 1,000 *g*. The resultant pellet was resuspended in lysis buffer (50 mM HEPES–KOH pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, protease inhibitors and Superase-In (added fresh each time)). The nuclear pellet was then sonicated with the Covaris S220 Sonicator for 45 min to an average fragment size of 500 to 1,000 bp and centrifuged at 8,000 *g* for 30 s at 4 °C. The supernatant was transferred to a new tube. A 10 µl portion of each sonicated sample was removed as the INPUT to obtain the RNA and DNA concentration. Chromatin (1 ml) was then mixed with 2 ml of hybridization buffer and 100 ml of biotinylated probe and incubated at 37 °C for 4 h with shaking. Streptavidin magnetic beads (Resyn Biosciences) were then washed three times in lysis buffer, added to hybridization reaction and incubated at 37 °C for 30 min with shaking. Beads were washed three times with 1 ml wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA pH 8, 150 mM NaCl, 20 mM Tris–HCl pH 8) and then one time with final wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA pH 8, 500 mM NaCl, 20 mM Tris–HCl pH 8). DNA from ChIRP and INPUT samples was purified using phenol:chloroform extraction and ethanol precipitated in the presence of 10 µl glycogen (5 mg ml^{-1}) and taken up in 100 µl H_2O. DNA samples were then analyzed by RT–qPCR, using primers that detect the chemokine promoters and other control gene promoters.

Repair construct. To generate double-stranded eGFP PCR product harboring the 20- and 18-bp homologous arms, the following primers were used: forward, 5′-TTGAACCGGGTTTTCCAGTCACATATGGTGAGCAAGGGCGA-3′; reverse, 5′-ACTATGAAGACTCTTGGGTCACTTGTACAGCTCGTCCA-3′. The PCR product was purified by QIAquick PCR Purification Kit (Qiagen) prior to transfection. To generate the HOTTIP repair template, a sequence harboring the exonic portion of HOTTIP, flanked by 352- and 237-bp mouse homologous arms, was synthesized and cloned into puc19 using the SmaI restriction enzyme site (Genscript).

ChIP. For each ChIP experiment, 1×10^7 cells were grown to ~80% confluence, fixed in 3.7% formaldehyde for 10 min at room temperature, then washed three times in PBS. Then glycine was added to a final concentration of 125 mM to the media and incubated with shaking for 5 min at room temperature. Cells were rinsed two times with 10 ml cold PBS, scraped into 5 ml cold PBS and centrifuged for 5 min at 1,000 *g*. The resultant pellet was resuspended in FA lysis buffer (50 mM HEPES–KOH pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, protease inhibitors (added fresh each time)). The nuclear pellet was then sonicated with the Covaris S220 Sonicator to an average fragment size of 500 to 1,000 bp and centrifuged for 30 s at 4 °C and 8,000 *g*. The supernatant was transferred to a new tube. A 50 µl portion of each sonicated sample was removed as the INPUT to obtain the DNA concentration. Protein (25 µg) was used per immunoprecipitation reaction. Protein concentration was calculated using the Bradford assay. Each sample was diluted 1:10 with RIPA buffer (50 mM Tris–HCl pH 8, 2 mM EDTA pH 8, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, protease inhibitors). Antibody was added at 1–2 µg per 25 µg of protein. The following ChIP-grade antibodies were used: anti-H3K4me3 (ab8580; Abcam), MED12 (A300–774A; Bethyl Laboratories), anti-RNA Pol II (Ser5) (ab5131; Abcam), anti-WDR5 (ab56919; Abcam), anti-CTCF (ab70303; Abcam), monoclonal mouse anti-NFAT1 (ab2722; Abcam) and monoclonal mouse anti-NFAT2 (ab2796; Abcam). Magnetic protein A/G beads (20 µl; Resyn Biosciences), pre-absorbed with sonicated single-stranded salmon sperm DNA at 1.5 µg per 20 µl beads, was added to all samples and incubated at 4 °C overnight with rotation. Beads were washed three times with 1 ml wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA pH 8, 150 mM NaCl, 20 mM Tris–HCl pH 8) and then one time with final wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA pH 8, 500 mM NaCl, 20 mM Tris–HCl pH 8). DNA was eluted with 120 µl elution buffer (1% SDS, 100 mM $NaHCO_3$) to the protein A/G beads and incubated at 37 °C for 15 min with rotation. DNA from ChIP and INPUT samples was purified using phenol:chloroform extraction, ethanol precipitated in the presence of 10 µl glycogen (5 mg ml^{-1}) and taken up in 50 µl H_2O.

Mouse repair experiment. To generate the murine HDR template a sequence harboring the endogenous *UMLILO* promoter and the exonic portion of *UMLILO* flanked by 500-bp murine homologous arms was synthesized and cloned into puc19 (Genscript). This HDR template and CRISPRs that target the murine super-enhancer were then electroporated into the murine RAW 264.7 macrophage cell line. The following Neon conditions were used: 1,680 V pulse voltage, 20 ms pulse width, 1 pulse, 5×10^6 cells ml^{-1} with a 10 µl tip. Successful integration of the HDR template was established using the following primers: forward, 5′-GCTTTCTTCCATGCCTGCAA-3′;reverse, 5′-TGTTCAGAGAAATTGCGCGG-3′. Once successful knock-in was established at the population level, clones were obtained by diluting to single cells. A negative clone from this experiment was used as a control for qPCR and H3K4me3 ChIP studies. To screen for the successful knock-in of the *UMLILO* sequence at both alleles the following primers that target the endogenous sequence were used: forward, 5′-ATGGGCTCAAAACACGCATG-3′; reverse, 5′-AGCAGTTTCAAAGCTTCAGGT-3′.

AAV–CRISPR constructs. The AAV vector construct used in this work permits co-expression of three distinct SpCas9 guide RNAs (gRNAs) from a single AAV vector backbone under the RNA Pol III promoters U6, H1 or 7SK. All gRNAs were embedded in an optimized scaffold reported previously[49]. Forward sequences of the gRNAs were as follows: U6 promoter, G*CTCCACATGTATAAT GTGAC; H1 promoter, G*CTAAACTGTATAGTCTGTAA; 7SK promoter, G*AAAAACCAATTATTGAACAA. Each gRNA was preceded by guanine, which is the preferred start nucleotide for the three promoters used. The gRNAs were first cloned under the individual promoters as annealed oligonucleotides with 5′-CACC-3′ and 5′-AAAC-3′ overhangs, before single gRNA expression cassettes were concatemerized using a Golden Gate assembly strategy, details of which will be reported separately (unpublished results, K.B. and D.G.).

The AAV vector plasmid expressing SpCas9 under the control of a cytomegalovirus promoter was derived from pSSV9, a plasmid routinely used for generation of single-stranded AAV vectors[50], and encoded the SpCas9 cDNA variant from the F. Zhang laboratory at the Massachusetts Institute of Technology (plasmid 49535; Addgene).

Production of crude AAV vector lysates. Small-scale AAV vector stocks ('crude lysates') were produced in six-well plates using a standard triple transfection protocol essentially as reported previously[51]. Briefly, HEK293T cells were seeded (3.5×10^5 cells per well) in a volume of 4 ml. Transfection was carried out 24 h later using TurboFect transfection reagent(Thermo Fisher Scientific) following the manufacturer's instructions for this format. The triple transfection mixture contained equal amounts (1.3 µg each) of the AAV helper plasmid encoding *rep* and *cap*, an adenovirus helper plasmid and the vector construct encoding the gRNA(s) of interest or SpCas9.

Here, *cap* was derived either from wild-type AAV6 or from a variant of AAV-DJ (a hybrid of serotypes 2, 8 and 9)[52] displaying a short (9 amino acids long) peptide on its surface that enhances cell transduction. Details of this peptide will be reported elsewhere (K.B. and D.B., manuscript in preparation). Another 72 h later, cells were collected, resuspended in 300 µl 1X PBS and subjected to five cycles of freezing and thawing. The AAV vector particle-containing supernatant was cleared from cell debris by 10 min centrifugation at 16,100 *g*, and the vectors were stored at −20 °C.

RNA-seq experiment. Quality-control analysis on RNA was performed using the Qubit DNA Assay Kit (Thermo Fisher Scientific), gel electrophoresis and an

Agilent 2100 bioanalyzer. All the RNA samples passed the quality control. RNA-seq libraries were prepared using the Illumina TruSeq Stranded mRNA Library Preparation Kit following the manufacturer's recommendations (Applied Biological Materials). An average of 30 million paired-end reads were generated per sample. Quality control of sequence data was performed using FastQC tool (v.0.11.5). Sequencing reads were then aligned to the human GRCh38 reference genome using Hisat2 (v.2.1.0). Fragments per kilobase of transcript per million for each sample was estimated using StringTie (v.1.3.3).

RNA immunoprecipitation and WDR5. UMLILO and HOTAIR-encoding cDNAs were cloned into a eukaryotic gene expression plasmid and co-transfected with pcDNA3.1 that did or did not express FlagWDR5. Exon 1, exon 2, exon 3, porcine UMLILO and the region encoding the unconserved region in pig were synthesized and cloned into pTwist Amp (Twist Biosciences) and co-transfected with pcDNA3.1 that did or did not express FlagWDR5. For each RNA immunoprecipitation experiment, 1×10^7 cells were grown to ~80% confluence, fixed in 3.7% formaldehyde for 10 min at room temperature, then washed three times in PBS. Then glycine was added to the medium to a final concentration of 125 mM, followed by incubation with shaking for 5 min at room temperature. Cells were lysed and immunoprecipitated as previously described, with modifications[28].

RNA immunoprecipitation with targeted mass spectrophotometry. For each experiment, 1×10^7 cells were grown to ~80% confluence, then washed three times in PBS. Briefly, cells were resuspended in Buffer A (10 mM HEPES pH 7.5, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 1.0 mM PMSF), lysed in 0.25% NP40 and fractionated by low-speed centrifugation. The nuclear fraction was resuspended and lysed in Buffer C (20 mM HEPES pH 7.5, 10% glycerol, 0.42 M KCl, 4 mM $MgCl_2$, 0.5 mM DTT, 1.0 mM PMSF). Either the nuclear or the cytoplasmic fraction was incubated with biotinylated oligonucleotide probes tiling exonic UMLILO in Hyb buffer (100 mM Tris–HCl pH 7.0, 500 mM NaCl, 10 mM EDTA pH 8.0, 15% formamide, 1 mM DTT, 1% SDS, 0.1 U μl^{-1} Superase) at 37 °C for 4 h. Magnetic protein A/G beads (20 µl; Resyn Biosciences) were added to all samples and incubated at 37 °C for 1 h with rotation. Beads were washed three times with wash buffer (2X SCC, 0.5% SD, 1 mM DTT). Proteins were eluted with SDS loading buffer and subject to SDS–PAGE analysis. Gels were stained using colloidal Coomassie, and protein bands of interest were in-gel trypsin digested as per the protocol described in Shevchenko et al.[53]. In short, gel bands were de-stained using 50 mM NH_4HCO_3/50% MeOH followed by in-gel protein reduction (50 mM DTT in 25 mM NH_4HCO_3) and alkylation (55 mM iodoacetamide in 25 mM NH_4HCO_3). Proteins were digested overnight at 37 °C using 5–50 µl of 10 ng μl^{-1} tryspin depending on the gel piece size. Digests were resuspended in 40 µl of 2% acetonitrile/0.2% formic acid and analyzed using a Dionex Ultimate 3000 rapid-separation liquid chromatography system coupled to an AB Sciex 6600 TripleTOF mass spectrometer. Peptides were first de-salted on an Acclaim PepMap C18 trap column (75 µm × 2 cm) for 8 min at 5 µl min^{-1} using 2% acetonitrile/0.2% formic acid, then separated on Acclain PepMap C18 RSLC column (75 µm × 15 cm, 2 µm particle size). Peptide elution was achieved using a flow rate of 500 nl min^{-1} with a gradient of 4–60% acetonitrile with 0.1% formic acid in 30 min. Nanospray was achieved using a NanoSpray III source assembled with a New Objective, PicoTip emitter. An electrospray voltage of 2.3 kV was applied to the emitter. Precursor spectra were acquired in the range 400–1,500 m/z (0.25 s accumulation time). Product ion spectra were acquired for 18 precursors reporting on 5 targeted proteins. Product scans were recorded in the mass range 100–1,500 m/z, in high-sensitivity mode, with an accumulation time of 0.1 s for a total cycle time of 2.1 s. Data processing was performed in Skyline v2.6 (ref. [53]). MS1 filtering was performed at 30,000 and MS2 at 15,000 resolution with a match tolerance of 0.025 m/z.

3C assay. The 3C experiments were performed as described previously by Hagege et al.[54] with some minor modifications. Briefly, a single-cell suspension of 8×10^6 HeLa cells was cross-linked with 1% formaldehyde in 10% (v/v) FBS/PBS for 10 min at room temperature and quenched by adding 0.125 M glycine and incubating for 5 min on ice. Cells were then pelleted by centrifugation at 225 g for 8 min at 4 °C and lysed on ice in 3C lysis buffer (10 mM Tris–HCl pH 7.5, 10 mM NaCl, 0.2% NP-40, 1X complete protease inhibitor (Roche)) for 30 min and subsequently dounce homogenized. Nuclei were collected by centrifugation at 400 g for 5 min at 4 °C. The pellets were resuspended in 1.2X Buffer R (Thermo Fisher Scientific) with 0.3% SDS and incubated at 37 °C for 1 h while being shaken at 250 r.p.m. The SDS was sequestered by adding 2% Triton X-100 and incubating this for 1 h at 37 °C with shaking. HindIII (550 units, 10 units μl^{-1}; Thermo Fisher Scientific) was then added and incubated overnight at 37 °C with shaking. The digestion was terminated with 1.3% SDS at 65 °C for 20 min. Next, 6.125 ml of 1.15X T4 ligation buffer (Thermo Fisher Scientific) with 1% Triton X-100 was added and incubated at 37 °C for 1 h with mild shaking (150 r.p.m.). Samples were allowed to cool to room temperature before 100 units T4 DNA ligase (Thermo Fisher Scientific) was added and incubated at 16 °C for 4 h with gentle shaking (90 r.p.m.), followed by an additional 30 min at room temperature. Cross-links were reversed with 300 µg proteinase K and overnight incubation at 65 °C. The following day, 300 µgRNase A was added to the sample and incubated at 37 °C for 30 min. Finally, genomic DNA was purified by phenol:chloroform extraction followed by isopropanol precipitation. Ligation events were detected using unidirectional, fragment-specific primers. qPCR reactions were performed with technical triplicates on a CFX Real-Time PCR Detection System (Bio-Rad) with SsoFast qPCR Supermix (Bio-Rad). A putative genomic interaction was used to correct for variations between 3C library preparations. Interaction frequencies were calculated relative to the random interaction frequencies as determined with a BAC template (cat. no. RPCI-11 447E20) spanning the genomic region of interest.

Statistics. Figure legends describe the statistical details of experiments, including the statistical tests used, the kind of replicates and the value of n. Asterisks define degree of significance as described in the figure legends ($^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$). All Student's t-tests and Fisher's exact tests were analyzed as two sided. All statistical analyses were performed using Prism (GraphPad Software v.7.0).

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

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. RNA-seq data are available in the Gene Expression Omnibus under accession number GSE120621.

References

42. Repnik, U., Knezevic, M. & Jeras, M. Simple and cost-effective isolation of monocytes from buffy coats. *J. Immunol. Methods* **278**, 283–292 (2003).
43. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
44. Li, G. et al. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.* **11**, R22 (2010).
45. Tennakoon, C. et al. BatMis: a fast algorithm for k-mismatch mapping. *Bioinformatics* **28**, 2122–2128 (2012).
46. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
47. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
48. Chu, C. & Chang, H. Y. Understanding RNA-chromatin interactions using chromatin isolation by RNA purification (ChIRP). *Methods Mol. Biol.* **1480**, 115–123 (2016).
49. Chen, B. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
50. Samulski, R. J. et al. A recombinant plasmid from which an infectious adeno-associated virus genome can be excised in vitro and its use to study viral replication. *J. Virol.* **61**, 3096–3101 (1987).
51. Grimm, D. Production methods for gene transfer vectors based on adeno-associated virus serotypes. *Methods* **28**, 146–157 (2002).
52. Grimm, D. et al. In vitro and in vivo gene therapy vector evolution via multi species interbreeding and retargeting of adeno-associated viruses. *J. Virol.* **82**, 5887–5911 (2008).
53. Shevchenko, A. et al. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2007).
54. Hagege, H. et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* **2**, 1722–1733 (2007).

nature research

Corresponding author(s): Musa M. Mhlanga

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	µmanager (1.0) open source microscope management software; PrESSto (Promoter Enhancer Slider Selector Tool, http://pressto.binf.ku.dk/)
Data analysis	Prism (7.0; GraphPad Software); Fiji (1.0); HISAT2 (v2.1.0. https://ccb.jhu.edu/software/hisat2/index.shtml); StringTie (v1.3.3); FastQC (0.11.5); Skyline (v2.6); Batmis (v3.0); Crispr design (v1) http://crispr.mit.edu/.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	As per common qRT-PCR and qChIP-PCR practice, at least three independent biological replicates were generated per treatment.
Data exclusions	Data was not excluded from analysis.
Replication	Experiments were repeated by two experimentalists and analyzed independently leading to the same conclusions.
Randomization	The majority of experiments were performed on cell lines. Therefore, randomization is not applicable for these experiments . Buffy coats were chosen at random.
Blinding	Not applicable. No group allocation was performed in this study.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	The following ChIP grade antibodies were used: anti-H3K4me3 (Abcam, ab8580), MED12 (Bethyl laboratories, A300-774A), anti-RNA Pol II (Ser5) (Abcam, ab5131), anti-WDR5 (Abcam, ab56919), anti-CTCF antibody (Abcam, ab70303), anti-NFAT1 mouse mAb (Abcam, ab2722), anti-NFAT2 mouse mAb (Abcam, ab2796).
Validation	As is stated on the manufacturers' website, each primary antibody has been individually validated to react with human, mouse and pig.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HUVECs (Lonza), HeLa, THP-1 and RAW cell lines (Sigma); Porcine corneal epithelial cells (ScienCell)
Authentication	HUVECs, THP-1, RAW and PCEs were authenticated by the supplier. HeLas were not authenticated.
Mycoplasma contamination	All cell lines tested negative for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	Due to the cell numbers required, HeLas were used to perform the 3C experiments.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Buffy coats from healthy donors were obtained after written informed consent (Western Province Blood Transfusion Services). Samples were chosen at random.
Recruitment	Samples were chosen at random.