nature genetics

# Hierarchy within the mammary STAT5-driven *Wap* super-enhancer

Ha Youn Shin[1,3], Michaela Willi[1–3], Kyung Hyun Yoo[1,3], Xianke Zeng[1], Chaochen Wang[1], Gil Metser[1] & Lothar Hennighausen[1]

**Super-enhancers comprise dense transcription factor platforms highly enriched for active chromatin marks. A paucity of functional data led us to investigate the role of super-enhancers in the mammary gland, an organ characterized by exceptional gene regulatory dynamics during pregnancy. ChIP-seq analysis for the master regulator STAT5A, the glucocorticoid receptor, H3K27ac and MED1 identified 440 mammary-specific super-enhancers, half of which were associated with genes activated during pregnancy. We interrogated the *Wap* super-enhancer, generating mice carrying mutations in STAT5-binding sites within its constituent enhancers. Individually, the most distal site displayed the greatest enhancer activity. However, combinatorial mutation analysis showed that the 1,000-fold induction in gene expression during pregnancy relied on all enhancers. Disabling the binding sites of STAT5, NFIB and ELF5 in the proximal enhancer incapacitated the entire super-enhancer. Altogether, these data suggest a temporal and functional enhancer hierarchy. The identification of mammary-specific super-enhancers and the mechanistic exploration of the *Wap* locus provide insights into the regulation of cell-type-specific expression of hormone-sensing genes.**



The main purpose of the mammary gland is to produce large quantities of milk to support newborns. The milk-secreting alveolar epithelium, absent in the virgin state, is established during pregnancy and is considerably remodeled upon cessation of lactation[1]. Proliferation and differentiation of mammary alveoli during pregnancy is controlled by the hormones progesterone[2] and prolactin[3,4] through the transcription factors STAT5 (refs. 5,6) and ELF5 (ref. 7). STAT5 (refs. 8,9) has emerged as the critical transcription factor activating genes encoding milk proteins, and other differentiation-associated proteins linked to milk secretion, by up to 10,000-fold during pregnancy[10]. The ten most abundant mRNAs encoding milk proteins account for more than 90% of the mRNA in mammary gland[11,12], and expression of the respective genes is synergistically induced by prolactin and glucocorticoids[13], making these genes an ideal system to investigate underlying transcriptional regulation.

Traditionally, transgenic mice have been used to identify sequences conveying mammary specificity and hormonal responsiveness[13–19]. Although informative, the transgenes used contained only limited promoter and upstream sequences, which conveyed mammary-restricted expression but normal hormonal regulation during pregnancy, and expression levels were not achieved. These results suggest that enhancers and insulators were located outside the sequences used for these transgenes. Putative regulatory elements in the genome can now be identified using ChIP-seq analyses for transcription factors and specific histone modifications such as acetylation of histone H3 at lysine 27 (H3K27ac)[20–23]. This permits a more accurate approach in the genetic analysis of these sequences.

The emerging concept of super-enhancers proposes that lineage-specific genes are under the control of special enhancer clusters, which are characterized by the presence of several densely occupied transcription factor platforms and extended H3K27ac marking[24,25]. Although super-enhancers have been reported in a plethora of cell types[26–46], there is scant genetic support of their unique biological relevance. Several hundred genes are expressed specifically in mammary tissue under the control of prolactin[10], making them an ideal test system for the concept of super-enhancers. Here we have integrated ChIP-seq analyses for the mammary master regulator STAT5A, glucocorticoid receptor (GR), the Mediator complex subunit MED1 and the active chromatin mark H3K27ac to identify putative mammary-specific super-enhancers. To investigate the biological relevance of super-enhancers, we focused on the one associated with *Wap*[47], a gene highly expressed in mammary tissue and induced by more than 1,000-fold during pregnancy[48]. We mutated constituent enhancers, individually and in combination, within the *Wap* super-enhancer, determined their respective importance, and identified a temporal and functional hierarchy among them.

## RESULTS

### Identification of mammary-specific super-enhancers

In search of mammary-specific super-enhancers, we conducted ChIP-seq experiments for STAT5A, a master regulator in mammary

[1]Laboratory of Genetics and Physiology, National Institute of Diabetes, Digestive and Kidney Diseases, US National Institutes of Health, Bethesda, Maryland, USA. [2]Division of Bioinformatics, Biocenter, Medical University of Innsbruck, Innsbruck, Austria. [3]These authors contributed equally to this work. Correspondence should be addressed to L.H. (lotharh@mail.nih.gov).

Received 14 March; accepted 3 June; published online 4 July 2016; doi:10.1038/ng.3606

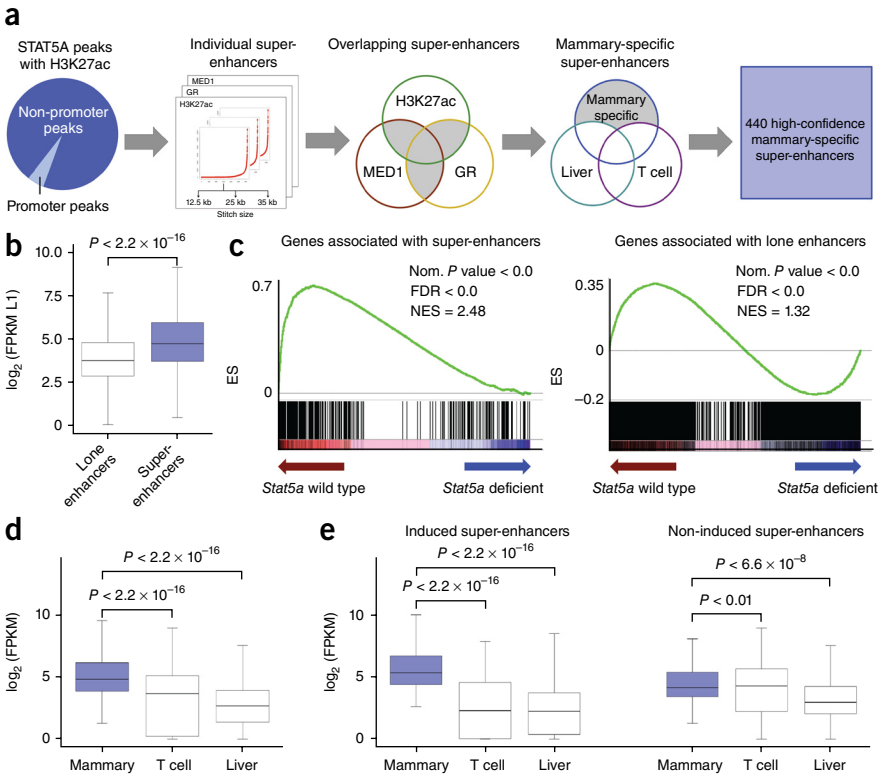


**Figure 1** Identification of mammary-specific super-enhancers. (**a**) Of 10,953 STAT5A peaks coinciding with H3K27ac marks, 549 are located in promoter regions and 10,404 are located in non-promoter regions. The 10,404 non-promoter peaks, together with H3K27ac, GR and MED1 peaks, served as the basis for super-enhancer analysis[37,46] using three stitch sizes as the parameter for the calculations. The final steps comprised overlapping super-enhancers by size and removing super-enhancers shared with T cells and liver. A total of 440 bona fide mammary-specific super-enhancers were identified. (**b**) The box plots depict significantly higher expression for 384 genes associated with super-enhancers (mean of ~5,931 FPKM) than for 4,384 genes linked to lone enhancers (mean of ~31 FPKM) at day 1 of lactation (L1) (genes under 5 FPKM were not considered). Median, middle bar inside each box; IQR (interquartile range), the box containing 50% of the data; whiskers, 1.5 times the IQR. (**c**) GSEA shows that super-enhancers are less enriched in *Stat5a*-deficient samples than in samples with wild-type STAT5A (70% reduction). Nom. *P* value, nominal *P* value; FDR, false discovery rate; NES, normalized enrichment score; ES, enrichment score. (**d**) The box plots show that expression of the 384 genes associated with the mammary-specific super-enhancer is significantly elevated at day 1 of lactation, when compared to T cells and liver tissue. (**e**) Super-enhancer-associated genes were categorized as genes induced by at least twofold between day 6 of pregnancy and day 1 of lactation (198) and genes that were not induced (186). Notably, induced genes exhibit lineage specificity, and their expression in T cells and liver is much lower than in mammary tissue.

epithelium[5,6], H3K27ac, GR and MED1. Mammary tissue was initially analyzed at lactation when mammary-specific genes are highly expressed. On the basis of the colocalization of transcription factor binding and H3K27ac marks and defined 'stitch' sizes[46] (**Fig. 1a** and **Supplementary Figs. 1–3**), we identified approximately 580 super-enhancers. After removing STAT5 super-enhancers common to mammary tissue and either liver or T cells, we obtained 440 mammary-specific super-enhancers, encompassing a total of 2,712 individual enhancers (**Fig. 1a** and **Supplementary Table 1**). These 440 super-enhancers were exclusively identified on the basis of ChIP-seq experiments, and their biological function has not been experimentally validated.

To assess the functional relevance of super-enhancers in lactating mammary tissue, we used RNA-seq data[10] and compared the expression of genes associated with super-enhancers to that of genes associated with lone STAT5 enhancers (**Fig. 1b**). Whereas the median expression level of genes linked to lone enhancers was ~14 FPKM (mean of ~31 FPKM), the median expression level was ~27 FPKM (mean of ~5,931 FPKM) for genes associated with super-enhancers. Gene set enrichment analysis (GSEA) demonstrated that super-enhancers were preferentially associated with genes with STAT5-dependent expression (**Fig. 1c**). The expression of approximately 35% of super-enhancer-associated genes was induced by more than twofold by STAT5 (**Supplementary Table 1**), and approximately 50% of the genes associated with super-enhancers were also induced during pregnancy. Notably, although most highly induced genes were associated with super-enhancers, some, such as *Aldoc*, were linked to lone enhancers (**Supplementary Fig. 3**).

Next, we examined whether the expression of super-enhancer-associated genes was specifically elevated in mammary tissue. We analyzed expression in mammary tissue at day 1 of lactation, in T helper type 1 ($T_H1$) cells and in liver tissue, all of which are targets of cytokines that activate STAT5. Whereas the expression level for super-enhancer-associated genes in mammary tissue averaged 27 FPKM (mean of ~5,931 FPKM), the average expression level was ~12 FPKM (mean of ~31 FPKM) in T cells and ~6 FPKM (mean of ~28 FPKM) in liver (**Fig. 1d**). Although these findings demonstrate a preference for mammary expression among genes associated with mammary super-enhancers, they also show that some of these genes are expressed in non-mammary cells under cytokine control. Approximately 50% of super-enhancer-associated genes were induced during pregnancy through STAT5, and their expression was highly enriched in mammary tissue (**Supplementary Table 1**) at a median of ~43 FPKM (mean of ~11,388 FPKM) (**Fig. 1e**, left). In contrast, the expression of genes not induced during pregnancy was only slightly enriched in mammary tissue (**Fig. 1e**, right).

## Establishment of mammary super-enhancers during pregnancy

Enhancers are characterized by the binding of several transcription factors at so-called hotspots[43,49]. A motif search (see the Online Methods for details) identified an enrichment of predicted binding sites for ELF5 and NFIB (**Fig. 2a**), transcription factors critical to mammary development[7,13,50]. ChIP-seq experiments validated binding of NFIB and ELF5 to enhancers also occupied by STAT5A, GR and MED1 (**Fig. 2b,c**), as exemplified by the *Wap* and *Olah* loci (**Fig. 2b** and **Supplementary Fig. 4**).

We hypothesized that mammary super-enhancers are established during pregnancy, a time frame during which the expression of mammary-specific genes greatly increases. To examine this hypothesis, we performed ChIP-seq for STAT5A, GR, NFIB, ELF5, MED1 and

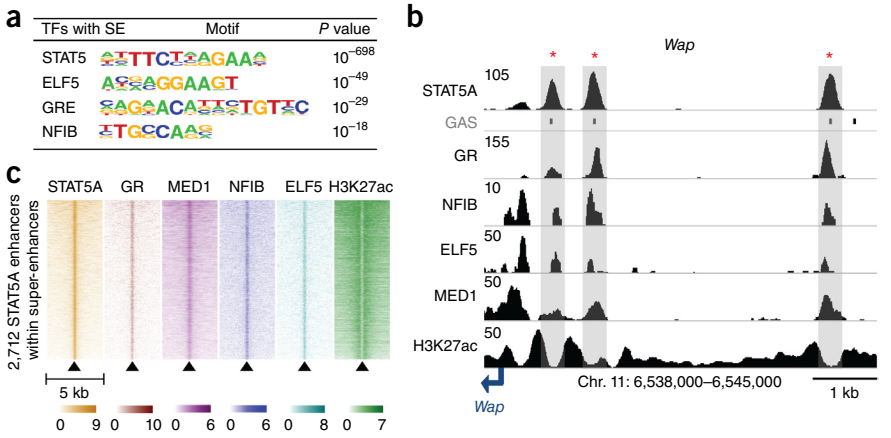


**Figure 2** Transcription factor binding signatures in mammary-specific super-enhancers. (**a**) Motif analysis in mammary-specific super-enhancers. Predicted motifs for transcription factors critical for mammary development were highly enriched in mammary-specific super-enhancers (±200 bp). TF, transcription factor; SE, super-enhancer. (**b**) Transcription factor binding profiles of the constituent enhancers in a mammary-specific super-enhancer. The *Wap* mammary-specific super-enhancer is shown as a representative example; constituent enhancers are indicated by red asterisks. The data are representative of biological duplicates. (**c**) Heat maps of H3K27ac marks and transcription factor binding for STAT5A, GR, MED1, NFIB and ELF5 in a 5-kb region centered on hotspots (triangles) in mammary-specific super-enhancers. Super-enhancers are ordered on the *y* axis by the row sum of the STAT5A signal intensities for all transcription factors.

H3K27ac on mammary tissue at day 13 of pregnancy, before the activation of key genes including *Wap*. Whereas 7% (32) of all super-enhancers were not occupied at day 13 of pregnancy and seemed to be established only at day 1 of lactation, some constituent enhancers within the remaining 93% (408) of super-enhancers showed limited transcription factor occupancy at day 13 of pregnancy (**Fig. 3a**). In 56% of the super-enhancers, less than half of the ChIP-seq peaks were established at day 13 of pregnancy, 32% of the super-enhancers had coverage for more than half of the peaks and in 5% of the super-enhancers all peaks were already fully established. The sensitivity of ChIP-seq experiments depends on the quality of the antibodies used, and it is possible that the data displayed in **Figure 3a** underestimate transcription factor binding. Next, we assessed to what extent the temporal gain in transcription factor binding to these enhancers translated into activation of associated genes (**Fig. 3b**). Genes associated with super-enhancers without transcription factor binding (group I) or that were only partially occupied (groups II and III) at day 13 of pregnancy were characterized by higher levels of induction than genes already fully occupied at this stage (group IV). Furthermore, genes associated with fully established super-enhancers at day 13 of pregnancy showed higher levels of expression (**Supplementary Fig. 5**). Super-enhancers without apparent STAT5A occupancy at day 13 of pregnancy showed the highest level of induction, followed by those with less than 50% occupancy and those with more than 50% occupancy (**Fig. 3c**, left). In contrast, super-enhancers associated with non-induced genes showed equivalent expression patterns across the groups (**Fig. 3c**, right).

The individual enhancers within a given super-enhancer were frequently not established in concert during pregnancy but were established in a seemingly defined temporal order, suggesting that they might differentially sense prolactin and, possibly, transcription factor concentrations. Notably, some individual enhancers within super-enhancers were already occupied during pregnancy, before the activation of the associated genes. To attain a better understanding of the temporal progression in the establishment of mammary-specific super-enhancers, we focused on the tripartite enhancer of *Wap*[47], a gene induced by more than 1,000-fold during pregnancy[13,51]. The *Wap* super-enhancer consists of three constituent enhancers, with one (E1) already occupied by STAT5A, GR, NFIB, ELF5 and MED1 at mid pregnancy (**Fig. 3d** and **Supplementary Fig. 6a**). In contrast, E2 and E3 were fully occupied only at the onset of lactation. Similarly, one of the three individual enhancers within the *Glycam1* super-enhancer was already established during pregnancy, before transcriptional activation of the gene (**Fig. 3e**). These findings suggest that individual enhancers within mammary-specific super-enhancers may have unique properties and differentially sense hormonal cues as pregnancy progresses. Moreover, enhancers already occupied in early pregnancy, such as E1 in the *Wap* super-enhancer, seem generally unable to activate mammary-specific gene expression by themselves. The window between day 13 of pregnancy and day 1 of lactation is characterized by epithelial differentiation with little proliferation, suggesting that the establishment of super-enhancers is the result of cellular differentiation. Expression of *Krt8* and *Krt18*, markers of mammary secretory epithelium, was equivalent between day 13 of pregnancy and day 1 of lactation (**Supplementary Table 1**). To further strengthen this evidence, we conducted STAT5A and H3K27ac ChIP-seq experiments at days 14 and 16 of pregnancy, stages that are distinguished by their differentiation status, as evidenced by the 100-fold activation of *Wap* gene expression (**Fig. 3f**,**g** and **Supplementary Fig. 6b**). Whereas E1, but not E2 and E3, was occupied by STAT5A at day 14 of pregnancy, binding at all sites was secured by day 16, indicating that the *Wap* super-enhancer is fully activated within this narrow window. Although the expansion of mammary epithelium between days 14 and 16 of pregnancy was negligible and the respective tissues had an equivalent appearance (**Supplementary Fig. 7**), we analyzed the gain of signal specific to mammary epithelium and still compared specific chromatin marks in intact tissue and enriched mammary epithelial cells (MECs; **Supplementary Fig. 8**).

### Hierarchy within the mammary-specific *Wap* super-enhancer

The progressive establishment of mammary-specific super-enhancers during pregnancy parallels the activation of some, but not all, associated genes, suggesting a possible causal relationship. Because the creation of mammary alveoli is strictly dependent upon STAT5 (refs. 6,52), it has not been possible to investigate the *in vivo* role of STAT5 in the establishment of individual mammary-specific enhancers[10] in tissue devoid of STAT5. To study the contribution of individual STAT5-binding sites within a mammary-specific super-enhancer, we focused on *Wap*, which is activated by more than 1,000-fold during pregnancy[13,51]. First, we individually deleted the γ-interferon–activated sequence (GAS) motif in each of the three putative STAT5-bound enhancers (E1–E3) (**Fig. 4a** and **Supplementary Fig. 9**). The proximal E1 site was mutated using CRISPR/Cas9 genome editing, and the E2 and E3 sites were deleted using TALEN technology (**Fig. 4b**). Mammary tissue from homozygous-mutant mice was analyzed at the onset of lactation. Deletion of the proximal STAT5 enhancer site (ΔE1) resulted in a 62% reduction in *Wap* mRNA levels, loss of E2 (ΔE2) resulted in a 48% reduction and inactivation of the distal site (ΔE3) led

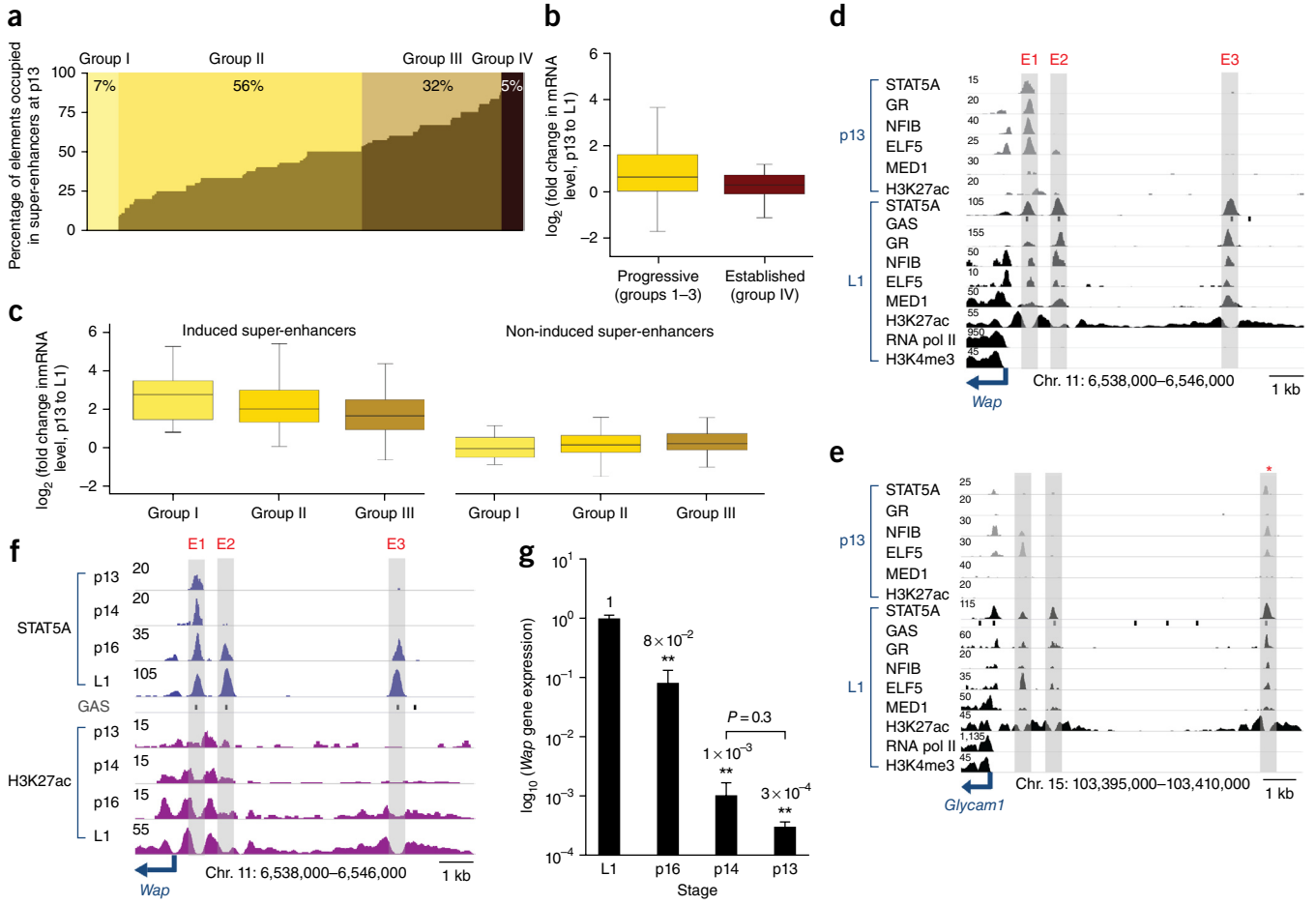


**Figure 3** Assembly at constituent enhancers in mammary-specific super-enhancers during pregnancy. (**a**) Seven percent of super-enhancers have no established peaks at day 13 of pregnancy, 56% have less than half their enhancers occupied, 32% have more than half their enhancers occupied and 5% are already established. (**b**) Genes associated with progressive super-enhancers (418) show higher levels of induction than those fully occupied at day 13 of pregnancy (22). Median, middle bar inside the box; IQR, 50% of the data; whiskers, 1.5 times the IQR. (**c**) Progressive enhancers associated with genes induced during pregnancy show higher levels of induction (left) than those associated with genes that are not induced (right). Super-enhancers having no established enhancers at day 13 of pregnancy show the highest level of induction, followed by those with less than 50% and more than 50% of their enhancers already occupied. (**d**) Establishment of enhancers in the mammary-specific *Wap* super-enhancer. Only E1 is occupied by mammary-enriched transcription factors at day 13 of pregnancy, whereas E2 and E3 are exclusively occupied at day 1 of lactation. Data are representative of biological duplicates. (**e**) Progressive establishment of the mammary-specific *Glycam1* super-enhancer. One of three individual enhancers (asterisk) is already occupied by mammary transcription factors at day 13 of pregnancy, and all three enhancers are fully occupied at lactation. (**f**) Establishment of individual enhancers in the mammary-specific *Wap* super-enhancer across different pregnancy stages. Only E1 is occupied by STAT5A and has H3K27ac marks at days 13 and 14 of pregnancy, whereas all three enhancers are fully occupied at day 16 of pregnancy and day 1 of lactation. (**g**) *Wap* mRNA levels in mammary tissue during different stages of pregnancy. A two-tailed unpaired Student *t* test was used to evaluate statistical significance: $^{**}P < 0.001$. L1, $n = 6$; p16, p14 and p13, $n = 3$. The *Csn* locus served as a control (**Supplementary Fig. 6**).

to a reduction of 91% (**Fig. 4c**). These findings indicate that the three STAT5-bound enhancer units, despite their similar transcription factor occupancy and H3K27ac profiles, possess different roles *in vivo*, suggesting unique contributions for STAT5 at these sites.

Although highly informative, this study did not explore possible interactions between the three constituent enhancers that would account for the 1,000-fold induction in *Wap* expression. To investigate such interactions, we conducted successive gene targeting using CRISPR/Cas9 editing and generated mice carrying different combinations of mutations. Because E1 is established before *Wap* gene activation, preceding the formation of E2, we explored the functional relationship between these enhancers. E1-mutant embryos were targeted at E2, and ΔE1a/ΔE2 mice were generated. The combined loss of these two STAT5-binding sites resulted in an 87% reduction in *Wap* expression (**Fig. 4c**), less than what had been observed for the E3 mutant. This finding provides further evidence of the prominent status of the distal E3 enhancer. An indication that the E2 and E3 enhancers might be the key to the extraordinary activation of *Wap* came from the observation that STAT5A occupancy of these sites coincided with the 100-fold activation of *Wap* between days 14 and 16 of pregnancy (**Fig. 3f,g**). To test this hypothesis, we targeted E2 in E3-mutant embryos and generated mice lacking both sites (ΔE2/ΔE3 mutants). *Wap* expression in these mice was reduced by ~95% relative to wild-type mice (**Fig. 4c**). Although the induction of *Wap* from these enhancers is extraordinary in comparison to that of other enhancer-dependent genes, this still does not account for the 1,000-fold induction in expression of endogenous *Wap*. To test whether the combination of all three sites is necessary for full induction, we generated mice devoid of all three STAT5-binding sites. E1-mutant embryos were targeted for E2 and E3, and homozygous mice (ΔE1a/ΔE2/ΔE3) were generated and analyzed. In the absence of the three

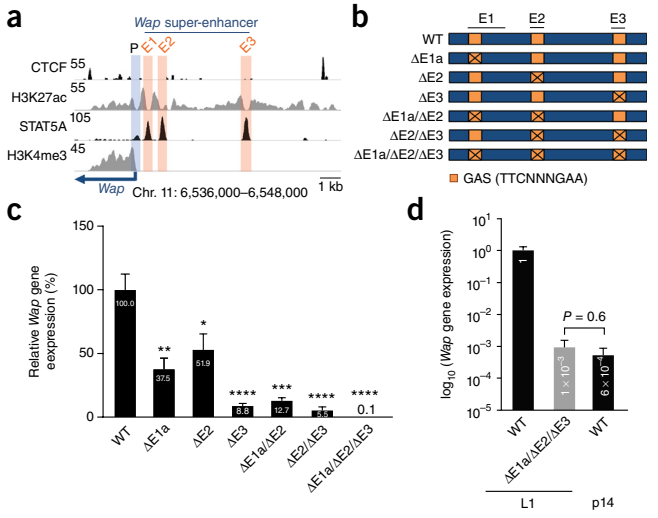


**Figure 4** *In vivo* functions of individual enhancers in the mammary-specific *Wap* super-enhancer. (**a**) Genomic features of E1, E2 and E3 in the *Wap* locus. H3K4me3, trimethylation of histone H3 at lysine 4. (**b**) Schematics of the genomes for single-enhancer mutants (ΔE1a, ΔE2 and ΔE3) and mutants with STAT5 sites deleted in combination (ΔE1a/ΔE2, ΔE2/ΔE3 and ΔE1a/ΔE2/ΔE3). The exact positions of the deletions are shown in **Supplementary Figure 9**. WT, wild type. (**c**) *Wap* mRNA levels in mammary tissues from mice carrying single enhancer mutations (ΔE1a, ΔE2 and ΔE3) or mutations in combination (ΔE1a/ΔE2, ΔE2/ΔE3 and ΔE1a/ΔE2/ΔE3) at day 1 of lactation. *Wap* mRNA levels were measured by qRT–PCR and normalized to *Gapdh* levels. Results are shown as the means ± s.e.m. of independent biological replicates (WT, $n = 9$; ΔE1a, $n = 7$; ΔE2, $n = 10$; ΔE3, $n = 7$; ΔE1a/ΔE2, $n = 3$; ΔE2/ΔE3, $n = 3$; ΔE1a/ΔE2/ΔE3, $n = 3$). A two-tailed unpaired *t* test was used to evaluate the statistical significance of differences between wild-type mice and each mutant group: $^{*}P < 0.05$, $^{**}P < 0.001$, $^{***}P < 0.0001$, $^{****}P < 0.00001$. *Wap* expression was reduced by approximately 91% in ΔE3 mutant mice and by over 99.9% in ΔE1a/ΔE2/ΔE3 mutant mice. (**d**) Comparison of *Wap* mRNA levels in mammary tissues from ΔE1a/ΔE2/ΔE3 mutant mice at day 1 of lactation and wild-type controls at different stages (day 1 of lactation and day 14 of pregnancy). Results are shown as $\log_{10}$-transformed means (error bars, s.e.m.; $n = 3$ mice per group). *Wap* expression was reduced by 1,000-fold in ΔE1a/ΔE2/ΔE3 mutant mice in comparison to wild-type mice at day 1 of lactation but was equivalent to that of wild-type mice at day 14 of pregnancy.

STAT5-binding sites, *Wap* expression at day 1 of lactation was reduced to ~0.1% in comparison to wild-type control (**Fig. 4c**). To determine whether this unprecedented enhancer activity mirrored the induction of *Wap* at mid pregnancy, the stage when E2 and E3 are bound by STAT5A and other transcription factors and acquire extensive H3K27ac, we compared *Wap* expression in tissues from ΔE1a/ΔE2/ΔE3 and control mice. In the triple mutant, *Wap* levels at day 1 of lactation were equivalent to those found for wild-type mice at day 14 of pregnancy (**Fig. 4d**), providing evidence that the tripartite STAT5-driven enhancer cluster is responsible for the activation of *Wap* at mid pregnancy.

Although our mutational analyses validated the exceptional role of the tripartite STAT5-driven super-enhancer, the consequences of individual and combined mutations on the binding of other transcription factors and the establishment of H3K27ac peaks remained unknown. We therefore investigated chromatin configuration and transcription factor binding at the *Wap* super-enhancer in mammary tissue derived from mice with the constituent enhancers mutated individually or in combination. The ΔE3 mutation not only led to loss of STAT5A binding but unexpectedly also led to the complete absence of GR binding and DNase I hypersensitivity and to lower H3K27ac levels at E3 (**Fig. 5a**, left). The mammary-specific *Lao1* gene served as a control and was not affected by the mutations (**Fig. 5a**, right). In contrast, some STAT5A binding was retained upon deletion of the GAS site in E1 (ΔE1a), and GR and MED1 binding were largely intact, as was the H3K27ac profile

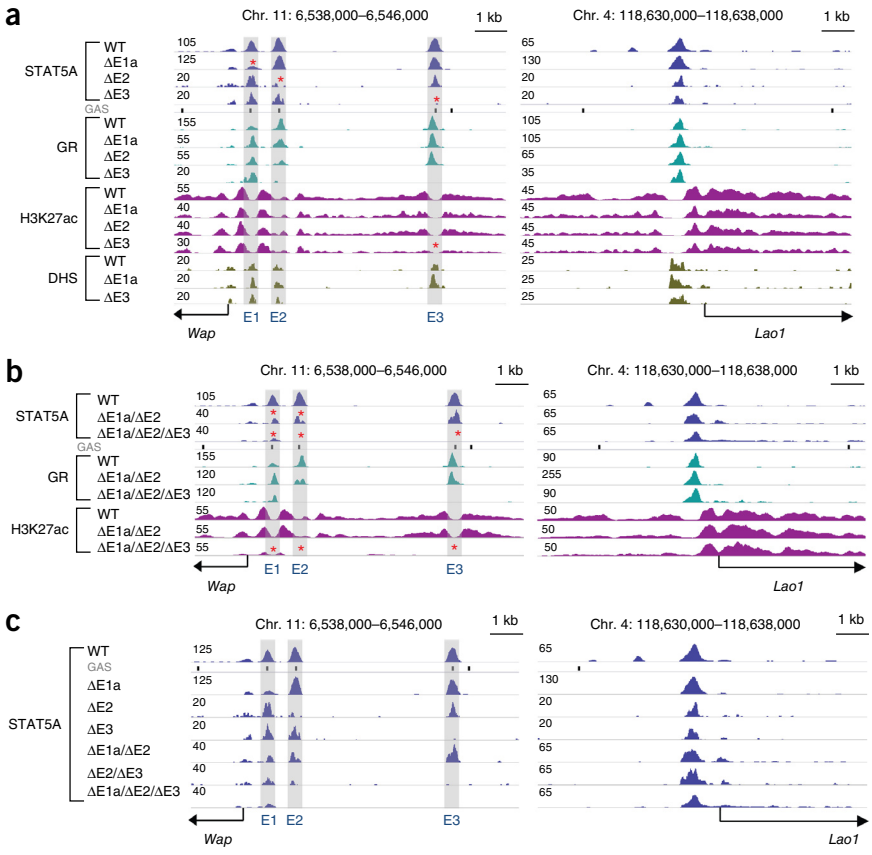


**Figure 5** Consequences of individual and combined loss of STAT5-binding sites at enhancers in the mammary-specific *Wap* super-enhancer. (**a**) ChIP-seq profiles and DNase I–hypersensitive sites (DHSs) at the *Wap* locus in mammary tissue from single-enhancer mutants at day 1 of lactation. The *Lao1* locus served as a ChIP-seq control. The data for STAT5A and H3K27ac ChIP-seq are representative of biological duplicates. (**b**) ChIP-seq profiles in mice with enhancers mutated in combination. The data for STAT5A, GR and H3K27ac ChIP-seq are representative of biological duplicates. STAT5A binding was reduced at E1 and E2, and GR binding and H3K27ac marks were retained in mammary tissue from ΔE1a/ΔE2 mutants. STAT5A and GR binding and H3K27ac marks were completely absent at E2 and E3 in mammary tissue from ΔE1a/ΔE2/ΔE3 mutants, although residual signal was retained at E1. (**c**) STAT5A ChIP-seq profiles in mice with enhancers mutated individually or in combination. STAT5A binding at E3 has the most prominent role in regulation of the *Wap* super-enhancer.

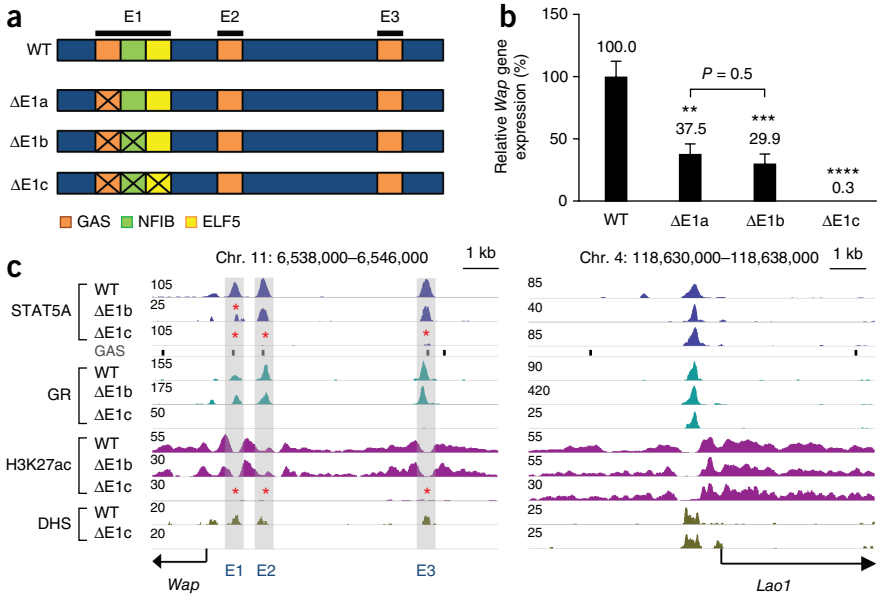


**Figure 6** *In vivo* function of the E1 epicenter in the *Wap* super-enhancer. (**a**) Diagram of mutations in E1 inactivating binding of STAT5 alone, STAT5 and NFIB, and STAT5, NFIB and ELF5. (**b**) *Wap* mRNA levels in mammary tissues from ΔE1a, ΔE1b and ΔE1c mutant mice at day 1 of lactation. *Wap* mRNA levels were measured by qRT–PCR and normalized to *Gapdh* levels. Results are shown as the means ± s.e.m. of independent biological replicates (WT, $n = 9$; ΔE1a, $n = 7$; ΔE1b, $n = 7$; ΔE1c, $n = 5$). A two-tailed unpaired $t$ test was used to evaluate the statistical significance of differences between wild-type mice and each mutant group: $^{**}P < 0.001$, $^{***}P < 0.0001$, $^{****}P < 0.00001$. *Wap* expression levels in ΔE1a and ΔE1b were not significantly different ($P = 0.5$). *Wap* expression was reduced by 99.7% in ΔE1c mutant mice. (**c**) Genomic features of ΔE1b and ΔE1c mutant mice. The data for STAT5A and H3K27ac ChIP-seq are representative of biological duplicates. STAT5A binding and H3K27ac marks were reduced at E1 in mammary tissue from ΔE1b mutants. STAT5A and GR binding, H3K27ac marks and DHSs were absent at the three individual enhancers (E1, E2 and E3) in mammary tissue from ΔE1c mutants. An asterisk indicates enhancer sites that showed prominent reduction in STAT5A and H3K27ac intensity.



(**Fig. 5a**), suggesting that this enhancer had retained limited integrity. Similarly, the chromatin landscape of E2, including H3K27ac, seemed largely intact upon deletion of the underlying STAT5-binding site (**Fig. 5a**). The retention of the epigenetic profiles of E1 and E2 in the respective mutants is in agreement with the relatively modest contribution of these enhancers to *Wap* expression. In contrast, E3 is exceptional in that loss of STAT5 binding precluded the establishment of active enhancer features at this site, in agreement with the more than 90% reduction in *Wap* activity observed in the ΔE3 mutant. We propose that STAT5 is a key factor in the establishment of E3, whereas E1 and E2 are established, at least partially, in the absence of STAT5. The rather modest functional consequences of the ΔE1a and ΔE2 mutations are mirrored in the retention of GR binding and H3K27ac marks in mutants (**Fig. 5b**). STAT5A ChIP-seq experiments indicated that loss of individual STAT5-binding sites had limited consequences for the remaining intact constituent enhancers (**Fig. 5c**). However, combined loss of the dominant E3 enhancer and E2 seemed to affect occupancy of the proximal E1 element. Most notably, loss of all three STAT5-binding sites largely abolished the establishment of active chromatin over the whole super-enhancer, particularly at E2 and E3 (**Fig. 5b**), in agreement with the complete loss of *Wap* activation during pregnancy in the corresponding mutant mice (**Fig. 4c**). The residual binding of GR in the E1 region was insufficient to convey any meaningful activation.

### A seed enhancer in the *Wap* super-enhancer

The importance of STAT5 at E1 is limited and loss of the STAT5-binding site in this enhancer did not affect the overall enhancer profile, suggesting that establishment of the E1 enhancer might require the presence of additional transcription factors, possibly NFIB and ELF5, which bind close to STAT5 (refs. 14,15,53,54) (**Fig. 6a** and **Supplementary Fig. 10**). CRISPR/Cas9 genome editing was used to generate mice lacking the GAS site and the juxtaposed NFIB motif (ΔE1b), and homologous recombination in embryonic stem cells was used to introduce point mutations into the GAS site and the NFIB and ELF5 motifs (ΔE1c) (**Fig. 6a**). Combined inactivation of the GAS site and the NFIB motif resulted in a reduction in *Wap* mRNA levels of approximately 70% (**Fig. 6b**), suggesting that NFIB is not essential for the full activity of the E1 enhancer. Although studies using a *Wap* transgene containing 800 bp of the 5′ flanking sequence, and thus lacking E2 and E3 (refs. 48,55), concluded that NFIB was essential for gene activity[14,15], this was not the case for the endogenous gene. ChIP-seq experiments confirmed loss of NFIB binding to mutant E1 and also demonstrated that ELF5 binding was unaffected (**Supplementary Fig. 11**).

The combined roles of STAT5, NFIB and ELF5 were investigated in mice carrying point mutations at all three sites (ΔE1c) (**Fig. 6a** and **Supplementary Fig. 11**). Unexpectedly, mutations in these three sites incapacitated the entire *Wap* locus (**Fig. 6b**). Moreover, E2 and E3 failed to be established during pregnancy, as demonstrated by the absence of transcription factor binding and H3K27ac marks (**Fig. 6c**). The complete absence of DNase I hypersensitivity suggests that the super-enhancer had not undergone any priming and was unable to respond to pregnancy hormones. Collectively, these data indicate that the joint binding of three transcription factors at E1 provides an epicenter required for the activation of this mammary-specific locus and the recruitment of additional enhancers in response to hormonal stimuli during pregnancy. The super-enhancer that underlies the 1,000-fold induction in *Wap* expression during pregnancy is built on STAT5 binding to three individual enhancers with distinct properties. Notably, the *Nfib* and *Elf5* genes also seem to be under the control of STAT5-driven enhancers (**Supplementary Fig. 12**).

## DISCUSSION

Although super-enhancers have been identified in diverse cell types[26–46,49,56–58], there has been scant genetic proof of their biological relevance *in vivo* (**Supplementary Table 2**). Similarly, the individual and combined contributions of constituent enhancers within super-enhancers have not been investigated through genetic dissection in mice[24,25,56]. Our study now provides compelling genetic evidence that a unique mammary-specific hormone-regulated super-enhancer controls expression of the *Wap* gene during pregnancy and that the embedded constituent enhancers making up this super-enhancer

provide distinct contributions to this regulation. In the case studied, the most distal enhancer seems to do the 'heavy lifting', whereas the proximal ones individually have more modest activity. Thus, our mouse study supports the notion of a functional hierarchy within the STAT5-driven *Wap* super-enhancer. Notably, however, all three STAT5-binding sites were required to achieve the 1,000-fold induction in *Wap* expression during pregnancy. Mutational studies of the constituent enhancers in an erythroid super-enhancer in a cell line also demonstrated that the most distal enhancer was the most potent[34]. Additional genetic studies are needed to determine whether this is a general strategy used by other cell-type-specific super-enhancers *in vivo*.

The establishment of a seed enhancer in the *Wap* super-enhancer during pregnancy depends on three mammary-enriched transcription factors, and the presence of this seed enhancer is required to launch the two accessory enhancers, leading to the full induction of gene expression. Although our data demonstrate that the formation of E2 and E3 depends on E1, suggesting that the *Wap* enhancer cluster could be viewed as a unique regulatory ensemble, we would argue that defining super-enhancers solely on the basis of ChIP-seq experiments and specific algorithms to identify peak patterns does not necessarily demonstrate that they are all special biological entities. Despite all super-enhancers having equivalent patterns of transcription factor binding, only half were associated with the induction of genes highly expressed during pregnancy by STAT5, the principal transcription factor used to define mammary-specific super-enhancers. Moreover, the range over which gene expression was induced by these super-enhancers covered four orders of magnitude, and there seemed to be no obvious correlation between transcription factor occupancy and induction of gene expression. Although increased and progressive occupancy of super-enhancers was observed during pregnancy, this did not necessarily parallel induction of expression for the associated genes. Careful examination and functional testing of other super-enhancers is necessary to draw any general conclusions regarding their biological relevance.

Although STAT family members are key in controlling cell lineages, likely through the cell-type-restricted activity of enhancers[10,56,59–62], studies with knockout mice have only provided limited insight into the contributions of individual STAT isoforms to specific super-enhancers and their embedded constituent enhancers. Loss of either *Stat3* or *Stat5a* and *Stat5b* from the mouse genome results in the absence of specific T cell populations[63–66] and mammary epithelium[6], making it difficult to assess the contributions of these transcription factors to lineage-specific super-enhancers. Using genome editing to target individual enhancers avoids the systemic pitfalls encountered in mice lacking transcription factors, either in the germ line or specific cell types. Such an approach allows for a functional appraisal of predicted enhancer structures in the context of an otherwise uncompromised organism[67]. This is particularly relevant for T cells and mammary epithelium, whose lineages fail to develop in the absence of specific cytokine- or hormone-sensing transcription factors. In summary, our study provides evidence that a genome-wide survey of putative super-enhancers, coupled with targeted genomic editing, can provide mechanistic insight into complex organ-specific and cytokine-regulated gene regulatory modules.

**URLs.** MIT CRISPR Design tool, http://crispr.mit.edu/; R Project for Statistical Computing, https://www.R-project.org/; dplyr, https://CRAN.R-project.org/package=dplyr; Power Analysis, https://CRAN.R-project.org/package=pwr.

## METHODS

Methods and any associated references are available in the online version of the paper.

**Accession codes.** All ChIP-seq and DNase-seq data sets have been deposited in the Gene Expression Omnibus (GEO) under accession GSE74826. Liver and T cell data are available under accessions GSE31578, GSE27158 and GSE31039. RNA-seq data from mammary tissue are available under accessions GSE70440 and GSE37646. The T cell and liver RNA-seq data are available under accessions GSE48138 and GSE66140.

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

### ACKNOWLEDGMENTS

We thank H. Smith from the NIDDK genomics core for never-ending help with next-generation sequencing and C. Liu from the NHLBI transgenic core for generating the CRISPR/Cas9-based mouse mutants. M.W. is a graduate student of the Individual Graduate Partnership Program (GPP) between NIH/NIDDK and the Medical University of Innsbruck. This work was performed in partial fulfillment of the graduation requirements for M.W. We thank K. Kang and S. Oh for discussions in the early stage of this project and Z. Trajanoski for advising M.W. during her graduate studies. This research was funded by the IPR of the NIDDK/NIH.

### AUTHOR CONTRIBUTIONS

H.Y.S. designed, executed and supervised genotyping, identified and validated all CRISPR/Cas9-based mutant founders, performed ChIP-seq, gene expression experiments and histological analysis, analyzed data and wrote the manuscript. M.W. designed experiments, analyzed ChIP-seq and RNA-seq data, performed computational and statistical analyses, and wrote the manuscript. K.H.Y. designed and conducted ChIP-seq, performed gene expression experiments and histological analysis, and analyzed data. X.Z. conducted genotyping, and performed and analyzed gene expression experiments for mutant mice. C.W. performed DNase-seq and ChIP-seq experiments and analyzed data. G.M. genotyped mutant mice, performed and analyzed gene expression experiments, and identified founder mice carrying mutations in combination. L.H. conceived and supervised the study, analyzed data and wrote the manuscript. H.Y.S., M.W. and L.H. wrote and finalized the manuscript, and all authors reviewed and approved the submitted version.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.



1. Hennighausen, L. & Robinson, G.W. Information networks in the mammary gland. *Nat. Rev. Mol. Cell Biol.* **6**, 715–725 (2005).
2. Lydon, J.P. *et al.* Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities. *Genes Dev.* **9**, 2266–2278 (1995).
3. Ormandy, C.J. *et al.* Null mutation of the prolactin receptor gene produces multiple reproductive defects in the mouse. *Genes Dev.* **11**, 167–178 (1997).
4. Horseman, N.D. *et al.* Defective mammopoiesis, but normal hematopoiesis, in mice with a targeted disruption of the prolactin gene. *EMBO J.* **16**, 6926–6935 (1997).
5. Liu, X. *et al.* Stat5a is mandatory for adult mammary gland development and lactogenesis. *Genes Dev.* **11**, 179–186 (1997).
6. Cui, Y. *et al.* Inactivation of *Stat5* in mouse mammary epithelium during pregnancy reveals distinct functions in cell proliferation, survival, and differentiation. *Mol. Cell. Biol.* **24**, 8037–8047 (2004).
7. Zhou, J. *et al. Elf5* is essential for early embryogenesis and mammary gland development during pregnancy and lactation. *EMBO J.* **24**, 635–644 (2005).
8. Wakao, H., Gouilleux, F. & Groner, B. Mammary gland factor (MGF) is a novel member of the cytokine regulated transcription factor gene family and confers the prolactin response. *EMBO J.* **13**, 2182–2191 (1994).
9. Liu, X., Robinson, G.W., Gouilleux, F., Groner, B. & Hennighausen, L. Cloning and expression of Stat5 and an additional homologue (Stat5b) involved in prolactin signal transduction in mouse mammary tissue. *Proc. Natl. Acad. Sci. USA* **92**, 8831–8835 (1995).
10. Yamaji, D., Kang, K., Robinson, G.W. & Hennighausen, L. Sequential activation of genetic programs in mouse mammary epithelium during pregnancy depends on STAT5A/B concentration. *Nucleic Acids Res.* **41**, 1622–1636 (2013).
11. Hennighausen, L.G. & Sippel, A.E. Characterization and cloning of the mRNAs specific for the lactating mouse mammary gland. *Eur. J. Biochem.* **125**, 131–141 (1982).

12. Richards, D.A., Rodgers, J.R., Supowit, S.C. & Rosen, J.M. Construction and preliminary characterization of the rat casein and α-lactalbumin cDNA clones. *J. Biol. Chem.* **256**, 526–532 (1981).
13. Pittius, C.W., Sankaran, L., Topper, Y.J. & Hennighausen, L. Comparison of the regulation of the whey acidic protein gene with that of a hybrid gene containing the whey acidic protein gene promoter in transgenic mice. *Mol. Endocrinol.* **2**, 1027–1032 (1988).
14. Li, S. & Rosen, J.M. Distal regulatory elements required for rat whey acidic protein gene expression in transgenic mice. *J. Biol. Chem.* **269**, 14235–14243 (1994).
15. Li, S. & Rosen, J.M. Nuclear factor I and mammary gland factor (STAT5) play a critical role in regulating rat whey acidic protein gene expression in transgenic mice. *Mol. Cell. Biol.* **15**, 2063–2070 (1995).
16. McKnight, R.A., Wall, R.J. & Hennighausen, L. Expression of genomic and cDNA transgenes after co-integration in transgenic mice. *Transgenic Res.* **4**, 39–43 (1995).
17. Burdon, T.G., Maitland, K.A., Clark, A.J., Wallace, R. & Watson, C.J. Regulation of the sheep β-lactoglobulin gene by lactogenic hormones is mediated by a transcription factor that binds an interferon-γ activation site–related element. *Mol. Endocrinol.* **8**, 1528–1536 (1994).
18. Greenberg, N.M., Reding, T.V., Duffy, T. & Rosen, J.M. A heterologous hormone response element enhances expression of rat β-casein promoter-driven chloramphenicol acetyltransferase fusion genes in the mammary gland of transgenic mice. *Mol. Endocrinol.* **5**, 1504–1512 (1991).
19. Gordon, K. *et al.* Production of human tissue plasminogen activator in transgenic mouse milk. 1987. *Biotechnology* **24**, 425–428 (1992).
20. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
21. Ong, C.T. & Corces, V.G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* **12**, 283–293 (2011).
22. Ong, C.T. & Corces, V.G. Enhancers: emerging roles in cell fate specification. *EMBO Rep.* **13**, 423–430 (2012).
23. Natoli, G. & Andrau, J.C. Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.* **46**, 1–19 (2012).
24. Heinz, S., Romanoski, C.E., Benner, C. & Glass, C.K. The selection and function of cell type–specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154 (2015).
25. Pott, S. & Lieb, J.D. What are super-enhancers? *Nat. Genet.* **47**, 8–12 (2015).
26. Adam, R.C. *et al.* Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature* **521**, 366–370 (2015).
27. Brown, J.D. *et al.* NF-κB directs dynamic super enhancer formation in inflammation and atherogenesis. *Mol. Cell* **56**, 219–231 (2014).
28. Chapuy, B. *et al.* Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777–790 (2013).
29. Chipumuro, E. *et al.* CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in *MYCN*-driven cancer. *Cell* **159**, 1126–1139 (2014).
30. Fang, Z. *et al.* Transcription factor co-occupied regions in the murine genome constitute T-helper-cell subtype-specific enhancers. *Eur. J. Immunol.* **45**, 3150–3157 (2015).
31. Gosselin, D. *et al.* Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327–1340 (2014).
32. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
33. Hnisz, D. *et al.* Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol. Cell* **58**, 362–370 (2015).
34. Huang, J. *et al.* Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis. *Dev. Cell* **36**, 9–23 (2016).
35. Li, Y. *et al.* CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. *PLoS One* **9**, e114485 (2014).
36. Liu, C.F. & Lefebvre, V. The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. *Nucleic Acids Res.* **43**, 8183–8203 (2015).
37. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
38. Mansour, M.R. *et al.* An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–1377 (2014).
39. Ohba, S., He, X., Hojo, H. & McMahon, A.P. Distinct transcriptional programs underlie Sox9 regulation of the mammalian chondrocyte. *Cell Rep.* **12**, 229–243 (2015).
40. Parker, S.C. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. USA* **110**, 17921–17926 (2013).
41. Pelish, H.E. *et al.* Mediator kinase inhibition further activates super-enhancer-associated genes in AML. *Nature* **526**, 273–276 (2015).
42. Pinz, S., Unser, S. & Rascle, A. Signal transducer and activator of transcription STAT5 is recruited to c-*Myc* super-enhancer. *BMC Mol. Biol.* **17**, 10 (2016).
43. Siersbæk, R. *et al.* Transcription factor cooperativity in early adipogenic hotspots and super-enhancers. *Cell Rep.* **7**, 1443–1455 (2014).
44. Thakurela, S., Sahu, S.K., Garding, A. & Tiwari, V.K. Dynamics and function of distal regulatory elements during neurogenesis and neuroplasticity. *Genome Res.* **25**, 1309–1324 (2015).
45. Vahedi, G. *et al.* Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature* **520**, 558–562 (2015).
46. Whyte, W.A. *et al.* Master transcription factors and Mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
47. Hennighausen, L.G. & Sippel, A.E. Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins. *Nucleic Acids Res.* **10**, 2677–2684 (1982).
48. Burdon, T., Sankaran, L., Wall, R.J., Spencer, M. & Hennighausen, L. Expression of a whey acidic protein transgene during mammary development. Evidence for different mechanisms of regulation during pregnancy and lactation. *J. Biol. Chem.* **266**, 6909–6914 (1991).
49. Siersbæk, R. *et al.* Extensive chromatin remodelling and establishment of transcription factor 'hotspots' during early adipogenesis. *EMBO J.* **30**, 1459–1472 (2011).
50. Robinson, G.W. *et al.* Coregulation of genetic programs by the transcription factors NFIB and STAT5. *Mol. Endocrinol.* **28**, 758–767 (2014).
51. Bayna, E.M. & Rosen, J.M. Tissue-specific, high level expression of the rat whey acidic protein gene in transgenic mice. *Nucleic Acids Res.* **18**, 2977–2985 (1990).
52. Miyoshi, K. *et al.* Signal transducer and activator of transcription (Stat) 5 controls the proliferation and differentiation of mammary alveolar epithelium. *J. Cell Biol.* **155**, 531–542 (2001).
53. Li, S. & Rosen, J.M. Glucocorticoid regulation of rat whey acidic protein gene expression involves hormone-induced alterations of chromatin structure in the distal promoter region. *Mol. Endocrinol.* **8**, 1328–1335 (1994).
54. McKnight, R.A. *et al.* An Ets site in the whey acidic protein gene promoter mediates transcriptional activation in the mammary gland of pregnant mice but is dispensable during lactation. *Mol. Endocrinol.* **9**, 717–724 (1995).
55. McKnight, R.A., Spencer, M., Wall, R.J. & Hennighausen, L. Severe position effects imposed on a 1 kb mouse whey acidic protein gene promoter are overcome by heterologous matrix attachment regions. *Mol. Reprod. Dev.* **44**, 179–184 (1996).
56. Witte, S., O'Shea, J.J. & Vahedi, G. Super-enhancers: asset management in immune cell genomes. *Trends Immunol.* **36**, 519–526 (2015).
57. González, A.J., Setty, M. & Leslie, C.S. Early enhancer establishment and regulatory locus complexity shape transcriptional programs in hematopoietic differentiation. *Nat. Genet.* **47**, 1249–1259 (2015).
58. Zhou, H. *et al.* Epstein–Barr virus oncoprotein super-enhancers control B cell growth. *Cell Host Microbe* **17**, 205–216 (2015).
59. Vahedi, G. *et al.* STATs shape the active enhancer landscape of T cell populations. *Cell* **151**, 981–993 (2012).
60. Yang, X.P. *et al.* Opposing regulation of the locus encoding IL-17 through direct, reciprocal actions of STAT3 and STAT5. *Nat. Immunol.* **12**, 247–254 (2011).
61. Kang, K., Yamaji, D., Yoo, K.H., Robinson, G.W. & Hennighausen, L. Mammary-specific gene activation is defined by progressive recruitment of STAT5 during pregnancy and the establishment of H3K4me3 marks. *Mol. Cell. Biol.* **34**, 464–473 (2014).
62. Li, P., Spolski, R., Liao, W. & Leonard, W.J. Complex interactions of transcription factors in mediating cytokine biology in T cells. *Immunol. Rev.* **261**, 141–156 (2014).
63. Yao, Z. *et al.* Stat5a/b are essential for normal lymphoid development and differentiation. *Proc. Natl. Acad. Sci. USA* **103**, 1000–1005 (2006).
64. Yao, Z. *et al.* Nonredundant roles for Stat5a/b in directly regulating Foxp3. *Blood* **109**, 4368–4375 (2007).
65. Laurence, A. *et al.* Interleukin-2 signaling via STAT5 constrains T helper 17 cell generation. *Immunity* **26**, 371–381 (2007).
66. Wei, L., Laurence, A., Elias, K.M. & O'Shea, J.J. IL-21 is produced by Th17 cells and drives IL-17 production in a STAT3-dependent manner. *J. Biol. Chem.* **282**, 34605–34610 (2007).
67. Metser, G. *et al.* An autoregulatory enhancer controls mammary-specific STAT5 functions. *Nucleic Acids Res.* **44**, 1052–1063 (2016).

## ONLINE METHODS

**Mice.** Six- to eight-week-old female C57BL/6 mice were purchased from Charles River and used as wild-type controls. CRISPR/Cas9-targeted mice were generated by the transgenic core of the National Heart, Lung, and Blood Institute (NHLBI), and TALEN-targeted mice were generated by Cyagen Biosciences. Mice carrying point mutations in the E1 region were generated by Ingenious Targeting Laboratory. All animal procedures were in accordance with National Institutes of Health, National Institute of Diabetes, Digestive and Kidney Diseases guidelines for the care and use of laboratory animals.

**ChIP-seq.** Frozen-stored mammary tissues collected at days 13, 14 and 16 of pregnancy and at day 1 of lactation were ground into powder with a mortar and pestle. Chromatin was fixed with 1% formaldehyde at room temperature for 10 min, and fixation was quenched with glycine at a final concentration of 0.125 M. Samples were processed as previously described[67]. The following antibodies were used for ChIP-seq: STAT5A (Santa Cruz Biotechnology, sc-1081), GR (Thermo Scientific, PA1-511A), NFIB (Santa Cruz Biotechnology, sc-5567), ELF5 (Santa Cruz Biotechnology, sc-9645), MED1 (Bethyl Laboratory, A300-793A), H3K27ac (Abcam, ab4729), H3K4me3 (Millipore, 17-614) and RNA polymerase II (Abcam, ab5408). Libraries for next-generation sequencing were prepared as previously described[67] and sequenced with a HiSeq 2000 instrument (Illumina).

**ChIP-seq data analysis.** ChIP-seq signals were trimmed using trimmomatic[68] (version 0.33) to filter out low-quality reads (using the following parameters: LEADING: 20, TRAILING: 20, SLIDINGWINDOW: 4:20, MINLEN: 20, HEADCROP: 15). The subsequent ChIP-seq reads were aligned to the mouse reference genome (mm9) using the Bowtie[69] aligner (version 1.1.2) with the -m 1 parameter to obtain only uniquely mapped reads. The average alignment rate was 93%, and around 24% were discarded because of the -m parameter (**Supplementary Table 3**). The correlation of all replicates was calculated using deepTools[70] with default parameters and Spearman correlation, as it is more stable if outliers occur, with the caveat that the correlation value is less sensitive. ChIP-seq data that are shown in this study were highly reproducible (Spearman's correlation coefficient > 0.7). HOMER software[71] (default settings) was used for visualization. To identify regions of ChIP-seq enrichment over background, MACS2 (ref. 72) peak finding algorithms (version 2.1.0) were used. As the data were from different resources, the *q* value/*P* value parameter was adjusted individually for each file to optimize STAT5 peak calling (STAT5A L1 replicate 1 *q*-value cutoff = $1 \times 10^{-5}$, STAT5A L1 replicate 2 *P*-value cutoff = $1 \times 10^{-2}$, STAT5A p13 replicate 1 *q*-value cutoff = $1 \times 10^{-2}$, STAT5A p13 replicate 2 *q*-value cutoff = $1 \times 10^{-3}$, STAT5A liver *q*-value cutoff = $1 \times 10^{-3}$, STAT5B T cell *q*-value cutoff = $5 \times 10^{-2}$). For all STAT5A L1 and p13 samples, adequate input files were also used. To obtain high-confident peaks, replicates were used for STAT5A L1 and p13, which were identified by overlapping the files using the bedOps[73] genome analysis toolkit (version 2.4.14) with at least 1 bp required for overlap. For H3K27ac in wild-type L1 tissue, the broad peak calling option with a *q*-value cutoff of $5 \times 10^{-2}$ was selected. Only STAT5A L1 peaks, validated using both replicates, which coincide with H3K27ac marks within ±500 bp were taken into account for further analysis. The final step comprised the extraction of promoter peaks defined by location within 500 bp of the transcription start site. The verified and filtered peaks were used for all subsequent analyses.

**Identification of super-enhancers.** The non-promoter STAT5 peaks served as the basis for super-enhancer analysis calculation applying the ROSE algorithm[37,46] with the default stitching size of 12.5 kb, as well as sizes of 25 kb and 35 kb because mammary-specific super-enhancers might be different from super-enhancers identified in embryonic stem cells. To obtain only mammary-specific super-enhancers, each size for stitched enhancers was used as a parameter for the calculations based on STAT5A, H3K27ac, GR and MED1 peaks. The next step comprised the overlap of H3K27ac, GR and MED1 super-enhancers for each size. The super-enhancers identified and reproduced for at least two factors were maintained for further analysis. The subsequent subtraction of liver and T cell super-enhancers (12.5 kb) was applied using bedOps with the requirement that a minimum of 30% of the mammary super-enhancer needed to overlap with the liver or T cell super-enhancers. In the final step, nested super-enhancers that had been annotated to the same gene were removed.

**RNA-seq data analysis and gene annotation.** RNA-seq data were trimmed in the same manner as the ChIP-seq data. Mapping was carried out using the STAR RNA-seq aligner[74] with default settings and Mus_musculus. NCBIM37.67 as a GTF file. To assign the super-enhancers and lone enhancers only to high-confidence genes, the GTF file was filtered to retain only protein-coding genes, and predicted genes (LOC, Rik and BC) were excluded. R (version 3.2.3), Bioconductor[75], and the packages Rsubread[76] (default settings) and DESeq2 (ref. 77; default settings) were used for RNA-seq analysis. Data from mammary tissue at days 6 and 13 of pregnancy and day 1 of lactation and from STAT5A-deficient tissue at day 1 of lactation as well as liver and T cell data were analyzed. Mammary-specific super-enhancers were annotated taking into account the two nearest genes and choosing the one with the higher FPKM value based on expression data obtained at day 1 of lactation.

**Additional bioinformatic analysis.** HOMER[71] was used for motif analysis using the default settings with an individually generated background, based on concatenated DNase-seq data from B cells, cerebellum, kidney, liver, lung, spleen and thymus. The heat map was created using a width of 5 kb. The analysis of enhancers specific for day 1 of lactation and enhancers already established at day 13 of pregnancy was carried out using the bedOps[73] genome analysis toolkit, with the requirement that the peaks be inside the super-enhancer. Graph plotting was performed using R with the packages dplyr and ggplot2 (ref. 78).

**Data.** STAT5 ChIP-seq data for liver and T cells were obtained from the Gene Expression Omnibus (GEO) under accessions GSE31578 and GSE27158, respectively. H3K27ac ChIP-seq data for liver and T cells were obtained from GEO under accession GSE31039. The RNA-seq data for wild-type controls at day 6 of pregnancy and at day 1 of lactation and for *Stat5a*-deficient mice at day 1 of lactation have been deposited under accession GSE37646. RNA-seq data for wild-type controls at day 13 of pregnancy were obtained from GSE70440. T cell and liver RNA-seq data are available under accessions GSE48138 and GSE66140.

**Generation of CRISPR/Cas9-targeted mice.** Single-guide RNA (sgRNA) constructs were designed to specifically target the GAS at E1 (0.7 kb upstream of the *Wap* TSS), E2 (1.4 kb upstream of the *Wap* TSS) and E3 (5.6 kb upstream of the *Wap* TSS). Off-target scores were evaluated by the online tool. The specific sgRNA sequences are described in **Supplementary Table 4**. Target-specific sgRNA and *Cas9* mRNA were *in vitro* transcribed and microinjected into the cytoplasm of fertilized eggs for founder mouse production.

**Generation of TALEN-targeted mice.** The GAS element (1.4 kb upstream of the *Wap* TSS or 5.6 kb upstream of the *Wap* TSS) was selected as a TALEN target site. TALEN mRNA was generated by *in vitro* transcription and then injected into fertilized eggs for founder mouse production.

**Generation of knock-in mice.** Point mutations were introduced in the GAS element, NFIB motif and ELF5 motif in E1 (0.7 kb upstream of the *Wap* TSS) using homologous recombination in embryonic stem cells.

**Generation of homozygous mice and genotyping.** Founder ($F_0$) CRISPR/Cas9-targeted mice and TALEN-targeted mice were bred to C57BL/6 wild-type mice to segregate the mosaicism and generate $F_1$ heterozygous mice, which were interbred to generate $F_2$ homozygous mice. Founders of knock-in mice were interbred for mouse colony expansion. All mice were genotyped by PCR amplification of genomic DNA isolated from mouse tail snips followed by Sanger sequencing. Details of the PCR primers and sequencing primers are provided in **Supplementary Table 5**. The two lines carrying deletions in the GAS at E1 were named ΔE1a (11-bp deletion) and ΔE1b (27-kb deletion). The knock-in mice were named ΔE1c. TALEN-targeted mice were named ΔE2 and ΔE3 according to the location of the specific GAS deletion sites. The mice that had combined mutations at GAS sites in E1, E2 and E3 introduced by CRISPR/Cas9 genome editing were named ΔE1a/ΔE2, ΔE2/ΔE3 and ΔE1a/ΔE2/ΔE3




doi:10.1038/ng.3606

according to the location of specific GAS deletion sites. The specific deletion sequences in each mouse are shown in **Supplementary Figure 9**.

**RNA isolation and qRT–PCR.** RNA was isolated using the PureLink RNA Mini kit (Ambion) according to the manufacturer's protocol. cDNA was synthesized from total RNA using SuperScript II (Invitrogen), and qPCR was performed using the TaqMan probe-based system (*Wap*, Mm00839913_m1; mouse *Gapdh* (used as an endogenous control), 4352339E, Applied Biosystems) on the CFX384 Real-Time PCR Detection System (Bio-Rad). *Wap* mRNA levels were measured by qRT–PCR and normalized to *Gapdh* levels.

**DNase-seq.** DNase-seq in the mammary glands of wild-type and mutant mice was performed as previously described[67].

**Statistical analyses.** All samples that were used for qRT–PCR, ChIP-seq and DNase-seq were randomly selected, and blinding was not applied. Significance for box plots was determined using a Student's *t* test with a one-sided alternative hypothesis in case of a two-sample comparison. For comparison of more samples, ANOVA was applied to determine whether a difference was significant. In the case of significance, an additional pairwise *t* test was used to compare the groups. Before any statistical analysis, a Shapiro–Wilk normality test was applied to validate the normal distribution of the data. Statistical power was calculated using R and the package pwr. An effect size was calculated for each comparison on the basis of the estimated mean and s.d. of each group, and the sample size for each group was therefore determined via power analysis of change in *Wap* expression with significance level of 0.05 and power level of 0.9. Gene expression data are presented as the mean of independent biological replicates. To evaluate whether gene expression was statistically different between wild type and each mutant group, a two-tailed unpaired *t* test was used.

**Mammary epithelial cell isolation.** Mammary tissue was obtained at day 13 of pregnancy and day 1 of lactation from wild-type mice. Tissue was digested for 2 h at 37 °C in complete EpiCult-B medium (EpiCult-B medium with 5% FBS) supplemented with 300 U/ml collagenase and 100 U/ml hyaluronidase. After lysis of red blood cells in $NH_4Cl$, single-cell suspensions were obtained by sequential dissociation with prewarmed 0.25% trypsin-EDTA followed by prewarmed 5 mg/ml Dispase and 0.1 mg/ml DNase I, and cells were filtered through a 70-μm cell strainer. All reagents were from Stemcell Technologies unless otherwise specified. The single-cell suspensions obtained were isolated to MECs by the removal of non-epithelial cell compartments with biotinylated antibodies against the cell surface antigens of mouse hematopoietic, endothelial and fibroblast cells (CD45, CD31, TER119 and BP-1) using the EasySep Mouse Epithelial Cell Enrichment kit (Stemcell Technologies).

**Histological analysis.** Mammary tissues from wild-type mice were collected at days 13, 14 and 16 of pregnancy and at day 1 of lactation. Isolated mammary tissues were fixed in 10% formalin and dehydrated in ethanol. Paraffin sections were stained with hematoxylin and eosin by standard methods (Histoserve).


68. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
69. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
70. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
71. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
72. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
73. Neph, S. *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
74. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
75. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
76. Liao, Y., Smyth, G.K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
77. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
78. Wikham, H. *Ggplot2: Elegant Graphics for Data Analysis* (2009).