Effect of natural genetic variation on enhancer selection and function

S. Heinz1*, C. E. Romanoski1*, C. Benner1,2,3, K. A. Allison1, M. U. Kaikkonen1,4, L. D. Orozco5 & C. K. Glass1,3,6

The mechanisms by which genetic variation affects transcription regulation and phenotypes at the nucleotide level are incompletely understood. Here we use natural genetic variation as an *in vivo* mutagenesis screen to assess the genome–wide effects of sequence variation on lineage–determining and signal–specific transcription factor binding, epigenomics and transcriptional outcomes in primary macrophages from different mouse strains. We find substantial genetic evidence to support the concept that lineage–determining transcription factors define epigenetic and transcriptional states by selecting enhancer–like regions in the genome in a collaborative fashion and facilitating binding of signal–dependent factors. This hierarchical model of transcription factor function suggests that limited sets of genomic data for lineage–determining transcription factors and informative histone modifications can be used for the prioritization of disease–associated regulatory variants.

Inter-individual genetic variation is a major cause of diversity in phenotypes and disease susceptibility. Although sequence variants in gene promoters and protein–coding regions provide obvious prioritization of disease–causing variants, most (88%) genome–wide association study (GWAS) loci are in non–coding DNA, suggesting regulatory functions. Prioritization of functional intergenic variants remains challenging, owing in part to an incomplete understanding of how regulation is achieved at the nucleotide level in different cell types and environmental contexts. Recent studies have described important roles for lineage–determining transcription factors (LDTFs), also referred to as pioneer factors or master regulators, in selecting cell–type–specific enhancers, but the sequence determinants that guide their binding are poorly understood. Previous findings in macrophages and B cells suggest a hierarchical model of regulatory function, in which a relatively small set of LDTFs collaboratively compete with nucleosomes to bind DNA in a cell–type–specific manner.

The binding of these factors is proposed to ‘prime’ DNA by initiating deposition of histone modifications that are associated with cis–active regulatory regions and enable concurrent or subsequent binding of signal–dependent transcription factors that direct regulated gene expression. Prior findings in macrophages and B cells suggest a hierarchical model of regulatory function, in which a relatively small set of LDTFs collaboratively compete with nucleosomes to bind DNA in a cell–type–specific manner. This analysis consistently identified consensus and degenerate motifs for the LDTFs PU.1, C/EBPα and AP-1 as the most highly enriched PWMs. Notably, the identified consensus and degenerate motifs (Fig. 1F).

**Direct effects of genetic variation**

First, we quantified genome–wide binding patterns of macrophage LDTFs PU.1 and C/EBPα from both mouse strains using chromatin immunoprecipitation followed by massively parallel sequencing (ChIP–Seq). These experiments identified a combined 82,154 PU.1 and 54,874 C/EBPα peaks, with less than 1% of sites exhibiting highly significant strain–specific binding (PU.1, n = 496; C/EBPα, n = 263; fourfold tag count ratio, false discovery rate (FDR) < 1 × 10–14, >90% located >3 kb to gene promoters) (Fig. 1B, C and Extended Data Fig. 1a). Strain–specific binding was defined using biological ChIP–Seq replicates, which yielded <0.2% empirical false positives (Extended Data Fig. 1b–g). Differential binding of PU.1 and C/EBPα was significantly correlated with differential expression of the nearest gene as measured by RNA–Seq (Fig. 1D). There were no apparent differences in genomic context for strain–similar and strain–specific binding at inter– or intragenic sites (>3 kb to promoters) as defined by CpG content, distance from nearest gene or repetitive element, or conservation score (Extended Data Fig. 2a). Instead, strain–specific binding was highly correlated with polymorphism frequency. We observed fivefold enrichment of polymorphisms at strain–specific versus strain–similar PU.1– and C/EBPα–bound regions (Fig. 1E and Extended Data Fig. 2b), with the greatest variant density at the peak centres (Extended Data Fig. 2c, d).

To investigate the direct effects of sequence variants on transcription factor binding, we identified the most enriched position weight matrices (PWMs) in genomic regions marked by histone H3 lysine 4 di–methylation (H3K4me2) or bound by PU.1 or C/EBPα (Extended Data Fig. 3a and Supplementary Table 1). This analysis consistently identified consensus and degenerate motifs for the LDTFs PU.1, C/EBPα and AP–1 as the most highly enriched PWMs. Notably, the frequency of mutations in these motifs increased with strain–specific binding of PU.1 and C/EBPα (Extended Data Fig. 2e, f). Excluding strain–specific loci without *cis–variation* (~11%), 41% of strain–specific PU.1 binding directly associated with strain–specific mutations in PU.1 motifs in the other strain. For C/EBPα, 44% of strain–specific binding associated with strain–specific C/EBPα motifs (Fig. 1F).

*These authors contributed equally to this work.

1Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, Mail Code 0651, La Jolla, California 92093, USA. 2Integrative Genomics and Bioinformatics Core, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. 3San Diego Center for Systems Biology, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92037, USA. 4Department of Biotechnology and Molecular Medicine, A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, PO Box 1627, 70211 Kuopio, Finland. 5Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. 6Department of Biotechnology and Molecular Medicine, A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, PO Box 1627, 70211 Kuopio, Finland. 7Department of Molecular Cell and Developmental Biology, University of California, Los Angeles, 3000 Terasaki Life Sciences Building, Los Angeles, California 90095, USA. 8Department of Medicine, University of California, San Diego, 9500 Gilman Drive, Mail Code 0651, La Jolla, California 92093, USA.
Although strain-specific binding of PU.1 and C/EBPα was highly linked to strain-specific motif mutations, strain-specific motif mutations were also associated with strain-similar binding (Extended Data Fig. 3c, d). This raised the question as to whether specific features of motif mutations could be used to predict strain-specific binding. Comparison of motif mutations in strain-specific and strain-similar peaks revealed three distinct attributes contributing to predictive power. First, mutated motifs within 20 base pairs (bp) of the experimentally defined binding centres were more highly associated with an effect on binding (PU.1, $P = 1.6 \times 10^{-4}$; C/EBPα, $P = 0.036$; Extended Data Fig. 4a–d). Second, the presence of alternative motifs within 100 bp of the PU.1 peak centres significantly buffered the effect of strain-specific PU.1 motifs (Extended Data Fig. 4e, f). Third, after removing peaks with alternative motifs, analysis of the nucleotides mutated enabled delineation of an empirically defined functional motif that revealed a strong relationship between ‘core’ mutations and altered binding (Fig. 1G and Extended Data Fig. 4g–i; $P = 2.6 \times 10^{-10}$). This enabled an algorithm incorporating these characteristics to be used to predict the effect of a specific motif mutation on transcription factor binding (Fig. 1H). Of the 14 million identifiable PU.1 motifs in the C57BL/6J reference genome, 18,322 contain SNPs that mutate the PU.1 motif in the NOD genome. A total of 1.6% of these mutations were associated with strain-specific binding (Fig. 1H). Of the 244 NOD PU.1 motif mutations located in PU.1-bound regions in C57BL/6J or BALB/cJ mice, 68% were associated with strain-specific binding. When considering all three variables (motif distance, alternative motif and motif core; Extended Data Fig. 5), 88% of the predicted functional mutations were consistent with impaired PU.1 binding in NOD (Fig. 1H).

**Figure 1** Genetic variation affects LDTF binding. A, Model in which LDTFs (PU.1 and C/EBPα) establish regulatory function (explained in text). B, C, ChiP-Seq-defined binding intensity for PU.1 (B) and C/EBPα (C) in resting macrophages derived from C57BL/6J (C57; $x$ axes) and BALB/cJ (BALB; $y$ axes) mice. Dots represent normalized tag counts in 200-bp peaks. PU.1, C/EBPα and AP-1 motifs that were mutated in one genome (distinguished by symbol; red denotes C57BL/6J, blue denotes BALB/cJ) are highlighted for peaks with strain-specific binding (fourfold, FDR $\leq 1 \times 10^{-4}$). D, RNA-Seq-determined expression for genes nearest to strain-specific PU.1 or C/EBPα peaks. P values are from one-tailed $t$-test. E, Variant frequency distributions for PU.1 binding ratio bins. Box midlines (D, E) are medians, boundaries are first and third quartiles, and whiskers extend to extremes. F, The percentage of polymorphic, strain-specific PU.1 and C/EBPα motifs with LDTF mutations. G, The observed position of SNPs generating strain-specific PU.1 motifs ($n = 359$) underlying differential (blue) or similar (red) PU.1 binding are shown. H, The proportion of NOD PU.1 motif mutations that abolished PU.1 binding for each group is shown (details in Extended Data Fig. 5). Alt., alternative.

**Variation and collaborative LDTF binding**

To investigate the potential effect of mutations in LDTF recognition motifs on collaborative binding, we analysed all strain-specific PU.1 or C/EBPα binding events in regions containing LDTF motif mutations. PU.1 motif mutations resulting in loss of PU.1 binding were frequently associated with a corresponding loss of nearby C/EBPα binding in the absence of C/EBPα motif mutations (Fig. 2a, top). Conversely, C/EBPα motif mutations resulting in a loss of C/EBPα binding were frequently associated with a corresponding loss of nearby PU.1 binding in the absence of PU.1 motif mutations (Fig. 2a, middle). Similar results were observed at locations containing strain-specific mutations in AP-1 binding motifs, but intact PU.1 and C/EBPα motifs (Fig. 2a, bottom).

We next considered the global relationships of mutations in PU.1, C/EBP and AP-1 motifs with strain-specific binding of PU.1 and C/EBPα, taking into account both consensus and ‘weak’ motifs for PU.1 and C/EBP. NF-κB8 motifs were included as controls that were not expected to affect PU.1 or C/EBPα binding in unstimulated macrophages (Extended Data Fig. 3a, b and Supplementary Table 1). Although mutations in PU.1 motifs had the strongest effect on strain-specific PU.1 binding, mutations exclusively in C/EBP and/or AP-1 motifs also significantly correlated with differential PU.1 binding relative to similarly bound loci (Fig. 2b). Similar relationships were observed for C/EBP (Extended Data Fig. 6a). The motif distance distributions for co-bound factors were broad (half-width $\sim 100$ nucleotides), and only a minor subset of sites exhibited defined distances expected for direct protein–protein
motif mutation (Fig. 1F). To test genetically whether these correlations are consistent with a collaborative binding model, we considered all LDTF motif mutations and evaluated their effects on PU.1 binding in macrophages derived from NOD mice. For polymorphic strain-specific PU.1 loci containing strain-specific LDTF motifs (n = 220), PU.1 binding profiles matched the strain with shared alleles for 91% of cases (Fig. 3a). At 8% (n = 17) of the loci, the NOD genome broke the C57BL/6J–BALB/cJ haplotypes, and in all cases, the NOD genotype at the LDTF motif variant matched the strain with similar binding (Supplementary Table 2), indicating that these variants are probably the cause of binding differences. An example is shown in Fig. 3b, in which PU.1 binds in C57BL/6J but not in BALB/cJ or NOD mice. Only one SNP in this region is associated with PU.1 binding exclusively in C57BL/6j; here, the T allele forms part of a neighbouring AP-1 motif in C57BL/6J that is mutated by the C allele in BALB/cJ and NOD mice. These findings provide genetic evidence that PU.1 binding to this location is dependent on collaborative interactions with AP-1.

To confirm that the allele-specific binding also occurs in heterozygous cells, we performed ChIP-Seq for PU.1 and C/EBPα in macrophages from CB6F1/J hybrid mice, which are F1 offspring of a C57BL/6J–BALB/c hybrid. In the most cases, alleles bound specifically in a parental strain were also bound preferentially in the F1 generation (Fig. 3c and Extended Data Fig. 6c).

Figure 2 | Genetic variation supports the LDTF collaborative binding model. a. Normalized ChIP-Seq signal at 342 loci defined by strain-specific PU.1 and/or C/EBPα binding, and containing LDTF motif mutations (rows) plotted for each factor/modification (columns). Left columns display SNPs as grey dots with mutated motifs highlighted by colour (LDTF mutation labels at left). b. Log2 odds ratios for observing strain-specific motif mutations at strain-specific (>2-fold tag ratio) and similar (≤2-fold tag ratio, middle bin) PU.1 peaks (details in Methods). c. Gene expression for genes nearest promoter-distal (≥3 kb), strain-specific H3K4me2 and H3K27ac peaks defined by strain-specific modification (next to promoter) and similarly (sim.) defined by strain-specific motif (next to promoter). d. Differential binding at sites with strain-similar C/EBPα motifs (rows) to strain-specific PU.1 and C/EBPα binding. Strain-specific motif mutations are indicated by symbol and colour. The distribution of H3K27ac strain ratios stratified by left and right bins (columns) versus log2 (PU.1 specific (≥2-fold tag ratio, left and right bins) and similarly (sim.) defined by strain-specific motif (next to promoter)). e. Enrichment significance (hypergeometric distribution testing, see Methods) of H3K27ac modification in eQTLs from different cell types is shown. Mac, macrophage.
Figure 4 | p65 binding is largely determined by LDTF binding. a, Strain-specific p65-bound regions were segregated into rows according to the bound strain (coloured side bar). Binding/modification is shown with and without 100 ng ml\(^{-1}\) KLA treatment (+/–, third header row). As in Fig. 2a, SNPs are indicated by grey dots and mutated motifs are highlighted by colour (labelled at bottom). b, The log odds ratio for observing strain-specific mutations is shown for bins of p65 binding as described in Fig. 2b. c, The percentage of polymorphic, differentially bound p65 loci containing LDTF or NF-κB motif mutations is shown. d, The ratio of variant counts in strain-specific versus strain-similar peaks (y axis) is shown relative to the peak centres for PU.1, C/EBPα and p65 and p65 peaks in 10-bp bins (x axis), smoothed using cubic spline. e, The relative amount of transcription (GRO-Seq) and mRNA production between strains after KLA treatment at the nearest gene to strain-specific p65 loci is shown. P values are from one-tailed \(t\)-test.

Given the genetic evidence that LDTFs collaborate to bind DNA, we next tested the extent to which strain-specific LDTF binding explained promoter-distal (>3 kb) strain-specific histone modification events, such as H3K4me2 and H3K27ac deposition, with the combined binding of both factors exhibiting even greater correlation than the individual factors (Extended Data Fig. 7a–f). Furthermore, LDTF motif mutations segregated with differential LDTF binding and histone modifications (Fig. 2d and Extended Data Fig. 7g). Together, these findings support the concept that LDTFs have quantitatively important roles in establishing these histone modifications, probably through initiating transcription in a combinatorial fashion\(^{23}\).

Expression quantitative trait loci (eQTLs) are polymorphic loci whose alleles are associated with individual RNA expression levels across a population\(^{22}\). Thus, eQTLs define active gene regulatory loci and provide an alternative method for assigning regulatory function to gene expression. To interrogate the relationship between histone modification and eQTLs, we analysed previously reported eQTL data from 85 inbred mouse strains in the hybrid mouse diversity panel in primary macrophages\(^{23}\) (see Methods). We found that eQTLs overlapped H3K4me2- or H3K27ac-marked regions at frequencies greater than expected by chance, supporting the role of histone modifications as landmarks of regulatory activity (hypergeometric test \(P\) values: H3K4me2 = \(1 \times 10^{-2.147}\), H3K27ac = \(1 \times 10^{-2.290}\). Next, given the highly cell-type-specific nature of gene regulation\(^{24}\), we proposed that eQTLs from different cell types would be reflected in the histone modification profiles in the same cell type. We examined liver and macrophage eQTLs for a set of ~130 k SNPs from the hybrid mouse diversity panel\(^{25}\) for overlap with H3K27ac loci defined in macrophages or in liver, pro-B or mouse embryonic stem cells\(^{20}\). Macrophage eQTLs were more significantly enriched for overlap with macrophage H3K27ac regions than liver H3K27ac regions. Similarly, liver eQTLs were most significantly enriched with liver H3K27ac relative to macrophage H3K27ac (Fig. 2e). Clustering of H3K27ac profiles revealed that liver and embryonic stem-cell H3K27ac profiles are most similar (Extended Data Fig. 7h), providing an explanation as to why liver eQTLs were highly enriched in mouse embryonic stem-cell H3K27ac regions.

**LDTF motif mutations affect NF-κB binding**

To evaluate the prediction that primed regulatory loci (containing H3K4me2) often require additional binding of signal-dependent transcription factors to achieve regulatory activity (Fig. 1A, c–e), we treated C57BL/6 and BALB/c macrophages with Kdo\(_2\)-lipid A (KLA), a potent and specific agonist of TLR4 (ref. 26). KLA treatment causes NF-κB to enter the nucleus, bind DNA and regulate several hundred target genes\(^{26,27}\). We performed ChIP-Seq for PU.1, C/EBPα and the p65 (also known as RelA) component of NF-κB in untreated and KLA-treated macrophages, and observed that 61% of sites that gained p65 were pre-bound by PU.1 and/or C/EBPα without KLA. De novo motif analysis indicated that an AP-1 motif was present in 42% of the remaining sites, suggesting that AP-1 is responsible for priming a large proportion of the p65 cistrome (Extended Data Fig. 8a), in line with previous reports\(^{16}\).

To interrogate the dependence of p65 on LDTFs further we focused on sites that gained p65 only in one strain (\(n = 932\), >90% promoter-distal; Extended Data Fig. 1a, 4a, fourth column). In most cases, PU.1 and/or C/EBPα were bound before KLA treatment only in the strain exhibiting p65 binding (Fig. 4a). In addition, strain-specific p65 binding primarily occurred at loci already marked by H3K4me2, and led to an increase of H3K27ac, consistent with the proposed model. To analyse the effects of genetic variation on transcription factor motifs, we performed strain-specific LDTF and NF-κB motif finding in polymorphic strain-specific p65-bound peaks (\(n = 750\)) (Extended Data Fig. 3b). Notably, p65 binding was influenced by mutations in individual LDTF motifs to a similar extent as mutations in the NF-κB motif itself (Fig. 4b and Extended Data Fig. 8b). For strain-specific p65 binding events, 34% could be attributed to one or more of the LDTF motifs, whereas 9% could be explained by mutations in the NF-κB motif itself (Fig. 4c). RelA is known to bind to degenerate and non-canonical motifs\(^{28}\) that might
not be captured by de novo motif analysis. To gain motif-independent insight into variant location and strain-specific transcription factor binding, we assessed the variant frequency relative to the centres of strain-specific p65 peaks. Similar to strain-specific PU.1 and C/EBPα peaks, strain-specific p65 peaks are in regions of higher variant density than strain-similar peaks (Extended Data Fig. 8c). In contrast to LDTFs, in which strain-specifically bound regions have a high variant density at their peak centres, the distribution of variants at strain-specific p65 peaks is significantly different from those of the LDTFs (Kolmogorov–Smirnov P < 0.013), as it contains fewer variants at the peak centres and is broader (Fig. 4d and Extended Data Fig. 8d–f). This is consistent with p65 binding being more affected by sequence variation in motifs of neighbouring factors than LDTFs.

Overall, strain-specific p65-bound regulatory sites were significantly correlated with nearby genetic transcription and messenger RNA production (Fig. 4e). We tested strain-specifically bound and epigenetically marked putative enhancer sequences with strain-specific mutations for differential enhancer function in transient and stable reporter assays (Fig. 5a, b and Extended Data Fig. 9a, b). We observed the predicted strain-specific enhancer activity for 18 out of 20 of these genomic sequences. Conversely, enhancer elements with sequence variation in non-core nucleotides that were not predicted to alter PU.1 or C/EBP binding and that exhibited strain-similar binding patterns exhibited similar enhancer activity (Extended Data Fig. 10a).

Lastly, we tested whether the predicted motif-disrupting variants could specifically explain strain-specific enhancer activity by swapping variants at the putative causative alleles in C57BL/6J and BALB/cJ while maintaining the genetic background for the remainder of the enhancer sequences. Representative examples in which reversal of such SNPs in PU.1, C/EBP and p65 motifs reversed strain-specific enhancer activity are illustrated in Fig. 5c and Extended Data Fig. 10b, c. By contrast, reversal of nearby SNPs not predicted to alter LDTF motifs had no effect on strain-specific enhancer activity (Extended Data Fig. 10c).

**Discussion**

Together, we have exploited natural genetic variation to test a collaborative model for enhancer selection and function, and conversely explored the ability of this model to explain strain-specific differences in transcription factor binding and epigenetic features associated with functional enhancers in macrophages. These studies provide genetic evidence that LDTFs are dependent on collaborative binding to variably spaced DNA recognition motifs to select enhancers and enable binding of signal-dependent transcription factors. Notably, the variable motif distances observed at loci co-bound by LDTF suggest that collaborative binding does not generally require direct protein–protein interactions between the involved transcription factors. The proposed hierarchical LDTF collaborative model provides a conceptual framework for prioritization of non-coding disease-associated regulatory variants. Although all cells express hundreds of transcription factors, a large fraction of functional enhancers (~70% in macrophages) are characterized by collaborative interactions involving relatively small sets of lineage-determining transcription factors (for example, PU.1, AP-1 and C/EBPs). The requirement for collaborative binding interactions provides an explanation for why transcription factor binding is lost at sites where mutations do not occur in the cognate recognition motif. In the case of NF-kB, for example, mutations in the motifs for LDTFs were approximately three times more likely to result in decreased binding of NF-kB than mutations in the NF-kB-binding site itself.

An essential step in leveraging the collaborative model to pinpoint potential disease-causing variants is the definition of relevant LDTF-binding sites and functionally important variants. At the current level of genome annotation, this cannot be achieved by analysis of DNA sequence alone. For example, there are ~1 × 10^6–2 × 10^6 identifiable PU.1 motifs in the human29 and mouse genomes, but less than 10% are actually occupied by PU.1 in macrophages. By experimentally defining strain-similar and strain-specific binding patterns for PU.1, the relevant sites at which mutations can result in altered function are identified. Comparison of PU.1 motif mutations associated with strain-specific versus strain-similar binding allowed the genetic definition of a functional binding matrix and additional distinguishing features that enabled accurate prediction of functional mutations in a third strain. Thus, by collecting a relatively limited set of genomic binding data for LDTFs and informative histone modifications, this analytical approach can be exploited to explain a greater extent of variation in enhancer selection and function than previously possible27,28. To increase the specificity and sensitivity for detecting functional variations further, identification of transcription factor motifs that permit binding but diverge from the consensus PWM, that is, ‘weak’ motifs, needs to be improved, as such sites are more likely to be affected by mutation29,30. In addition, transcription factors less abundant than LDTFs probably have individually small but collectively significant roles. At a larger scale, non-cis-acting, long-range epigenetic mechanisms may also be important for enhancer selection. A major goal for the future will be to extend these approaches to understanding natural genetic variation associated with human disease.

**METHODS SUMMARY**

**Cell culture.** Peritoneal macrophages from male 6–8-week-old mice were thiglycollate-elicited and collected 4 days after injection, plated overnight, and incubated for 1 h in fresh media with or without 100 ng ml^-1^ KLA.

**ChIP-Seq and feature identification.** ChIP-Seq was performed based on published protocols31,32, on either native chromat in an MNase digestion (H3K4me2) or fixed, sonicated chromatin (PU.1, C/EBPs, H3K27ac and p65). ChIP-Seq libraries were sequenced for 51 cycles on a HiSeq 2000 sequencer (Illumina). Reads from C57BL/6J and BALB/cJ were mapped with low stringency to both the mm9 reference (C57BL/6J) genome and the BALB/cJ contigs3, and the 98% of all reads that mapped to both genomes was kept for further analysis. NOD data were mapped to the mm9 reference with low stringency. For C6B6F1/J ChIP-Seq experiments,
14. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the

12. Garber, M.

10. Reddy, T. E.

9. McDaniell, R.

6. Heinz, S.

et al.

transfections) or total protein content (stable transfections).

macrophages after 16 h stimulation with or without 100 ng ml

promoter containing luciferase reporter plasmid were assayed in RAW264.7

Genomic

Reporter assays.

492 | NATURE | VOL 503 | 28 NOVEMBER 2013

4. Gaffney, D. J.

et al.

3. Degner, J. F.

2. Cowper-Sallari, R.

1. Hindorff, L. A.

Data are available in the online version of the paper; references unique to these

Online Content Any additional Methods, Extended Data display items and Source

Data are available in the online version of the paper; references unique to these

sections appear only in the online paper.

Received 29 January; accepted 29 August 2013.

Published online 13 October 2013.

1. Hindorff, L. A. et al. Potential etiologic and functional implications of genome-wide

association loci for human diseases and traits. Proc. Natl Acad. Sci. USA 106,

9362–9367 (2009).

2. Cowper-Sallari, R. et al. Breast cancer risk-associated SNPs modulate the affinity

of chromatin for FOXA1 and alter gene expression. Nature Genet. 44, 1191–1198

(2012).

3. Degner, J. F. et al. DNase I sensitivity QTLs are a major determinant of human

expression variation. Nature 482, 390–394 (2012).

4. Gaffney, D. J. et al. Dissecting the regulatory architecture of gene expression QTLs.

Genome Biol. 13, R7 (2012).

5. Gautron, K. J. et al. A map of open chromatin in human pancreatic islets. Nature

Genet. 42, 255–259 (2010).

6. 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).

7. Kasowski, M. et al. Variation in transcription factor binding among humans.

Science 328, 232–235 (2010).

8. Mauroso, M. T., Wang, H., Kutyavin, T. & Stamatoyannopoulos, J. A. Widespread

site-dependent buffering of human regulatory polymorphism. PLoS Genet. 8, e1002599 (2012).

9. McDaniel, R. et al. Heritable individual-specific and allele-specific chromatin

signatures in humans. Science 328, 235–239 (2010).

10. Reddy, T. E. et al. Effects of sequence variation on differential allelic transcription

factor occupancy and gene expression. Genome Res. 22, 860–869 (2012).

11. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease

associations with regulatory information in the human genome. Genome Res. 22,

1745–1759 (2012).

12. Garber, M. et al. A high-throughput chromatin immunoprecipitation approach

reveals principles of dynamic gene regulation in mammals. Mol. Cell 47, 810–822

(2012).

13. Mullen, A. C. et al. Master transcription factors determine cell-type-specific

responses to TGF-b signaling. Cell 147, 567–576 (2011).

14. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the

pluripotency reprogramming factors’ initial engagement with the genome. Cell

151, 994–1004 (2012).

15. Trompouki, E. et al. Lineage regulators direct BMP and Wnt pathways to cell-

specific programs during differentiation and regeneration. Cell 147, 577–589

(2011).

16. Ghisletti, S. et al. Identification and characterization of enhancers controlling the

inflammatory gene expression program in macrophages. Immunity 32, 317–328

(2010).

17. Keane, T. M. et al. Mouse genomic variation and its effect on phenotypes and gene

regulation. Nature 477, 289–294 (2011).

18. Mirny, L. A. Nucleosome-mediated cooperativity between transcription factors.

Proc. Natl Acad. Sci. USA 107, 22534–22539 (2010).

19. He, H. H. et al. Nucleosome dynamics define transcriptional enhancers. Nature

Genet. 42, 343–347 (2010).

20. Creighton, M. P. et al. Histone H3K27ac separates active from poised enhancers

and predicts developmental state. Proc. Natl Acad. Sci. USA 107, 21931–21936

(2010).

21. Kaikkonen, M. U. et al. Remodeling of the enhancer landscape during macrophage

activation is coupled to enhancer transcription. Mol. Cell 51, 310–325 (2013).

22. Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. Nature Rev.

Genet. 7, 862–872 (2006).

23. Gaffney, D. J. et al. Unraveling inflammatory responses using systems genetics and

gene-environment interactions in macrophages. Cell 151, 658–670 (2012).

24. Song, L. et al. Open chromatin defined by DNase I and FAIRE identifies regulatory

elements that shape cell-type identity. Genome Res. 21, 1757–1767 (2011).

25. Bennett, B. J. et al. A high-resolution association mapping panel for the dissection

of complex traits in mice. Genome Res. 20, 281–290 (2010).

26. Raetz, C. R. et al. Kdo2-lipid A of Escherichia coli, a defined endotoxin that activates

macrophages via TLR-4. J. Lipid Res. 47, 1097–1111 (2006).

27. Smale, S. T. Transcriptional regulation in the innate immune system. Curr. Opin.

Immunol. 24, 51–57 (2012).

28. Wong, D. et al. Extensive characterization of NF-kB binding uncovers non-
canonical motifs and advances the interpretation of genetic functional traits.

Genome Biol. 12, R70 (2011).

29. Pham, T. H. et al. Mechanisms of in vivo binding site selection of the hematopoietic

master transcription factor PU.1. Nucleic Acids Res. 41, 6391–6402 (2013).

30. Jolma, A. et al. DNA-binding specificities of human transcription factors. Cell

152, 327–339 (2013).

31. Barski, A. et al. High-resolution profiling of histone methylations in the human

genome. Cell 129, 823–837 (2007).

32. Wang, D. et al. Reprogramming transcription by distinct classes of enhancers

functionally defined by eRNA. Nature 474, 390–394 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. J. Luisi for providing access to eQTL data (http://
systems.genetics.ucla.edu/) and for productive conversations. We thank D. Pollard for
discussions and suggestions, and L. Bautista for assistance with figure preparation.

These studies were supported by National Institutes of Health (NIH) grants DK091183,
CA17390 and DK063491 (C.K.G.). M.U.K. was supported by the Foundation Leducq
Career Development award and grants from Academy of Finland, Finnish Foundation
for Cardiovascular Research and Finnish Cultural Foundation, North Savo Regional
government. C.E.R. was supported by the American Heart Association Western States
Awards (12POST11750017) and the NIH (R35DK070494).

Author Contributions S.H., C.K.G. and C.E.R. designed the study; S.H., C.E.R., K.A.A.,
M.U.K. and L.D.O. performed experiments; C.E.R. performed all

gene-variation-related analysis; C.B. wrote custom code for HOMER2 and analysed

data; K.A.A. and S.H. analysed data; C.E.R., S.H. and C.K.G. wrote the manuscript.

Author Information Data are available in the Gene Expression Omnibus (GEO) under

accession GSE46494. Reprints and permissions information is available at

www.nature.com/reprints. The authors declare no competing financial interests.

Readers are welcome to comment on the online version of the paper. Correspondence
and requests for materials should be addressed to C.K.G. (ckg@ucsd.edu).
METHODS

Animals and cell culture. Thiglycato-colored peritoneal macrophages were collected 4 days after injection from male 6–8 week C57BL/6J, BALB/cJ or CB6F1/J LPS-stimulated hybrid mice, and plated at 20 × 10^6 cells per 15-cm Petri dish in RPMI61640 plus 10% FBS and 1% penicillin-streptomycin. One day after plating, cells were treated with fresh media with or without 100 ng ml⁻¹ KLA for 1 h, and then directly used for downstream analyses. All animal experiments were performed in compliance with the ethical guidelines set forth by University of California, San Diego’s Institutional Care and Use Committee (IUCAC).

ChIP-Seq and feature identification. Media were decanted and cells were fixed at room temperature with either 1% formaldehyde in PBS for 10 min (for PU.1, C/EBPβ, H3K27ac ChiPs) or 2 mM disuccinimidylglutarate (DSS, Pierce) and 10% dimethylsulfoxide (DMSO) in PBS for 30 min followed by 1% formalde- hyde in PBS for another 15 min (p65). After quenching the reaction by adding glycerine to final concentration 0.125 M, cells were washed twice with PBS and snap-frozen in dry-ice and methanol. ChIPs for PU.1 (Santa Cruz, sc-352) and C/EBPβ (Santa Cruz, sc-61) were performed as described previously³. The H3K27ac (Abcam, ab7292) ChIP was performed in the presence of 1 mM butyric acid. For p65 (Santa Cruz, sc-372), immunoprecipitation conditions were identical to those described before, except that pre-clearing was omitted, and the ChIP was previously constructed. To control for open chromatin and library biases, input chromatin replicates were prepared using magnetic beads similar to described procedures².

Reference (C57BL/6J) genome and the BALB/cJ contigs were aligned to using Bowtie0.12.7 on a HiSeq 2000 sequencer (Illumina) using CASAVA1.7 or 1.8.

Animals and cell culture. For polyA-Seq, hybrid mice, and plated at 20 × 10^6 cells per 15-cm Petri dish in RPMI61640 plus 10% FBS and 1% penicillin-streptomycin. One day after plating, cells were treated with fresh media with or without 100 ng ml⁻¹ KLA for 1 h, and then directly used for downstream analyses. All animal experiments were performed in compliance with the ethical guidelines set forth by University of California, San Diego’s Institutional Care and Use Committee (IUCAC).

ChIP-Seq and feature identification. Media were decanted and cells were fixed at room temperature with either 1% formaldehyde in PBS for 10 min (for PU.1, C/EBPβ, H3K27ac ChiPs) or 2 mM disuccinimidylglutarate (DSS, Pierce) and 10% dimethylsulfoxide (DMSO) in PBS for 30 min followed by 1% formalde- hyde in PBS for another 15 min (p65). After quenching the reaction by adding glycerine to final concentration 0.125 M, cells were washed twice with PBS and snap-frozen in dry-ice and methanol. ChIPs for PU.1 (Santa Cruz, sc-352) and C/EBPβ (Santa Cruz, sc-61) were performed as described previously. The H3K27ac (Abcam, ab7292) ChIP was performed in the presence of 1 mM butyric acid. For p65 (Santa Cruz, sc-372), immunoprecipitation conditions were identical to those described before, except that pre-clearing was omitted, and the ChIP was previously constructed. To control for open chromatin and library biases, input chromatin replicates were prepared using magnetic beads similar to described procedures.

Reference (C57BL/6J) genome and the BALB/cJ contigs were aligned to using Bowtie0.12.7 on a HiSeq 2000 sequencer (Illumina) using CASAVA1.7 or 1.8.

Strain-specific motifs. To study strain-specific ChIP-Seq differences, we used strain-specific motifs. De novo motif finding in ChIP-Seq enriched regions from both inbred mouse strains was used to define PWMs for transcription factors of interest (Extended Data Fig. 3a, b and Supplementary Table 1). These PWMs were used to define strain-specific motifs by using the options ‘homem2 -find < individual genome sequence >=’ in Homer for each genome sequence for the regions of interest. The positions of the identified motifs were compared between strains, taking into account shifts caused by indels relative to peak start coordinates and which DNA strand matched the identified motifs. Motifs with alignments only in one genome were considered strain specific.

PolyA-Seq. For each condition, RNA was isolated from 5 × 10^6 thiglycato- colored macrophages with Trizol LS, and 15 μg RNA was DNase-treated using TURBO DNase (Ambion) according to the manufacturer’s instructions. Poly-A RNA was sequenced using the MicroPoly(A) Purist kit (Ambion), according to the manufacturer’s instructions. Isolated RNA was hydrolysed in a total volume of 20 μl with 2 μl RNA fragmentation buffer (Ambion) for 10 min at 70 °C. The reaction was stopped with stop buffer, and buffer was exchanged to Tris, pH 8.5, using P30 size-exclusion columns (Bio-Rad). The fragmented RNA (30 ng) was 5′-decapped in a total volume of 21 μl containing 0.5 μl tobacco acid pyrophosphatase (TAP, Epicentre), 2 μl 10× TAP buffer and 1 μl SUPERase-In, and incubated for 2 h at 37 °C. To dephosphorylate RNA 3′ ends, 0.5 μl TAP buffer, 1.5 μl water, 0.5 μl of 0.25 M MgCl₂ (4.17 mM final; 1 mM EDTA for maximum phosphatase activity), and 0.5 μl of 10 mM ATP (0.2 μM final to protect PKN) where added, and the reaction was incubated with 1 μl PKN (Enzymaks) for 50 min at 37 °C. RNA fragments were 5′-phosphorylated by adding 10 μl 10× T4 DNA ligase buffer, 63 μl water and 2 μl PKN, and incubated for 60 min at 37 °C. RNA fragments were isolated using Trizol LS, precipitated in the presence of 300 mM sodium acetate and 2 μl glycoblue (Ambion), washed twice with 80% ethanol and dissolved in 4.5 μl water.

To prepare sequencing libraries, 0.5 μl of 9 μM 5′-adenylated 3′MPX adapter/ 5′Phos/AGATCGGAAGACGACGCTTC/3′/AMMO/ (IDT, desalted; adenylated CACGAGCGGCTTC/3′/AMMO/). The C57BL/6J and BALB/cJ sequence using Bowtie2 were performed as described previously. The libraries for the replicates were prepared using magnetic beads similar to described procedures². Libraries for the initial p65 ChIPs were prepared as described previously³, libraries for the replicates were prepared using magnetic beads similar to described procedures². Libraries for the initial p65 ChIPs were prepared as described previously³, libraries for the replicates were prepared using magnetic beads similar to described procedures².
GRO-Seq. GRO-Seq was performed as described previously using 10^7 cells per condition. RNA at RefSeq transcripts was quantified for GRO-Seq and polyA-RNA-Seq by counting the normalized tags (to 10 million tags per experiment) in annotated exons for each RefSeq transcript.

Odds ratio calculations and statistical testing. Odds ratios for observing C57BL/6j-specific motif mutations relative to BALB/c-specific motif mutations in different classes of bound/modified loci (for example, Fig. 2b) were calculated using \((p_1/(1 - p_1))/(p_2/(1 - p_2))\), in which \(p_1\) is the frequency of C57BL/6j-specific motifs, and \(p_2\) is the frequency of BALB/c-specific motifs. For Extended Data Fig. 4i, \(k\), \(p_1\) is the frequency of indicated events occurring in differentially bound loci and \(p_2\) is the frequency in similarly bound loci. Unless otherwise indicated, \(t\)-tests were two-sided assuming unequal variance.

eQTL analysis. eQTL analysis was performed as previously described. In brief, thioglycolate-elicited peritoneal macrophages were collected from 85 strains of mice. RNA was processed and hybridized to Affymetrix Genome HT_MG-430A. There were 22,416 probe sets analysed after removing individual probes overlapping SNPs and probe sets with 8 or more probes overlapping SNPs. Expression data was RMA normalized.

A total of 3,918,755 SNPs with a minimum minor allele frequency of 10% originating from mouse Perlegen variation data set was imputed across the strains, and filtered to 3,695,041 SNPs based on proximity (<2 Mb) to transcription start sites of transcripts detectable by the microarray. Gene expression for each transcript was associated to SNPs within 2 Mb using the efficient mixed-model association approach to multiple testing.

Association mapping of eQTLs was overlapped with H3K4me2 and H3K27ac regions. To test for enrichment of significant eQTL in H3K4me2 regions we made use of the hypergeometric distribution function:

\[
P(X = k) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}
\]

in which \(k\) successes represents the number of significant eQTL in (or in linkage disequilibrium with) H3K4me2 regions; \(m\) denotes the number of SNPs with significant eQTL; \(N\) denotes the total number of SNPs; and \(n\) denotes the total number of SNPs in H3K4me2 regions.

Macrophage eQTL enrichment in enhancers from other cell types. The short-read archive files were downloaded from the GEO under accession GSE24164 (ref. 20) for ChIP-sequencing for the H3K27ac mark in whole liver (Sequence Read Archive accession SRX027340), pro-B cells (SRX027345), and embryonic stem cells (SRX027331 and SRX027332), and input chromatin as background (liver: SRX027343, pro-B: SRX027348, stem cells: SRX027352). Sequencing reads were mapped to the C57BL/6j genome. H3K27ac regions were identified where tag pile-ups exceeded four times the input tags using HOMER, and interrogated for enrichment of significant macrophage eQTLs as described for macrophage H3K4me2 and H3K27ac regions above.

**Reporter assays and mutation analysis.** One-kilobase enhancers were PCR-amplified from C57BL/6j and BALB/c genomic DNA using genomic primers not overlapping variants that introduced terminal BamHI, BglII or BclI sites on one end and SalI or XhoI sites on the other end of the PCR products, depending on the restriction site content of the enhancer. These were digested with the respective restriction enzymes and ligated into a modified, BamHI- and SalI-digested pGL4.10 luciferase reporter plasmid (Invitrogen) containing a minimal HSV-TK promoter derived from pTAL-Luc (Clontech) (see Fig. 5a). Alternatively, 1-kb fragments were amplified using primers that introduced overhangs identical to the sequences flanking the BamHI/SalI tandem site in the pGL4.10 plasmid. Fragments were purified from the PCR reaction by STRIP using magnetic beads and cloned into the BamHI/SalI-cut reporter plasmid described above using Gibson Assembly master mix (NEB) according to the manufacturer’s instructions. Mutations were introduced by PCR amplification with complementary primers containing the mutation to be introduced in the centre, followed by DpnI digestion of the template and transformation of bacteria. All constructs were confirmed by sequencing. For each reporter assay, 300 ng plasmid was transfected into RAW264.7 macrophages using SuperFect (Qiagen) together with 300 ng UB6 promoter-driven β-galactosidase reporter for transfection normalization in 24-well plates seeded with 1 × 10^5 cells 24 h before transfection. Twenty-four hours after transfection, media alone (RPMI plus 10% FBS) or media containing 100 ng/ml KLA was added for an additional 16 h. Luciferase activity was measured 24 h after transfection using a Veritas microplate luminometer (Turner Biosystems), and normalized to β-galactosidase activity (Applied Biosystem) for transfection efficiency. Each experiment was performed at least three independent times, with each reaction done in triplicates. Data represented as mean ± s.d., and statistical significance was determined by a one-sided \(t\)-test.

Stable transfected cell lines were made by transient co-transfection of the linearized reporter plasmids together with linearized neomycin resistance-expressing pCDA3 vector as described above, followed by incubation with 275 μg/ml G418 (G418 Sulphate, Invitrogen) for 2–3 weeks. Bulk cells from stably transfected colonies were tested for luciferase activity and normalized to total protein concentration (DC Protein Assay, BioRad).

33. 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).
34. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359 (2012).
35. Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. A Stat. Soc. 57, 289–300 (1995).
36. Frazer, K. A. et al. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. Nature 448, 1050–1053 (2007).
37. Kirby, A. et al. Fine mapping in 94 inbred mouse strains using a high-density haplotype resource. Genetics 185, 1081–1095 (2010).
38. Kang, H. M. et al. Efficient control of population structure in model organism association mapping. Genetics 178, 1709–1723 (2008).
39. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploviz: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265 (2005).
Extended Data Figure 1 | ChIP-Seq data characteristics. a, Summary of ChIP-Seq features identified. The number of ChIP-seq regions/peaks identified in untreated primary thioglycolate-elicited macrophages is tabulated for H3K4me2, H3K27ac, PU.1 and C/EBPα. Peaks for p65 were quantified in macrophages treated with 100 ng ml\(^{-1}\) KLA for 1 h. Unless otherwise noted, modification and binding were considered strain-specific at ≥fourfold difference between strains in sequenced tags, and the FDR was based on Poisson cumulative distribution testing and Benjamini and Hochberg correction.

b–e, Reproducibility and strain-specific binding. Two separate pools of thioglycolate-elicited macrophages from mice from each strain (C57BL/6J and BALB/cJ) were treated with KLA for 1 h. ChIP-seq for p65 was performed separately on each pool (see Methods). The number of normalized sequencing tags at the union of peaks identified in the indicated experiments is shown. Peaks highlighted in red were deemed experiment-specific using criteria applied throughout this study (fourfold, and FDR based on the cumulative Poisson distribution and Benjamini and Hochberg FDR estimation). The number of experiment-specific peaks is indicated (red) relative to the total number of peaks (black).

g, Heat map showing pairwise correlation between all p65 experiments. Pearson correlation coefficients are given for each comparison.
Extended Data Figure 2 | Strain-specific LDTF binding correlates with variant density and location in LDTF motifs but not with genomic context.

a, Genomic features do not distinguish between strain-similar and strain-specific LDTF binding. Peaks were restricted to promoter-distal peaks (>3 kb to gene start sites). Genomic features (distance to nearest gene, distance to nearest repeat, CpG content and conservation score) were compared among three pairs of strain-similarly bound and strain-specifically bound PU.1 and/or C/EBPα loci (listed as groups 1–6). Box midlines are medians, boundaries are first and third quartiles. Whiskers extend to the extreme data points. CpG content and conservation were quantified in 1-kb regions centred on the LDTF peak. *P* values from two-sided *t*-test are given if below 0.05.

b, Strain-specific C/EBPα binding occurs in regions with increased variant density. ChIP-Seq tag counts in 200-bp peak regions were stratified into five bins according to log2 ratios of peak tag counts in BALB/cJ versus C57BL/6J mice (x axis, log2 ratio), and the variant density distributions are shown per bin. c, d, Variant density distribution in strain-specific peaks. Mean variant densities within 10-bp bins were plotted against the logarithmic ratio of PU.1 peak tag counts in each strain (binding ratio) (x axis). The frequency of motifs that were mutated in BALB/cJ are plotted in red and those mutated in C57BL/6J in blue. e, The analogous relationship as shown in e for PU.1 is plotted for C/EBPα motif mutations versus C/EBPα strain binding ratio.
Extended Data Figure 3 | Strain-specific PU.1 and C/EBPα binding correlates with strain-specific LDTF motifs. a, Top and degenerate motifs enriched in H3K4me2 and PU.1 or C/EBPα ChIP-Seq peaks. These motifs were used to query individual genome sequences and identify strain-specific motifs in subsequent analysis. Degenerate ‘weak’ motif occurrence numbers for a given factor include ChIP-Seq peaks containing ‘strong’ motifs. Position weight matrices and log-odds score thresholds for each motif are given in Supplementary Table 1.

b, NF-κB consensus and degenerate motifs enriched in p65 ChIP-Seq peaks.

c, d, Mutations in LDTF motifs affect PU.1 (c) and C/EBPα (d) binding. Left panels show scatterplots for the ChIP-Seq-defined binding of PU.1 (c) and C/EBPα (d) between C57BL/6J (x axes) and BALB/c (y axes). Strain-specific motifs were queried within 100-bp of each peak position. Red symbols designate binding events at loci where a polymorphism mutated a C/EBP, PU.1 or AP-1 motif in the C57BL/6J genome, whereas the motif was intact in the BALB/cJ genome. Blue points highlight mutations in these motifs in the BALB/cJ genome only. Violin plots in the right panels show the effect of each motif mutation (along x axes: PU.1, C/EBP, AP-1 and NF-κB) on the ratio of PU.1 (c) and C/EBPα (d) binding between mouse strains, (y axes: positive values denote BALB/cJ-specific, negative values denote C57BL/6J-specific). Tag ratio distributions for peaks overlapping C57BL/6J motif mutations are on the left (light colours), those for peaks overlapping BALB/cJ motif mutations are on the right (dark colours). The fold-difference between mean binding ratios is indicated under the pair of distributions for each motif. The grey distribution indicates PU.1- or C/EBPα-bound loci not overlapping strain-specific motifs.
Extended Data Figure 4 | Effects of cognate motif distance from peak centre, variant position within a motif and the presence of alternative motifs on strain-differential binding of PU.1 and C/EBPα. a–d, PU.1 and C/EBPα motif mutations near the experimentally derived peak centre are associated with impaired binding. a, c. The ratios of the frequencies of variant-containing motifs at the given distances from strain-differentially versus strain-similarly bound peak centres (twofold versus twofold tag count ratio) for 570 PU.1 (a) and 278 C/EBPα (b) variant-containing motifs are shown, respectively. b, d. The distribution of absolute strain peak tag count ratios of peaks whose centre is at the given distances from mutated PU.1 (b) or C/EBPα (d) motifs. Box midlines are medians, and boundaries are first and third quartiles. Whiskers extend to the extreme data points. P-values are from two-sided t-test. e, f. Effects of alternative PU.1 and C/EBPα motifs and core mutations on binding. The number of non-mutated ‘alternative’ PU.1 or C/EBPα motifs in the strain with a PU.1 or C/EBPα motif mutation was counted, and the absolute respective PU.1 or C/EBPα log2 strain binding ratio is shown. g. Defining the C/EBPα motif core by comparing differential versus similar C/EBPα binding. Sequence variants within C/EBPα motifs located in loci devoid of alternative C/EBPα motifs (n = 178) were counted according to whether they were in differential (blue) or similar (red) C/EBPα-bound peaks. h. The distribution of PU.1 binding strain log2 ratios (x axis) is shown for PU.1 mutations located in the PU.1 core and non-core nucleotides (defined in Fig. 1g). i. The C/EBPα binding strain log2 ratio is shown for C/EBPα core and non-core mutations as defined in g. j, k. Motif mutations predominately occur at differentially bound loci. The odds ratios (x axis; equation shown in box) describing the relative effect of the indicated characteristics of mutated motifs on differential binding relative to similar binding are shown for PU.1 (j) and C/EBPα (k). Whiskers show 95% confidence intervals. nt, nucleotides. l, m, The percentage of respective motif mutations consistent with altered PU.1 (l) and C/EBPα (m) binding is shown for the indicated categories of motif mutations.
Extended Data Figure 5 | Analysis pipeline for predicting functional PU.1 mutations in NOD. Data are shown in Fig. 1H.
Extended Data Figure 6 | LDTF motif mutations are enriched at strain-specific C/EBPα-bound loci relative to strain-similar loci. 

a, The log_2 odds ratio for observing a C57BL/6J-specific versus BALB/cJ-specific mutation in the indicated three bins of C/EBPα binding ratios: similar (middle bin), or strain-specifically C/EBPα bound (left and right bins). Details are in the Methods.

b, Collaborative binding is largely not mediated by direct protein–protein interactions. A total of 14,199 loci bound by PU.1 and C/EBPα were centred on the PU.1 weak motif (0 on x axes) and cumulative instances of C/EBP and AP-1 motifs were plotted at each position relative to the central PU.1 motif. Interferon response factor (IRF) half-sites are plotted as control for a factor that requires direct protein–protein interactions with PU.1 for DNA binding. The motifs in each comparison showing overlapping sequence and base pair distances are indicated to the right. Peak distances from the central PU.1 motif are indicated in the histograms. RC denotes reverse complement.

c, Allele-specific C/EBPα binding in F1 heterozygotes is similar to binding in homozygous parental strains. C/EBPα ChIP-seq reads from CB6F1/J hybrid F1 macrophages were mapped with no mismatches to both parental genome sequences to identify allele-specific reads. C/EBPα log_2 peak tag ratios between the parental strains (BALB/cJ versus C57BL/6J) are shown on the x axis, and the log_2 ratio of allele-specific reads in the F1 hybrids are shown on the y axis (BALB/cJ allele versus C57BL/6J allele). C57BL/6J-specific C/EBPα regions are blue, BALB/cJ-specific C/EBPα regions are red, and strain-similar C/EBPα regions are black. Strain-specific or similar regions were defined from parental data.
Extended Data Figure 7 | Strain-specific epigenetic marks correlate with LDTF binding, and LDTF mutations segregate with altered H3K4me2 deposition. a–f, Strain-specificity of LDTF binding and epigenetic marks. The relative amount of H3K4me2 (a–c) and H3K27ac (d–f) between C57BL/6J and BALB/cJ (x axes) is highly correlated with the amount of bound PU.1, C/EBPa or product (PU.1 x C/EBPa). The log2 ratios of the peak tag counts for PU.1, C/EBPa and PU.1 x C/EBPa in each strain are shown relative to the log2 of the peak tag count ratios for H3K4me2 or H3K27ac. Loci containing strain-specific LDTF motifs in a differentially PU.1- or C/EBPa-bound peak are highlighted. Correlation coefficients (Pearson) are indicated for each comparison. g, LDTF mutations segregate with altered H3K4me2 deposition. The log2 of the ratio of the product of the normalized peak tag counts for PU.1 and C/EBPa in 200 bp in each strain (x axis) is compared to the log2 H3K4me2 peak tag ratio in 1 kb (y axis) for loci containing at least a PU.1 or C/EBPa peak. Strain-specific LDTF motif mutations are indicated by the designated symbols and coloured by the mutated strain (C57BL/6J red, BALB/cJ blue). The distribution of H3K4me2 strain ratios stratified by corresponding LDTF strain mutations is shown to the right, with P value from a two-sided t-test. h, Relationships between H3K27ac patterns in different cell types. ES, embryonic stem. Hierarchical clustering of H3K27ac-positive regions as determined by ChIP-Seq and analysis with HOMER. The number of ChIP-seq tags in each of the 86,264 H3K27ac-marked regions used for comparison with eQTL data in Fig. 2e that were detected in at least one cell type was clustered using Euclidean distance.
Extended Data Figure 8 | LDTFs prime the p65 cistrome. a, The 69,517 regions that gained p65 in C57BL/6J after KLA treatment were analysed for binding of PU.1 and C/EBPα with and without KLA treatment as shown in the pie charts. Loci not bound by PU.1 or C/EBPα after KLA treatment were analysed by de novo motif finding. The most enriched motif was AP-1, and the second-most enriched motif was NF-κB. b, Violin plots of the p65 strain ratios of mean-normalized p65 binding for p65-bound peaks stratified by motifs mutations present in either BALB/cJ or C57BL/6J. Mutated motifs included PU.1 (strong and weak), C/EBP (strong and weak), C/EBP:AP-1 heterodimers, AP-1 and NF-κB. The effect on p65 binding per group is shown by comparing the mean-normalized p65 tag binding ratio along the y axis (log2(BALB/cJ–C57BL/6J); positive values denote BALB/cJ-specific, negative values denote C57BL/6J-specific). White circles indicate the distribution means, and the average fold change associated with C57BL/6J-mutating and BALB/cJ-mutating SNPs in the respective motifs is given beneath. One-sided t-test P values between each pair of distributions ranged from $1 \times 10^{-29}$ to $1 \times 10^{-14}$. c, Variant density in strain-specific and strain-similar p65 peaks. Mean variant density within 10-bp bins relative to p65 ChIP-Seq peak centres in strain-similar (red) or strain-specific peaks (blue). d–e, The variant density distribution in strain-specific p65 peaks is broader than those for PU.1 or C/EBPα. Fold enrichment of variant densities in strain-specific relative to strain-similar peaks (y axes) for PU.1 (d), C/EBPα (e) and p65 (f) is shown relative to the peak centres (x axes). Ratios plotted in d and e are from data in Extended Data Fig. 2c and d, respectively.
Extended Data Figure 9 | Validation of strain-specific enhancer activity.

**a**, Enhancer activity in transient reporter assays correlates with strain-specific LDTF and p65 binding. Luciferase assay results for 24 loci (20 strain-specific enhancers with strain-specific motifs, 1 positive control with strain-similar enhancer activity (row 7, column 3), 2 negative controls lacking enhancer activity in both strains (row 8, columns 1 and 2), and 1 strain-specific enhancer lacking a strain-specific motif (row 8, column 3)) in transiently transfected RAW264.7 cells 48 h after transfection. Each 1-kb locus is represented by the horizontal midline within a box (see Fig. 5). ChIP-seq tag pile-ups are shown for PU.1 (green), C/EBPα (blue), p65 (red), H3K27ac (purple) and H3K4me2 (orange) for C57BL/6J (above midline) and BALB/cJ (below midline) with identical scales. Binding/modification data are shown after treatment with 100 ng ml⁻¹ KLA. Vertical black lines indicate SNP locations. Horizontal bars indicate average luciferase (enhancer) activity of the empty vector (blue, no enhancer), activity of a locus cloned from either strain in grey C57BL/6J (above) and BALB/cJ (below) under basal conditions, or after overnight stimulation with 100 ng ml⁻¹ KLA (pink). Luciferase values from transiently transfected cells were normalized to the activity measured for a co-transfected UB6 promoter-β-galactosidase reporter construct. Empty vector values were scaled to 0.5 for the first four loci, and to 1 for the remaining loci. Constructs in which the predicted motif-disrupting variant alleles were swapped are denoted by 'M', with mutations causing a significant effect in at least two out of three replicates being denoted by an additional asterisk (*P* < 0.05, one-sided *t*-test). Error bars show s.d. from three biological replicates, average values are indicated next to each bar. Experiments were replicated at least three times. Significant strain-specific enhancer activity is indicated by a dagger (grey without treatment, red after KLA treatment, one-tailed *t*-test, *P* < 0.05).

**b**, Chromatinization is necessary for the strain specificity of a subset of enhancers. RAW264.7 cells were stably transfected with the two constructs containing the loci that showed strain-specific binding but lacked strain-specific enhancer activity in transient reporter assays (row 4, column 1 and row 1, column 3, marked by an asterisk). Luciferase activity measured in lysates of stably transfected cells was normalized to total protein content. RLU, relative light units.
Extended Data Figure 10 | Motif analysis identifies causal SNPs in enhancers. Regions of ∼1 kb size centred on PU.1 or C/EBPa ChIP-Seq peaks of similar tag count in C57BL/6J and BALB/cJ (<twofold difference) that contain a variant in a motif for the respective factor within 100 bp of the peak centre were cloned into a luciferase reporter plasmid containing a minimal HSV-TK derived promoter. Three independent transient transfection experiments were performed in RAW264.7 cells, with triplicate transfections of each construct. Where indicated, variant nucleotides in a motif were mutated to that present in the other strain, and the resulting enhancer activity was scored relative to the wild-type allele. Shown are the ratios of the normalized luciferase activity of the C57BL/6J versus BALB/cJ alleles from a representative experiment. Luciferase values from transiently transfected cells were normalized to the activity measured for a co-transfected UB6 promoter β-galactosidase reporter construct. Error bars represent derived s.d. calculated by Gaussian error propagation. Constructs exhibiting significantly different activity ratios in two out of three experiments as determined by two-sided t-test (P < 0.05) are marked with an asterisk. Strain and motif mutated by a variant are indicated below denoted by the ‘m’ prefix. In the table, plus signs indicate whether a tested enhancer contains an alternative motif for the same factor, a variant at a motif position that is not located at a motif core as defined in Fig. 1g and Extended Data Fig. 4g, or a variant in a motif located less than 20 bp away from the peak centre. Characteristics of the loci and primer sequences are in Supplementary Table 3. b, Identifying causal variants by motif analysis. Left panels show the ChIP-Seq pile-ups and SNP locations as in Extended Data Fig. 9. Right panels plot the relative enhancer reporter luciferase activities of the loci shown on the left, either in the wild-type configuration or when swapping the SNP indicated by a black triangle by site-directed mutagenesis. Motifs mutated by the indicated SNPs are shown above, with the mutation underlined and in red. c, To confirm that the centrally located PU.1 motif is essential for the C57BL/6J-specific activity, a 1-kb fragment of the locus from C57BL/6J or BALB/cJ was cloned into the luciferase reporter as described in Fig. 5 and the effects of swapping alleles at the predicted causal PU.1 SNP and flanking control 5’ and 3’ SNPs on enhancer activity are shown. Swapping alleles at the PU.1 SNP reversed strain-specific enhancer activity, whereas swapping alleles at either flanking SNP had no significant effect.