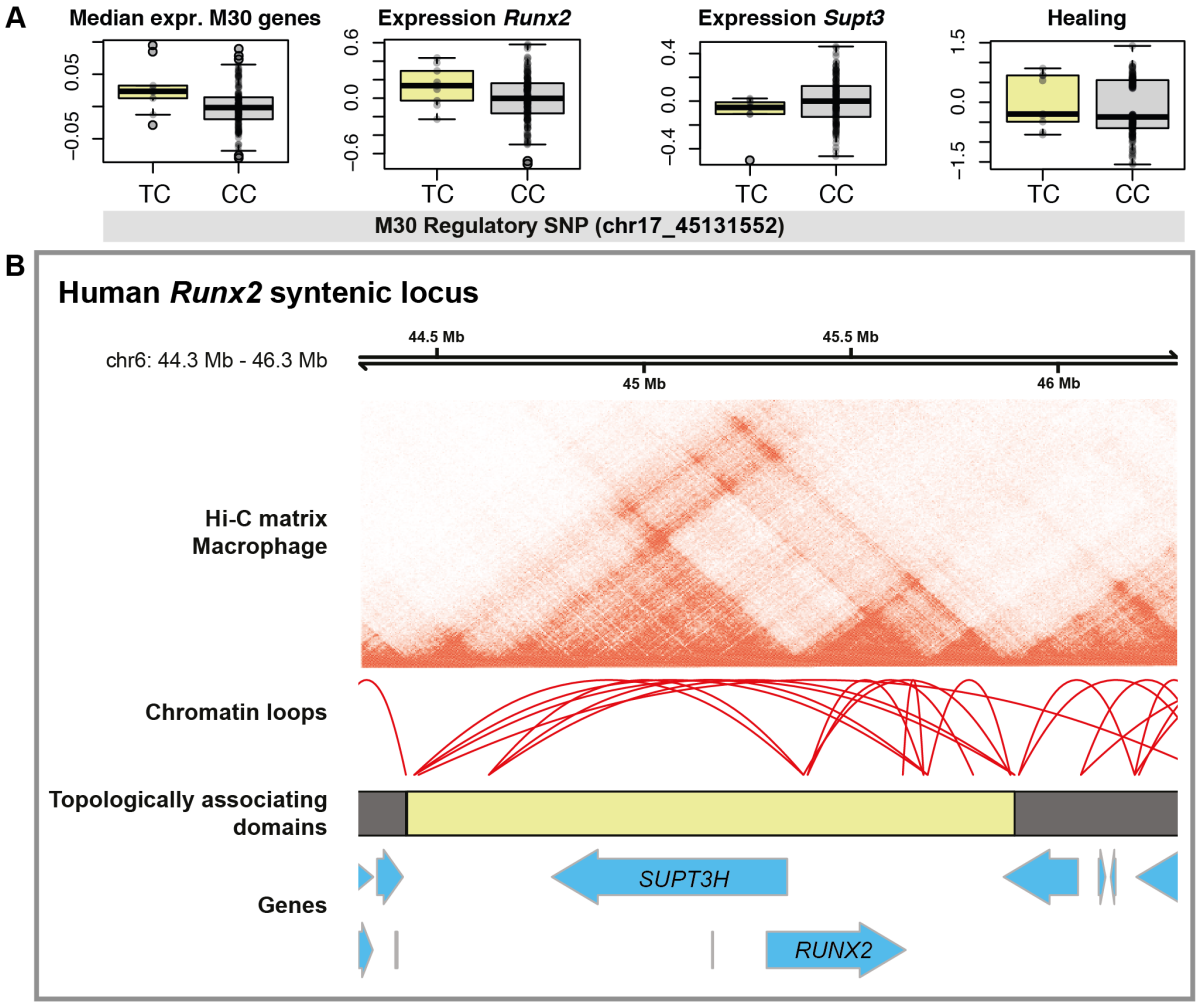
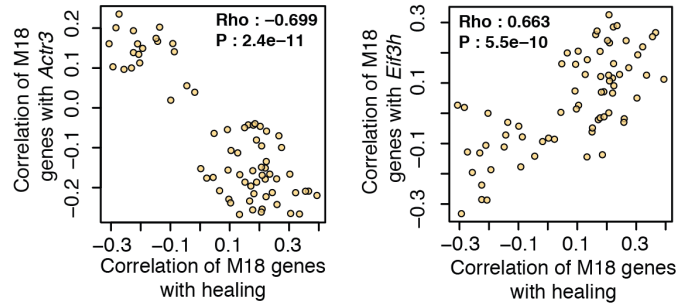
Systems genetics identifies a macrophage cholesterol network associated with physiological wound healing (by Bagnati M. *et al.*)

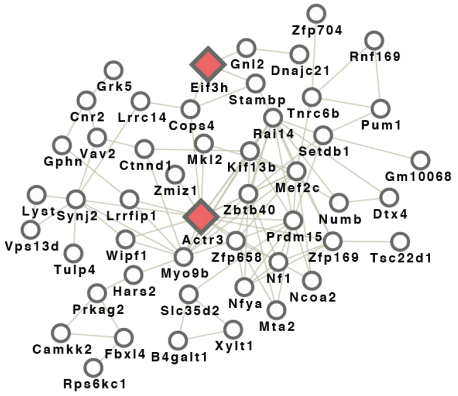
Supplemental Figures



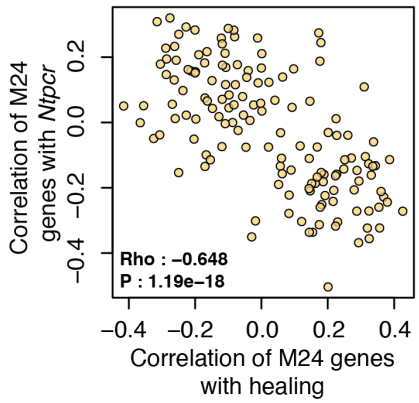
Supplemental Figure 1. *Related to Figure 2.* **A**. Box plot graphs showing the median expression of all the genes included in M30 (macrophage-mediated healing network, MMHN), *Runx2*, *Supt3* and ear healing distribution (*y-axis*, left to right respectively) according to the genotype of the M30 regulatory SNP (chr17_45131552, either TC or CC, *x-axis*). The expression level plotted corresponds to the normalised variance-stabilised gene counts (VST) after correcting for covariate effects. For each graph (left to right), the non-parametric Mann–Whitney U test p-values are P=0.025, P=0.102, P=0.167 and P=0.337. **B**. DNA topologically associated domain (TAD) containing the human *RUNX2*/*SUPT3H* locus obtained from human macrophage Hi-C data (Phanstiel et al., 2017). This TAD which contains the mouse regulatory SNP associated with the macrophage-mediated healing network (highlighted in yellow), is highly conserved across tissues and species (Barutcu et al., 2014; Harmston et al., 2017; Robertson et al., 2009).

A Co-expression module M18

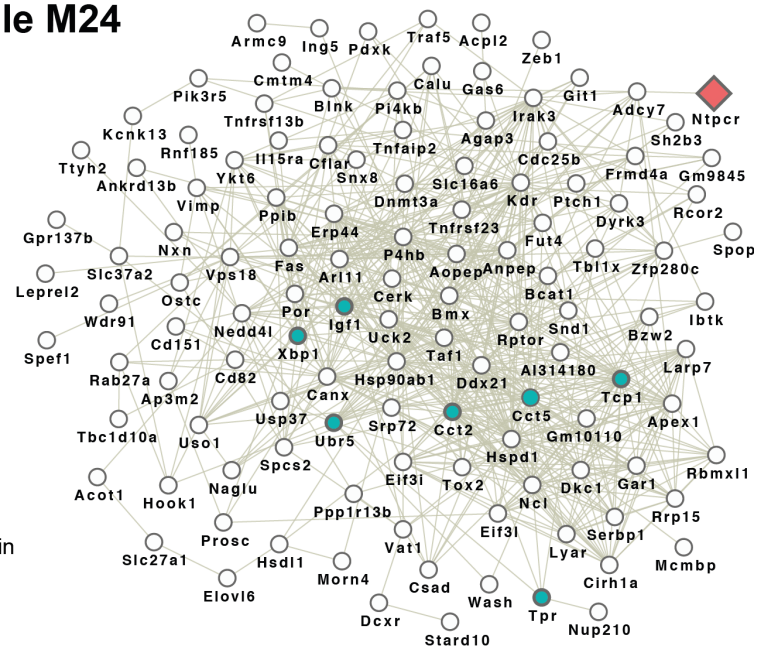




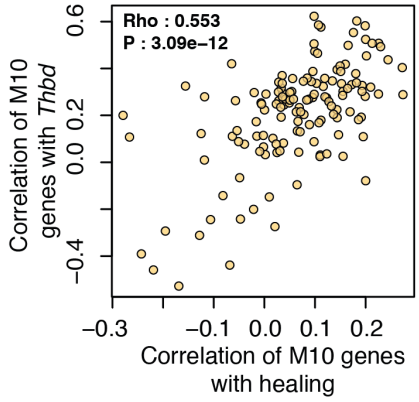
B Co-expression module M24



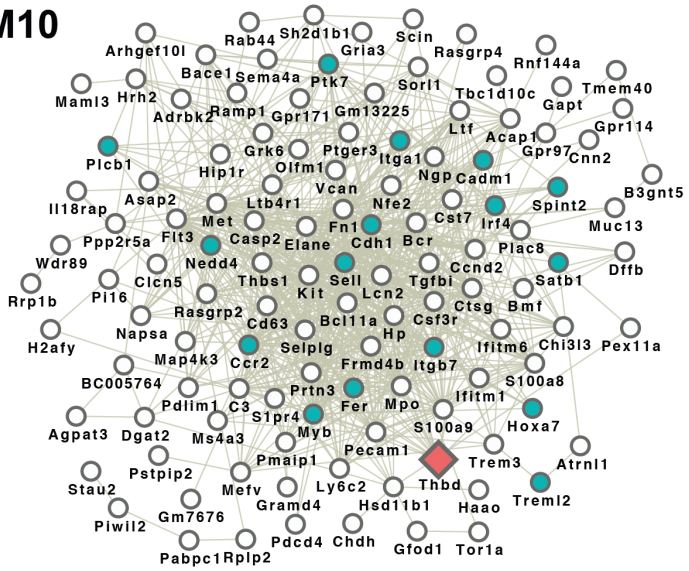
Gene annotated as involved in “positive regulation of protein localization to nucleus”

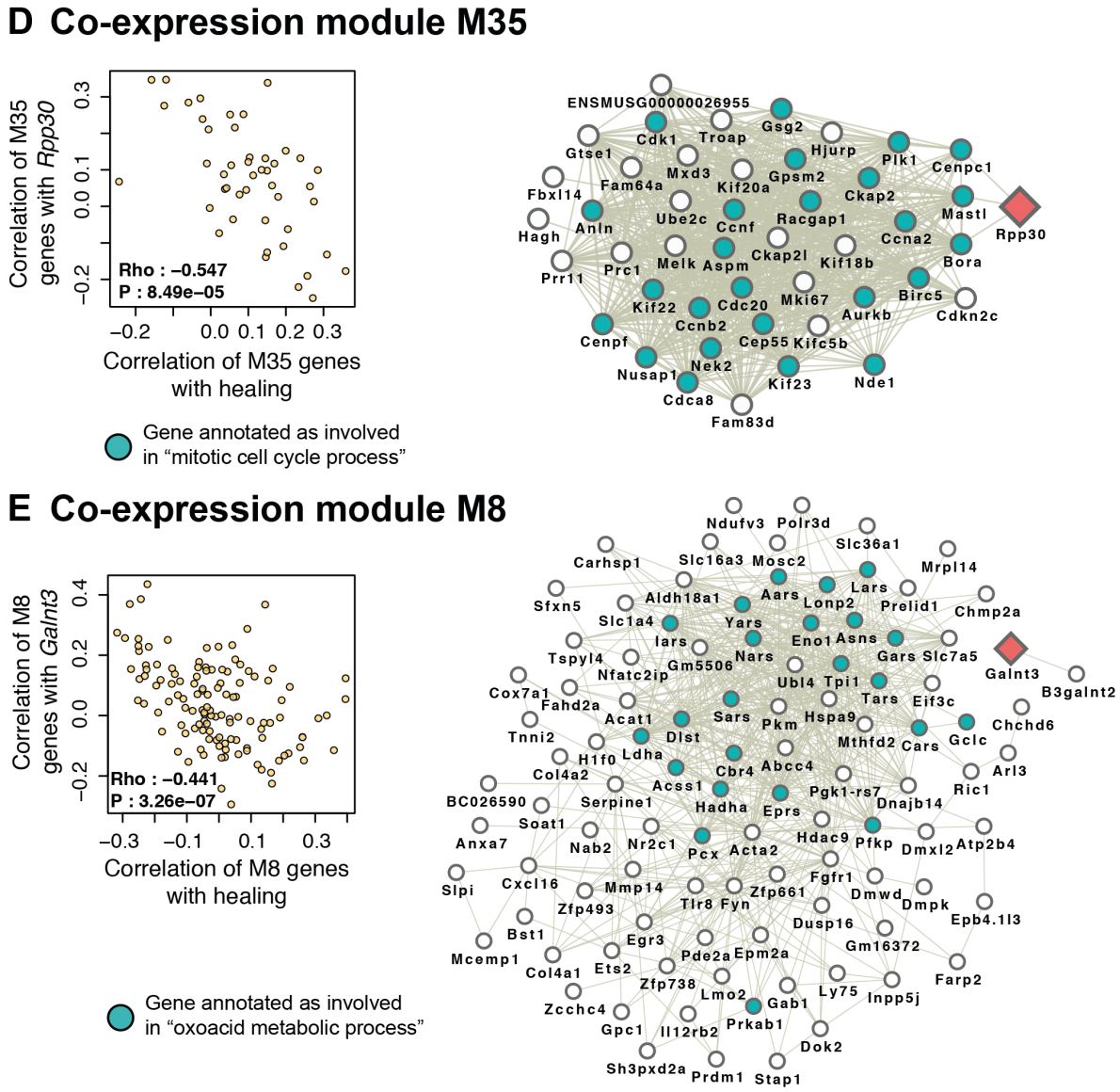


C Co-expression module M10

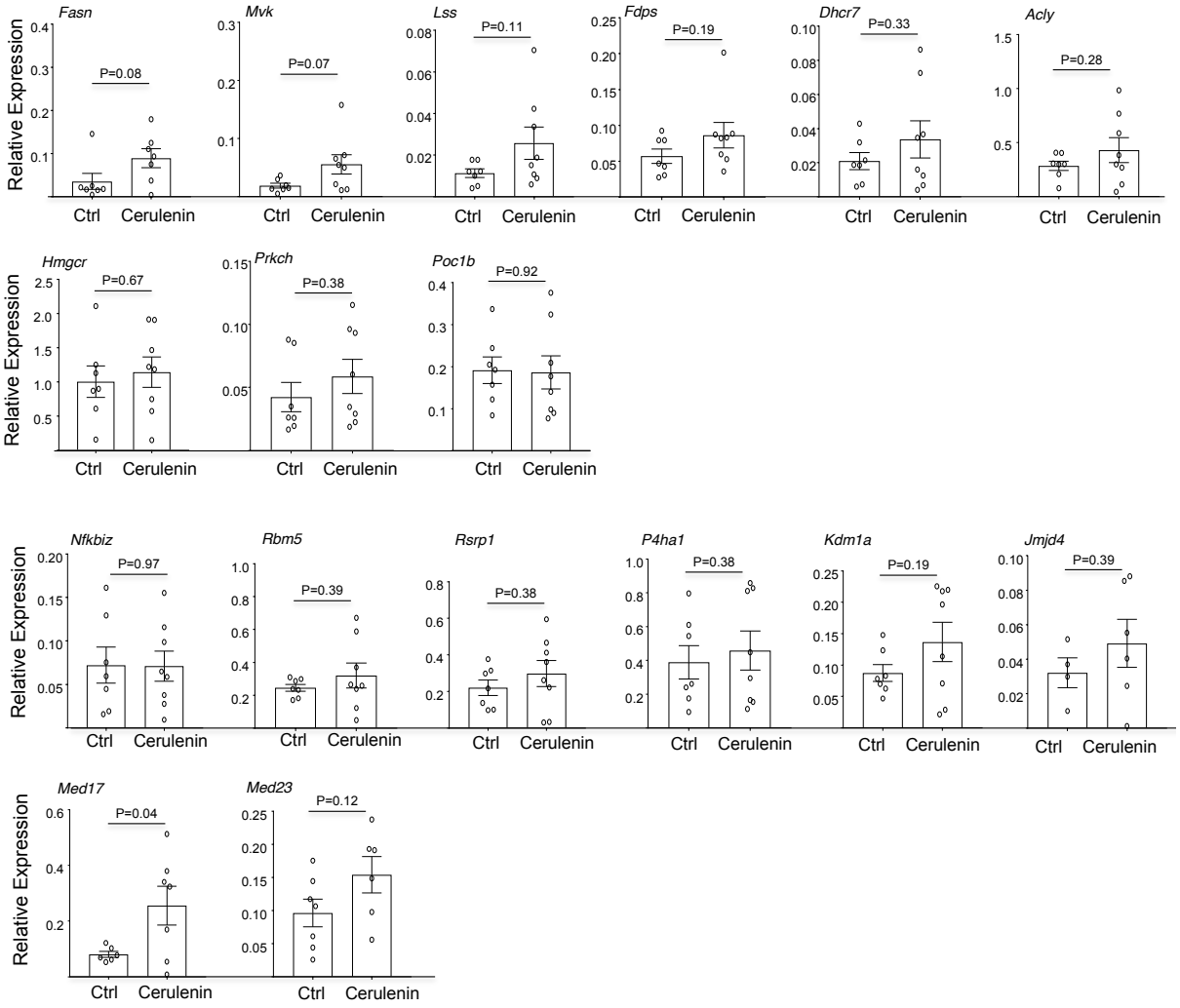


Gene annotated as involved in “biological adhesion”





Supplemental Figure 2. *Related to Figure 2 and Supplemental Table 3.* Nearest gene approach identifies candidate *trans*-acting regulators of healing co-expression modules (**A-E**). *Left*, shows a graph with the relationship between predicted regulatory gene-module and healing-module. Y-axis shows the correlation of the expression level of each gene in the co-expression module with the expression levels of the predicted *trans*-regulator by the nearest gene approach. X-axis shows the correlation between the expression levels of each gene in module and the rate of healing. *Right*, network graphs with the genes (nodes) in each module highlighting STRING protein database connections (the largest connected component is shown). Genes annotated with the top Gene Ontology (GO) functional term in each module are highlighted in green (See also Supplemental Table 2; M18 no enrichment, M24 "positive regulation of protein localization to nucleous", M10 "biological adhesion", M35 "mitotic cell cycle process" and M8 "oxoacid metabolic process"). *trans*-acting regulatory genes are shown in diamond shape.



Supplemental Figure 3. *Related to Figure 4.* M30 gene expression at day 8 following wounding. qRT-PCR for a subset of MMHN network genes in control (vehicle) and cerulenin-treated rats (n=7 controls, n=8 cerulenin, two-tailed Student's t-test p-value).

Supplemental table files

Supplemental Table 1. Ear wound phenotyping measurements in 146 mice. Related to supplemental methods.

Supplemental Table 2. List of all the co-expression modules computed from 146 outbred mice macrophages' transcriptome. Related to Supplemental methods.

Supplemental Table 3. Co-expression modules genome-wide Bayesian mapping results in 146 outbred mice. The mapping was generated using the first principal component of the modules and 10,778 mouse SNPs. Only associations with a Bayes factor higher than 100

are included. Closest gene to the associated SNPs (nearest-gene-approach trans-regulatory gene) are also included. Spearman correlation and correlation P-value between candidate *trans*-regulatory gene and healing associated genes in the module are also shown. Related to Supplemental methods.

Supplemental Table 4. Annotation of the genes belonging to the macrophage-mediated healing network (M30). Related to Supplemental methods.

Supplemental methods

Animals and phenotyping

The 1378 outbred mice used were part of a study comprising a total of 2,117 outbred mice (Crl:CFW(SW)-US_P08 (CFW); 1,065 males and 1,052 females), purchased from Charles River Laboratories, at 4–7 weeks of age over a period of 2 years. Mice were selected from the breeding colony to avoid the selection of siblings and half-siblings. Shipment and husbandry details are previously described (Nicod et al., 2016).

Wound healing phenotype was studied using the ear punch model. A 2mm ear punch was performed on the ears of each mice and the reduction in wound size has been measured after 5 weeks as ear area after healing. We tested the effect of all potential covariates on the variance in the ear area to regress them for the expression and mapping analyses. We found that the strongest effect was dependent on sex and year of measurement. The year of measurement refers to the year (2012, 2014) during which the ear area was measured from 2117 outbred mice. Indeed, there was a difference in the ear area according to the year of measurement, possibly because the time the collected tissues spent in formalin before the measurement was performed. Ear area measurements were therefore corrected for sex and year of measurement using a linear regression model. 146 mice were then selected at the extreme of the ear area distribution (62 fast healers and 84 slow healers) and their BMDMs profiled for RNA sequencing.

C57BL/6J mice used for BMDMs extraction for *Runx2 in vitro* blockage were purchased from Jackson Laboratory, UK. Lewis (LEW) rats for *in vivo* wound healing experiment were purchased from Charles River, UK. All mice and rats were used source by housing them until the appropriate experimental age. All procedures were performed in accordance to institutional guidelines and procedures approved by the UK Home Office (United Kingdom Animals Scientific Procedures Act, 1986).

Genotyping and imputation

We used the genotypes published by (Nicod et al., 2016) in a 1378 outbred mice population. We downloaded the imputed allelic dosages from http://outbredmice.org/ (359,559 single nucleotide polymorphisms, SNPs) and selected the samples for which macrophage RNA-sequencing was performed (146 mice in total). To avoid the presence of outlier SNPs in the subpopulation, we carried out a linkage disequilibrium (LD) and minor allele frequency (MAF) extra filtering steps. To perform these filtering steps, we first recoded the imputed allelic dosages (0-1) as 0, 1, 2 (representing the number of copies of the minor

allele). We used the function *snpgdsLDpruning* from R package SNPRelate 1.8.0 (Zheng et al., 2012) and the method “r”, with a LD threshold of 0.8 and a MAF threshold of 0.01 was applied. This resulted in a final number of 10,778 SNPs that were used for genetic mapping.

Sample selection and macrophage culture

Average ear area (average between left and right ear) was normalized for sex and year of measurement using a linear regression model. Only mice kept in cage density of 3 and having consistent measurements of ear area between left and right ears were considered. 146 mice showing extreme phenotypes in their rate of healing were chosen for subsequent analyses (62 fast healers and 84 slow healers, Supplemental Table 1).

Bone marrow-derived macrophages (BMDMs) from 1378 genetically outbred mice were isolated by allowing bone marrow cells to differentiate in DMEM (Thermo Fisher Scientific, Waltham, MA) containing 25 mM HEPES buffer (Sigma), 25% L929-conditioned medium, 25% fetal bovine serum (Labtech, batch 40811), penicillin (100 units/ml; Thermo Fisher Scientific) and streptomycin (100 µg/ml; Thermo Fisher Scientific), and cultured for 5 days in Petri dishes (Nunc) (Behmoaras et al., 2015; Lai et al., 2014).

RNA extraction and RNA sequencing (RNA-seq)

Total RNA was extracted from BMDMs using Trizol (Invitrogen) and RNeasy mini kit (Qiagen) according to manufacturer's instructions, with an additional purification step by on-column DNase treatment using the RNase-free DNase Kit (Qiagen) to ensure elimination of any genomic DNA. The integrity and quantity of total RNA was determined using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific) and Agilent 2100 Bioanalyzer (Agilent Technologies). In total 500 ng of total RNA was used to generate RNA-seq libraries using TruSeq RNA sample preparation kit (Illumina) according to the manufacturer's instructions. Briefly, RNA was purified and fragmented using poly-T oligo-attached magnetic beads using two rounds of purification followed by the first and second cDNA strand synthesis. Next, cDNA 3' ends were adenylated and adapters ligated followed by 15 cycles of library amplification. Finally, the libraries were size selected using AMPure XP Beads (Beckman Coulter) purified and their quality was checked using Agilent 2100 Bioanalyzer. Samples were randomized to avoid batch effects and multiplexed libraries were run on a single lane (6 samples/lane) of the HiSeq 2500 platform (Illumina) to generate

100bp paired-end reads. An average coverage of 64M reads per sample was achieved.

Raw reads were mapped to the reference mouse genome (GRCm38/mm10, Ensembl version v74) using TopHat 2.0.11 (Trapnell et al., 2009). Read counts per gene were calculated for each sample using HTseq 0.6.1 (Anders et al., 2015), (only genes with the "gene_biotype" type "protein coding" were considered for quantification). The average mapping percentage was >80%. Sequencing and mapping were controlled for quality using the FastQC software. A filtering criterion was added removing lowly expressed genes (i.e. only genes with more than 5 counts in all samples were considered for further analysis; 10,893 genes). Gene counts were normalised and variance-stabilized transformed (VST) by using DESeq2 1.14.1 R package (Love et al., 2014). VST-normalised gene counts were adjusted for *batch*, *lane*, *year of measuring* and *lane* by taking the residuals of a linear model in which the normalized gene counts were explained by these four variables.

Co-expression module inference and functional enrichment

To infer gene co-expression modules in the macrophage mouse transcriptome (10,893 genes), we used the WGCNA 1.61 R package (Langfelder and Horvath, 2008). WGCNA was run using the soft threshold beta value automatically generated by the *pickSoftThreshold* function (beta 6). We used Spearman ranked correlations and the "ward.D2" agglomeration method. To avoid extremely large clusters, the *deepSplit* parameter was set to the maximum, 4. Minimum module size was set to 30 and the module merging parameter (*MEDissThres*) was set to 0.15. This resulted in 40 co-expression modules each containing a range of transcript sets from 30 to 1,151 genes. The obtained co-expression modules with the assigned genes are included in Supplemental Table 2.

Gene Ontology (GO) (Ashburner et al., 2000) functional enrichment of all the inferred modules was computed by using the function *gprofiler* from R package gProfileR 0.6.1 (Reimand et al., 2016). The background was set to the input set of genes in all modules and non-clustered gene identified by WGCNA. Electronic annotations were excluded, the p-value correction method was set to "fdr" and only results with FDR smaller than 0.01 were considered. The full list of enriched terms in each co-expression module can be found in Supplemental Table 2. M30 GO enriched terms were visualized in a graph (Figure 3A, to avoid some redundancy, only terms with relative.depth=3 and the term with the largest overlap ("lipid biosynthetic process") are presented.

Genetic mapping of co-expression modules

Graphical Unit Evolutionary Stochastic Search (GUESS) genetic mapping tool was used to map the co-expression modules to the mouse genome (GUESS version 1.1). GUESS is a sparse Bayesian multiple linear regression method in which one outcome variable is regressed against all SNPs to identify the minimum (non-redundant) set of SNPs that predict the variability. For each SNP-outcome variable pair, GUESS returns a Marginal Posterior Probability of Inclusion (MPPI) which can be interpreted as the posterior strength of association between a single SNP and the outcome variable (Bottolo et al., 2013). Thus, we map the expression levels of the genes in each co-expression module summarized by the first principal component (1st PC). This first principal component was computed on the covariate-adjusted normalized counts by using the R function *prcomp.* Independent jobs of the algorithm were run for each co-expression module, each time for 20,000 sweeps and 5,000 burn in. From the output MPPI, we computed the Bayes Factor (BF) for each 1st PC-SNP pair. BF is defined as the ratio between the posterior and prior odds. The prior odds in GUESS is defined as $\pi = {E(p_g)}/{p}$, where p is the input number of SNPs and $E(p_g)$ is the expected number of control points for the gth outcome (in our case the first principal component of the co-expression module). In GUESS, $E(p_g)$ is set by default to 2. Thus the BF formula becomes: $= \frac{MPPI_{gj}/(1-MPPI_{gj})}{\pi/(1-\pi)}$, where $MPPI_{gj}$ is the marginal posterior probability of inclusion for the gth outcome and the ith SNP. The BF of the most highly associated SNP to each co-expression module can be found in Supplemental Table 3 (only modules with a BF higher than 100 are shown). Locus fine mapping was carried out for the module with the strongest genetic control point (macrophage-mediated healing network or M30). We used Hierarchical Evolutionary Stochastic Search (HESS) model (Bottolo et al., 2011; Lewin et al., 2016), which implements a hierarchical regression model in a Bayesian framework using a stochastic search algorithm. This allows jointly regressing a set of response variables (i.e. genes in a co-expression module) against a set of SNPs. As output HESS computes an individual MPPI for each gene and SNP, HESS was run for 25,000 sweeps and 5,000 burn in. We mapped the individual expression levels of all genes in M30 (VST gene counts after adjusting for covariates effects) to the SNPs present in a $\pm$1Mb window from the location of the most significant SNP identified by ESS analysis for M30 (region mapped: mouse chromosome 17 from 43,997,787 to 46,194,647, comprising 14 SNPs in our data). The MPPI for each gene and SNP can be found in Supplemental Table 4 (find also annotation of M30 gene network).

TAD computation

Processed Hi-C data for mouse ESC (Bonev et al., 2017) and human macrophages (Phanstiel et al., 2017) were obtained from JuiceBox (Durand et al., 2016). TADs were identified using the directionality index calling algorithm implementation in tadtool (Kruse et al., 2016). Gviz and GenomicInteractions (Harmston et al., 2015) were used for visualisation purposes.

Functional analysis of macrophage-mediated healing network MMHN

We predicted Runx2 transcription factor binding sites (TFBS) in the promoter of the genes in MMHN by using the R package TFBSTools 1.10.3 (Tan and Lenhard, 2016). This package queries JASPAR database and provides tools to predict TFBS in a list of provided sequences. As there is no matrix for the *Runx2* mouse gene in JASPAR database, we investigated one-to-one human orthologs of all murine genes present in the module and computed genes with binding sites for *RUNX2* human gene. Promoter sequences were defined as 200 bp upstream of the 5' flanking region of each gene. Ortholog genes and promoter sequences were retrieved from Ensembl v74 using the R library biomaRt (Durinck et al., 2009). In the function *searchSeq* of TFBSTools package, both strands were considered (strand parameter="*") and the minimum score was set to 80%. This resulted in 70 genes predicted to carry a RUNX2 TFBS in the module.

To inspect the relationship between Runx2 transcriptional regulation and healing, we first correlated the VST gene counts of the genes in the module with Runx2 expression levels. We then correlated the VST gene counts of the module genes with the negative sex and year-adjusted average ear areas (i.e. rate of wound healing). The correlation of these two outputs resulted in $\rho = 0.74$ ($P = 1.04^{-31}$). The correlations (and P-values) were computed with the R function *corAndPvalue*. In all cases Spearman's ranked correlations (ρ) were taken into account.

The 177 genes of the module were entered into STRING protein-protein interaction database 10.0 (Szklarczyk et al., 2015) (queried on the 27/02/2018). Experimental, co-expression and databases connections with a minimum interaction score of 0.15 were retrieved and the largest connected component was visualized using Cytoscape (Smoot et al., 2011) (Figure 3A). In the module graph, genes annotated with the functional term "lipid biosynthetic process" (22 genes) were colored in green and genes predicted to have a RUNX2 TFBS were highlighted with yellow border color. Gene node size was mapped to the probability of association of each gene with the SNP chr17_45131552 (HESS output).

Identification of candidate trans-regulators of co-expression modules by nearest gene approach

We inspected *trans*-regulatory genes for the transcriptional programs through an associated SNP (BF>100) in the GUESS Bayesian analysis (Supplemental Table 3). We carried out the nearest gene approach and annotated each of these networks with the nearest gene expressed in our macrophage RNA-seq data (i.e. candidate *trans*-regulator gene, n=146, 10,893 genes). To further inspect the association of the candidate gene transcriptional regulation and healing, we first correlated the VST gene counts of the genes in the module with candidate *trans*-regulator gene expression levels. We then correlated the VST gene counts of the module genes with the sex and year-adjusted average ear areas (i.e. rate of wound healing). These results are included in Supplemental Table 3. For the top 5 modules (ranked by absolute correlation with healing and without considering M30), we visualised these correlations (Supplemental Figure 2, right) and further inspected known connections between the genes in the Modules and the candidate *trans*-regulatory genes (Supplemental Figure 3). In this analysis, we input all genes in each of these 5 modules in addition to the *trans*-regulatory candidate gene into STRING protein-protein interaction database 10.0 (Szklarczyk et al., 2015) (queried on the 27/11/2018). Experimental, co-expression, database and text-mining connections with a minimum interaction score of 0.15 were retrieved and the largest connected component was visualized using Cytoscape (Smoot et al., 2011) (Supplemental Figure 2, right). In each graph, genes annotated with the top enriched functional term in the module (i.e. most significant FDR as in Supplemental Table 2) were colored in green. The *trans*-regulatory candidate gene was highlighted with diamond shape and orange color.

In vitro blockage of Runx2 and qRT-PCR

BMDMs were cultured using tibias and femurs isolated from C57BL/6J mice for 4 days in Petri dishes, after which they were seeded in 6-well plates (0.5 million cells/well). The following day cells were treated with CADD522 for 48h (Chembridge Corporation) at a concentration of 20 μM in full culture medium.

Complementary DNA (cDNA) was obtained from 500 ng of total RNA using the Bio-Rad iScript kit (Bio-Rad, UK) according to manufacturer's instructions. qRT-PCR reactions were performed using the Viaa 7 Real-Time PCR system (Life Technologies). A total of 10 ng of cDNA per sample was used for PCR using Brilliant II SYBR Green qPCR Master Mix (Agilent). QuantStudio Real Time PCR Software (Life Technologies) was used for the determination of Ct values. Results were analyzed using the comparative Ct method (Schmittgen and Livak, 2008) and each sample was normalized to the reference gene (HPRT), to account for any cDNA loading differences.

Wound healing and histological analysis

13-week-old Lewis (LEW) rats were divided into two groups as controls (n=6) and cerulenin treated (n=8) animals. Hair was removed from the back of the rats using a depilatory cream at least one day prior to surgery. Surgeries were performed under anesthesia using pre-operative analgesic (0.1 mg/kg Buprenex). One 10mm full thickness wound was excised from the dorsum of the rat using a biopsy punch along the midline. A donut-shaped silicone splint was placed around the wound and attached to the skin with interrupted sutures. Splints are required to promote healing via epithelialization rather than contraction. Following secure attachment, the wounds/splints were covered with appropriate Tegaderm dressings to minimize the risk of infection. Cerulenin (Sigma Aldrich, UK) (300µg in 100µl of propylene glycol) was applied topically at days 0, 3 and 6 after excision. Wounds were monitored and imaged every 3 days and wound tissue was collected at day 8 after excision for total RNA extraction and histological analysis.

Tissues were fixed in formalin for 48h. H&E slides and unstained slides were taken at all levels for the analysis of the healing tissue. Data analysis was performed using Image J. CD68 immunohistochemistry was performed on paraffin-embedded sections with rat anti-CD68 antibody (Biorad) and developed using EnVision+ System-HRP (K4007, Dako). Pictures were taken with Leica Microscope Camera DFC7000T. Pictures were further merged using Adobe Photoshop and analysed using Image J software. Reported values represent the average of the quantification of 5 different High Power Field (HPF) per animal.

Data availability

Mouse macrophage RNA-seq data has been deposited at GEO database under accession number GSE112171. Phenotype data is available in Supplemental Table 1.

Statistics overview

Two-tailed Student's t test was used in the experimental comparisons. See statistical methods used in the RNA-seq data analysis in the section “RNA extraction and RNA sequencing (RNA-seq)**”.**

Study approval

This study was performed in accordance with the Home Office Guidance on the Operation of the Animals (Scientific Procedures) Act 1986, published by Her Majesty's Stationery Office (London, United Kingdom). All animal protocols were approved both by Imperial College's Animal Welfare and Ethical Review Body (AWERB) and the Home Office.

References

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

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Barutcu, A.R., Tai, P.W.L., Wu, H., Gordon, J.A.R., Whitfield, T.W., Dobson, J.R., Imbalzano, A.N., Lian, J.B., van Wijnen, A.J., Stein, J.L., et al. (2014). The bone-specific Runx2-P1 promoter displays conserved three-dimensional chromatin structure with the syntenic Supt3h promoter. Nucleic Acids Res. *42*, 10360–10372.

Behmoaras, J., Diaz, A.G., Venda, L., Ko, J.-H., Srivastava, P., Montoya, A., Faull, P., Webster, Z., Moyon, B., Pusey, C.D., et al. (2015). Macrophage epoxygenase determines a profibrotic transcriptome signature. J. Immunol. *194*, 4705–4716.

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.-P., Tanay, A., et al. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell *171*, 557–572.e24.

Bottolo, L., Petretto, E., Blankenberg, S., Cambien, F., Cook, S.A., Tiret, L., and Richardson, S. (2011). Bayesian detection of expression quantitative trait loci hot spots. Genetics *189*, 1449–1459.

Bottolo, L., Chadeau-Hyam, M., Hastie, D.I., Zeller, T., Liquet, B., Newcombe, P., Yengo, L., Wild, P.S., Schillert, A., Ziegler, A., et al. (2013). GUESS-ing polygenic associations with multiple phenotypes using a GPU-based evolutionary stochastic search algorithm. PLoS Genet. *9*, e1003657.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. *3*, 99–101.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191.

Harmston, N., Ing-Simmons, E., Perry, M., Barešić, A., and Lenhard, B. (2015). GenomicInteractions: An R/Bioconductor package for manipulating and investigating chromatin interaction data. BMC Genomics *16*, 963.

Harmston, N., Ing-Simmons, E., Tan, G., Perry, M., Merkenschlager, M., and Lenhard, B. (2017). Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. Nat. Commun. *8*, 441.

Kruse, K., Hug, C.B., Hernández-Rodríguez, B., and Vaquerizas, J.M. (2016). TADtool: visual parameter identification for TAD-calling algorithms. Bioinformatics 32, 3190–3192.

Lai, P.-C., Chiu, L.-Y., Srivastava, P., Trento, C., Dazzi, F., Petretto, E., Cook, H.T., and Behmoaras, J. (2014). Unique regulatory properties of mesangial cells are genetically determined in the rat. PLoS One 9, e111452.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559.

Lewin, A., Saadi, H., Peters, J.E., Moreno-Moral, A., Lee, J.C., Smith, K.G.C., Petretto, E., Bottolo, L., and Richardson, S. (2016). MT-HESS: an efficient Bayesian approach for simultaneous association detection in OMICS datasets, with application to eQTL mapping in multiple tissues. Bioinformatics *32*, 523–532.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Nicod, J., Davies, R.W., Cai, N., Hassett, C., Goodstadt, L., Cosgrove, C., Yee, B.K., Lionikaite, V., McIntyre, R.E., Remme, C.A., et al. (2016). Genome-wide association of multiple complex traits in outbred mice by ultra-low-coverage sequencing. Nat. Genet. *48*, 912–918.

Phanstiel, D.H., Van Bortle, K., Spacek, D., Hess, G.T., Shamim, M.S., Machol, I., Love, M.I., Aiden, E.L., Bassik, M.C., and Snyder, M.P. (2017). Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. Mol. Cell *67*, 1037–1048.e6.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. *44*, W83-9.

Robertson, A.J., Larroux, C., Degnan, B.M., and Coffman, J.A. (2009). The evolution of Runx genes II. The C-terminal Groucho recruitment motif is present in both eumetazoans and homoscleromorphs but absent in a haplosclerid demosponge. BMC Res. Notes *2*, 59.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. *3*, 1101–1108.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics *27*, 431–432.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447-52.

Tan, G., and Lenhard, B. (2016). TFBSTools: an R/bioconductor package for transcription factor binding site analysis. Bioinformatics *32*, 1555–1556.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., and Weir, B.S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics *28*, 3326–3328.