Title: Sequence-Dependent DNA Shape Features Regulate Chromatin Accessibility in Alternatively Activated Macrophages

Authors: Mei San Tang[1,6§], Emily R. Miraldi[2,3], Natasha M. Girgis[1], Richard A. Bonneau[4,5], P'ng Loke[1§]

Affiliations:

[1]Department of Microbiology, New York University School of Medicine, New York, NY 10016, USA.

[2] Divisions of Immunobiology and Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA.

[3] Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, 45267, USA.

[4] Department of Biology, Center for Genomics and Systems Biology, New York University, New York, NY 10003, USA.

[5] Simons Center for Data Analysis, Simons Foundation, New York, NY 10011, USA.

[6] Current address: Department of Pathology and Immunology, Washington University in St Louis, Saint Louis, MO 63110, USA.

[§]Correspondence to:
meisan@wustl.edu
Png.Loke@nyulangone.org

Abstract: Interleukin-4 (IL-4) activates macrophages to adopt a distinct phenotype important for tissue repair and helminth infection, but the molecular basis of chromatin remodeling in response to IL-4 stimulation is not understood. We find that IL-4 activation of terminally differentiated macrophages in mice is accompanied by macrophage-specific chromatin remodeling in regions enriched with binding motifs of similar transcription factors, including the pioneer factor PU.1. However, the PU.1 motifs display macrophage-specific sequence features, including macrophage-specific DNA shape profiles. Mutation studies based on natural genetic variations between C57BL/6 and BALB/c mouse strains demonstrate that accessibility of these IL-4 induced regions can be regulated through differences in DNA shape, where accessibility is only disrupted through a sufficiently large change in DNA shape. Furthermore, the chromatin remodeling process following long-term IL-4 stimulation is accompanied by a change in response to a secondary IL-4 stimulation. We propose a model whereby DNA sequence variations alter accessible chromatin landscape by inducing a large DNA shape change, leading to differences in secondary responses with different host genetics.

Introduction

Macrophage activation is a process by which macrophages transition from a resting state to adopt different phenotypes, in response to specific external stimuli that can either be danger signals or homeostatic and metabolic signals (*1*). The macrophage activation process is accompanied by changes in transcriptional activities and histone modifications genome-wide, orchestrated by combinatorial actions of different transcription factors (TFs) that include lineage-determining pioneer factors (LDTF) (such as PU.1) and stimulus-dependent TFs (such as the STAT and IRF proteins) (*1-5*). Such molecular events have been most extensively characterized for bone marrow derived macrophages (BMDMs) in response to toll-like receptor (TLR) signaling, as a reductionist model for the type 1 immune response to bacterial infections. In contrast, alternatively activated macrophages (AAMs) induced by the type 2 cytokines interleukin-4 (IL-4) and IL-13, adopt a distinct phenotype that can promote expulsion and resolution of helminth infections, as well as tissue repair during chronic inflammation (*6*). Notably, different lineages and phenotypes of tissue macrophages will respond differently to IL-4 stimulation *in vivo* (*7, 8*). Hence, a key question is how and why do different types of macrophages respond differently to the same stimuli? The accessible chromatin landscape of terminally differentiated cells, shaped during differentiation, can determine cellular functions and phenotype (*9, 10*). Different types of tissue macrophages with distinct functions have distinct accessible chromatin profiles (*11*). However, the remodeling of the chromatin landscape in response to external stimulations for different macrophages is poorly understood, especially in the context of alternative activation in response to type 2 cytokines.

We address this question here, where we induced a state of alternative activation using IL-4 stimulation on different types of mouse peritoneal macrophages and used ATAC-Seq (*12*)

64 to profile the open chromatin landscape in these macrophages. We observed that IL-4 activation
65 is accompanied by cell-type-specific chromatin remodeling in regions enriched with similar TF
66 DNA binding motifs, but these similar TF motifs have macrophage-specific sequence
67 characteristics. Notably, PU.1 and AP-1 motifs in the IL-4 induced regions of AAM^{res} and
68 AAM^{mono} have different DNA shape profiles. Leveraging the genetic variations between
69 C57BL/6 and BALB/c mice to perform mutational studies on the accessible regions in AAMs,
70 we demonstrate that regulatory regions will only experience changes in accessibility if they
71 undergo sufficiently large changes in DNA shape induced by sequence variants. In a recent study
72 that characterized TLR4 activation in BMDMs derived from different mouse strains, "silent
73 mutations" were observed in 20-95% of accessible chromatin regions (i.e. chromatin regions that
74 contained sequence variants but experienced no change in chromatin accessibility) but were
75 unexplained, despite the large variety of genomic approaches used (*13*). We propose that DNA
76 shape features may be an important contributor to this unexplained observation.

77 Furthermore, long-term exposure to IL-4 stimulation not only leads to a change in the
78 accessible chromatin landscape, but is also accompanied by persistent changes in cellular
79 phenotype, including a dampening of response to a repeat IL-4 stimulation. While there have
80 been considerable number of studies that described the opposing effects of different cytokines
81 (*14, 15*), none has so far demonstrated this dampening of response with repeated stimulations of
82 the same kind. Notably, this dampening in response to the second IL-4 stimulation occurs to
83 different degrees in C57BL/6 and BALB/c mouse strains. In light of the increasing interest in
84 macrophage alternative activation, it is important to realize that much of what has been defined
85 for alternative activation (such as the expression of *Arg1*) has been based on studies on the

86 C57BL/6 mouse strain and many of these findings might not hold true on a different genetics
87 background.

88

89 **Results**

90

91 *IL-4 stimulation leads to remodeling of open chromatin landscape in peritoneal macrophages*

92 To examine chromatin remodeling on different types of tissue macrophages, we injected
93 recombinant IL-4–antibody complex (IL-4c) into the peritoneal cavity of C57BL/6 mice to
94 induce accumulation of alternatively activated F4/80^{hi}CD206^{-} macrophages of embryonic origin
95 (AAM^{res}) and compared these with F4/80^{int}CD206^{+} macrophages derived from
96 Ly6C^{hi} inflammatory blood monocytes (AAM^{mono}) from mice injected with IL-4c and
97 thioglycollate (*8*). We then used ATAC-seq (*12*) to profile the open chromatin landscape of these
98 macrophages, in comparison to unstimulated F4/80^{hi}CD206^{-} macrophages of naïve mice and
99 F4/80^{int}CD206^{+} macrophages from thioglycollate-treated mice (*8*).

100 The overall differences in accessible chromatin landscape (a total of 61,713 open
101 chromatin regions) could be attributed mainly to the type of macrophage (27% of total variance),
102 but alternative activation by IL-4 also altered the accessible chromatin profiles (Figures 1A, B).
103 *Arg1* and *Ucp1*, which are known to be IL-4 inducible (*8*), had constitutively accessible
104 chromatin regions, whereas *Retnla*, another IL-4 inducible gene, had chromatin regions that
105 gained accessibility in response to IL-4 (Figure 1B). This IL-4 induced chromatin remodeling
106 process can be cell-type-specific at certain regions (e.g. regions adjacent to the loci of *Tgfb2*,
107 *Ccl2*) (Figure 1B). Of the 61,713 total accessible regions, we identified 1572 regions induced by
108 IL-4 for AAM^{res} and 1462 regions for AAM^{mono} (Figure S1A). IL-4-dependent regions also had

109 the largest contribution to the differences in open chromatin profiles between unstimulated and
110 IL-4 stimulated macrophages (Figure 1C).

111 The IL-4 induced regions almost all (99% in AAM^{res} and 97% in AAM^{mono}) gained
112 accessibility from undetectable levels at baseline (Figure 1D). When compared to constitutively
113 accessible regions (Figure S1B), IL-4 induced regions have distinct sequence properties. They
114 were more likely to be non-coding intronic elements (binomial enrichment, $p = 2.53 \times 10^{-5}$ in
115 AAM^{res} and $p = 8.54 \times 10^{-4}$ in AAM^{mono}) (Figure 1E). In the monocyte-derived macrophages, IL-
116 4 induced regions were closer to IL-4 induced genes (two-sided Mann-Whitney test, $p = 3.41 \times$
117 10^{-25} in AAM^{mono}) (Figure 1F). In tissue-resident macrophages, IL-4 induced regions were lower
118 in GC content (two-sided Mann-Whitney test, $p = 8.65 \times 10^{-46}$ in AAM^{res}) (Figure 1G, left).
119 These regions were also less likely to overlap with a CpG island (Figure 1G, right). These
120 differences between constitutively accessible and IL-4 induced regions were observed in both
121 monocyte-derived and tissue resident macrophages.

122

123 *IL-4 induced regions contain similar TF motifs with macrophage-specific sequence*
124 *characteristics*

125 Even though both AAM^{mono} and AAM^{res} received the same stimulation within the
126 peritoneal tissue environment, the regions that were remodeled by IL-4 were largely
127 macrophage-specific (Figure 2A). Of all the 2855 IL-4 induced regions, only 179 regions (6% of
128 total IL-4 induced regions) were shared between both AAM^{mono} and AAM^{res} (Figure S1C).
129 While the IL-4 induced regions from AAM^{mono} and AAM^{res} were largely distinct, the DNA
130 motifs discovered from these distinct regions were grouped into similar families of TFs, which
131 included PU.1, KLF and the AP-1 family of motifs (Figure 2B). We next used an over-

132 representation approach to further identify additional motifs that were selectively enriched in the
133 IL-4 induced regions of AAM^{res} vs. AAM^{mono} (Data S1). Since there were some overlaps
134 between the motifs discovered by the *de novo* discovery method and over-representation method,
135 we combined the two sets of motifs for a clustering analysis and merged motifs that were
136 redundant. Using this approach, we were able to identify macrophage-specific motifs that were
137 not initially identified by the *de novo* motif discovery method (Figure 2C). Motifs of the GATA
138 family and of the bHLH class (primarily resembling MYC, with consensus sequence CACGTG),
139 were specific to IL-4 induced regions of AAM^{res}. In contrast, the NFY and STAT motifs were
140 specific to IL-4 induced regions of AAM^{mono}. However, these macrophage-specific motifs were
141 only detected in a minority of IL-4 induced regions (236 of 1572 IL-4 induced peaks in AAM^{res}
142 and 317 of 1462 IL-4 induced peaks in AAM^{mono}), while the PU.1, KLF and AP-1 motifs
143 discovered by the *de novo* method were present in approximately 75% of IL-4 induced peaks in
144 AAM^{res} and AAM^{mono}. This suggests that the families of TFs involved in the IL-4 induced
145 chromatin remodeling process (PU.1, KLF and AP-1) were largely similar across AAM^{res} and
146 AAM^{mono}.

147 We further delineated the proportion of accessible regions containing different
148 combinations of the PU.1, KLF and AP-1 motifs. Overall, 50% of IL-4 induced regions
149 contained only a single TF motif (one of either PU.1, KLF or AP-1 only) while 25% of IL-4
150 induced regions contained multiple TF motifs (Figure 2D, left panel). This trend was shared
151 across AAM^{res} and AAM^{mono}. However, AAM^{res} had significantly higher number of accessible
152 regions with the KLF motif (two-sided Fisher's exact test, $p = 9.21 \times 10^{-8}$), while AAM^{mono} had
153 significantly higher number of accessible regions with the AP-1 motif (two-sided Fisher's exact
154 test, $p = 9.59 \times 10^{-10}$), suggesting the preferential use of these TFs by AAM^{res} vs. AAM^{mono}

155 during chromatin remodeling in the context of IL-4 induced alternative activation (Figure 2D,
156 right panel). Such differences were not observed with the PU.1 motifs. We next examined the
157 expression levels of these TF families to determine if specific members within each family could
158 be differentially expressed between AAM^{res} and AAM^{mono}. The 21 TFs of PU.1, KLF and AP-1
159 families that were highly expressed in peritoneal macrophages almost all demonstrated lineage-
160 specific expression, both at baseline (Clusters 1 and 2) and with IL-4 stimulation (Clusters 3 and
161 4) (Figure 2E, Figure S2A).

162 Members of the same TF families recognize and bind to similar short DNA sequences
163 genome-wide to cause downstream cellular changes. Since IL-4 induced chromatin remodeling
164 occurred at largely distinct sites in AAM^{res} vs. AAM^{mono}, additional sequence characteristics
165 beyond the DNA motif sequences must be associated with the macrophage-specific IL-4 induced
166 regions. While PU.1 is better characterized as a pioneer factor that shapes the chromatin
167 landscape during macrophage development (*16, 17*), it can also unveil latent enhancers during
168 macrophage activation, through cooperative binding with stimulus-dependent TFs (such as
169 STAT1, STAT6) (*4*). In this study, it was the motif most frequently discovered from IL-4
170 induced regions (Figure 2D, right panel). We focused on these predicted PU.1 binding sites and
171 characterized their local sequence features in both types of macrophages. We first quantified the
172 accuracy of PU.1 motif prediction with actual PU.1 binding, by comparing our predicted PU.1
173 motifs from thioglycollate-elicited macrophages with ChIP-seq-defined PU.1 binding sites of
174 thioglycollate-elicited macrophages (*5, 18*). We found that 78% of the PU.1 motif sites predicted
175 from thioglycollate-induced macrophages in our study (4,282 of total 5,492 predicted PU.1 motif
176 sites) overlapped with a PU.1 binding site defined by ChIP-seq (Figure S2B).

177 The PU.1 motifs discovered from the IL-4 induced regions of AAM^{res} and AAM^{mono} were
178 prominently different in the nucleotides flanking the core GGAA nucleotides, which were
179 recognized as the nucleotides directly involved in TF-DNA binding (*19, 20*) (Figure 3A). In
180 particular, the core-flanking nucleotides on the 5' end of the PU.1 motif in AAM^{mono} appeared to
181 have higher AT content. We asked if this was likely due to different collaborating TFs binding
182 adjacent to PU.1 in AAM^{res} vs. AAM^{mono}. To examine this possibility, we selected only for
183 regions of the IL-4-induced PU.1 motif along with ±25bp flanking sequences from both
184 AAM^{mono} and AAM^{res} (Figure S2C) and determined which TF motifs were specific for AAM^{res}
185 vs. AAM^{mono}. We observed a greater diversity in TF families specifically enriched around the IL-
186 4 induced PU.1 motifs of AAM^{mono} (Figure 3B). We also found that different motifs from the
187 family of STAT factors were selectively enriched around the IL-4 induced PU.1 regions from
188 AAM^{res} vs. AAM^{mono} (Figure 3B), suggesting that while STAT proteins could similarly bind
189 cooperatively with PU.1 during the IL-4 induced chromatin remodeling process in both types of
190 macrophages, there could be macrophage-specific variations in the STAT proteins that were
191 involved with this process.

192 Since functional PU.1 binding can also be accurately determined by the local three-
193 dimensional shape of the DNA at PU.1 binding sites (*20*), we next determined if such sequence
194 properties could be associated with differences in IL-4 induced PU.1 regions between AAM^{mono}
195 and AAM^{res}. We computationally predicted four DNA shape configurations (minor groove width
196 - MGW, propeller twist – ProT, helical twist – HelT and roll) at the IL-4 induced PU.1 regions
197 (*21*) and found that IL-4 induced PU.1 regions of AAM^{mono} and AAM^{res} differed significantly in
198 shape profiles (Figure 3C). In addition, the motif scores (commonly used as a proxy of TF-DNA

199 binding affinity) of PU.1 in AAM^{res} were significantly higher than that in AAM^{mono} (two-sided
200 Mann-Whitney test, $p = 3.00 \times 10^{-8}$) (Figure 3D).

201 To determine if cell-type-specific local DNA shape of PU.1 binding sites could be
202 observed in an independent setting, we predicted DNA shape profiles of PU.1 binding sites
203 defined by ChIP-seq from different cell types (splenic B cells, thioglycollate-elicited-peritoneal
204 macrophages and a microglial cell line) (*22, 23*) (Figure S2D). The shape of PU.1 binding sites
205 specific to each of these cell types were significantly distinct (Figure 3E). These results point
206 towards the contribution of local DNA shape on the cell-type-specific pattern of PU.1 motif
207 accessibility and binding.

208 Finally, we also characterized local sequence features for the KLF and AP-1 family
209 motifs. While KLF motif scores were significantly higher in AAM^{mono} (two-sided Mann-
210 Whitney test, $p < 2.2 \times 10\text{-}16$), there were no differences in the DNA shape around the KLF
211 motifs of AAM^{res} vs. AAM^{mono} (Figure S3A, S3C). In contrast, AP-1 motifs appeared to have
212 similar motif scores in AAM^{res} and AAM^{mono}, but displayed macrophage-specific DNA shape in
213 the propeller twist configuration (two-sided Kolmogorov-Smirnov test, $p = 3.91 \times 10^{-9}$) (Figure
214 S3B, S3D). Together, all these findings suggest that, while chromatin remodeling during
215 alternative activation of peritoneal macrophages from different lineages are associated with TFs
216 from the same families, the lineage-specific chromatin remodeling pattern is most likely
217 achieved by recognition of distinct sequence features flanking the short nucleotide sequences
218 directly involved in TF-DNA binding. However, depending on the TF family, the sequence
219 features used to recognize potential binding sites could be different.

220

221 *Accessibility of IL-4 induced PU.1 regions can be altered through DNA shape change induced*
222 *by sequence mutation, without disruption of the PU.1 motif*

223 To determine whether altering the local DNA shape would change accessibility of the
224 PU.1 motifs, we designed a mutation study using natural variants between the C57BL/6 and
225 BALB/c mouse strains, by performing ATAC-seq on AAM^{mono} and AAM^{res} generated from mice
226 of these two different genetics backgrounds. We first defined "strain-specific" (significant
227 differences in accessibility counts between C57BL/6 and BALB/c at FDR 10%) or "strain-
228 common" (no significant differences in accessibility count between C57BL/6 and BALB/c) IL-4
229 induced PU.1 regions from both AAM^{mono} and AAM^{res}. We next focused only on strain-specific
230 IL-4 induced PU.1 regions with SNP(s) flanking the PU.1 motif (Figure S4A). This allowed us to
231 ask if the differences in chromatin accessibility between strains were only due to the flanking
232 sequence mutation(s) and not confounded by disruption of a potential PU.1 binding site. We then
233 correlated these strain-specific differences in accessibility with phenotypic differences in gene
234 expression.

235 We highlight this DNA shape dependent regulatory mechanism at two independent
236 genomic regions that we infer to be regulating biologically relevant functions. Our first example
237 is an intergenic accessible region in AAM^{res}, where two different SNPs occurring at the 19th and
238 24th base pairs upstream of a PU.1 motif led to changes in DNA shape, most drastically with the
239 ProT and HelT configurations (Figure 4A, Table 1). This was accompanied by decrease in
240 chromatin accessibility counts ($\log_2$ fold change BALB/c vs. C57BL/6 = -2.57, adjusted p-value
241 = 0.01) and reduced expression of the nearby *Slc30a4* gene in the BALB/c mice ($\log_2$ fold
242 change BALB/c vs. C57BL/6 = -3.43, adjusted p-value = 0.006). *Slc30a4* is a zinc exporter that
243 affects the survival of *Histoplasma capsulatum* in IL-4 stimulated macrophages (*24*). Our second

244 example is an intronic region on the *P4hb* gene in AAM^{mono}, where SNPs occurring at the 8th and
245 24th base pairs upstream of a PU.1 motif led to changes in DNA shape, with the former causing
246 larger shifts in the MGW and roll configurations (Figure 4B, Table 1). This was, however,
247 accompanied by increased accessibility instead ($\log_2$ fold change BALB/c vs. C57B6L/6 = 1.87,
248 adjusted p-value = 0.0085) and increased expression of the *P4hb* gene ($\log_2$ fold change BALB/c
249 vs. C57BL/6 = 0.34, adjusted p-value = 0.048).

250 Overall, PU.1 motif-flanking variants occurred at higher frequency then variants located
251 directly within the motif (24.3% vs. 5.4% in AAM^{res} and 20.7% vs. 6.9% in AAM^{mono}) (Figure
252 4C). This suggests that DNA shape readout (a proxy for local environment), instead of direct
253 base readout during PU.1 binding, is likely to be a more important sequence property that could
254 influence PU.1 binding. The frequency of local variants was also greater in strain-specific
255 genomic elements when compared to strain-common regions (two-sided Fisher's exact test, $p =$
256 4.59×10^{-4} in AAM^{res}, $p = 8.12\times10^{-6}$ in AAM^{mono}), even though strain-common PU.1 motifs could
257 also have sequence variants present within ±25bp of the PU.1 motif (Figures 4D; Table 2).

258

259 *A sufficiently large change in DNA shape is required to alter accessibility of chromatin regions*

260 To identify potential mechanisms that could alter PU.1 accessibility without disrupting
261 the nucleotide sequences directly involved with TF protein binding, we asked if the SNPs
262 flanking PU.1 motif could reside in non-PU.1 TF motifs, since PU.1 binding events are likely
263 associated with a combinatorial set of multiple different TFs. While we found that most of the
264 SNPs flanking the PU.1 motifs did reside in TF motifs, the proportion of motif-disrupting SNPs
265 appeared to have no significant differences between strain-specific and strain-common PU.1

266 regions (Table 3), suggesting that disruption of motifs for collaborating TFs alone was unlikely
267 to be the sole mechanism of chromatin accessibility alteration.

268 We next determined if there could be differences in the magnitude of DNA shape change
269 experienced by strain-specific vs. strain-common regions. To answer this question, we identified
270 18 strain-specific and 63 strain-common IL-4 induced PU.1 regions across both AAM^{res} and
271 AAM^{mono} that all contained SNPs flanking the PU.1 motif (but not within the motif itself) and
272 determined the magnitude of DNA shape change induced by the sequence variants in each of
273 these IL-4 induced PU.1 regions (Figure S4B). We found that strain-specific IL-4 induced PU.1
274 regions underwent a greater magnitude in DNA shape change induced by the presence of SNPs
275 (Figure 5A, Table 4), particularly in the DNA minor groove width (two-sided Mann-Whitney
276 test, $p = 0.03$). In other words, simply having sequence variants alone was insufficient to cause
277 an alteration in chromatin accessibility, but the sequence variants must cause a sufficiently large
278 change in DNA shape in order to alter the accessibility profile of a chromatin region.

279 Based on the observation made on sequence variants proximal to PU.1 motif, we further
280 extended the analysis to a genome-wide level to determine the features of accessible regions
281 most likely to be affected by presence of sequence variants. We identified all accessible regions
282 carrying SNP(s), regardless of the presence of a PU.1 motif or IL-4 inducibility, and categorized
283 them into 920 strain-specific regions and 4769 strain-common regions, using the definition
284 described above (Figure S4B). The strain-specific regions had a significantly greater change in
285 DNA shape secondary to sequence variants and this was most significant in the MGW, ProT and
286 roll DNA shape configurations (Figure 5B), similar to what was observed for IL-4 induced PU.1
287 regions, albeit with greater levels of statistical significance due to the higher number of regions
288 used for comparisons. We also asked if the degree of change in DNA shape could be associated

289 with differences in SNP frequency. Indeed, SNP frequency positively correlated with the amount
290 of SNP-induced DNA shape change, especially for the shape features MGW and ProT (Figure
291 5C). Hence, the observation that chromatin regions must undergo a sufficiently large change in
292 DNA shape for alteration of accessibility to occur can be generalized across the genome-wide
293 chromatin landscape in AAMs.
294
295 *IL-4 induced accessible regions contribute more to the differences in chromatin landscape*
296 *between C57BL/6 and BALB/c AAMs*
297 We next used a logistic regression model to systematically examine the contribution of
298 various genomic features (IL-4 inducibility, % GC, SNP frequencies and whether the region was
299 non-coding) in altering chromatin accessibility (Figure S5A). We found regions with higher SNP
300 frequency were more likely to have strain-specific accessibility, which was consistent with our
301 findings described above (ANOVA p-value $< 2.2 \times 10^{-16}$, Figure S5A). However, we
302 unexpectedly identified IL-4 inducibility as another significant genomic feature predictive of
303 strain-specific accessibility (ANOVA p-value $= 4.97 \times 10^{-5}$, Figure S5A). In other words,
304 accessible regions that were IL-4-inducible were more likely to be strain-specific.
305 76 of the 920 strain-specific regions (8.26%) examined in this analysis were IL-4 induced
306 regions, as compared to 231 of 4769 strain-common regions (4.84%) (two-sided Fisher's exact
307 test, $p = 6.28 \times 10^{-5}$). We also examined all strain-specific regions (regardless of the presence of
308 sequence variants) and found enrichment for IL-4 induced regions (Figure S5B). Together, these
309 findings suggest that stimulation-dependent genomic elements were more likely to contribute to
310 differences in the accessible chromatin landscape of activated macrophages on different genetics

311 background, further supporting the functionality of these IL-4-induced genomic elements during
312 alternative activation.
313
314 *AAMs from C57BL/6 and BALB/c mice are functionally distinct*
315 We next examined whether macrophages from different genetic macrophages were functionally
316 distinct, by comparing the transcriptional profiles of C57BL/6 and BALB/c AAMs. Even though
317 most of the differences in transcriptional profiles were predominantly driven by the types of
318 macrophages (45% of total variance), strain differences also contributed to the considerable
319 variation in transcriptional profiles (18% of total variance) (Figure 6A). Consistent with this
320 finding, most of the macrophage-specific functions were conserved across mouse strains and not
321 affected by differences in genetics backgrounds (Figure 6B, left panel). In contrast, functional
322 differences secondary to genetics were largely specific to the different types of macrophages
323 (Figure 6B, right panel). We next examined the strain-specific functional differences in AAM^{res}
324 and AAM^{mono}, respectively (Figure 6C). Notably, BALB/c AAM^{res} expressed lower levels of
325 cell-cycle-related genes, in line with the previously reported observation that peritoneal AAM^{res}
326 have lower proliferation capacity during *Litomosoides sigmodontis* infection(*25*). Furthermore,
327 the expressions of PD-L2 (*Pdcd1lg2*) (Figure 6D) and MHCII molecules (Figure 6E), which are
328 cellular markers typically used to characterize alternative activation (*8*) in AAM^{mono} of C57BL/6
329 background, were significantly reduced in BALB/c AAM^{mono}. Notably, while 3 of the 5 MHCII
330 genes (*H2-Aa*, *H2-Ab1*, *H2-Eb1*) had significantly higher expression in C57BL/6 AAM^{mono}, *H2-*
331 *Ea-ps* expression was specific to BALB/c AAM^{mono}, indicating differences in antigen
332 presentation between C57BL/6 and BALB/c AAM^{mono}.
333

334 *AAMs from C57BL/6 and BALB/c mice respond differently to secondary stimulation by IL-4*

335 Primary stimulation of macrophages could lead to the formation of *de novo* enhancer and
336 transcriptional memory, which consequently hastens the kinetics of activation upon a second
337 repeated stimulation (*4, 26*). We hypothesized that the chromatin remodeling that occurred with
338 IL-4 stimulation in AAMs may translate into enhanced secondary responses with repeated IL-4
339 stimulation. To test this hypothesis, we compared the responsiveness of *in vivo* derived
340 F4/80^{int}CD206^{+} AAM^{mono} and M^{mono} to a secondary *ex vivo* stimulus of 24 hours (Figure 7A).
341 The secondary stimulation was performed after the harvested macrophages were rested overnight
342 in the absence of IL-4. These differences in secondary responses were also compared between
343 C57BL/6 and BALB/c mice.

344 AAM^{mono} from C57BL/6 mice were transcriptionally more distinct from M^{mono} after 24
345 hours of *ex vivo* culture without IL-4 stimulation, when compared to AAM^{mono} from BALB/c
346 mice (Figure 7B). In general, across both mouse strains, *in vivo* IL-4 stimulation led to greater
347 transcriptional changes than *ex vivo* IL-4 stimulation (Figure 7B-7C). AAM^{mono} were also less
348 responsive to the *ex vivo* IL-4 stimulation, as compared to M^{mono}, indicating that the pre-
349 treatment with *in vivo* IL-4 led to an overall unresponsiveness to a secondary *ex vivo* IL-4
350 stimulation and this was contrary to what we had hypothesized. However, the degree of
351 unresponsiveness to the secondary *ex vivo* IL-4 stimulation was more pronounced in AAM^{mono}
352 from C57BL/6 mice than that in BALB/c AAM^{mono} (Figure 7D).

353 By taking a supervised approach, we determined that 80.6% of the genes that were
354 inducible by *ex vivo* IL-4 in C57BL/6 macrophages would no longer be responsive if the
355 macrophages were pre-stimulated with *in vivo* IL-4. This was significantly greater (two-sided
356 Fisher's exact test, $p = 1 \times 10^{-9}$) than that observed for BALB/c macrophages, where 53.9% of

genes that were inducible by *ex vivo* IL-4 would no longer be responsive after pre-treatment with *in vivo* IL-4 (Figure 7E). Some of the genes that were unresponsive to the secondary IL-4 stimulation were genes that remained persistently upregulated overnight in the absence of IL-4 stimulation (45% of unresponsive genes in C57BL/6 AAM^{mono} and 35% of unresponsive genes in BALB/c AAM^{mono}). On the other hand, BALB/c macrophages had significantly higher number of genes that were enhanced in upregulation upon IL-4 restimulation (23.8% in BALB/c vs. 6.0% in C57BL/6, two-sided Fisher's exact test, $p < 2.2 \times 10^{-16}$) (Figure 7E). When we examined the subset of genes that were only upregulated in BALB/c macrophages with repeated IL-4 stimulation (Figure 7F, Cluster C6), this included *Arg1* (Figure 7G). In AAM^{mono} of C57BL/6 mice, *Arg1* was downregulated overnight in the absence of IL-4 and was equally responsive to *ex vivo* IL-4 stimulation in both M^{mono} and AAM^{mono}. In contrast, in macrophages of BALB/c mice, *Arg1* expression was only induced with repeated IL-4 stimulations (Figure 7G). Together, these findings suggest that *in vivo* IL-4 activation did not only lead to remodeling of the accessible chromatin landscape, but also led to persistent differences in cellular response that were different between C57BL/6 and BALB/c mice.

Discussion

In this study, we define and characterize IL-4 induced chromatin accessibility with *in vivo* alternative activation of AAM^{res} and AAM^{mono}. Although the pioneer factor PU.1 motif is similarly associated with IL-4 induced regions in both macrophages lineages, the local DNA shape features at these PU.1 motifs are macrophage-specific. By leveraging natural genetic variation between mouse strains, we demonstrate that the accessibility of putative PU.1 binding sites can be altered by DNA sequence mutations that cause a sufficiently large change in DNA

381 shape, without disruption of the PU.1 motif itself. We show that these IL-4-induced elements are
382 more likely to be specific to either AAMs from C57BL/6 and BALB/c mice. Finally, we show
383 that *in vivo* derived AAMs from C57BL/6 and BALB/c mice are functionally distinct and more
384 importantly, they have different degree of responsiveness to secondary stimulation. BALB/c
385 macrophages are more responsive to the secondary stimulation, when compared to C57BL/6
386 macrophages. Enhanced secondary response to stimulation is a hallmark of trained immunity,
387 which is also associated with epigenetic changes in macrophages (*27-29*). Hence, variation in
388 DNA sequences resulting from different genetic backgrounds in mice may affect whether
389 secondary responses are enhanced or dampened.

390 Most studies on regulation of immune cell differentiation or activation states are focused
391 on the presence or absence of specific TFs. However, less is known about how these TFs read
392 the genetic code in order to bind DNA and cause changes to cellular phenotypes. Here, we find
393 that similar TF motifs (i.e. putative TF binding sites consisting of similar short DNA sequences)
394 can have macrophage-specific sequence features, indicating that similar TFs recognize DNA
395 binding sites through sequence features beyond the core DNA-protein interacting base pairs.
396 Furthermore, by specifically examining the IL-4 induced PU.1 regions, we find that the PU.1
397 regions in both types of macrophages were not only selectively enriched with different TFs, but
398 were also associated with different local DNA shape profiles, specifically in the DNA propeller
399 twist and helical twist configurations. However, while the DNA propeller twist and helical twist
400 are associated with TF family-specific binding pattern, the DNA minor groove width, on the
401 other hand, appeared to be associated with accessible vs. non-accessible PU.1 regions. These
402 observations are consistent with findings from a study that determined how bHLH proteins
403 (Tye7 vs. Cbf1) with similar DNA binding motifs could recognize distinct binding regions in the

404 yeast genome. DNA propeller twist was predictive of binding preferences between Tye7 vs.
405 Cbf1, while DNA minor groove width was predictive of high vs. low affinity binding sites (*30*).
406 In a different study, DNA minor groove width was identified as an important predictor of actual
407 PU.1 binding in BMDMs (*20*). However, questions remain as to how differences in DNA shape
408 mechanistically translate into differences in chromatin remodeling. Other than the possibility of
409 different co-binding factors, different variants of the PU.1 protein could also be present in
410 AAM^{res} vs. AAM^{mono}, leading to different binding preferences genome-wide. Since DNA shape
411 is a consequence of DNA sequence, both modes of DNA recognition are confounded and
412 difficult to dissociate from one another. It is, therefore, important to identify which amino acid
413 residue(s) in the PU.1 protein could be involved with this DNA shape readout, as mutating these
414 residues could potentially be a strategy to identify PU.1 binding sites that are solely dependent
415 on DNA shape readout, without being confounded by DNA sequences (*31*).

416 While our initial goal was to explore how different lineages of macrophages remodel
417 their chromatin in response to the same IL-4 stimuli, our findings on the potential role of DNA
418 shape features in regulating chromatin accessibility around binding motifs led us to test our
419 model utilizing natural genetic variations between C57BL/6 and BALB/c mice. We know by
420 now that most genetic variants reside within non-coding regions of the genome, but less is
421 known about why some of these regulatory elements are affected by the presence of sequence
422 variants and some are not. For example, 20-95% of accessible regions in BMDMs with TLR4
423 activation could contain a sequence variant, yet the accessibility of these regions remained
424 unaffected (*13*). In our comparison between C57BL/6 and BALB/c mice, we find that a
425 regulatory element containing multiple sequence variants is far more likely to experience an
426 alteration in chromatin accessibility and this is due to a greater change in DNA shape. This

427 observation is not specific to PU.1, but can also be observed with all accessible chromatin
428 regions in AAMs genome-wide. While the presence of multiple sequence variants within a single
429 regulatory element indicates the need to further decipher which of these variants are truly causal,
430 this finding nonetheless suggests that DNA shape change could represent a generalizable means
431 to prioritize which regulatory elements are more likely to be affected by sequence variants.
432 However, we are still unable to draw a convincing conclusion on how this disruption of
433 chromatin accessibility leads to actual functional differences. This challenge of associating
434 chromatin structure with gene expression has already been described by several other groups and
435 it is now generally thought that the structural changes (chromatin organization and TF binding)
436 induced by DNA sequence variants are often more pronounced than the associated functional
437 changes (*13, 32, 33*).

438 Several studies have recently described the combinatorial effects of cytokines on
439 macrophage activation, although most of these studies have focused on the opposing effects of
440 type 1 and type 2 cytokines (*14, 15*). When repeated stimulation with the same cytokine was
441 performed and demonstrated "transcriptional memory" (*4, 26*), these studies were conducted
442 using short-term stimulation in BMDMs. As such, no comparable studies have so far
443 demonstrated the dampening of response to repeated cytokine stimulation after *in vivo* alternative
444 activation. Whereas an enhanced secondary response in macrophage function is now considered
445 to be a key component of "trained immunity", our results here suggest that such "training" may
446 also include a dampening of responses under certain circumstances that may be dependent on the
447 genetic background of the host. It would be important to determine if this dampening of response
448 was specific to *in vivo* IL-4 stimulation, particularly in the setting of a physiological helminth
449 infection model, which should have continuous production of type 2 cytokines at high abundance

450 and also over a chronic infection course. Furthermore, given that the dampening of response
451 occurred to different degree between the C57BL/6 and BALB/c mice, it would be important to
452 determine if similar observations can be made within the human population and how immune
453 cells from individuals of different genetics backgrounds and with different cytokine production
454 capacities would respond differently to repeated environmental stimulations.

455

456 **Supplementary Materials**

457 Materials and Methods

458 Fig. S1. IL-4 stimulation leads to remodeling of open chromatin landscape in peritoneal
459 macrophages.

460 Fig. S2. Motif analyses on IL-4 induced regions.

461 Fig. S3. Sequence features of KLF and AP-1 motifs.

462 Fig. S4. Leveraging on natural genetic variation in C57BL/6 and BALB/c mouse strains to
463 perform a mutation study.

464 Fig. S5. Genomic features associated with strain-specificity.

465 Data file S1. List of motifs identified by over-representation approach (related to Figure 2C).

466 Data file S2. DNA shape plots of all IL-4 induced PU.1 motif sites with flanking SNPs and
467 strain-specific chromatin accessibility (related to Figures 4A).

468 Data file S3. Coordinates and sequences of all strain-specific IL-4 induced PU.1 regions (related
469 to Figures 4A-B).

470 Data file S4. List of pathways enriched among strain-specific genes from AAM^{res} and AAM^{mono}
471 (related to Figure 6C).

472

473 **References and Notes:**

474 1. C. K. Glass, G. Natoli, Molecular control of activation and priming in macrophages.
475 *Nature immunology* **17**, 26 (2016).
476 2. S. Ghisletti *et al.*, Identification and Characterization of Enhancers Controlling the
477 Inflammatory Gene Expression Program in Macrophages. *Immunity* **32**, 317-328 (2010).
478 3. G. Natoli, S. Ghisletti, I. Barozzi, The genomic landscapes of inflammation. *Genes &*
479 *development* **25**, 101-106 (2011).
480 4. R. Ostuni *et al.*, Latent enhancers activated by stimulation in differentiated cells. *Cell*
481 **152**, 157-171 (2013).
482 5. M. U. Kaikkonen *et al.*, Remodeling of the enhancer landscape during macrophage
483 activation is coupled to enhancer transcription. *Molecular cell* **51**, 310-325 (2013).

484 6. J. E. Allen, R. M. Maizels, Diversity and dialogue in immunity to helminths. *Nature*
485 *Reviews Immunology* **11**, 375 (2011).
486 7. U. M. Gundra *et al.*, Vitamin A mediates conversion of monocyte-derived macrophages
487 into tissue-resident macrophages during alternative activation. *Nature immunology* **18**,
488 642-653 (2017).
489 8. U. M. Gundra *et al.*, Alternatively activated macrophages derived from monocytes and
490 tissue macrophages are phenotypically and functionally distinct. *Blood* **123**, 22 (2014).
491 9. R. E. Thurman *et al.*, The accessible chromatin landscape of the human genome. *Nature*
492 **489**, 75 (2012).
493 10. A. B. Stergachis *et al.*, Developmental fate and cellular maturity encoded in human
494 regulatory DNA landscapes. *Cell* **154**, 888-903 (2013).
495 11. Y. Lavin *et al.*, Tissue-resident macrophage enhancer landscapes are shaped by the local
496 microenvironment. *Cell* **159**, 1312-1326 (2014).
497 12. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of
498 native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-
499 binding proteins and nucleosome position. *Nature methods* **10**, 1213-1218 (2013).
500 13. V. M. Link *et al.*, Analysis of Genetically Diverse Macrophages Reveals Local and
501 Domain-wide Mechanisms that Control Transcription Factor Binding and Function. *Cell*
502 **173**, 1796-1809.e1717 (2018).
503 14. V. Piccolo *et al.*, Opposing macrophage polarization programs show extensive
504 epigenomic and transcriptional cross-talk. *Nature immunology* **18**, 530-540 (2017).
505 15. K. Kang *et al.*, Interferon-gamma Represses M2 Gene Expression in Human
506 Macrophages by Disassembling Enhancers Bound by the Transcription Factor MAF.
507 *Immunity* **47**, 235-250.e234 (2017).
508 16. J. C. Walsh *et al.*, Cooperative and antagonistic interplay between PU.1 and GATA-2 in
509 the specification of myeloid cell fates. *Immunity* **17**, 665-676 (2002).
510 17. S. R. McKercher *et al.*, Targeted disruption of the PU.1 gene results in multiple
511 hematopoietic abnormalities. *The EMBO Journal* **15**, 5647-5658 (1996).
512 18. S. Heinz *et al.*, Effect of natural genetic variation on enhancer selection and function.
513 *Nature* **503**, 487-492 (2013).
514 19. C. R. Escalante *et al.*, Crystal structure of PU.1/IRF-4/DNA ternary complex. *Molecular*
515 *cell* **10**, 1097-1105 (2002).
516 20. I. Barozzi *et al.*, Coregulation of transcription factor binding and nucleosome occupancy
517 through DNA features of mammalian enhancers. *Molecular cell* **54**, 844-857 (2014).
518 21. T. Zhou *et al.*, DNAshape: a method for the high-throughput prediction of DNA
519 structural features on a genomic scale. *Nucleic acids research* **41**, W56-W62 (2013).
520 22. A. Crotti *et al.*, Mutant Huntingtin promotes autonomous microglia activation via
521 myeloid lineage-determining factors. *Nature neuroscience* **17**, 513 (2014).
522 23. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime
523 cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**,
524 576-589 (2010).
525 24. K. S. Vignesh *et al.*, IL-4 induces metallothionein 3-and SLC30A4-dependent increase in
526 intracellular Zn2+ that promotes pathogen persistence in macrophages. *Cell reports* **16**,
527 3232-3246 (2016).
528 25. S. M. Campbell *et al.*, Myeloid cell recruitment versus local proliferation differentiates
529 susceptibility from resistance to filarial infection. *eLife* **7**, (2018).

530 26. B. Daniel *et al.*, The IL-4/STAT6/PPARγ signaling axis is driving the expansion of the
531 RXR heterodimer cistrome, providing complex ligand responsiveness in macrophages.
532 *Nucleic Acids Research*, gky157-gky157 (2018).
533 27. B. Novakovic *et al.*, beta-Glucan Reverses the Epigenetic State of LPS-Induced
534 Immunological Tolerance. *Cell* **167**, 1354-1368.e1314 (2016).
535 28. I. Mitroulis *et al.*, Modulation of Myelopoiesis Progenitors Is an Integral Component of
536 Trained Immunity. *Cell* **172**, 147-161.e112 (2018).
537 29. E. Kaufmann *et al.*, BCG Educates Hematopoietic Stem Cells to Generate Protective
538 Innate Immunity against Tuberculosis. *Cell* **172**, 176-190.e119 (2018).
539 30. R. Gordân *et al.*, Genomic regions flanking E-box binding sites influence DNA binding
540 specificity of bHLH transcription factors through DNA shape. *Cell reports* **3**, 1093-1104
541 (2013).
542 31. N. Abe *et al.*, Deconvolving the recognition of DNA shape from sequence. *Cell* **161**, 307-
543 318 (2015).
544 32. K. Alasoo *et al.*, Shared genetic effects on chromatin and gene expression indicate a role
545 for enhancer priming in immune response. *Nature genetics*, (2018).
546 33. M. Gasperini *et al.*, A Genome-wide Framework for Mapping Gene Regulation via
547 Cellular Genetic Screens. *Cell* **176**, 377-390.e319 (2019).
548 34. S. J. Jenkins *et al.*, Local macrophage proliferation, rather than recruitment from the
549 blood, is a signature of TH2 inflammation. *Science* **332**, 1284-1288 (2011).
550 35. T. Hashimshony *et al.*, CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq.
551 *Genome biology* **17**, 77 (2016).
552 36. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Meth* **9**,
553 357-359 (2012).
554 37. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
555 *(Oxford, England)* **25**, 2078-2079 (2009).
556 38. M. R. Breese, Y. Liu, NGSUtils: a software suite for analyzing and manipulating next-
557 generation sequencing datasets. *Bioinformatics (Oxford, England)* **29**, 494-496 (2013).
558 39. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV):
559 high-performance genomics data visualization and exploration. *Briefings in*
560 *Bioinformatics* **14**, 178-192 (2013).
561 40. M. T. McCarthy, C. A. O'Callaghan, PeaKDEck: a kernel density estimator-based peak
562 calling program for DNaseI-seq data. *Bioinformatics (Oxford, England)* **30**, 1302-1304
563 (2014).
564 41. R. Stark, G. Brown. (2011).
565 42. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for
566 RNA-seq data with DESeq2. *Genome biology* **15**, 550 (2014).
567 43. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing
568 and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
569 44. R. Gentleman, V. Carey, W. Huber, F. Hahne. (2016).
570 45. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic
571 features. *Bioinformatics (Oxford, England)* **26**, 841-842 (2010).
572 46. F. Hsu *et al.*, The UCSC Known Genes. *Bioinformatics (Oxford, England)* **22**, 1036-1046
573 (2006).
574 47. T. L. Bailey *et al.*, MEME SUITE: tools for motif discovery and searching. *Nucleic Acids*
575 *Res* **37**, W202-208 (2009).

48. K. Karwacz *et al.*, Critical role of IRF1 and BATF in forming chromatin landscape during type 1 regulatory cell differentiation. *Nature immunology* **18**, 412 (2017).
49. S. Mahony, P. E. Auron, P. V. Benos, DNA Familial Binding Profiles Made Easy: Comparison of Various Motif Alignment and Clustering Strategies. *PLoS computational biology* **3**, e61 (2007).
50. P. V. Benos, S. Mahony, STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Research* **35**, W253-W258 (2007).
51. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: scanning for occurrences of a given motif. *Bioinformatics (Oxford, England)* **27**, 1017-1018 (2011).
52. E. Wingender, T. Schoeps, M. Haubrock, M. Krull, J. Dönitz, TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Research* **46**, D343-D347 (2018).
53. T.-P. Chiu *et al.*, DNAshapeR: an R/Bioconductor package for DNA shape prediction and feature encoding. *Bioinformatics (Oxford, England)* **32**, 1211-1213 (2016).
54. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics (Oxford, England)* **27**, 2156-2158 (2011).
55. S. Neph *et al.*, BEDOPS: high-performance genomic feature operations. *Bioinformatics (Oxford, England)* **28**, 1919-1920 (2012).
56. J. Rozowsky *et al.*, AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol Syst Biol* **7**, 522 (2011).
57. S. C. J. Parker, L. Hansen, H. O. Abaan, T. D. Tullius, E. H. Margulies, Local DNA Topography Correlates with Functional Noncoding Regions of the Human Genome. *Science* **324**, 389 (2009).

Acknowledgments: We thank the NYU School of Medicine Genome Technology Center and Cytometry and Cell Sorting Laboratory core facilities. These shared resources are partially supported by the Laura and Isaac Perlmutter Cancer Center support grant P30CA016087. We also thank the NYU IT High Performance Computing resources, services, and staff expertise. **Funding**: This work was supported through the NIH, NIAID grants AI093811 and AI094166 (P.L.) and NIDDK grant DK103788 (P.L.), American Association of Immunologists (M.S.T.), Vilcek Foundation (M.S.T.). **Author contributions**: Conceptualization, M.S.T. and P.L.; Methodology, M.S.T, E.R.M., N.M.G., R.A.B. and P.L.; Formal Analysis, M.S.T.; Investigation, M.S.T. and N.M.G.; Writing – Original Draft, M.S.T. and P.L.; Writing – Review and Editing, E.R.M. and P.L.; Visualization, M.S.T.; Data Curation, M.S.T.; Supervision, R.A.B. and P.L.; Funding acquisition, P.L. All authors read and approved of the final draft. **Competing interests**: None **Data and materials availability**: ATAC-Seq and RNA-Seq data have been deposited on NCBI database Gene Expression Omnibus (GEO) under the SuperSeries GSE116108 (ATAC-Seq subseries GSE116107 and RNA-Seq subseries GSE116105).

Figures:

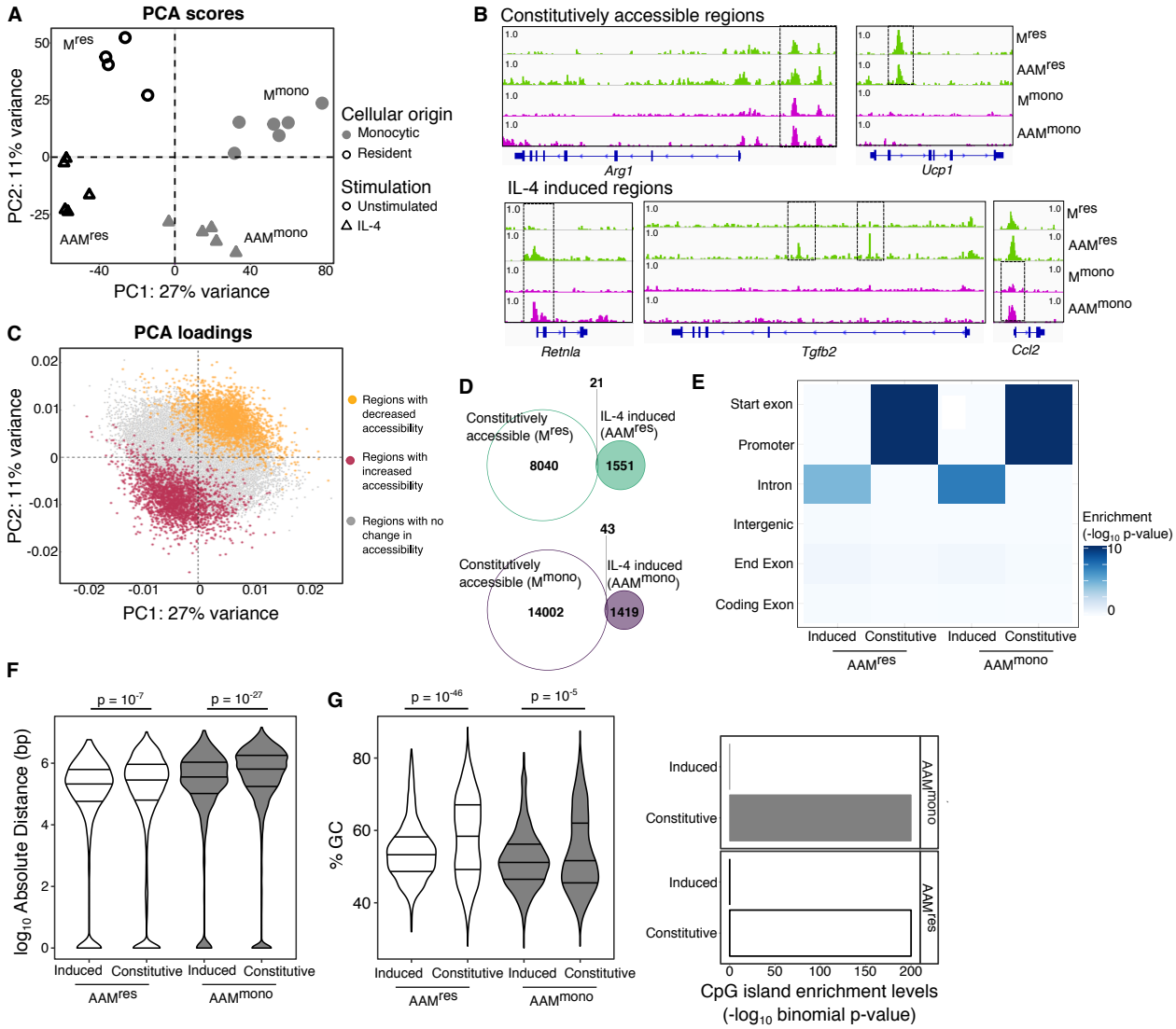


Fig. 1. IL-4 stimulation leads to remodeling of open chromatin landscape in peritoneal macrophages. (**A**) PCA scores of individual ATAC-seq samples. PCA was performed using rlog-transformed ATAC-seq read counts of 30,856 regions with high variance (only regions with variance inter-quartile range > 0.5 were retained). N = 4-6 mice per macrophage population. Data points represent independent biological replicates. (**B**) Genome browser views of representative (boxed) constitutively accessible and IL-4 induced regions. (**C**) The contributions of individual accessible regions to PCs 1 and 2 are represented in the PCA loadings plot. Each data point is color-coded based on the direction of its IL-4 dependency. Hence, IL-4 induced regions (red) are highly associated with IL-4 stimulated macrophages, while IL-4 repressed regions (yellow) are highly associated with unstimulated macrophages. We compared (**D**) enrichment levels for different types of genomic elements, (**E**) distance from a closest IL-4-induced gene and (**F**) G/C content between constitutively accessible and IL-4-induced regions in AAM^{mono} and AAM^{res}, respectively. G/C content information is represented in two different ways – percentage of G/C bases in an accessible region (F, left panel) and CpG island

634 enrichment for a given group of accessible regions (F, right panel). Number of IL-4
635 induced regions = 1572 in AAM^{res} and 1462 in AAM^{mono}; number of constitutively
636 accessible regions = 8061 in AAM^{res} and 14045 in AAM^{mono}. Enrichment p-values are
637 from binomial test, while two-class comparison p-values are from two-sided Mann-
638 Whitney test.
639

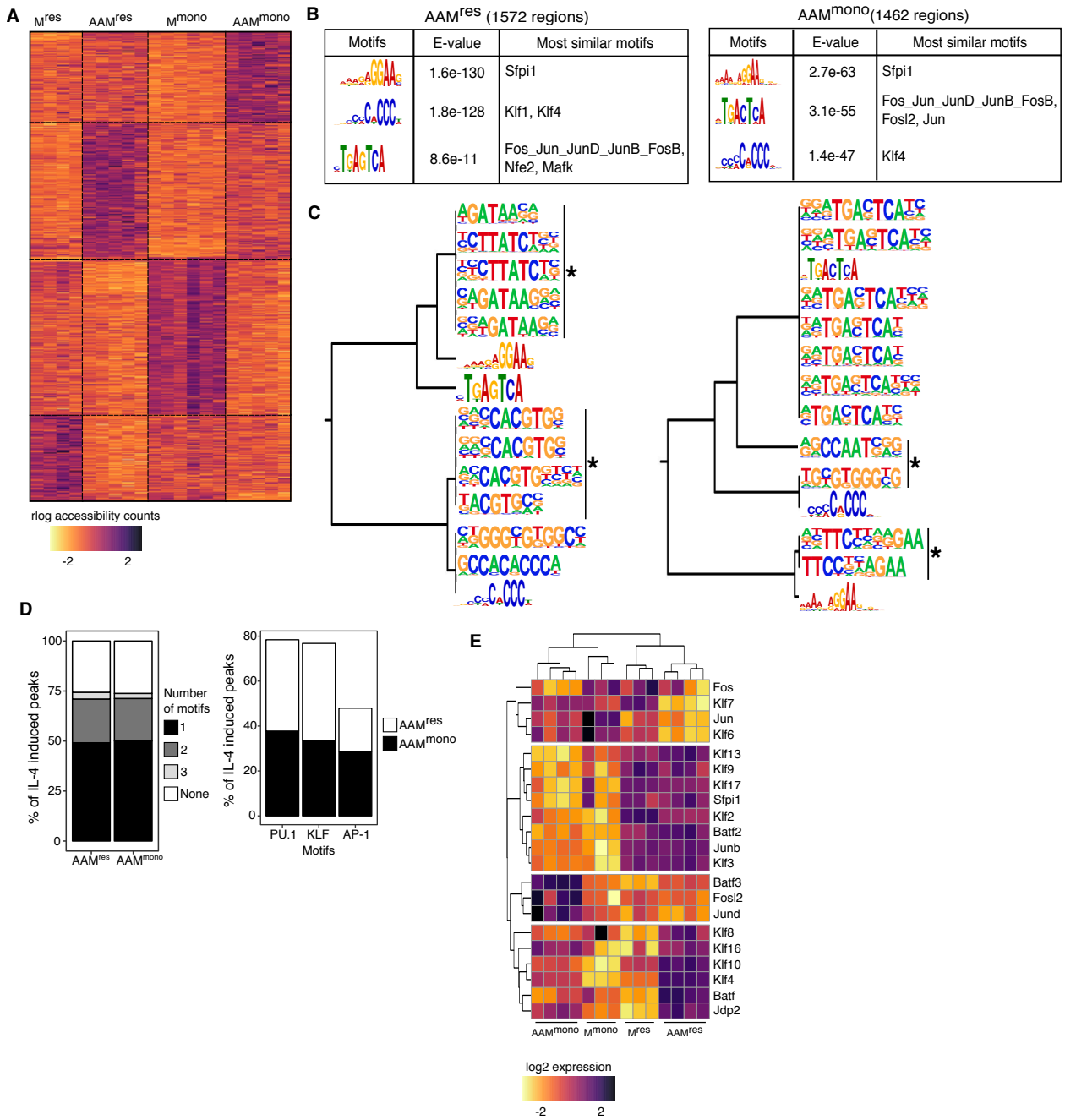


Fig. 2. IL-4 induced regions are associated with similar families of TF motifs. (**A**) Heatmap visualizing the macrophage-specific IL-4 dependent regions. Each row represents one of the 2855 IL-4 dependent regions and each column a unique sample. Values are rlog-transformed, batch-subtracted read counts, scaled using a z-score transformation for each region. (**B**) Motifs discovered *de novo* by MEME from 1572 IL-4 induced regions in AAM^{res} and 1462 IL-4 induced regions in AAM^{mono}. (**C**) Clustering analysis of motifs discovered *de novo* and macrophage-specific motifs identified using an over-representation approach, from the IL-4 induced regions of AAM^{res} (left) and AAM^{mono} (right), respectively. Asterisks indicate macrophage-specific motifs uniquely identified via the over-representation approach. (**D**) Frequencies of IL-4 induced peaks categorized

651 by the number of *de novo* motifs detected per region (left) and the presence of any PU.1,
652 KLF or AP-1 motif (right). (**E**) Expression values of individual TF members that were
653 highly expressed and belonged to either PU.1, KLF or AP-1 family. Expression values
654 were obtained from (*8*).
655

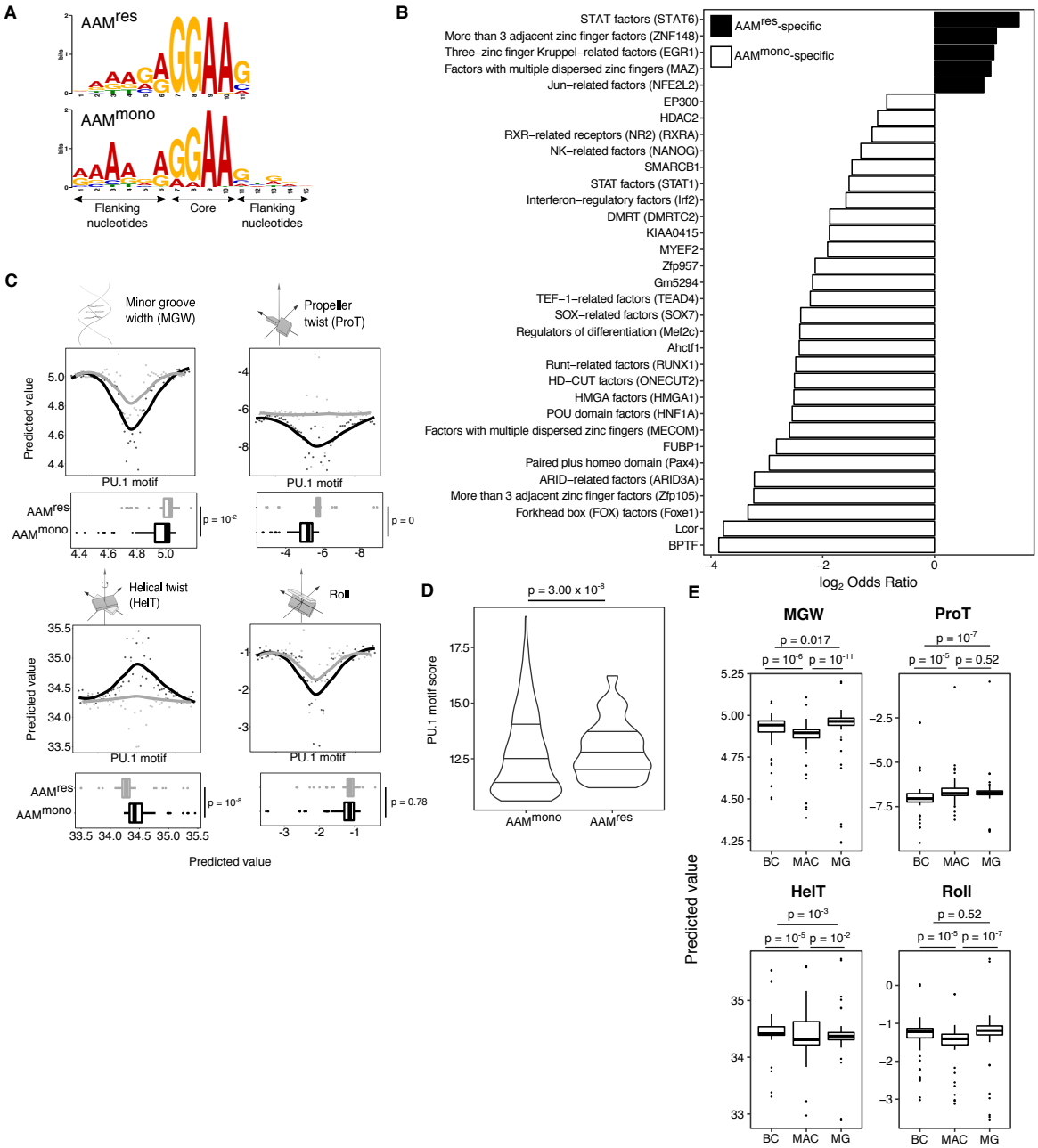
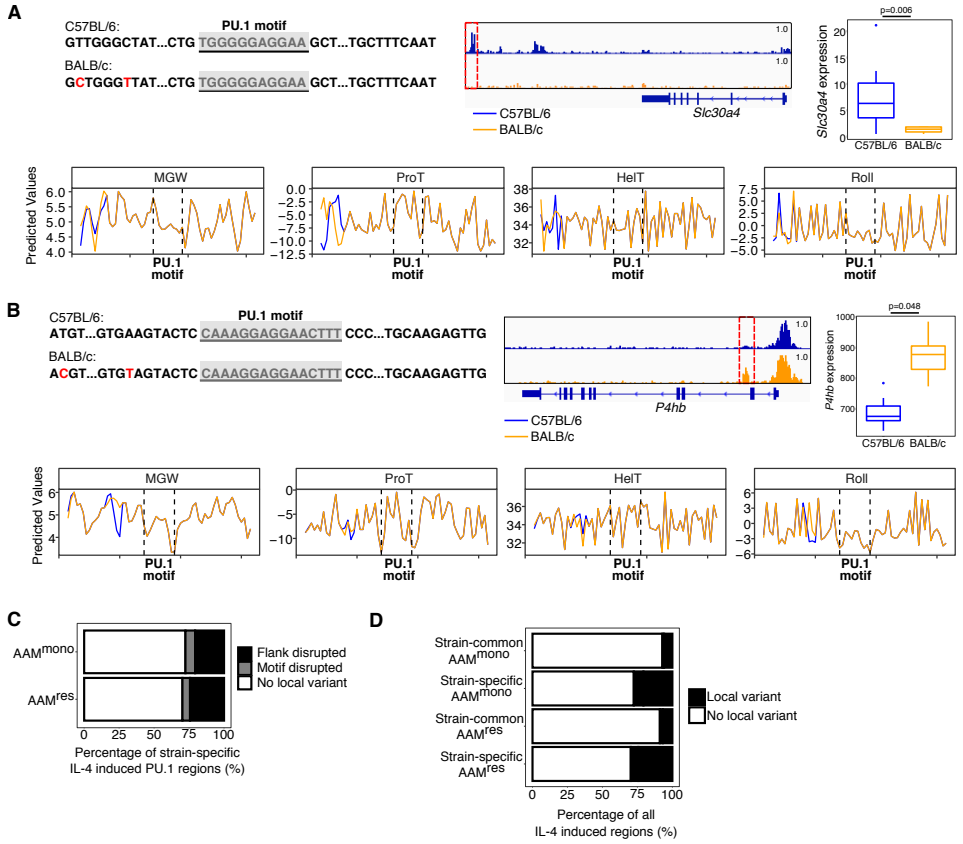


Fig. 3. PU.1 motifs are associated with macrophage-specific sequence features. (**A**) PU.1 motif discovered from IL-4 induced regions of AAMres (top) and AAMmono (bottom), with the core and flanking nucleotides highlighted. (**B**) Non-ETS TF motifs that were detected in the IL-4 induced PU.1 regions (PU.1 motif ± 25bp) and were macrophage-specific, as determined using a two-sided Fisher's exact test (false discovery rate < 10%). TF motifs were summarized at the family level, with the maximum absolute $\log_2$ odds ratio of the family visualized and the specific TF member with the maximum absolute $\log_2$ odds ratio value is stated in parenthesis. Where TF family annotation was not available, the $\log_2$ odds ratio of the specific TF itself is used. (**C**) Average of DNA shape features at IL-4 induced PU.1 regions of AAM^{mono} and AAM^{res} (MGW = Minor Groove Width, ProT =

667 Propeller Twist, HelT = Helical Twist). Scatter plots are centered on IL-4 induced PU.1
668 motifs (x-axes), with the solid lines representing Loess fit of predicted shape values at
669 single-nucleotide resolution (dots). Boxplots are average predicted shape values over the
670 PU.1 motif ± 25bp windows and p-values from two-sided Kolmogorov-Smirnov test. (**D**)
671 FIMO-derived scores of PU.1 motifs discovered from IL-4 induced regions. P-value from
672 two-sided Mann-Whitney test. (**E**) Average predicted DNA shape values over PU.1 motif
673 ± 25bp derived from published PU.1 ChIP-Seq data (*22, 23*). P-values from two-sided
674 Kolmogorov-Smirnov test. BC = B cell, MAC = thioglycollate-induced peritoneal
675 macrophages, MG = microglial BV2 cell line.
676



677

678 **Fig. 4.** Accessibility of IL-4 induced PU.1 regions can be altered through DNA shape change
679 induced by sequence mutation, without disruption of the PU.1 motif. Selected examples
680 from (A) AAM^{res} and (B) AAM^{mono} of an IL-4 induced PU.1 motif with flanking BALB/c
681 SNPs. In both examples, sequences represent a unique IL-4 induced PU.1 motif region,
682 where the PU.1 motif is underlined and shaded in grey, while BALB/c SNPs are
683 highlighted with red font. Line graphs of predicted shape values are centered on the PU.1
684 motif (sequences between the vertical lines). Genome browser tracks illustrate the strain-
685 specific accessibility of the region containing the matching sequence (boxed in red).
686 Boxplots represent size-factor normalized read counts in C57BL/6 and BALB/c
687 macrophage. P-values are from DESeq2 (N = 4-8 mice per group) and adjusted by
688 Benjamini-Hochberg procedure. (C) Frequency of different types of sequence variants in
689 strain-specific IL-4 induced PU.1 regions and (D) all IL-4 induced PU.1 regions
690 (bottom). Number of strain-specific IL-4 induced PU.1 regions = 58 in AAM^{mono} and 37
691 in AAM^{res}; number of strain-common IL-4 induced PU.1 regions = 754 in AAM^{mono} and
692 828 in AAM^{res}.
693

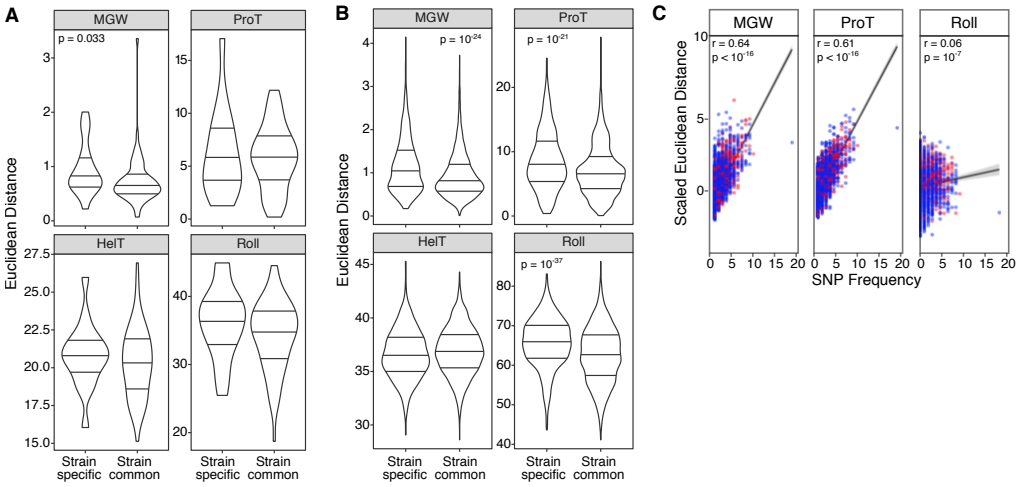


Fig. 5. Accessibility is most likely to be disrupted in chromatin regions that experienced a large change in DNA shape. Comparison of the amount of SNP-induced DNA shape change in (**A**) 18 strain-specific vs. 63 strain-common PU.1 regions that contained SNPs and (**B**) 922 strain-specific vs. 4776 strain-common accessible regions that contained SNPs. The shift in DNA shape for each sequence is quantitated using Euclidean distance, where a larger Euclidean distance value indicates a larger shift in shape. P-values from two-sided Mann-Whitney tests. (**C**) Scatter plots illustrating the positive correlations between SNP frequency and amount of DNA shape change undergo by a specific region. Red = strain-specific regions, blue = strain-common regions. Euclidean distances here are scaled across each DNA shape feature for visualization purpose. Correlation coefficient and p-values are based on the Spearman's test using un-scaled Euclidean distances.

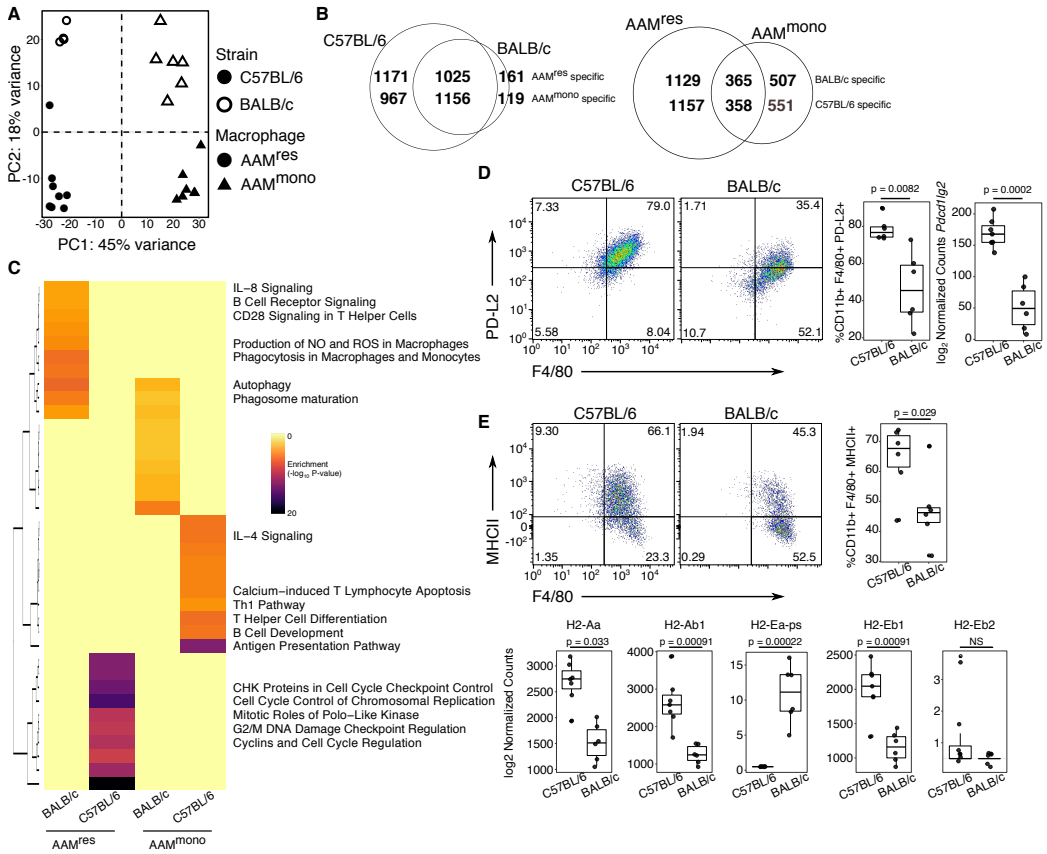


Fig. 6. AAMs from C57BL/6 and BALB/c are functionally distinct. (**A**) PCA of 7431 genes with high variance (only genes with variance inter-quartile range > 0.5 were retained). N = 4-8 mice per group. Data points represent independent biological replicates. (**B**) Venn diagrams indicating the number of genes that were commonly and uniquely identified as significantly differential in different comparisons. Left panel: Identification of macrophage-specific genes in C57BL/6 and BALB/c AAMs, respectively. Right panel: Identification of strain-specific genes in AAM^{res} and AAM^{mono}, respectively. (**C**) Enrichment values from Ingenuity Pathway Analysis visualized as $-\log_{10}$ p-value for the four different groups of genes – (1) BALB/c specific in AAM^{res}, (2) C57BL/6 specific in AAM^{res}, (3) BALB/c specific in AAM^{mono} and (4) C57BL/6 in AAM^{mono}. Only the top 10 pathways (as defined by enrichment p-values) are included in this visualization. Specific pathways are highlighted for clarity purpose. The complete list of enriched pathways is provided as Supplemental Materials Data S2. Differential expression of (**D**) PD-L2 and (**E**) MHCII, as determined by flow cytometric analysis of surface expression and transcriptional profiling. Representative flow plots are chosen based on the median values of the double positive populations. Values within each quadrant represent % CD11b+ population.

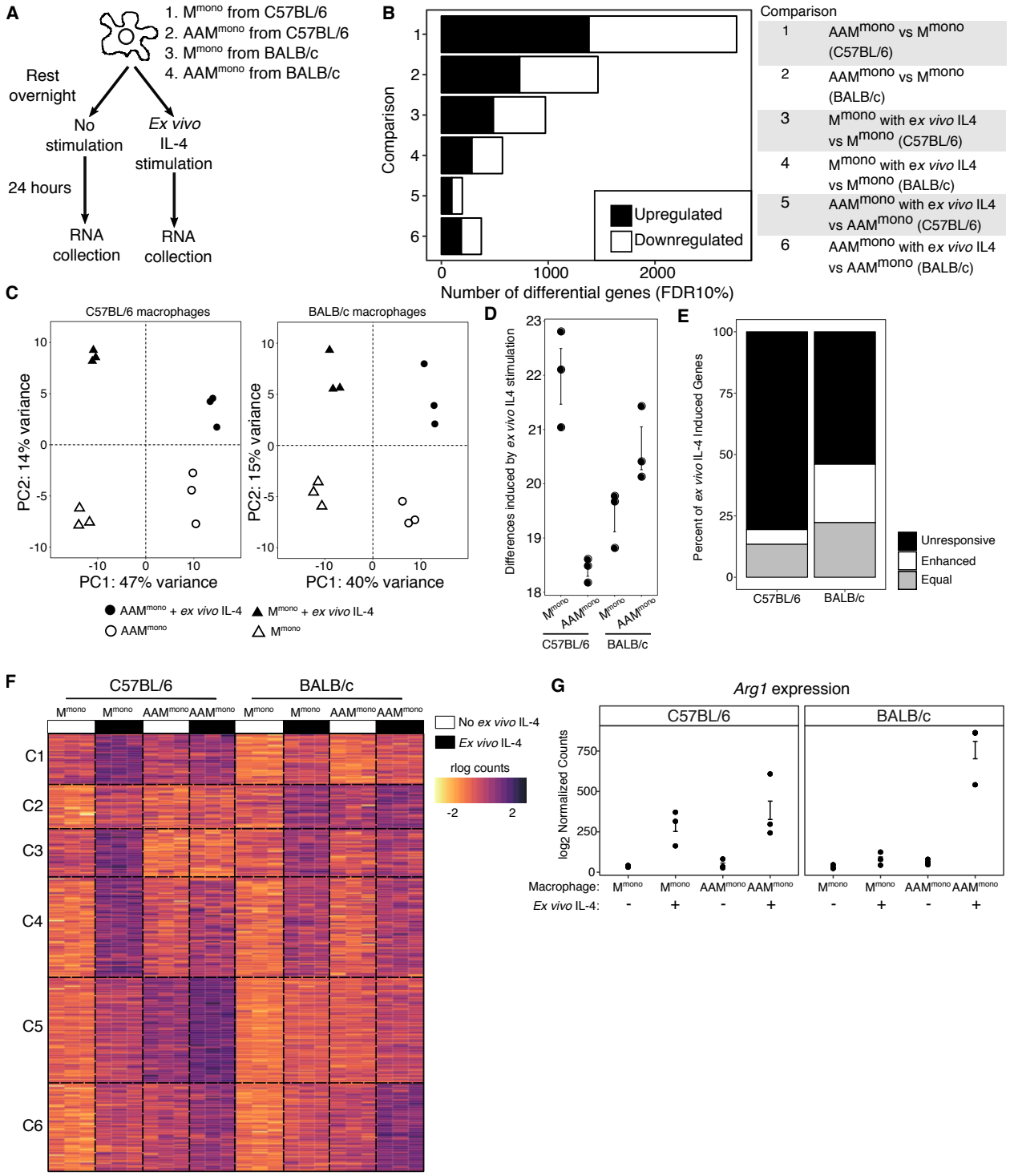


Fig. 7. Initial alternative activation dampens the macrophage's response to a secondary stimulation. (A) Schematic of the restimulation experiment. (B) Number of differential genes for each of the listed comparison. Statistical significance threshold is set at FDR < 10%. N = 3 mice per group. (C) PCA of 6528 genes with high variance (only genes with variance inter-quartile range > 0.5 were retained), in C57BL/6 (left) and BALB/c (right) macrophages. (D) Magnitude of transcriptional changes caused by *ex vivo* IL-4

stimulation, indicated by Euclidean distance, where a larger Euclidean distance value indicates greater changes in transcriptional profiles. (E) Frequency of genes categorized as “unresponsive”, “enhanced” and “equal”, based on changes in transcriptional levels pre- and post-*ex vivo* IL-4 stimulation. (F) K-means clustering of 334 genes upregulated by *ex vivo* IL-4 stimulation in C57BL/6 and BALB/c monocyte-derived macrophages. Each row represents one of the 334 genes and each column a unique sample. Values are rlog-transformed and scaled using a z-score transformation for each gene. (G) $\log_2$ DESeq2 normalized counts for Arg1 across different mouse strains and conditions.

742 **Table 1.** Full length sequences of PU.1 motif with flanking SNPs used as example in Figures
743 4A-B. SNPs are highlighted in red font, while PU.1 motifs are underlined.

744

AAM^{res} (Figure 4A)
C57BL/6 (chr2:122656227-122656288) GTTGGGCTATGGTCACACTGTCCTG TGGGGGAGGAA GCTGAGATCATTGTCTGCTTTCAAT **BALB/c (chr2:122639812-122639873)** GCTGGGTTATGGTCACACTGTCCTG TGGGGGAGGAA GCTGAGATCATTGTCTGCTTTCAAT
AAM^{mono} (Figure 4B)
C57BL/6 (chr11:120571348-120571413) ATGTGACACTTCCTGTGAAGTACTC CAAAGGAGGAACTTT CCCCAGCTCCCACATGCAAGAGTTG **BALB/c (chr11:120542029-120542094)** ACGTGACACTTCCTGTGTAGTACTC CAAAGGAGGAACTTT CCCCAGCTCCCACATGCAAGAGTTG

745
746

747 **Table 2.** Frequency count of different types of sequence variants in all IL-4 induced PU.1
748 regions.
749

Type of variants	Strain-specific IL-4 induced PU.1 regions in AAM^{res}	Strain-common IL-4 induced PU.1 regions in AAM^{res}	Strain-specific IL-4 induced PU.1 regions in AAM^{mono}	Strain-common IL-4 induced PU.1 regions in AAM^{mono}
Motif disrupting	2	16	4	9
Motif flanking	9	53	12	45
No local variants	26	702	42	699

750 **Table 3.** Non-ETS family TF motif occurrences were identified within IL-4 induced PU.1 motif
751 ± 25bp regions and the number of SNPs that did and did not disrupt these motif occurrences
752 were counted. There were no significant differences between the proportions of motif disrupting
753 SNPs in strain-specific vs. strain-common IL-4 induced PU.1 regions, based on the two-sided
754 Fisher's test.

Type of IL-4 induced PU.1 regions	Type of SNPs	Number of SNPs
Strain-specific	Motif disrupting	20
Strain-specific	Non-motif disrupting	3
Strain-common	Motif disrupting	64
Strain-common	Non-motif disrupting	10

755

756 **Table 4.** Euclidean distances represent the magnitude of shape change caused by the presence of
757 BALB/c SNPs, where a larger Euclidean distance value represents a greater shift in DNA shape.
758 Values presented here are for 18 strain-specific IL-4 induced PU.1 regions and 63 strain-
759 common IL-4 induced PU.1 regions across AAM^{res} and AAM^{mono}. Only IL-4 induced PU.1
760 regions with flanking SNPs are included in this analysis.

Shape feature	Mean Euclidean Distance ± SEM (18 strain-specific regions)	Mean Euclidean Distance ± SEM (63 strain-common regions)
MGW	0.95 ± 0.12	0.74 ± 0.06
ProT	6.17 ± 0.99	5.82 ± 0.39
HelT	20.81 ± 0.52	20.34 ± 0.29
Roll	36.14 ± 1.27	34.25 ± 0.65

761

762 **Supplementary Materials:**

763

764 **Materials and Methods**

765

766 Experimental methods

767 **Mice** Wild type (WT) C57BL/6 mice were purchased from Jackson laboratory and bred onsite
768 for the first set of experiments that compared the effect of IL-4 stimulation on the accessible
769 chromatin profiles from M^{res}, AAM^{res}, M^{mono} and AAM^{mono}. For experiments directly comparing
770 AAMs of C57BL/6 and BALB/c backgrounds, mice of both strains were purchased from Jackson
771 laboratory and immediately used for experiments. Mice were age (7-8 weeks of age) and gender
772 matched. IL-4/anti-IL-4 monoclonal antibody (mAb) complexes (IL-4c) were prepared as
773 described previously (*34*). To induce AAM^{res}, mice were injected intraperitoneally (i.p.) with IL-
774 4c on days 0 and 2. Mice were also treated with 4% thioglycollate alone (to induce M^{mono}) or in
775 combination with IL-4c to induce AAM^{mono} (*8*). All animal procedures were approved by the
776 New York University Institutional Animal Care and Use Committee (IACUC) under protocol
777 numbers 131004 and 130504.

778

779 **Peritoneal cell isolation and cell sorting** Peritoneal cells were isolated by washing the
780 peritoneal cavity twice with cold PBS 1x. Peritoneal exudate were then treated with ACK Lysis
781 buffer to lyse red blood cells and washed once with PBS. Cells were then re-suspended to single-
782 cell suspensions for staining with fluorescently conjugated antibodies at 1:100 dilutions, unless
783 otherwise noted. Antibodies were diluted using 2% fetal bovine serum (FBS). Cells were stained
784 with one of either LIVE/DEAD™ Blue (Invitrogen) or LIVE/DEAD™ Near-IR (Invitrogen),
785 blocked with 4µg/ml anti-CD16/32 (2.4G2; Bioxcell) and stained with anti-CD11b Pacific Blue
786 (M1/70; Biolegend), F4/80 PECy7 (BM8; Biolegend), CD206 APC (C068C2; Biolegend),

787 Siglec-F PE (E50-2440; BD Biosciences), CD3 PE (145-2C11; Biolegend), CD19 PE (6D5;
788 Biolegend), CD49b PE (DX5; Biolegend), Ly6G (1A8; Biolegend), PD-L2 (PerCP-Cy55;
789 Miltenyi; diluted at 1:20), MHCII (APC-Cy7; Biolegend). Cells were gated on singlet, live,
790 Dump-negative (CD3−, CD19−, DX5−, Siglec-F−, Ly6G−), CD11b+, then subsequently gated
791 on their M^{res} and AAM^{res} (F4/80hi, CD206−) or M^{mono} and AAM^{mono} (F4/80int, CD206+)
792 phenotype. Cell surface expression of PD-L2 and MHCII were acquired for analysis. Cells were
793 sorted using 100µm nozzle into FBS, on either BD FACSAriaII or SONY HAPS1, depending on
794 instrument availability.
795
796 **Assay for Transposase-Accessible Chromatin with Sequencing (ATAC-seq)** ATAC-seq was
797 performed as described by Buenrostro et al (*12*). 50,000 FACS-purified cells per sample were
798 spun down at 400g for 5 min at 4°C and washed once with 50µl cold PBS. Cells were lysed with
799 50µl lysis buffer (10 mM Tris-HCl, pH7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-630)
800 and immediately spun down at 1500rpm for 10 min at 4°C. The isolated cell nuclei were then
801 incubated for 30 min at 37°C with 50µl of transposase reaction, which contained 25µl Tagment
802 DNA buffer (Illumina), 2.5µl Tagment DNA enzyme (Illumina) and 22.5µl nuclease-free water.
803 The transposed DNA was immediately purified using the Qiagen MinElute PCR Purification Kit
804 (Qiagen) following manufacturer's guide and eluted at 10ul volume. PCR amplification of the
805 transposed DNA was done using a low-cycle number protocol and with primers published by
806 Buenrostro et al (*12*). Each PCR mix contained of 25µl of NEB 2x PCR Mix (New England
807 Biolabs), 2.5µl of 25µM forward primer (Primer Ad1_noMX), 2.5µl of 25µM reverse barcoded
808 primer, 0.3µl of 100x SYBR Green (Invitrogen) and 10µl of transposed DNA. PCR was carried
809 out with the cycling protocol: 72°C for 5 min, 98°C for 30s, followed by 5 cycles of 98°C for

810 10s, 63°C for 30s, 72°C for 1 min. The reaction was held at 4°C after the 5th cycle. A side qPCR
811 was set up using the PCR product from these 5 cycles of amplification. Each qPCR mix
812 contained 5µl NEB 2x PCR Mix, 0.25µl 25µM forward primer, 0.25µl 25µM reverse barcoded
813 primer, 0.06µl 100x SYBR Green, 4.44µl nuclease-free water and 5µl of the PCR-amplified
814 product. qPCR was carried out using the cycling protocol: 98°C for 30s, followed by 25 cycles
815 of 98°C for 10s, 63°C for 30s, 72°C for 1 min and plate read. The qPCR amplification plot was
816 then used to calculate the additional number of cycles needed for the PCR to achieve maximum
817 amount of product without going into saturation. Each sample was amplified for a total of 14-16
818 cycles. The amplified libraries were then purified using Qiagen MinElute PCR Purification kit
819 following manufacturer's guide and eluted at 20µl volume. Libraries were sequenced on the
820 HiSeq 2000 with 2 x 50 cycles and for an average of 50 million paired-end reads per sample. We
821 performed the IL-4c stimulation experiment twice and generated two independent sets of
822 libraries to obtain an optimal number of biological replicates for each macrophage population.
823 The two independent sets of libraries are referred here after as "Run 1" and "Run 2",
824 respectively. ATAC-seq libraries for C57BL/6 and BALB/c AAMs were generated using the
825 same protocol and sequenced in a single run.
826
827 **Transcriptional profiling of BALB/c and C57BL/6 AAMs** 100,000 cells were sorted per
828 sample as described above. FACS-purified cells were spun down and washed once with PBS
829 before lysis with 350µl of Buffer RLT from the RNeasy Mini Kit (QIAgen). RLT lysates were
830 homogenized by 1 minute of vortexing and were immediately stored at -80°C until RNA
831 isolation. RNA was isolated using the RNeasy Mini Kit (QIAgen) based on manufacturer's
832 protocol, with an additional DNase digestion step using the RNase-free DNase set (QIAgen).

833 Transcriptional profiling was done using the CEL-seq2 protocol (*35*) and library preparation was
834 performed at the NYU School of Medicine Genome Technology Center core facility. CEL-seq2
835 libraries were sequenced on the HiSeq 4000 with 2 x 50 cycles. While CEL-seq2 was originally
836 developed as a single-cell assay, we used this protocol in this study as a bulk transcriptional
837 profiling assay and use the more commonly-used terminology "RNA-seq" to describe data
838 generated from this assay.
839
840 **Re-stimulation experiment** 300,000 cells (M^{mono} or AAM^{mono} from C57BL/6 and BALB/c
841 mice, induced using the injection protocols described above) were sorted per sample as described
842 above. FACS-purified cells were washed twice, then re-suspended to 1ml PBS for cell counting
843 on the hemocytometer. Cells were spun down again and re-suspended to a single-cell suspension
844 of 100,000 cells per 200µl, using DMEM with 10% FBS and 1% Penicillin/Streptomycin. Two
845 aliquots of 100,000 cells were obtained from each sample and plated separately onto wells of a
846 48-well tissue-culture plate in 200µl. Cells were incubated overnight (approximately 18 hours) at
847 37°C, 5% CO_2. After the overnight incubation, the initial culture media was discarded and
848 replaced with fresh media. An *ex vivo* dose of IL-4 (20ng/ml) was added for designated wells.
849 Control wells received fresh media only. After 24 hours of stimulation, media was aspirated
850 completely from each well and 350µl of Buffer RLT was immediately added for cell lysis. The
851 lysed cells of each sample was transferred to a 1.5ml Eppendorf tube, vortexed for 1 minute and
852 immediately stored at -80°C until RNA isolation. RNA isolation and transcriptional profiling by
853 CEL-seq2 were performed as described above.
854

855 Bioinformatics and computational methods

856 **ATAC-seq sequence processing** Raw ATAC-Seq reads were aligned to the reference mouse
857 genome mm9 using bowtie2 (v2.2.9) (*36*), with the parameters --maxin 2000 and --local, while
858 keeping all other parameters at default settings. To keep only highly unique alignments, reads
859 with MAPQ score less than 30 were removed. We further removed all duplicate reads, as well as
860 reads mapping to mitochondrial DNA and chromosome Y. Read filtering steps were done using
861 the suite of tools from samtools (v1.2 and v1.3.1) (*37*), ngsutils (v0.5.9) (*38*) and picard-tools
862 (http://broadinstitute.github.io/picard/, v1.1.1 and v2.8.2). After all filtering steps, reads were
863 merged across all replicates from the same macrophage population. This resulted in a median
864 depth of 15,235,324 reads per macrophage population in Run 1 and 9,865,310 reads per
865 macrophage population in Run 2. For visualization of accessibility reads on the Integrative
866 Genomics Viewer (IGV), we merged reads from the same macrophage population across
867 samples from both runs, generated tiled data format (TDF) files using IGVtools and finally
868 normalized the merged reads to reads per million (RPM) (*39*).

869

870 **Identification of accessible chromatin regions** We used the merged reads for each macrophage
871 population to identify accessible chromatin regions, using the PeaKDEck (v1.1) peak calling
872 algorithm, which measures signal density from randomly sampled bins genome-wide before
873 generating a data set-specific probability distribution to identify regions with significant signal
874 enrichment (*40*). We ran PeaKDEck using sampling bins that consist of a 75bp central bin (-bin)
875 and a 10000bp background bin (-back). Sampling along the genome was done in steps (-STEP)
876 of 25bp and the background probability distribution was generated using 100000 randomly
877 selected sites (-npBack). Significance was defined using a p-value of less than 0.0001 and

878 regions with significant p-values were defined as a "peak" (i.e. an accessible chromatin region).
879 Peak calling was done independently on libraries generated from Run 1 and Run 2.
880
881 **Generation of a union set of accessible chromatin regions** We next counted the number of
882 reads present at each accessible region in order to analyze the ATAC-Seq data using quantitative
883 approaches downstream. To do this, we first generated a set of consensus peaks across the data
884 set by taking the union of peaks called from each macrophage population. Peaks were merged if
885 overlapping by 1bp or more. The number of reads at each peak within the union peak sets were
886 then counted for each sample. Finally, each peak was re-centered ±100bp on its summit, defined
887 as the position with maximum pile up of reads. Re-centering on peak summits was performed as
888 this should coincide with the binding event of a transcription factor within an accessible
889 chromatin region. We implemented the read counting and peak summit re-centering steps
890 directly using the dba.count function from the Bioconductor package DiffBind (version 1.14.2)
891 (*41*). The final count matrix, which consisted of 61,713 peaks, was used for downstream
892 analyses.
893
894 **Quantitative ATAC-seq analysis** ATAC-seq read counts were transformed using the
895 regularized logarithmic (rlog) transformation as implemented in the Bioconductor package
896 DESeq2 (*42*). To manage batch effect from the two separate libraries, we first modeled the rlog
897 accessibility read counts to the batch variable using a linear model and subtracted out the
898 coefficient contributed by the batch variable – this was implemented directly using the
899 removeBatchEffect function in limma (*43*). We next chose a set of 30,856 regions with high
900 variance, using the varFilter function in the genefilter package with default parameters, which

901 keeps only features with variance inter-quartile range > 0.5 (*44*). We performed principal
902 component analysis (PCA) using the batch-subtracted rlog read counts of these regions with high
903 variance using the prcomp function in R.

904 To identify IL-4 dependent accessible regions, we directly compared the ATAC-seq
905 profiles of IL-4 stimulated macrophages to their reference unstimulated macrophages, using a
906 differential analyses workflow directly implemented through DESeq2. We fit the negative
907 binomial model in DESeq2 using the raw accessibility reads from all 61,713 regions, with the
908 model ~ Batch + Population**,** where Batch is a variable describing if a sample belonged in Run 1
909 or Run 2, while Population is a variable describing if the sample is M^{res}, AAM^{res}, M^{mono} or
910 AAM^{mono}. IL-4 dependency was defined using a significance threshold of False Discovery Rate
911 (FDR) of 10%. To visualize IL-4 dependent regions, we scaled the batch-subtracted rlog read
912 counts of these IL-4 dependent regions by z-score transformation and next performed k-means
913 clustering on these scaled, rlog-transformed reads (K = 4). The clustered matrix was visualized
914 as a heatmap.

915

916 **Comparison of sequence properties between constitutively accessible and IL-4 induced**
917 **regions** *Identification of constitutively accessible and IL-4 induced regions*: To define a set of
918 constitutively accessible regions, we used only peaks from M^{res} and M^{mono}, respectively, that
919 were identified in both Run 1 and Run 2. This resulted in 8061 constitutively accessible regions
920 in M^{res} and 14,045 constitutively accessible regions in M^{mono}. IL-4 induced peaks were defined
921 using the differential analysis outlined above. All region overlap analyses throughout this study
922 were performed using the intersect function from the BEDTools suite (*45*) and overlaps were
923 defined as any regions overlapping by at least 1bp, unless otherwise noted.

924

925 *Genomic elements enrichment analysis*: We downloaded genome-wide annotations of five
926 different genomic elements (promoter, start exon, coding exon, end exon, intron) from the UCSC
927 Known Gene database for mm9 (*46*). We defined promoter elements as the 200bp-region
928 upstream of a transcriptional start site (TSS). We next assigned each of the 61,713 accessible
929 regions in our data set to a unique genomic element label. Where an accessible region overlapped
930 two different types of genomic elements, we assigned it to the element with higher number of
931 overlapping base pairs. Finally, any chromatin regions not assigned to one of these five genomic
932 elements were labeled as intergenic. To determine the enrichment levels of a particular type of
933 genomic element G within a given set of accessible regions A (either constitutively accessible or
934 IL-4 induced regions in AAM^{res} or AAM^{mono}), we used the binomial cumulative probability
935 distribution, $b(x; n,p)$, where x = number of success, n = number of trials and p = background
936 probability of success. We used the pbinom function in R. We defined x to be the number of
937 accessible regions in A that were labeled as the genomic element G that was being tested, n to be
938 the total number of genomic elements G detected in our combined data set and p to be the
939 proportion of the accessible region A to the total 61,713 accessible regions. This then gave the
940 enrichment levels of G in A, relative to all the accessible regions identified across the different
941 macrophage populations.

942

943 *G/C content analysis*: To calculate percentage GC, we first used the hgGcPercent function from
944 the kentTools suite (v20170111, UCSC Genome Bioinformatics Group,
945 https://github.com/ucscGenomeBrowser/kent) to quantitate the number of G or C bases in each
946 accessible region. This value was next normalized using the length of the accessible region. CpG

947 island track was downloaded from the UCSC Genome Annotation Database for mm9
948 (http://hgdownload.soe.ucsc.edu/goldenPath/mm9/database/). Enrichment levels of CpG island
949 in a given set of accessible regions *A* was based on the binomial cumulative probability as
950 described above, where x = number of accessible regions in *A* that overlapped a CpG island, n =
951 number of CpG island in the total data set of 61,713 regions and p = proportion of *A* to the total
952 61,713 regions.
953
954 *Calculation of distance to IL-4 induced genes*: IL-4 induced genes were first identified for
955 AAM^{res} and AAM^{mono} from the microarrays generated by Gundra and Girgis *et al* (*8*), using the
956 linear model and empirical Bayes statistics as implemented in limma, with genes significantly
957 induced by IL-4 defined using the thresholds FDR 10% and $\log_2$ fold change > 1. The distance
958 between each accessible region and its closest IL-4 induced gene body was calculated using the
959 closest function in BEDTools.
960
961 **Transcription factor (TF) motif analysis** Sequences of IL-4 induced regions were fetched
962 using the BEDTools getfasta function for TF motif analysis with the MEME Suite tools (*47*).
963 Whole genome fasta file for mm9 was downloaded from the Illumina igenome database
964 (https://support.illumina.com/sequencing/sequencing_software/igenome.html). Background file
965 was generated using the function fasta-get-markov in MEME, based on the total 61,713
966 accessible regions at a Markov model order of 3. TF motif databases (which included mouse and
967 human TF motifs) were curated as described in (*48*). We performed *de novo* motif discovery by
968 running MEME (as part of MEMEChIP, which randomly sampled 600 sequences) with the
969 parameters: -mod zoops -nmotifs 3 -minw 6 -maxw 30. Over-representation analysis to identify

970 macrophage-specific TF motifs was performed by running HOMER (*23*) using sequences from
971 the opposing macrophages as background sequences (i.e. to identify TF motifs specific to
972 AAM^{res}, sequences of IL-4 induced regions from AAM^{mono} were used as background sequences)
973 and the parameters: -mask, -size 8,10,12,16, -mset vertebrates, -nlen 3. For motifs from each
974 macrophage lineage, we combined all three *de novo* motifs discovered by MEME and motifs
975 with enrichment $\log_2$ p-values < -15 (approximately p-values < 10^{-6}) by HOMER for clustering
976 analysis (we used the known motifs output from HOMER). This resulted in 14 motifs for
977 AAM^{res} and AAM^{mono}, respectively. Clustering was done using STAMP (*49, 50*), with the
978 frequency matrices of motifs and the default parameters of: column comparison metric – Pearson
979 correlation coefficient, alignment method – ungapped Smith-Waterman, tree-building algorithm
980 – UPGMA, multiple alignment strategy – iterative refinement.

981 We identified IL-4 induced PU.1 motif sites by performing motif scanning with FIMO
982 (*51*), using the PU.1 motif discovered *de novo* from the IL-4 induced peaks of AAM^{res} and
983 AAM^{mono}, respectively. FIMO was run with a p-value threshold of 0.0001 (as part of
984 MEMEChIP). These IL-4 induced PU.1 motif sites were then extended ±25bp using the
985 BEDTools slop function. These PU.1 motif sites ± 25bp are hereafter referred to as "PU.1
986 regions". To identify potential co-factors for PU.1, PU.1 regions were specifically subjected to
987 motif scanning by FIMO. To determine which of these were macrophage-specific, the odds ratio
988 of a motif being detected in the PU.1 regions of AAM^{res} vs. AAM^{mono} were calculated. P-value
989 was determined using two-sided Fisher's test, with the null hypothesis of a motif being equally
990 likely to be detected in the IL-4 induced PU.1 regions of AAM^{res} and AAM^{mono} (i.e. $\log_2$ odds
991 ratio of zero). Regions with different motifs of the same TF from the same organism (one of
992 either mouse or human) were combined into a single test. All TFs from the ETS family (of which

993 PU.1 belongs to) were excluded from this analysis. To account for the multiple hypotheses
994 testing performed over 612 mouse non-ETS TFs and 807 human non-ETS TFs, the Benjamini-
995 Hochberg procedure was used to perform p-value adjustment by calculating the FDR and
996 significance threshold was set at FDR 10%. Hence, significant TFs were macrophage-specific,
997 with $\log_2$ odds ratio > 0 AAM^{res}-specific and $\log_2$ odds ratio < 0 AAM^{mono}-specfic. For
998 visualization of significant results, TFs were summarized at the family level as defined in (*52*)
999 (Figure 3B). The maximum absolute $\log_2$ odds ratio of the family was visualized and the specific
1000 TF with the maximum absolute $\log_2$ odds ratio value was stated in parenthesis. Where TF family
1001 annotation was not available, the $\log_2$ odds ratio of the specific TF itself was used.

1002 DNA shape features of these PU.1 motif sites ± 25bp regions were analyzed using the
1003 DNAshape algorithm (*21, 53*) for 4 different DNA shape configurations at single nucleotide
1004 resolution. The same analysis steps were repeated for the KLF motif ± 25bp regions and AP-1
1005 motif ± 25bp regions as presented in Supplemental Figure 3. Motif scores were calculated as part
1006 of the motif scanning procedure using FIMO.

1007

1008 **Comparisons between predicted PU.1 motif and ChIP-seq defined PU.1 binding sites** PU.1
1009 ChIP-seq regions identified in thioglycollate-induced macrophages, generated from two different
1010 experiments, were downloaded as BED files that had been directly deposited on Gene
1011 Expression Omnibus (GEO) (GSM1131238 and GSM1183968) (*5, 18*). A set of 55,386
1012 reproducible PU.1 binding sites were defined by intersecting these two sets of PU.1 ChIP-seq
1013 regions. We used the PU.1 motif discovered *de novo* from all the constitutively accessible
1014 regions in M^{mono} and ran FIMO to identify all PU.1 motif sites from M^{mono}, using a p-value
1015 threshold of 0.0001 and the background file generated as described above. Since the published

1016 PU.1 ChIP-seq regions were of 200bp length, we extended the predicted PU.1 motifs from M^{mono}
1017 by ±100bp to match the comparison. Overlapping rate was calculated as total number of
1018 predicted PU.1 motifs from M^{mono} overlapping a reproducible PU.1 ChIP-seq region / total
1019 number of predicted PU.1 motifs from M^{mono} × 100%.
1020
1021 **Identification of cell-type-specific PU.1 binding sites** We first downloaded PU.1 ChIP-seq
1022 regions as BED files directly deposited on GEO (GSM537983, GSM537989, GSM1315477) (*22,*
1023 *23*) and lifted over regions annotated on the mm8 mouse reference genome to the mm9 mouse
1024 reference genome using the UCSC liftover tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver). We
1025 next defined cell type specific PU.1 binding sites as PU.1 ChIP-seq regions identified in only one
1026 cell type and not in the other two. Finally, we fetched the sequences for each set of cell type
1027 specific PU.1 binding sites using the getfasta function in BEDtools. We performed *de novo* motif
1028 discovery using MEME (as part of MEMEChIP) with the parameters -mod zoops -nmotifs 3 -
1029 minw 6 -maxw 30 and a background Markov model of order 3. We used the PU.1 motif
1030 discovered *de novo* and ran FIMO to identify cell type specific PU.1 motif sites that were ChIP-
1031 seq defined. FIMO was run using a raw p-value threshold of 0.0001 and only the top 1000 motif
1032 sites (based on motif score) were included. Finally, these cell type specific PU.1 motif sites were
1033 extended ±25bp for DNA shape analysis. We used only the top 1000 motif sites to keep the
1034 number of motif sites comparable to the number of IL-4 induced PU.1 motif sites from our own
1035 study.
1036
1037 **ATAC-seq sequence processing from C57BL/6 vs. BALB/c experiment** Raw ATAC-seq
1038 reads were aligned to the mm10 mouse reference genome (for compatibility with available

1039 BALB/c sequence variant data, details below). Reads with MAPQ < 30, as well as duplicate and
1040 mitochondrial reads were all removed before peak calling. Peak calling in PeaKDEck was done
1041 using the same parameters in the IL-4c stimulation experiment, except for a p-value threshold of
1042 0.00001. A consensus count matrix consisting of 40,981 accessible regions from all samples was
1043 generated as described above using DiffBind and used to identify strain-specific regions using
1044 the differential analysis workflow implemented through DESeq2. We fit the negative binomial
1045 model using ~ Strain + CellType + Strain:CellType, where Strain is one of either C57BL/6 or
1046 BALB/c and CellType is one of either AAM^{res} or AAM^{mono}, and extracted strain-specific regions
1047 from the comparisons: (1) BALB/c vs. C57BL/6 in AAM^{res} and (2) BALB/c vs. C57BL/6 in
1048 AAM^{mono}. Significance was defined at a threshold of FDR 10%.
1049
1050 **Characterizing different types of BALB/c variants at PU.1 motif regions** BALB/c-specific
1051 sequence variants, including single nucleotide polymorphisms (SNPs) and insertion/deletion
1052 (indels), were downloaded as variant call format (VCF) files from the Sanger Mouse Project
1053 (ftp://ftp-mouse.sanger.ac.uk/). Variants that were not labeled as "PASS" under the Filter flag
1054 were removed using vcftools (*54*). These VCF files were then converted to the BED file format
1055 using the vcf2bed function from bedops (*55*) for overlap analysis with the IL-4 induced PU.1
1056 regions. Since these BALB/c variants were identified on the mm10 genome, we lifted over IL-4
1057 induced PU.1 regions defined initially on the mm9 coordinates to the mm10 coordinates. PU.1
1058 motif sites that directly overlapped sequence variants were identified using the BEDTools
1059 intersect function. To define PU.1 motif sites with flanking variants, we use the BEDTools
1060 closest function to calculate distances between a PU.1 motif and its closest BALB/c variant and
1061 identified motifs with a closest variant at an absolute distance of <25bp, but not overlapping a

1062 variant directly. We also limited the DNA shape prediction to SNPs only, to avoid the shift in
1063 position due to the insertions/deletions.
1064
1065 **Generation of BALB/c whole genome fasta** To construct a BALB/c whole genome fasta for
1066 DNA shape prediction, we took the list of BALB/c variants filtered from above and used the
1067 vcf2diploid tool (*56*) to integrate these BALB/c variants into the mm10 reference genome. To
1068 annotate PU.1 regions on the BALB/c coordinates, we used the accompanying chain file
1069 generated from the vcf2diploid tool in the above step to liftover the PU.1 regions from mm10
1070 coordinates to BALB/c coordinates. The mm10 and BALB/c whole genome fasta, as well as
1071 PU.1 regions on corresponding C57BL/6 and BALB/c coordinates, were then used with the
1072 getfasta function in BEDTools to fetch the sequences of specific IL-4 induced PU.1 regions with
1073 flanking SNPs for shape prediction with DNAShape algorithm.
1074
1075 **Quantitating the amount of shape change induced by SNP** To quantitate the amount of shape
1076 change induced by the presence of BALB/c variant(s) at an accessible region, we calculated the
1077 Euclidean distance between the two vectors of shape values that corresponded to shape values
1078 predicted from a C57BL/6 region and a BALB/c region, respectively (*57*). A larger Euclidean
1079 distance would indicate a greater shift in shape. We first identified all accessible regions that
1080 carried SNP(s) only from across the data set (i.e. regions that did not contain any sequence
1081 variants, or regions that contained both SNPs and insertion/deletions were discarded), as
1082 including insertion/deletions in the analysis would result in comparisons between sequences of
1083 unequal length and it would not be feasible to calculate the Euclidean distance at these regions.
1084 We then grouped regions as “strain-specific” and “strain-common” based on results from the

1085 differential analysis (922 strain-specific and 4776 strain-common chromatin regions; 18 strain-
1086 specific and 63 strain-common PU.1 regions). Shape prediction was done using the DNAShape
1087 algorithm and Euclidean distance was calculated for each pair of shape values using the dist
1088 function in R.

1089

1090 **Identification of predictors for strain-specific and strain-common regions** To identify
1091 features associated with the 922 strain-specific and 4776 strain-common SNP-carrying accessible
1092 regions, we first assessed each of these accessible regions for the following sequence
1093 characteristics:

1094 - $GCPercent_{B6}$ = % GC calculated using C57BL/6 sequences
1095 - SnpFrequency = Number of SNPs carried by the region
1096 - IL4Inducibility = A binary variable (Yes / No) indicating if the region was IL-4 inducible
1097 - NonCoding = A binary variable (Yes / No) indicating if the region is a non-coding
1098 regulatory element (Yes – one of either "Intron", "Intergenic" or "Promoter" accessible
1099 regions; No – one of either "Start exon", "End exon" or "Coding")

1100 We then fit a logistic regression model as follows using the glm function in R, with the
1101 parameter family = "binomial":

1102 $$Class \sim GCPercent_{B6} + GenomicElements + SnpFrequency + IL4Inducibility$$

1103 The dependent variable "Class" is a label describing if the region is "strain-specific" or "strain-
1104 common". P-value of predictors were determined by analysis of variance for the coefficients of
1105 the predictors.

1106

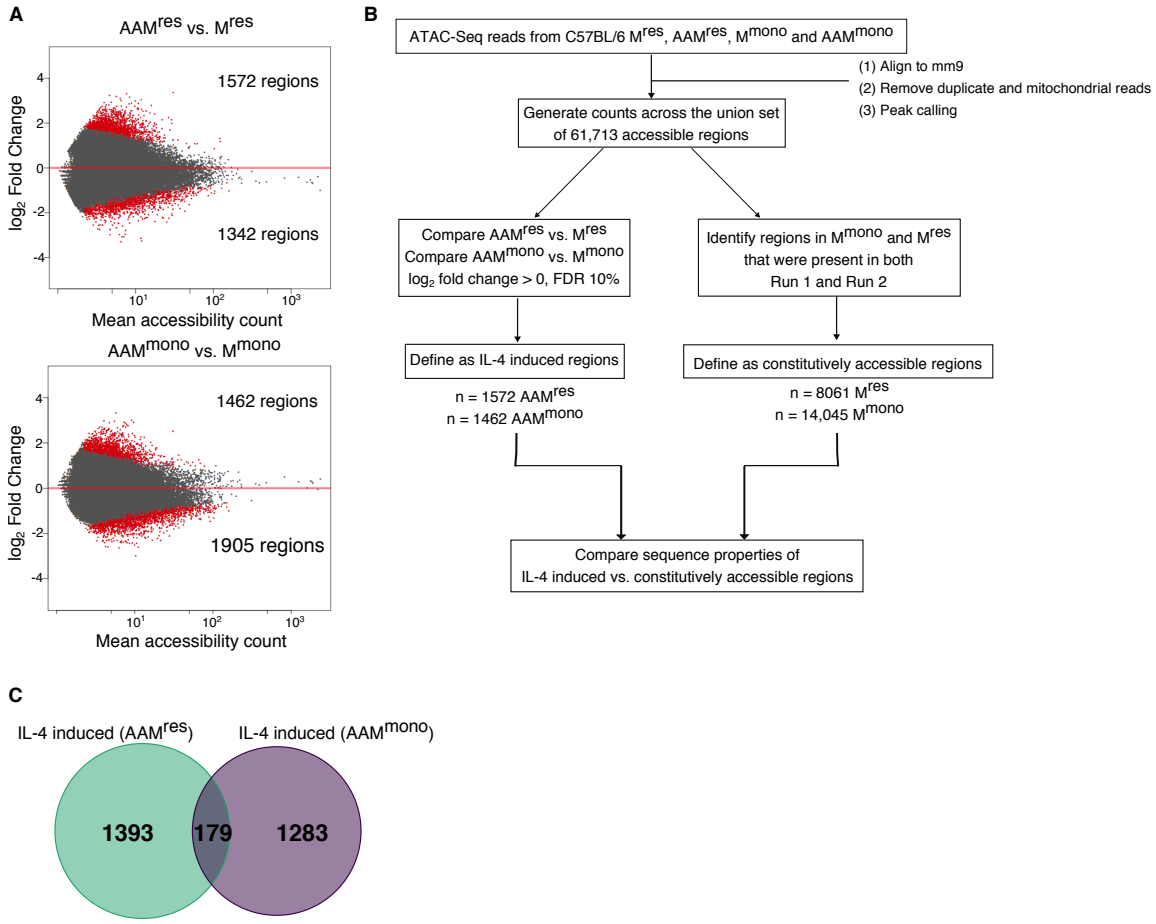
1107 **Processing of CEL-seq reads** CEL-seq reads were first demultiplexed using the bc_demultiplex
1108 script from https://github.com/yanailab/CEL-Seq-pipeline (*35*). Demultiplexed reads were
1109 aligned to the mm10 mouse reference genome using bowtie2 (version 2.2.9). Aligned reads were
1110 counted for each gene using a modified htseq-count script (from
1111 https://github.com/yanailab/CEL-Seq-pipeline) adapted for CEL-seq reads with unique
1112 molecular identified (UMI). We included only reads with MAPQ score > 30 and removed
1113 singleton genes. This resulted in a final median read depth of 737,848 reads per sample, covering
1114 a median of 11,096 genes per sample. PCA was performed using 7431 genes with high variance,
1115 defined using the varFilter function in the genefilter package with default parameters, which
1116 keeps only features with variance inter-quartile range > 0.5. Differential analysis was done using
1117 DESeq2 by fitting the negative binomial model using ~ Strain + CellType + Strain:CellType.
1118 Significantly differential genes were extracted using a threshold of FDR 10% for the four
1119 different comparisons of: (1) AAM^{res} vs. AAM^{mono} in C57BL/6, (2) AAM^{res} vs. AAM^{mono} in
1120 BALB/c, (3) BALB/c vs. C57BL/6 in AAM^{res} and (4) BALB/c vs. C57BL/6 in AAM^{mono}.
1121 Overlapping genes were defined as genes identified as differential in two different comparisons.
1122 Pathway enrichment analysis was done using strain-specific genes in AAM^{res} and AAM^{mono},
1123 respectively, through Ingenuity Pathway Analysis (IPA) with the parameter Organism = Mouse
1124 and keeping all other parameters at default settings.
1125
1126 **Defining gene classes from re-stimulation experiment** CEL-seq reads were processed as
1127 described above. This resulted in a final median read depth of 979,547 reads per sample,
1128 covering a median of 10,820 genes per sample. PCA was separately performed on the rlog read
1129 counts from 6528 high variance genes from C57BL/6 and BALB/c, respectively. The differences

1130 in transcriptional profiles induced by *ex vivo* IL-4 stimulation were represented as Euclidean
1131 distances, which were calculated between transcriptional profiles (rlog read counts of all non-
1132 singleton genes) of cell aliquots from the same cell type that either received a secondary *ex vivo*
1133 IL-4 stimulation or cultured in control media. Differential analysis was done using DESeq2 by
1134 fitting the negative binomial model using ~ Group, where Group is a factor consisting of 8
1135 different levels (M^{mono} from C57BL/6, AAM^{mono} from C57BL/6, M^{mono} with *ex vivo* IL-4
1136 stimulation from C57BL/6, AAM^{mono} with *ex vivo* IL-4 stimulation from C57BL/6, M^{mono} from
1137 BALB/c, AAM^{mono} from BALB/c, M^{mono} with *ex vivo* IL-4 stimulation from BALB/c, AAM^{mono}
1138 with *ex vivo* IL-4 stimulation from BALB/c). Significantly differential genes were extracted
1139 using a threshold of FDR 10% and the comparisons are described on Figure 7B.

1140 To define "Unresponsive", "Enhanced" and "Equal" genes, we identified genes that were
1141 either uniquely or commonly differential in the comparisons "AAM^{mono} with *ex vivo* IL-4 vs.
1142 AAM^{mono}" and "M^{mono} with *ex vivo* IL-4 vs. M^{mono}" (*29*). "Unresponsive" genes were defined as
1143 genes that were uniquely differential in "M^{mono} with *ex vivo* IL-4 vs. M^{mono}", while "Enhanced"
1144 genes were uniquely differential in "AAM^{mono} with *ex vivo* IL-4 vs. AAM^{mono}". For genes that
1145 were differential in both comparisons, we further examined the magnitude of fold change – if the
1146 fold change (defined as a differences in $\log_2$ fold change of >1) is greater in the "M^{mono} with *ex*
1147 *vivo* IL-4 vs. M^{mono}" comparison, it is defined as an "Unresponsive" genes and if the fold change
1148 is greater in the "AAM^{mono} with *ex vivo* IL-4 vs. AAM^{mono}" comparison, it is defined as an
1149 "Enhanced" gene. Finally, if the magnitude of fold change is similar in both comparisons
1150 (defined as a differences in $\log_2$ fold change of <1), the gene is defined as "Equal". This
1151 categorization process of different gene classes was performed for C57BL/6 and BALB/c
1152 macrophages, respectively. For visualization of these different gene groups, the rlog read counts

1153 from all of these genes were extracted, scaled to a mean of 0 and standard deviation of 1 across
1154 each sample, and subjected to k-means clustering, with K=6. Heatmap was visualized using the
1155 R package pheatmap.
1156
1157

1158 **Supplemental figures**



1159

1160 **Fig. S1.** IL-4 stimulation leads to remodeling of open chromatin landscape in peritoneal
1161 macrophages. (**A**) Comparisons between the accessible chromatin profiles of IL-4
1162 stimulated macrophages and unstimulated macrophages, presented as MA plots (left –
1163 AAM^{res}; right – AAM^{mono}). Differential chromatin regions (FDR 10%, |LFC| > 0) are
1164 highlighted in red. (**B**) Venn diagram illustrating the minimal overlap between IL-4
1165 induced regions in AAM^{res} and AAM^{mono}. Values in Venn diagrams represent the
1166 number of unique accessible regions in each corresponding set. (**C**) Schematic outlining
1167 the workflow to identify IL-4 induced regions and constitutively regions for
1168 comparisons of the sequence properties associated with these two classes of genomic
1169 elements.
1170

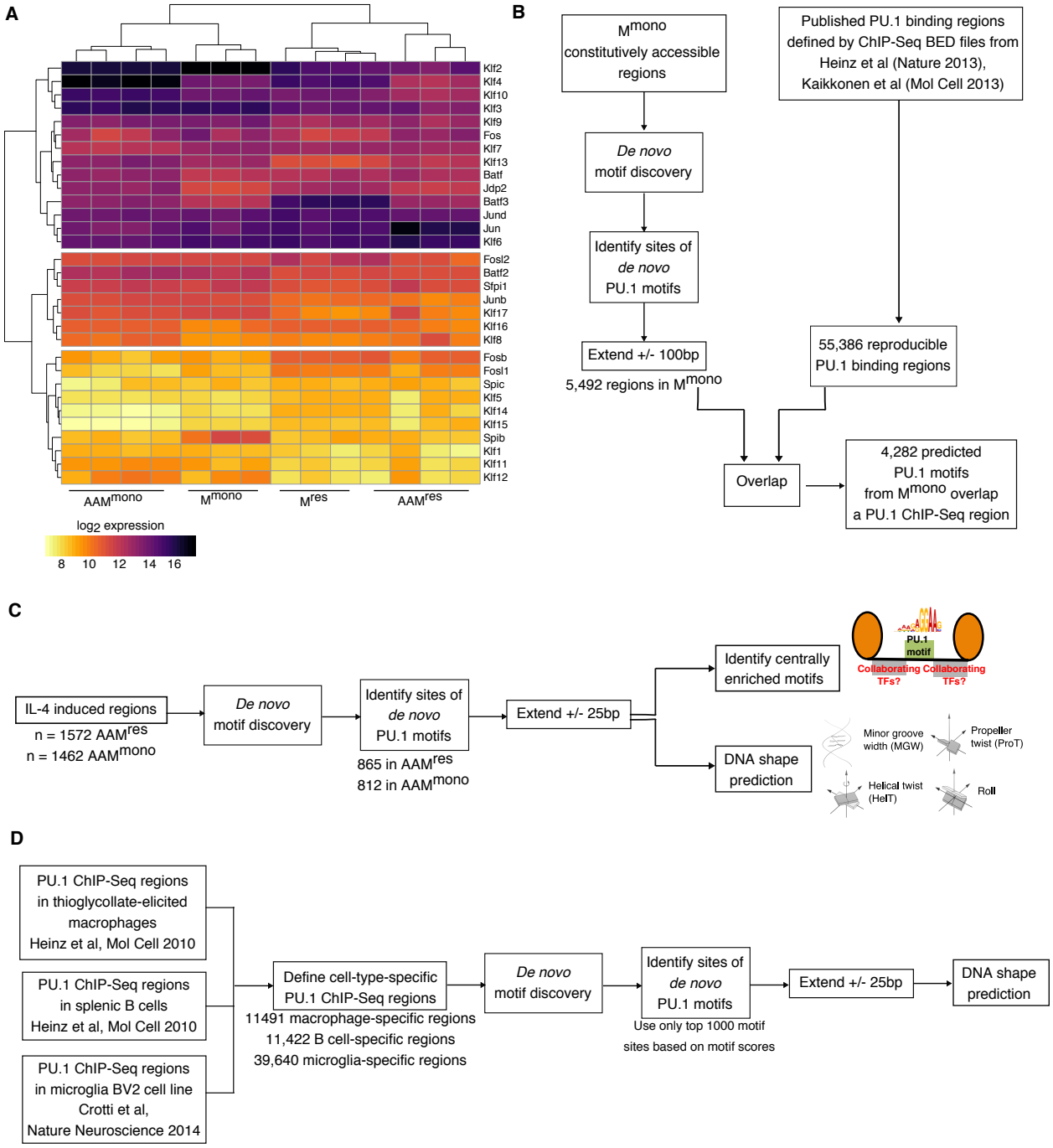


Fig. S2. Motif analyses on IL-4 induced regions. (**A**) Expression values of individual TF members that belonged to either PU.1, KLF or AP-1 family. Expression values were obtained from (*8*) and represents $\log_2$ intensity values. Only TFs in clusters 1 and 2 are considered highly expressed and included in Fig. 2E. (**B**) Schematic outlining the workflow used to verify the accuracy of predicted PU.1 motif sites by comparing to published ChIP-Seq data sets. (**C**) Schematic outlining the workflow used to analyze features of IL-4 induced PU.1 motif sites. (**D**) Schematic outlining the workflow used to identify cell-type-specific PU.1 binding regions from published ChIP-Seq data sets for DNA shape prediction.

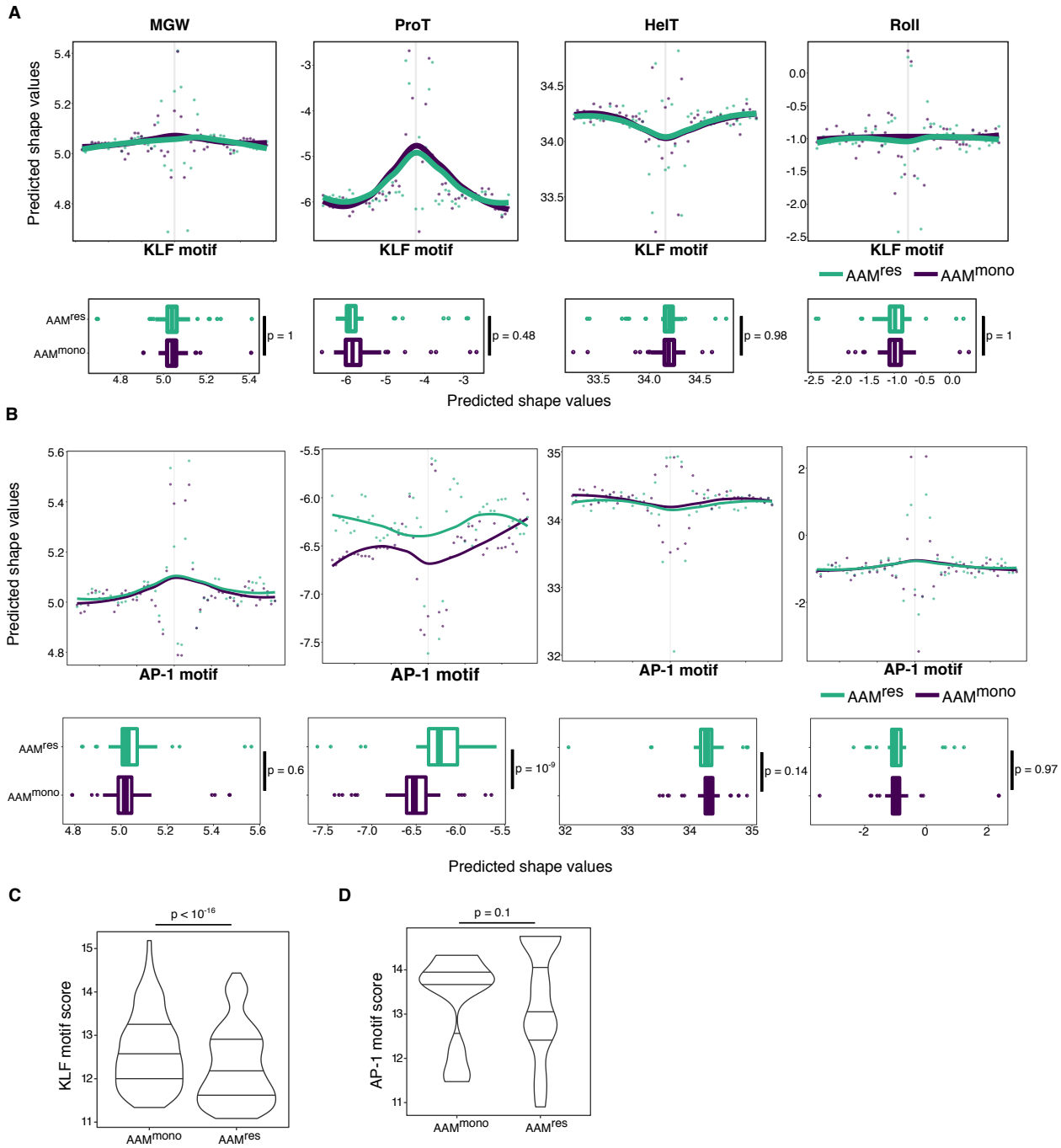


Fig. S3. Sequence features of KLF and AP-1 motifs. Average of DNA shape features at IL-4 induced (**A**) KLF regions of AAM^{res} (902 regions) and AAM^{mono} (627 regions) and (**B**) AP-1 regions of AAM^{res} (336 regions) and AAM^{mono} (497 regions). Scatter plots are centered on IL-4 induced KLF motifs (x-axes), with the solid lines representing Loess fit of predicted shape values at single-nucleotide resolution (dots). Boxplots are average predicted shape values over the KLF motif ± 25bp windows and p-values are from two-sided Kolmogorov-Smirnov test. FIMO-derived scores of (**C**) KLF and (**D**) AP-1 motifs discovered from IL-4 induced regions. P-values from two-sided Mann-Whitney test.

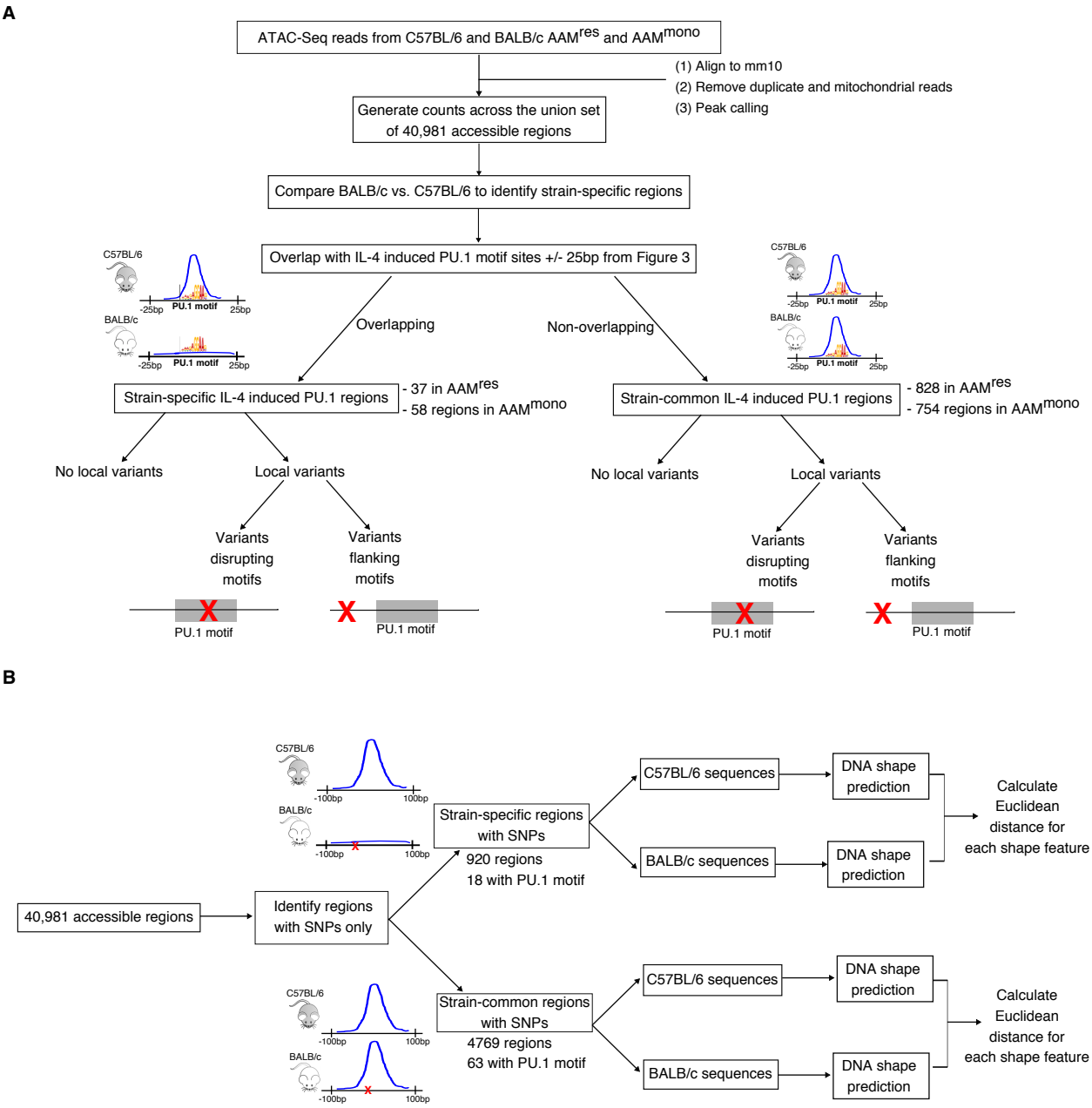


Fig. S4. Leveraging on natural genetic variation in C57BL/6 and BALB/c mouse strains to perform a mutation study. (**A**) Schematic with graphical illustrations outlining the strategy used to identify IL-4 induced PU.1 motif sites with altered chromatin accessibility due to flanking SNP(s). Blue line indicates read pile-ups from an accessible region, while the "X" in red represent a motif-flanking SNP. (**B**) Schematic with graphical illustrations outlining the strategy used to classify strain-specific and strain-common IL-4 induced PU.1 regions and all accessible regions carrying SNP(s), as well as to quantitate the amount of DNA shape change induced by SNP(s) in these regions.

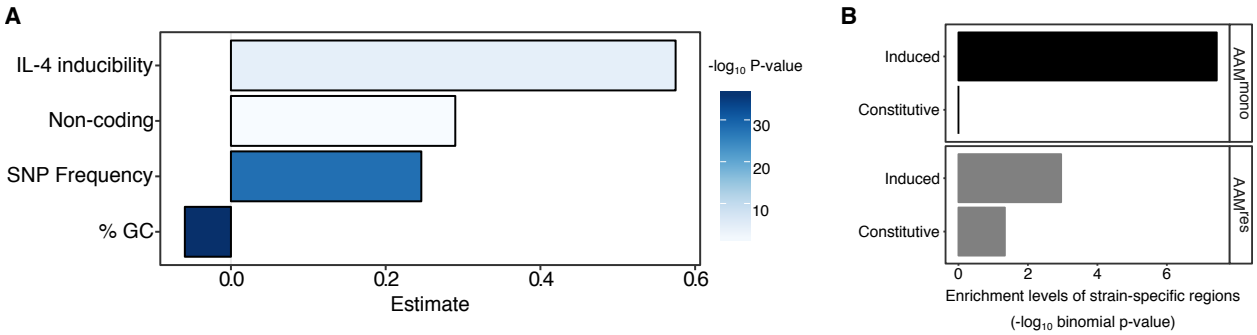


Fig. S5. Genomic features associated with strain-specificity. (**A**) Coefficient values representing the relative contribution of each significant predictor in distinguishing between strain-specific and strain-common accessible regions. P-values are from analysis of variance. (**B**) Enrichment values of IL-4 induced and constitutively accessible regions in strain-specific regions for AAMmono and AAMres. Enrichment p-value is based on binomial test.