



**HAL**  
open science

# Characterising the genetic basis of immune response variation to identify causal mechanisms underlying disease susceptibility

Maxime Rotival

► **To cite this version:**

Maxime Rotival. Characterising the genetic basis of immune response variation to identify causal mechanisms underlying disease susceptibility. *HLA: Immune Response Genetics*, 2019, 94 (3), pp.275-284. 10.1111/tan.13598 . pasteur-02868915

**HAL Id: pasteur-02868915**

**<https://pasteur.hal.science/pasteur-02868915>**

Submitted on 17 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38

**Characterising the genetic basis of immune response variation to identify causal mechanisms underlying disease susceptibility**

Short title: genetic basis of immune response variation

**Maxime Rotival<sup>1,\*</sup>**

<sup>1</sup>Unit of Human Evolutionary Genetics, CNRS UMR2000, Institut Pasteur,  
28 rue du Docteur Roux, 75015 Paris, France;

\*Correspondence: [maxime.rotival@pasteur.fr](mailto:maxime.rotival@pasteur.fr)  
Tel : +33 1 40 61 36 34

**Keywords:** GWAS, immune response, eQTL, regulatory elements, population genetics, selective constraint, adaptation.

39 **Abstract (250 words)**

40 Over the last ten years, genome-wide association studies (GWAS) have identified  
41 hundreds of susceptibility loci for autoimmune diseases. However, despite increasing  
42 power for the detection of both common and rare coding variants affecting disease  
43 susceptibility, a large fraction of disease heritability has remained unexplained. In  
44 addition, a majority of the identified loci are located in non-coding regions, and  
45 translation of disease-associated loci into new biological insights on the aetiology of  
46 immune disorders has been lagging. This highlights the need for a better  
47 understanding of non-coding variation and new strategies to identify causal genes at  
48 disease loci. In this review, I will first detail the molecular basis of gene expression  
49 and review the various mechanisms that contribute to alter gene activity at the  
50 transcriptional and post-transcriptional level. I will then review the findings from 10  
51 years of functional genomics studies regarding the genetics on gene expression, in  
52 particular in the context of infection. Finally, I will discuss the extent to which genetic  
53 variants that modulate gene expression at transcriptional and post-transcriptional  
54 level contribute to disease susceptibility and present strategies to leverage these  
55 information for the identification of causal mechanisms at disease loci in the era of  
56 whole genome sequencing.

57

58

59 **Introduction**

60 The host immune response to stress exhibits considerable variation, both at the  
61 individual and population levels<sup>1</sup>. While many non heritable factors contribute to this  
62 variability, including age, gender or past exposures to pathogens<sup>2-5</sup> (see<sup>1</sup> for a review  
63 on these aspects). There is also widespread evidence that genetics plays a  
64 significant role in determining the efficacy of the human immune response, in  
65 particular in the context of innate immunity<sup>6-12</sup>. Genome-wide association studies  
66 (GWAS) have been a considerable asset in defining the genomic regions involved in  
67 immune disease susceptibility<sup>13</sup>. However, identifying the causal mechanisms  
68 through which GWAS variants contribute to disease susceptibility has proven to be  
69 extremely challenging. Indeed, it was soon realized that most GWAS pointed toward  
70 gene deserts or non-coding regions<sup>14,15</sup>, highlighting the strong need for  
71 characterization of the molecular mechanisms through which such non-coding  
72 regions may alter immune disease susceptibility<sup>16</sup>. Furthermore, while large GWAS  
73 meta-analyses have granted increasing power to capture the effect of common  
74 variants on disease susceptibility, a large fraction of disease heritability has  
75 remained unaccounted for<sup>17</sup>. There is now increasing evidence to suggest that rare  
76 variants, non-coding in particular, may contribute to explain this missing heritability  
77 <sup>18-20</sup>. Therefore, there is a growing need to understand how non-coding variants,

78 whether rare or common, contribute to the variability of immune phenotypes in  
79 humans. In this review, I will first present the various mechanisms through which non-  
80 coding elements may contribute to regulate gene expression, either at the  
81 transcriptional level, or post-transcriptionally through RNA splicing and degradation.  
82 Then, I will review current knowledge of how common genetic variants drive immune  
83 gene expression at the transcriptional and post-transcriptional level. Finally, I will  
84 discuss strategies to decipher the impact of rare regulatory variants on immune  
85 response and characterize the causal mechanisms that determine immune disease  
86 susceptibility.

87

### 88 ***The molecular bases of gene expression***

89 In multi-cellular organisms, tight regulation of gene expression programs is essential.  
90 First, during development, cells must coordinate expression of various transcription  
91 factors to divide and differentiate into future organs. Then, in adult life, regulatory  
92 sequence need to allow for cell-type specific gene expression profiles, that are both  
93 (i) robust to genetic variations and (ii) able to react appropriately to external stimuli.  
94 In human, such combination of resilience and adaptability is achieved through a  
95 network of regulatory features, the complexity of which largely exceeds that of the  
96 protein coding genome. Indeed, while coding sequence represent only 1.5% of the

97 human genome, there is evidence that at least 20% of the genome is covered by  
98 regulatory features, and 10% is actively bound by known transcription factors<sup>21</sup>.

99

### 100 ***Promoters and enhancers shape transcriptional activity across tissues***

101 The regulatory features that control transcription are characterized by epigenetic  
102 histone modifications (see **Table 1** and ref.<sup>22</sup>) and can be decomposed in two main  
103 categories (**Figure 1a**). At the transcription start site, promoters recruit RNA  
104 polymerase, along with a variety of other transcription factors, to initiate gene  
105 transcription<sup>15,23,24</sup>. While 96% of promoters are active in more than one tissue or cell  
106 type, only ~20% are ubiquitously expressed indicative of strong tissue-specificity of  
107 gene regulation<sup>23</sup>. Tissue-specificity is further increased by distal enhancer regions  
108 — located up to 1Mb away from the promoters they regulate — which, when  
109 activated by tissue-specific transcription factors, are brought in contact with their  
110 associated promoter by looping of the DNA molecule (**Figure 1b,c**) to modulate their  
111 activity<sup>15,24,25</sup>. Activation of both promoters and enhancers is associated to an  
112 opening of the chromatin and acetylation of H3K27 histone residues (**Figure 1d,e**),  
113 together with transcription of non-coding RNAs<sup>22,25,26</sup>. This transcription occurs either  
114 in the opposite direction of transcriptional elongation at promoters (upstream  
115 antisense RNAs), or bi-directionally for enhancers (enhancer RNAs). Although such

116 RNAs are rapidly degraded by exosomes, they allow precise quantification of  
117 enhancer/promoter activity by sequencing techniques that target the 5' end of  
118 transcripts such as CAGE-seq or GRO-Cap<sup>23,25,27</sup>.

119 Over the last 10 years, large Consortia — such as Fantom5<sup>23,25</sup>, ENCODE<sup>26</sup>,  
120 Blueprint<sup>28-31</sup> or Roadmap Epigenomic<sup>22</sup> — have combined these techniques with  
121 Chromatin immuno-precipitation (ChIP-seq) and chromatin accessibility Assays  
122 (DNase-Seq, FAIRE-Seq, or ATAC-Seq) to precisely map promoters and enhancers  
123 along the genome and characterize of their activity across tissues. In doing so,  
124 these studies have provided valuable insights on the regulatory architecture of the  
125 response to immune stimuli. Notably, immune enhancers were shown to have a  
126 distinct sequence composition, DNA structure and activity patterns<sup>32</sup>, with their  
127 activation preceding that of their proximal promoters during immune response,  
128 revealing their crucial role in the early transcriptional response to immune stimuli<sup>33</sup>.

129

### 130 ***Three-dimensional chromatin conformation orchestrates transcription***

131 Because enhancers are located far from the genes they regulate,  
132 understanding how they are brought into contact with their target gene is of primary  
133 importance to decipher the complex networks of that underlies transcriptional  
134 regulation in humans. Chromosome Conformation Capture techniques (Hi-C, and

135 Promoter-capture Hi-C, Chia-PET) have recently enabled the detection of DNA-DNA  
136 interactions along the genome<sup>30,31,34,35</sup>, allowing the assignment of enhancers to their  
137 target genes. In practice, a single gene is often regulated by multiple enhancers, and  
138 one enhancer can regulate multiple genes<sup>15,24</sup>. Notably, ~70% of these interactions  
139 are restricted to specific regions, known as Topologically associating domains or  
140 TADs, which can spread across hundreds of kilobases<sup>24,31</sup>. The boundaries of these  
141 TADs are genetically encoded — with evidence of TADs being modified by targeted  
142 deletions and genetic variants occurring at the boundary<sup>35,36</sup> — and are enriched in  
143 are binding sites of the insulating factor CTCF. Note, however, that CTCF binding  
144 sites have also been observed within TADs<sup>35</sup>, indicating that CTCF binding is not  
145 sufficient to induce a TAD boundary. Growing evidence suggests that TADs act to  
146 facilitate efficient and insulated co-regulation of functionally related genes. In  
147 particular, they were reported to segregate promoters of immune gene together to  
148 facilitate their epigenetic priming<sup>37</sup>.

149 Within TADs, promoter/enhancer conformation exhibits a wide variability  
150 across tissues, and was proposed to underlie cell type-specificity of enhancer  
151 functions<sup>31</sup>. In particular, several studies have reported the existence of enhancers  
152 acting as hubs for chromatin-chromatin interactions, linking promoters to other  
153 enhancers<sup>30,38,39</sup>. These hubs — sometimes referred as super-enhancers<sup>38</sup>,



154 frequently interacting regions<sup>30</sup> or initiator enhancers<sup>39</sup> — are tissue-specific<sup>30</sup>,  
155 preferentially found next to cell identity genes<sup>30,38</sup> and enriched in tissue-specific  
156 eQTL<sup>39</sup>. More importantly, super enhancers are strongly enriched in disease-related  
157 SNPs and largely contribute to auto-immunity<sup>30,38</sup>. For instance, up to 22% of SLE  
158 GWAS loci fall into B-cell specific super enhancers, with disease-linked super-  
159 enhancers being observed at key immune genes such as HLA-DRB1, HLA-DQA1,  
160 the interferon response factor IRF5 or the NF- $\kappa$ B inhibitor TNFAIP3<sup>38</sup>.

161

### 162 ***Post-transcriptional regulation through alternative splicing***

163 Beyond transcription, the control of gene expression is also ensured at the  
164 post-transcriptional level through alternative splicing (AS) and miRNA-mediated  
165 degradation. AS can change the function of the protein isoform that is being  
166 expressed, through addition/removal of exons encoding functional domains. The  
167 most iconic example of such a mechanism is the splicing transition of CD45 antigen  
168 during training of adaptive immunity, where T-cells switch from the CD45RO epitope  
169 that characterizes naïve T-cells to the CD45RA epitope associated with memory T-  
170 cells<sup>40</sup>. AS can also repress gene expression through the introduction of premature  
171 stop codons into transcribed sequences, leading to non-sense mediated decay<sup>41,42</sup>.  
172 Although our understanding of the splicing code remains incomplete, there is a wide  
173 agreement that it involves the recognition by the spliceosome of donor/acceptor

174 sequences located at the intron boundaries, together with a branchpoint sequence  
175 — typically located ~30 nucleotide upstream of the acceptor splice site<sup>41,42</sup>. This  
176 leads to the formation of a lariat and the removal of the intron. AS is induced by  
177 competition between splice sites and is regulated through a wide array of RNA  
178 binding proteins able to recognize both exonic and intronic regulatory sequences<sup>42-</sup>  
179 <sup>45</sup>. These sequences can act as to either repress or promote, the recruitment of the  
180 spliceosome around a specific splice site or exon. In addition, alternative splicing can  
181 also be regulated co-transcriptionally, with exon inclusion being associated to a  
182 lower transcriptional elongation rate around the alternatively spliced exon<sup>41</sup>.

183

#### 184 ***Impact of miRNA-mediated degradation on immune response***

185 Another layer of post-transcriptional regulation occurs through the binding of  
186 miRNAs to 3'UTRs, and the targeting of bound mRNAs to the RNA induced silencing  
187 complex (RISC)<sup>46,47</sup>. The RISC then functions to repress gene expression through  
188 degradation of transcripts and inhibition of translation. For example, upon TLR4  
189 stimulation, the NF- $\kappa$ B-activated miRNAs mir-146a and miR-155 are strongly up-  
190 regulated and bind to 3'UTRs of the TLR4 signal transducers, TRAF6 and IRAK4  
191 and the TLR4 inhibitors SHIP1 and SOCS1<sup>48</sup>, respectively. These mRNAs are then  
192 targeted to the RISC for degradation, thereby contributing to the balance of

193 inflammatory responses to TLR4 stimulation. Several mechanisms can further  
194 modulate the impact of miRNAs on gene expression. For instance, usage of  
195 alternative UTRs and poly-adenylation sites can change the set of miRNA binding  
196 sites available for regulation of gene expression<sup>49</sup>. Such changes have been shown  
197 to be widespread upon stimulation<sup>49</sup> and to play an important role in allowing key  
198 regulators of immune response, such as IRF5 or Map Kinases MAPKAP1 and  
199 MAP2K4, to escape repression by immune miRNAs<sup>49</sup>. Formation of miRNA isoforms  
200 — known as isomiRs — could also contribute to immune response variability.  
201 Indeed, it was shown that post-transcriptional modifications of miRNAs through  
202 editing of the miRNA seed sequence or shifts in 5'/3' boundaries of the mature  
203 miRNA, can alter their target sites or stability and lead to a decrease in their ability to  
204 for regulate mRNAs<sup>46,50</sup>. Furthermore, extensive changes in isomiRs have been  
205 observed upon stimulation, including for key immune miRNAs<sup>51</sup>. For instance, upon  
206 stimulation by interferon, mir-222 expresses shorter isoforms, which are associated  
207 to lower apoptosis<sup>52,53</sup>, highlighting how changes of miRNA isoform can contribute to  
208 regulate the response to immune stimuli.

209

210 ***The genetic bases of gene expression***

211 One major challenge of functional genomics is the understanding of the regulatory  
212 code that dictates gene expression from the genetic information encoded by DNA  
213 sequences. One approach to learn this regulatory code is through the use of  
214 naturally-occurring genetic variation. Studies of expression quantitative trait loci  
215 (eQTL) aim to identify genetic variants that are associated to gene expression levels  
216 in human populations.

217

218 ***Genetic control of transcriptional activity is highly stimulation-specific***

219 eQTL studies have first focused of steady-state expression measurements on  
220 lymphoblastoid cell lines or whole blood, from a wide range of populations<sup>54-58</sup>.  
221 These studies are revealed that the majority of eQTLs are located in *cis*, i.e. less  
222 than 1Mb away of the gene they control, and that population-specific eQTLs are  
223 driven by changes in allelic frequency rather than gene × gene or gene ×  
224 environment interactions<sup>59</sup>. While a few *trans*-associations — involving genes  
225 located on different chromosomes — were detected, these associations were either  
226 weak or driven by cross mapping errors<sup>60</sup> and changes in cellular composition<sup>61</sup>. The  
227 strong potential of eQTL for biomedical applications<sup>62</sup> has rapidly lead to extend this  
228 framework to multiple tissues<sup>63</sup> or purified immune cell types<sup>29,64</sup>, revealing a strong  
229 tissue specificity of eQTLs. Furthermore, exposing human cells to immune and

230 infectious challenges has allowed the identification of response eQTLs that are  
231 associated with changes in gene expression specifically upon immune stimulation  
232 **(Figure 1f,g)**<sup>6,9,12,65,66</sup>. Interestingly, response-eQTLs were found to be highly  
233 variable across stimuli and time points<sup>8,9,12</sup>. For instance, the rs9266257 variant was  
234 found to control expression of HLA-C in macrophages specifically upon infection by  
235 Salmonella, but not Listeria<sup>65</sup>. Interestingly, this phenomenon appeared even more  
236 striking among trans-eQTLs which were often revealed by immune stimulation<sup>8</sup>. An  
237 outstanding example of this is the TLR1 locus, where the European-specific  
238 missense variant rs5743618 is associated with a decreased inflammatory response  
239 to stimulation of the TLR1/2 complex<sup>9</sup>. Interestingly, the European-specific allele of  
240 this variant is protective against leprosis<sup>67</sup> and overlaps one of the strongest  
241 signature of natural selection in the human genome<sup>9</sup>, highlighting how selective  
242 pressures induced by pathogens can lead to strong inter-individual and inter-  
243 population differences in immune phenotypes.

244

#### 245 ***The subtle effect of enhancer variants of gene expression***

246 Studies of the mechanistic bases of eQTLs<sup>9,57</sup> have shown that promoter variants  
247 are largely overrepresented among eQTLs. However, enhancer variants are  
248 increasing likely to be found when focusing on secondary, weaker eQTLs<sup>63</sup>. The

249 increasing power to detect small effect size and highly localized effects offered by  
250 eQTL meta-analyses<sup>61</sup> and analyses of eQTLs at single cell resolution<sup>68</sup> will likely  
251 improve our understanding of the contribution of enhancer variants to gene  
252 expression. Another approach to detect subtle effects of enhancers variants on gene  
253 expression consists in assaying directly the effect of genetic variants on regulatory  
254 activity, based on chromatin accessibility<sup>66,69</sup>, histone modifications<sup>29,69</sup>, or short-live  
255 transcriptional products<sup>70</sup>. In doing so, Li *et al.* estimated that QTLs for regulatory  
256 elements could account for ~60% of eQTLs, but only <25% of enhancers QTLs were  
257 associated to an effect on transcription nearest gene<sup>69</sup>, suggesting that only a small  
258 fraction of variants that alter regulatory activity lead to a measurable effect on gene  
259 expression. However, variants that alter chromatin accessibility with no visible effect  
260 on gene expression at basal state can contribute to prime enhancers<sup>66</sup>. Upon  
261 stimulation, this priming favours the binding of stimulus-specific transcription factors,  
262 such as NF- $\kappa$ B or Stat2, leading to a stronger immune response<sup>37,66</sup>. Together, these  
263 observations highlight the potential of eQTL meta-analyses and direct assays of  
264 regulatory activity to understand inter-individual variability in immune response.

265

266 ***Genetic regulation of post-transcriptional variability in the immune response***

267 Perhaps unsurprisingly given their strong potential to alter protein function,

268 splicing quantitative trait loci (sQTLs, **figure 2a**) have turned out to be consistently

269 less frequent than eQTLs across the human genome. Indeed, eQTLs appear to be

270 from 1.5 to 6 time more frequent than sQTLs, across studies<sup>56,57,58,65,71,72</sup> (**Table 2**).

271 In addition, sQTLs displayed a stronger overlap between conditions of stimulation

272 than eQTLs<sup>65,71</sup>, suggesting a limited impact of stimulation on the genetic regulation

273 of splicing. Nevertheless, disease-causing sQTL have been observed at many key

274 immune genes including *OAS1*, *IRF7*, *IL7R*, *IFI44L*, *TYK2* and *ERAP2*<sup>65,71,73-75</sup>,

275 highlighting the importance of considering splicing variation when searching for the

276 causal mechanisms underlying GWAS loci. Interestingly, several of these loci,

277 including *OAS1* or *ERAP2* present haplotypic signatures of positive and balancing

278 selection, suggesting that these sQTL have conferred a selective advantage to

279 human populations in the past, likely through increased resistance to pathogens.<sup>71,73</sup>

280 The regulation of miRNA expression has also been extensively studied, both at basal

281 state<sup>76,77</sup> and in response to infection<sup>78</sup>. Similar to splicing, miRNA QTLs studies

282 (**figure 2b**) have generally suggested a limited genetic control of miRNAs, consistent

283 with expectation of strong selective constraints on miRNA expression due to their

284 widespread impact on gene expression<sup>76,78</sup>. miRNA QTLs appear to be strongly  
285 enriched in variants located in their promoter and eQTLs of their host gene<sup>79</sup>.  
286 However, there is also strong evidence that enhancers and loci that are bound by  
287 transcription factor (TF) contribute to miRNA expression<sup>77</sup>. Recently, a study has  
288 used models of SNP-miRNA interactions to quantify the impact of SNPs on the  
289 regulation by miRNAs (reg-eQTL, **figure 2c**). This has led to the identification of  
290 ~40 SNP that disrupt miRNA binding sites<sup>80</sup>. The proposed approach, however,  
291 suffers from the lack of power usually associated with interaction testing, and may  
292 largely underestimate the true number of such reg-eQTL. Larger studies combining  
293 expression, genetic and miRNA data are now needed to fully characterize the impact  
294 of genetic variants on miRNA-mediated degradation in the immune response.

295

### 296 ***The contribution of regulatory variants to disease susceptibility***

297 In support of the importance of eQTLs in disease susceptibility, they have been  
298 repeatedly reported to be enriched in disease loci<sup>8,9,12,57,63</sup>. For instance, the  
299 association of the rs4077515 variant with Inflammatory bowel disease is mediated by  
300 activation of a distal enhancer of *CARD9* leading to an increase of its transcription,  
301 supporting the causal role of *CARD9* in Inflammatory bowel disease<sup>29</sup>. Similarly,  
302 there is increasing support to the notion that sQTL or miRNA QTLs can also



303 contribute to common disease risk<sup>69,71,81</sup>. For instance, the Multiple sclerosis  
304 associated variant rs6897932, located within an exon 6 of IL7R, has been shown to  
305 promote exon skipping. As a result, a soluble form of IL7R is being produced that is  
306 associated with exacerbated auto-immune phenotypes in mice<sup>75</sup>. However, despite  
307 evidence that 90% of auto-immune loci act through regulatory variants, 60% of which  
308 are located at enhancers<sup>82</sup>, it was suggested that only a limited fraction of the QTLs  
309 detected at GWAS loci play a causal role in disease susceptibility<sup>63,83</sup>. This  
310 observation, which may reflect the robustness of disease gene expression induced  
311 by the high redundancy of their enhancers<sup>84</sup>, highlights the importance of taking  
312 linkage disequilibrium into account when assessing the co-localization between  
313 GWAS loci and eQTLs, or sQTLs<sup>85</sup>.

314

### 315 **Deciphering the regulatory code to predict the effect of rare variants**

316 Recently, the development of massively parallel reporter assays (MPRA) has  
317 enabled the direct assessment of cis-regulatory activity associated to specific DNA  
318 fragments, either synthetic<sup>86</sup>, or naturally occurring<sup>87,88</sup>. The general principle of  
319 MPRA techniques (**figure 3**) consists in transfecting libraries of barcoded plasmids  
320 into live cells and counting the occurrence of each barcode in the transfected cells,  
321 as a measure of the activity of the associated regulatory sequence (after

322 normalization by the barcode frequency in the initial library). By measuring the  
323 activity of both alleles at heterozygous loci, these approaches further allow  
324 quantifying the impact of genetic variants, or haplotypes, from a single individual.  
325 Thus, they constitute a privileged tool for assessing the impact of rare variants on  
326 regulatory regions, or identifying causal variants among sets of linked SNPs  
327 identified through eQTL analyses<sup>88,89</sup>. For instance, dissection of regulatory  
328 mechanisms at a 87 nucleotide-long *IFNB1* enhancer indicated that 83 substitutions,  
329 out the 261 possible, resulted in altered activity in virus-infected cells<sup>88</sup>.

330 Massively parallel assays have also been used in the context of both splicing<sup>90,91</sup>  
331 and miRNA-mediated regulation<sup>92,93</sup>, revealing, for instance, that up to 16% of splice  
332 disrupting variants are located in deep intronic regions<sup>90</sup>. With the development of  
333 deep learning frameworks for sequence-based predictions<sup>94,95</sup>, data generated by  
334 these assays, will now fuel the construction of predictive models. These models will,  
335 in turn, allow quantifying the regulatory impact of novel genetic variants on  
336 transcriptional activity<sup>88</sup>, alternative splicing<sup>96</sup>, or miRNA-mediated regulation<sup>93</sup>.

337

### 338 **Concluding remarks: a strategy for the future.**

339 Elucidating the mechanisms by which non-coding variants contribute to the inter-  
340 individual variability of immune response is now an essential step in order to

341 translate the findings of epidemiological genetic studies into the clinics<sup>13</sup>. The  
342 characterization the human epigenome across a wide variety of tissues and  
343 conditions has allowed establishing a first map of the regulatory landscape of the  
344 human genome. Further efforts are now needed to fully define the cascade of  
345 regulatory events that take place during the immune response, for a wide spectrum  
346 of stimuli and with a high temporal resolution. This will allow classifying regulatory  
347 elements according to their activity in distinct cell types, and in response to specific  
348 stimuli. Then, using heritability partitioning techniques<sup>97,98</sup> that decompose the  
349 genetic variance of complex traits across several SNP categories, we will be able to  
350 identify the cell types and stimuli that most strongly contribute to the development of  
351 immune disorders. Holistic studies that consider both transcriptional and post-  
352 transcriptional aspects of gene expression variation<sup>41,66,74</sup>, and combine RNA seq  
353 with sequencing of small RNAs or ATAC-seq, constitute a promising approach to  
354 understand the aetiology of immune disorders. Such approaches, however, are only  
355 the first step toward the identification of the causal mechanisms that underlie disease  
356 susceptibility. Validation of the proposed mechanisms, through colocalization  
357 analysis<sup>85</sup>, mendelian randomization approaches<sup>99</sup> and randomized clinical trials<sup>100</sup>,  
358 are then required to formally establish causality. Finally, as whole genome  
359 sequencing is becoming increasingly common<sup>19</sup>, the ability to identify rare regulatory

360 variants at known disease genes will become a crucial step toward personalized  
361 treatments of both rare and common immune disorders. In this context, the  
362 increased knowledge of the regulatory code granted by functional genomics studies,  
363 and MPRA in particular, will constitute a precious tool to predict and assess the  
364 effect of rare variants on intermediate disease phenotypes.

365

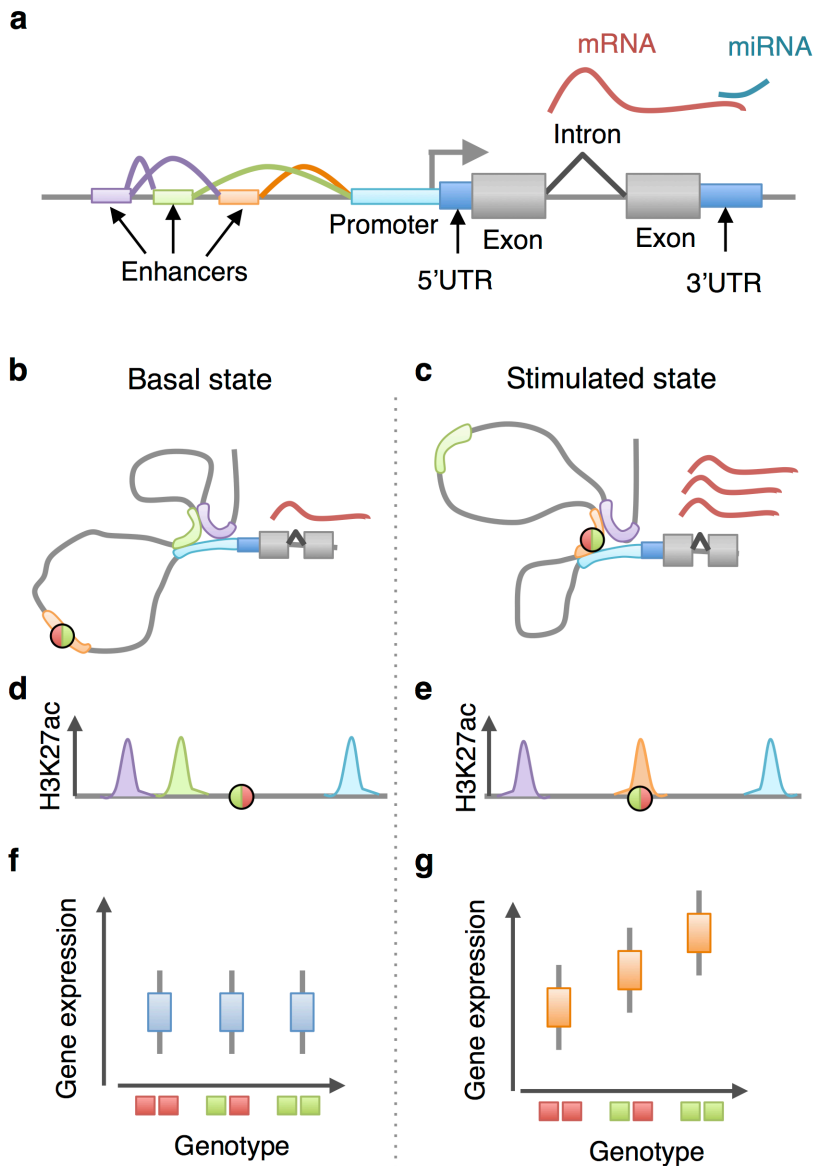
366

367 **Table 1. Epigenetic marks associated with gene regulation**

<b>Epigenetic mark</b>	<b>Associated function</b>	<b>Assay</b>
H3K4me1	Enhancers	Chip-Seq
H3K4me3	Promoters	Chip-Seq
H3K27Ac	Active regulatory regions (promoter/enhancer)	Chip-Seq
H3K9Ac	Active regulatory region (promoter/enhancer)	Chip-Seq
H3K36me3	transcribed regions /exon definition	Chip-Seq
H3K9me3	repressed region (heterochromatin)	Chip-Seq
H3K27me3	repressed region (polycomb)	Chip-Seq
Cytosine methylation	repressed regulatory regions (promoter/enhancer)	BisulfiteSeq
Open Chromatin	Active regulatory region (promoter/enhancer)	DNase-Seq/ ATAC-Seq

368 **Table 2. Compared number of eQTL and sQTL detected across studies**

<b>Study</b>	<b>Sample used</b>	<b>Number of eQTL detected</b>	<b>Number of sQTL detected</b>	<b>Metric used</b>
Pickrell, 2010 <sup>56</sup>	lymphoblastoid cell lines (n=69)	929	187	Exon
Lappalainen, 2013 <sup>57</sup>	lymphoblastoid cell lines (n= 462)	3,773	639	Transcript
Battle, 2014 <sup>58</sup>	Whole blood (n=922)	10,914	2,851	Transcript
Nedelec, 2016 <sup>65</sup>	macrophages (n=175 x 3 conditions)	1,647	1,120	Transcript
Chen, 2016 <sup>29</sup>	Monocytes, neutrophils, CD4+ T-cells (n=197)	6513, 5845, 5799	3044, 2034, 3347	Transcript, Exon
Ye, 2018 <sup>71</sup>	Dendritic cells (3 conditions, n= 99+227+250)	6,694	2,763	Transcript

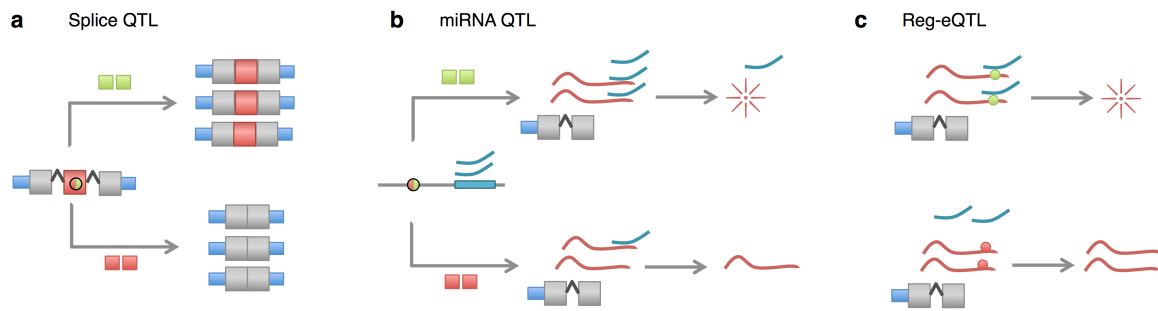


370

371 **Figure 1: The molecular bases of response-eQTLs.**

372 **a.** Genomic dissection of a genic region. Grey arrow represents the transcription  
 373 start site. Purple, green and orange lines represent possible chromatin-chromatin  
 374 interactions between enhancers, or between enhancers and the promoter.  
 375 Green/red circle represents a mutation of the orange enhancer. The same color

376 code is used throughout the figure. **b-c.** 3-dimensional representation of  
377 chromatin interactions leading to gene expression. At basal state (**b**), gene  
378 expression is mediated by the green enhancer, which comes in contact with the  
379 promoter (lightblue) and recruits the purple enhancer. Upon stimulation (**c**), the  
380 green enhancers is sequestered far from the promoter and replaced by the  
381 orange enhancer, allowing the mutation to alter gene expression. **d-e.** Activity of  
382 regulatory elements in the region, at basal state (**d**) and upon stimulation (**e**).  
383 Peaks indicate H3K27 acetylation with measures activity at promoters and  
384 enhancers. **f-g.** Boxplots of gene expression for each genotype, at basal state (**f**)  
385 and upon stimulation (**g**). Alleles at the locus are colour coded (red: low  
386 expression, green: high expression allele). The impact of the genetic variant  
387 located in the orange enhancer is visible only in the stimulated condition, where  
388 the enhancer is active.  
389



390

391

392 **Figure 2: Genetic regulation at the post-transcriptional level.**

393 **a.** sQTL alter gene function by changing their splicing, leading to a different protein

394 isoform. **b.** miRNA QTL that modify the abundance of miRNAs can act in *trans* in to

395 alter gene expression of genes located elsewhere on the genome through miRNA-

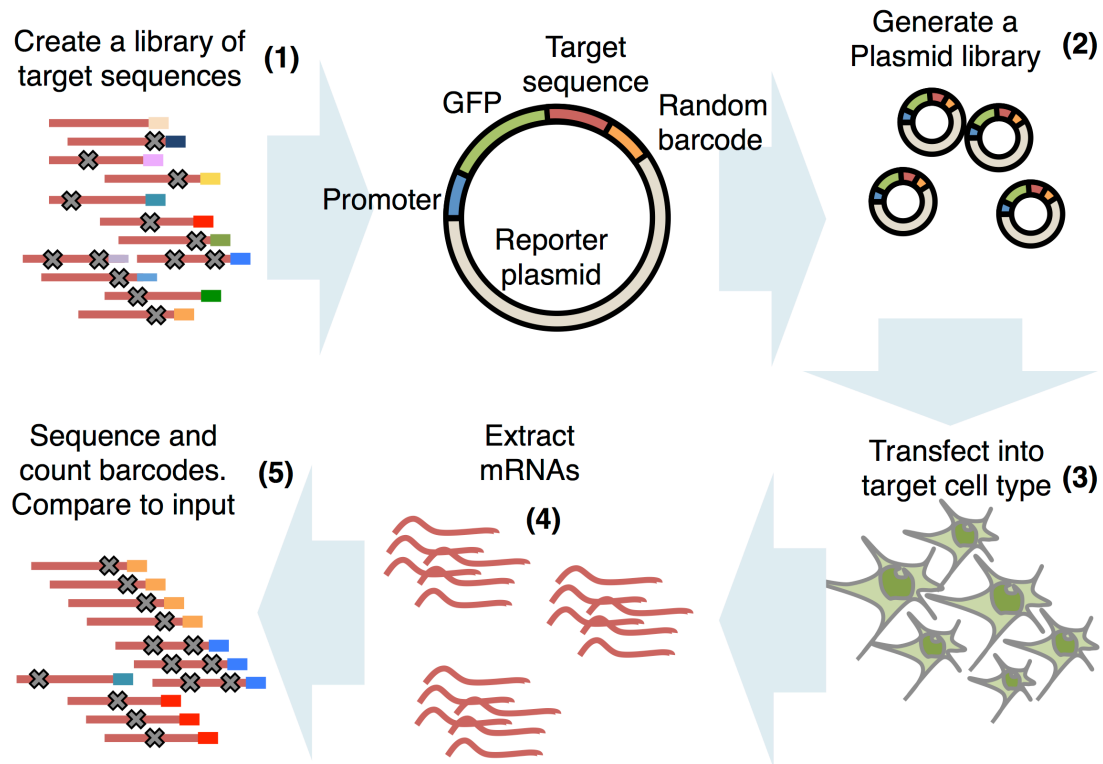
396 mediated degradation. **c.** SNPs that disturb (or create) miRNA binding sites in the

397 3'UTR of a gene can also change their level of expression, by acting in *cis* on

398 miRNA-mediated degradation. Such SNPs are referred as reg-eQTLs.

399





400

401 **Figure 3: Principle of massively parallel reporter assays (MPRA).**

402 MPRA consists in forming libraries of putative regulatory sequences, where each

403 sequence is associated to a specific, random barcode (1). The barcoded sequence

404 is then introduced in a reporter plasmid, where it is ligated to the 3' of a green

405 fluorescent protein (GFP) associated with a basic promoter sequence (2). The

406 plasmid library is then transfected into a cell line representative of the tissue of

407 interest (3). RNA from GFP positive cell are harvested and mRNAs of the reporter

408 gene are captured and sequenced (4). The regulatory activity of each sequence is

409 then measured by comparing the counts of its associated barcode in the transfected

410 cell library with that of the initial sequence library (5).

411

412 **Acknowledgements**

413 Research in the Human Evolutionary Genetics lab is funded by the Institut Pasteur,  
414 the Centre National de la Recherche Scientifique (CNRS), and the Fondation pour la  
415 Recherche Médicale (Equipe FRM DEQ20180339214).

416

417 **Biosketch.**

418

419 Dr Maxime Rotival obtained his PhD in statistical genetics in 2011 from University of  
420 Paris-Sud/XI. His PhD work regarded the genetic determinants of monocyte gene  
421 expression in humans and their contribution to metabolic and cardiovascular  
422 disorders. He did his first postdoc at Imperial College London, where he studied the  
423 genetics of macrophage dysfunction in a rat model of chronic kidney inflammatory  
424 disease, before joining the Human Evolutionary Genetics laboratory of Institut  
425 Pasteur (CNRS UMR2000) in 2014. There he studied how genetic variants that alter  
426 response to immune stimuli have contributed to the adaptation of modern human  
427 population to pathogens present in their environment. He now has a researcher  
428 position at institut Pasteur and is interested in the contribution of genetic diversity to  
429 the inter-individual variability in innate immune response and how selective  
430 pressures induced by pathogens have shaped this diversity. His work was awarded  
431 with the Julia Bodmer Award for Young Scientists from the European Federation for  
432 Immunogenetics in 2018.

433

434 **References**

- 435 1. Brodin P, Davis MM. Human immune system variation. *Nat Rev Immunol.*  
436 2017;17(1):21-29.
- 437 2. Piasecka B, Duffy D, Urrutia A, et al. Distinctive roles of age, sex, and  
438 genetics in shaping transcriptional variation of human immune responses to  
439 microbial challenges. *Proc Natl Acad Sci U S A.* 2018;115(3):E488-E497.

- 440 3. Alpert A, Pickman Y, Leipold M, et al. A clinically meaningful metric of immune  
441 age derived from high-dimensional longitudinal monitoring. *Nat Med*.  
442 2019;25(3):487-495.
- 443 4. Brodin P, Jojic V, Gao T, et al. Variation in the human immune system is  
444 largely driven by non-heritable influences. *Cell*. 2015;160(1-2):37-47.
- 445 5. Netea MG, Joosten LA, Latz E, et al. Trained immunity: A program of innate  
446 immune memory in health and disease. *Science*. 2016;352(6284):aaf1098.
- 447 6. Barreiro LB, Tailleux L, Pai AA, Gicquel B, Marioni JC, Gilad Y. Deciphering  
448 the genetic architecture of variation in the immune response to  
449 *Mycobacterium tuberculosis* infection. *Proc Natl Acad Sci U S A*.  
450 2012;109(4):1204-1209.
- 451 7. Lee MN, Ye C, Villani AC, et al. Common genetic variants modulate  
452 pathogen-sensing responses in human dendritic cells. *Science*.  
453 2014;343(6175):1246980.
- 454 8. Fairfax BP, Humburg P, Makino S, et al. Innate immune activity conditions the  
455 effect of regulatory variants upon monocyte gene expression. *Science*.  
456 2014;343(6175):1246949.
- 457 9. Quach H, Rotival M, Pothlichet J, et al. Genetic Adaptation and Neandertal  
458 Admixture Shaped the Immune System of Human Populations. *Cell*.  
459 2016;167(3):643-656 e617.
- 460 10. Patin E, Quintana-Murci L. The demographic and adaptive history of central  
461 African hunter-gatherers and farmers. *Curr Opin Genet Dev*. 2018;53:90-97.
- 462 11. Nédelec Y, Sanz J, Baharian G, et al. Genetic Ancestry and Natural Selection  
463 Drive Population Differences in Immune Responses to Pathogens. *Cell*.  
464 2016;167(3):657-669 e621.
- 465 12. Kim-Hellmuth S, Bechheim M, Putz B, et al. Genetic regulatory effects  
466 modified by immune activation contribute to autoimmune disease  
467 associations. *Nat Commun*. 2017;8(1):266.
- 468 13. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery:  
469 Biology, Function, and Translation. *Am J Hum Genet*. 2017;101(1):5-22.
- 470 14. Pal LR, Yu CH, Mount SM, Moulton J. Insights from GWAS: emerging landscape  
471 of mechanisms underlying complex trait disease. *BMC Genomics*. 2015;16  
472 Suppl 8:S4.
- 473 15. Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting  
474 enhancers. *Nature*. 2009;461(7261):199-205.
- 475 16. Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to  
476 Function. *Am J Hum Genet*. 2018;102(5):717-730.
- 477 17. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of  
478 complex diseases. *Nature*. 2009;461(7265):747-753.

- 479 18. Auer PL, Lettre G. Rare variant association studies: considerations,  
480 challenges and opportunities. *Genome Med.* 2015;7(1):16.
- 481 19. Wainschtein P, Jain DP, Yengo L, et al. Recovery of trait heritability from  
482 whole genome sequence data. *bioRxiv.* 2019:588020.
- 483 20. Hunt KA, Mistry V, Bockett NA, et al. Negligible impact of rare autoimmune-  
484 locus coding-region variants on missing heritability. *Nature.*  
485 2013;498(7453):232-235.
- 486 21. Kellis M, Wold B, Snyder MP, et al. Defining functional DNA elements in the  
487 human genome. *Proc Natl Acad Sci U S A.* 2014;111(17):6131-6138.
- 488 22. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al.  
489 Integrative analysis of 111 reference human epigenomes. *Nature.*  
490 2015;518(7539):317-330.
- 491 23. FANTOM Consortium, the RP, Clst, et al. A promoter-level mammalian  
492 expression atlas. *Nature.* 2014;507(7493):462-470.
- 493 24. Gondor A, Ohlsson R. Enhancer functions in three dimensions: beyond the  
494 flat world perspective. *F1000Res.* 2018;7.
- 495 25. Andersson R, Gebhard C, Miguel-Escalada I, et al. An atlas of active  
496 enhancers across human cell types and tissues. *Nature.* 2014;507(7493):455-  
497 461.
- 498 26. ENCODE consortium, An integrated encyclopedia of DNA elements in the  
499 human genome. *Nature.* 2012;489(7414):57-74.
- 500 27. Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. Analysis of  
501 nascent RNA identifies a unified architecture of initiation regions at  
502 mammalian promoters and enhancers. *Nat Genet.* 2014;46(12):1311.
- 503 28. Stunnenberg HG, International Human Epigenome C, Hirst M. The  
504 International Human Epigenome Consortium: A Blueprint for Scientific  
505 Collaboration and Discovery. *Cell.* 2016;167(5):1145-1149.
- 506 29. Chen L, Ge B, Casale FP, et al. Genetic Drivers of Epigenetic and  
507 Transcriptional Variation in Human Immune Cells. *Cell.* 2016;167(5):1398-  
508 1414 e1324.
- 509 30. Schmitt AD, Hu M, Jung I, et al. A Compendium of Chromatin Contact Maps  
510 Reveals Spatially Active Regions in the Human Genome. *Cell Rep.*  
511 2016;17(8):2042-2059.
- 512 31. Javierre BM, Burren OS, Wilder SP, et al. Lineage-Specific Genome  
513 Architecture Links Enhancers and Non-coding Disease Variants to Target  
514 Gene Promoters. *Cell.* 2016;167(5):1369-1384 e1319.
- 515 32. Lecellier CH, Wasserman WW, Mathelier A. Human Enhancers Harboring  
516 Specific Sequence Composition, Activity, and Genome Organization Are  
517 Linked to the Immune Response. *Genetics.* 2018;209(4):1055-1071.

- 518 33. Arner E, Daub CO, Vitting-Seerup K, et al. Transcribed enhancers lead waves  
519 of coordinated transcription in transitioning mammalian cells. *Science*.  
520 2015;347(6225):1010-1014.
- 521 34. Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: a  
522 comprehensive technique to capture the conformation of genomes. *Methods*.  
523 2012;58(3):268-276.
- 524 35. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional  
525 organization of genomes: interpreting chromatin interaction data. *Nat Rev*  
526 *Genet*. 2013;14(6):390-403.
- 527 36. Delaneau O, Zazhytska M, Borel C, et al. Intra- and inter-chromosomal  
528 chromatin interactions mediate genetic effects on regulatory networks.  
529 *bioRxiv*. 2017:171694.
- 530 37. Fanucchi S, Fok ET, Dalla E, et al. Immune genes are primed for robust  
531 transcription by proximal long noncoding RNAs located in nuclear  
532 compartments. *Nat Genet*. 2019;51(1):138-150.
- 533 38. Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell  
534 identity and disease. *Cell*. 2013;155(4):934-947.
- 535 39. Song W, Ovcharenko I. Dichotomy in redundant enhancers points to presence  
536 of initiators of gene regulation. *BMC Genomics*. 2018;19(1):947.
- 537 40. Lynch KW. Consequences of regulated pre-mRNA splicing in the immune  
538 system. *Nat Rev Immunol*. 2004;4(12):931-940.
- 539 41. Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, Munoz MJ. Alternative  
540 splicing: a pivotal step between eukaryotic transcription and translation. *Nat*  
541 *Rev Mol Cell Biol*. 2013;14(3):153-165.
- 542 42. Park E, Pan Z, Zhang Z, Lin L, Xing Y. The Expanding Landscape of  
543 Alternative Splicing Variation in Human Populations. *Am J Hum Genet*.  
544 2018;102(1):11-26.
- 545 43. Ray D, Kazan H, Cook KB, et al. A compendium of RNA-binding motifs for  
546 decoding gene regulation. *Nature*. 2013;499(7457):172-177.
- 547 44. Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. Using positional  
548 distribution to identify splicing elements and predict pre-mRNA processing  
549 defects in human genes. *Proc Natl Acad Sci U S A*. 2011;108(27):11093-  
550 11098.
- 551 45. Van Nostrand EL, Freese P, Pratt GA, et al. A Large-Scale Binding and  
552 Functional Map of Human RNA Binding Proteins. *bioRxiv*. 2018:179648.
- 553 46. Gebert LFR, MacRae IJ. Regulation of microRNA function in animals. *Nat Rev*  
554 *Mol Cell Biol*. 2019;20(1):21-37.
- 555 47. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*.  
556 2009;136(2):215-233.

- 557 48. Mehta A, Baltimore D. MicroRNAs as regulatory elements in immune system  
558 logic. *Nat Rev Immunol.* 2016;16(5):279-294.
- 559 49. Pai AA, Baharian G, Page Sabourin A, et al. Widespread Shortening of 3'  
560 Untranslated Regions and Increased Exon Inclusion Are Evolutionarily  
561 Conserved Features of Innate Immune Responses to Infection. *PLoS Genet.*  
562 2016;12(9):e1006338.
- 563 50. Neilsen CT, Goodall GJ, Bracken CP. IsomiRs--the overlooked repertoire in  
564 the dynamic microRNAome. *Trends Genet.* 2012;28(11):544-549.
- 565 51. Siddle KJ, Tailleux L, Deschamps M, et al. bacterial infection drives the  
566 expression dynamics of microRNAs and their isomiRs. *PLoS Genet.*  
567 2015;11(3):e1005064.
- 568 52. Nejad C, Pillman KA, Siddle KJ, et al. miR-222 isoforms are differentially  
569 regulated by type-I interferon. *RNA.* 2018;24(3):332-341.
- 570 53. Yu F, Pillman KA, Neilsen CT, et al. Naturally existing isoforms of miR-222  
571 have distinct functions. *Nucleic Acids Res.* 2017;45(19):11371-11385.
- 572 54. Stranger BE, Montgomery SB, Dimas AS, et al. Patterns of cis regulatory  
573 variation in diverse human populations. *PLoS Genet.* 2012;8(4):e1002639.
- 574 55. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, et al. Transcriptome  
575 genetics using second generation sequencing in a Caucasian population.  
576 *Nature.* 2010;464(7289):773-777.
- 577 56. Pickrell JK, Marioni JC, Pai AA, et al. Understanding mechanisms underlying  
578 human gene expression variation with RNA sequencing. *Nature.*  
579 2010;464(7289):768-772.
- 580 57. Lappalainen T, Sammeth M, Friedlander MR, et al. Transcriptome and  
581 genome sequencing uncovers functional variation in humans. *Nature.*  
582 2013;501(7468):506-511.
- 583 58. Battle A, Mostafavi S, Zhu X, et al. Characterizing the genetic basis of  
584 transcriptome diversity through RNA-sequencing of 922 individuals. *Genome*  
585 *Res.* 2014;24(1):14-24.
- 586 59. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG.  
587 Common genetic variants account for differences in gene expression among  
588 ethnic groups. *Nat Genet.* 2007;39(2):226-231.
- 589 60. Saha A, Battle A. False positives in trans-eQTL and co-expression analyses  
590 arising from RNA-sequencing alignment errors. *F1000Res.* 2018;7:1860.
- 591 61. Vösa U, Claringbould A, Westra H-J, et al. Unraveling the polygenic  
592 architecture of complex traits using blood eQTL meta-analysis. *bioRxiv.*  
593 2018:447367.
- 594 62. Montgomery SB, Dermitzakis ET. From expression QTLs to personalized  
595 transcriptomics. *Nat Rev Genet.* 2011;12(4):277-282.

- 596 63. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, et  
597 al. Genetic effects on gene expression across human tissues. *Nature*.  
598 2017;550(7675):204-213.
- 599 64. Schmiedel BJ, Singh D, Madrigal A, et al. Impact of Genetic Polymorphisms  
600 on Human Immune Cell Gene Expression. *Cell*. 2018;175(6):1701-1715  
601 e1716.
- 602 65. Nedelec Y, Sanz J, Baharian G, et al. Genetic Ancestry and Natural Selection  
603 Drive Population Differences in Immune Responses to Pathogens. *Cell*.  
604 2016;167(3):657-669 e621.
- 605 66. Alasoo K, Rodrigues J, Mukhopadhyay S, et al. Shared genetic effects on  
606 chromatin and gene expression indicate a role for enhancer priming in  
607 immune response. *Nat Genet*. 2018;50(3):424-431.
- 608 67. Johnson CM, Lyle EA, Omuetti KO, et al. Cutting edge: A common  
609 polymorphism impairs cell surface trafficking and functional responses of  
610 TLR1 but protects against leprosy. *J Immunol*. 2007;178(12):7520-7524.
- 611 68. van der Wijst MGP, Brugge H, de Vries DH, et al. Single-cell RNA sequencing  
612 identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat Genet*.  
613 2018;50(4):493-497.
- 614 69. Li YI, van de Geijn B, Raj A, et al. RNA splicing is a primary link between  
615 genetic variation and disease. *Science*. 2016;352(6285):600-604.
- 616 70. Garieri M, Delaneau O, Santoni F, et al. The effect of genetic variation on  
617 promoter usage and enhancer activity. *Nat Commun*. 2017;8(1):1358.
- 618 71. Ye CJ, Chen J, Villani AC, et al. Genetic analysis of isoform usage in the  
619 human anti-viral response reveals influenza-specific regulation of ERAP2  
620 transcripts under balancing selection. *Genome Res*. 2018;28(12):1812-1825.
- 621 72. Alasoo K, Rodrigues J, Danesh J, Freitag DF, Paul DS, Gaffney DJ. Genetic  
622 effects on promoter usage are highly context-specific and contribute to  
623 complex traits. *Elife*. 2019;8.
- 624 73. Sams AJ, Dumaine A, Nedelec Y, et al. Adaptively introgressed Neandertal  
625 haplotype at the OAS locus functionally impacts innate immune responses in  
626 humans. *Genome Biol*. 2016;17(1):246.
- 627 74. Odhams CA, Cunninghame Graham DS, Vyse TJ. Profiling RNA-Seq at  
628 multiple resolutions markedly increases the number of causal eQTLs in  
629 autoimmune disease. *PLoS Genet*. 2017;13(10):e1007071.
- 630 75. Galarza-Munoz G, Briggs FBS, Evsyukova I, et al. Human Epistatic  
631 Interaction Controls IL7R Splicing and Increases Multiple Sclerosis Risk. *Cell*.  
632 2017;169(1):72-84 e13.

- 633 76. Lappalainen T, Sammeth M, Friedlander MR, et al. Transcriptome and  
634 genome sequencing uncovers functional variation in humans. *Nature*.  
635 2013;501(7468):506-511.
- 636 77. Huan T, Rong J, Liu C, et al. Genome-wide identification of microRNA  
637 expression quantitative trait loci. *Nat Commun*. 2015;6:6601.
- 638 78. Siddle KJ, Deschamps M, Tailleux L, et al. A genomic portrait of the genetic  
639 architecture and regulatory impact of microRNA expression in response to  
640 infection. *Genome Res*. 2014;24(5):850-859.
- 641 79. Budach S, Heinig M, Marsico A. Principles of microRNA Regulation Revealed  
642 Through Modeling microRNA Expression Quantitative Trait Loci. *Genetics*.  
643 2016;203(4):1629-1640.
- 644 80. Wilk G, Braun R. regQTLs: Single nucleotide polymorphisms that modulate  
645 microRNA regulation of gene expression in tumors. *PLoS Genet*.  
646 2018;14(12):e1007837.
- 647 81. Lord J, Gallone G, Short PJ, et al. Pathogenicity and selective constraint on  
648 variation near splice sites. *Genome Res*. 2018.
- 649 82. Farh KK, Marson A, Zhu J, et al. Genetic and epigenetic fine mapping of  
650 causal autoimmune disease variants. *Nature*. 2015;518(7539):337-343.
- 651 83. Chun S, Casparino A, Patsopoulos NA, et al. Limited statistical evidence for  
652 shared genetic effects of eQTLs and autoimmune-disease-associated loci in  
653 three major immune-cell types. *Nat Genet*. 2017;49(4):600-605.
- 654 84. Wang X, Goldstein DB. Enhancer redundancy predicts gene pathogenicity  
655 and informs complex disease gene discovery. *bioRxiv*. 2018:459123.
- 656 85. Hormozdiari F, van de Bunt M, Segre AV, et al. Colocalization of GWAS and  
657 eQTL Signals Detects Target Genes. *Am J Hum Genet*. 2016;99(6):1245-  
658 1260.
- 659 86. Smith RP, Taher L, Patwardhan RP, et al. Massively parallel decoding of  
660 mammalian regulatory sequences supports a flexible organizational model.  
661 *Nat Genet*. 2013;45(9):1021-1028.
- 662 87. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-wide  
663 quantitative enhancer activity maps identified by STARR-seq. *Science*.  
664 2013;339(6123):1074-1077.
- 665 88. Melnikov A, Murugan A, Zhang X, et al. Systematic dissection and  
666 optimization of inducible enhancers in human cells using a massively parallel  
667 reporter assay. *Nature Biotechnology*. 2012;30:271.
- 668 89. Tewhey R, Kotliar D, Park DS, et al. Direct Identification of Hundreds of  
669 Expression-Modulating Variants using a Multiplexed Reporter Assay. *Cell*.  
670 2016;165(6):1519-1529.



- 671 90. Cheung R, Insigne KD, Yao D, et al. A Multiplexed Assay for Exon  
672 Recognition Reveals that an Unappreciated Fraction of Rare Genetic Variants  
673 Cause Large-Effect Splicing Disruptions. *Mol Cell*. 2019;73(1):183-194 e188.
- 674 91. Soemedi R, Cygan KJ, Rhine CL, et al. Pathogenic variants that alter protein  
675 code often disrupt splicing. *Nat Genet*. 2017;49(6):848-855.
- 676 92. Rabani M, Pieper L, Chew GL, Schier AF. A Massively Parallel Reporter  
677 Assay of 3' UTR Sequences Identifies In Vivo Rules for mRNA Degradation.  
678 *Mol Cell*. 2017;68(6):1083-1094 e1085.
- 679 93. Vainberg Slutskin I, Weingarten-Gabbay S, Nir R, Weinberger A, Segal E.  
680 Unraveling the determinants of microRNA mediated regulation using a  
681 massively parallel reporter assay. *Nat Commun*. 2018;9(1):529.
- 682 94. Movva R, Greenside P, Shrikumar A, Kundaje A. Deciphering regulatory DNA  
683 sequences and noncoding genetic variants using neural network models of  
684 massively parallel reporter assays. *bioRxiv*. 2018:393926.
- 685 95. Zhou J, Theesfeld CL, Yao K, Chen KM, Wong AK, Troyanskaya OG. Deep  
686 learning sequence-based ab initio prediction of variant effects on expression  
687 and disease risk. *Nat Genet*. 2018;50(8):1171-1179.
- 688 96. Xiong HY, Alipanahi B, Lee LJ, et al. RNA splicing. The human splicing code  
689 reveals new insights into the genetic determinants of disease. *Science*.  
690 2015;347(6218):1254806.
- 691 97. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by  
692 functional annotation using genome-wide association summary statistics. *Nat*  
693 *Genet*. 2015;47(11):1228-1235.
- 694 98. Finucane HK, Reshef YA, Anttila V, et al. Heritability enrichment of specifically  
695 expressed genes identifies disease-relevant tissues and cell types. *Nat*  
696 *Genet*. 2018;50(4):621-629.
- 697 99. Porcu E, Rüeger S, Lepik K, Santoni FA, Reymond A, Kutalik Z. Mendelian  
698 Randomization integrating GWAS and eQTL data reveals genetic  
699 determinants of complex and clinical traits. *bioRxiv*. 2019:377267.
- 700 100. Nelson MR, Tipney H, Painter JL, et al. The support of human genetic  
701 evidence for approved drug indications. *Nat Genet*. 2015;47(8):856-860.
- 702