



OPEN

## Multiple sclerosis genetic and non-genetic factors interact through the transient transcriptome

Renato Umeton<sup>1,2,3,4,14</sup>✉, Gianmarco Bellucci<sup>5,14</sup>, Rachele Bigi<sup>5,6</sup>, Silvia Romano<sup>5</sup>, Maria Chiara Buscarinu<sup>5,6</sup>, Roberta Reniè<sup>5</sup>, Virginia Rinaldi<sup>5</sup>, Raffaella Pizzolato Umeton<sup>7,8,9,10</sup>, Emanuele Morena<sup>5</sup>, Carmela Romano<sup>5</sup>, Rosella Mechelli<sup>11,12</sup>, Marco Salvetti<sup>5,13</sup>✉ & Giovanni Ristori<sup>5,6</sup>✉

A clinically actionable understanding of multiple sclerosis (MS) etiology goes through GWAS interpretation, prompting research on new gene regulatory models. Our previous investigations suggested heterogeneity in etiology components and stochasticity in the interaction between genetic and non-genetic factors. To find a unifying model for this evidence, we focused on the recently mapped transient transcriptome (TT), that is mostly coded by intergenic and intronic regions, with half-life of minutes. Through a colocalization analysis, here we demonstrate that genomic regions coding for the TT are significantly enriched for MS-associated GWAS variants and DNA binding sites for molecular transducers mediating putative, non-genetic, determinants of MS (vitamin D deficiency, Epstein Barr virus latent infection, B cell dysfunction), indicating TT-coding regions as MS etiopathogenetic hotspots. Future research comparing cell-specific transient and stable transcriptomes may clarify the interplay between genetic variability and non-genetic factors causing MS. To this purpose, our colocalization analysis provides a freely available data resource at [www.mscoloc.com](http://www.mscoloc.com).

A large body of literature agrees that regulatory genomic intervals, especially those encompassing enhancers, are enriched with disease-associated DNA elements. Most of this evidence comes from genome wide association studies (GWAS) based on single polymorphism nucleotides (SNPs) representing common variants<sup>1–5</sup>, even though a recent study showed that low-frequency and rare coding variants may somewhat contribute to multifactorial diseases<sup>6</sup>. Several characteristics of regulatory disease-associated genetic variants complicate GWAS interpretation, prompting research on new gene regulatory models: (i) SNPs are chosen as haplotypes to spare the genotyping work needed for the large number of samples used in GWAS, therefore fine mapping and epigenetic studies are required to integrate GWAS data<sup>7–10</sup>; (ii) a fraction of supposedly causal disease-associated variants directly alters recognizable transcription factor binding motifs as it might be expected, according to their regulatory function<sup>4</sup>; (iii) the identified GWAS signals are likely to exert highly contextual (i.e., time- and position-dependent) regulatory effects, that may change according to the tissue and to the time when they receive an input from inside or outside the cell. In summary, current gene regulatory models help only in part to fully detail which disease-associated SNP signals are causal, and by which exact mechanisms they are causal. Recent studies on the

<sup>1</sup>Department of Informatics and Analytics, Dana-Farber Cancer Institute, Boston, MA, USA. <sup>2</sup>Department of Biological Engineering, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>3</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. <sup>4</sup>Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY, USA. <sup>5</sup>Department of Neurosciences, Mental Health and Sensory Organs, Centre for Experimental Neurological Therapies (CENTERS), Sapienza University of Rome, Rome, Italy. <sup>6</sup>Neuroimmunology Unit, IRCCS Fondazione Santa Lucia, Rome, Italy. <sup>7</sup>Department of Neurology, UMass Memorial Health Care, Worcester, MA, USA. <sup>8</sup>University of Massachusetts Medical School, Worcester, MA, USA. <sup>9</sup>Department of Neurology, Massachusetts General Hospital, Boston, MA, USA. <sup>10</sup>Harvard Medical School, Boston, MA, USA. <sup>11</sup>IRCCS San Raffaele Pisana, Rome, Italy. <sup>12</sup>San Raffaele Roma Open University, Rome, Italy. <sup>13</sup>IRCCS Istituto Neurologico Mediterraneo Neuromed, Pozzilli, Italy. <sup>14</sup>These authors contributed equally: Renato Umeton and Gianmarco Bellucci. ✉email: [umeton@mit.edu](mailto:umeton@mit.edu); [marco.salvetti@uniroma1.it](mailto:marco.salvetti@uniroma1.it); [giovanni.ristori@uniroma1.it](mailto:giovanni.ristori@uniroma1.it)

biological spectrum of human DNase I hypersensitive sites (DHSs), that are disease-associated markers of regulatory DNA, may help to better rework GWAS data and particularly to contextualize the genomic variants according to tissue/cell states and to gene body colocalization of DHSs<sup>11</sup>. In this context, the latest version of the Genotype-Tissue Expression project may provide further insights into the tissue specificity of genetic effects, supporting the link between regulatory mechanisms and traits or complex diseases<sup>12</sup>.

Such layers of complexity are in agreement with our previous investigations on the interplay between genetic and non-genetic factors contributing to multiple sclerosis (MS) development: twin pairs studies<sup>13,14</sup>; disease modelling suggested stochastic phenomena (i.e., random events not necessarily resulting in disease in all individuals) contributing to the disease onset and progression<sup>15,16</sup>; bioinformatics analyses determined a significant enrichment of binding motifs for Epstein-Barr virus (EBV) nuclear antigen 2 (EBNA2) and vitamin D receptor (VDR) in genomic regions containing MS-associated GWAS variants<sup>17</sup>. We also demonstrated that genomic variants of *EBNA2* resulted to be MS-associated<sup>18</sup>, and other groups expanded our findings showing that enrichment of EBNA2-binding regions on GWAS DNA intervals is involved in the pathogenesis of autoimmune disorders, including MS<sup>19</sup>. The role of EBV proteins and/or VDR as key transcriptional regulators in MS falls within well-known sero-epidemiological evidences on the virus as risk factor for disease development<sup>20</sup>, and on the vitamin deficiency associated to different disease prevalence in diverse geographic areas<sup>21</sup>. Recent works have even reinforced the EBV causal role and its mechanistic link for MS development<sup>22,23</sup>.

A recent sequencing innovation (namely, TT-seq) allowed to map the transient transcriptome that has a typical half-life within minutes, compared to stable RNA elements, such as protein-coding mRNAs, long-noncoding RNAs, and micro-RNAs, that persists at least a few hours<sup>24–26</sup>. The transient transcriptome (TT) includes mostly enhancer RNAs (eRNA), short intergenic non-coding RNAs (sincRNA) and antisense RNAs (asRNA). eRNAs are bidirectionally transcribed by mammalian genome regions having specific histone modifications to finely regulate chromatin conformation and transcription. Unlike promoters, enhancers can execute their functions regardless of orientation, position and spatial segregation from target genes that may be affected both in cis and in trans by eRNAs. Most sincRNAs are defined as genomic regions located within 10 kbp of a GENOME mRNA transcription start site. Overall, these transient RNAs (trRNA) are relatively short in length, generally lack a secondary structure, and would not present those chemical modifications that characterize unidirectional and polyadenylated stable RNAs<sup>24,27</sup>. Other recent works based on time-resolved analysis, agree on the eRNAs very rapid functional dynamics model while interacting with the transcriptional co-activator acetyltransferase CBP/p300 complex<sup>28,29</sup>. This confirms the highly contextual role of eRNAs through the control of transcription burst frequencies, which are known to influence cell-type-specific gene expression profiles<sup>30</sup>. Along these lines, a recent study showed that T cells selectively filter oscillatory signals within the minute timescale<sup>31</sup>, further supporting the aforementioned model.

On these bases, we leveraged the recent sequencing innovations in the mapping of the transient transcriptome, in particular the work by Michel et al. on T lymphocytes (that are known to play a major role in MS pathogenesis), that applied both the TT-seq and the RNA-seq protocol in Jukart cells during their immediate response to the stimulation with ionomycin and phorbol 12-myristate 13-acetate (PMA). Michel's study allowed to compare the trRNAs and mRNAs with high temporal resolution, showing that TT-seq, but not RNAseq, caught rapid changes in transcriptional activity just after 15 min after stimulation<sup>24,25</sup>. We hypothesized that MS-associated GWAS signals prevalently fall within regulatory regions of DNA coding for trRNAs. In theory, the genomic intervals coding for this transient transcriptome may be the hotspots where temporospatial occurrences may coalesce and so contribute to physiological (developmental and/or adaptive) outcomes, or possibly give rise to disease onset or progression. This study is aimed at verifying this working hypothesis through a colocalization analysis and its further dissection in the context of MS.

## Results

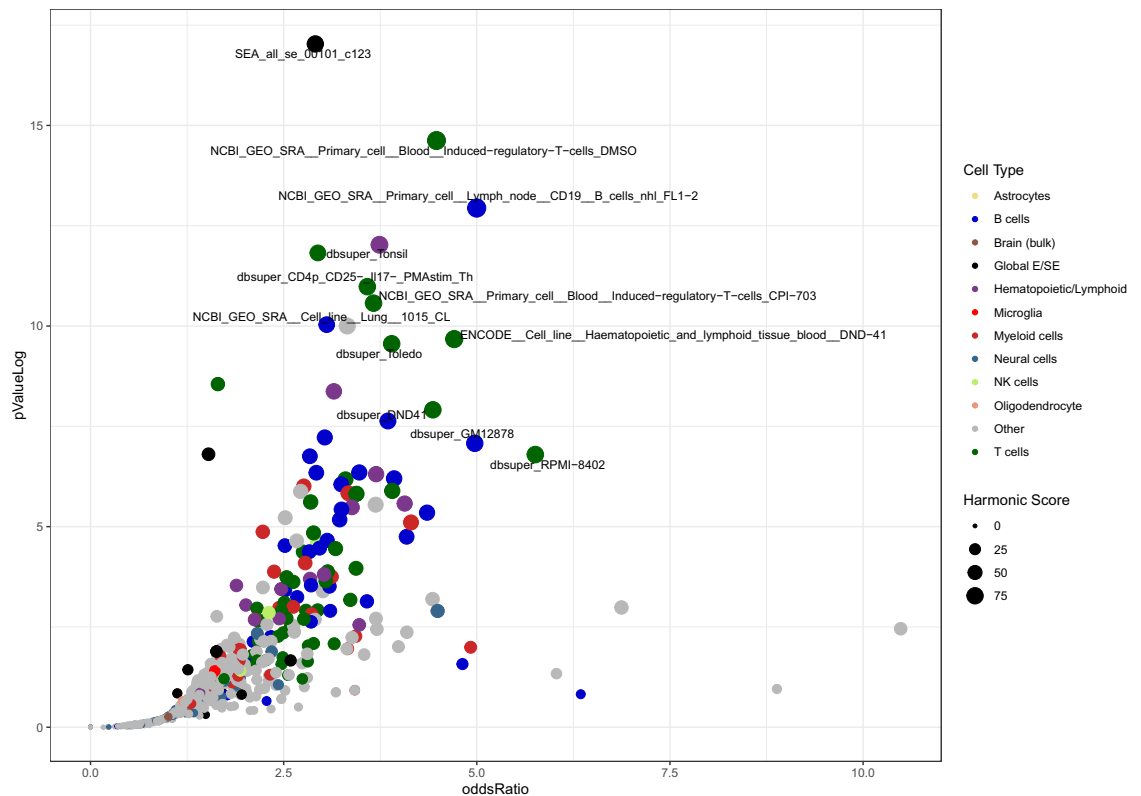
### MS-associated GWAS signals colocalize with regulatory regions of DNA plausibly coding for trRNAs.

We set up our region-of-interest (ROI) inside GWAS catalogue<sup>32</sup> by considering all MS GWAS that were published, extracting all SNP positions, and creating a single set of genomic coordinates that therefore encompass all GWAS-derived or GWAS-verified signal for MS. We then refined the SNP list by pruning out about 1.5% of the SNPs as they did not contain intelligible genomic annotations or were duplicates. The final ROI list is reported in Additional File: Table S1 and consists of 603 unique single-nucleotide regions; to provide a “threshold” against which the match ROI < > Database would be benchmarked, we used 107,423 regions as Universe, that corresponded to the signals coming from the entire GWAS Catalog.

Next, we matched through colocalization analyses our ROI with lists of regions resulting from the work by Michel et al., which mapped the transient and stable transcriptome captured by TT-seq after T cell stimulation<sup>24</sup>. We found a significant enrichment of MS-associated genetic variants in the transient transcriptome ( $p$ -value =  $2.80 \times 10^{-9}$ ; Table 1). Of note, when we split the transcriptome list in two subsets for long ( $\geq 60$  min) and short ( $< 60$  min) half-life, we found that only the short half-life subset significantly colocalized with the ROI ( $p$ -value  $2.06 \times 10^{-8}$  vs. 0.09). This finding was indicative of the relationship between MS-associated GWAS signals and the regulatory regions of DNA coding for trRNAs.

List	$-\log(p\text{-value})$	$p\text{-value}$	Odds ratio	Support	List size
Whole transient transcriptome	8.55	<b><math>2.80 \times 10^{-9}</math></b>	1.65	241	22,126
Short half-life transcripts	7.68	<b><math>2.06 \times 10^{-8}</math></b>	1.63	209	20,143
Long half-life transcripts	1.05	0.09	1.29	35	1993

**Table 1.** Enrichment of MS-associated genetic variants in lists of T-cell transient transcripts extracted from Michel et al., 2017. The whole transcriptome list was split in two sub-lists depending on the transcripts' half-life: short ( $< 60'$ ) and long ( $\geq 60'$ ), respectively. Results are considered significant at  $p < 0.05$  and are highlighted in bold.



**Figure 1.** Enrichment of MS-associated SNPs in databases of regulatory elements, sorted by experiment/cell lines. X-axis shows the Odds Ratio, y-axis shows the  $-\log(p\text{Value})$ ; dot size is proportional to the Harmonic Score (HS), a comprehensive estimation of the relevance of hits, as derived by merging and balancing the OR,  $p\text{Value}$  and Support (i.e., the number of hits resulting from the colocalization analysis) of each match. Thus, prioritized hits are represented by dots that occupy the upper-right area of the chart. Dots are coloured by cell type. Labeled points have  $HS > 50$ . Labels were arbitrarily designated according to the database of origin and the cell lineage where the enrichment occurred. Supplementary Table S4 displays in detail the results that generated this plot.

When we further dissected the mapping of the ROI colocalization signals, we found a significant excess of intergenic and intron regions (as anticipated), as well as their prevalent distribution away from the transcription start site (TSS; Figure Supplement 1A). Notably, when we extended this analysis to GWAS data coming from other multifactorial diseases or traits, dividing immune-mediated and other complex conditions, we found highly comparable profiles (supplementary Fig. S1B–C, Additional File: Table S2), suggesting that the colocalization between MS-associated DNA intervals and intergenic or intronic sequences, plausibly referring to trRNA coding regions, is shared by the genetic architecture of most multifactorial disorders.

To consolidate this result and gain a deeper biological insight, we extended the colocalization analysis matching the ROI with a vast set of databases of regulatory DNA regions, including enhancers and super-enhancers, derived from experiments on diverse tissue types (a total of 4,697,782 DNA regions, plausibly coding for trRNA, were extracted from a wide variety of raw data sources; referenced in Additional File: Table S3). To improve interpretability of the results through ranking, we implemented a harmonic score

	± 50 KB				± 100 KB				± 200 KB			
	–log (pValue)	Odds ratio	Support	Harmonic score	–log (pValue)	Odds ratio	Support	Harmonic score	–log (pValue)	Odds ratio	Support	Harmonic score
EBNA2 (6880)	<b>10.658</b>	<b>1.790</b>	<b>158</b>	<b>45.544</b>	<b>8.616</b>	<b>1.509</b>	<b>239</b>	<b>38.327</b>	<b>15.444</b>	<b>1.542</b>	<b>421</b>	<b>41.913</b>
EBNA3C (3335)	0.614	1.108	55	11.765	<b>1.647</b>	<b>1.227</b>	<b>109</b>	<b>20.956</b>	<b>3.448</b>	<b>1.294</b>	<b>199</b>	<b>28.098</b>
AID (4823)	<b>4.963</b>	<b>1.596</b>	<b>99</b>	<b>35.793</b>	<b>3.890</b>	<b>1.374</b>	<b>153</b>	<b>30.259</b>	<b>13.924</b>	<b>1.619</b>	<b>309</b>	<b>43.308</b>
VDR (23,409)	<b>19.348</b>	<b>1.575</b>	<b>474</b>	<b>43.564</b>	<b>19.181</b>	<b>1.422</b>	<b>767</b>	<b>39.635</b>	<b>32.090</b>	<b>1.424</b>	<b>1329</b>	<b>40.872</b>

**Table 2.** Enrichment of MS-GWAS regions (at ± 50,100,200 kb range of extension) in lists (number in brackets in the right-most column) of DNA binding sites of human and viral molecular transducers; significant results ( $p < 0.05$ , corresponding to a  $-\log(p) > 1.301$ ) in bold.

(HS), based on the Odd Ratio, the  $-\log(p\text{-value})$ , and the support of each match. Statistically significant results came from sets included in SEA, seDB, dbSuper and other single lists of enhancers and non-coding RNAs derived from experiments on diverse tissue types, not necessarily belonging to the immune cells lineages (Fig. 1, black dots; Additional File: Table S4). On another hand, we found a strong enrichment of MS-associated genetic variants in cell lines of hematopoietic lineage, including CD19+ and CD20+ B lymphocytes, CD4+ T helper cells, and CD14+ monocytes. This is in line with the GWAS data and the known immunopathogenesis of the disease, as well as with the fact that we consider a TT collection coming from a lymphoid line for the co-localization analysis. Moreover, among the significant hits, we found collections coming from brain resident cell, in particular microglial-specific enhancers, which is in line with recent reports on brain cell type-specific enhancer-promoter interactome activities, and the latest GWAS on MS genomic mapping<sup>33,34</sup>. Non-relevant tissues serving as controls (such as kidney, muscle, glands, etc.) scored low in the ranking, crowding the bottom-left corner of Fig. 1 (grey dots; see also Additional File: Table S4).

**Genetic and non-genetic factors for MS etiology converge in genomic regions plausibly coding for the transient transcriptome.** Independent studies support the fact that MS GWAS intervals are enriched with DNA binding regions (DBRs) for protein ‘transducers’ mediating non-genetic factors of putative etiologic relevance in MS, such as vitamin D deficiency or EBV latent infection<sup>17,19</sup>. Therefore, we further inquired whether DNA regions plausibly coding for trRNA would share these features (i.e., they colocalize with such DRBs). We set up 4 new ROIs corresponding to the DBRs for VDR, activation-induced cytidine deaminase (AID), EBNA2, and Epstein Barr nuclear antigen 3 (EBNA3C), chosen among viral or host’s nuclear factors potentially associated to MS etiopathogenesis<sup>35–37</sup>. The DBRs for each nuclear factor were derived from recent literature (Additional File: Table S5) and matched with the GWAS-derived MS signals to confirm and expand previous results. We found statistically significant results for VDR, EBNA2, and AID for all the SNP position extensions (± 50, 100, 200 kb up- and down-stream), while for EBNA3C significant results came out at extension of ± 100 and 200 kilobases. This finding suggests that several DBRs can impact on the MS-associated DNA intervals through colocalization (Table 2).

Building once again on the work by Michel et al.<sup>25</sup>, we inquired whether there was a colocalization between genomic regions containing MS-associated variants, DBRs for VDR, EBNA2, EBNA3C, AID, and DNA intervals plausibly coding for trRNA. To this end, we considered the transient transcriptome that proved to be enriched with MS-associated variants (Table 1), and we then matched the corresponding coding regions with the DBRs for the four molecular transducers. For this analysis DBRs for EBNA2 (6880 regions), EBNA3C (3835 regions), AID (4823 regions), and VDR (23,409 regions), represented the ROI, while the ENCODE database of Transcription Factors Binding Sites served as Universe (13,202,334 regions; Fig. 2a). We report the results of this analysis in Table 3, which shows the significant colocalization of DNA regions plausibly coding for trRNAs with both MS-relevant GWAS signals, and DBR of 3 out of 4 factors active at nuclear level, and potentially associated with MS. The DBR for EBNA3C did not reach statistical significance, though it showed higher values of support for short half-life transcripts.

To review and confirm previous colocalizations, we considered the genomic regions resulting from the above reported match between the MS-associated GWAS intervals and the databases of regulatory DNA regions, containing enhancers and super-enhancers, plausibly enriched in trRNA-coding sequences (Fig. 2 and the online data resource). We therefore matched these DNA regions with the DBR for VDR, EBNA2, EBNA3C and AID, finding significant enrichments that allow to contextualize and prioritize genomic positions, cell/tissue identity or cell status associated to MS. Considering the harmonic score obtained from these colocalization analyses, the top hits in EBNA2, EBNA3C, and AID involved lymphoid (CD19+ B cell lines and lymphomas; T regulatory cells; tonsils) and monocyte-macrophage lineages (peripheral macrophages; dendritic cells) from experiments included in the ENCODE, dbSuper, roadmapEpigenomics databases; however, also global collections of superenhancers/enhancers and brain resident lineages appeared far from the bottom-left corner of Fig. 2 (the control datasets) (Fig. 2A–C, see also Additional File: Table S6 and the online resource). Even though immune cells prevailed also in VDR top hits, a less stringent polarization was seen, somehow reflecting the wide-spreading actions of

this transducer in human biology (Fig. 2D). However, with a more stringent cutoff of Harmonic Score > 40 that selects the most significant hits (Fig. Supplement 2), a core subset of MS-relevant cell lineages, shared across all four examined transducers, became evident (Additional File: Table S7).

**A data resource for future research on transcriptional regulation in MS.** A public web interface for browsing the results of our colocalization analysis is freely available at [www.mscoloc.com](http://www.mscoloc.com). This is a comprehensive genomic atlas disentangling specific aspects of MS gene-environment interactions to support further research on transcriptional regulation in MS. It includes the whole list of results derived from ROI, DBRs and database matches (Fig. 3a) across all performed experiments that yielded significant results. The user can navigate across the results and perform tailored queries searching and filtering for a variety of parameters, including MS-associated variant, DBR, experimental cell type, other match details (see Fig. 3b for all available search and filter modalities). Moreover, personalized HS, *p*-value, support and Odd Ratio threshold can easily be set to screen results, that are readily displayed in tabular format. To provide an example, we select “AID, EBNA2, EBNA3C, VDR” in the ‘Matched DBR region (s)’ panel and obtain the list of MS-associated SNPs (that proved to be enriched in genomic regions plausibly coding for tRNA) targeted by all four transducers (Fig. 3b,c). Through this approach we searched for MS-associated regions shared by the DBRs analyzed, and we were able to prioritize 275 genomic regions (almost half of the MS-associated GWAS SNPs) capable of binding at least 2 molecular transducers. These regions are ‘hotspots’ of interactions between genetic and nongenetic modifier of MS risk/protection: all four proteins (VDR, AID, EBNA2, EBNA3C) proved to target 24 regions, 3 of them 115 regions, and 2 of them 136 regions. A detailed legend and more example queries may be found on the online data resource website.

Finally, to obtain a functional mapping of MS-TrRNA regions, we attempted to identify MS-relevant genes by integrating our results with the ‘activity-by-contact’ (ABC) model (Fulco et al., 2019; Nasser et al., 2021), which was recently developed to define cell-specific, gene-enhancers connections according to chromatin conformation and accessibility, as well as to histone acetylation-methylation status. We retrieved a total of 77 gene-enhancers pairings (Additional File: Table S8), enriched in *IL6-JAK-STAT3*, *IL-18*, *IL2RB* pathways. Among these, we focused on MS variants-tRNA colocalization hotspots targeted by all four (AID, EBNA2, EBNA3C, VDR, *n* = 24) or three (AID, EBNA2, VDR, *n* = 60; see also Fig. 3c) molecular transducers, excluding EBNA3C, as it did not reach statistical significance in previous analysis (Table 3): ABC gene-enhancers connections were found for 10 out of 84 hotspot SNPs, corresponding to 31 genes (Table 4 and Fig. 4). As expected from the pleiotropy of enhancer activity, many MS-tRNA hotspots were linked to multiple genes differentially regulated in distinct cell types: for example, the MS-tRNA hotspot in *rs11026091* was linked to *MARGPRE* in T cells and *MARGPRE-AS1* in B cells (see also Additional File: Table S9). Results included regulators of immune cell activity (*MAP3K8*, *GIMAP8*, *TMEM176A*, *TMEM176B*), ion channels and solute carriers (*KCNH2*, *KCNMA1*, *SLC25A42*), and transcriptional modulators (*ICE2*, *SIN3B*, *NWD1*).

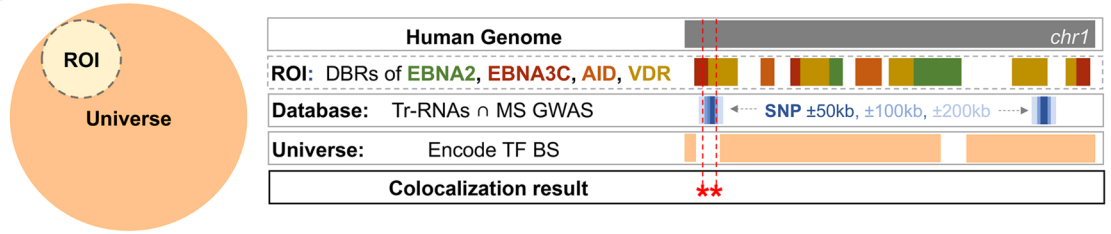
Moreover, in most cases, the ABC-identified genes differed from the candidate genes reported in MS GWAS, underscoring the relevance of integrative approaches to annotate statistical genomic associations.

## Discussion

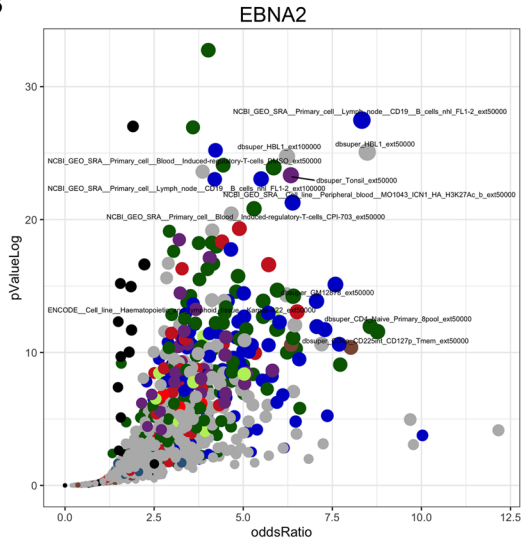
Our study supports the hypothesis that investigations on the transient transcriptome may contribute to clarify how the GWAS signals affect the etiopathogenesis of MS and possibly of other complex disorders. Specifically, we show that genomic regions coding for the transient transcriptome recently described in T cells<sup>25</sup>, are significantly enriched for both MS-associated GWAS variants, as well as for DNA binding sites for protein ‘transducers’ of non-genetic signals, chosen among those plausibly associated to MS. The colocalization of GWAS intervals and some DNA-binding factors involved in MS etiology has already been reported<sup>17–19</sup>. Here we reinforce this premise and extend the result to AID, whose DBRs were not previously correlated to MS-associated genetic signals. The result is of relevance considering the role of AID in B cell biology and the high effectiveness of B cell-depleting approaches recently introduced in clinical practice to tackle the disease progression (Cencioni et al. 2021). Our colocalization analysis suggests a model in which tRNA-coding regions are hotspots of convergence between genetic and non-genetic factors of risk/protection for MS. These hotspots are shared by two or more of the chosen transducers, indicating possible additive pathogenic effects or a multi-hits model to reach the threshold for MS development (see Fig. 3c and Additional File: Table S4). This model may reconcile previous evidences coming from ours and others’ studies on MS etiology: genetic susceptibility plausibly exerts a soft effect (with the notable exception of the major histocompatibility complex variants, that are known to directly shape the repertoire of the (auto-)immune effectors); in fact, single base changes in GWAS loci could conceivably lead to subtle changes in TT expression, and twin studies in Mediterranean areas showed a disease concordance as low as 1 out of 10 identical twin pairs (Ristori et al. 2006). A likely higher weight has the non-genetic component, that seems to be multiple and heterogeneous (with the notable exception of EBV, the most recurring and convincing risk factor for MS development; Ascherio et al., 2001), and that may favor stochastic events, by prevalently acting on genome regions coding for TT.

In homeostatic conditions, it can be hypothesized that DNA sequences coding for tRNA are composed of regulatory regions where genetic variability and non-genetic signals interact to finely regulate the gene expression according to cell identity, developmental or adaptive states, and time-dependent stimuli. As a matter of fact, the sequence variability of these regions and the strict time-dependence of their transcription could be instrumental to adaptive features; however, these same features make these regions susceptible to become dysfunctional or to

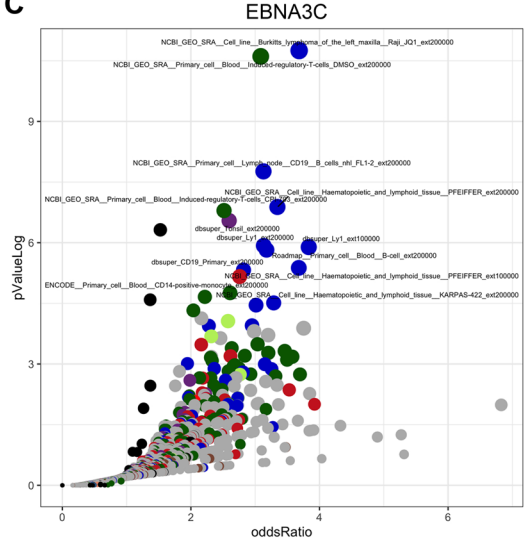
**A**



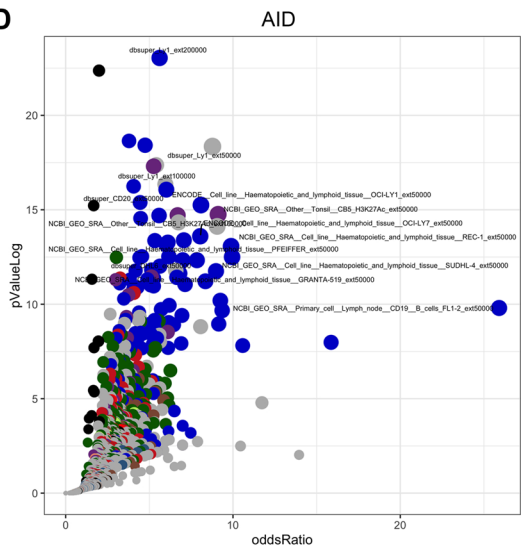
**B**



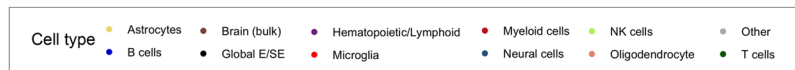
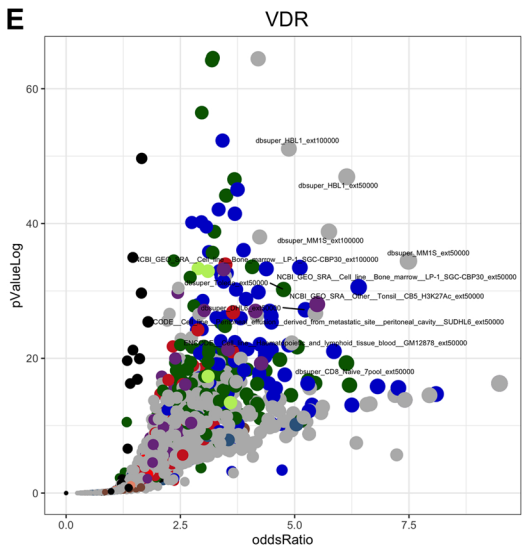
**C**



**D**



**E**



◀ **Figure 2.** Colocalization analysis of DBRs for human and viral molecular transducers, MS-associated SNPs and DNA regulatory regions derived from databases. **(A)** Schematic representation of the colocalization analysis. (ROI region of interest, DBR DNA binding regions, ENCODE TFBS transcription factor binding site). The figure shows the tracks we considered for the colocalization analyses. In brief, the ROI included the DBRs of MS-related viral and human transducers and was matched with MS-associated SNPs extended by 50, 100, and 200 kilobases that colocalize with regions plausibly coding for trRNAs (Database). As a control (Universe), we took from ENCODE the entire list of transcription factors binding sites. Results were considered significant if a colocalization was found across ROI and a Databases element without occurring in the Universe as a statistically significant match. **(B–E)**, Colocalization results of EBNA2, EBNA3C, AID, VDR. The charts display results of all matches, i.e., with MS-associated SNPs and their extension at  $\pm 50, 100, 200$  kb. X-axis shows the Odd Ratio, y-axis shows the  $-\log(p\text{Value})$ ; dot size is proportional to the Harmonic Score (HS). Thus, prioritized hits are represented by dots that occupy the upper-right area of the chart. Dots are coloured by cell type. Top-scoring hits in each subgroup are labeled; labels were arbitrarily designated according to the database of origin and the cell lineage where the enrichment occurred.

be the targets of pathogenic interaction. In some instances, these detrimental interactions come from outside the cell, such as in the case of EBV interference with host transcription<sup>38,39</sup>, and the pathogenic consequences of vitamin D deficiency; in other cases, the dysfunction develops within the cell, such as the tumorigenic activity of AID in B cells<sup>40,41</sup>.

To support the relationship between trRNA and transcription of regulatory DNA regions, we matched a large dataset of enhancers and super-enhancers with MS-GWAS signals and DBR for VDR, EBNA2, EBNA3C and AID. The significant enrichment in cell lines and cell status coming from the hematopoietic lineages and the CNS-specific cell subsets corroborates data coming from recent reports showing the relevance of contextualizing and prioritizing the role of MS-associated GWAS signals<sup>33,34,42,43</sup>. Our analysis supports the pivotal regulatory role of enhancer transcription (i.e., a main component of transient transcriptome) that was recently reported as not dispensable for gene expression at the immunoglobulin locus and for antibody class switch recombination<sup>44</sup>, though more research is needed to unravel such topic at a finer grain.

Reports on the dynamics of time-course data are a recent area of focus within the analysis of gene expression, specifically in immune cells. Although current studies use methods that investigate time points related to the stable transcriptome (RNA-seq performed with time spans of hours), they clearly show that gene expression dynamics may influence allele specificity, regulatory programs that seem to depend on autoimmune disease-associated loci, and different transcriptional profiles based on cell status after stimulation<sup>45</sup>. A recent work showed that an *IL2ra* enhancer, which harbors autoimmunity risk variants and was one of the first MS-associated loci from GWAS, has no impact on the gene level expression, but rather affects gene activation by delaying transcription in response to extracellular stimuli<sup>46</sup>. The importance of the timing in the gene expression control emerges also from several studies implicating enhancers and super-enhancers in the process of phase separation and formation of nuclear condensates, where the transcriptional apparatus steps-up to drive robust genic responses (Sabari et al., 2018). The overall process seems to be highly dynamic, with time spans of seconds or minutes, and hence compatible with the temporal features of the transient transcriptome, which could somehow contribute to the formation of these phase-separated condensates.

We suggest that studies on transient transcriptomes may integrate previous RNA-seq data in accounting for the interplay between genetic variability and non-genetic etiologic factors leading to MS development. Possible correlation between transient and persisting transcriptome obtained in ex-vivo and in-vivo experimental settings of neuroinflammation may help to better decipher the genomic regulatory syntax driven by non-coding DNA variants. In this context our results on ‘hotspots’, MS-associated trRNAs, and those obtained in the paper describing ABC mapping (Nasser et al., 2021) are concordant in identifying regulated additional genes, besides those resulting from current interpretations on GWAS data (Table 4 and Additional File: Table S6), thus revealing a complex scenario in cell-specific gene-enhancers interaction that supports the need of a wider approach in characterizing plausibly causal genes.

Components of a more-complex-than-anticipated regulation of gene expression could include transcriptional noise, transitory time-courses, erratic dynamics, and highly flexibility of some DNA regions, possibly oscillating between bistable states of enhancer and silencer<sup>47</sup>. Our analysis provides a platform for future studies on transient transcriptome, which we support by making our data resource available at [www.mscoloc.com](http://www.mscoloc.com). New gene regulatory models may emerge from this approach in order to better evaluate the meaning of GWAS in complex

	± 50 KB ± 50 KB				± 100 KB				± 200 KB			
	−log (pValue)	Odds ratio	Support	Harmonic Score	−log (pValue)	Odds ratio	Support	Harmonic score	−log (pValue)	Odds ratio	Support	Harmonic score
<b>EBNA2</b>												
Long half-life	0.023	0.478	3	0.644	0.062	0.717	8	1.708	<b>1.879</b>	<b>1.531</b>	33	<b>24.679</b>
Short half-life	<b>6.163</b>	<b>1.920</b>	<b>69</b>	<b>43.011</b>	<b>3.241</b>	<b>1.433</b>	<b>95</b>	<b>29.496</b>	<b>8.945</b>	<b>1.610</b>	<b>189</b>	<b>40.642</b>
<b>EBNA3C</b>												
Long half-life	0.064	0.572	2	1.669	0.006	0.321	2	0.185	0.182	0.914	11	4.500
Short half-life	0.070	0.794	16	1.923	0.023	0.752	28	0.661	0.066	0.875	58	1.841
<b>AID</b>												
Long half-life	0.089	0.682	3	2.303	0.283	1.024	8	6.477	0.051	0.726	11	1.432
Short half-life	<b>1.769</b>	<b>1.465</b>	<b>37</b>	<b>23.531</b>	<b>1.346</b>	<b>1.267</b>	<b>59</b>	<b>19.367</b>	<b>3.954</b>	<b>1.442</b>	<b>119</b>	<b>31.416</b>
<b>VDR</b>												
Long half-life	<b>1.737</b>	<b>1.502</b>	<b>32</b>	<b>23.571</b>	0.845	1.187	45	14.646	<b>2.315</b>	<b>1.322</b>	<b>97</b>	<b>25.031</b>
Short half-life	<b>2.221</b>	<b>1.239</b>	<b>152</b>	<b>23.734</b>	<b>2.336</b>	<b>1.181</b>	<b>267</b>	<b>23.460</b>	<b>11.478</b>	<b>1.367</b>	<b>548</b>	<b>36.561</b>

**Table 3.** Colocalization of human and viral transducer DBRs and MS-GWAS positions (at ± 50,100,200 kb range of extension) in DNA regions coding for transient transcripts; significant results ( $p < 0.05$ , corresponding to a  $-\log(p) > 1.301$ ) in bold. The transcript half-life is considered short if  $< 60'$  and long if  $\geq 60'$ , respectively.

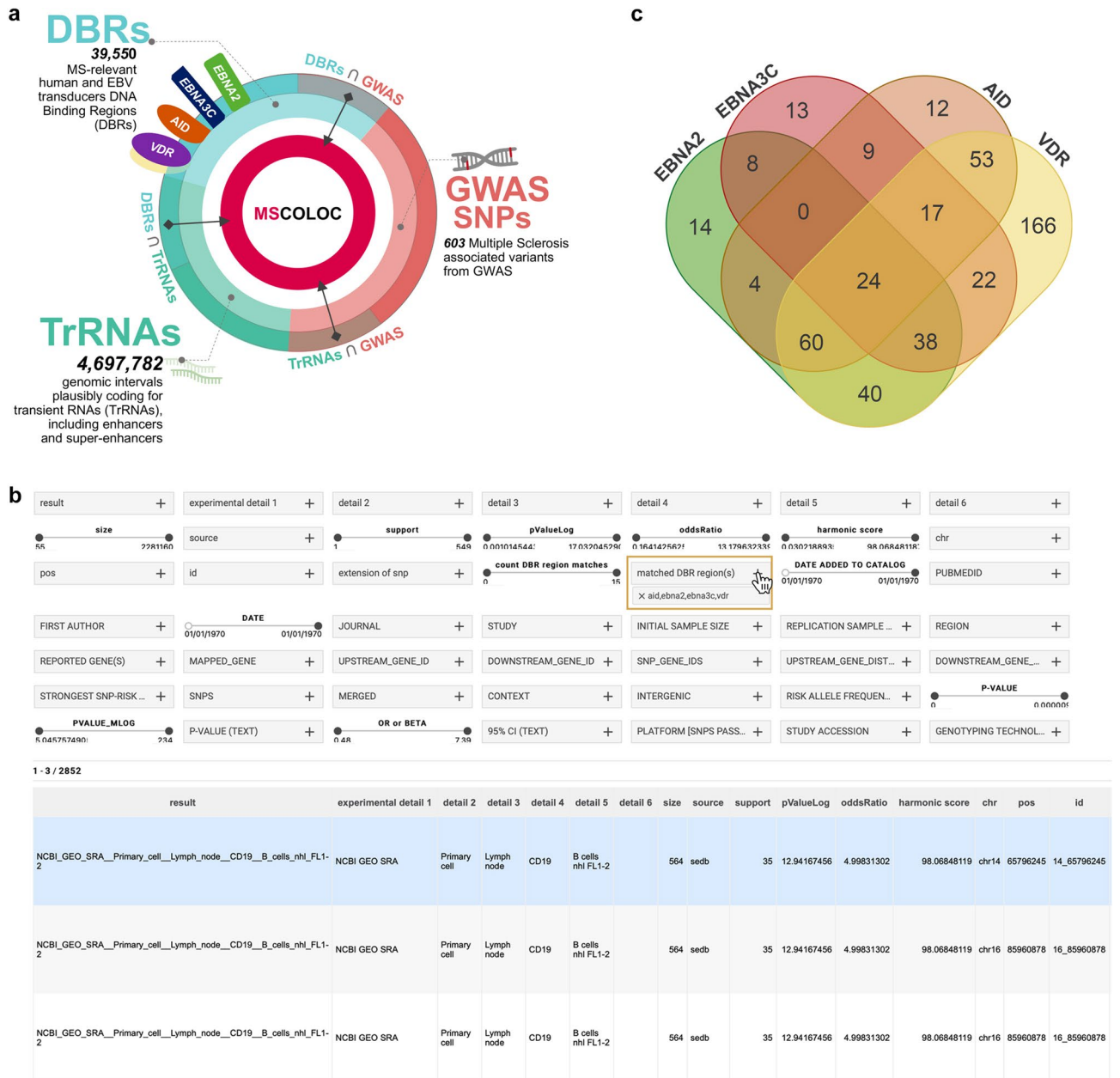
traits and the impact of the enhancer transcription<sup>44</sup>, which was recently reported as an ancient and conserved, yet flexible, genomic regulatory syntax<sup>48</sup>.

## Methods

**Data sources.** Analyses were performed in Python and R. A data freeze was applied on 3/1/2020. All GWAS data was gathered from the GWAS Catalog through its REST API<sup>32</sup>; about 1.5% of this data was filtered out as part of a QC process aimed at homogenizing legacy and more recent data. The MS GWAS regions were extracted from the overall GWAS Catalog data filtering by trait EFO\_0003885. All Transcription Factor Binding Site regions (TFBS) were obtained from the ENCODE portal<sup>49</sup>. All data was organized in various databases and data pipelines as detailed below. A modular and parallel data pipeline was created to: (1) readily generate and evaluate all experiments in the paper, (2) manage and organize all data coming from various region collections (42,075 ROI regions; 4,697,782 regions plausibly coding for trRNAs; 13,309,757 Universe regions), multiple ROIs (MS GWAS, EBNA2, EBNA3C, VDR, AID, etc.), databases of vast background regions as they were populated with the data obtained from GWAS Catalog, ENCODE, and other raw data sources, (3) provide overlaps and intersection among various data elements, annotate them with the original MS GWAS loci that generated the signal, and (4) generate the overarching data resource available at [www.mscoloc.com](http://www.mscoloc.com).

**ABC functional mapping.** The Activity-By-Contact model was applied to map genes regulated by selected MS-trRNA colocalization hotspots. Briefly, this model identifies gene-enhancers connection taking into account chromatin accessibility (ATAC-seq and DNase-seq experiments), histone modifications (H3K27ac ChIP-seq), and chromatin conformation (Hi-C)<sup>50</sup>. ABC analysis was performed using the ABC pipeline outputs for 131 cell types and tissues<sup>51</sup>. Gene-enhancers maps were produced through <https://flekschas.github.io/enhancer-gene-vis/>. Pathway and process enrichment analysis of mapped genes with the highest ABC score for each colocalization region was performed through Metascape<sup>52</sup>, using the entire human genome as background and the following ontology





**Figure 3.** A comprehensive genomic atlas on gene-environment interactions regulating transcription in MS. (a) Searchable results at mscoloc.com derive from the matches of GWAS MS regions, DNA binding regions of selected genomic transducers, and more than 4 million of regions annotated as plausible transient RNAs. (b) The user interface includes text panels and range sliders allowing extremely personalized queries, that combine statistical significance level (including Odd Ratio, pValue, support, and Harmonic Score), study source, SNP or reported gene, and so on. Filtered results are shown as tables ranked by HS, that can be saved, printed or shared through URL. In the example, the cursor selects ‘AID, EBNA2, EBNA3C, VDR’ in the ‘matched DBR region (s)’ panel looking for MS-associated SNPs (from the ROI, Additional File: Table S1) and their extensions at ± 50, 100, 200 kb that colocalized within DNA binding regions of the molecular transducers. The top hit represents the colocalization of the DBRs, a super-enhancer region derived from experiments on CD19 + B cells included in sedb, and the rs8007846 MS-associated SNP on chromosome 14. (c) The Venn diagram shows the number of non-redundant MS-associated SNPs derived from the query: for each transducer, SNPs are considered only once if present in more than one match. Intersections show the numbers of regions colocalizing with DBRs of multiple transducers. For instance: 8 regions colocalize with both EBNA2 and EBNA3C DBRs, but not with AID nor VDR DBRs; 24 regions colocalize with all four DBRs, and could be identified as regulatory “hotspots” in MS.

MS SNP-TrRNA hotspot	Transducers	Chr	Position	GWAS catalog reported gene	ABC linked gene
<i>rs11026091</i>	4/4	11	3,238,579	MRGPRE	<b>MRGPRG-AS1</b> , MRGPRG, <b>MRGPRE</b>
<i>rs11008218</i>	4/4	10	30,744,074	SVILP1	<b>MAP3K8</b>
<i>rs12048904</i>	4/4	1	100,865,980	EXTL2	GPR88, <b>CDC14A</b> , RTCA-AS1, RTCA-AS1
<i>rs11666377</i>	3/4	19	17,007,623	CPAMD8	<b>F2RL3</b> , HAUS8, MYO9B, NWD1, SIN3B
<i>rs11669861</i>	3/4	19	19,166,593	MEF2B, MEF2BNB-MEF2B	<b>SLC25A42</b>
<i>rs12909611</i>	3/4	15	60,865,204	RORA	ICE2, RORA-AS1, <b>RORA-AS2</b>
<i>rs1250551</i>	3/4	10	79,299,578	ZMIZ1	<b>KCNMA1</b>
<i>rs1152430</i>	3/4	14	90,906,615	RPS6KA5	CALM1, <b>LINC00642</b> , LOC105370619, LOC105370622
<i>rs11762408</i>	3/4	7	150,676,640	GIMAP6	ABCB8, AOC1, GIMAP8, <b>KCNH2</b> , TMEM176A, TMEM176B
<i>rs10931933</i>	3/4	2	201,247,748	CASP8	<b>SPATS2L</b> , KCTD18

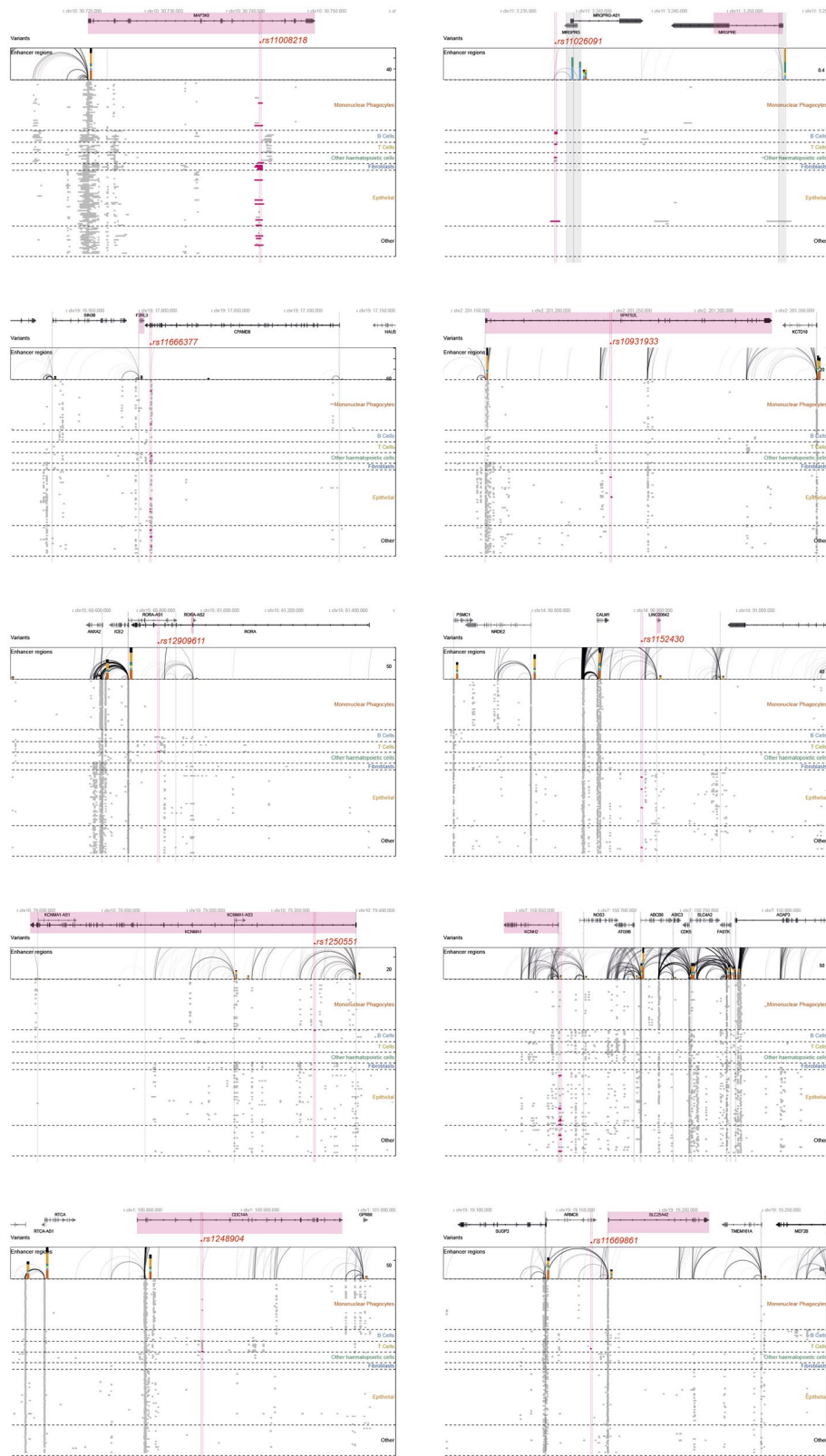
**Table 4.** Activity-By-Contact (ABC) functional mapping of MS-trRNA hotspots bound by 4/4 MS-relevant transducers or only 3/4 (AID, EBNA2, VDR). Genes who scored the highest ABC score are highlighted in bold.

sources were used: GO Biological Processes, KEGG Pathway, Reactome Gene Sets, Hallmark Gene Sets, Canonical Pathways, BioCarta Gene Sets and WikiPathways.

**Statistical analysis.** For SNP overlaps and region colocalization, we used LOLA<sup>53</sup> and Fisher's exact test with False Discovery Rate (Benjamini-Hochberg) to control for multiple testing. Linkage disequilibrium was considered as described in Sheffield & Bock, 2016. Resulting  $-\log(p\text{-value})$ , support, and Odds Ratio (OR) were combined into a single score inspired by the harmonic mean<sup>54</sup> and multi-objective optimization<sup>55</sup> with the formula below, where the spacing parameter  $k_p$  was set to 10.0 and we consider all three contributors equally, setting therefore weights  $w_i$  to 1.0. Statistical significance was taken at  $p < 0.05$ .

$$\text{HarmonicScore} = k_p * \frac{\sum_i w_i}{\frac{w_1}{-\log P} + \frac{w_2}{\text{Supp}} + \frac{w_3}{\text{OR}}}$$

For pathway analysis in Metascape, enrichment p-values were calculated based through the accumulative hypergeometric distribution, q-values were calculated using the Benjamini-Hochberg method to account for multiple testing.



**Figure 4.** ABC gene-enhancer mapping of MS-trRNAs regulatory hotspots. The figure displays the positional mapping of selected MS-trRNAs-DBRs hotspots, their colocalization with regions encoding for regulatory elements active in specific cell types (*Mononuclear phagocytes, B cells, T cells, Other haematopoietic cells, epithelial cells, Other*), and the enhancer-gene connections that determine the ABC mapping. The gene with the highest ABC score for each hotspot is highlighted in pink.

## Data availability

The dataset supporting the conclusions of this article is available at the website [www.mscoloc.com](http://www.mscoloc.com).

Received: 14 October 2021; Accepted: 22 April 2022

Published online: 09 May 2022

## References

- Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**(7345), 43–49 (2021).
- Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**(6099), 1190–1195 (2012).
- Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* **95**(5), 535–552 (2014).
- Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**(7539), 337–343 (2015).
- Vahedi, G. *et al.* Super-enhancers delineate disease-associated regulatory nodes in t cells. *Nature* **520**(7548), 558–562 (2015).
- chris.cotsapas@yale.edu IMSGCEa, Consortium IMSG. Low-frequency and rare-coding variation contributes to multiple sclerosis risk. *Cell* **180**(2):403 (2020).
- Mumbach, M. R. *et al.* Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* **49**(11), 1602–1612 (2017).
- van Arensbergen, J. *et al.* High-throughput identification of human SNPS affecting regulatory element activity. *Nat. Genet.* **51**(7), 1160–1169 (2019).
- Calderon, D. *et al.* Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat. Genet.* **51**(10), 1494–1505 (2019).
- Ohkura, N. *et al.* Regulatory t cell-specific epigenomic region variants are a key determinant of susceptibility to common autoimmune diseases. *Immunity* **52**(6), 1119–1132.e1114 (2020).
- Meuleman, W. *et al.* Index and biological spectrum of human dnase I hypersensitive sites. *Nature* **584**(7820), 244–251 (2020).
- Consortium G. The gtex consortium atlas of genetic regulatory effects across human tissues. *Science* **369**(6509), 1318–1330 (2020).
- Ristori, G. *et al.* Multiple sclerosis in twins from continental italy and sardinia: A nationwide study. *Ann. Neurol.* **59**(1), 27–34 (2006).
- Fagnani, C. *et al.* Twin studies in multiple sclerosis: A meta-estimation of heritability and environmentality. *Mult. Scler.* **21**(11), 1404–1413 (2015).
- Bordi, I. *et al.* A mechanistic, stochastic model helps understand multiple sclerosis course and pathogenesis. *Int. J. Genom.* **2013**, 910321 (2013).
- Bordi, I. *et al.* Noise in multiple sclerosis: Unwanted and necessary. *Ann. Clin. Transl. Neurol.* **1**(7), 502–511 (2014).
- Ricigliano, V. A. *et al.* Ebna2 binds to genomic intervals associated with multiple sclerosis and overlaps with vitamin d receptor occupancy. *PLoS One* **10**(4), e0119605 (2015).
- Mechelli, R. *et al.* Epstein-barr virus genetic variants are associated with multiple sclerosis. *Neurology* **84**(13), 1362–1368 (2015).
- Harley, J. B. *et al.* Transcription factors operate across disease loci, with ebna2 implicated in autoimmunity. *Nat. Genet.* **50**(5), 699–707 (2018).
- Ascherio, A. *et al.* Epstein-barr virus antibodies and risk of multiple sclerosis: A prospective study. *JAMA* **286**(24), 3083–3088 (2001).
- Simon, K. C., Munger, K. L. & Ascherio, A. Vitamin d and multiple sclerosis: Epidemiology, immunology, and genetics. *Curr. Opin. Neurol.* **25**(3), 246–251 (2012).
- Bjornevik, K. *et al.* Longitudinal analysis reveals high prevalence of epstein-barr virus associated with multiple sclerosis. *Science* **375**(6578), 296–301 (2022).
- Lanz, T. V. *et al.* Clonally expanded b cells in multiple sclerosis bind ebv ebna1 and glialcam. *Nature* **603**(7900), 321–327 (2022).
- Schwalb, B. *et al.* Tt-seq maps the human transient transcriptome. *Science* **352**(6290), 1225–1228 (2016).
- Michel, M. *et al.* Tt-seq captures enhancer landscapes immediately after t-cell stimulation. *Mol. Syst. Biol.* **13**(3), 920 (2017).
- Villamil, G., Wachutka, L., Cramer, P., Gagneur, J. & Schwalb, B. Transient transcriptome sequencing: Computational pipeline to quantify genome-wide rna kinetic parameters and transcriptional enhancer activity. *bioRxiv* 659912 (2019).
- Natoli, G. & Andrau, J. C. Noncoding transcription at enhancers: General principles and functional models. *Annu. Rev. Genet.* **46**, 1–19 (2012).
- Bose, D. A. *et al.* RNA binding to CBP stimulates histone acetylation and transcription. *Cell* **168**(1–2), 135–149.e122 (2017).
- Weinert, B. T. *et al.* Time-resolved analysis reveals rapid dynamics and broad scope of the CBP/p300 acetylome. *Cell* **174**(1), 231–244.e212 (2018).
- Larsson, A. J. M. *et al.* Genomic encoding of transcriptional burst kinetics. *Nature* **565**(7738), 251–254 (2019).
- O'Donoghue, G. P., Bugaj, L. J., Anderson, W., Daniels, K. G., Rawlings, D. J., Lim, W. A. T cells selectively filter oscillatory signals on the minutes timescale. *Proc Natl Acad Sci U S A* **118**(9) (2021).
- Buniello, A. *et al.* The nhgri-ebi gwas catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucl. Acids Res.* **47**(D1), D1005–D1012 (2019).
- Nott, A. *et al.* Brain cell type-specific enhancer-promoter interactome maps and disease. *Science* **366**(6469), 1134–1139 (2019).
- Consortium IMSG. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365**(6460) (2019).
- Marcucci, S. B. & Obeidat, A. Z. Ebna1, ebna2, and ebna3 link epstein-barr virus and hypovitaminosis d in multiple sclerosis pathogenesis. *J. Neuroimmunol.* **339**, 577116 (2020).
- Bäcker-Koduah, P. *et al.* Vitamin d and disease severity in multiple sclerosis-baseline data from the randomized controlled trial (evidims). *Front. Neurol.* **11**, 129 (2020).
- Sun, Y. *et al.* Critical role of activation induced cytidine deaminase in experimental autoimmune encephalomyelitis. *Autoimmunity* **46**(2), 157–167 (2013).
- Mechelli, R. *et al.* Viruses and neuroinflammation in multiple sclerosis. *Neuroimmunol. Neuroinflammation* **8**, 269–83 (2021).
- Park, A. *et al.* Global epigenomic analysis of kshv-infected primary effusion lymphoma identifies functional. *Proc. Natl. Acad. Sci. U S A* **117**(35), 21618–21627 (2020).
- Meng, F. L. *et al.* Convergent transcription at intragenic super-enhancers targets aid-initiated genomic instability. *Cell* **159**(7), 1524–1537 (2014).
- Qian, J. *et al.* B cell super-enhancers and regulatory clusters recruit aid tumorigenic activity. *Cell* **159**(7), 1524–1537 (2014).
- Orrù, V. *et al.* Complex genetic signatures in immune cells underlie autoimmunity and inform therapy. *Nat. Genet.* **52**(10), 1036–1045 (2020).
- Factor, D. C. *et al.* Cell type-specific intralocus interactions reveal oligodendrocyte mechanisms in MS. *Cell* **181**(2), 382–395.e321 (2020).

44. Fitz, J. *et al.* Spt5-mediated enhancer transcription directly couples enhancer activation with physical promoter interaction. *Nat. Genet.* **52**(5), 505–515 (2020).
45. Gutierrez-Arcelus, M. *et al.* Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. *Nat. Genet.* **52**(3), 247–253 (2020).
46. Simeonov, D. R. *et al.* Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549**(7670), 111–115 (2017).
47. Halfon, M. S. Silencers, enhancers, and the multifunctional regulatory genome. *Trends Genet.* **36**(3), 149–151 (2020).
48. Wong, E. S. *et al.* Deep conservation of the enhancer regulatory code in animals. *Science* **370**(6517), eaax8137 (2020).
49. Sloan, C. A. *et al.* Encode data at the encode portal. *Nucl. Acids Res.* **44**(D1), D726–732 (2016).
50. Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**(12), 1664–1669 (2019).
51. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**(7858), 238–243 (2021).
52. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**(1), 1523 (2019).
53. Sheffield, N. C. & Bock, C. Lola: Enrichment analysis for genomic region sets and regulatory elements in R and bioconductor. *Bioinformatics* **32**(4), 587–589 (2016).
54. Wilson, D. J. The harmonic mean. *Proc. Natl. Acad. Sci. U S A* **116**(4), 1195–1200 (2019).
55. Umeton, R., Sorathiya, A., Liò, P., Papini, A., Nicosia, G. Design of robust metabolic pathways. In *Proceedings of the 48th Design Automation Conference (DAC '11)*. ACM, New York, NY, USA, 747–752 (2011).

## Acknowledgements

Authors would like to thank Dr. Adem Albayrak for the editorial suggestions that improved this manuscript.

## Author contributions

R.U., G.B., R.B., R.M., M.S., and G.R. conceived and planned the analysis. R.R., V.R., E.M., C.R., S.R., and M.C.B. guided data engineering and database generation from raw data. R.U., R.P.U. and G.B. developed the data resource and analyzed the data. R.U., G.B., R.P.U., GR and M.S. wrote the manuscript. R.U, G.B, R.B., and R.M. created all tables and figures. R.B., R.M., M.S. and G.R. supervised the project. All the authors, including S.R., M.C.B., R.R., V.R., E.M. and C.R., contributed to fortnight discussion for data interpretation and new analysis planning. All the authors reviewed and approved the manuscript.

## Funding

This work was supported by “Progetti Grandi Ateneo” 2020, Sapienza University of Rome. MS and GR are supported by CENTERS, a Special Project of, and financed by, FISM—Fondazione Italiana Sclerosi Multipla. RPU is supported by the National MS Society.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-11444-w>.

**Correspondence** and requests for materials should be addressed to R.U., M.S. or G.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022