Defective Integrator activity shapes the transcriptome of patients with multiple sclerosis

Yevhenia Porozhan*[(1)], Mikkel Carstensen*[(2)], Sandrine Thouroude[(1)], Mickael Costallat[(1)], Christophe Rachez[(1)], Eric Batsché[(1)], Thor Petersen**[(3)], Tove Christensen**[(2)(4)], Christian Muchardt[(1)(5)]

* Participated equally to this work, ** Participated equally to this work

(1) Institut de Biologie Paris-Seine (IBPS), CNRS UMR 8256, Biological Adaptation and Ageing, Sorbonne Universités, 75005, Paris, France.
(2) Department of Biomedicine, Aarhus University, Skou Building, Høegh-Guldbergsgade 10, DK-8000, Aarhus C, Denmark
(3) Department of Neurology, Hospital of Southern Jutland and Research Unit in Neurology, Department of Regional Health Research, University of Southern Denmark.
(4) Department of Clinical Medicine, Aarhus University, Forum, Palle Juul-Jensens Blvd. 5, 8200 Aarhus N, Denmark

(5) Correspondance : christian.muchardt@sorbonne-universite.fr

Abstract

While the role of the immune system in multiple sclerosis (MS) is widely acknowledged, our comprehension of the transcriptional foundations of this prevalent neurological disorder remains largely incomplete. Here, we conducted high-depth RNA sequencing on monocytes from a cohort of patients with MS to explore rare RNA species associated with the disease. In a subset of the patients, a markedly altered transcriptome was observed, coinciding with a decreased expression of the epigenetic silencer HP1α/CBX5. These patients also exhibited diminished expression of multiple genes encoding subunits of the Integrator complex. Alongside, there were clear indications of decreased Integrator activity, including impaired U snRNA and enhancer-RNA maturation/degradation and dysregulated RNA polymerase II (RNAPII) pause-release. Inactivation of Cbx5 in the mouse recapitulated the defects in Integrator activity and resulted in hypersensitivity to Experimental Autoimmune Encephalomyelitis (EAE). While these data implied an unexpected function for HP1α in regulating RNAPII pause-release, they also showed that the dysregulation of numerous genes in MS patients could be attributed to either the biased distribution of transcription toward the 5' end of genes or the heightened elongation and stability of enhancer RNAs. We therefore propose that impaired Integrator activity, which has been hinted at by mutations within Integrator subunits previously linked to an increased risk of MS, can provide a well-defined mechanistic understanding of the transcriptional anomalies associated with the disease.

Multiple sclerosis (MS) is an acquired demyelinating disorder of the central nervous system (CNS) characterized by chronic inflammation and the formation of scar tissue (sclerosis) in various areas of the CNS. It is the most common disabling neurological disease of young adults, and currently available treatments focus primarily on managing the symptoms.

During the course of the disease, immune cells traverse the blood-brain barrier (BBB), and attack the myelin sheath protecting the neurons. This results in numbness in various body parts, paresis, problems with coordination and balance, blurred vision, slurred speech, and cognitive changes. MS can manifest in at least four distinct forms. Clinical isolated syndrome (CIS) is the first clinical presentation of a disease that shows characteristics of inflammatory demyelination that could be MS, but has yet to fulfill criteria for dissemination in time. When dissemination in time is fulfilled, then Relapsing-Remitting MS (RRMS) is the most common form of MS, characterized by clearly defined relapses or flare-ups of symptoms followed by periods of partial or complete recovery. This stage is often followed by Secondary Progressive MS (SPMS) when the disease gradually worsens without distinct relapses or remissions. Primary Progressive MS (PPMS) is a form of MS characterized by a gradual worsening of symptoms from the beginning, without distinct relapses or remissions (Lublin et al., 2014).

The etiology of the disease remains essentially unknown and is highly debated. Mostly, the CNS directed autoimmunity is considered the result of a complex interaction between genetic susceptibility, hormonal factors, and environmental cues. These environmental cues, which also refer to lifestyle, include EBV infection, exposure to tobacco smoke and organic solvents, obesity in adolescence, limited sun exposure/low vitamin D, and more. Some of these cues seems to be relevant only when they occur during adolescence, particularly EBV serology, obesity, and possibly vitamin D deficit (Christensen & Muchardt, 2019). As the age of onset ranges from 20 to 40, there seems to be a memory mechanism involved. In that respect, modulation of the peripheral adaptive immune response has received the most attention (Hemmer et al., 2015).

A less explored complementary memory mechanism may rely on epigenetic processes causing long-lasting changes in the transcriptional landscape in immune cells, but possibly also in cells of the CNS and of the BBB. Previously, we have explored this possibility by examining the activity of HP1α/CBX5, a transcriptional repressor with affinity for both chromatin and RNA (Maison et al., 2002, 2011; Muchardt et al., 2002), and encoded by a gene sharing one of its promoters with hnRNPA1, a gene harboring SNPs previously associated with elevated risk of MS (Clarke et al., 2021). CBX5 directly binds the tail of histone H3 when it is tri-methylated on lysine 9 (H3K9me3), a modification associated with transcriptional repression and present on the promoter of many cytokine genes in the absence of pro-inflammatory stimulation and on many DNA repeats, including endogenous retroviruses (HERVs). In our initial study, we found that the affinity of CBX5 for H3K9me3 was compromised upon conversion of the nearby arginine 8 into a citrulline (H3R8 to H3Cit8) by nuclear PADI enzymes (PADI2 and PADI4). Furthermore, we described a significant increase in H3Cit8 levels among MS patients, suggesting that reduced HP1 activity contributed to the persistent state of inflammation and the often-reported re-expression HERV sequences in association with MS (Sharma et al., 2012).

In parallel, other members of the HP1 family have been described as regulators of RNA metabolism, including a role for SWI6 in the destruction of heterochromatic transcripts in yeast (Keller et al., 2012), and a role for HP1γ/CBX3 in the regulation of alternative splicing in mammals (Mata-Garrido et al., 2022; Rachez et al., 2021; Saint-André et al., 2011; Smallwood et al., 2012; Yearim et al., 2015). We therefore wished to re-explore the possible role of CBX5

in MS in the light of genome-wide RNA-seq data from MS patient monocytes, with a depth sufficient to detect rare RNA species. This approach has revealed that, for MS patients, low levels of HP1α/CBX5 is associated with numerous transcriptional defects all suggestive of reduce Integrator (INTcom) activity, a multi-subunit assembly that participates in both maturation of non-coding RNAs (ncRNAs) and regulation of RNAPII elongation (Kirstein et al., 2021). The discovery of this transcriptional defects that can be reproduced in a mouse model inactivated for Cbx5 provides mechanistic insight in multiple gene-deregulation events associated with MS.

A subset of MS patients displays extensive transcriptional deregulation in monocytes

With the objective of obtaining high quality RNA-seq allowing detection of rare RNA species, we collected monocytes from a cohort of 18 MS patients and 7 control patients (Figure 1A). To explore eventual changes in the RNA populations induced by the disease progression, we selected a very heterogeneous set including patients with familial MS (#68, #125, and #130 – one family, 3 generations), under immune modulating treatment (#125, #127, #135, #136, and #139), or with comorbidities (psoriasis, IDDM, Graves, Crohn and ulcerative colitis). The control group, designated as symptomatic controls (SCs), comprised individuals who had sought medical assistance for symptoms suggestive of MS, but who, upon examination, exhibited no objective clinical or paraclinical findings conclusive for a specific neurological disease at the time of sample collection (Figure 1A). The RNA-seq analysis was conducted using a stranded paired-end protocol, yielding an output of approximately 110 million reads per sample. Principal Component Analysis (PCA) on the RNA-seq data, profiling the 500 most variable genes, identified a substantial proportion of variance in gene expression across the samples. It further revealed the existence of two distinct biological subgroups, with PC1 scores that were either lower or significantly higher than those of the SCs (Figure 1B).

We further noted that expressions levels of CBX5 varied among patients and that ranking the patients according to their respective CBX5 expression level essentially recapitulated the PCA. Indeed, patients with CBX5 expression levels in the lower quartile (Lo-CBX5) largely coincided with the patients strongly affected in their transcriptome, while patients with CBX5 expression levels in the upper quartile (Hi-CBX5) remained transcriptionally close to the SCs (Figure 1A, 1B, and 1C). The highly transcriptionally divergent/Lo-CBX5 patients were all either RRMS, SPMS, or PPMS, while the Hi-CBX5 group was enriched in early-stage patients (Figure 1D). None of the other criterion proved relevant (patient age, gender, or comorbidities).

Pathway analysis on the genes differentially expressed between Hi-CBX5 patients and Scs (38 upregulated genes and 109 down-regulated genes, 2-fold or more, adj. pVal<0.05, baseMean>10 – Figure 1E) revealed that down-regulated genes were enriched in Vitamin D regulated genes (analysis of the ChEA database, Sup. Figure 1A). This observation may reflect the importance of Vitamin D in early phases of the disease.

In contrast to the Hi-CBX5 patients, Lo-CBX5 patients displayed a dramatically high number genes with modified expression when compared to the SCs, with a total of 1395 upregulated and 1350 down-regulated genes (2-fold or more, adj. pVal<0.05, baseMean>10 – Figure 1F). As anticipated from earlier studies, pathway analysis using the KEGG database on upregulated genes identified activation of the IL17-, TGF-beta-, NFkappaB-, and MAPK-pathways as well as genes associated with hypoxia (Elkjaer et al., 2022) (Sup. Figure 1B). The pathway analysis also detected a significant enrichment in genes associated with ferroptosis. This form of regulated cell death characterized by the iron-dependent accumulation of lipid peroxides, has previously been linked to neurodegeneration in MS (Luoqian et al., 2022). It is induced either through

inactivation of the glutathione peroxidase, GPX4, the enzyme responsible for detoxifying lipid hydroperoxides, or by functional inhibition of the cystine-glutamate antiporter, a two-subunit complex composed of SLC7A11 and SLC3A2. Yet, in the monocytes of the Lo-CBX5 patients, while p53 (TP53), which exerts a protective role against ferroptosis was strongly down-regulated (2-fold, Adj. pVal<10^{-4}), we observed only a moderate upregulation of enzymes promoting lipid peroxidation (ACSL4 and LPCAT3). In addition, the detoxifying pathways (SLC7A11, SLC3A2, and GPX4) were markedly upregulated, and markers of ferroptosis were either down-regulated PTGS2 or unchanged (CHAC1) (Figure 1G and Sup. Figure 1C). Although transcriptional activity is only indicative on enzymatic activities, this suggests that pro-ferroptotic signals are activated in the Lo-CBX5 monocytes, but are offset by compensatory mechanisms inhibiting cell death.

Further analysis of the upregulated genes using the WikiPathway database surprisingly identified an enrichment in genes associated with adipogenesis (Sup. Figure 1D). In that context, we noted also an upregulation of p16/CDKN2A and p21/CDKN1A, possibly indicative of at least partial growth arrest or differentiation (Figure 1H).

Analysis of the down-regulated genes with either Reactome, BioPlanet, or WikiPathway databases revealed a significant enrichment of genes associated with Type I interferon signaling (Figure 1I and Sup. Figure 1E). This is consistent with the observed benefits of administering interferon alpha/beta to patients with MS (Filipi & Jack, 2020).

In conclusion, most of the traits commonly linked with individuals with MS, such as heightened activation of stress pathways and diminished activity of type I Interferon, are predominantly observed in patients who exhibit lower levels of CBX5 expression, in association with a significantly disrupted transcriptome. In contrast, patients expressing high levels of CBX5, enriched in patients in early phases of the disease, display only a minor deregulation of their transcriptome, exhibiting reduced activity of vitamin D reactive genes as their main characteristic.

Evidence for reduced Integrator activity in Lo-CBX5 patients

To investigate the regulatory mechanism underlying the extensive transcriptome reprogramming observed in Lo-CBX5 patients, we confronted the list of genes upregulated in Lo-CBX5 patients (relative to the SCs) with the ENCODE Transcription Factor Data Bank, which serves as a comprehensive repository of information about transcription factors and their involvement in controlling gene expression. This approach very clearly designated NELFE targets as enriched among the upregulated genes (Figure 2A). NELFE is a subunit of the Negative Elongation Factor (NELF) complex, which controls the speed and efficiency of RNAPII movement along the DNA, and regulates gene expression by causing RNAPII pausing downstream of the transcription start site (TSS). As all RNAPII-transcribed genes require NELF, annotated NELF-targets designate genes at which promoter escape rather than transcription initiation is the rate-limiting step (Gilchrist et al., 2008).

The activity of NELF depends on the presence of at least two other complexes, the Integrator complex (INTcom) and the DRB Sensitivity-Inducing Factor (DSIF) complex. In addition, the negative impact of these complexes on transcription is balanced out by factors having a positive effect on elongation, including the Positive Transcription Elongation Factor b (P-TEF-b) and the transcription factor MYC.

Examination of the transcriptome did not indicate any clear variations in expression of the subunits of NELF. In contrast, 6 out 12 subunits of the INTcom were significantly (adj pVal<0.05) down-regulated (Figure 2B). Among these, INTS8 was previously reported to

harbor single nucleotide polymorphisms (SNPs) associated with an elevated risk of MS (Briggs & Sept, 2021). The INTcom was initially characterized for its function in the 3'-end processing of U snRNAs, during which it recognizes the 3' cleavage site and facilitates the cleavage reaction (Yamamoto et al., 2014). In that context, we noted that an earlier study had reported aberrant U snRNA polyadenylation in patients with MS, a known manifestation of a defective maturation of these RNA species (Spurlock et al., 2015). Examination of our patient RNA-seq data with a genome browser revealed increased accumulation of 3' extensions at U snRNA transcription at multiple loci, characteristic of a hampered maturation (Figure 2C and Sup. Figure 2A). A compromised production of U snRNAs was also suggested by a defect in the maturation of several replication-dependent histone mRNAs, a process that relies on the U7 snRNA (Mowry & Steitz, 1987) (Figure 2D).

In parallel, the INTcom also promotes enhancer RNA (eRNA) instability, recognizing and cleaving their 3'-end (Lai et al., 2015). We therefore quantified RNA-seq reads within annotated monocyte enhancers, revealing a significant increase in eRNA accumulation in Lo-CBX5, but not in Hi-CBX5 patients (Figure 2D). In this analysis, we also quantified RNA-seq reads mapping to DNA repeats, and neither SINEs, LINEs, or endogenous retroviruses (HERVs) were upregulated genome-wide in the patients (neither Lo-CBX5s, nor Hi-CBX5s) (Sup. Figure 2B, 2C, 2D). The unaltered transcription of DNA repeats also documents that the low levels of CBX5 did not result in an overall restructuring of chromatin-mediated transcriptional silencing. We noted however an upregulation of HERVs coopted as enhancers, with transcription occasionally bleeding into HERV-encoding DNA sequences (Sup. Figure 2E and example Figure 2F). This suggested that the commonly observed upregulation of HERVs in patients with MS might be attributed to a deficiency in terminating or processing eRNAs that originate from these repetitive elements. A similar phenomenon also seemed responsible for ectopic expression of protein coding genes. For example, we noted that the leptin (LEP) gene, normally detected only in adipose tissue and encoding a secreted hormone, was strongly activated in the Lo-Cbx5 patient monocytes (6-fold increased expression, Adj. pVal<0.03), apparently as a consequence of eRNAs originating from an upstream enhancer elongating into the LEP gene body, and undergoing splicing (Figure 2G). Thus, the modified turn-over of eRNAs may participate in the activation of adipogenesis pathways mentioned in Sup. Figure 1D.

Modified RNAPII pause-release affects expression and splicing of genes relevant for MS

CBX5 activity is best documented in heterochromatin. However, CBX5 ChIP-seq data from HepG2 cells documented a clear enrichment of this chromatin-bound protein at the transcription start site (TSS) of protein-coding genes, suggesting that the chromatin regulator may also play a role in transcription initiation (Sup. Figure 3A). In parallel, during transcription initiation at protein coding genes, INTcom participates in the regulation of RNAPII pause-release, and reduced activity of this complex results in RNAPII entering the gene with reduced competence for elongation. This translates into exacerbated transcription downstream of the TSS, a phenomenon best observed with sequencing approaches detecting nascent pre-mRNA (Stein et al., 2022). However, visualization is also possible with RNA-seq data (Stadelmayer et al., 2014). To favor visualization of nascent RNAs in our RNA-seq patient data, we selected a set of genes with first introns exceeding 20kb in length, allowing the monitoring of reads originating from pre-mRNA over a long, uninterrupted region. Average transcription profiles over these genes revealed consistently increased accumulation of reads downstream of the TSS in the Lo-CBX5 patients as compared to SCs and Hi-CBX5 patients, thereby mirroring the

pattern observed in earlier *in vitro* studies (Figure 3A and 3B, and (Stein et al., 2022)). This accumulation was not seen at the 3' end of genes.

As an alternative method to visualize the phenomenon, we generated meta-genes with and without intronic sequences. When intronic sequences are excluded, the profiles show almost exclusively mature mRNA, more stable and therefore more abundant than pre-mRNA. Inversely, when including all sequences, the mature mRNAs account for only a minor fraction of the signal as exons are in average 300 nucleotides in length, while introns are several kilobases. These profiles confirmed increased transcription in the 5' half of the genes in the Lo-CBX5 patients as compared to SCs and Hi-CBX5 patients (Figure 3C). In contrast, the mRNA output of the genes was in average equivalent in all donors (Figure 3D). Finally, we noted that genes activated in Lo-CBX5 patients were in average shorter than the down-regulated genes, a possible manifestation of globally increased transcription of regions located immediately downstream of the TSS (Sup. Figure 3B).

Consistent with these profiles, examination of the RNA-seq data with a genome browser allowed identifying multiple genes relevant for MS and displaying increased transcription over the first exon and the first intron, with consequences on their global level of expression. For example, the IRF2 gene, involved in repressing the type I interferon response, showed a clear increase in transcription down-stream of the TSS ("More" arrow, Figure 3E), while transcription was reduced from exon 2 and onwards ("Less" arrow, Figure 3E). This suggested that reduced control of RNAPII pause-release at this gene resulted in decreased production of full-length pre-mRNA. At other genes, uncontrolled promoter escape resulted in a net increase in the production of mature mRNA. This was exemplified by the PPARG and SLC7A11 genes, mostly transcribed in their 5'-regions, yet overall activated respectively 6-folds and 3-folds according to DeSeq2 quantification (Adj. pVal<0.05 - see representative patients in Figure 3F and Sup. Figure 3C). Increased expression by this mechanism of PPARG, a transcription factor involved in adipocyte differentiation, may, together with leptin, participate in the ectopic activation of adipogenesis in the monocytes detected above (Sup. Figure 1D). The strong initiation/poor elongation behavior of the RNAPII also modified expression of regulatory RNAs when embedded in larger genes. For example, it caused increased expression of mir3188 previously described as an MS risk factor (Afrasiabi et al., 2021; Rhead et al., 2019) (Figure 3G), while resulting in decreased expression of MIR4435-2, a miR located further inside its host gene (Figure 2H).

From earlier studies, MS is known also to affect alternative splicing (AS) at many genes (Evsyukova et al., 2010). Therefore, we examined whether the MS patients exhibited any alterations in overall splicing levels, using the RNA-seq data to quantify junction reads (reads produced when the sequenced RNA fragments span over exon-intron boundaries - such reads serve as evidence of a splicing event). This approach showed that, on average, less splicing events were detected in Lo-CBX5 patients when compared to Hi-CBX5 and SC donors, with a more pronounced impact at the 5'-end of genes (profiles Figure 3I, red below green and blue). Further analysis of the RNA-seq data with the RMATs package confirmed that alternative splicing was affected in the MS patients, whether displaying Lo- or Hi-CBX5 expression. Yet, it also documented that the reduced splicing level in the Lo-CBX5 patients essentially translated into exon skipping, consistent with the fact that fewer splicing events are needed when fewer exons are included (Figure3J). Inversely, Hi-Cbx5 patients displayed reduced exon skipping when compared to the SCs (Figure 3K). Interestingly, increased exon skipping in Lo-CBX5 patients was observed at the INTS6 gene, suggesting that reduced Integrator activity is a self-

reinforcing phenomenon (Figure 3L). Increased exon skipping was also observed at MBP and GAPDH genes encoding MS auto-antigens (Figure 3M, and Sup. Figure 3D).

Thus, reduced Integrator activity translates into either up- or down-regulation of numerous genes associated with multiple sclerosis (MS), the outcome possibly varying depending on the extent to which these genes rely on RNAPII pause-release for their regulation. In addition, gene expression is further impacted by overall reduced splicing.

Inactivation of Cbx5 in the mouse promotes inflammation and exacerbates EAE.

To explore further a possible causative role for reduced CBX5 activity in the MS pathogenesis, we implemented a mouse model inactivated for the Cbx5 gene. This mouse model was viable, although animals homozygote for the Cbx5 mutation (Cbx5-/-) displayed high perinatal mortality.

We first challenged the mouse model with Experimental Autoimmune Encephalomyelitis (EAE), a protocol widely used to mimic some aspects of MS by triggering an autoimmune response with injections of a myelin-derived peptides followed by an induction of inflammation by administration of pertussis toxin (Figure 4A). Cbx5-/- mice displayed an exacerbated reaction to that protocol, with an earlier onset of the symptoms followed by a rapid reaching of high grade EAE, beyond recovery (Figure 4B). The early onset and the exacerbated symptoms were also observed with heterozygote Cbx5+/- mice and these animals experience a more extensive loss of body weight than the wild-types during the protocol (Sup. Figure 4A and 4B). Unlike the Cbx5-/-, the Cbx5+/- mice eventually transitioned into a recovery phase, although with a one-day delay compared to the wild-types (Sup Figure 4A).

We next carried out RNA-seq on CD4+ T cells from either WT (n=2), Cbx5+/- (n=1), Cbx5-/- (n=3), EAE (n=3), and Cbx5-/- EAE (n=3). As anticipated from the Lo-CBX5 MS patients, inactivation of Cbx5 had an extensive impact on the transcriptome, with a total of 6478 differentially regulated genes (3224 up and 3248 down genes 2-fold or more, Adj pVal<0.05 – Sup. Figure 4C and 4D). Among the 3224 upregulated genes, 703 genes were also upregulated by EAE, corresponding to approximately half of the genes upregulated by EAE (Figure 4C). This observed intersection was approximately 6-fold higher than expected by chance (Sup Figure 4E). Pathway analysis on the 703 shared genes revealed that Cbx5-/- mice activated IL-17 and NFkappaB pathways at steady state, which may participate in the exacerbated response to the EAE protocol (Figure 4D). We noted also that the 703 genes upregulated both by Cbx5 inactivation and by EAE were enriched in Myc target genes (Figure 4E). This observation suggested a potential reliance of these genes on regulation through RNAPII pause-release mechanisms.

Gene ontology analysis of the 3224 genes upregulated following Cbx5 inactivation identified strong enrichment scores for transmembrane transport through the blood-brain barrier, underscoring the modified expression of numerous solute carrier (SLC) genes (Figure 4F). These genes encode membrane transport proteins involved in the movement of ions, nutrients, and metabolites and frequently associated with inflammation and autoimmunity disease (Sheng et al., 2022). We noted that among the 373 SLC genes annotated in the mouse genome, 117 displayed a modified expression (min. 2-fold, Adj. pVal<0.05), while the 77 upregulated SLC genes were enriched in amino acid transporters (Figure 4G). These observations suggest that SLC genes are particularly sensitive to Cbx5 regulation, and that the

modified expression of the SLC genes may promote T cell activation, heavily relying on amino acid transportation (Wang & Zou, 2020).

To mirror our study on the MS patients, we also examined the ferroptosis pathway. We here noted a strong down-regulation of Gpx4 (10-fold) and a corresponding upregulation of the ferroptosis markers Ptgs2 and Chac1 (Figure 4H). Consistent with a possibly increased cell death by ferroptosis, we noted also that down-regulated genes (down 2-fold or more, Adj pVal<0.05, 3248 genes) were enriched in genes involved in fatty acid beta oxidation (FAO – Sup. Figure 4F). A down-regulation of this pathway is anticipated to favor ferroptosis, as FAO normally consumes the fatty acids, leading to a reduction in the rate of lipid peroxidation. Analysis of the down-regulated genes with either Reactome, or BioPlanet databases revealed that down-regulated genes were also enriched in genes involved in cell cycle, DNA replication, and mitochondrial ATP synthesis, consistent with a subset of the cells possibly withdrawing from the cell cycle (Sup. Figure 4G and 4H).

The Cbx5-/- mouse model recapitulates the defect in Integrator activity.

Finally, we investigated whether Cbx5-/- mice recapitulated some of the transcriptional anomalies observed in Lo-CBX5 MS patients. Examination of the transcriptome revealed that, as observed for the Lo-CBX5 patients, regulators of snRNA/eRNA processing, and RNAPII pause-release were affected, with reduced expression of several INTcom subunits (Ints3, Ints4, Ints9, Ints10, and Ints11/Cpsf3l), and of the DSIF subunit Supt4h1/Spt4, while expression of Myc was increased (Figure 5A). The RNA-seq data also suggested that the functions of these regulators were affected. At U snRNA and histone genes, we noted that despite a decrease in overall expression, the read-count in the 3' regions remained at a consistent level. This relatively higher proportion of 3' transcription compared to the rest of the gene was strongly suggestive of a reduced 3’ end processing (see examples Figure 5B and Sup. Figure 5A). To probe for defects in the regulation of RNAPII pause-release, we plotted the average distribution of reads at a set of genes harboring a first intron of more than 20kb, thereby using the same approach as for the MS patient samples. For the 3 Cbx5-/- samples and the one Cbx5+/-, alike what we observed in the Lo-CBX5 patients, accumulation of reads was increased over the region downstream of the TSS, while unaffected or reduced over 3’ region of the genes (Figures 5C and 5D). The increased transcription at the TSS-proximal regions was also visualized when plotting the distribution over an average (meta) gene including both exons and introns (Figure 5E). In contrast, plotting the read distribution over exons only, revealed an overall reduced production of mature mRNA, consistent with the mouse model recapitulating a reduced efficiency of splicing (Figure 5F). The Ints6 gene provided a clear illustration of this transcriptional defect, with accumulation of intronic reads, associated with a reduced signal at exons, particularly toward the 3’ end of the gene (Figure 5G). Analysis of the data with the RMATS package confirmed an extensive impact of Cbx5 inactivation on alternative splicing, with, as observed in the Lo-CBX5 patients, an increase in exon-skipping events. Yet, consistent with an accumulation of non-maturated pre-mRNA species, the most favored type of alternative splicing was intron retention (Figure 5H). EAE alone did not induce accumulation of reads on the 5’ region of genes, but appeared to cause a global reduction in transcription (Sup. Figure 5B). Yet, accumulation of reads on the 5’ region upon Cbx5 inactivation was preserved in EAE (Sup. Figure 5C).

In conclusion, like in the LoCBX5 MS patients, inactivation of Cbx5 in the mouse resulted in reduced regulation of RNAPII promoter escape.

Discussion:

In this study, we generated RNA-seq data from a very heterogeneous group of patients, some with comorbidities. Principal component analysis (PCA) of the sequencing data revealed the presence of two patient groups, one mapping close to the symptomatic controls, and another group with significantly different overall gene expression. Neither comorbidity, age, nor gender were distinguishing features between the two groups. In contrast, ranking the patients according to their level of CBX5 expression allowed to better define the groups, the Hi-CBX5 (upper quartile) and the Lo-CBX5 patients (lower quartile) mapping clearly at each extremity of axis 1 of the PCA. The Hi-CBX5 group included a patient with a clinically isolated syndrome (CIS) and one of the 3 patients with a radiologically isolated syndrome (RIS). This group also exhibited diminished expression of genes controlled by vitamin D. The correlation between the level of serum vitamin D and disease activity still needs to be further investigated (Sintzel et al., 2018). Yet, our observation suggests that Hi-CBX5 patients might benefit from a diet enriched with this vitamin.

The Lo-CBX5 group did not included any of the patients with an isolated syndrome, and instead encompassed primary and secondary progressive stages, leading us to propose that this group may correspond to patients in phases of steady decline in neurological function. In this group of patients, we observed defects in U snRNA and enhancer-RNA maturation/degradation and in RNA polymerase II (RNAPII) pause-release, transcriptional anomalies that are all indicative of reduced Integrator activity. An implication of the Integrator complex in the MS pathology has hitherto not been investigated, although several indications have been pointing in that direction. Firstly, single nucleotide polymorphisms (SNPs) within the Integrator subunit INTS8 have previously been associated with an elevated risk of MS (Briggs & Sept, 2021). Another study has also detected association between MS and SNPs in genes associated with “abortive elongation”, including the NELF subunits NELFE and NELFCD, and the DSIF subunit SUPT4H1 (Misicka et al., 2021). Finally, patients with MS have been reported to show aberrant U snRNA polyadenylation, a known manifestation of a defective maturation of these RNA species (Spurlock et al., 2015). More generally, we note also that the integrator complex is a known player in neurological diseases, with mutations in INTS1 and INTS8 being associated with rare recessive human neurodevelopmental syndromes (Oegema et al., 2017).

Reduced Integrator activity affects expression of hundreds of protein-encoding genes independently of its impact on snRNA maturation. Typically, by promoting pause-release of RNAPII enzymes not competent for elongation, it reduces production of full-length pre-mRNAs. In parallel, the lack of regulation of RNAPII pausing upregulates basal transcription of (often short) inducible genes (Mendoza-Figueroa et al., 2020). Therefore, the chronic activation of short intron-less pro-inflammatory genes such as c-Jun and JunD may be directly related to this mechanism, consistent with up-regulated genes being in average shorter than down-regulated genes in the Lo-CBX5 patients.

Deregulated elongation of eRNAs, another characteristic of reduced Integrator activity, also appeared as a mechanism resulting in ectopic gene expression. The first illustration of this seems to be the increased transcription of HERVs. The promoter regions of a subset these sequences of viral origin have been coopted as enhancers. Increased elongation of the eRNAs thus mechanically results in transcription of downstream viral sequence. The impact of elongated eRNAs was also exemplified by the leptin gene that we found transcribed from an upstream enhancer bleeding into the LEP gene. This adipokine normally secreted by adipose tissues regulates systemic metabolism and appetite, while also signaling directly to immune

cells to promote inflammation (Kiernan & MacIver, 2021). It has been found to be present at increased levels in patients with MS, and has been associated with MS risk (Biström et al., 2021; Fahmi et al., 2021). This may directly illustrate how defective Integrator activity can participate in the pathogenesis of MS.

Reduced Integrator activity was the main transcriptional phenotype of the mouse T cells inactivated for Cbx5, suggesting that the encoded HP1α protein is directly involved in the regulation of RNAPII activity. This is in contrasts with previously documented heterochromatic activity of this protein. Yet, examination of the available ChIP-seq data revealed a localization of HP1α at the TSSs of genes. We note also that HP1α is known to dimerize with TRIM28/KAP1 previously reported to participate in the regulation of RNAPII elongation (Bunch et al., 2014). Finally, HP1 proteins copurifies with the CBC cap binding complex, consistent with a role in early phases of transcription initiation (Winczura et al., 2018).

The mechanism causing Cbx5 inactivation to promote EAE still needs to be explored at an immunological level to identify the involved cell populations. Yet, we note that the many genes up-regulated both by EAE and Cbx5 inactivation were enriched in Myc targets. The Myc transcription factor is indispensable for T cell activation, playing a role in the amino acid supply required for protein production associated with the increased cellular activity (Marchingo et al., 2020). The amino acid transporter Slc7a5, a key target of Myc, is also among the numerous SLC genes up-regulated by Cbx5 inactivation (> 4-fold, Adj. Pval=10^{-80}). This apparently exacerbated sensitivity of genes of the SLC family to deregulated RNAPII pause-release may possibly explain some aspects of the increased responsiveness of Cbx5 null mice to EAE.

While displaying numerous similarities, the transcriptional defects observed in Lo-CBX5 patients and in Cbx5 null mice were not entirely identical. In particular, in the mouse model, we noted a reduced expression of U snRNAs and a more pronounced disruption of pre-mRNA maturation. It is plausible that these two phenomena are related. The significant scarcity of U snRNAs in the mouse model could be a major factor contributing to the extensive intron retention. In contrast, in patients, where the deficit is milder, a relatively less affected splicing machinery may cause only increased exon skipping.

Whether reduced CBX5 activity is a trigger of MS symptoms remains an open question. We noted that in the Cbx5 KO mouse model, inactivation of just one Cbx5 allele was sufficient to obtain both the reduced Integrator activity and the exacerbated reaction to EAE. Nevertheless, the mice did not manifest EAE symptoms spontaneously, suggesting that Cbx5 inactivation generates a conducive terrain, but that the neuroinflammatory symptoms necessitate an external trigger to develop. Although our data do not offer insights into the potential nature of this trigger, they may provide valuable information regarding subsequent transcriptional events. Indeed, several observations suggest that reduced Integrator activity is a self-amplifying mechanism. First, mutations in INTS1 and INTS8 were previously shown to result in the down-regulation of multiple subunits of the INTcom (Oegema et al., 2017). A similar phenomenon was observed upon Cbx5 inactivation in our mouse model. The reduced expression of these INTcom genes may be due to a sensitivity of these genes to their own activity, as we documented for the INTS6 gene in both human and mouse. This suggests that external stimuli hitting on RNAPII pause-release may trigger a feed-forward loop resulting in reduced INTcom activity. In this context, we note that the Epstein Barr virus (EBV) was found associated to Multiple Sclerosis (MS), first described 30 years ago (Haahr & Höllsberg, 2006) and has continued to show evidence as a causal role in MS (Soldan & Lieberman, 2022). This virus is known to target INTS6 (Lei et al., 2013), while also being a promoter of ferroptosis (Burton et al., 2022). Together with the clear evidence for ferroptosis in the Cbx5-/- mice, this

may suggest that ferroptosis is the normal fate of lymphocytes experiencing a defective Integrator activity. The seemingly abortive ferroptosis in the MS patients may further suggest that this defense mechanism is not entirely functional in these patients, possibly participating in the exacerbated neuroinflammation.

MATERIAL AND METHODS

Ethics statement

The study was conducted in accordance with the Ethical Declaration of Helsinki and all patients gave written, informed consent. The study and the material for informed consent were approved by The Central Denmark Region Committee on Biomedical Research Ethics (journal number: 1-10-72-334-15).

Patients and controls

Patients admitted to the MS clinic, Department of Neurology, Aarhus University Hospital, were consecutively included from January 2017 to January 2018. A full diagnostic workup included medical history, clinical examination, MRI of the entire neural axis, cerebrospinal fluid (CSF) analyses (cells, protein, IgG index, oligoclonal bands) and evoked potentials as recommended (Thompson et al., 2018). CSF and MRI examination were evaluated according to the revised MacDonald criteria from 2017 (Thompson et al., 2018) and an Extended Disability Status Scale score was assessed according to Kurtzke (Kurtzke, 1983). Patients were excluded if they had other neurologic disease or received glucocorticoids within the month preceding sampling. Total number of MRI white matter lesions were registered by fluid-attenuated inversion recovery sequences on MRI. Demographics and paraclinical findings of patients with CIS, RRMS, PPMS, RIS and SCs are summarized in Table 1. Patients included as SC have neurological symptoms, but have no objective clinical or paraclinical findings to define a specific neurological disease. This specific definition is described in detail by Teunissen et al (Teunissen et al., 2013) and they do not represent early MS.

Patients included as RIS have no neurological symptoms and are only referred to diagnostic workup owing to the presence of incidental white matter lesions in MRI suggestive of MS. Diagnostic criteria for RIS were proposed in 2009 and include the number, shape and location of the brain lesions (Thompson et al., 2018). Lesions are ovoid and well circumscribed with a size greater than 3 mm, show dissemination in space and can be juxtaposed to the corpus callosum. Lesions should not follow a vascular distribution and do not account for any other pathologic processes.

Data download

We downloaded the Cbx5 ChIP-seq data acquired in hepatocarcinoma-derived HepG2 cells from the ENCODE portal, output type was “signal p-value” in bigwig format, file ENCFF408KOE (https://www.encodeproject.org/)(Sloan et al., 2016).

RNA sequencing

Monocytes were obtained from patient PBMCs using a Pan Monocyte Isolation Kit (Miltenyi Biotech, ref. 130-096-537) following the kit protocol, allowing for negative selection of unstimulated monocytes. CD4+ mouse T cells were purified from spleenocytes using the Miltenyi kit 130-104-454. For both cell types, total RNA was extracted by Trizol LS (Thermo Fisher Scientific, ref. 10296028), according to the manufacturer's protocol. Total RNA library preparation and sequencing were performed by Novogene Co., Ltd, as a lncRNA sequencing service, including lncRNA directional library preparation with rRNA depletion (Ribo-Zero Magnetic Kit), quantitation, pooling and PE 150 sequencing (30G raw data) on Illumina HiSeq 2500 platform. Filtering and trimming of the RNA seq data left around 230-300 million reads pairs /sample.Mapping was carried out with STAR (v2.6.0b) (parameters: --

outFilterMismatchNmax 1 --outSAMmultNmax 1 --outMultimapperOrder Random --outFilterMultimapNmax 30) (Dobin et al., 2013). The reference genomes were hg19 homo sapiens and mm10 mus musculus primary assemblies from Ensembl. The SAM files were converted to BAM files and sorted by coordinate using samtools (v1.7) (Danecek et al., 2021).

BigWig files, heatmaps, and profiles

Bigwigs files were generated from .bam files with bamCoverage (parameter: --normalizeUsing CPM) from Deeptools (v3.1.3) (Ramírez et al., 2016). Heatmaps and profiles were also generated with Deeptools (v3.1.3). Matrices were generated with computeMatrix followed by plotProfile or plotHeatmap as appropriate.

Data quantification

Read quantification was carried out with featureCounts (v1.6.1) from the Subread suite (Liao et al., 2014). For repeated element, files were obtained by extracting in .bed file format, entries annotated “SINE”, “LINE” or “LTR” in the “RepClass” field from RepeatMasker.

Analysis of splicing

Differential splicing analysis was done using rMATS (v4.1.0) (Shen et al., 2014) with the parameters: --libType fr-firststrand –novelSS. The occurrence of each type of splicing event was then counted.

Pathway analysis and data visualization

Pathway analysis was carried out with Enrichr from the Ma’ayan lab (Chen et al., 2013; Xie et al., 2021). The Integrative Genomics Viewer software (IGV) was used to examine specific loci (Robinson et al., 2011).

Cbx5 null mouse model

C57BL/6N-Atm1Brd Cbx5tm1a(EUCOMM)Wtsi/WtsiOrl strain inactivated for Cbx5 was received from the EUCOMM Consortium in the context of a Standard Material Transfer Agreement. It was transferred via the TAAM – CNRS Orléans for breeding at the Institut Pasteur - Paris. The experiments involving the mouse model were carried out in strict accordance with the approved research protocols and in compliance with all applicable institutional and regulatory guidelines (CETEA Institut Pasteur approval dap170030).

Experimental Autoimmune Encephalomyelitis (EAE)

EAE was carried out as previously described (Porozhan et al., 2012). Briefly, to prepare the MOG35-55 emulsion for immunization, we first calculated the total volume required and then prepared 1.5 to 2 times that amount to account for potential losses during the process. Each mouse received a subcutaneous injection of 200 µl of a mixture containing MOG35-55 peptide solution and Complete Freund's Adjuvant (CFA) in a 1:1 ratio. The MOG35-55 peptide solution was prepared by diluting lyophilized 200 µg of the peptide per mouse with ddH2O to achieve a final concentration of 2 mg/mL, which was then stored at -20°C. CFA was prepared by grinding 100 mg of desiccated Mycobacterium tuberculosis H37RA and adding 10 mL of Incomplete Freund's Adjuvant (IFA) to create a 10 mg/mL CFA stock solution that could be stored at 4°C. Just prior to immunization, CFA was diluted with IFA to reach a final concentration of 2 mg/mL, with thorough mixing. The MOG35-55 and CFA were then mixed in a 1:1 ratio until the final concentration of 1 mg/mL was achieved. This emulsion, a critical

step for immunization, was carefully prepared to ensure proper emulsification. After verification of its stability, the emulsion was drawn into syringes for subsequent use. In addition to the MOG35-55 emulsion, pertussis toxin was prepared and administered. For each mouse, 400 ng of pertussis toxin in 200 µl of PBS was injected intraperitoneally on the day of immunization and repeated two days later. Pertussis toxin was reconstituted by dissolving 50 µg in 500 µl of ddH2O to create a 100 µg/mL stock solution, which was stored at 4°C. To achieve the required concentration for immunization, the stock solution was diluted 1:50 with PBS. Finally, to ensure proper identification and monitoring, individual mice were marked, typically by using color markings on the tail base. A second dose of pertussis toxin was administered on day two after the initial immunization to complete the immunization process.

Blood brain barrier permeability

In the conducted experiments, mice were administered Fluorescein-cadaverine via tail vein injections at a dosage of 6 µg/g. Following a 2-hour waiting period, the mice were anesthetized, and their brains were perfused with ice-cold PBS for 7 minutes. The cortex was subsequently isolated and weighed, and cortical tissues were homogenized using a 1% Triton X-100 in PBS solution. After centrifugation at 12,000 × g for 20 minutes at 4 °C, the resulting supernatant was used to measure fluorescence with specific excitation and emission wavelengths. Importantly, to ensure the accuracy of the fluorescence measurements, the obtained values were normalized by dividing them by the respective cortical weights. This meticulous methodology allowed for precise and standardized assessment of tracer leakage and fluorescence in these experiments, contributing to the reliability of the results.

REFERENCES

Afrasiabi, A., Fewings, N. L., Schibeci, S. D., Keane, J. T., Booth, D. R., Parnell, G. P., & Swaminathan, S. (2021). The interaction of human and epstein–barr virus mirnas with multiple sclerosis risk loci. *International Journal of Molecular Sciences*, *22*(6), 1–15. https://doi.org/10.3390/ijms22062927

Biström, M., Hultdin, J., Andersen, O., Alonso-Magdalena, L., Jons, D., Gunnarsson, M., Vrethem, M., & Sundström, P. (2021). Leptin levels are associated with multiple sclerosis risk. *Multiple Sclerosis Journal*, *27*(1), 19–27. https://doi.org/10.1177/1352458520905033

Briggs, F. B. S., & Sept, C. (2021). Mining Complex Genetic Patterns Conferring Multiple Sclerosis Risk. *Int. J. Environ. Res. Public Health*, *18*, 2518. https://doi.org/10.3390/ijerph18052518

Bunch, H., Zheng, X., Burkholder, A., Dillon, S. T., Motola, S., Birrane, G., Ebmeier, C. C., Levine, S., Fargo, D., Hu, G., Taatjes, D. J., & Calderwood, S. K. (2014). TRIM28 regulates RNA polymerase II promoter-proximal pausing and pause release. *Nature Structural and Molecular Biology*, *21*(10), 876–883. https://doi.org/10.1038/nsmb.2878

Burton, E. M., Voyer, J., & Gewurz, B. E. (2022). Epstein–Barr virus latency programs dynamically sensitize B cells to ferroptosis. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(11), 1–12. https://doi.org/10.1073/pnas.2118300119

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*, 128. https://doi.org/10.1186/1471-2105-14-128

Christensen, T., & Muchardt, C. (2019). Chapter 5 - The epigenetics of multiple sclerosis. In O. B. T.-C. S. and N. D. Binda (Ed.), *Translational Epigenetics* (Vol. 12, pp. 97–118). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-813796-3.00005-5

Clarke, J. P., Thibault, P. A., Salapa, H. E., & Levin, M. C. (2021). A Comprehensive Analysis of the Role of hnRNP A1 Function and Dysfunction in the Pathogenesis of Neurodegenerative Disease. *Frontiers in Molecular Biosciences*, *8*(April), 1–19. https://doi.org/10.3389/fmolb.2021.659610

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), 1–4. https://doi.org/10.1093/gigascience/giab008

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Elkjaer, M. L., Röttger, R., Baumbach, J., & Illes, Z. (2022). A Systematic Review of Tissue and Single Cell Transcriptome/Proteome Studies of the Brain in Multiple Sclerosis. *Frontiers in Immunology*, *13*(March), 1–25. https://doi.org/10.3389/fimmu.2022.761225

Evsyukova, I., Somarelli, J. A., Gregory, S. G., & Garcia-Blanco, M. A. (2010). Alternative splicing in multiple sclerosis and other autoimmune diseases. *RNA Biology*, *7*(4), 462–473. https://doi.org/10.4161/rna.7.4.12301

Fahmi, R. M., Kamel, A. E., Elsayed, D. A., Zidan, A. A., & Sarhan, N. T. (2021). Serum levels of leptin and adiponectin in patients with multiple sclerosis. *Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, *57*(1). https://doi.org/10.1186/s41983-021-00369-2

Filipi, M., & Jack, S. (2020). Interferons in the treatment of multiple sclerosis: A clinical

efficacy, safety, and tolerability update. *International Journal of MS Care*, *22*(4), 165–172. https://doi.org/10.7224/1537-2073.2018-063

Gilchrist, D. A., Nechaev, S., Lee, C., Ghosh, S. K. B., Collins, J. B., Li, L., Gilmour, D. S., & Adelman, K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes & Development*, *22*(14), 1921–1933. https://doi.org/10.1101/GAD.1643208

Haahr, S., & Höllsberg, P. (2006). Multiple sclerosis is linked to Epstein-Barr virus infection. *Reviews in Medical Virology*, *16*(5), 297–310. https://doi.org/10.1002/RMV.503

Hemmer, B., Kerschensteiner, M., & Korn, T. (2015). Role of the innate and adaptive immune responses in the course of multiple sclerosis. *The Lancet Neurology*, *14*(4), 406–419. https://doi.org/10.1016/S1474-4422(14)70305-9

Keller, C., Adaixo, R., Stunnenberg, R., Woolcock, K. J., Hiller, S., & Bühler, M. (2012). HP1 Swi6 Mediates the Recognition and Destruction of Heterochromatic RNA Transcripts. *Molecular Cell*, *47*(2), 215–227. https://doi.org/10.1016/j.molcel.2012.05.009

Kiernan, K., & MacIver, N. J. (2021). The Role of the Adipokine Leptin in Immune Cell Function in Health and Disease. In *Frontiers in Immunology* (Vol. 11). https://doi.org/10.3389/fimmu.2020.622468

Kirstein, N., Gomes Dos Santos, H., Blumenthal, E., & Shiekhattar, R. (2021). The Integrator complex at the crossroad of coding and noncoding RNA. *Current Opinion in Cell Biology*, *70*, 37–43. https://doi.org/10.1016/J.CEB.2020.11.003

Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, *33*(11), 1444–1452. https://doi.org/10.1212/wnl.33.11.1444

Lai, F., Gardini, A., Zhang, A., & Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature*, *525*(7569), 399–403. https://doi.org/10.1038/NATURE14906

Lei, T., Yuen, K. S., Xu, R., Tsao, S. W., Chen, H., Li, M., Kok, K. H., & Jin, D. Y. (2013). Targeting of DICE1 tumor suppressor by Epstein–Barr virus-encoded miR-BART3* microRNA in nasopharyngeal carcinoma. *International Journal of Cancer*, *133*(1), 79–87. https://doi.org/10.1002/IJC.28007

Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lublin, F. D., Reingold, S. C., Cohen, J. A., Cutter, G. R., Sørensen, P. S., Thompson, A. J., Wolinsky, J. S., Balcer, L. J., Banwell, B., Barkhof, F., Bebo, B., Calabresi, P. A., Clanet, M., Comi, G., Fox, R. J., Freedman, M. S., Goodman, A. D., Inglese, M., Kappos, L., … Polman, C. H. (2014). Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology*, *83*(3), 278–286. https://doi.org/10.1212/WNL.0000000000000560

Luoqian, J., Yang, W., Ding, X., Tuo, Q. zhang, Xiang, Z., Zheng, Z., Guo, Y. jie, Li, L., Guan, P., Ayton, S., Dong, B., Zhang, H., Hu, H., & Lei, P. (2022). Ferroptosis promotes T-cell activation-induced neurodegeneration in multiple sclerosis. *Cellular and Molecular Immunology*, *19*(8), 913–924. https://doi.org/10.1038/s41423-022-00883-0

Maison, C., Bailly, D., Peters, A. H. F. M., Quivy, J. P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., & Almouzni, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nature Genetics*, *30*(3), 329–334. https://doi.org/10.1038/ng843

Maison, C., Bailly, D., Roche, D., De Oca, R. M., Probst, A. V., Vassias, I., Dingli, F., Lombard,

B., Loew, D., Quivy, J. P., & Almouzni, G. (2011). SUMOylation promotes de novo targeting of HP1 alpha to pericentric heterochromatin. *Nature Genetics*, *43*(3), 220–227. https://doi.org/10.1038/ng.765

Marchingo, J. M., Sinclair, L. V., Howden, A. J., & Cantrell, D. A. (2020). Quantitative analysis of how myc controls t cell proteomes and metabolic pathways during t cell activation. *ELife*, *9*, 1–23. https://doi.org/10.7554/eLife.53725

Mata-Garrido, J., Xiang, Y., Chang-Marchand, Y., Reisacher, C., Ageron, E., Guerrera, I. C., Casafont, I., Bruneau, A., Cherbuy, C., Treton, X., Dumay, A., Ogier-Denis, E., Batsché, E., Costallat, M., Revêchon, G., Eriksson, M., Muchardt, C., & Arbibe, L. (2022). The Heterochromatin protein 1 is a regulator in RNA splicing precision deficient in ulcerative colitis. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-34556-3

Mendoza-Figueroa, M. S., Tatomer, D. C., & Wilusz, J. E. (2020). The Integrator Complex in Transcription and Development. *Trends in Biochemical Sciences*, *45*(11), 923–934. https://doi.org/10.1016/j.tibs.2020.07.004

Misicka, E., Davis, M. F., Kim, W., Brugger, S. W., Beales, J., Loomis, S., Bronson, P. G., & Briggs, F. B. S. (2021). A higher burden of multiple sclerosis genetic risk confers an earlier onset. *Https://Doi.Org/10.1177/13524585211053155*, *28*(8), 1189–1197. https://doi.org/10.1177/13524585211053155

Mowry, K. L., & Steitz, J. A. (1987). Identification of the human U7 snRNP as one of several factors involved in the 3' end maturation of histone premessenger RNA's. *Science (New York, N.Y.)*, *238*(4834), 1682–1687. https://doi.org/10.1126/SCIENCE.2825355

Muchardt, C., Guillemé, M., Seeler, J. S., Trouche, D., Dejean, A., & Yaniv, M. (2002). Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1α. *EMBO Reports*, *3*(10), 975–981. https://doi.org/10.1093/embo-reports/kvf194

Oegema, R., Baillat, D., Schot, R., van Unen, L. M., Brooks, A., Kheradmand Kia, S., Jeannette Hoogeboom, A. M., Xia, Z., Li, W., Cesaroni, M., Lequin, M. H., van Slegtenhorst, M., Dobyns, W. B., M de Coo, I. F., Verheijen, F. W., Kremer, A., van der Spek, P. J., Heijsman, D., Wagner, E. J., … S Mancini, G. M. (2017). *Human mutations in integrator complex subunits link transcriptome integrity to brain development*. https://doi.org/10.1371/journal.pgen.1006809

Porozhan, E. A., Gol'tsev, A. N., Babenko, N. N., & Dubrava, T. G. (2012). Level of Sensibilization of Immunocompetent Cells as a Criterion of Efficacy of Introduction of Cryopreserved Fetal Nerve Cells in Experimental Allergic Encephalomyelitis. *Neurophysiology*, *44*(4).

Rachez, C., Legendre, R., Costallat, M., Varet, H., Yi, J., Kornobis, E., & Muchardt, C. (2021). HP1γ binding pre-mRNA intronic repeats modulates RNA splicing decisions. *EMBO Reports*, *22*(9), 1–17. https://doi.org/10.15252/embr.202052320

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165. https://doi.org/10.1093/nar/gkw257

Rhead, B., Shao, X., Graves, J. S., Chitnis, T., Waldman, A. T., Lotze, T., Schreiner, T., Belman, A., Krupp, L., Greenberg, B. M., Weinstock–Guttman, B., Aaen, G., Tillema, J. M., Rodriguez, M., Hart, J., Caillier, S., Ness, J., Harris, Y., Rubin, J., … Barcellos, L. F. (2019). miRNA contributions to pediatric-onset multiple sclerosis inferred from GWAS. *Annals*

of Clinical and Translational Neurology, *6*(6), 1053–1061. https://doi.org/10.1002/acn3.786

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. In *Nature Biotechnology* (Vol. 29, Issue 1, pp. 24–26). https://doi.org/10.1038/nbt.1754

Saint-André, V., Batsché, E., Rachez, C., & Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1γ favor inclusion of alternative exons. *Nature Structural & Molecular Biology*, *18*. https://doi.org/10.1038/nsmb.1995

Sharma, P., Azebi, S., England, P., Christensen, T., Møller-Larsen, A., Petersen, T., Batsché, E., & Muchardt, C. (2012). Citrullination of Histone H3 Interferes with HP1-Mediated Transcriptional Repression. *PLoS Genetics*, *8*(9), 1–15. https://doi.org/10.1371/journal.pgen.1002934

Shen, S., Park, J. W., Lu, Z. X., Lin, L., Henry, M. D., Wu, Y. N., Zhou, Q., & Xing, Y. (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(51), E5593–E5601. https://doi.org/10.1073/PNAS.1419161111/-/DCSUPPLEMENTAL/PNAS.201419161SI.PDF

Sheng, L., Luo, Q., & Chen, L. (2022). Special Section on New Era of Transporter Science: Unraveling the Functional Role of Orphan Transporters-Minireview Amino Acid Solute Carrier Transporters in Inflammation and Autoimmunity. In *Drug Metabolism and Disposition* (Vol. 50, Issue 9, pp. 1228–1237). American Society for Pharmacology and Experimental Therapeutics. https://doi.org/10.1124/dmd.121.000705

Sintzel, M. B., Rametta, M., & Reder, A. T. (2018). Vitamin D and Multiple Sclerosis: A Comprehensive Review. *Neurology and Therapy*, *7*(1), 59–85. https://doi.org/10.1007/S40120-017-0086-4/TABLES/3

Sloan, C. A., Chan, E. T., Davidson, J. M., Malladi, V. S., Strattan, J. S., Hitz, B. C., Gabdank, I., Narayanan, A. K., Ho, M., Lee, B. T., Rowe, L. D., Dreszer, T. R., Roe, G., Podduturi, N. R., Tanaka, F., Hong, E. L., & Cherry, J. M. (2016). ENCODE data at the ENCODE portal. *Nucleic Acids Research*, *44*(D1). https://doi.org/10.1093/nar/gkv1160

Smallwood, A., Hon, G. C., Jin, F., Henry, R. E., Espinosa, J. M., & Ren, B. (2012). CBX3 regulates efficient RNA processing genome-wide. *Genome Research*, *22*(8), 1426–1436. https://doi.org/10.1101/GR.124818.111

Soldan, S. S., & Lieberman, P. M. (2022). Epstein–Barr virus and multiple sclerosis. *Nature Reviews Microbiology 2022 21:1*, *21*(1), 51–64. https://doi.org/10.1038/s41579-022-00770-5

Spurlock, C. F., Tossberg, J. T., Guo, Y., Sriram, S., Crooke, P. S., & Aune, T. M. (2015). Defective structural RNA processing in relapsing-remitting multiple sclerosis. *Genome Biology*, *16*(1), 1–13. https://doi.org/10.1186/s13059-015-0629-x

Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., Raffel, R., Sobhian, B., Severac, D., Rialle, S., Parrinello, H., Cuvier, O., & Benkirane, M. (2014). Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nature Communications*, *5*. https://doi.org/10.1038/NCOMMS6531

Stein, C. B., Field, A. R., Mimoso, C. A., Zhao, C. C., Huang, K. L., Wagner, E. J., & Adelman, K. (2022). Integrator endonuclease drives promoter-proximal termination at all RNA polymerase II-transcribed loci. *Molecular Cell*, *82*(22), 4232-4245.e11. https://doi.org/10.1016/j.molcel.2022.10.004

Teunissen, C., Menge, T., Altintas, A., Álvarez-Cermeño, J. C., Bertolotto, A., Berven, F. S.,

Brundin, L., Comabella, M., Degn, M., Deisenhammer, F., Fazekas, F., Franciotta, D., Frederiksen, J. L., Galimberti, D., Gnanapavan, S., Hegen, H., Hemmer, B., Hintzen, R., Hughes, S., … Bartos, A. (2013). Consensus definitions and application guidelines for control groups in cerebrospinal fluid biomarker studies in multiple sclerosis. *Http://Dx.Doi.Org/10.1177/1352458513488232*, *19*(13), 1802–1809. https://doi.org/10.1177/1352458513488232

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., … Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. In *The Lancet Neurology* (Vol. 17, Issue 2, pp. 162–173). Elsevier. https://doi.org/10.1016/S1474-4422(17)30470-2

Wang, W., & Zou, W. (2020). Amino Acids and Their Transporters in T Cell Immunity and Cancer Therapy. *Molecular Cell*, *80*(3), 384–395. https://doi.org/10.1016/j.molcel.2020.09.006

Winczura, K., Schmid, M., Iasillo, C., Molloy, K. R., Harder, L. M., Andersen, J. S., LaCava, J., & Jensen, T. H. (2018). Characterizing ZC3H18, a Multi-domain Protein at the Interface of RNA Production and Destruction Decisions. *Cell Reports*, *22*(1), 44–58. https://doi.org/10.1016/j.celrep.2017.12.037

Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., Lachmann, A., Wojciechowicz, M. L., Kropiwnicki, E., Jagodnik, K. M., Jeon, M., & Ma'ayan, A. (2021). Gene Set Knowledge Discovery with Enrichr. *Current Protocols*, *1*(3), e90. https://doi.org/10.1002/cpz1.90

Yamamoto, J., Hagiwara, Y., Chiba, K., Isobe, T., Narita, T., Handa, H., & Yamaguchi, Y. (2014). DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nature Communications 2014 5:1*, *5*(1), 1–10. https://doi.org/10.1038/ncomms5263

Yearim, A., Gelfman, S., Shayevitch, R., Melcer, S., Glaich, O., Mallm, J. P., Nissim-Rafinia, M., Cohen, A. H. S., Rippe, K., Meshorer, E., & Ast, G. (2015). HP1 Is Involved in Regulating the Global Impact of DNA Methylation on Alternative Splicing. *Cell Reports*, *10*(7), 1122–1134. https://doi.org/10.1016/J.CELREP.2015.01.038

FIGURE LEGENDS

Figure 1 : Extensive transcriptional deregulation in monocytes from patients with reduced expression of CBX5

(A) List of patients with MS with indications on the stage of the disease and eventual comorbidities. Symptomatic controls (SC) are also indicated. Donors are ranked according to levels of CBX5 expression after normalization by DeSeq2. Patients highlighted in red express CBX5 at high levels (upper quartile), while patients indicated in green express this gene at low levels (lower quartile). (B) PCA analysis using the 500 most variable genes. (C) Graphic representation of the CBX5-HNRNPA1 locus using the IGV genome browser. Bar charts represent the normalized expression levels of the two genes in the indicated patients. The 2 top lanes use a scale suited for the visualization of CBX5, while the 2 following lanes are at a scale fitted for HNRNPA1. RefSeq allows visualization of splice variants of the two genes annotated in Hg19 version of the human genome. (D) Within each group of patients defined by either high, low, or intermediate CBX5 levels, the percentage representation of each stage of the MS disease is indicated. (E-F) Vulcano plots illustrating differential gene expression between Hi-CBX5 patients and SCs (E), or between Lo-CBX5 patients and SCs (F). (H) Histograms representing the fold-activation of the indicated genes between SCs and Lo-CBX5 patients. (G) Schematic representation of the KEGG ferroptosis pathway. Genes up- or down-regulated between SC and Lo-CBX5 are indicated in red or blue respectively. (I) Analysis of genes down-regulated between SC and Lo-CBX5 patients using the Reactome pathway databank revealing functional enrichment.

Sup. Figure 1 : Extensive transcriptional deregulation in monocytes from patients with reduced expression of CBX5

(A) Analysis of genes up-regulated between SC and Hi-CBX5 patients using the ChEA (ChIP-Seq Enrichment Analysis) database providing information on transcription factors likely to regulate these genes. (B) Analysis of genes up-regulated between SC and Lo-CBX5 patients using the KEGG database providing information about the molecular functions and biological pathways of genes. (C) Bar graph reporting variation in expression of the indicated genes based on RNA-seq data in from SCs and Lo-Cbx5 patients. Levels in SCs set at 1. *: Adj. pVal<0.05, **: Adj. pVal<0.01, ***: Adj. pVal<0.001. (D) Analysis of genes up-regulated between SC and Lo-CBX5 patients using the WikiPathway database. (E) Analysis of genes down-regulated between SC and Lo-CBX5 using the Reactome, BioPlanet, or WikiPathway databases as indicated.

Figure 2: Evidence for a defective Integrator activity in Lo-CBX5 patients

(A) Analysis of genes down-regulated between SC and Lo-CBX5 patients using the ENCODE TF ChIP-seq data resource providing insights into the binding patterns of various transcription factors. (B) Bar graph reporting variation in expression of the indicated genes based on RNA-seq data in from SCs and Lo-Cbx5 patients. Levels in SCs set at 1. *: Adj. pVal<0.05, **: Adj. pVal<0.01, ***: Adj. pVal<0.001. (C-D) Screen shots from the IGV genome browser representing expression of the U3 snRNA copy in the TEX14 gene and of histone H1-3 (HIST1H1D) in SCs and Lo-CBX5 patients as indicated. (E) Quantification of reads mapping inside regions annotated as enhancers in monocytes by the Epigenome Roadmap. (F-G) Screen shots from the IGV genome browser representing expression of enhancers producing eRNAs either covering several repeated sequences, including a HERV (F), or resulting in transcription of the entire LEP gene (G).

Sup. Figure 2: Evidence for a defective Integrator activity in Lo-CBX5 patients
(A) Screen shots from the IGV genome browser representing expression of the U2 snRNA copy upstream of the WDR74 gene. (B-E) Quantification of reads mapping inside regions annotated as SINEs, LINEs, or LTRs in ReapeatMasker (B-D), or both as LTRs in ReapeatMasker and as enhancers in monocytes by the Epigenome Roadmap (E).

Figure 3: Defective Integrator activity results in extensive gene deregulation
(A-D) Average distribution of RNA-seq reads at genes containing an initial intron exceeding 20 kilobases in length. Profiles are either anchored on the Transcription Start Site (TSS in A), on the Transcription End Site (TES in B), or are plotted over the entire metagene either including (C), or excluding (D) intronic sequences. SC: symptomatic controls, Hi-CBX5: patients in the upper quartile for CBX5 expression levels, Lo-CBX5: patients in the lower quartile for CBX5 expression levels. (E-H) Screen shots from the IGV genome browser representing the impact of deregulated RNAPII pause-release at the indicated genes. (I) Average distribution of junction reads from the RNA-seq data across a metagene as in (C). (J-K) The RNA-seq was analyzed with the Rmats to quantify alternative splicing occurring between SC and Lo-CBX5 patients (J), or between SC and Hi-CBX5 patients (K). Histograms indicate number of significant events (Pval<0.05). A3SS: alternative 3' splice site, A5SS: alternative 5' splice site, SE: skipped exon, MXE: mutually exclusive exons, RI: retained introns. (L-M) Screen shots from the IGV genome browser representing at the indicated genes, splicing events either upregulated (green), or down-regulated (red) in the Lo-CBX5 (top track) or in the Hi-CBX5 (bottom track) when compared to SCs (Pval<0.05).

Sup. Figure 3: Defective Integrator activity results in extensive gene deregulation
(A) Heatmap of HP1a/CBX5 ChIP-seq signal in HepG2 cells (SRR14105564). The heatmap is anchored on the Transcription Start Site of genes with an exon 1 longer than 20kb from the Hg19 version of the genome. (B) Bar graph representing the average size in nucleotides of the genes either upregulated (1395 genes), or down-regulated (1350 genes) in Lo-CBX5 patients compared to SCs. (C) Screen shots from the IGV genome browser representing the impact of deregulated RNAPII pause-release at the SLC7A11 gene. (D) Screen shots from the IGV genome browser representing at the GAPDH gene, splicing events either upregulated (green), or down-regulated (red) in the Lo-CBX5 (top track) or in the Hi-CBX5 (bottom track) when compared to SCs (Pval<0.05).

Figure 4: Inactivation of Cbx5 in the mouse exacerbates EAE.
(A) Schematic representation of the EAE protocol. (B) Graphic representation of the clinical scores observed for the mice included in the EAE protocol. 0: No clinical symptoms. 1: Flaccid tail (loss of tail tone). 2: Hind limb weakness or partial paralysis. 3: Complete hind limb paralysis. 4: Complete hind limb paralysis and partial forelimb paralysis. 5: Moribund or dead. Indicated data are averaged from 5 females, 10 weeks of age. (C) Venn diagram representing the number of genes up-regulated either in Cbx5 null mice compared to WT, in mice exposed to an EAE protocol compared to controls, or in both conditions. (D-E) Analysis of genes up-regulated both upon Cbx5 inactivation and by the EAE protocol using the KEGG database providing information about the molecular functions and biological pathways of genes (D), and the ENCODE TF ChIP-seq data resource providing insights into the binding patterns of various transcription factors (E). (F) Analysis of genes up-regulated upon Cbx5 inactivation using the "Biological Process" category in the GO database focusing on describing biological

events, pathways, and processes. (G) Analysis of the 77 SLC genes up-regulated (Min. 2-fold, Adj pVal<0.05) upon Cbx5 inactivation using the "Molecular Function" category in the GO database focusing the molecular activities of gene products. (H) Bar graph reporting variation in expression of the indicated genes based on RNA-seq data in from WT and Cbx5-/- mice. Levels in WTs set at 1. *: Adj. pVal<0.05, **: Adj. pVal<0.01, ***: Adj. pVal<0.001.

Sup. Figure 4: Inactivation of Cbx5 in the mouse exacerbates EAE.
(A) Graphic representation of the clinical scores observed for the mice included in the EAE protocol either WT or Cbx5+/-. 0: No clinical symptoms. 1: Flaccid tail (loss of tail tone). 2: Hind limb weakness or partial paralysis. 3: Complete hind limb paralysis. 4: Complete hind limb paralysis and partial forelimb paralysis. 5: Moribund or dead. Indicated data are averaged from 10 females, 10 weeks of age. (B) Graphic representation of the body weight of either WT or Cbx5+/- mice included in the EAE experiment. Indicated data are averaged from 10 females, 10 weeks of age. (C-D) Vulcano plots illustrating differential gene expression between Cbx5-/- and WT mice either at steady state (C), or involved in an EAE protocol (D). (E) Intersection between genes upregulated upon Cbx5 inactivation and the 1423 genes upregulated by EAE was compared to the intersection between genes upregulated upon Cbx5 inactivation and a list of 1423 randomly selected genes (average of 10 iterations). (F) Analysis of genes down-regulated in Cbx5-/- mice compared to WT using the Go Biological Process, Reactome, or BioPlanet databases as indicated.

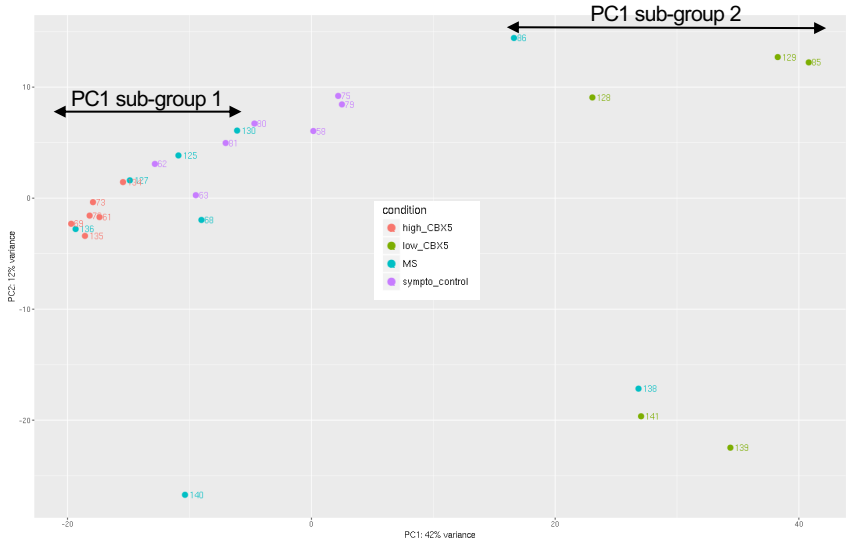
Figure 5: The mouse model confirms the impact of Cbx5 on Integrator activity
(A) Bar graph reporting variation in expression of the indicated genes based on RNA-seq data in from WT and Cbx5-/- mice. **: Adj. pVal<0.01, ***: Adj. pVal<0.001. (B) Screen shots from the IGV genome browser representing expression of at a series of U2 snRNA copies in WT, Cbx5+/-, and Cbx5-/- mice as indicated. Schematic shows two different scales to allow visualization of the reduced expression of the U2 snRNAs (top tracks) and of the downstream reads accumulating at similar levels in all samples (bottom tracks). (C-F) Average distribution of RNA-seq reads at genes containing an initial intron exceeding 20 kilobases in length. Profiles are either anchored on the Transcription Start Site (TSS in C), on the Transcription End Site (TES in D), or are plotted over the entire metagene either including (E), or excluding (F) intronic sequences. (G) Screen shots from the IGV genome browser representing the impact of deregulated RNAPII pause-release at the Ints6 gene. (H) Histograms indicate number of significant events (Pval<0.05). A3SS: alternative 3' splice site, A5SS: alternative 5' splice site, SE: skipped exon, MXE: mutually exclusive exons, RI: retained introns.

Sup. Figure 5: The mouse model confirms the impact of Cbx5 on Integrator activity
(A) Screen shots from the IGV genome browser representing expression of histone gene H2ac25 in WT, Cbx5+/-, and Cbx5-/- mice as indicated. Schematic shows two different scales to allow visualization of the reduced expression of the histone (top tracks), and of the downstream reads accumulating at similar levels in all samples (bottom tracks). (B-C) Average distribution of RNA-seq reads at genes containing an initial intron exceeding 20 kilobases in length. Profiles are plotted over the entire metagene and compare either WT and WT EAE (B), or Cbx5-/- and Cbx5-/- EAE (C).

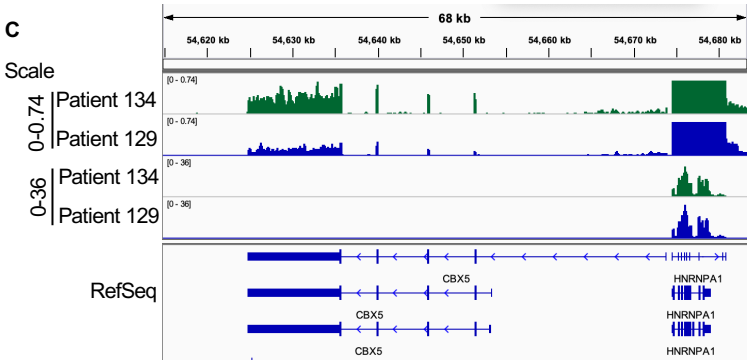
A

	Patient ID	Diagnosis	Sub-group	CBX5 level
Low-CBX5	129	RRMS	2	8.1533E-06
	139	RRMS + Crohns	2	9.5622E-06
	85	RRMS	2	9.6777E-06
	128	PPMS + Crohns	2	9.8989E-06
	141	SPMS + Psoriasis	2	1.1405E-05
	68	RIS	1	1.1617E-05
	138	RRMS	2	1.1999E-05
	86	RIS	2	1.2733E-05
	130	SPMS	1	1.3274E-05
	136	RRMS	1	1.5187E-05
	127	RRMS + Psoriasis	1	1.5283E-05
	125	RRMS	1	1.6386E-05
	140	RRMS + Ulcerative colitis	1	1.7088E-05
High-CBX5	69	CIS + Psoriasis	1	1.9183E-05
	76	CIS	1	2.0406E-05
	73	CIS + Graves	1	2.0511E-05
	61	RRMS + IDDM	1	2.0806E-05
	134	RIS	1	2.1532E-05
	135	RRMS + Psoriasis	1	2.1711E-05
Controls	75	SC		8.7405E-06
	63	SC		1.3461E-05
	79	SC		1.4064E-05
	62	SC		1.5061E-05
	81	SC		1.5828E-05
	58	SC		1.5914E-05
	80	SC		1.6128E-05

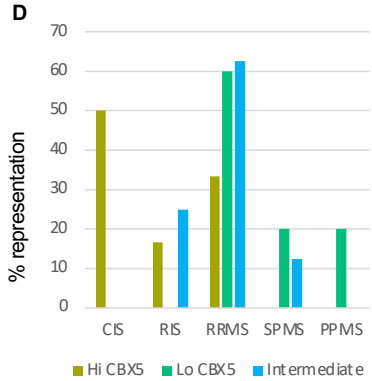
B



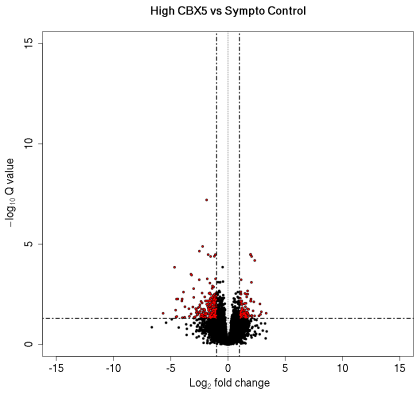
C



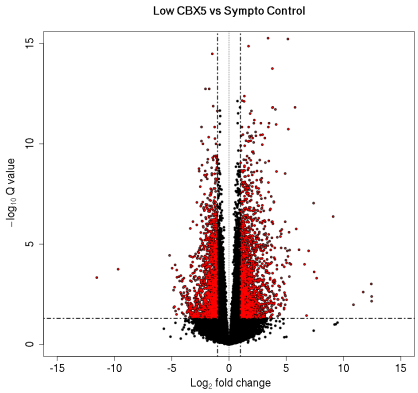
D



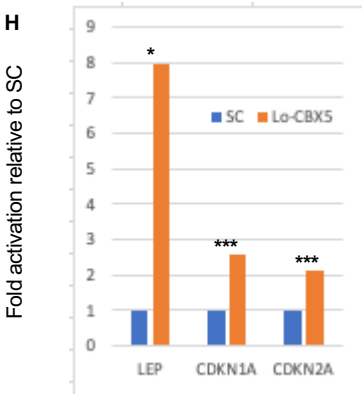
E



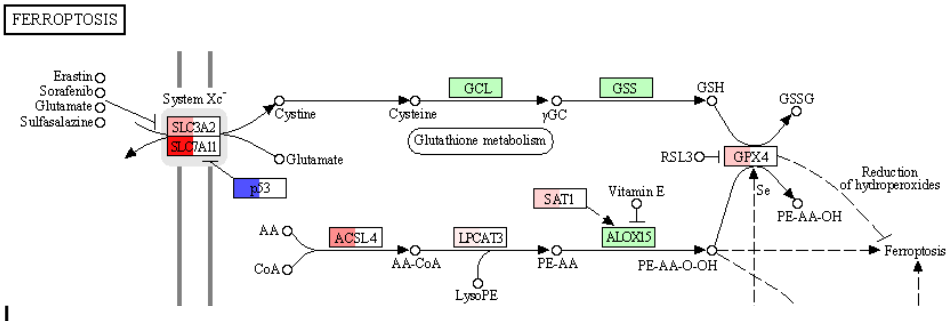
F



H



G



I

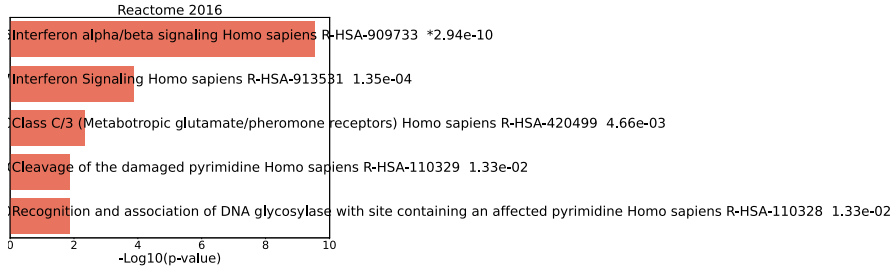


Figure 1

A Genes up in **High** CBX5 compared to **SC**: 109 genes

ChEA 2022 Bar Graph **Table** Clustergram Appyter

Hover each row to see the overlapping genes.

10 entries per page Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	VDR 24763502 ChIP-Seq THP-1 Human	0.00004294	0.02894	3.43	34.48
2	RELB 30642670 ChIP-Seq CTB1 Human Placenta Inflammation	0.0001139	0.03838	8.65	78.58
3	VDR 24787735 ChIP-Seq THP-1 Human	0.0009890	0.2222	3.31	22.92
4	FOXM1 32153563 ChIP-Seq SKOV3 Human Ovary OvarianCancer	0.001558	0.2625	6.37	41.18
5	ESR2 21235772 ChIP-Seq MCF-7 Human	0.004162	0.5610	3.69	20.24
6	KLF5 18264089 ChIP-ChIP MESCs Mouse	0.01125	0.9482	6.84	30.68
7	KLF2 18264089 ChIP-ChIP MESCs Mouse	0.01125	0.9482	6.84	30.68
8	KLF4 18264089 ChIP-ChIP MESCs Mouse	0.01125	0.9482	6.84	30.68
9	PPARG 19300518 ChIP-PET 3T3-L1 Mouse	0.02063	1.000	4.01	15.58
10	ESR1 22217937 ChIP-Seq MCF-7 Human BreastCancer	0.03625	1.000	4.30	14.27

B Genes up in **Low** CBX5 compared to **SC**: 1395 genes

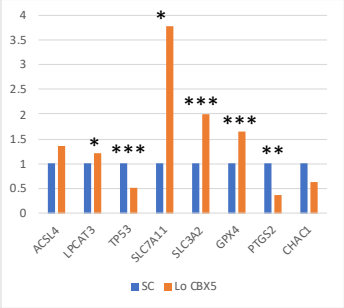
KEGG 2021 Human Bar Graph **Table** Clustergram Appyter

Hover each row to see the overlapping genes.

10 entries per page Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Rheumatoid arthritis	0.0002152	0.02023	3.01	25.40
2	HIF-1 signaling pathway	0.0001802	0.02023	2.84	24.49
3	MAPK signaling pathway	0.0001743	0.02023	2.01	17.37
4	Ferroptosis	0.0004017	0.02832	4.33	33.83
5	IL-17 signaling pathway	0.0007432	0.03342	2.76	19.86
6	TGF-beta signaling pathway	0.0007432	0.03342	2.76	19.86
7	NF-kappa B signaling pathway	0.0008295	0.03342	2.63	18.63
8	Legionellosis	0.001715	0.04836	3.21	20.42
9	AGE-RAGE signaling pathway in diabetic complications	0.001471	0.04836	2.56	16.69
10	Fluid shear stress and atherosclerosis	0.001574	0.04836	2.26	14.58

C



D **WikiPathway 2021 Human** Bar Graph **Table** Clustergram Appyter

Hover each row to see the overlapping genes.

10 entries per page Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Adipogenesis WP236	2.537e-7	0.0001223	3.55	53.84
2	Hypertrophy Model WP516	0.000003161	0.0007619	10.98	139.01
3	Nuclear Receptors Meta-Pathway WP2882	0.000004969	0.0007984	2.23	27.23
4	Apoptosis-related network due to altered Notch3 in ovarian cancer WP2864	0.00001100	0.001326	4.83	55.10
5	IL-18 signaling pathway WP4754	0.00001420	0.001369	2.27	25.31
6	Imatinib and Chronic Myeloid Leukemia WP3640	0.00003238	0.002601	8.94	92.39
7	Glucocorticoid Receptor Pathway WP2880	0.00007835	0.005395	3.67	34.66
8	Transcriptional cascade regulating adipogenesis WP4211	0.0001275	0.007682	11.48	102.92
9	Pathways in clear cell renal cell carcinoma WP4018	0.0002303	0.01233	3.12	26.11
10	Signal transduction through IL1R WP4496	0.0003173	0.01239	5.03	40.50

E

Reactome 2022

Interferon Alpha/Beta Signaling R-HSA-9097
Interferon Signaling R-HSA-913531
OAS Antiviral Response R-HSA-8983711
Class C/3 (Metabotropic Glutamate/Pherom
Sensory Perception Of Sweet, Bitter, And Un

BioPlanet 2019

Interferon alpha/beta signaling
Interferon signaling
Type II interferon signaling (interferon-gam
Beta-alanine metabolism
Taste transduction

WikiPathway 2021 Human

Type I interferon induction and signaling du
Immune response to tuberculosis WP4197
Host-pathogen interaction of human corona
Type II interferon signaling (IFNG) WP619
Pathways of nucleic acid metabolism and in

Sup. Figure 1

A ENCODE TF ChIP-seq 2015

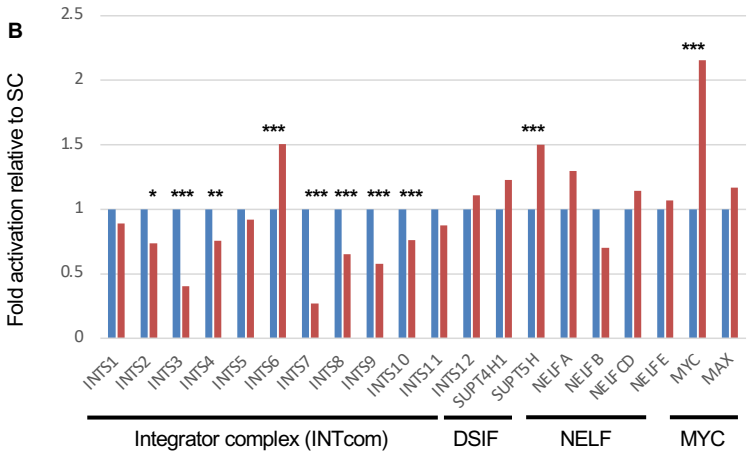
Bar Graph | **Table** | Grid | Network | Clustergram | Appyter

Hover each row to see the overlapping genes.

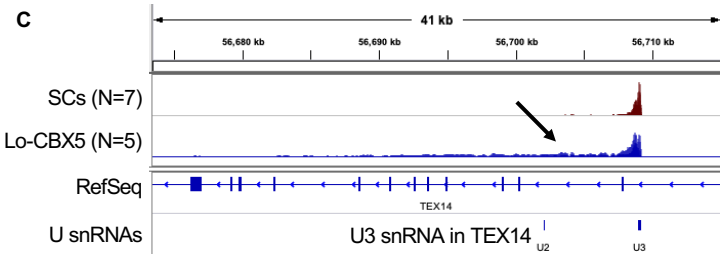
10 entries per page Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	NELFE K562 hg19	2.984e-14	2.435e-11	3.33	103.75
2	FOSL1 K562 hg19	2.754e-7	0.0001105	1.82	27.47
3	EP300 HeLa-S3 hg19	4.064e-7	0.0001105	1.55	22.84
4	JUND HCT116 hg19	0.00001009	0.002058	2.12	24.39
5	FOSL2 SK-N-SH hg19	0.00002964	0.004837	1.88	19.61
6	RXRA SK-N-SH hg19	0.00009109	0.01157	2.67	24.86
7	TCF12 SK-N-SH hg19	0.00009923	0.01157	1.58	14.59
8	MEF2C GM12878 hg19	0.0001871	0.01908	2.47	21.19
9	JUND HeLa-S3 hg19	0.0004750	0.04306	1.39	10.61
10	FOSL1 HCT116 hg19	0.0007362	0.05461	1.79	12.88

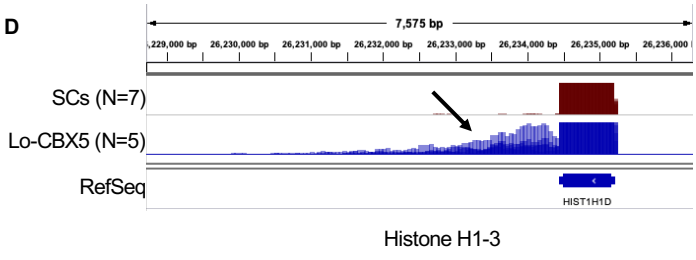
B



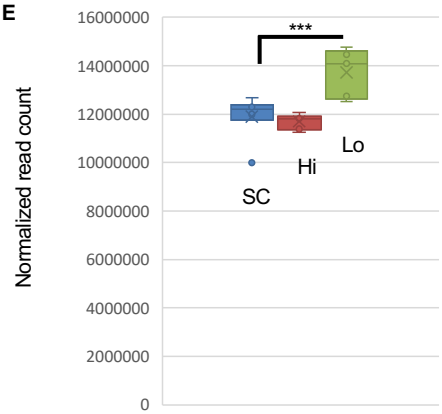
C



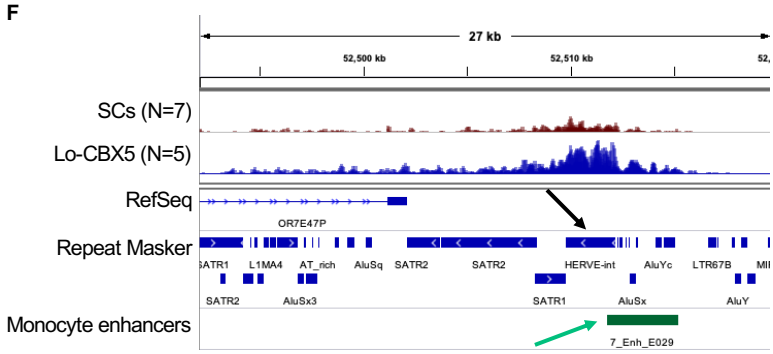
D



E

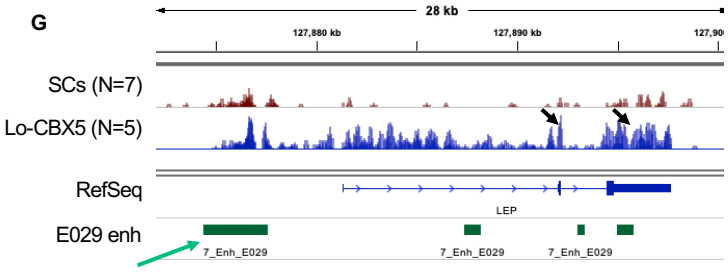


F



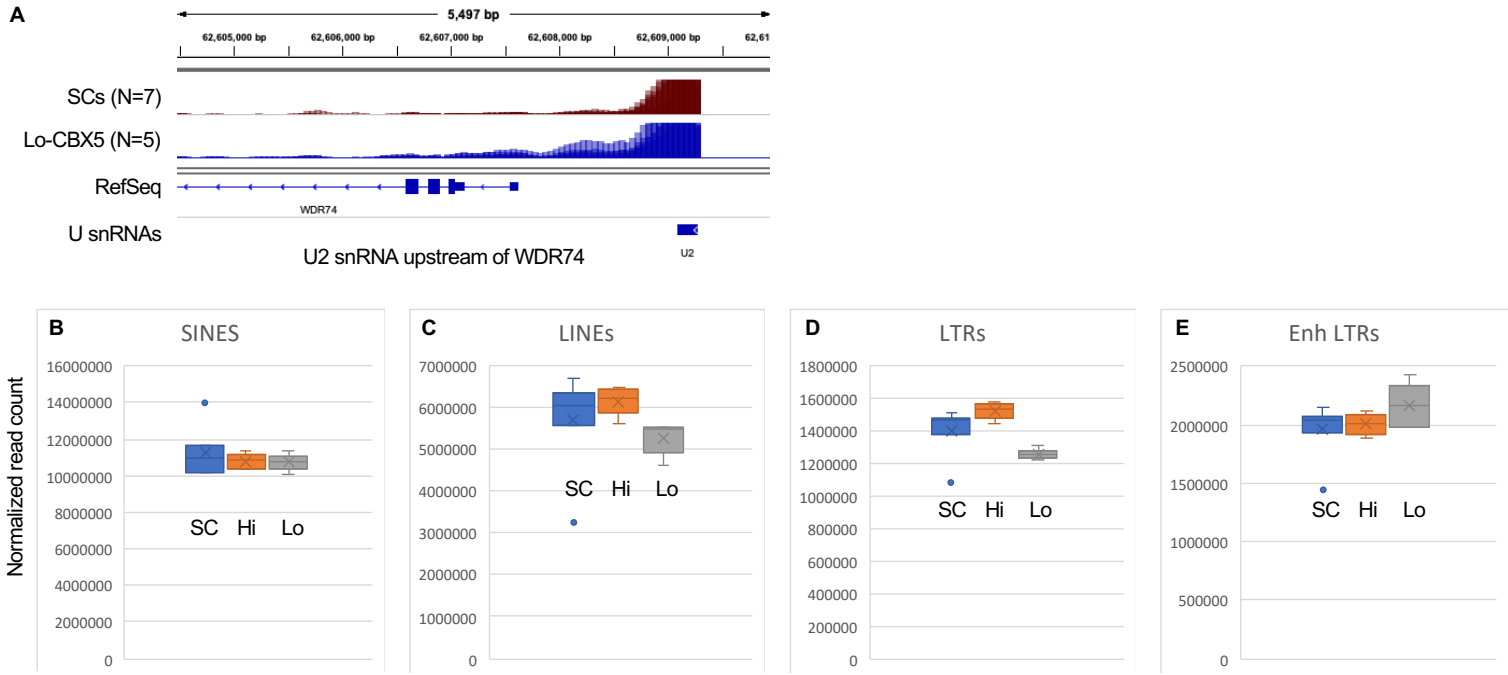
Representative enhancer transcription bleeding into a HERV

G



LEPTIN

Figure 2


A
5,497 bp
62,605,000 bp
62,606,000 bp
62,607,000 bp
62,608,000 bp
62,609,000 bp
SCs (N=7)
Lo-CBX5 (N=5)
RefSeq
WDR74
U snRNAs
U2
U2 snRNA upstream of WDR74
Normalized read count
B
SINES
16000000
14000000
12000000
10000000
8000000
6000000
4000000
2000000
0
SC Hi Lo
C
LINEs
7000000
6000000
5000000
4000000
3000000
2000000
1000000
0
SC Hi Lo
D
LTRs
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
200000
0
SC Hi Lo
E
Enh LTRs
2500000
2000000
1500000
1000000
500000
0
SC Hi Lo

Sup. Figure 2

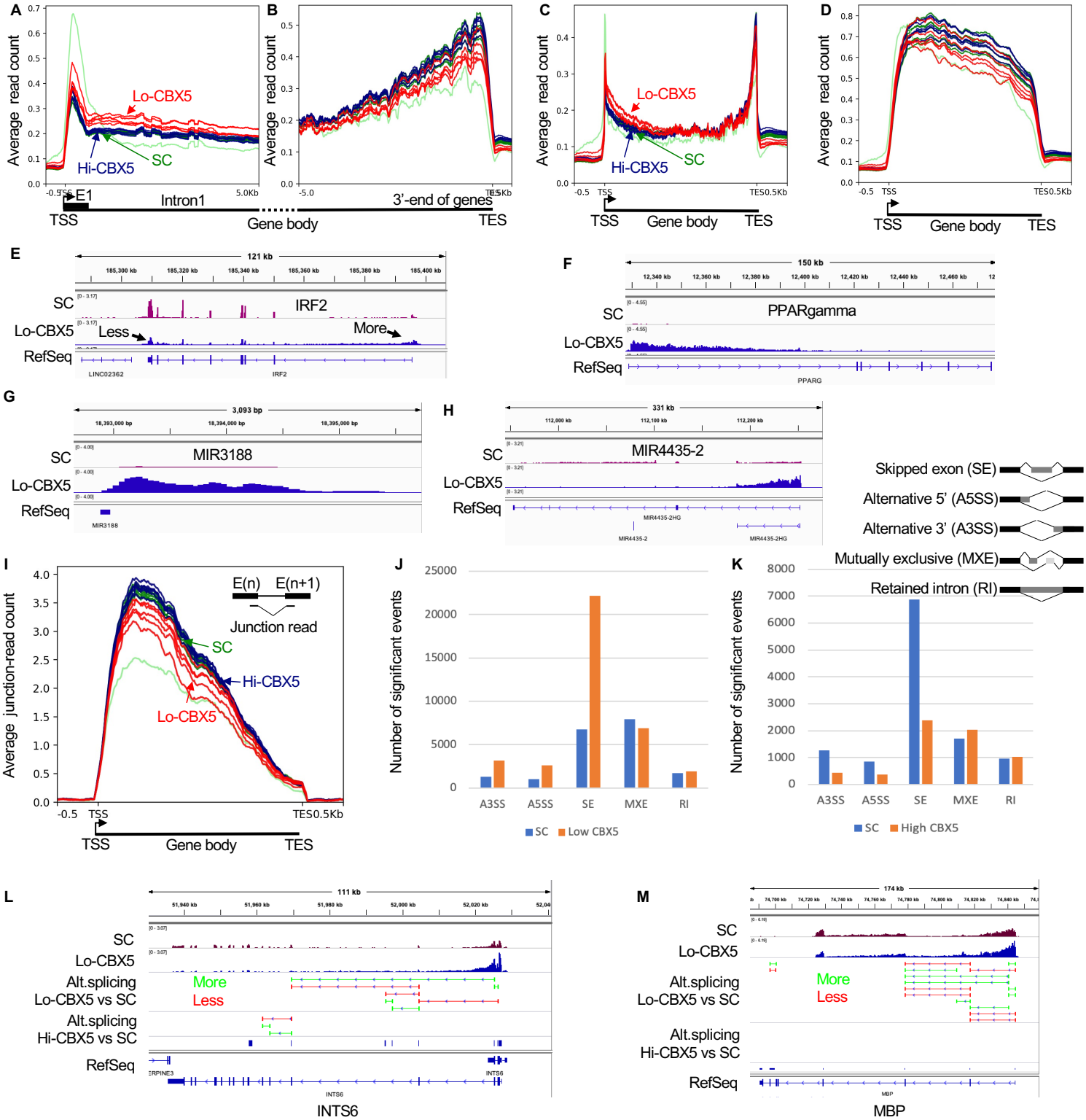
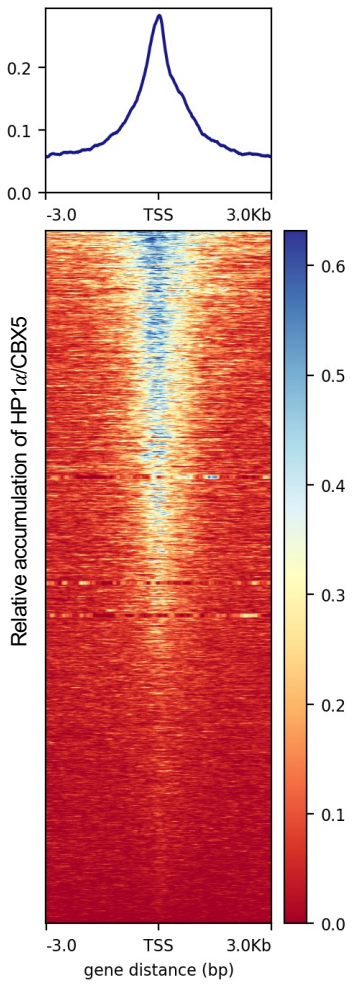
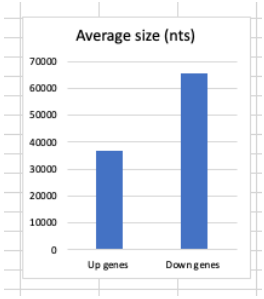

A
B
C
D
Average read count
Lo-CBX5
Hi-CBX5
SC
E1
Intron1
3'-end of genes
TSS
Gene body
TES
E
F
G
H
SC
Lo-CBX5
RefSeq
IRF2
Less
More
PPARgamma
MIR3188
MIR4435-2
Skipped exon (SE)
Alternative 5' (A5SS)
Alternative 3' (A3SS)
Mutually exclusive (MXE)
Retained intron (RI)
I
J
K
Average junction-read count
E(n)
E(n+1)
Junction read
Number of significant events
A3SS
A5SS
SE
MXE
RI
Low CBX5
High CBX5
L
M
Alt.splicing
Lo-CBX5 vs SC
Hi-CBX5 vs SC
INTS6
MBP

Figure 3

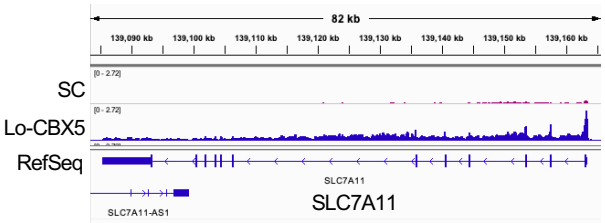
A HP1α/CBX5 ChIP-seq data from HepG2 cells



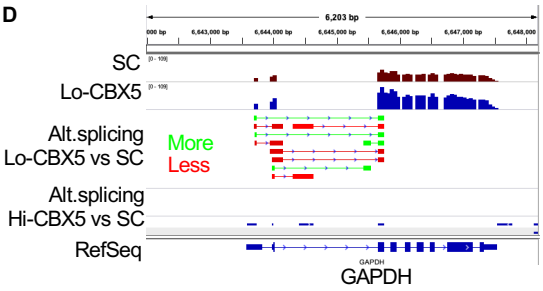
B



C

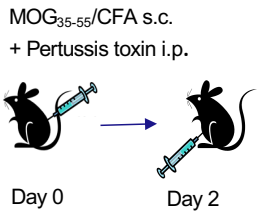


D

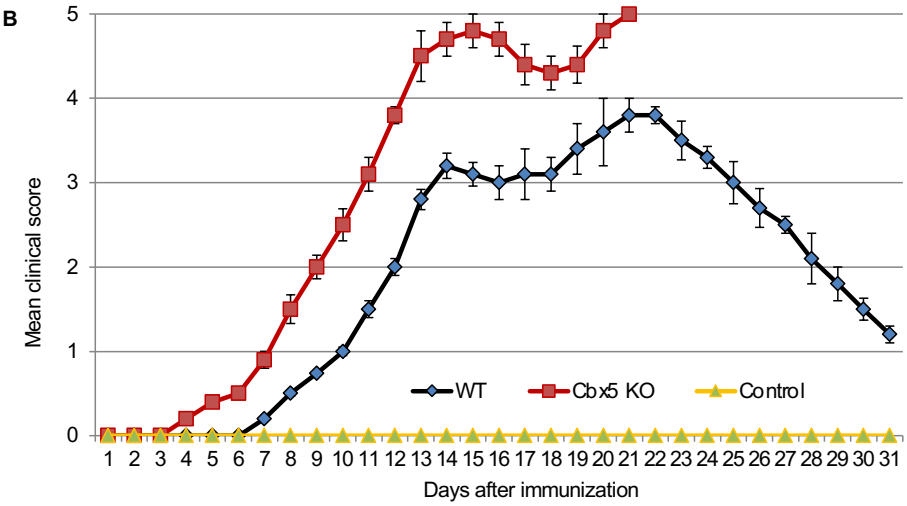


Sup. Figure 3

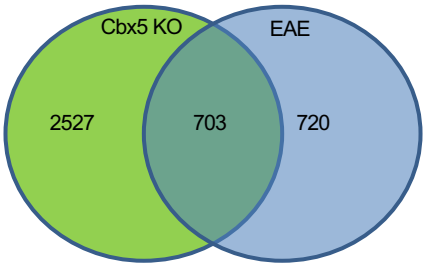
A



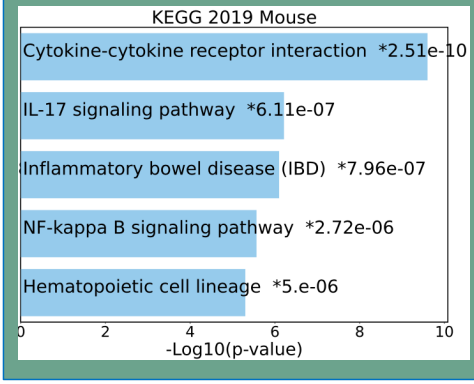
B



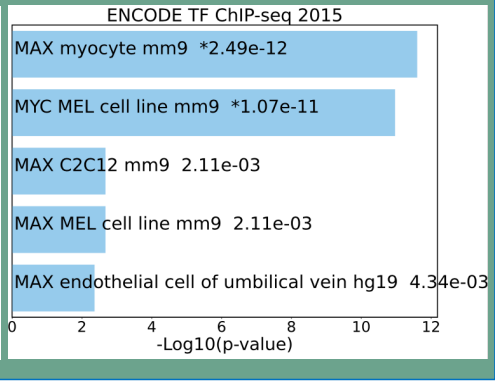
C Up-genes



D 703 common up genes



E



F 3230 genes up in Cbx5-/- vs WT (Min. 2-fold, adj. pVal<0.05)

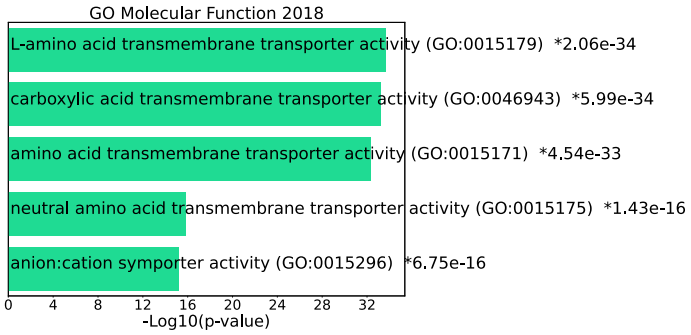
GO Biological Process 2021 Bar Graph **Table** Clustergram Appyter

Hover each row to see the overlapping genes.

10 entries per page Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	transport across blood-brain barrier (GO:0150104)	1.610e-7	0.0007384	3.52	55.02
2	vascular transport (GO:0010232)	3.097e-7	0.0007384	3.47	52.02
3	regulation of calcidiol 1-monooxygenase activity (GO:0060558)	0.00005708	0.09072	35.41	345.97
4	regulation of smooth muscle cell proliferation (GO:0048660)	0.00009781	0.09349	3.43	31.71
5	inflammatory response (GO:0006954)	0.00009804	0.09349	1.87	17.24
6	amino acid transport (GO:0006865)	0.0001323	0.1051	3.33	29.71
7	L-alpha-amino acid transmembrane transport (GO:1902475)	0.0001764	0.1052	4.73	40.85
8	response to molecule of bacterial origin (GO:0002237)	0.0001741	0.1052	2.72	23.57
9	amino acid import across plasma membrane (GO:0089718)	0.0002358	0.1249	5.00	41.75
10	L-ornithine transmembrane transport (GO:1903352)	0.0003389	0.1469	29.50	235.67

G 77 SLC genes upregulated (Min. 2-fold, Adj pVal<0.05)



H

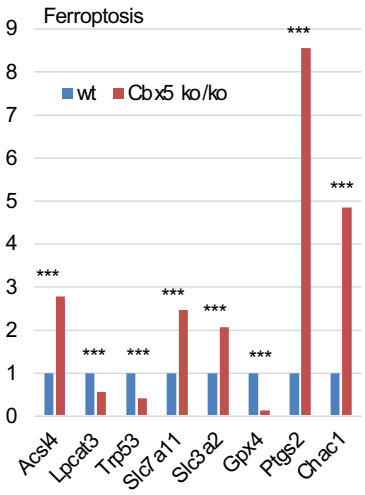
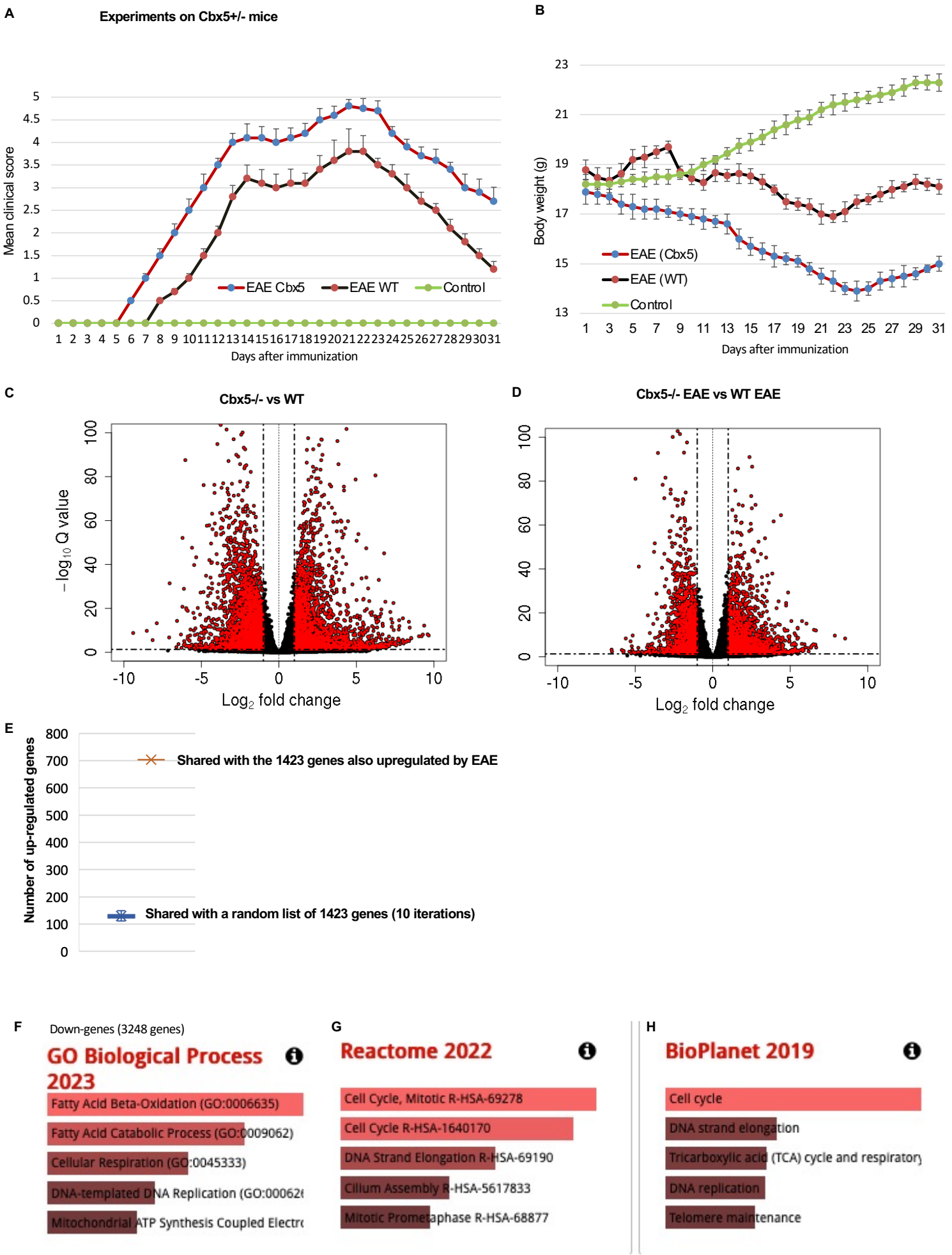


Figure 4


A
Experiments on Cbx5+/- mice
Mean clinical score
Days after immunization
EAE Cbx5
EAE WT
Control
B
Body weight (g)
Days after immunization
EAE (Cbx5)
EAE (WT)
Control
C
Cbx5-/- vs WT
$-\log_{10}$ Q value
Log_2 fold change
D
Cbx5-/- EAE vs WT EAE
Log_2 fold change
E
Number of up-regulated genes
Shared with the 1423 genes also upregulated by EAE
Shared with a random list of 1423 genes (10 iterations)
F
Down-genes (3248 genes)
GO Biological Process 2023
Fatty Acid Beta-Oxidation (GO:0006635)
Fatty Acid Catabolic Process (GO:0009062)
Cellular Respiration (GO:0045333)
DNA-templated DNA Replication (GO:00062
Mitochondrial ATP Synthesis Coupled Electr
G
Reactome 2022
Cell Cycle, Mitotic R-HSA-69278
Cell Cycle R-HSA-1640170
DNA Strand Elongation R-HSA-69190
Cilium Assembly R-HSA-5617833
Mitotic Prometaphase R-HSA-68877
H
BioPlanet 2019
Cell cycle
DNA strand elongation
Tricarboxylic acid (TCA) cycle and respiratory
DNA replication
Telomere maintenance

Sup. Figure 4

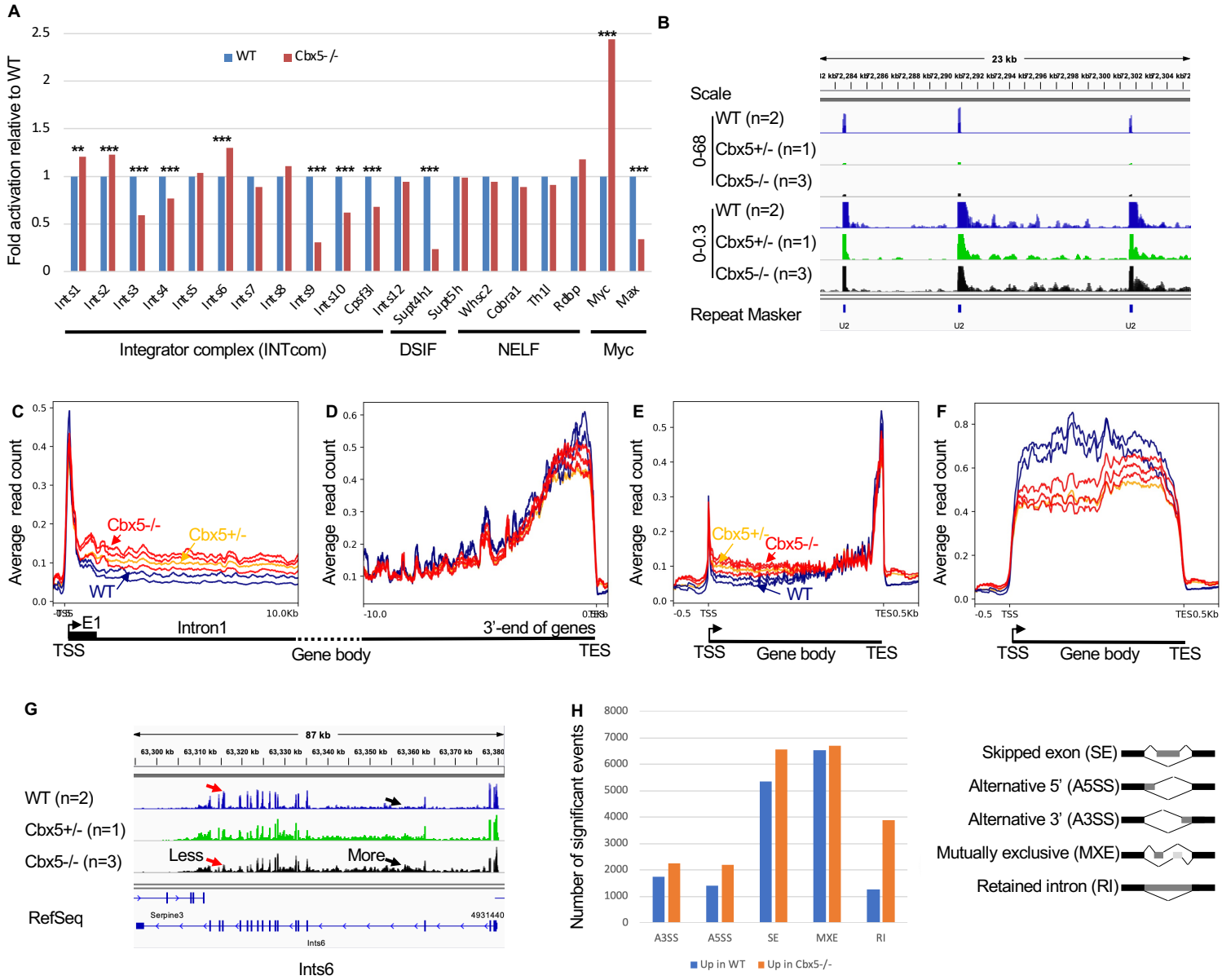
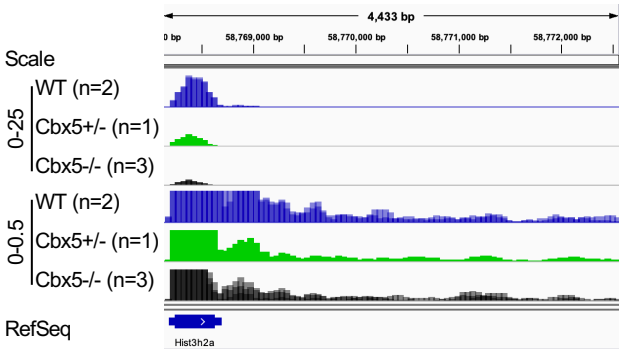

A
Fold activation relative to WT
WT
Cbx5-/-
Integrator complex (INTcom)
DSIF
NELF
Myc
B
Scale
23 kb
WT (n=2)
Cbx5+/- (n=1)
Cbx5-/- (n=3)
0-68
0-0.3
Repeat Masker
C
D
E
F
Average read count
Cbx5-/-
Cbx5+/-
WT
E1
Intron1
TSS
Gene body
3'-end of genes
TES
G
87 kb
WT (n=2)
Cbx5+/- (n=1)
Cbx5-/- (n=3)
Less
More
RefSeq
Serpine3
Ints6
H
Number of significant events
A3SS
A5SS
SE
MXE
RI
Up in WT
Up in Cbx5-/-
Skipped exon (SE)
Alternative 5' (A5SS)
Alternative 3' (A3SS)
Mutually exclusive (MXE)
Retained intron (RI)

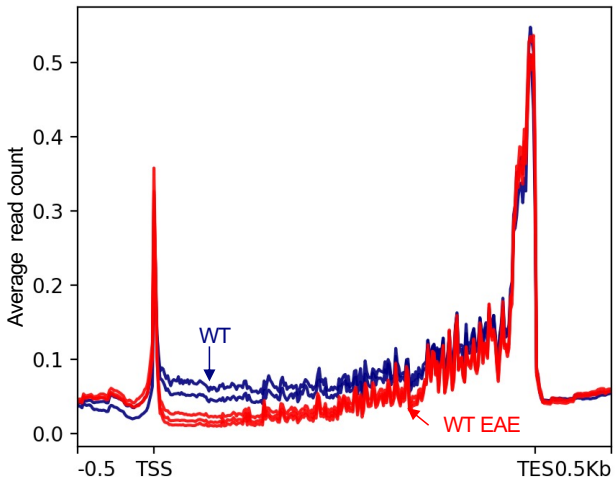
Figure 5

A

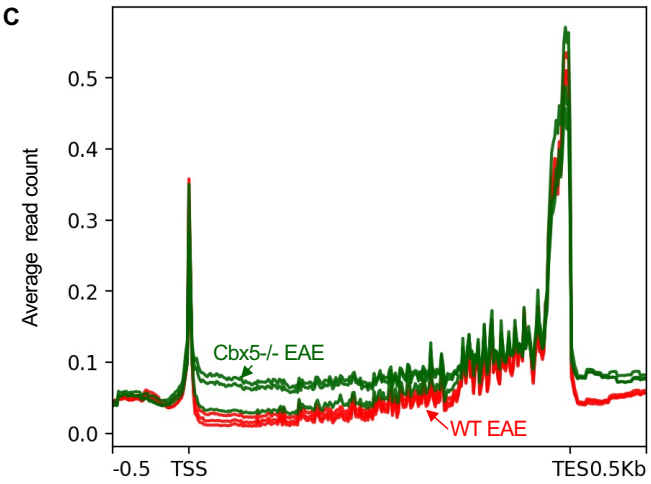


Histone H2ac25

B



C



Sup. Figure 5