Eryk Duński

Evolution of regulome in human astrocytes based on iPSC-derived astrocyte models

PhD thesis

Completed in the Dioscuri Centre

for Chromatin Biology and Epigenomics

of the Nencki Institute of Experimental Biology

Polish Academy of Sciences

Supervisor:

Dr. Aleksandra Pękowska, Ph.D., D.Sc.

Warsaw, 2025

This work was generously funded under Dioscuri, a programme initiated by the Max Planck Society, jointly manged with the National Science Centre in Poland, and mutually funded by Polish Ministry of Science and Higher Education and German Federal Ministry of Education and Research (UMO-2018/01/H/NZ4/00001).

Part of the project was funded by National Science Center, Poland under Preludium 22 grant (UMO-2023/49/N/NZ2/04241).

NARODOWE
CENTRUM
NAUKI

MAX-PLANCK-GESELLSCHAFT

Ministerstwo Nauki
i Szkolnictwa Wyższego

Federal Ministry
of Education
and Research

I would like to dedicate this work to my late mother, Dr. Ewa Jakubowska (1957-2025), who was the first person to teach me about genetics.

I would like to extend my special thanks to:

my supervisor, Dr. Aleksandra Pękowska, Ph.D., D.Sc., for scientific supervision and generous funding,

Debadeep Chaudhury for indispensable help with bioinformatic analyses,

colleagues from Dioscuri Centre for Chromatin Biology and Epigenomics, especially Aleksandra Piotrowska and Dr. Katarzyna Ciuba, for vital assistance with managing moods of *in vitro* astrocytes,

colleagues from Dioscuri Centre for Metabolic diseases, sequencing facility and doctoral community at Nencki Institute for collegial support and friendly atmosphere,

P.P.U. ADARA sp. z o. o. for sharing crucial reagents,

my Family – father Witold, brother Olaf, grandmothers Elżbieta and Barbara and my late grandfather Witold for supporting me throughout my education and work,

and finally to my fiancée Sandra for being the Sun that illuminates my days, and the Moon that guides me in the dark.

Nasze życie to wędrówka. Poprzez Zimę, poprzez Noc. My szukamy sobie przejścia w Niebo, gdzie nie świeci nic.

Contents

List of abbreviations

iPS (cell) – induced pluripotent stem (cell)

NP (cell) – neural progenitor (cell)

iAstrocyte – astrocyte derived from induced pluripotent stem cells

KD – knock-down

IF – immunofluorescence

GFAP – glial fibrillary acidic protein

ATAC-Seq – Assay for Transposase-Accessible Chromatin using sequencing)

ChIP-seq – chromatin immunoprecipitation-sequencing

TSS – transcription start site

HS – *Homo sapiens*

NHP – Non-human primate

PT – *Pan troglodytes*

RM – Rhesus Maccaque (*Maccaca mulatta*)

TF – transcription factor

TFBS – transcription factor binding site

USF – universal stripe factor. See Zhao et al., (2022) for a more detailed description

1Mb – one megabasepair, 1 000 000 basepairs for genomic distance

bp – basepair

AD – Alzheimer’s disease

SCZ – schizophrenia

MS – multiple sclerosis

PD – Parkinson’s disease

ASD – autism spectrum disorders

NGS – next-generation sequencing

SNP – single nucleotide polymorphism

Abstract

Astrocytes, a type of glial cells, are essential for proper brain function, and disruptions in their homeostasis are a hallmark of numerous pathological conditions. Human astrocytes exhibit unique features compared to their mouse and primate counterparts, which are believed to be driven by alterations in the DNA sequences of regulatory regions, particularly enhancers, which increase gene expression. The precise functional contribution of individual evolutionary changes in regulatory elements active in astrocytes, however, remains unknown.

The main objective of my doctoral research was to gain a deeper understanding of the role of astrocytes in shaping the characteristics of the human brain and to determine whether, and how, evolutionary changes in enhancers may have contributed to uniquely human gene expression in astrocytes and, consequently, to the emergence of traits of the human mind during evolution.

In this work, I used astrocyte models generated in vitro through the differentiation of induced pluripotent stem cells (iPSCs) derived from humans, chimpanzees (*Pan troglodytes*), and rhesus macaques (*Macaca mulatta*). The resulting *in vitro* astrocyte model (so-called iAstrocytes) exhibited a transcriptional profile corresponding to fetal astrocytes. Based on this model, I characterized the transcriptome, regulome, and chromatin conformation using omics techniques including RNA-seq, ATAC-seq, ChIP-seq, and Intact Hi-C.

I demonstrated that sequence variants associated with brain morphology are significantly enriched in genomic regions that bear the signature of active enhancers. Moreover, I identified enhancers active in human astrocytes whose sequences undergo exceptionally rapid evolutionary changes compared to other primates. I also showed that human-specific enhancers frequently contain DNA variants that distinguish modern humans from archaic hominins. These results support broader conclusions regarding the mechanisms underlying brain evolution and human cognitive abilities.

In the subsequent part of the dissertation, to identify functional regulatory elements, I employed a high-throughput reporter assay (MPRA, Massively Parallel Reporter Assay) to measure the activity of 10,241 unique DNA sequences that represent potential enhancers in primate iAstrocytes. Using MPRA, I assessed the effects of mutations within sequences recognized by selected transcription factors, as well as the contributions of enhancer variants associated with human traits and diseases.

In particular, I demonstrated that the presence of binding sites for USF (Universal Stripe Factors) transcription factors is essential for the activity of enhancers located near genes with

higher expression in human iAstrocytes than in chimpanzee and macaque iAstrocytes. Furthermore, I showed that USF binding sites underwent recent evolutionary changes and described associations between these evolutionary modifications in the regulome and disease-linked genetic variants.

In summary, the studies presented in this dissertation deepen our understanding of the role of astrocytes in the development and evolution of the human brain and provide new insights into gene regulatory mechanisms that may have contributed to the emergence of uniquely human traits.

Streszczenie

Astrocyty, będące komórkami gleju, są niezbędne dla poprawnego funkcjonowania mózgu, a zaburzenia w ich homeostazie są charakterystyczną cechą wielu zmian chorobowych. Ludzkie astrocyty wykazują unikalne cechy w porównaniu z ich mysimi i małpimi odpowiednikami. Tymi zmianami sterują ewolucyjne modyfikacje w sekwencji regionów regulatorowych DNA, zwłaszcza w tzw. wzmacniaczach („enhancerach"), które zwiększają ekspresję genów. Dokładna funkcjonalna rola poszczególnych zmian ewolucyjnych w elementach regulatorowych aktywnych w astrocytach pozostaje jednak nieznana.

Głównym celem mojej pracy doktorskiej było lepsze zrozumienie roli astrocytów w kształtowaniu cech charakterystycznych dla ludzkiego mózgu oraz ustalenie, czy i w jaki sposób ewolucyjne zmiany we wzmacniaczach mogły przyczynić się do specyficznej dla człowieka regulacji ekspresji genów w astrocytach, a co za tym idzie, do kształtowania się cech ludzkiego umysłu w toku ewolucji.

W pracy wykorzystano modele astrocytów otrzymane *in vitro* poprzez różnicowanie indukowanych pluripotentnych komórek macierzystych (ang. *induced pluripotent stem cells,* iPSC) pochodzących od człowieka, szympansa (*Pan troglodytes*) oraz makaka królewskiego (*Macaca mulatta*). Uzyskany w ten sposób model astrocytów *in vitro* (tzw. iAstrocyty) miał profil transkrypcyjny odpowiadającym astrocytom płodowym. W oparciu o ten model scharakteryzowano transkryptom, regulom i konformację chromatyny za pomocą technik omicznych: RNA-seq, ATAC-seq, ChIP-seq i Intact Hi-C.

Na podstawie analizy ludzkich regionów regulatorowych w niniejszej pracy wykazałem, że warianty sekwencji związane z morfologią mózgu są istotnie wzbogacone w regionach genomowych posiadających sygnaturę aktywnych wzmacniaczy. Ponadto, zidentyfikowałem wzmacniacze aktywne w ludzkich astrocytach, których sekwencje ulegają wyjątkowo szybkim zmianom ewolucyjnym w porównaniu z innymi naczelnymi. Pokazałem również, że wzmacniacze specyficzne dla człowieka często zawierają warianty DNA odróżniające ludzi od hominidów. Wyniki te sugerują szersze wnioski na temat mechanizmów związanych z ewolucja mózgu i ludzkich zdolności poznawczych.

W dalszej części rozprawy, w celu identyfikacji funkcjonalnych elementów, zastosowałem wysokoprzepustowy test reporterowy (MPRA, *Massively Parallel Reporter Assay*). Do zbadania aktywności 10,241 unikatowych sekwencji DNA, które są potencjalnymi wzmacniaczami w iAstrocytach naczelnych. Za pomocą MPRA oszacowałem wpływ mutacji w

sekwencjach rozpoznawanych przez wybrane czynniki transkrypcyjne oraz efekty wariantów wzmacniaczy związanych z cechami ludzkimi i chorobami.

W szczególności wykazałem, że obecność miejsc wiązania czynników transkrypcyjnych z rodziny USF (ang. *Universal Stripe Factors*). jest niezbędna dla aktywności wzmacniaczy zlokalizowanych w pobliżu genów wykazujących wyższą ekspresję w ludzkich iAstrocytach niż w iAstrocytach szympansa i makaka. Ponadto dowiodłem, że miejsca wiązania USF uległy niedawnym zmianom ewolucyjnym oraz opisałem powiązania między tymi zmianami a wariantami genetycznymi związanymi z chorobami.

Reasumując, badania przedstawione w niniejszej dysertacji pogłębiają nasze rozumienie roli astrocytów w rozwoju i ewolucji ludzkiego mózgu oraz dostarczają nowych informacji na temat mechanizmów regulacji genów, które mogły przyczynić się do powstania unikalnie ludzkich cech.

1. Introduction

1.1. Discovery of the importance of astrocytes

The brain is composed of two major cell types: neurons and glial cells. Neurons, which are generally regarded as the brain's principal functional units. Glial cells, often referred to as glia or neuroglia, encompass multiple subtypes and carry out essential roles such as providing structural and metabolic support, insulating neural elements, and guiding developmental processes[2]. In the brain, glia can be divided into macroglia, which include astrocytes and oligodendrocytes and are of ectodermal origin, and microglia, which are brain resident macrophages of mesodermal origin[3]. Astrocytes possess a highly branched morphology, forming an extensive arborization of fine processes that permeate the neural tissue. The number of such processes can be up to one hundred thousand in the mouse, while in the human brain, it exceeds one million[4]. Through these fine processes, astrocytes contact other astrocytes and neurons, forming a network which contributes to intercellular communication and homeostatic regulation in the central nervous system (CNS).

Due to this highly branched morphology, astrocytes were named after stars appearance (from ancient Greek - *ástron*, a star), by one of the pioneers of neurobiology, Santiago Ramón y Cajal. Although astrocytes, like other glial cell types, were described by early pioneers of neuroanatomy such as Cajal, Virchow, and Golgi[5], their crucial role in brain function was elucidated only relatively recently. This is striking, taking into account their great abundance – according to the most recent estimates, glia comprise approximately half of the total number of cells in the human brain, and astrocytes approximately 20 percent[6].

Astrocytes were first shown to be indispensable for neuronal growth *in vitro* more than 40 years ago[7–9]. Since then, research has demonstrated they play essential roles in maintaining the proper homeostasis of the extracellular milieu of the brain parenchyma, being critical for glutamatergic synaptic transmission in the central nervous system (CNS)[10,11].

The discoveries that calcium signalling in astrocytes, which is closely linked to glutamatergic activity in the CNS, constitutes a complete signalling network that is largely independent of synaptic transmission[12,13], indicate that astrocytes play important roles in shaping neural activity in the brain. Subsequent research, most notably by Benjamin Barres[14,15] and Maiken Nedergaard[16], has demonstrated the importance of astrocytes in synapse formation, maintenance, and pruning, which altogether shape neuron maturation[17]. Co-culturing acutely isolated retinal ganglion cells (RGCs) from rats with glial cells promotes synapse formation and maintenance, dramatically enhancing neuronal firing rate[14]. Inversely, he removal of astrocytes from co-cultures significantly reduced synapse density[15]. Those experiments established astrocytes as an essential component for proper neuronal function.

1.2. Processes regulated by astrocytes

Astrocytes are multifunctional cells[18,19]. They support neuronal networks by maintaining ionic and metabolic homeostasis, regulating nutrient uptake, and contributing to blood–brain barrier (BBB) integrity[18,20–29]. They also crucially influence the formation and maintenance of synapses through secreted factors[15,30–41], and modulation of extracellular matrix composition and physical contact[42–47]. Astrocytes also crucially control neurotransmitter balance[48–55]. Due to the role of astrocytes in synapse function and neurotransmission, they are an especially interesting target for study, as those processes are the basis of cognition and higher brain function

1.2.1. Astrocytes in the formation and elimination of synapses

Astrocytes, through a combination of secreted factors and direct cellular interactions, exert powerful crucial influence on the formation, maturation, and refinement of synapses. They produce various soluble signals, including thrombospondins, tumour necrosis factor-alpha (TNF-α), transforming growth factor-beta (TGF-β), estradiol, and brain-derived neurotrophic factor (BDNF), which may impact synapse biology.

Thrombospondins are crucial for synapse formation and orchestrate synapse organization through the α2δ neuronal receptor[35,36]. TNF-α enhances synaptic efficacy by increasing levels of AMPA receptors on the surface of neurons[37]. TNF-α is required for the continual maintenance of synaptic strength in hippocampal neuron cultures[38]. Likewise, the activation of TGF-β signalling through fibulin-2 located at small extracellular vesicles secreted by astrocytes induces synaptogenesis in mouse cortical neurons[40] in an activity and D-serine-dependent manner [39]. Furthermore, estradiol secreted by astrocytes was shown to increase the number of synapses and the rate of synaptic firing in co-cultures of rat neonatal astrocytes and neurons[41]. Another crucial factor, BDNF, which is abundantly produced in the brain, supports neuronal survival, differentiation, and synaptic maturation. Astrocytes both sense and recycle BDNF[56], influencing synaptic strength and plasticity. Consequently, astrocyte-derived BDNF is essential for dendritic growth, spine density, and memory function, as shown in experiments performed in a mouse model[57]. An interesting case of astrocyte-secreted factor that drives synapse maturation is Chordin-like 1 (Chrdl1), which promotes substitution of calciumpermeable AMPA glutamate receptors to their calcium-impermeable form[37]. Fascinatingly, mutations disrupting Chrdl1 function, while deleterious for overall health, result in higher cognitive ability[58].

Astrocytes also shape synapses by providing essential components of the extracellular matrix (ECM). Glypicans 4 and 6 (Gpc4 and Gpc6), produced by astrocytes, promote excitatory synapse formation through recruitment of GluA1 AMPA receptors[31]. In mice, knockout (KO) of Gpc4 impairs synapse formation[33], and was shown to induce the formation of excitatory synapses *in vitro*[32]. Very recently, another glypican, Gpc5, is important for synapse maturation by astrocyte-specific KO in mice[59]. Hevin (or SPARCL-1)[60] interacts with adhesion molecules: presynaptic neurexin-1α and postsynaptic neuroligin-1B, and serves as a “bridge” to allow synapse formation[61]. Astrocytes modulate the balance of synapse formation mediated by hevin by secreting SPARC, which counters its synaptogenic effects[60].

Astrocytes can also enhance synapse formation through physical contact. This was shown in experiments with acutely isolated rat embryonic neurons, where stimulation of integrin receptors by astrocytic contact triggered synapse formation via activation of the protein kinase C (PKC) pathway[42]. γ-protocadherins, neuroligins, and syndecans constitute adhesion molecules that promote synaptogenesis. Garrett and Weiner, (2009) demonstrated that synaptogenesis *in vitro* depends on the expression of γ-protocadherins in both astrocytes and neurons. Furthermore, studies in protocadherin-deficient mice show that specific proteins from this group regulate distinct pre- and postnatal stages of brain development[44]. Some of the protocadherins influence synapse formation through interactions with neuroligins[45,46]. Deletion of neuroligin-2 (NL2) specifically in astrocytes in mouse embryos results in a shifted balance of inhibitory and excitatory synapse formation[47].

Astrocytes produce a myriad of projections that contact and ensheathe nearby synapses[62]. These astrocytic protrusions are essential for the proper neuronal communication as they sense and respond to synaptic activity and, in turn, modulate neurotransmission[63–66]. In this context, one of the most crucial functions of astrocytes is the release of “gliotransmitters, discussed in more detail in the subsequent section. This essential role of astrocytes in synaptic function has led to the formulation of the concept of the “tripartite synapse”, comprising not only the pre- and post-synaptic terminals of two neurons, but also the associated astrocytic projections[65,67]

In addition to their role in promoting synaptic growth, astrocytes are essential for modelling the brain connectome through synapse elimination via multiple pathways, and in ways which are brain area- and circuit- specific[68]. Astrocytes were shown to mediate synapse elimination during brain maturation by activating the complement cascade[69]. Interleukin-33 (IL-33), secreted by astrocytes, regulates synapse number by promoting microglial synaptic pruning during development. It also enhances excitatory synapse formation and homeostatic plasticity in the adult hippocampus, contributing to learning and memory[70,71]. The cross-talk with microglia is two-sided – secretion of Wnt ligands by microglia modulates astrocyte-synapse contact to allow synapse pruning by microglia[72]. Astrocytes also utilize MEGF10 and MERTK phagocytic pathways to remove redundant synapses, as demonstrated during mouse brain development[73]. Knockout experiments in mice lacking type 2 inositol 1,4,5-trisphosphate receptor (IP3R2) showed that IP3R2-dependent ATP release is another pathway for astrocyte-mediated synaptic regulation[74]. Another particularly interesting example of astrocytic control over synapse formation is the secretion of ephrin-B1, which competes with neuronally derived ephrin-B1[75]. Overexpression of ephrin-B1 in mouse astrocytes reduced excitatory synapse formation and led to impaired memory retention. Hence, astrocytes may shape the brain connectome.

The astrocyte-microglia cross-talk is important for synapse pruning. In the postnatal mouse cortex, microglia use Wnts to direct astrocytes to reduce their contact with synapses before synapse engulfment by microglia. This mechanism allows microglia to instruct astrocyte-synapse interactions to facilitate synapse removal and remodelling.

Together, these findings highlight the multifaceted ways in which astrocytes regulate the formation, maturation, and elimination of synapses through both molecular signalling and physical interactions. By integrating these diverse mechanisms, astrocytes play a central role in shaping neural circuitry and maintaining synaptic balance throughout brain development and function.

1.2.2. Gliotransmission: a complex interplay between neurotransmitter processing and astrocyte metabolism

One of the most crucial functions of astrocytes is the release of “gliotransmitters” – the most important being glutamate, ATP, D-serine, and γ-aminobutyric acid (GABA)[76,77]. Release of gliotransmitters forms a complex, interdependent system linking astrocytic signal transduction, neuronal support, and astrocyte metabolism[78].

The “textbook” function of astrocytes is to clear the excess glutamate, the major neurotransmitter in the brain, from the synaptic cleft after synaptic firing. This function is performed through glutamate transporters EAAT1 (GLAST) and EAAT2 (GLT1). EAAT2 is responsible for the majority of glutamate uptake in the forebrain, while EAAT1 is more prominently expressed in the cerebellum[79,80]. KO of astrocyte transporters in rats results in neurodegeneration and progressive paralysis[53], which is caused by hyperexcitation and neuronal death, due to the process of “glutamate excitotoxicity”[81]. Hence, the removal of glutamate by astrocytes plays an important role in maintaining a healthy brain connectome.

Astrocytic function in glutamate significantly impacts their metabolism. After uptake, the majority of glutamate is converted to glutamine by astrocyte-specific glutamine synthase[82,83]. Glutamine is then secreted from astrocytes through SNAT3 and SNAT5 transporters. Next, the amino acid is taken up by neurons by SNAT1 and SNAT4, where it acts as a glutamate precursor. This process is referred to as the glutamate-glutamine cycle[84].

The remainder of the astrocytic glutamate is metabolised to α-ketoglutarate, which is a substrate for the production of ATP[85], a potent neuromodulator through which astrocytes impact the excitability of neurons.

Metabolising glutamate offsets the high energetic cost of its uptake from the synaptic cleft[86]. Glutamate import is coupled to Na^+ uptake, which activates glycolysis and glycogenolysis[87] to provide energy to astrocytes. Glycolysis is the source of lactate, which is then released into the extracellular space to be taken up as the main energy source by neurons[88]. Interestingly, glutamate uptake by astrocytes varies spatially within the brain, which suggests that astrocytes may control differential activation of synapses in different cortical regions[89]. Disruption of the glutamate cycle was shown to be important for Alzheimer’s disease (AD)[90], amyotrophic lateral sclerosis (ALS)[91] epilepsy[92].

The release of ATP, another key gliotransmitter, is primarily mediated through Ca^{2+}-dependent exocytosis[48–52]. ATP acts on both ionotropic P2X and metabotropic P2Y receptors located on neurons and glia[50,93]. It modulates synaptic transmission and plasticity by, for instance, causing glutamate release. Glutamate exocytosis in turn influences the density of NMDA and AMPA receptors impacting synaptic maturity[94–96]. ATP also modulates inhibitory transmission, for example, by enhancing GABAergic signalling at certain interneuron synapses[97,98]. In the extracellular space, ATP is rapidly degraded to adenosine by ectonucleotidases present at the astrocytic surface. Adenosine acts predominantly on A1 receptors to suppress presynaptic glutamate release and dampen neuronal excitability[94,99]. Recent studies also reveal that astrocyte-derived ATP can influence network excitability in a cell-type-specific manner. Astrocytic ATP affects interneurons and pyramidal neurons differently, and even triggers further ATP release from microglia, underscoring the complexity of glial communication in the brain beyond neurons[51,97]. Behavioural effects of astrocytic ATP and adenosine include modulation of fear expression, demonstrated in mice[100], modulation of the dopaminergic reward system[101], and depression like-symptoms[102].

D-serine was previously thought the be a gliotransmitter[103]. Yet, more recent data challenge this view by revealing that astrocytes produce and secrete L-serine. This amino acid is subsequently shuttled to neurons, where it is used for the synthesis of D-serine, the enzyme serine racemase[77]. Both inhibition of astrocytic D-serine production[54,104] and synthesis of L-serine from D-serine precursor[105] were shown to impair NMDA receptor activation, which is one of the main mechanisms of memory formation[106].

Gamma aminobutyric acid (GABA), the primary inhibitory neurotransmitter in the brain, is synthesized in astrocytes by glutamate decarboxylase[55,107], or is taken-up from the synaptic space by both neurons and astrocytes, by GABA transporters (GATs) from solute carrier 6 (SLC6) transporter family[108]. GABA is connected to the glutamate cycle, since its breakdown through the tricarboxylic cycle[109,110], yields glutamate[111]. Uptake of GABA from the perisynaptic space by astrocytes modulates glutaminergic neurotransmission[99]. Release of GABA by astrocytes regulates tonic inhibition - a persistent, background form of inhibitory signalling that continuously regulates the excitability of neurons[112,113]. Genetic modifications and pharmacological inhibition in mice helped reveal the implication of this mechanism in motor coordination[114]. Notable, KO of $GABA_B$ receptors in prefrontal cortex astrocytes was shown to affect goal-directed behaviours in mice[115], highlighting the importance of astrocytic sensing of GABA in the modulation of behaviour. The functions of GABA signalling in astrocytes extend beyond impacting motor and goal-related behaviours. Indeed, astrocytic GABA modulates the circadian clock, as shown by an experiment in which an astrocyte-specific disruption of a circadian clock protein was rescued by GABA receptor antagonists[116].

Single, highly branched astrocytes contacting multiple synapses may form microdomains and play a role in communication between multiple synapses[117]. Those branches form microdomains[118], which can respond to diverse neurotransmitter cues, and allow astrocytes to integrate information from multiple synapses, creating a slower-acting mechanism of neuromodulation[119]. Astrocytes may play modulatory roles across larger scales and communicate not only using soluble factors (e.g., gliotransmitters or extracellular vesicles) but also by forming direct intercellular contacts. In fact, astrocytes are highly interconnected and form a syncytium, which may connect cortical areas or brain regions[12]. Connexin 30 or 43 (Cx30, Cx43) are essential for the formation of channels between cells through which astrocytes may exchange ions or secondary messengers, including calcium ions (Ca^{2+}), to impact each other. Ca^{2+} waves were first shown to propagate in cultured astrocytes in response to glutamate[12]; and were replicated *in situ*[18,120,121]. Such waves can propagate both within single astrocytes and across astrocytic networks via gap junctions and extracellular ATP diffusion, influencing large populations of neurons[118,122–124]. While such signalling is markedly slower than the communication between neurons, it might constitute another layer of information storage or processing that could impact higher-level brain functions, as suggested over 30 years ago[125].

Tunnelling nanotubes (TNTs) are thin, F-actin-dependent structures that allow direct transfer of ions, proteins, organelles, and even electrical signals between cells[126–128]. TNTs may constitute means of direct transport between astrocytes and neurons in the brain[129].

Together, these findings highlight the remarkable versatility of astrocytes as active regulators of neuronal communication, metabolism, and network dynamics. By coordinating gliotransmitter release, calcium signalling, and direct intercellular exchange, astrocytes integrate synaptic activity across spatial and temporal scales, highlighting their role as essential modulators of brain function and plasticity.

1.2.3. Astrocytes in blood-brain barrier transport and maintenance

Astrocytes are crucial for communication across the largely impermeable BBB and interface the CNS with the bloodstream. There are multiple cell types related to the BBB, which from "perivascular unit" (barrier endothelial cells (BEC), pericytes, astrocytes, microglia, and neurons), and astrocytes act as the "integrator"[20,22], connecting these diverse cell types.

The BBB itself is composed of BECs [23]. BECs form close interactions with astrocytic end feet through membrane proteins, including aquaporin 4 (AQP4) or Inwardly Rectifying Potassium Channel 4.1 (KIR4.1). Leveraging these channels, astrocytic endfeet cover over 90% of the brain vasculature[24] and are critical for the exchange of various substances between the blood and the brain parenchyma. Thereby, astrocytes are critical for the transport of water and nutrients, including glucose[25] and fatty acids[26]. Likewise, astrocytic endfeet are essential for the import of chemical signals[28] and exchange of ions [27]. Astrocytes take up glucose through GLUT-1, a crucial transporter of glucose into the CNS[130,131]. Non-esterified fatty acids are imported through fatty acid transport proteins (FATP)[132]. Water balance in the brain is maintained by astrocytic aquaporin 4 (AQP4)[21,29]

Astrocytes impact blood vessels directly, and the dialogue between these structures is bidirectional. In the developing brain, astrocytes release various factors, including (but not restricted to) transforming growth factor-β (TGF-β), which lead to the consolidation of the emerging capillaries and venules[133–136]. Mature astrocytes secrete angiopoietin 1 and angiotensin that restrict BBB permeability by supporting efficient organization of tight junctions[137,138]. Siddharthan (et al., 2007) has shown that BECs exposed to indirect astrocytic signals display elevated expression of tight junction proteins, which are responsible for the maintenance of barrier properties. Furthermore, in an *in vivo* experiment, a sustained removal of astrocytes using conditional expression of diphtheria toxin A subunit in mice has led to a sustained leakage of the BBB for several weeks[140].

Many factors expressed in astrocytes affect the integrity of the BBB and may play essential functions in disease. For example, KO of Wnt in astrocytes results in decreased expression of Wnt/β-catenin target genes in BECs, brain edema, altered coverage of blood vessels by astrocyte end-feet, and swelling of end-feet[141]. KO of Neogenin (Neo-1) in astrocytes in mice causes defects in the BBB: an increase in dextran leakage, disrupted basal membrane, and reduced pericyte coverage[142].

The glymphatic system, a term coined in the laboratory of Maiken Nedergaard[143], relates to a perivascular network of tunnels that facilitates the exchange of cerebrospinal fluid (CSF) to the interstitial fluid (ISF), thereby orchestrating the clearance of metabolic waste and proteins from the brain parenchyma[144]. Apart from purifying functions, the glymphatic system also contributes to the organ-wide transport of nutrients, neuromodulators, and metabolic compounds in the brain[144]. Astrocytic AQP4 water channels, densely localized to perivascular endfeet, enable efficient fluid inflow along para-arterial routes and efflux along the paravenous pathways, thereby underlying the flow of CSF and ISF and the function of this intricate system. The glymphatic system operates most efficiently during sleep, when interstitial space volume expands and neuronal activity decreases, enhancing waste removal[145]. Astrocyte-mediated glymphatic flow is crucial for clearing neurotoxic waste products, such as amyloid β, α-synuclein, and tau associated with AD[143,146].

Together, these findings underscore astrocytes as active regulators of the BBB, coordinating nutrient transport, ion and water homeostasis, and signalling to maintain barrier integrity. Recent discoveries show that astrocytes not only maintain the bidirectional exchange between the CNS and bloodstream, but also drive glymphatic clearance, and reinforce their essential role in brain health and protection against neurotoxic accumulation of harmful proteins or toxic compounds.

1.2.4. Astrocytes in brain metabolism

Astrocytes are the central element of brain metabolism – they import nutrients across the BBB, and process them into forms required by other cell types[147], and synthesize more complex macromolecules from basic building blocks[148,149]. Crucially, astrocytes are the site of metabolic processing and breakdown of molecules that accumulate in neurons due to their activity, and are toxic to them, including glutamate[81] and peroxidized fatty acids[150,151].

Glucose, after being taken up from the bloodstream by astrocytes, is stored by them in the form of glycogen. This metabolic route, called "glycogen shunt", is energetically unfavourable, but essential for coordinating multiple astrocytic functions. Glycogen storage is important for neurotransmitter balance, since it serves as a substrate of glutamate synthesis[152,153]. This underscores the close integration of synaptic and metabolic functions in astrocytes. Reduction of astrocytic glycogen levels has been shown to cause seizures, as observed in Lafora disease[154]. Glycogen storage is mobilised through stimulation of adrenergic receptors[155,156], allowing for increased glutamate and lactate production under stress conditions. Lactate, which is produced from astrocytic glycogen and transported to neurons via monocarboxylate transporters[157], serves as an essential energy source for neurons[158].

While in astrocytes, glucose is primarily metabolized through glycolysis[159,160], neurons, due to inhibition of key glycolytic enzymes (for example, Pfkfb3)[161], produce energy mainly through oxidative phosphorylation. Due to the susceptibility of neurons to oxidative damage, much of the glucose in this cell type is routed into the pentose phosphate pathway, generating an antioxidant, reduced glutathione[162]. Consequently, neurons rely on astrocyte-derived lactate as a major energy source. This metabolic coupling is referred to as "astrocyte-neuron lactate shuttle"[159]. Supply of astrocytic lactate to neurons has important implications for cognition, as it is critical for long-term potentiation in synapses and learning[163].

While neurons are capable of fatty acid synthesis[164], astrocytes are the site of fatty acid recycling from neurons. Due to the oxidative burden of beta oxidation, which neurons are maladjusted to bear[150], oxidised fatty acids are transported to astrocytes, mainly through lipoprotein particles including the AD-related ApoE[165].

Since BBB is impermeable to high-density lipids, astrocytes synthesize the majority of cholesterol in the brain[166]. Cholesterol is crucial for proper brain function, including synapse formation[167] and the synthesis of the myelin sheath[168]. Astrocytes synthesize cholesterol directly from Acetl-CoA[169] and transport it to neurons through ApoE-dependent lipoprotein shuttle[170]. The importance of astrocyte cholesterol synthesis has been shown in mouse studies, where a KO of Serpin mRNA Binding Protein 2 (SERBP2) gene, which is an RNA-binding regulation of protein translation in astrocytes, leads to impaired brain development and neurite outgrowth[166].

These metabolic functions highlight astrocytes as essential integrators of nutrient supply and brain energy homeostasis in the brain. Disruption of astrocytic metabolism can therefore have profound effects on cognition, synaptic function, and overall brain health.

1.3. Role of astrocytes in disease

The essential role of astrocytes is frequently disrupted in disease states. Dysregulated interaction of astrocytes with synapses has been implicated in multiple neuropsychiatric diseases, mainly degenerative, such as AD[171,172], multiple sclerosis (MS)[173,174], Parkinson's disease (PD)[175,176], but also in neurological[177,178] and psychiatric diseases such as schizophrenia (SCZ)[179,180] and depression[181,182].

A crucial element of astrocyte biology in the disease context is the state of "reactivity", which they enter in response to damage or injury. Reactive astrocytes, generally described as being larger in size and expressing abnormally high levels of glial acidic fibrillary protein[18], constitute a wide spectrum of phenotypes[183]. This category is nevertheless a useful hallmark of astrocyte dysfunction, since activated astrocytes can feature hampered neuronal support leading to neurotoxicity[184]. "Reactive astrogliosis" and "astroglial reactivity" are used interchangeably to describe astroglial responses to pathology[185,186].

Astrocytes, due to their crucial support roles, have been a point of focus in the context of neurodegenerative diseases characterised by neuronal death, such as AD. Unsurprisingly, many aspects of AD are linked to astrocyte dysfunction. Even early, presymptomatic stages of the disease show signs of astrocyte reactivity, as demonstrated in the mouse model[187]. The development of AD is widely thought to be driven by the generation and accumulation of the β-amyloid (Aβ) peptide. Even picomolar concentrations of amyloid-beta (Aβ) may trigger a response in astrocytes[188]. Acutely isolated rat astrocytes challenged with Aβ[189] exhibited signs of mitochondrial dysfunction. Aβ also disrupts glutamate uptake by astrocytes[190] and alters Ca^{2+} signalling in astrocytes[191]. Single-nucleus transcriptomic analysis from astrocytes from AD disease brains showed downregulation of genes involved in cell-cell interactions, neuronal development, and synapse regulation[192]. Astrocyte-mediated synapse elimination has also been implicated in synapse loss in AD mouse models[193]. Crucially, apolipoprotein ApoE, a major genetic risk factor for AD, is essential for metabolic coupling between astrocytes and neurons, as discussed previously[165,170].

The functions of astrocytes are also disrupted in PD. The BBB, supported by astrocytic endfeet, is disputed in PD[194]. Pathogenic stimuli, such as α-synuclein aggregates, activate astrocytic NF-κB signalling, leading to the secretion of TNF-α, IL-1β, and other proinflammatory cytokines that may exacerbate neurodegeneration[195]. Reactive astrocytes formed in response to injury are also implicated in PD[196], since they secrete TNF-α and interleukin-1 beta (IL-1β), which are neurotoxic at elevated concentrations[197,198]. Such a secretory subtype of astrocytes is induced by activated microglia[184]. Furthermore, blocking of this switch was shown to prevent PD-like neurodegeneration and behavioural deficits[199].

In MS, astrocytes undergo severe morphological alterations[173] reflected by a distinct transcriptional signature[200] including the upregulation of proinflammatory factors. These molecules secreted by reactive astrocytes were shown to contribute to MS by disrupting tight junctions that maintain the BBB[201]. Studies using an *in vitro* astrocyte model have shown that suppression of crucial astrocytic NF-κB signalling pathway limits MS-like pathology[202]. An intracellular protein produced by astrocytes, Sterile alpha and TIR motif containing 1 (SARM1), promotes neuroinflammation and axonal demyelination through inhibiting glial-derived neurotrophic factor (GDNF[203]).

Astrocytes are also involved in psychiatric diseases, such as SCZ and depression. Cell-level changes in those diseases are much more subtle than the alterations observed in neurodegeneration. However, astrocyte manipulation has been shown to produce SCZ-like symptoms in mice[204], and disruption of the frontal cortex astrocytes has been shown to result in cognitive dysfunction resembling SCZ[205]. Transcriptome studies have indicated that gene expression is dysregulated in astrocytes in SCZ brains, especially in terms of gene isoform changes[206,207]. Furthermore, changes in astrocyte morphology have been described in SCZ patients[208], and astrocyte models derived from IPSCs obtained from SCZ patients have an immunological response and effect on vascularization[209]. This evidence points towards a nuanced, but crucial, role of astrocytes in SCZ.

Collectively, these findings highlight astrocytes as central regulators of brain homeostasis, whose dysfunction contributes to both neurodegenerative and psychiatric disorders. Understanding the mechanisms underlying the disruption of the diverse roles of astrocytes is therefore essential for developing targeted therapies aimed at restoring neural function and preventing disease progression.

1.4. Role of astrocytes in cognition

Astrocytes are pivotal in synaptic plasticity, the process by which activity-dependent, long-lasting changes occur in the efficacy of synaptic transmission. Classically, this is expressed as long-term potentiation (LTP) and long-term depression (LTD), processes that respectively strengthen or weaken synaptic efficacy depending on the pattern and context of pre- and postsynaptic activity[210].

Astrocytes were shown to stabilise hippocampal circuits during learning in mice[211,212]. Their control of neurotransmitter clearance and recycling, via high-affinity glutamate transporters (EAAT/GLT family) and the glutamate–glutamine shuttle discussed earlier, influences glutamatergic NMDA[213] and AMPA[214] receptors, which are central to synaptic plasticity. Furthermore, astrocytes provide metabolic support that is essential for long-term synaptic change. Astrocyte-derived lactate, shuttled to neurons via monocarboxylate transporters (MCT1/4 in astrocytes, MCT2 in neurons), is required for long-term memory formation and maintenance of LTP *in vivo*[163]. Lactate acts both as an energy substrate and as a signalling molecule, promoting synaptic activity-dependent gene expression and supporting the expression of genes involved in learning[215].

Extracellular Ca^{2+} must be maintained within a precise range for LTP to occur[216]. Astrocytes also mediate LTP, and consequently learning, by impacting the extracellular calcium levels, which are crucial for LTP[217]. They achieve this in part through the release of the S100β protein[218], a Ca^{2+}-binding protein frequently used as an astrocytic marker[219]. Experimental evidence regarding the effects of S100β is mixed: mice lacking S100β expression have shown increased LTP[220], whereas optogenetic inactivation of S100β decreased cognitive flexibility in more recent studies[221].

Another pathway through which astrocytes modulate LTP is through the release of D-serine, which is crucial for long-term potentiation[54]. Astrocytes provide neurons with D-serine, the precursor of L-serine, which in turn acts as a co-agonist of NMDA receptors[77,105], through a process dependent on macrovesicle transport via the Soluble N-ethylmaleimide–Sensitive Factor Attachment Protein Receptor (SNARE) proteins[222].

Furthermore, several fascinating studies have demonstrated that astrocytes can be directly manipulated to influence memory formation and decision-making[223]. A group led by Inbal Goshen showed that memory in mice can be enhanced by altering astrocytic calcium dynamics[224]. This was achieved through chemical stimulation of lentivirally-delivered Gq-coupled receptor hM3Dq in CA1 astrocytes of the hippocampus[225], a process dependent on the expression of the transcription factor c-Fos. Similarly, the same laboratory revealed that the activation of the Gi-coupled designer receptor MD4i in CA1 astrocytes impaired the retrieval of remote memories, also associated with reduced c-Fos expression[226]. In a recent study further highlighting the importance of astrocytes for memory formation, spatial transcriptomic analysis of the mouse brain identified memory-specific gene expression signatures in neurons and astrocytes, revealing a population termed “engram astrocytes”, which is implicated in the storage of information[227].These findings collectively underscore that astrocytes are essential regulators of synaptic plasticity and memory encoding.

1.5. Evolution of astrocytes

In order to fulfil the multitude of crucial functions described above, astrocytes have undergone remarkable evolutionary changes, largely paralleling the complexity of nervous systems across species.

Astrocyte-like multifunctional glial cells first appeared in early metazoans. In invertebrates such as Drosophila, glial cells exhibit some astrocyte-like properties, including ion regulation and synaptic support, but lack the full morphological and molecular complexity seen in vertebrates[228–230]. In vertebrates, astrocytes became increasingly specialized. In fish and amphibians, radial glia are the main astroglial type in the brain, and they fulfil both neural stem cell and astrocytic functions[231,232]. Comparative transcriptomic analyses mapping single neural cell diversity in lamprey[233] and in zebrafish[232] suggest that the core astrocytic functions of neurotransmitter uptake, metabolic support, and ionic balance maintenance emerged in ancestral radial glia and later diversified into distinct astrocyte lineages in vertebrates. In zebrafish, glial cells express GFAP, GS, and AQP-4[234], but do not display a reactive phenotype upon injury[235].

While the brains of some sharks contain radial glia alongside numerous astrocyte-like cells[236], genuine parenchymal glial cells (i.e., astrocytes or astrocyte-like cells) appear to have emerged as brain thickness increased during evolution[230]. This might in turn creates the need for more complex supportive cells. In reptiles and birds, which share an evolutionary origin, astrocytes show greater morphological complexity, as well as regional specialization reflecting the requirements for more advanced synaptic regulation and plasticity[237], which supports this view.

Exact characteristics of astrocytes differ between mammalian phylogenetic branches[230,238–240], although detailed studies of glia from non-model species remain scarce. It appears that astrocytes underwent substantial morphological alterations during mammalian evolution. These include morphological diversification and the emergence of new cellular morphotypes of astrocytes. Interlaminar astrocytes constitute a prime example of such a phenomenon: this unique type of astrocyte appears to have emerged in primates[241–243]. Other morphotypes - protoplasmic, varicose projection, and fibrous astrocytes are shared between mammals, yet differ in complexity. For instance, human protoplasmic astrocytes are 2.6-fold larger in diameter and possess ten times more processes than their murine counterparts[243,244]. The interspecies differences in astrocyte morphology far exceed the structural differences observed between neurons from different species[18,245]. Studies of human and chimpanzee organoids, as well as acutely isolated human and primate brain slices have shown that astrocytes have the greatest number of genes expressed in human relative to other compared primate species[246,247]. Gene expression analyses in brain slices also revealed that astrocytes and oligodendrocyte progenitor cells exhibit the largest excess of human-specific expression differences across brain regions[246], with similar results reported by Jorstad et al., (2023) and Ma et al., (2022). The functional significance of evolutionary changes in human astrocytes was further suggested in a remarkable study in which human glial progenitor cells were engrafted into mouse brains; this modification of the recipient animals enhanced their cognitive performance[250].

1.5.1. Astrocyte regulome, and its role in brain evolution

Despite being highly similar in the genomic DNA sequences (98.7% identity between the coding sequences in humans and chimpanzees), humans differ markedly from other primates, particularly in behavioural and cognitive traits [251] and one of the primary tasks of fundamental biology is to delineate which changes in our genomes led to the expansion of the cognitive capacity of our minds. However, the genetic changes that drive the evolutionary expansion of astrocyte morphology, size, and function remain unclear. It is currently not known which mechanisms allow human astrocytes to exert the effect on cognition suggested by Han et al., (2013).

To understand how changes in human and non-human primates (NHP) and other mammals are obtained, we need to identify gene changes that modulate astrocyte functions in evolution. A few genes that emerged in the human lineage and affect human neocortex size have been identified: *ARHGAP11B*[252], *NOTCH2NL*[253], *SRGAP2C*[254], and *TBC1D3*[255]. These loci are predominantly expressed in neural precursors. Remarkably, mutations in several loci gained in the human lineage are also linked to neuropsychiatric disorders, suggesting an intimate relationship between the enhanced cognitive capacity of our brains and disorders of the mind - a link I reviewed previously[256]. However, to date, the list of genes that are specific to the human lineage and active primarily in astrocytes has not been proposed. Likewise, the potential functions of genes that are human-specific and active in astrocytes have not been defined.

Evolutionary changes in gene activity and function may emerge from alterations in its coding sequence[257] or in the DNA regulatory elements that orchestrate its expression[258,259]. As a result, genes with similar expression across species may still differ in function due to divergence in coding sequence[260], while genes with conserved coding sequence may vary in expression level between species due to alterations in regulatory program[247,261,262]. Notably, while changes in protein-coding sequences are rare throughout primate evolution[263,264], species differ in the amounts of proteins produced from the otherwise unaltered loci. Therefore, it has been postulated that a large share of the observed interspecies differences stems from evolutionary changes in the sequence of DNA regulatory elements that orchestrate gene expression[265–268]. Indeed, the non-coding part of genomes of humans and primates contains positively selected sequence differences[269–272], including changes in regulatory elements affecting the activity of loci related to corticogenesis[273]. These findings suggest that evolutionary differences in human cognition likely arise from subtle but consequential genetic and regulatory changes that drive differences in gene expression levels.

The spatiotemporal pattern of gene expression is orchestrated by DNA regulatory elements, which can be classified into four groups: promoters, silencers, insulators, and enhancers[274]. Promoters are DNA sequences flanking the transcription start site (TSS) of a gene. Silencers are regulatory elements negatively affecting gene expression. Enhancers, which are the focus of this work, are a class of regulatory elements that possess the ability to positively modulate gene expression in a cell type-specific manner. Insulators are elements that in mammals are bound by the CTCF transcription factor[275]. Insulators preclude interaction between other regulatory elements. Enhancers are located distally from the genes they regulate ("target genes"). Enhancers are identified using two chromatin features: open chromatin, which can be described using Assay for Transposase-Accessible Chromatin (ATAC-seq)[276], and acetylation of lysine 27 in the tail of histone 3, which is assayed using chromatin immunoprecipitation followed by sequencing (ChIP-seq) [277,278]. Enhancers act as platforms for transcription factor binding[279] and, according to the current model, enhancers come into physical contact with their target promoters[280]. Such interactions are mapped using conformation capture maps, such as Hi-C assays[281]. Chromatin conformation capture experiments have revealed partitioning of the mammalian genomes into regions of strong self-contacts termed topologically associated domains (“TADs”)[282–284]. Importantly, cognate promoter–enhancer pairs tend to reside within the same TAD, supporting the view that TADs constitute genuine functional units of genome organisation[285]. The regulome changes that drive the evolutionary expansion of astrocyte morphology, size, and function remain unclear and **constituted the main motivation for the study presented in this work.**

In a recently published study, which forms part of this work, we identified crucial changes in the astrocyte regulome that are specific to human astrocytes[286]. In this study, we employed an iPSC-derived astrocyte model (iAstrocytes) from humans and two primate species - chimpanzee (*Pan troglodytes*) and rhesus macaque (*Macaca mulatta*) - to identify transcription factors driving differential gene expression between human and chimpanzee astrocytes.

1.6. Modelling astrocytes and their regulome

1.6.1. IPS technology

The ability to reprogram fibroblasts into induced pluripotent stem cell technology (iPSC)[287] enables the establishment of *in vitro* models of cell types that would be otherwise difficult to culture or extract. The iPSC-based cell models have found application in disease modelling and drug discovery[288]. In the context of evolutionary molecular biology, iPSC models allow the otherwise impossible analyses of cells from endangered and rare species, such as NHP, including the phylogenetic branch that is closest to humans, the chimpanzees. This is especially pertinent due to the very restrictive Convention on International Trade in Endangered Species of Wild Fauna and Flora[289]. Furthermore, iPSC-derived cells allow application of the most up-to-date techniques of genome modification, perturbation, and engineering[290], allowing direct verification of biological hypotheses on the molecular level, and testing predictions of bioinformatic analyses, which are mostly correlative, for example, using CRISPR-based screens[291]. Over the past six years, our laboratory has developed unique iPSC-derived models of astrocytes (iAstrocytes), which, in combination with the most modern genome engineering techniques, allow us to perform state-of-the-art genetic experiments verifying enhancer function. Those models were published in a publication that I co-authored[286]**.**

1.6.2. Massively Parallel Reporter Assay (MPRA)

Evolutionary changes in predicted enhancer activity are detected based on analysis of ATAC-seq and ChIP-seq data[273], including advanced single-cell techniques[247], as well as by intersecting genomic data with more detailed sequence characterisation[292]. However, functional verification of enhancer function is by and large lacking, precluding broader inference, limiting the ability to draw more general conclusions. **This work aims to help bridge that gap.**

High-throughput reporter assays (Massively Parallel Reporter Assay, MPRA) allow for testing the activity of thousands of putative enhancer sequences in parallel. Different variations of MPRA rely on the same principle. In short, candidate enhancer fragments are inserted into a reporter construct, creating a library which, upon delivery into cells, allows tracking of levels of the reporter transcript depending on which candidate sequence is inserted into a given plasmid. Different kinds of MPRA were used to reveal the activity of enhancers in neural progenitor cells[293], during neuronal differentiation[294], and allowed for comparison of enhancer activity in human and chimpanzee neural cells[295].

1.6.3. CRISPRi technology

Enhancer perturbation, which complements and verifies observations inferred from MPRA, can be achieved either by directly editing the genome using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 technology, or by using CRISPR-interference (CRISPRi) epigenetic modifiers[296]. The CRISPRi approach utilises a catalytically dead Cas9 nuclease (dCas9) to target chosen chromatin modifiers to a specific genomic locus. The first version of this system employed a covalently bound Krüppel-associated box (KRAB) domain[297], which recruits cofactors that deposit histone H3 lysine 9 trimethylation and induce local histone deacetylation, collectively silencing the target sequence. This system has since been enhanced by including MeCP2 binding, which strengthens repression[298]. The dCas9-Krab-MeCP2 lentiviral construct has been successfully used to abolish activity of enhancers, in example the enhancer of *NSMCE3* gene in acute myeloid leukaemia model cell line (Kasumi-1)[299], enhancers of *DUSP5*, *SMAD5*, *EDN1* in human pulmonary artery endothelial cells[300] and enhancer of *Myh* mouse muscle progenitor cells[301].

During my doctoral project, I applied the tools described above, particularly MPRA to study enhancers in human iAstrocytes. This generated results forming the foundation of this work, in which I aim to determine which enhancer elements gained in the human lineage -identified through our genomics assays (ChIP-seq, RNA-seq, ATAC-seq, and Hi-C) - are active enhancers capable of driving reporter gene expression, and to infer genome-wide trends that govern their activity.

2. Aims of this work

Human brains have evolved to support complex cognitive functions, yet the genetic and cellular mechanisms underlying their exceptional size and complexity are still unclear. Human astrocytes are unusually large and morphologically complex, suggesting a potential role in shaping uniquely human brain features. However, it is not known how astrocytes contribute to brain development and morphology, and which genetic changes drive the evolutionary expansion of astrocyte size, form, and function. Changes in regulatory elements, particularly in enhancers, which increase gene expression, is thought to play a key role.

The transcription factors central to astrocyte gene regulation are not fully known. Previous work from our laboratory suggested that the gain of binding sites for Universal Stripe Factors (USFs) in enhancers near genes more highly expressed in human then non-human primate astrocytes[286]. Such changes may underlie the emergence of human-specific regulatory activity, but this hypothesis required experimental validation. Using induced pluripotent stem (iPS) cell-derived astrocytes (iAstrocytes) from human, chimpanzee, and macaque, combined with high-throughput genomic and bioinformatic tools, this work directly addresses these questions.

In particular, I aimed to answer the following questions:

a. Which transcription factors (TFs) best predict enhancer activity in foetal astrocytes, and what is the role of USF in astrocyte regulome?
b. Which brain traits and disorders are linked to foetal astrocyte functions?
c. Can we link disease-associated single-nucleotide polymorphisms (SNPs) with possible changes in the activity of astrocyte enhancers?
d. Is the gain of USF binding sites truly responsible for the increased activity of human-specific enhancers?

3. Materials and methods

iAstrocyte differentiation and genomic characterization was performed by Dr. Katarzyna Ciuba and Aleksandra Piotrowska. Intact Hi-C was performed by Dr. Anna Stroynowska-Czerwińska. Debadeep Chaudhury contributed to computational processing and performed differential RNA-seq analysis.

3.1. Materials

3.1.1. List of reagents

Reagent or resource	**Source**	**Identifier**
dNTP Set (100 mM)	ThermoFisher Scientific	Cat#10297018
Platinum™ SuperFi II DNA Polymerase	ThermoFisher Scientific	Cat#12361010
UltraPure™ DNase/RNase-Free Distilled Water	Invitrogen/ ThermoFisher Scientific	Cat#10977035
Agarose-out kit	EurX	Cat#E3540-1
Quick Ligation™ Kit	New England Biolabs (New England Biolabs)	Cat#M2200S
New England Biolabs® Stable Competent E. coli	New England Biolabs (New England Biolabs)	Cat#C3040I
Corning® Treated Culture Dishes	Corning	
DMEM/F-12, GlutaMAX™ supplement	Gibco/ ThermoFisher Scientific	Cat#10565018
Fetal Bovine Serum	SIGMA	Cat#F9665
hESC-qualified Matrigel	Corning	Cat#354277
Accutase	Merck	Cat#A6964
DMEM Low Glucose w/ Stable Glutamine w/Sodium Pyruvate	Biowest	Cat#L0066-500
Penicillin-Streptomycin (10,000 U/mL)	ThermoFisher Scientific	Cat#15140122
Gelatin	Merck	Cat#G9391

High-glucose DMEM	Biowest	Cat#L0104-500
Trypsin-EDTA	Merck	Cat#T3924
ReproRNA-OKSGM Kit	Stemcell Technologies	Cat#050930
Advanced DMEM	ThermoFisher Scientific	Cat#12491015
L-glutamine	ThermoFisher Scientific	Cat#A2916801
Opti-MEM Reduced-Serum Medium	ThermoFisher Scientific	Cat#31985070
STEMdiff Neural Progenitor Medium	Stemcell Technologies	Cat#05833
STEMdiff Neural Induction Medium	Stemcell Technologies	Cat#08581
STEMdiff Neural Rosette Selection Reagent	Stemcell Technologies	Cat#05832
BV421 Mouse Anti-Human CD271	BD Biosciences	Cat#562562; RRID: AB_2737657
Mouse monoclonal CD133/1 Antibody, anti-human, PE	Miltenyi Biotec	Cat#130-113-108; RRID: AB_2725937
APC Mouse Anti-Human CD184 Clone 12G5	BD Biosciences	Cat#555976; RRID: AB_398616
ScienCell AM	Sciencell	Cat#1801
Rabbit polyclonal anti-S100 Beta	Proteintech	Cat#15146-1-AP / AB_2254244
Mouse monoclonal anti-Vimentin	DSHB	Cat#AMF-17b / AB_528505
Rabbit polyclonal anti-SLC1A3	Proteintech	Cat#20785-1-AP / AB_2878738
Formaldehyde solution, 37%	SIGMA - MERCK	Cat#158127
Triton X100	SIGMA - MERCK	Cat#X100-100ML
Albumin, Bovine Serum	BioShop	Cat#ALB001
Goat anti-Mouse IgG (H+L), Alexa Fluor 488	ThermoFisher Scientific	Cat#A-11001 / AB_2534069
Goat anti-Mouse IgG (H+L), Alexa Fluor 568	ThermoFisher Scientific	Cat#A-11011 / AB_143157
Goat anti-Mouse IgG (H+L), Alexa Fluor 647	ThermoFisher Scientific	Cat#A-21244 / AB_2535812

Hoechst	ThermoFisher Scientific	Cat#34580
Prolong Diamond Antifade Mountant	ThermoFisher Scientific	Cat#P36961
Hanks' Balanced Salt solution Modified, with sodium bicarbonate, without phenol red, calcium chloride and magnesium sulphate, liquid, sterile-filtered, suitable for cell culture	SIGMA	Cat#H6648-500ML
DPBS	Biowest	
Fast SYBR Green	ThermoFisher Scientific	Cat#4385616
Glutamate-Glo Assay kit	Promega	Cat#J7021
Fluo-4AM	ThermoFisher Scientific	Cat#F14201
Fluo-4AM	ThermoFisher Scientific	Cat#F14201
ATP	ThermoFisher Scientific	Cat#R0441
Trans-well inserts	VWR	Cat#734-2723
AraC	Sigma Aldrich	Cat#C6645
LIVE/DEAD Viability/Cytotoxicity Kit	ThermoFisher Scientific	Cat#L3224
Polyethylenimine, Linear, MW 25000,Transfection Grade (PEI 25K)	Polysciences Europe GmbH	Cat#23966-100
Lenti-X™ Concentrator	TaKaRa	Cat#631232
ViralBoost reagent	Alstem	Cat#960-VB100
Polybrene	Merck	Cat#TR-1003-G
2-butanol	ThermoFisher Scientific	Cat#107700010
Agencourt AMPure XP	Beckman Coulter)	Cat#A63881
TegoSoft DEC	Evonik	Cat#P021B09302
ABIL WE	Evonik	Cat#E522311955
Mineral Oil	SIGMA	Cat#M5904
PEG 8000 Molecular Biology Grade	Promega	Cat#V3011
Buffer EB (250 ml)	Qiagen	Cat#19086
LB	AA Biotechnology	Cat#2020-250
Carbenicillin, disodium salt	VWR	Cat#A1491.0005
Plasmid Midi AX	AA Biotechnology	Cat#092-10

SfiI	New England Biolabs	Cat#R0123
BmgBI	New England Biolabs	Cat#R0628
MluI	EurX	Cat#E2288-01
10-beta Electrocompetent E. Coli	New England Biolabs	Cat#C3020K
rCutSmart Buffer (10x)	New England Biolabs	Cat#B6004S
GeneArt™ Gibson Assembly HiFi Master Mix	Invitrogen/ThermoFisher Scientific	Cat#A46628
Gene Pulser/MicroPulser Electroporation Cuvettes, 0.2 cm gap	Bio-Rad	Cat#1652086
SOC medium	ThermoFisher Scientific	Cat#15544034
Plasmid MIDI AX kit	AA Biotechnology	Cat#092-10
MfeI-HF	New England Biolabs	Cat#R3589
KpnI-HF	New England Biolabs	Cat#R3142
New England BiolabsNext Ultra II Q5® Master Mix	New England Biolabs	Cat#M0544S
Q5® High-Fidelity DNA Polymerase	New England Biolabs	Cat#M0491
Q5® buffer	New England Biolabs	Cat#B9027
AsiSI	New England Biolabs	Cat#R0630
Exonuclease V (RecBCD)	New England Biolabs	Cat#M0345
ATP	New England Biolabs	Cat#P0756S
New England Biolabsuffer™ 4	New England Biolabs	Cat#B7004S
Lipofectamine™ Stem Transfection Reagent	ThermoFisher Scientific	Cat#STEM00015
SuperScript III Reverse transcriptase	Life Technologies	Cat#18080093
Direct-zol™ RNA MiniPrep kit	Zymo	Cat#R2052
SUPERase•In™ RNase Inhibitor	ThermoFisher Scientific	Cat#AM2694
SYBR™ Green I Nucleic Acid Gel Stain, 10,000X concentrate in DMSO	ThermoFisher Scientific	Cat#S7567
Hard-Shell® 384-Well PCR Plates	Bio-Rad	Cat#HSP3905

New England BiolabsNext® High-Fidelity 2X PCR Master Mix	New England Biolabs	Cat#M0541
TRI-reagent	SIGMA - MERCK	Cat#T9424
Agilent RNA 6000 Nano Kit	Agilent	Cat#5067-1511
KAPA mRNA HyperPrep kit	Roche	Cat#8098115702
LEGENDplex HU Proinflam. Chemokine Panel	Biolegend	Cat#740985
Tape station 4200	Agilent	Cat#G2991BA
UMI in xGen UDI-UMI Adapters	IDT	Cat#10005903
ATAC seq kit	Active Motif	Cat#53150
Rabbit polyclonal anti-histone H3 (tri methyl K4)	Abcam	Cat#ab213224; RRID: AB_2923013
Rabbit monoclonal anti-histone H3 (acetyl K27)	Cell Signalling	Cat#8173S; RRID: AB_10949503
Ovation Ultralow System V2	Tecan	Cat#0344NB-32
NEBuilder® HiFi DNA Assembly Master Mix	New England Biolabs	Cat#E2621L
repliQa HiFi ToughMix	Quantabio	Cat#95200-025)
T4 DNA Ligase Reaction Buffer	New England Biolabs	Cat#B0202S
T4 Polynucleotide Kinase	New England Biolabs	Cat#M0201S
Quick CIP	New England Biolabs	Cat#M0525S
Dual-Luciferase Reporter Assay System	Promega	Cat#E1910
BbsI-HF	New England Biolabs	Cat#R3539L
BsmBI-v2	New England Biolabs	Cat#R0739L
Buffer 3.1	New England Biolabs	Cat#B6003S
Buffer CutSmart	New England Biolabs	Cat#B6004S
Puromycin	ThermoFisher Scientific	Cat# A1113803
Blasticidin	VWR	Cat#CAYM14499-10
High-Capacity cDNA Reverse Transcription Kit	ThermoFisher Scientific	Cat#4368814

16% methanol-free formaldehyde	ThermoFisher Scientific	Cat# 28908
EGTA	SIGMA	E4378-10G
Biotin-11-dUTP 5 x 200 µl (1 mM)	Jena Bioscience	Cat# NU-803-BIOX-L
Polymerase I Klenow Fragment	New England Biolabs	Cat# M0210S
dNTP Mix	ThermoFisher Scientific	Cat# R0192
Dynabeads™ MyOne™ Streptavidin T1	ThermoFisher Scientific	Cat#65601
NEB Next II Ultra Prep kit	New England Biolabs	Cat# E7645S
KAPA HiFi HotStart ReadyMix	Roche	Cat#KK2601
DSG	ThermoFisher Scientific	Cat#A35392
ChIC/CUT&RUN pAG-MNase	Symbios	Cat#53181
IGEPAL® CA-630	Merck	Cat# I3021-100ML
cOmplete™, EDTA-free Protease Inhibitor Cocktail	Roche	Cat#11873580001

3.1.2. Cell lines

Cell lines used in this work were derived by Dr. Katarzyna Ciuba and Aleksandra Piotrowska. They were published in Ciuba et al., (2025).

3.1.2.1. Induced Pluripotent Stem cell lines

All the iPS cell lines were female and fibroblast-derived. Human iPS cell lines: "ELE10" (IIMCBi001-A) and "ELE30" (IIMCBi002-A)[302] were a gift from Prof. J. Jaworski and "AG9319", and "AG9429" iPS cell lines were a gift from Prof. F. Gage[303]. The Chimpanzee iPS cell line "SandraA" was a gift from Prof. S. Pääbo. Chimpanzee iPS cell lines "Mandy04" and "Mandy06" were generated in Dioscuri Centre for Chromatin Biology and Epigenomics in Nencki Institute, in collaboration with the Warsaw ZOO. Rhesus macaque iPS cell line "Becky" was a gift from Prof. R. Behr (German Primate Centre, Göttingen).

3.1.2.1. Cancer lines

Cancer lines used of MPRA screens were commercially available LN-229 (ATCC catalogue number CRL-2611). The lines were a shared by A. Ellert Miklaszewska from Laboratory of Molecular Microbiology at Nencki Institute. HEK 293T cells were a gift from M. Krakowczyk (Laboratory of Protein Homeostatsis, Nencki Institute/National Institute of Oncology).

3.1.3. Plasmids

The following plasmids were used in this study:

Plasmids lenti_dCas9-KRAB-MeCP2,, pMPRAv3:Δluc:Δxbal, pMPRAv3:minP-GFP pLS-SV40-mP-EGFP were ordered from Addgene (reference numbers 122205, 109035, 109036, 137724, respectively).

Plasmid px_laila was a gift from Laila El Khattabi (Paris Brain Institute).

Plasmids VSV.G and PsPax2 were a gift from Toufic Kassouf (Nencki Institute).

Plasmids pGL4.10{luc2} and pRL_SV40 were obtained from Promega.

Plasmid px_laila_BFP was generated by restriction digest of px-Laila plasmid with restriction enzymes BmgBI (New England Biolabs), and MluI (EurX), purified on TAE agarose gel, excised, and purified using Agarose-out kit (EurX), and eluted in water. The insert, encoding blue fluorescent protein BFPTag2, was amplified from plasmid pHAGE-EFS-MCP-3XBFPnls (Addgene, 75384), in three parts (primer pairs ED228_BFP_TA2A_F/ED234_BbsI_removal_R, ED233_BbsI_removal_F/ED232_BsmBi_removal_R, and ED231_BsmBi_removal_F/ED230_BFP_for_Laila_R). PCR reaction was assembled on ice as follows F primer (5 µM), 1 µl, R_primer (5 µM), 1 µl, dNTPs (5 µM), 1.6 µl, 5X SuperFi II Buffer (ThermoFisher Scientific), 8 µl, Platinum SuperFi II DNA Polymerase (ThermoFisher Scientific), 0.4 µl, plasmid pHAGE-EFS-MCP-3XBFPnls, 50 ng, UltraPure water (Invitrogen) to 50 µl. The reaction mixture was then incubated in a thermocycler according to the following conditions: step 1: 98°C, 30 seconds. step 2: 98°C, 10 seconds, step 3: 60°C, 30' seconds, step 4: 72°C for 20 seconds, return to step 2, 3x, step 5: 72°C, 5 minutes. Each reaction mixture was then ran on 2% TAE agarose gel, extracted, and purified using Agarose-out kit (EurX), and eluted in water.

Overlap extension ("stitch") PCR was then performed as follows: PCR reaction mix was assembled on ice as follows: primer ED228_BFP_TA2A_F (5µM), 0.5µl, ED230_BFP_for_Laila_R (5 µM), 0.5 µl, dNTPs (5 µM), 0.8 µl, 5X SuperFi II Buffer (ThermoFisher Scientific), 4 µl, Platinum SuperFi II DNA Polymerase (ThermoFisher Scientific), 0.2 µl, 1 µl of each amplicon (ED228_BFP_TA2A_F/ED234_BbsI_removal_R, ED233_BbsI_removal_F/ED232_BsmBi_removal_R, and ED231_BsmBi_removal_F/ED230_BFP_for_Laila_R), UltraPure water (Invitrogen) to 25 µl. The reaction mixture was then incubated in a thermocycler according to the following conditions: step 1: 98°C, 30 seconds. step 2: 98°C, 10 seconds, step 3: 60°C, 30' seconds, step 4: 72°C for 20 seconds, return to step 2, 34x, step 5: 72°C, 5 minutes. Amplicon size was tested on agarose gel, and rest of the mixture was then purified using AMpure XP beads using 1.8:1 bead:sample ratio. The insert was then digested with MluI restriction enzyme (EurX) purified on TAE agarose gel, excised, and purified using Agarose-out kit (EurX), and eluted in water.

Plasmid was ligated using Quick Ligation™ Kit (New England Biolabs) according to manufacturer's instructions. The ligated plasmid was transformed into New England Biolabs® Stable Competent E. coli (New England Biolabs).

Plasmid 109035_SV40_GFP was generated by restriction digest of pGL4 MP with restriction enzyme SfiI (New England Biolabs) purified on TAE agarose gel, excised, and purified using Agarose-out kit (EurX), and eluted in water. The insert was prepared as follows: SV40 was amplified from plasmid pLS-SV40-mP-EGFP (Addgene, 137724) using primers R71_SV40_for_109035_R/R72_SV40_for_109035_F. PCR reaction was assembled on ice as follows F primer (10 µM), 2.5 µl. R_primer (10µM), 2.5 µl, dNTPs (10 µM), 1 µl, Q5® buffer (New England Biolabs), 0.5 µl of Q5 polymerase (New England Biolabs) 50 ng of plasmid pLS-SV40-mP-EGFP, UltraPure water (Invitrogen) to 50 µl. The reaction mixture was then incubated in a thermocycler according to the following conditions: step 1: 98°C, 30 seconds. step 2: 98°C, 10 seconds, step 3: 70°C, 30' seconds, step 4: 72°C for 30 seconds, return to step 2, 34x, step 5: 72°C, 5 minutes. Part of the reaction mixture was then ran on 1% TAE agarose gel to check size, and the rest was purified using), Agencourt AMPure XP beads (Beckman Coulter) at 1.8:1 bead:sample ratio. Minimal promoter-EGFP cassette was amplified as follows: 0.5 µM of primer R21_ED_MPRA_GFP_F, 0.5 µM of primer R22_ED_MPRA_GFP_R, 0.2 µM of dNTPs (ThermoFisher Scientific). 5x Q5® buffer (New England Biolabs), 10 µl, Q5® High-Fidelity DNA Polymerase (New England Biolabs), 0.5 µl, pMPRAv3:minP-GFP, 50ng, UltraPure water (Invitrogen) to 50 µl. The reaction mixture was then incubated in a thermocycler according to the following conditions: step 1: 98°C, 30 seconds. step 2: 98°C, 10 seconds, step 3: 69°C, 30 seconds, step 4: 72°C for 30 seconds, return to step 2, 35x, step 5: 72°C, 2 minutes. 5µl of the amplicon mixture was tested on agarose gel – band of 843 bp was expected. The rest was purified using Agencourt AMPure XP beads (Beckman Coulter)) according to manufacturer's protocol for purification of small fragments, using 0.65:1 bead:sample ratio. DNA was eluted in buffer EB (Qiagen). The two fragments were then combined using overlap extension ("stitch") PCR using primers R73_MPRA_GFP_for_109035/R72_SV40_for_109035_F. PCR reaction was assembled on ice as follows: F primer (10 µM), 2.5 µl, R_primer (10 µM), 2.5 µl, dNTPs (10 µM), 1 µl, Q5® buffer (New England Biolabs), 0.5 µl of Q5 polymerase (New England Biolabs), fragments SV40 and mP-EGFP, UltraPure water (Invitrogen) to 50 µl. The reaction mixture was then incubated in a thermocycler according to the following conditions: step 1: 98°C, 30 seconds. step 2: 98°C, 10 seconds, step 3: 66°C, 45' seconds, step 4: 72°C for 30 seconds, return to step 2, 34x, step 5: 72°C, 5 minutes. Part of the reaction mixture was then ran on 1% TAE agarose gel to check size, and the rest was purified using Agencourt AMPure XP beads (Beckman Coulter) at 1.8:1 bead:sample ratio.

Plasmid was ligated using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) according to manufacturer's instructions. The ligated plasmid was transformed into New England Biolabs® Stable Competent E. coli (New England Biolabs).

Plasmid pGL4 MP was generated by restriction digest of pGL4 MP with restriction enzyme SfiI (New England Biolabs). The insert was amplified using primers Y65_mP_r/Y66_blunt_end_luc_plasmid. PCR reaction was assembled on ice as follows: F primer (3 µM), 3 µl, R primer (3 µM), 3 µl, Template 2 µl, 100 ng of template plasmid 109035_SV40_GFP, UltraPure water (Invitrogen) to 50µl, 50µl of repliQa HiFi ToughMix (Quantabio). The reaction was split into two 50µl aliquots, and incubated in a thermocycler according to the following conditions: step 1: 98°C, 30 seconds. step 2: 98°C, 10 seconds, step 3: 55°C, 10 seconds, step 4: 68°C 10 seconds, return to step 2, 34x, step 5: 68°C, 2 minutes. The amplicon was purified on TAE agarose gel, excised, and purified using Agarose-out kit (EurX), and eluted in water.

Plasmid was ligated using New England Biolabsuilder® HiFi DNA Assembly Master Mix (New England Biolabs) according to manufacturer's instructions. The ligated plasmid was transformed into New England Biolabs® Stable Competent E. coli (New England Biolabs).

Preparation of CRISPRi plasmids is described below:

The following scrambled sequences were used: non-targeting_00000 F, non-targeting_00000 R, non-targeting_00001 F, non-targeting_00001 R[304].

Plasmids were cloned as follows: oligonucleotides were ordered as extended primers in which the sgRNA sequences were incorporated as reverse-complement inserts positioned between constant flanking regions. The primers are designed such that, following amplification, the resulting product reconstitutes the dual-guide architecture of the px-laila plasmid, with the two sgRNAs oriented away from each other. Cloning was carried out as follows: backbone was digested: 6 ug px-laila-BFP, 2µl of BbsI-HF (New England Biolabs) and BsmBI-v2 (New England Biolabs), 10µl of buffer 3.1 (New England Biolabs) and UltraPure water (Invitrogen) to 100 µl. The mixture was incubated at 37°C overnight. The following day, the cut plasmid analysed on TAE agarose gel. Two bands were extracted: the backbone: 8906 bp, and the insert 578bp, and purified using Agarose-out kit (EurX), and eluted in water. The insert band was used as template to add sgRNA sequences. The PCR reaction was assembled as follows: : F primer (3µM), 1,5µl, R primer (3µM), 1,5µl, Template 1µl, 20ng of template plasmid 109035_SV40_GFP, UltraPure water (Invitrogen) to 25µl, 25µl of repliQa HiFi ToughMix (Quantabio). Reactions were incubated in a thermocycler according to the following conditions: step 1: 98°C, 30 seconds. step 2: 98°C, 10 seconds, step 3: 55°C, 10 seconds, step 4: 68°C 10 seconds, return to step 2, 34x, step 5: 68°C, 2 minutes. The amplicon was purified on TAE agarose gel, excised, and purified using Agarose-out kit (EurX), and eluted in water. The insert was then cloned back into previously purified backbone by NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) according to manufacturer's instructions. The ligated plasmids were transformed into New England Biolabs® Stable Competent E. coli (New England Biolabs).

Luciferase assay plasmids were amplified using primers specified in the table below. In case of blunt end cloning, primers were pre-phsphorylated as follows: 2µl of 100uM F primer, 2µl of 100uM R primer, 2µl of 10x T4 ligatse buffer (New England Biolabs), 1µl of T4 polynucleotide kinase, UltraPure water (Invitrogen) to 20 µl, and incubated for 30 minutes at 37°C, and for 5 minutes at 95°C.

PCR reactions were assembled on ice as follows: F primer (3µM), 1,5µl, R primer (3µM), 1,5µl, Template 1µl, 20ng of template plasmid 109035_SV40_GFP, UltraPure water (Invitrogen) to 25µl, 25µl of repliQa HiFi ToughMix (Quantabio). Reactions were incubated in a thermocycler according to the following conditions: step 1: 98°C, 30 seconds. step 2: 98°C, 10 seconds, step 3: 55°C, 10 seconds, step 4: 68°C 10 seconds, return to step 2, 34x, step 5: 68°C, 2 minutes. The amplicon was purified on TAE agarose gel, excised, and purified using Agarose-out kit (EurX), and eluted in water. I case of constructs Rho_E3_HS and Rho_E1_HS_PT1PT2, 100 ng of DNA string was directly phosphorylated (2µl of 10x T4 ligase buffer (New England Biolabs), 1µl of T4 polynucleotide kinase, UltraPure water (Invitrogen) to 20µl, incubation for 30minutes at 37°C, and for 5 minutes at 95°C). Destination plasmid pGL4 MP was prepared by overnight digest, as follows: 3ug of plasmid and 2μl of BmgBI (New England Biolabs) and 1μl of CIP in buffer 3.1 (New England Biolabs) were incubated overnight at 37°C. Cut plasmid was purified on TAE agarose gel, excised, and purified using Agarose-out kit (EurX), and eluted in water. Plasmids were cloned either using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) according to manufacturer's instructions or using Quick Ligation™ Kit (New England Biolabs). The ligated plasmids were transformed into New England Biolabs® Stable Competent E. coli (New England Biolabs).

Plasmid	Template	F primer	R primer
Cloned using blunt end cloning			
Rho_E1_HS_PT1	Rho_E1_HS_PT1	AC45_Rho_R1_HS_F	AC46_Rho_R1_HS_R
Rho_E1_HS_PT1PT4_PCR_gen	Rho_E1_HS_PT1	AC45_Rho_R1_HS_F	AC66_Rho_R1_HS_PT4_R
Rho_E1_HS_PT2	Rho_E1_HS_PT2	AC45_Rho_R1_HS_F	AC46_Rho_R1_HS_R
Rho_E1_HS_PT2PT4_PCR_gen	Rho_E1_HS_PT2	AC45_Rho_R1_HS_F	AC66_Rho_R1_HS_PT4_R
Rho_E1_HS_PT1PT2PT4_PCR_gen	Rho_E1_HS_PT1PT2	AC45_Rho_R1_HS_F	AC66_Rho_R1_HS_PT4_R
Rho_E1_HS_PT4_PCR_gen	Rho_E1_HS	AC45_Rho_R1_HS_F	AC66_Rho_R1_HS_PT4_R
Cloned using NEBuilder			
Rho_E1_HS_PT2PT3	Rho_E1_HS_PT2PT3	AC47_Rho_R1_PT3_R	AC48_Rho_E1_HS_New England Biolabs_F
Rho_E1_HS_PT3	Rho_E1_HS_PT3	AC47_Rho_R1_PT3_R	AC48_Rho_E1_HS_New England Biolabs_F
Rho_E1_HS_PT3PT4_PCR_gen	Rho_E1_HS_PT3	AC64_Rho_R1_PT3_PT4_R	AC48_Rho_E1_HS_New England Biolabs_F
Rho_E1_HS_PT1PT2PT3	Rho_E1_HS_PT1PT2PT3	AC47_Rho_R1_PT3_R	AC48_Rho_E1_HS_New England Biolabs_F

Rho_E1_HS_PT1PT3PT4_PCR_gen	Rho_E1_HS_PT1PT3	AC64_Rho_R1_PT3_PT4_R	AC48_Rho_E1_HS_New England Biolabs_F
Rho_E1_HS_PT2PT3PT4_PCR_gen	Rho_E1_HS_PT2PT3	AC64_Rho_R1_PT3_PT4_R	AC48_Rho_E1_HS_New England Biolabs_F
Rho_E1_HS_PT1PT2PT3PT4_PCR_gen	Rho_E1_HS_PT1PT2PT3	AC64_Rho_R1_PT3_PT4_R	AC48_Rho_E1_HS_New England Biolabs_F

3.1.3.1 List of plasmids

Plasmid	**Source/Backbone**	**Use**
Externally obtained plasmids		
px-laila	Laila El Khattabi (Paris Brain Institute).	Backbone for px-laila-BFP
pHAGE-EFS-MCP-3XBFPnls	Addgene, plasmid number 75384	Source of BFPTag2 sequence
lenti_dCas9-KRAB-MeCP2	Addgene, plasmid number 122205	CRISPRi lentiviral construct
pMPRAv3:Δluc:Δxbal	Addgene, plasmid number 109035	MPRA destination plasmid
pMPRAv3:minP-GFP	Addgene, plasmid number 109036	Source of mP-GFP insert in preparation of MPRA libraries
pLS-SV40-mP-EGFP	Addgene, plasmid number 137724	Source plasmid for DNA fragments.
VSV.G	Toufic Kassouf	Lentiviral envelope plasmid
psPAX2	Toufic Kassouf	Lentiviral packaging plasmid
pGL4.10{luc2}	Promega, E6651	Backbone for luciferase reporter assay

pRL_SV40	Promega, E223A	Normalisation of luciferase experiments
Plasmids generated for this study		
px_laila_BFP	px-laila	Backbone for sgRNA perturbation
109035_SV40_GFP	pMPRAv3:Δluc:ΔxbaI	Control plasmid for MPRA trasnfections
pGL4 MP	pGL4.10{luc2}	Backbone for luciferase assay
Rho_E1_HS	pGL4 MP	Luciferase assay
Rho_E1_HS_PT1	pGL4 MP	Luciferase assay
Rho_E1_HS_PT1PT2	pGL4 MP	Luciferase assay
Rho_E1_HS_PT1PT3	pGL4 MP	Luciferase assay
Rho_E1_HS_PT1PT4_PCR_gen	pGL4 MP	Luciferase assay
Rho_E1_HS_PT2	pGL4 MP	Luciferase assay
Rho_E1_HS_PT2PT3	pGL4 MP	Luciferase assay
Rho_E1_HS_PT2PT4_PCR_gen	pGL4 MP	Luciferase assay
Rho_E1_HS_PT3	pGL4 MP	Luciferase assay
Rho_E1_HS_PT3PT4_PCR_gen	pGL4 MP	Luciferase assay
Rho_E1_HS_PT1PT2PT3	pGL4 MP	Luciferase assay
Rho_E1_HS_PT1PT3PT4_PCR_gen	pGL4 MP	Luciferase assay
Rho_E1_HS_PT2PT3PT4_PCR_gen	pGL4 MP	Luciferase assay
Rho_E1_HS_PT1PT2PT4_PCR_gen	pGL4 MP	Luciferase assay
Rho_E1_HS_PT1PT2PT3PT4_PCR_gen	pGL4 MP	Luciferase assay
Rho_E1_HS_PT4_PCR_gen	pGL4 MP	Luciferase assay
X54_X58_110792_1_110792_2_bfp	px_laila_BFP	CRISPRi
X57_X61_rhobtb3_negcontr_bfp	px_laila_BFP	CRISPRi
Snp plasmid	px_laila_BFP	CRISPRi

3.1.4. List of primers

Name	Sequence	Use
ED228_BFP_TA2A_F	GTGGAGGAGAATCCCGGCCCTATGGTGTCTAAGGGCGAAGAGC	Amplification of TagBFP2 cassette
ED234_BbsI_removal_R	CACGCCCCCGTCCTCGTATG	Amplification of TagBFP2 cassette, site-directed mutagenesis of BbsI site
ED233_BbsI_removal_F	CATACGAGGACGGGGGCGTG	Amplification of TagBFP2 cassette, site-directed mutagenesis of BbsI site
ED232_BsmBi removal_R	CGGGGTACAGCGTTTCGGTGA	Amplification of TagBFP2 cassette, site-directed mutagenesis of BsmBI site
ED231_BsmBi removal_F	TCACCGAAACGCTGTACCCCG	Amplification of TagBFP2 cassette, site-directed mutagenesis of BsmBI site
ED230_BFP_for_Laila_R	ATCCAGAGGTTGATTGTCGACTTAACGCGTTTAATTAAGCTTGTGCCCCAGTTTGC	Amplification of TagBFP2 cassette
R19_MPRA_v3_20I_R	CCGACTAGCTTGGCCGCCGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNTCTAGAGGTTCGTCGACGCGATCGCAGG	Emulsion PCR, barcoding
R20_MPRA_v3_F	GCCAGAACATTTCTCTGGCCTAACTGGCCGCTTGACG	Emulsion PCR
R21_ED_MPRA_GFP_F	CACTGCGGCTCCTGCGATGGCGGCCAAGCTTAGACAC	Amplification of minimal promoter - GFP cassette
R22_ED_MPRA_GFP_R	TCTAGAGGTTCGTCGACGCGA	Amplification of minimal

	TTTACTTGTACAGCTCGTCCATGCC	promoter - GFP cassette
R48_MPRAv3_Illu_GFP_F	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGCCCTGAGCAAAGACC	Quantification of cDNA transcripts
R44_MPRA_v3_Amp2Sc_R	CCGACTAGCTTGGCCGC	Second strand synthesis
R45_MPRA_v3_Amp2Sa_Illu_F	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACTGGCCGCTTGACG	Amplification of insert-barcode association libraries
R46_Illu_univ_adapter	ATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	Amplification of insert-barcode association libraries
S69_Illumina_Multiplex_v1	CAAGCAGAAGACGGCATACGAGATCGCATGATGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
S70_Illumina_Multiplex_v2	CAAGCAGAAGACGGCATACGAGATCGCATGATGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
T17_Illumina_Multiplex_v4	CAAGCAGAAGACGGCATACGAGATAGAGTAGCGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
T18_Illumina_Multiplex_v5	CAAGCAGAAGACGGCATACGAGATTCGAGAGTGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
T19_Illumina_Multiplex_v6	CAAGCAGAAGACGGCATACGAGATCTAGCTCAGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
T20_Illumina_Multiplex_v7	CAAGCAGAAGACGGCATACGAGATATCGTCTCGTGACTGGAG	Illumina sequencing barcode

	TTCAGACGTGTGC	
T21_Illumina_Multiplex_v8	CAAGCAGAAGACGGCATACGAGATTCGACAAGGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
T22_Illumina_Multiplex_v9	CAAGCAGAAGACGGCATACGAGATATCATGCGGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
T23_Illumina_Multiplex_v10	CAAGCAGAAGACGGCATACGAGATTGTTCCGTGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
T36_Illumina_Multiplex_v3	CAAGCAGAAGACGGCATACGAGATATCATGCGGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
Z61_Illu_Multiplex_v11	CAAGCAGAAGACGGCATACGAGATTGTTCCGTGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
Z62_Illu_Multiplex_v12	CAAGCAGAAGACGGCATACGAGATACGATCAGGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
Z63_Illu_Multiplex_v13	CAAGCAGAAGACGGCATACGAGATATTAGCCGGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
Z64_Illu_Multiplex_v14	CAAGCAGAAGACGGCATACGAGATCGATCGATGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
Z65_Illu_Multiplex_v15	CAAGCAGAAGACGGCATACGAGATGATCTTGCGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
Z66_Illu_Multiplex_v17	CAAGCAGAAGACGGCATACGAGATAGGATAGCGTGACTGGAG	Illumina sequencing barcode

	TTCAGACGTGTGC	
Z67_Illu_Multiplex_v16	CAAGCAGAAGACGGCATACGAGATGTAGCGTAGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
Z70_Illu_Multiplex_v20	CAAGCAGAAGACGGCATACGAGATGCTCAGTTGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
Z71_Illu_Multiplex_v21	CAAGCAGAAGACGGCATACGAGATGTCCTAAGGTGACTGGAGTTCAGACGTGTGC	Illumina sequencing barcode
R71_SV40_for_109035_R	GCAGGAGCCGCAGTGCAGCCATGGGGCGGAGAATG	Assembly of plasmid 109035_SV40_GFP
R72_SV40_for_109035_F	GCCAGAACATTTCTCTGGCCTAACAAAGATTGATATAAACCATGCTTCTTGTGTATATCC	Assembly of plasmid 109035_SV40_GFP
R73_MPRA_GFP_for_109035	CCGACTAGCTTGGCCGCCGACGCTCTTCCGATCTTCTAGAGGTTCGTCGACGCG	Assembly of plasmid 109035_SV40_GFP
Y65_mP_r	GGTGGCTTTACCAACAGTACCGG	Assembly of pGL4 MP plasmid
Y66_blunt_end_luc_plasmid	GCCAGAACATTTCTCTGGCCTCACGTCGGCTCCTGCGATGGCGGCCAAGCTTAGACAC	Assembly of pGL4 MP plasmid
AC45_Rho_E1_HS_F	CTTTCTCCTTGTTCACTGGCACAC	Amplification and cloning of luciferase constructs
AC46_Rho_E1_HS_R	ACACAATAGGAATGGCTTTTATCGGCTC	Amplification and cloning of luciferase constructs
AC47_Rho_E1_PT3_R	ACACAATAGGAATGGCTTTTATCTGCTC	Amplification and cloning of luciferase constructs

AC48_Rho_E1_HS_NEB_F	TGCCAGAACATTTCTCTGGCCTCACCTTTCTCCTTGTTCACTGGCACAC	Amplification and cloning of luciferase constructs
AC49_Rho_E1_HS_NEB_R	CTTGGCCGCCATCGCAGGAGCCGACACACAATAGGAATGGCTTTTATCGGCTC	Amplification and cloning of luciferase constructs
AC50_Rho_E1_PT3_NEB_R	CTTGGCCGCCATCGCAGGAGCCGACACACAATAGGAATGGCTTTTATCTGCTC	Amplification and cloning of luciferase constructs
AC51_Rho_E3_HS_F	CCTCTGGTCAACCTCACTGCTC	Amplification and cloning of luciferase constructs
AC52_Rho_E3_HS_R	GTCATGGGCTGCTCGACTGAG	Amplification and cloning of luciferase constructs
AC53_Rho_E3_PT1_F	CCTCTGGTCAGCCTCACTGCTC	Amplification and cloning of luciferase constructs
AC54_Rho_E3_HS_NEB_F	TGCCAGAACATTTCTCTGGCCTCAC CCTCTGGTCAACCTCACTGCTC	Amplification and cloning of luciferase constructs
AC55_Rho_E3_HS_NEB_R	CTTGGCCGCCATCGCAGGAGCCGACGTCATGGGCTGCTCGACTGAG	Amplification and cloning of luciferase constructs
AC56_Rho_E3_PT1_NEB_F	TGCCAGAACATTTCTCTGGCCTCACCCTCTGGTCAGCCTCACTGCTC	Amplification and cloning of luciferase constructs
AC57_sult_E2_HS_F	CAGCTCCCACATTCTAGCCTAGATATTTG	Amplification and cloning of luciferase constructs
AC58_sult_E2_HS_R	AGTTTACTTTCTGCAGAAAGGGTGCTG	Amplification and cloning of luciferase constructs

AC59_sult_E2_HS_NEB_F	CTTGGCCGCCATCGCAGGAGCCGACCAGCTCCCACATTCTAGCCTAGATATTTG	Amplification and cloning of luciferase constructs
AC60_sult_E2_HS_NEB_R	TGCCAGAACATTTCTCTGGCCTCACAGTTTACTTTCTGCAGAAAGGGTGCTG	Amplification and cloning of luciferase constructs
AC61_sult_E2_PT_var_R	CTAATGCTTGCTTGAACCGGTTCAGGCATGC	Amplification and cloning of luciferase constructs
AC62_sult_E2_PT_var_F	GCATGCCTGAACCGGTTCAAGCAAGCATTAG	Amplification and cloning of luciferase constructs
AC63_Rho_R1_PT3_PT4_NEB_R	CTTGGCCGCCATCGCAGGAGCCGACACACAATAGGAATGGCTTGTATCTGCTC	Amplification and cloning of luciferase constructs
AC64_Rho_R1_PT3_PT4_R	ACACAATAGGAATGGCTTGTATCTGCTC	Amplification and cloning of luciferase constructs
AC65_Rho_R1_HS_PT4_NEB_R	CTTGGCCGCCATCGCAGGAGCCGACACACAATAGGAATGGCTTGTATCGGCTC	Amplification and cloning of luciferase constructs
AC66_Rho_R1_HS_PT4_R	ACACAATAGGAATGGCTTGTATCGGCTC	Amplification and cloning of luciferase constructs
X54_110792_1_two_oligos_F	CTATGCACTCTTGTGCTTAGCTCTGAAACGATGCCCTACACACACCCTGGGTGTTTCAAACAAGGCTTTTCTCCAAG	Plasmid for perturbation of enhancer R1
X58_110792_2_two_oligos_R	CCTTATTTTAACTTGCTATTTCTAGCTCTAAAACATTGTGTTCTGTCTAACCTCGGTGTTTCGTCCTTTCCACAAGATATATAAAGC	Plasmid for perturbation of enhancer R1

X57_303forw_negcontr_two_oligos_F	CTATGCACTCTTGTGCTTAGCTCTGAAACCTGCACCTCTTGCACTTTGTGGTGTTTCAAACAAGGCTTTTCTCCAAG	Plasmid for perturbation of negative control for enhancer R1
X61_547rev_negcontr_two_oligos_R	CCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAAATCATGCATTTTACGCAGGTGTTTCGTCCTTTCCACAAGATATATAAAGC	Plasmid for perturbation of negative control for enhancer R1
Z51_scram_1_2_neb_F	CTATGCACTCTTGTGCTTAGCTCTGAAACCCGTCTACGCATCACGACACGGTGTTTCAAACAAGGCTTTTCTCCAAG	Scrambled sgRNAs
Z52_scram_1_2_neb_R	CCTTATTTTAACTTGCTATTTCTAGCTCTAAAACACGGAATGCTCCTTGATGACGGTGTTTCGTCCTTTCCACAAGATATATAAAGC	Scrambled sgRNAs
A37_ACTBpzF	TGTCCCCCAACTTGAGATGT	Beta actin qPCR primer F
A38_ACTBpzR	TGTGCACTTTTATTCAACTGGTC	Beta actin qPCR primer R
Y31_Rhobtb3_qpcr_1_F	TGACAGTGATTGGTACACTTCTCG	RhoBTB3 qPCR primer F
Y32_Rhobtb3_qpcr_1_R	TGTAGAAGTCATCAAGGCTGTGG	RhoBTB3 qPCR primer R
AC24_rs2473351_1_F	CTATGCACTCTTGTGCTTAGCTCTGAAACGGCCTCGGTCCTTACTCCGTGGTGTTTCAAACAAGGCTTTTCTCCAAG	Plasmid for perturbation of rs2473351

AC35_rs2473351_2_R	CCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGCTGAGGCTGCAAAGTGACTGGTGTTTCGTCCTTTCCACAAGATATATAAAGC	Plasmid for perturbation of rs2473351

3.1.5. List of DNA fragments

Name	Sequence
Rho_E1_HS	CTTTCTCCTTGTTCACTGGCACACTCACTTCTGTACCCCCTGAGCCACCATGTAAGATGTCCAACTACCCTGAGGCCTCTATGCTGTGAGGAATGCAAGGGCAGGGTGTGTGTAGGGCATCTGGTTATCAGCCCCAGCTGAAGTCCTAGCCAACCGCAAGCATCAACTGTCAGACAAGAAAGTGAAGGTGCCTCAGGATGATTCCAGCCTGAGTCACTGGGCCACTTGAAGATGTTGAATCACCACAAGTCATTTGAGTTTTCCCACCTGAGGCCCCAGACATCATGGAGCCGATAAAAGCCATTCCTATTGTGT
Rho_E1_HS_PT1	CTTTCTCCTTGTTCACTGGCACACTCACTTCTGTACCCCCTGAGCCACCATGTAAGATGTCCAACTACCCTGAGGCCTCTATGCTGTGAGGAATGCAAGGGCAGGGTATGTGTAGGGCATCTGGTTATCAGCCCCAGCTGAAGTCCTAGCCAACCGCAAGCATCAACTGTCAGACAAGAAAGTGAAGGTGCCTCAGGATGATTCCAGCCTGAGTCACTGGGCCACTTGAAGATGTTGAATCACCACAAGTCATTTGAGTTTTCCCACCTGAGGCCCCAGACATCATGGAGCCGATAAAAGCCATTCCTATTGTGT
Rho_E1_HS_PT1PT2	CTTTCTCCTTGTTCACTGGCACACTCACTTCTGTACCCCCTGAGCCACCATGTAAGATGTCCAACTACCCTGAGGCCTCTATGCTGTGAGGAATGCAAGGGCAGGGTATGTGTAGGGCATCTGGTTATCAGCCCCAGCTGAAGTCCTAGCCAACCGCAAGCATCAACTGTCAGACAAGAAAGTGAAGGTGCCTCAGGATGATTCCAGCCTCGAGTCACTGGGCCACTTGAAGATGTTGAATCACCACAAGTCATTTGAGTTTTCCCACCTGAGGCCCCAGACATCATGGAGCCGATAAAAGCCATTCCTATTGTGT
Rho_E1_HS_PT1PT2PT3	CTTTCTCCTTGTTCACTGGCACACTCACTTCTGTACCCCCTGAGCCACCATGTAAGATGTCCAACTACCCTGAGGCCTCTATGCTGTGAGGAATGCAAGGGCAGGGTATGTGTAGGGCATCTGGTTATCAGCCCCAGCTGAAGTCCTAGCCAACCGCAAGCATCAACTGTCAGACAAGAAAGTGAAGGTGCCTCAGGATGATTCCAGCCTCGAGTCACTGGGCCACTTGAAGATGTTGAATCACCACAAGTCATTTGAGTTTTCCCACCTGAGGCCCCAGACATCATGGAGCAGATAAAAGCCATTCCTATTGTGT

Rho_E1_HS_PT2	CTTTCTCCTTGTTCACTGGCACACTCACTTCTGTACCCCCTGAGCCACCATGTAAGATGTCCAACTACCCTGAGGCCTCTATGCTGTGAGGAATGCAAGGGCAGGGTGTGTGTAGGGCATCTGGTTATCAGCCCCAGCTGAAGTCCTAGCCAACCGCAAGCATCAACTGTCAGACAAGAAAGTGAAGGTGCCTCAGGATGATTCCAGCCTCGAGTCACTGGGCCACTTGAAGATGTTGAATCACCACAAGTCATTTGAGTTTTCCCACCTGAGGCCCCAGACATCATGGAGCCGATAAAAGCCATTCCTATTGTGT
Rho_E1_HS_PT2PT3	CTTTCTCCTTGTTCACTGGCACACTCACTTCTGTACCCCCTGAGCCACCATGTAAGATGTCCAACTACCCTGAGGCCTCTATGCTGTGAGGAATGCAAGGGCAGGGTGTGTGTAGGGCATCTGGTTATCAGCCCCAGCTGAAGTCCTAGCCAACCGCAAGCATCAACTGTCAGACAAGAAAGTGAAGGTGCCTCAGGATGATTCCAGCCTCGAGTCACTGGGCCACTTGAAGATGTTGAATCACCACAAGTCATTTGAGTTTTCCCACCTGAGGCCCCAGACATCATGGAGCAGATAAAAGCCATTCCTATTGTGT
Rho_E1_HS_PT3	CTTTCTCCTTGTTCACTGGCACACTCACTTCTGTACCCCCTGAGCCACCATGTAAGATGTCCAACTACCCTGAGGCCTCTATGCTGTGAGGAATGCAAGGGCAGGGTGTGTGTAGGGCATCTGGTTATCAGCCCCAGCTGAAGTCCTAGCCAACCGCAAGCATCAACTGTCAGACAAGAAAGTGAAGGTGCCTCAGGATGATTCCAGCCTGAGTCACTGGGCCACTTGAAGATGTTGAATCACCACAAGTCATTTGAGTTTTCCCACCTGAGGCCCCAGACATCATGGAGCAGATAAAAGCCATTCCTATTGTGT
Rho_E1_HS_PT1PT3	CTTTCTCCTTGTTCACTGGCACACTCACTTCTGTACCCCCTGAGCCACCATGTAAGATGTCCAACTACCCTGAGGCCTCTATGCTGTGAGGAATGCAAGGGCAGGGTATGTGTAGGGCATCTGGTTATCAGCCCCAGCTGAAGTCCTAGCCAACCGCAAGCATCAACTGTCAGACAAGAAAGTGAAGGTGCCTCAGGATGATTCCAGCCTGAGTCACTGGGCCACTTGAAGATGTTGAATCACCACAAGTCATTTGAGTTTTCCCACCTGAGGCCCCAGACATCATGGAGCAGATAAAAGCCATTCCTATTGTGT

3.2. Methods

3.2.1. iPS cell culture

iPS cells were routinely cultured on hESC-qualified Matrigel (Corning) coated plastic multi-well plates and maintained in mTeSR1 medium (STEMCELL Technologies). For routine passaging, cell colonies were detached, gently dissociated into small cell aggregates using ReLeSR (STEMCELL Technologies) and seeded onto fresh Matrigel-coated plates at a density of 1/10.

3.2.2. LN229 cell line cell culture

LN229 were routinely cultured on cell culture treated plastic dishes (Corning). Cells were maintained in DMEM/F-12, GlutaMAX™ (Gibco) supplemented with 10% Fetal bovine serum (SIGMA). Cells were fed every 3 days. Cells were routinely passaged upon reaching confluence using accutase (Marck) and seeded onto fresh plates at density of 1:8 to 1:4.

3.2.3. HEK293 cell line culture

HEK cells were routinely cultured on cell-culture treated plastic dishes (Corning). Cells were maintained in DMEM Low Glucose w/ Stable Glutamine w/Sodium Pyruvate (BioWest) supplemented with 5% Foetal bovine serum (SIGMA) and 1% of Penicillin-Streptomycin (ThermoFisher Scientific).

3.2.4. Generation of chimpanzee iPS cells from fibroblasts

Small bits of leftover skin tissue, from a scheduled medical procedure performed in the Warsaw ZOO, were obtained from 27 years old chimpanzee ‘‘Mandy’’ (female, 55 kg, signature S6471). The skin fragments were washed, cut into small pieces, and plated on wells of 6-well plate (3-4 explants per well), coated with 0.2% gelatin (Merck). Explants were covered with 400ml of culture medium (CM) (High-glucose DMEM (Biowest) with 20% FBS (Merck) and 1% Penicilin-Streptomycin (ThermoFisher Scientific). CM was changed every other day until the explants were attached to the culture plate. Then, the fibroblasts migrated out of the explants and attached to the culture plastic. After 3 weeks of culture, fibroblasts were passaged with Trypsin-EDTA (Merck) and used for reprogramming at 1st passage. 5000 fibroblasts/cm^2 were seeded on hESC-qualified Matrigel (Corning) - coated wells. Reprogramming was initiated on the 5th-day post-plating, using ReproRNA-OKSGM Kit (STEMCELL Technologies). Briefly, CM was changed to the growth medium (GM) (Advanced DMEM (ThermoFisher Scientific) with 10% FBS (Merck), 2 mM L-glutamine (ThermoFisher Scientific) and 175 ng/ml recombinant B18R Protein (STEMCELL Technologies). Fibroblasts were transfected (day 0) with freshly prepared ReproRNA Cocktail (2 ml ReproRNATM-OKSGM (STEMCELL Technologies, 05931), 4 ml ReproRNA Transfection Supplement (STEMCELL Technologies), 4 ml ReproRNA Transfection Reagent (STEMCELL Technologies), in 200 ml Opti-MEM Reduced-Serum Medium (ThermoFisher Scientific).

3.2.5. Generation and culture of the iPS cell derived neural progenitors (iNP)

To initiate the differentiation of iPS cells to iNP cells, the iPS cell colonies were detached and dislodged into single cells with Accutase(Merck; Day 0). 3x106 iPS cells were then resuspended in 1 ml of STEMdiff Neural Induction Medium + SMADi (STEMCELL Technologies) and seeded into single well of 24-well AggreWell 800 plate, in the presence of 10 mM Y-27632 (Adooq Bioscience), to formembryoid bodies (EBs). 3=4 of the medium was replaced daily between days 1-4. On the 5th day, EBs were collected using wideboar tips and reversible 37 mM cell strainers (STEMCELL Technologies) and plated into a single well of 6-well plate, coated withhESc-qualified Matrigel, in STEMdiff Neural Induction Medium + SMADi. Full medium change was performed daily (days 6-11).On day 12, neural rosettes were selected using STEMdiff Neural Rosette Selection Reagent (STEMCELL Technologies) and replatedinto a single well of 6-well plate, coated with Matrigel. Cells were cultured with daily medium change until day 14 (macaque and chimpanzee) or day 19 (human). Then, single-cell suspension of iNPs was obtained using Accutase, cells were plated in Matrigel-coated plates, and cultured in STEMdiff Neural Progenitor Medium (STEMCELL Technologies), with 1/4 Accutasse passage every 5-6 days.

3.2.6. Generation and culture of iPS-derived astrocytes (iAstrocytes)

Single cell suspension of the iNP cells were immunolabelled for CD271 (BD, dilution 1:20), CD133 (Miltenyi Biotec, dilution 1:100),CD184 (BD, dilution 1:5). The CD271-

/CD133+/CD184+ iNPs were purified using flow cytometry.119 iAstrocyte differentiation was induced between passages 2-5 after flow cytometry assisted purification, as previously described33 with minor modifications. Briefly, the amplified iNPs were dissociated to single cells with Accutase; 150,000 cells/cm2 were plated, on Matrigel-coated wells in AM medium (ScienCell). Cells were differentiated for a total of 30 days (on day 14 of the differentiation, cells were dissociated using Accutase and passaged to the initial seeding density). During the differentiation, ½ medium was changed every other day. To examine the presence of astrocyte markers, cells were immunolabelled between DIV35 and DIV45 for Vimentin (DSHB, dilution1:20); GFAP (Merck, dilution 1:300); S100B (Proteintech, dilution 1:100); SLC1A3 (Proteintech, dilution 1:100). After completed differentiation, iAstrocytes were propagated in Astrocyte Medium (ScienCell).

3.2.7. Immunofluorescence staining

Cells were fixed for 15 min. at room temperature (RT) with 4% PFA (Merck), washed with Dulbecco's phosphate buffered saline(DPBS; 3 times for 5 min.), and permeabilized (10 min., at RT) with 0.5 % Triton X-100 (Merck) in DPBS. Then, the permeabilised cells were washed with (DPBS; 3 times 5 min. each wash). Next, the cells were incubated in blocking solution: 0.5 % BSA (BioShop) in DPBS for 1 hr at RT. Subsequently, cells were incubated with primary antibodies, diluted accordingly in blocking solution, for 1hour at RT. Following primary antibody incubation, cells were washed (DPBS; 3 washes for 5 min. each) and incubated with fluorescently labelled secondary Alexa Fluor 488/568/647 antibodies (ThermoFisher Scientific). The preparations were counterstained with Hoechst (ThermoFisher Scientific) for 1 hour at RT, when applicable. After three washes with DPBS, cells were mounted on slides using Prolong Diamond Antifade Mountant (ThermoFisher Scientific) and analysed without delay by Zeiss LSM800 confocal microscope with Airyscan detector or Olympus scanning microscope with camera Hamamatsu ORCA Flash4.0 V2.

3.2.8. Glutamate uptake assay

For glutamate uptake assay, 15,000 iAstrocytes/cm2 were plated in Matrigel-coated wells and grown for 10 days, with routine medium changes but without passaging. On day 10, cells were washed once with Hank's Balanced Salt Solution (HBSS, ThermoFisher Scientific) and incubated for 1 hour with 30 mM glutamate (Promega) in HBSS buffer at 37C. Then, the glutamate concentration in supernatant was assessed using luminescence-based Glutamate-Glo Assay kit (Promega), according to the Manufacturer's instructions. Luminescence was read using a TECAN-Infinite M1000 PRO microplate reader.

3.2.9. Imaging of ATP-evoked calcium waves

To assess the potential changes in the cytosolic level of calcium ions (Ca2+) upon addition of extracellular ATP, 15,000 iAstrocytes/cm2 were plated on Matrigel-coated, 35 mm glass-bottom dishes (VWR). On the next day, cells were incubated with 1 mM Fluo-4AM(ThermoFisher Scientific), pre-mixed in a 1:1 ratio with Pluronic F-127 (ThermoFisher Scientific), for 20 min in culture medium at 37C. Following the incubation, cells were washed once with pre-warmed live cell imaging solution (LCIS; 140 mM NaCl; 2.5 mM KCl; 1.8 mM CaCl2;1 mM MgCl2; 10 mM HEPES pH = 7.5 in MiliQ H2O) and kept in fresh LCIS for additional 15 min. to equilibrate. Changes in Fluo4fluorescence intensity, indicating raises in the cytosolic Ca2+, were recorded in constant flow of the LCIS. During imaging, 100mM ATP (ThermoFisher Scientific) in LCIS was administered for 30s. Time-lapse images were acquired every 500 ms, for the total time of300 s, using an inverted Leica DMI7000 microscope, equipped with Andor Zyla 5.5 sCMOS camera, using HCX PL APO CS 10x/0.40 DRY objective. Fluorescence was recorded in the green channel (I3 filter;(ex) 450-490 510; (em) 515). For the analysis of the Ca2+ waves, either a single cell (foreground) or a similarly sized area region with no cells (background) were our regions of interest (ROIs). We chose 50 foreground and one background ROIs per recording were manually selected in Fiji (2.1.0/1.53c). Fluorescence intensity over time for each ROI was exported for further data analysis. Background subtraction and intensity normalization to the basal level (DF/F0), was applied in Microsoft Excel (Microsoft). Graphs were generated using GraphPad Prism (DotMatics).

3.2.10. Co-cultures of iAstrocytes with primary rat neurons

Primate iAstrocytes were co-cultured with cortical primary rat foetal neurons (E18) in indirect manner (Live/Dead assessment). Briefly, iAstrocytes were plated either on Matrigel-coated trans-well inserts (VWR) in 1:2 astrocyte to neuron ratio. Co-cultures were maintained in NBA-B27 medium, supplemented with 5.5 mM AraC (ThermoFisher Scientific). Half of the culture medium was replaced every 3rd day, until culture termination on the day 10. Coverslips with neurons growing underneath iAstrocytes in indirect mode were used for viability assessment.

Neurons were washed with DPBS and stained with calcein-AM from LIVE/DEAD Viability/Cytotoxicity Kit (ThermoFisher Scientific), and Propidium Iodide (ThermoFisher Scientific), according to manufacturer’s instructions. Imaging was performed immediately after staining, using ZOE FluorescentCell Imager (Bio-Rad).

3.2.11. Lentivirus production

HEK 293 cells were grown to 60-80% confluence on cell culture-treated plastic dishes (corning). For a single 10cm petri dish of cells, medium was changed at least 30 minutes prior to transfection with virus components. 7ml of medium was added per dish. Cell were transfected as follows: 5ug of PsPAX2, 5ug VsVG (plasmids and 10ug of transfer plasmid were suspended in 1ml of serum-free medium (Opti-MEM Reduced-Serum Medium, ThermoFisher Scientific) mixed with 26µl of PEI 1mg/ml (Polysciences Europe GmbH) The mixture was then incubated for 20 minutes at room temperature, and added to cell growth medium drop by drop.

Medium was changed after 20 to 28 hours, and fresh growth medium was added. Medium supplemented with ViralBoost reagent (Alstem). Cells were grown in incubator for 44 to 52 hours more. On the third day from transfection, 7 ml medium was collected. In order to remove detached cells and debris, collected virus-containing medium was diluted with 1ml of ice-cold PBS, and spun for 5 minutes at 1000 g. 6ml of supernatant was collected. 1ml of ice-cold PBS was added for the second time, and the spin was repeated for the second time. 6ml of supernatant was collected and concentrated using Lenti-X™ Concentrator (TaKaRa), according to manufacturer's instructions. The resulting viral pellet was suspended in ice-cold PBS, aliquoted into 1/20 portions and stored at -80 for later use. This protocol was scaled to other cell culture dish format according to the surface area ratio.

3.2.12. Viral transduction

Cells were plate at density of 62 500 cells/cm2. Two protocol variants were employed for viral infection, one volving spinfection and without it. In the spinfection variant, 2 ml of growth medium was added to cells, and supplemented with polybrene (Merck) to 10ug/ml. Subsequently, 1/20 of a virus suspension collected form a single 10cm petri dish was added. If the infection involved two viruses, then the same portion of each virus was added to the dish. Dishes were then sealed with parafilm. Cultures were then spun in a rotating bucket centrifuge for 90min at 300g at 37°C and, after removing parafilm seal, grown in incubator overnight. the medium used for infection was then changed the next day. In the variant of the protocol without spinfection, the spinfection step was mitted, but cells were treated the same otherwise. Perturbation of RhoBTB3 enhancer was performed with spinfection. Perturbation of enhancer overlapping rs2473351 was performed without spinfection.

3.2.13. CRISPRi

Cells were infected as described above, with sgRNA virus and CRISPRi virus prepared from plasmid lenti_dCas9-KRAB-MeCP2 (Addgene #122205). Cells were selected with 15µg/ml of blasticidin (VWR) and 1.5µg/ml of puromycin (ThermoFisher Scientific) for two weeks, with medium changed daily.

3.2.14 RT-qPCR

RNA was isolated using Direct-zol RNA MiniPrep Kit (Zymo), according to the manufacturer’s instructions. cDNA was prepared from 200ng of RNA using High-Capacity cDNA Reverse Transcription Kit (ThermoFisher Scientific). Standard qPCR was performed using Fast Sybr Green (ThermoFisher Scientific).

3.2.15. Luciferase assay

Cells were plated the day before transfection on Matrigel-coated 96-well breated plates (Corning). 24000 cells were plated per well. On the day of transfection, fresh medium (33.28 µL) was added to each well. per well. Cells were transfected using Lipofectamine Stem reagent, according to manufacturer’s protocol. For each well, a 100.8 ng of each construct was used, and – the luciferase reporter and pRL_SV40 plasmid, and 0.32 µL per well. DNA was suspended in 4.2µl of medium, and lipofectamine in 4.2µl of medium – the combined volume added to cells was 8.4ul. After overnight incubation, luciferase assay was performed using Dual-Luciferase? Reporter Assay System (Promega) and measured using Tecan 1000 plate reader. Normalised luciferase reading was obtained by dividing the luciferase raw value by renilla raw value. For each construct, 4 readouts were obtained.

3.2.16. sgRNA design

sgRNAs were designed using CRISPOR [305] or picked from database described by Yao et al., (2024). For scrambled control sequences, sgRNAs described by Horlbeck et al., (2016) were used.

3.2.17. MPRA

3.2.17.1 MPRA library synthesis

Both libraries were obtained from Twist Bioscience. Final structure of each library was as follows:

`ACTGGCCGCTTGACG`-200 bp sequence to be synthesized-`CACTGCGGCTCCTGC`

SV40 region used in MPRA

```
GGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTC
CCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGC
CCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTG
```

CMV region used in MPRA

```
CCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAAC
GCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTAC
ATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTG
```

3.2.17.2. Library cloning

MPRA cloning was performed according to a modified protocol by Tewhey et al, (2016), described in detail below.

3.2.17.2.1. Library amplification

Library was ordered from Twists biosciences a lyophilized oligonucleotide pool. Each oligonucleotide was designed as a 200-basepair stretch of DNA flanked by universal amplification adapters. Final library structure followed the pattern: ACTGGCCGCTTGACG[200 basepairs]CACTGCGGCTCCTGC. A more detailed description of oligonucleotide design is described in “results” section. Upon receipt, DNA was spun and suspended in UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen) to concentration of 10ng/µl, and then further diluted to 1ng/µl. The suspended library was stored at -80°C.

The library amplification was performed in emulsion PCR. For a library of approximately 6000 fragments, 3 separate emulsion PCR reactions were prepared on ice, each comprising: 0.5µM of primer R20_MPRA_v3_F, 0.5µM of primer R19_MPRA_v3_20I_R 0.5µM, 1.86ng of oligonucleotide library, 25µl of New England BiolabsNext Ultra II Q5® Master Mix (New England Biolabs), 0.5µl of Q5® High-Fidelity DNA Polymerase (New England Biolabs), 2ng of BSA, UltraPure™ DNase/RNase-Free Distilled Water to 50µl. For each 50µl of reaction, the following reagents were added to create an emulsion: 220µl of TegoSoft DEC (Evonik), 60µl of ABIL WE (Evonik), 20µl of Mineral Oil (SIGMA), for a total of 300µl per single PCR reaction mix. To emulsify the reaction, the mixture was then vortexed for 5’ at 4°C. The mixture was then distributed into 96 well plate on ice, in 50µl portions. The plate was then incubated in thermocycler according to the following conditions: step 1: 95°C, 30 seconds, step 2: 95°C, 20 seconds, step 3: 60'°C, 10 seconds, 72°C, step 4: 15 seconds, return to step 2 35 times, step 5: 72°C, 5 minutes. The emulsion was subsequently broken up by addition of 1ml of 2-butanol (Thermo Fisher Scientific), 50µl of Agencourt AMPure XP beads (Beckman Coulter)), 80µl of binding buffer (2.5M NaCl, 20% PEG-8000 (Promega)) per 350µl of emulsion mix, followed by vigorous vortexing. The mixture was then incubated at room temperature for 10 minutes and spun for 5 minutes at 2900 g. The resulting solution contained two clearly visible phases. The upper, organic phase was carefully collected and discarded. The remaining solution was then placed on a magnetic rack for 20 minutes. Supernatant was removed, and beads were washed: once with 1 ml of 2-butanol, and three times with freshly made 80% ethanol. The beads were subsequently dried at room temperature for 5 minutes, and eluted using buffer EB (Qiagen).

The resulting concentration of DNA was estimated using NanoDrop. Size distribution was tested using TapeStation. Distributions with multiple peaks were deemed correct as long as at least one peak had a size of approximately 328 bp.

3.2.17.2.2. First cloning step

Cloning of barcoded enhancer library was performed a follows. Firstly, the destination plasmid pMPRAv3:∆luc:∆xbaI was isolated from overnight culture in LB (AA Biotechnology) supplemented with 100ug/ml of carbenicillin (VWR) using Plasmid MIDI AX kit (AA Biotechnology,092-10)). The DNA was quantified using NanoDrop, and digested with with SfiI restriction enzyme (New England Biolabs). An example reaction contained: 10µl of rCutSmart buffer (New England Biolabs, 2µl (40U) of SfiI, 10ug of DNA and UltraPure water (Invitrogen) to 100µl. The mixture was then incubated at 37°C overnight ,and purified on 1% agarose gel. A band corresponding to 2499 bp was then extracted, and purified using Agarose-out kit (EurX), and eluted with 30µl of water. To create the mpraΔorf library, the barcoded enhancer library was ligated into digested plasmid pMPRAv3:∆luc:∆xbaI using gibson assembly. The reaction mixture was composed of: 550ng of amplified and barcoded library, 500ng of digested plasmid, UltraPure water (Invitrogen) to 10µl, 10µl of GeneArt™ Gibson Assembly HiFi Master Mix (Ivitrogen). Ligation was carried out by incubating the mixture for 60 minutes at 50°C, followed by SPRI purification using Agencourt AMPure XP beads (Beckman Coulter)), according to manufacturer's instructions using 1.8:1 bead:sample ratio. Ligated plasmid was eluted from beads in 20µl of buffer EB (Qiagen) and quantified using NanoDrop. 25ng of the solution was then transformed per 50µl of 10-beta Electrocompetent E. Coli (New England Biolabs) by electroporation (Gene Pulser/MicroPulser Electroporation Cuvettes, 0,2cm gap, (BioRad), electroporator: electroporation settings 2kV, 200ohm, 25uF). In total, 100µl of bacteria was used. Serial dilution was performed after electroporation to estimate the number of CFU. For the first library, the number of CFU obtained was 5300000 For the second library the number of CFU obtained was 4900000. After electroporation, 1ml of SOC medium (ThermoFisher Scientific) was added per 50µl of bacteria, and the cultures were regenerated for 1 hour at 37°C, shaking at 300 RPM. Then, each 1ml portion was added to 2x 20ml of LB supplemented with 100 mg/mL of carbenicillin (VWR), and incubated for 6.5h at 37°C shaking at 180 RPM. After than bacteria was spun and purified using Plasmid MIDI AX kit (AA Biotechnology).

3.2.17.2.3. Preparation of association library

After cloning the first sequencing step to determine the association between oligo sequence and barcode was performed. Library purification consisted of two rounds of PCR, each followed by SPRI purification.

For the first step of library preparation, two 50µl PCR reactions were set up on ice as follows: 200ng of plasmid library prepared in the previous step, 50µl of New England BiolabsNext Ultra II Q5® Master Mix (New England Biolabs), 0,5 µM of R45_MPRA_v3_Amp2Sa_Illu_FR primer, 0,5 µM of R46_Illu_univ_adapter, UltraPure H2O to 100µl. The reaction mixture was then incubated in a thermocycler according to the following conditions: step 1: 95°C, 20 seconds. step 2: 95°C, 20 seconds, step 3: 62°C, 15' seconds, step 4: 72°C for 30 seconds, return to step 2, 8x, step 5: 72°C, 2 minutes. The mixture was then purified using AMpure XP beads according to manufacturer's protocol for purification of small fragments, using 0.6:1 bead:sample ratio. DNA was eluted with 15µl of EB. Size distribution was examined with TapeStation DNA ScreenTape & Reagents (Agilent) on TapeStation4200 Device (Agilent), with the product peak expected at 365bp.

For the second step of library preparation, index-containing primers were used (primers S69-S70, T17-T23, T36, Z61-Z67, Z70-Z71). The PCR mixture was set-up on ice as follows: 10µl of eluted DNA from the first amplification step, 50µl of New England BiolabsNext Ultra II Q5® Master Mix (M0544S), 0,5µM of barcoding primer (one of primers S69-S70, T17-T23, T36, Z61-Z67, Z70-Z71), 0,5µM of R46_Illu_univ_adapter, UltraPure water to 100µl. The reaction mixture was then incubated in a thermocycler according to the following conditions: step 1: 95°C, 20 seconds. step 2: 95°C, 20 seconds, step 3: 64°C, 30 seconds, step 4: 72°C for 30 seconds, return to step 2, 6x, step 5: 72°C, 2 minutes. The mixture was then purified using Agencourt AMPure XP beads (Beckman Coulter) according to manufacturer's protocol, using 1.8:1 bead:sample ratio. DNA was eluted with 15µl of buffer EB (Qiagen). Size distribution was examined with TapeStation DNA ScreenTape & Reagents (Agilent) on TapeStation4200 Device (Agilent), with the product peak expected at 397bp. Libraries were sequenced (2x150 bp, paired-end) using NovaSeq6000 device (Illumina).

3.2.17.2.4. Second cloning step

In this step the minimal promoter-GFP cassette was cloned into mpraΔorf library. The insert cassette was prepared as follows. First, plasmid Addgene 109036 was digested with restriction enzymes. The reaction mixture was set-up on ice as follows: MfeI-HF(New England Biolabs), 2µl, KpnI-HF (New England Biolabs), 1µl, 10x CutSmart Bufer (New England Biolabs), 5µl, plasmid 109036 pMPRAv3::minP-GFP, 10ug, UltraPure water (Invitrogen) to 50µl, and incubated for 1 hour at 37°C. excised fragment was then purified on 1% agarose gel. A band corresponding to 1022 bp was then extracted, purified using Agarose-out kit (EurX), and eluted with 30µl of UltraPure water (Invitrogen).

To amplify the minimal promoter-GFP construct, multiple PCR reactions were prepared to produce sufficient amount of DNA for cloning. A single PCR reaction was set up on ice as follows: 0,5µM of primer R21_ED_MPRA_GFP_F, 0,5µM of primer R22_ED_MPRA_GFP_R, 0,2µM of dNTPs (ThermoFisher Scientific). 5x Q5® buffer (New England Biolabs), 10µl, Q5® High-Fidelity DNA Polymerase (New England Biolabs), 0,5µl, excised promoter-GFP cassette, 50ng, UltraPure water (Invitrogen) to 50µl. The reaction mixture was then incubated in a thermocycler according to the following conditions: step 1: 98°C, 30 seconds. step 2: 98°C, 10 seconds, step 3: 69°C, 30 seconds, step 4: 72°C for 30 seconds, return to step 2, 35x, step 5: 72°C, 2 minutes. 5µl of the amplicon mixture was tested on agarose gel – band of 843bp was expected. The rest was purified using Agencourt AMPure XP beads (Beckman Coulter)) according to manufacturer's protocol for purification of small fragments, using 0.65:1 bead:sample ratio. DNA was eluted in buffer EB (Qiagen).

To linearise mpraΔorf library, the following reaction was set-up on ice: 10ug of mpraΔorf library, 15µl of 10x CutSmart buffer (New England Biolabs), 10µl (100U) of AsiSI (New England Biolabs), UltraPure water (Invitrogen) to 150µl. The reaction mix was then incubated for 3 hours at 37°C. 10µl of plasmid digest mix was then tested on a gel- band of 2287 was expected. Rest of the mixture was frozen at -20°C.

To assemble the final plasmid, the following reaction was set-up on ice: 2,5ug of linearized mpraΔorf library, 8,25ug of GFP amplicon (primers R21_ED_MPRA_GFP_F and R22_ED_MPRA_GFP_R), UltraPure water (Invitrogen) to 75µl. To this mix, 75µl of 2x GeneArt™ Gibson Assembly HiFi Master Mix (Invitrogen) was added. The reaction mix was then incubated for 90 minutes at 50°C. The mixture was then purified using Agencourt AMPure XP beads (Beckman Coulter) according to manufacturer's protocol, using 1.8:1 bead:sample ratio. DNA was eluted with 25µl of buffer EB (Qiagen). In order to remove uncut plasmid, the following reaction was set-up on ice: 2,5µl (25U) of AsiSI (New England Biolabs), 2,5µl (2.5U) of Exonuclease V (RecBCD) (New England Biolabs) 5µl of BSA (2ug/µl), 5µl of 10mM ATP (New England Biolabs, P0756S) (final concentration 1mM), 5µl of 10x New England Biolabsuffer™ 4 (New England Biolabs), UltraPure water (Invitrogen) to 50µl. The reaction mixture was then incubated at 37°C for 6 hours, and purified using Agencourt AMPure XP beads (Beckman Coulter) according to manufacturer's protocol, using 1.8:1 bead:sample ratio, eluted in 20µl of EB (Qiagen) and quantified using NanoDrop.

The plasmid library was then transformed into New England Biolabs stable electrocompetent bacteria (New England Biolabs). Around 30ng of the library was transformed per 50µl of 10-beta Electrocompetent E. Coli (New England Biolabs) by electroporation electroporation (Gene Pulser/MicroPulser Electroporation Cuvettes, 0,2cm gap, (BioRad), electroporator: electroporation settings 2kV, 200ohm, 25uF). In total, 100µl of bacteria was used. Serial dilution was performed after electroporation to estimate the number of CFU.

After electroporation, 4ml SOC medium (ThermoFisher Scientific) was added to bacteria, and then 8ml more for total of 6 aliquots of 2ml each. Cultures were regenerated for 1 hour at 37°C, shaking at 300 RPM. Then, each 2ml portion was added to 500ml of LB supplemented with 100 mg/mL of carbenicillin, and incubated for 9 to 10h at 37°C shaking at 180 RPM, until optical density of the culture (OD 600) as measured using NanoDrop reached 0.4-0.5. After than bacteria was spun and purified using Plasmid MIDI AX kit (AA Biotechnology, 092-10). To re-transform libraries, 100 ng ready library was added per 50µl of bacteria, and electroporated as described above. In total, 400µl of bacteria was used, After electroporation, 4ml SOC medium (ThermoFisher Scientific) was added to bacteria, and then 8ml more for total of 6 aliquots of 2ml each Cultures were then regenerated, cultivated and harvested as described above. CFU was estimated as described above.

3.2.17.2.5. Transfection

Transfection of cell cultures with the library was performed as follows: cells were seeded at density of 75000 cells/cm2 of petri dish the day before transfection, in 75cm2 cell culture bottles. Line LN229 was seeded directly on cell-culture treated plastic dishes. Line U87 was seeded on cell-culture treated plastic dishes coated with Matrigel (Corning). On the day of transfection, medium was changed approximately 30 minutes before transfection; 7.8ml of fresh growth medium was added. For each 75cm2 bottle, transfection as performed as follows: in two separate tubes, DNA mix (1972.5µl of Opti-MEM Reduced-Serum Medium, (ThermoFisher Scientific), 47.5 ug of plasmid library) and lipofectamine mix (1972.5µl of Opti-MEM Reduced-Serum Medium, 157.5µl of Lipofectamine™ Stem Transfection Reagent). DNA mixture as then added to lipofectamine mixture, mixed gently, and incubated at room temperature for 10 minutes before being added to cells drop by drop. One day after transfection, medium was changed. Cells were collected on the second day from transfection. First, the medium was removed and kept. Then, cells were washed once with PBS, and the wash-out was kept. Then, Accutase (SIGMA) was added. Spent medium wash then spun for 5 minutes at 500g, and discarded, leaving the pellet. Then, the PBS wash-out was added to the same tube, and spun as before, and supernatant was discarded. After cells were detached, Accutase was collected, added to pellet of previously collected cells, and the plate was once more washed with PBS, and the was -out was collected in the same tube. The, the tube was then spun for 5 minutes at 500g. Most of supernatant was then collected, and split into two equal potions in 1.5 Eppendorf tubes. Each portion was spun as above, and washed two more time with 1ml of PBS. After final spin, as much of supernatant as possible was removed, and cell pellets were frozen. For LN229 line the transfection was performed on two 75cm2 bottles.

3.2.17.2.6. cDNA synthesis

cDNA synthesis was performed using SuperScript III Reverse transcriptase kit (Life Technologies). First, RNA was extracted from each of aliquot of frozen cells using Direct-zol™ RNA MiniPrep (Zymo) kit according to manufacturer's instructions. RNA was stored at -80°C. cDNA reaction was set-up on ice as follows: 2 pmol of primer R44_MPRA_v3_Amp2Sc_R, 500ng of RNA, 1µl of 10 mM dNTP Mix (ThermoFisher Scientific), UltraPure water (Invitrogen) to 10µl. The mixture was then incubated at 65°C for 5 minutes, and then on ice for at least 1 minute. Contents of the tube was collected by brief centrifugation. In ice, the following components were added to the reaction 4µl of 5X First-Strand Buffer (from Super Script II kit, Life Technologies), 1µl of 0.1 M DTT (from Super Script II kit, Life Technologies), 1µl of 1µl of SUPERase•In™ RNase Inhibitor, 20 U/µl (ThermoFisher Scientific), 1µl of SuperScript™ III RT (Life Technologies). The reaction was mixed gently by pipetting up and down, and incubated for 80 minutes at 47°C, and subsequently inactivated by incubation for 15minutes at 70°C. Purified using Agencourt AMPure XP beads (Beckman Coulter)) according to manufacturer's protocol, using 1.8:1 bead:sample ratio, and eluted in 10µl of UltraPure water (Invitrogen).

3.2.17.2.7. Preparation of readout library

In order to standardise library concentrations and minimise amplification bias, samples are amplified by qPCR to estimate relative concentration of GFP cDNA. cDNA or plasmid library diluted 500x were used as a template. The following reaction was set-up on ice, protected from light, in Hard-Shell® 384-Well PCR Plates (Bio-Rad) : 1µl of cDNA, diluted, 2x or plasmid library, diluted 500x. 5µl of New England BiolabsNext Ultra II Q5® Master Mix (New England Biolabs, 1µl of SYBR™ Green I Nucleic Acid Gel Stain (ThermoFisher Scientfic, diluted

1:10000, R48_MPRA_v3_Illu_GFP_F to concentration of 0.5μM, R46_Illu_univ_adapter to concentration of 0.5μM, (of 10μM), UltraPure water (Invitrogen) 10μl. Plates were then vortexed and spun. qPCR reaction was ran in Opus Real-Time PCR System according to the following conditions: step 1: 95°C, 20 seconds. step 2: 95°C, 20 seconds, step 3: 65°C, 20 seconds, step 4: 72°C for 30 seconds, return to step 2, 40x, step 5: 72°C, 2 minutes. cDNAs and plasmid solutios nwere diluted to march the relative concentration of the most dilute sample.

For the first step of library preparation, a 50μl PCR reaction was set up on ice for each sample, as follows: 2.5μl of cDNA/plasmid library diluted to match the concentration of the most dilute sample. 25μl of New England BiolabsNext® High-Fidelity 2X PCR Master Mix (New England Biolabs), primer R45_MPRA_v3_Amp2Sa_Illu_F primer to final concentration of 0,5 μM, primer R46_Illu_univ_adapter to final concentration of 0,5 μM, UltraPure water (Invitrogen) to 50μl. The reaction mixture was then incubated in a thermocycler according to the following conditions: step 1: 95°C, 20 seconds. step 2: 95°C, 20 seconds, step 3: 65°C, 20' seconds, step 4: 72°C for 30 seconds, return to step 2, 10x(11cycles in total), step 5: 72°C, 2 minutes. The mixture was then purified using AMpure XP beads according to manufacturer's protocol for purification of small fragments, using 0.6:1 bead:sample ratio. DNA was eluted with 15μl of buffer EB (Qiagen).

For the second step of library preparation, index-containing primers were used (primers S69-S70, T17-T23, T36, Z61-Z67, Z70-Z71, see primer list). The PCR mixture was set-up on ice as follows: 10μl of eluted DNA from the first amplification step, 50μl of New England BiolabsNext® High-Fidelity 2X PCR Master Mix (New England Biolabs), barcoding primer (one of primers S69-S70, T17-T23, T36, Z61-Z67, Z70-Z71) to concentration of 0,5μM of, primer R46_Illu_univ_adapter to concentration of 0,5μM, UltraPure water to 100μl. The reaction mixture was then incubated in a thermocycler according to the following conditions: step 1: 95°C, 20 seconds. step 2: 95°C, 20 seconds, step 3: 64°C, 30 seconds, step 4: 72°C for 30 seconds, return to step 2, 7x (in total, 8 cycles), step 5: 72°C, 2 minutes. The mixture was then purified using Agencourt AMPure XP beads (Beckman Coulter) according to manufacturer's protocol, using 1.8:1 bead:sample ratio. DNA was eluted with 15μl of buffer EB (Qiagen). Size distribution was examined with TapeStation DNA ScreenTape & Reagents (Agilent) on TapeStation4200 Device (Agilent), with the product peak expected at 268bp. Libraries were sequenced (2x150 bp, paired-end) using NovaSeq6000 device (Illumina).

3.2.18. Genomic characterisations

3.2.18.1. RNA-Seq

For RNA isolation, pellets of 500,000 cells were lysed in TRI-reagent (Sigma). Brain tissue pieces were homogenized in TRI-reagent, using Tissue Lyser II (Qiagen). RNA was isolated with Direct-zol RNA MiniPrep kit (Zymo), according to manufacturer's instructions. RNA quality was examined with Agilent RNA 6000 Nano Kit (Agilent). Samples featuring RIN>8 were used for library preparation. RNA-seq libraries were prepared using KAPA mRNA HyperPrep kit (Roche), with 1 mg of RNA as the starting material. Library preparation was performed according to manufacturer's instructions, with use of UMI in xGen UUDI-UMI Adapters (IDT). Library sizes were examined with TapeStation DNA ScreenTape & Reagents (Agilent) on TapeStation4200 Device (Agilent). Libraries were sequenced (2x100 bp, paired-end) using NovaSeq6000 device (Illumina).

3.2.18.2. ATAC-Seq

ATAC-seq libraries were prepared using ATAC-Seq kit (Active Motif), according to Manufacturer's instructions using 100,000 Accutase-detached iAstrocytes. Libraries were sequenced (2 x 100 bp, paired-end) using NovaSeq6000 device (Illumina).

3.2.18.3. ChIP-Seq

Chip-Seq experiments were performed as described before [307]. Chromatin corresponding to 3 million cells (H3K27ac, H3K4Me3) was used in the experiments. Following antibodies were used in ChIP: H3K27ac (Cell Signaling), H3K4Me3 (Abcam). Sequencing libraries were prepared using Ovation Ultralow V2 DNA-Seq Library Preparation Kit (Tecan), according to manufacturer's instructions. Libraries were sequenced (2 x 100 bp, paired-end) using NovaSeq6000 device (Illumina).

3.2.18.4. Intact HI-C

Intact Hi-C was performed by dr. Anna Stroynowska-Czerwińska according to the protocol published by ENCODE consortium (https://www.encodeproject.org/documents/4dfd0b02-ed3a-4461-b0f5-9ef51570af1f/).

Cells were detached using accutase or pelleted, washed in PBS and counted. At least 10,000,000 cells were taken for crosslinking the chromatin for Intact HiC. First, freshly prepared solution of 300 mM DSG (ThermoFisher Scientific) dissolved in DMSO was 100x diluted in PBS. Next, cell pellet was resuspended in 3mM DSG solution at a concentration of 5 000 000 cells/ml and incubated for 35 minutes at room temperature on rotating wheel. Afterwards, 16% methanol-free formaldehyde (ThermoFisher Scientific) was added dropwise to a final concentration of 1% and incubated for 10 minutes at room temperature with rotation. To quench crosslinking reaction, dropwise 1M Tris pH 7.5 was added to final concentration 0.375 M and incubated for 5 minutes at room temperature with rotation. Cells were then centrifuged for 5 minutes at 850xg at 4°C and pellet was washed once with 1x ice-cold PBS. Crosslinked cells were counted, divided for 5 000 000 portions, centrifuged again, supernatant was aspirated and cellular pellet was snap frozen and kept in -80°C until further use.

Before first experiments, crosslinked chromatin was titrated for MNase cleavage on 1 000 000 sample of cells. Later on the amount of MNase (Symbios) resulting in efficient cleavage was used.

To start Intact HiC with nuclei isolation, fresh complete MB#1 buffer 50 mM naCl, 10mM Tris, 4 mM $MgCl_2$, 1 mM $CaCl_2$) supplemented with 0.2% IGEPAL (Merck) and 1x Protease Inhibitor Cocktail (Sigma) was prepared. Cell pellets ($5x10^6$ cells) was thawed on ice, resuspend in complete MB#1 at a concentration of $1x10^6$ cells/100 µl and incubated for 20 min on ice. Nuclei were centrifuged at 1,750 g for 5 min at 4°C, washed once in complete MB#1 at a concentration of $1x10^6$ cells/100 µl, centrifuged again at 1,750 g for 5 min at 4°C. Chromatin was resuspended in complete MB#1, containing appropriate amount of MNase and digested 20 min at 37°C with shaking at 1,000 RPM using a thermomixer. To ensure complete Mnase inactivation, reaction was stopped by adding 500 mM EGTA to a final concentration of 4 mM and 10 minutes incubation at 65°C. Next, chromatin was centrifuged at 1,750 g for 5 min at 4°C, washed in 200 ul Hi-C Wash Buffer (10 mM Tris 8.0, 10 mM NaCl), centrifuged at 2000 g for 5 minutes and supernatant was discarded conservatively.

To repair MNase-digested DNA ends the chromatin was resuspended in 80 µl of MNase Repair Master Mix (67 µl H2O, 8 µl 10X T4 DNA Ligase Reaction Buffer (New England Biolabs), 5 µl of 10 U/µl T4 Polynucleotide Kinase (New England Biolabs)), incubated at 37ºC for 30 minutes, centrifuged at 2000 g for 5 minutes and the supernatant was discarded conservatively. For simultaneous biotinylation and ligation colocalizing DNA fragments, the nuclear pellet was resuspended in 100 µl of Ligase Master Mix (36 µl H2O, 10 µl 1 mM Biotin-11-dUTP (Jena Bioscience), 10 µl 1mM dATP/dCTP/dGTP (ThermoFisher Scientific), 10 µl 10X T4 DNA Ligase Reaction Buffer (New England Biolabs), 4 µl 5U/µl DNA Polymerase I Klenow Fragment (NEB), 10 µl 400 U/µl T4 DNA Ligase (New England Biolabs)), incubated at 25ºC for 1.5 hours and centrifuged at 2000 x g for 5 minutes. To digest proteins and remove crosslinks the nuclear pellet was resuspended in 100µl of Proteinase Master Mix (74 µl H_2O, 1 µl 1 M Tris pH 8.0, 5 µl of 20% (w/v) SDS, 10 µl 5 M NaCl, 5 µl of 0.8 U/µl Proteinase K (ThermoFisher Scientific) and incubated at 55ºC for 10 minutes and next at 75ºC for 1 hour.

DNA was purified using 1.8x AmPure beads, washed twice with freshly prepared 80% ethanol, eluted in 130µl of 10 mM Tris 8.0 and its concentration was measured using Nanodrop. To make the biotinylated DNA suitable for high-throughput sequencing, 40 ug DNA in 130 ul 10 mM Tris 8.0 was sonicated in Pre-Slit Snap-Cap 6x16 mm glass microTUBE vial in Cavaris M220 Focused-ultrasonicator.

To enrich DNA sample in biotinylated fragments, Dynabeads MyOne Streptavidin T1 beads (ThermoFisher Scientific) were washed, resuspended in 3xTWB (3 M NaCl, 15 mM Tris 8.0, 1.5 mM EDTA, 0.15% Tween20), mixed with sheared DNA to final 1xTWB and incubated at room temperature for 30 minutes. Next, beads were washed once with 55°C preheated 1XTWB and twice with 10mM Tris Buffer.

For library preparation NEB Next II Ultra Prep kit (New England Biolabs) was used, according to manufacturer's protocol. Briefly, beads were resuspended in 50 ul of 10 mM Tris 8.0, mixed with End Repair Mix, incubated at 20ºC for 30 minutes followed by 65ºC for 30 minutes. Next, Ligation Mix containing sample-specific IDT indexes was added and incubated at 20ºC for 15 minutes. To remove free adaptors from the solution, the beads were washed once with 55°C preheated 1X TWB and once with 10 mM Tris Buffer. To amplify the library, the beads were resolved in Kapa HiFi HotStart PCR Mix containing standard Illumina i5 and i7 primers by incubation in thermocycler according to the following conditions: step 1: 98°C, 45 seconds, step 2: 98°C, 15 seconds, step 3: 55°C, 30 seconds,, step 4: 72°C, 15 seconds, return to step 2 7 times, step 5: 72°C, 1 minute.

The final library was purified with two-side selection AmPure beads (0.6x and final 1.1x), washed with freshly prepared 80% ethanol and eluted into 10 mM Tris pH 8.0. The final DNA concentration was measured with Qubit and fragment size distribution with TapeStation.

3.2.19. Quantification and statistical analysis

Gene models used in this study were retrieved from ensemble using BioMart package [308] using the following command: useEnsembl(biomart="genes", dataset="hsapiens_gene_ensembl", mirror="useast").

3.2.19.1. General statistical analysis

All statistical analyses were carried out in R environment (version 4.5.1 (2025-06-13)), in RStudio environment. Fisher's Exact Test was calculated using fisher.test from base R package "stats"; t-test was calculated using t.test from base R package "stats"; Wilcoxon Rank Sum Test was calculated wilcox.test command from base R package "stats", except for figures 25d, f, 26d, f, 27d, 32c in which wilcox_test command from rstatix R package was used.

3.2.19.2. GO term analysis

GO term analysis was performed using clusterProfiler[309], using the following command: enrichGO(gene = gene_df$ENTREZID, universe = gene_df_ref$ENTREZID, OrgDb = org.Hs.eg.db, keyType = "ENTREZID", ont = "ALL", pAdjustMethod = "BH", pvalueCutoff = 0.05, qvalueCutoff = 0.05).

3.2.19.3. RNA-seq data pre-processing

Raw RNA-Seq reads (this study or published RNA-seq data from GSE97904, GSE73721, GSE127898) were trimmed using TrimGalore ver. 0.6.7[310], using parameters '–paired -q 30 –stringency 3 –length 30'. Alignment against the consensus genome was implemented using STAR version 2.7.10(Dobin *et al.*, 2013) with default parameters and 'outFilterMultimapNmax 1'. We used featureCounts version 2.0.3[311] with parameters '-p -O –countReadPairs -t exon -g gene_id' to obtain per gene RNA-seq read counts using ''Homo_sapiens. GRCh38.101.gtf'' from Ensembl's release version 101 as a reference[246,312,313].

3.2.19.4. ATAC-seq and ChIP-seq data processing

Raw reads were trimmed using TrimGalore version 0.6.7[310,314] using parameters '–paired -q 30 –stringency 3 –length 30' and alignment was performed using bowtie2[315] using parameters '–very-sensitive -X 2000'. All the ATAC-Seq, H3K27ac ChIP-Seq, H3K4me3 ChIP-Seq, data were aligned to the reference genomes of the individual species in humans (hg38 genome), chimpanzee (panTro6 genome) and the rhesus macaque (rheMac10) genomes. CTCF data was also aligned against the consensus genome (above). The alignments were filtered to remove duplicates using alignmentSieve (using parameters '–minFragmentLength 40 –ignoreDuplicates') which is available as a part of the deeptools package version 3.5.1[316]

Peak calling was performed using MACS2 (Model-based Analysis for ChIP-Seq) ver. 2.2.7.1 z[317] using parameters '–no-model'. The effective genome size required as one of the input parameters for the program was kept at default for humans and calculated separately for chimpanzee and rhesus macaque using faCount from UCSC Kent's tools.

3.2.19.5. Hi-C data pre-processing

Raw Hi-C reads were trimmed using TrimGalore version 0.6.7[310,314]. The fastq files were processed using the Juicer Pipeline version 2.13.07[318], using default options and the Hg38 and PanTro06 genome assemblies.. Intact Hi-C data was visualised in Juicebox 2.15.07 with "Balanced ++" normalisation.

3.2.19.6. Differential Expression Analysis

Two of the human iAstrocyte lines (ELE10, ELE30) obtained in this study along with data from 3 other female iAstrocyte from a published study[313] TCW - 3651_Astros (GSM2580014), 9319_Astros (GSM2580016) and 9429_Astros (GSM2580017) were analysed. Three chimpanzee iAstrocyte lines (SandraA, Mandy04 and Mandy06) and one Rhesus macaque iAstrocyte line (Becky) were also included. Counts from technical replicates of individual iAstrocyte lines were summed.

Differentially Expressed Genes were identified using DESeq2[319] version 1.36.0. DESeq2 was run with the following model: '~ 0 + species (human, chimpanzee or rhesus) and the fitType parameter set to "local". Differentially expressed genes (DEG), (cutoff: p-adj.<0.01) were identified in two separate groups (human versus chimpanzee and human versus macaque).

DEGs with an aligned change in expression in both comparisons were further studied (upregulated DEGs = 677; downregulated DEGs = 486 genes).

3.2.19.7. Identification of differentially open regions

ATAC-Seq peaks were identified using MACS2 in human, chimpanzee, and macaque cells. To enhance discovery rate and coverage, peaks were called using merged BAM files combining reads from all the samples for each species. “narrowPeak” files were considered. Differential openness was determined for peaks from autosomes and X chromosome (150642 peaks) that have a minimum of 50 % of the bases conserved in the other species (determined using liftOver with parameters” -minMatch=0.5 -bedPlus=6”). For regions produced by liftOver to chimpanzee and macaque, the ATAC-Seq reads BAM files were counted in the final ATAC-Seq peak list using featureCounts[311]. DESeq2[319] was used to identify differentially open regions (DORs). Regions which are considered to be DORs are a sum of two sets: human ATAC-seq peaks that have log2 fold change>0 compared to regions in chimpanzee and RM (at p-adj. < 0.316 in separate analyses, for a final at p-adj. <0.01) and peaks that have log2 fold change>0 compared to regions in PT (at p-adj. <0.01). This strategy identified 15218 DORs, and 777 peaks which were not matched due to different sequences.

Peaks are filtered for presence of methylation signal in human and active promoters (promoters are defined as 2kb regions centred on possible TSS based on ensembl gene models) that are have H3K4me3 mark in human chimpanzee or rhesus macaque.

3.2.19.8. Identification of loops from Hi-C data and normalisation

Loop calling was done using the HiCCUPS tools from Juicer 2.13.07[318] with default parameters using the merged HiC data from ELE30 and AG94 human cell lines. All loops greater than 3MB were removed from the analysis. This yielded 34578 domains.

The HiC data was normalised using Iterative Proportional Fitting for further analysis as described in published literature[320–322]

3.2.19.9. Genome-wide TFBS annotation

The complete HOCOMOCO v11 database[323] with 769 transcription factor (TF) motifs was used with FIMO[324] to scan the whole human (hg38) and chimpanzee (PanTro6) genomes for the presence transcription factor binding sites (TFBS), using parameter '– text'. The positions of TFBS in the chimpanzee were lifted over to the human genome assembly using liftOver with default parameters. The TFBS information from the FIMO analysis along with the sequence alteration information were overlapped using BEDTools[325] function intersect. Only human and chimpanzee sequences were included in the analysis of evolution of transcription factor motifs.

3.2.19.10. The human-chimpanzee-macaque consensus genome

Human (GRCh38 Release 101), chimpanzee (panTro6 genome) and the rhesus macaque (rheMac10) genome assemblies were considered to obtain consensus genome as described earlier[247,326,327]. In brief, we used available pairwise genome alignment files from UCSC Genome Browser and constructed a multiple genome alignment using multiz[328]. We masked all sites that were divergent between human and chimp or between human and macaque genomes along with their 6bp flanks (3bp on both ends of a mismatched position).

3.2.19.11. Defining genes expressed in foetal astrocytes

For genes expressed in astrocytes, I considered a published dataset corresponding to RNA-seq profiles of acutely purified human astrocytes[312], and processed it using RNA-seq analysis pipeline described above, and selected genes which had average TPM-normalised expression value across foetal astrocyte samples > 1. For genes specific in foetal astrocytes, I considered expression values given by Zhang et al., (2016) and selected genes for which the mean expression in foetal astrocytes was at least two times higher than in any other cell type analysed in the study.

3.2.19.12. Determination of potential disruption by a SNP

For each SNP mapping within ATAC-seq peak overlapping a H3K27ac peak that is not a promoter, I identified all nucleotide variants and exported 200bp sequences flanking them. I used those sequences to identify motifs on those sequences using FIMO[324] with the following parameters: --verbosity 1 –thresh 1.0E-4. The complete HOCOMOCO v11 database was used[323]. Essentiality score for SNPs was derived by determining the highest score for any basepair within motifs described in HOCOMOCO v11 database.

3.2.19.13. Design of MPRA libraries

MPRA regions were selected as described in the "Results" section. Sequence from the leading strand was picked. "Linked" and "Non-linked" enhancers were selected from DORs, based on proximity to promoters of DEGs described above. Enhancers surrounding specific genes were picked based on manually curated Intact-C domains. HARs were picked based on summary described by Girskis et al., 2021; only ATAC-seq peaks which, after resizing to 200 bp centred around the summit, were completely enclosed within a HAR were selected. "Linked" enhancers representing a subset of 769 linked enhancers which differed from their homologous chimpanzee counterparts by binding sites of USF only were selected based on global TFBS annotation described above; their variants featured sequences modified with the chimpanzee version of USF binding sites.

3.2.19.14. MPRA pre-processing

3.2.19.14.1. Association library reconstruction

Analysis was conducted as described by Tewhey et. al, (2016). The analysis was described in the original reference as follows: Paired-end 150 bp reads from the sequencing of the mpraΔorf library were merged into single amplicons using Flash v1.2.7 (flags: -r 150, -f 220, -s 10)[329]. Amplicon sequences were kept if the 5' adapter matched with a levenshtein distance of 3 or less and 2 bp at the edges of both the 5' and internal constant sequences matched perfectly. Oligo sequences from the passing reads were then mapped back to the expected oligo sequences using BWA mem version 0.7.9a (flags: -L 100 -k 8 -O 5)[330]. Alignment scores were calculated as matching bases divided by the expected oligo size and reads with alignment scores of less than 0.95 were discarded. Remaining oligo/barcode pairs were then merged and barcodes attributable to multiple oligo sequences were marked as conflicting and removed from further analysis. Barcodes matched with more than one enhancer were then removed using an R script, and barcodes were converted to reverse complements. The final association library consisted of a fasta list of enhancers candidate with individual barcodes.

3.2.19.14.2. Pre-processing of readout library

Analysis was conducted similarly to the method described in Tewhey et. al, (2016). Only R1 reads were considered. Reads were trimmed to 30 bp using cutadapt[310] (flags -l 30 -o <sample_name>.fastq). Subsequently, they were filtered according to the conditions specified by[331]: levenshtein distance of 4 or less was required within the constant sequence at the end of the tag-seq read with the two bases directly adjacent to the barcode (base 21 & 22) required to match perfectly. The output was then trimmed again, this time removing 11bp at the 5’ end using cutadapt[310] (flags -l 11 -o <sample_name>.fastq).

3.2.19.14.3. Alignment of readout library

Sequences were then aligned to a fasta list of barcodes generated previously. Alignment was performed using bwa (flags: -L 0 -k 19 -T 0)[330]. Relevant lines were then extracted using samtools (samtools view <library_name>.sam | cut -f 3 | grep _v “*” > <output>.txt and samtools view <library_name>_aligned.sam | cut -f 3, 10 | grep -v "*" > <output.txt

3.2.19.15. Analysis of MPRA data

For each library, sum of all aligned barcodes associated with a given CRE was treated as counts, and analysed using DEseq2[319].

3.2.19.15.1. Correlation with ChIP-seq and ATAC-seq signals

In this analysis, unique MPRA locations from each species were considered. For locations measured in both experiments, MPRA values were averaged out. Mean signal at each genomic coordinate was retrieved from RPGC-normalised data (.bigwig format) using tools developed by Chris Seidel (http://research.stowers.org/cws/CompGenomics/Resources/). Enhancers and repressors were only considered as such if they were detected in agreement across both MPRA experiments. Negative controls were not considered.

3.2.19.15.2. Correlation with G4 quadruplexes and CG/AT dinucleotides

In this analysis, unique MPRA sequences were considered. For sequences measured multiple times, MPRA values were averaged out. G4 quadruplexes were annotated using pqsfinder[332], with minimal score set to 20. Enhancers and repressors were only considered as such if they were detected in agreement across both MPRA experiments. Negative controls were not considered.

3.2.19.15.3. Correlation with Hi-C loops

In this analysis, ATAC-seq peaks corresponding to active and poised enhancers in human that overlapped MPRA enhancers and repressors were considered. Unique sequences measured using MPRA were considered. For sequences measured multiple times, MPRA values were averaged out. Enhancers and repressors were only considered as such if they were detected in agreement across both MPRA experiments.Only signal for actual human sequences was considered. Overlap with loop anchor was determined using GenomicRanges package in R. Anchors of loop domains determined using Juicer were used, as described before.

3.2.19.15.4. Correlation of MPRA activity with individual TFBS

In this analysis, only MPRA CREs corresponding to active and poised enhancers in human were considered. Only unique sequences from MPRA library were considered. Enhancers and repressors were only considered as such if they were detected in agreement across both MPRA experiments. Motifs were called on RE sequences using FIMO[324] with the following parameters: --verbosity 1 –thresh 1.0E-4. The complete HOCOMOCO v11 database was used[323]. Only USFs and TFs expressed in astrocytes according to Zhang et al., (2016) were considered.

3.2.19.15.5. Correlation with number of USF sites

In this analysis, unique MPRA sequences were considered. For sequences measured multiple times, MPRA values were averaged out. G4 quadruplexes were annotated using pqsfinder[332], with minimal score set to 20. Enhancers and repressors were only considered as such if they were detected in agreement across both MPRA experiments. Negative controls were not considered. Only USFs and TFs expressed in astrocytes according to Zhang et al., (2016) were considered.

3.2.19.15.6. Sequence comparison of human and chimpanzee CREs, and WT and ΔUSF CREs

Alignment was performed using msa package in R[333] with default settings.

3.2.19.15.7. Differences in USF number and corresponding CRE activity

In those analyses, motifs were called using FIMO[324] with the following parameters: --verbosity 1 –thresh 1.0E-4. The complete HOCOMOCO v11 database was used[323]. Only USFs and TFs expressed in astrocytes according to Zhang et al., (2016) were considered. Enhancers and repressors were only considered as such if they were detected in agreement across both MPRA experiments.

3.2.19.15.8. Comparison of MPRA activity between “linked” and “non-linked” enhancers

In those analyses, motifs were called using FIMO[324] with the following parameters: --verbosity 1 –thresh 1.0E-4. The complete HOCOMOCO v11 database was used[323]. Only USFs and TFs expressed in astrocytes according to Zhang et al., (2016) were considered. Average PhastCons 30-way score was calculated for 5 bp bins.

3.2.19.15.9. Comparison between SNP variants

For each SNP for which more than two variants were assessed using MPRA, the most active variant was considered the MaxVar. To perform human-chimpanzee sequence comparison, all nucleotides corresponding to SNP positions were mapped to pt6 reference genome using liftOver (-minMarch=0.1 -bedPlus=6). Sites which differed in nucleotide, or were not mapped, were considered to be different between the two species. For pair-wise significance testing, normalised reads from DeSeq2 were considered. Significance was tested using one-sided paired t-test. To determine TFBS change between variants, motifs called using FIMO[324] with the following parameters: --verbosity 1 –thresh 1.0E-4 were considered for ach variant. The complete HOCOMOCO v11 database was used[323] for FIMO annotation. Essentiality score for SNPs was derived by determining the highest score for any basepair within motifs described in HOCOMOCO v11 database. PhyloP scores were retrieved for SNP positions. Only USFs and TFs expressed in astrocytes according to Zhang et al., (2016) were considered.

3.2.19.16. Characterisation of HS-PT differences in USF in enhancers

Globally called TFBS, as described above, were considered. SNCs and USF sites were annotated in TFBS overlapping active and poised enhancers. Only USFs and TFs expressed in astrocytes according to Zhang et al., (2016) were considered. To perform human-chimpanzee sequence comparison, all nucleotides within detected motifs were mapped to pt6 reference genome using liftOver (-minMarch=0.1 -bedPlus=6). Sites which differed nucleotide, or were not mapped, were considered to be different between the two species. For TFBS, the percentage of such differences was calculated.

3.2.19.17. Analysis of RT-qPCR

Standard curve was constructed for each qPCR measurement. Expression values were calculated from average Cq values from two technical replicates. Expression value of each gene was normalised to expression of beta actin in respective sample. The expression was then normalised to expression in “scrambled” sample for each respective replicate.

3.3. External data used in analyses

In this section, the data not generated in our laboratory which was used in this study is listed.

File name	Contents	Source/internet link
gwas_catalog_v1.0.2-associations_e113_r2024-11-20	GWAS SNP database, according to European Bioinformatics Institute	https://www.ebi.ac.uk/gwas/docs/file-downloads
HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HoCoMoCo database of TFB motifs	https://hocomoco13.autosome.org/downloads_v11
Girskis_HAR_coordinates_Data_S2.csv	Coordinates of HARs from multiple studies re-mapped to hg38 genome assembly	10.1016/j.neuron.2021.08.005
HumanDerived_SNC_bothgq30.all_combined_maxsco_ranked	single-nucleotide changes (SNCs) that differ between humans and ancient hominids.	http://cdna.eva.mpg.de/neandertal/altai/catalog/HumanCatalog/
fe-wp-dataset-124.csv	Comparison of gene expression between cell types Zhang et al., (2016)	https://brainrnaseq.org/wp-content/uploads/2022/09/fe-wp-dataset-124.csv
granges1.1.r	Tools for granges	http://research.stowers.org/cws/CompGenomics/Resources
hg38.phastCons30way.bw	PhastCons 30-way sequence conservation score	https://hgdownload.soe.ucsc.edu/goldenPath/hg38/phastCons30way/
hg38.phyloP30way.bw	PhyloP 30-way sequence conservation score	https://hgdownload.soe.ucsc.edu/goldenPath/hg38/phyloP30way/

4. Results

Part of this work has been published as:

Ciuba, K., Piotrowska, A., Chaudhury, D., Dehingia, B., **Duński, E.**, Behr, R., et al. (2025). Molecular signature of primate astrocytes reveals pathways and regulatory changes contributing to human brain evolution. Cell Stem Cell 32, 426-444.e14. doi: 10.1016/j.stem.2024.12.011.

The manuscript describing results from Massively Parallel Assay (MPRA) described below (Duński et. al) is currently in preparation.

4.1. Astrocyte model

To generate an *in vitro* model suitable for studying the astrocyte biology, I employed iAstrocytes previously established in our laboratory, as described in our publication, where I am a co-author Ciuba et al., (2025). The iAstrocytes were produced following the procedure described by TCW et al., (2017) and modified by our group. Briefly, iPS cell lines were differentiated into NP cells by inhibiting TGF-β/BMP-dependent SMAD signalling. NP cells were then purified using flow cytometry to exclude cells expressing the mesenchymal lineage marker CD271[334] and enriched for those expressing the neural stem cell markers CD184 and CD133[335]. The resulting NP cells were subsequently differentiated into iAstrocytes as outlined in the Materials and Methods section (Figure 1a).

Cell identity was confirmed by immunofluorescence, which was used to test for expression of canonical astrocyte markers, including calcium-binding protein S100β, glial fibrillary acidic protein (GFAP), vimentin (VIM), and the excitatory amino acid transporter (EAAT1, also referred to as solute carrier 1 A 3, SLC1A3; Figure 1b). Furthermore, functional assays demonstrated hallmark astrocytic properties that are crucial for astrocyte function in the brain: iAstrocytes responded to extracellular ATP by generating intracellular calcium (Ca^{2+}) spikes (Figure 1c) and efficiently removed glutamate from the culture medium (Figure 1d). When co-cultured with rat embryonic cortical neurons, the iAstrocytes promoted neuronal survival and synapse formation (Figure 1e). It is important to note that iAstrocytes generated using this protocol exhibit a foetal-like gene expression profile[286].

Based on these phenotypic and functional characterisations, I concluded that the iAstrocytes represented a suitable model for subsequent experiments.

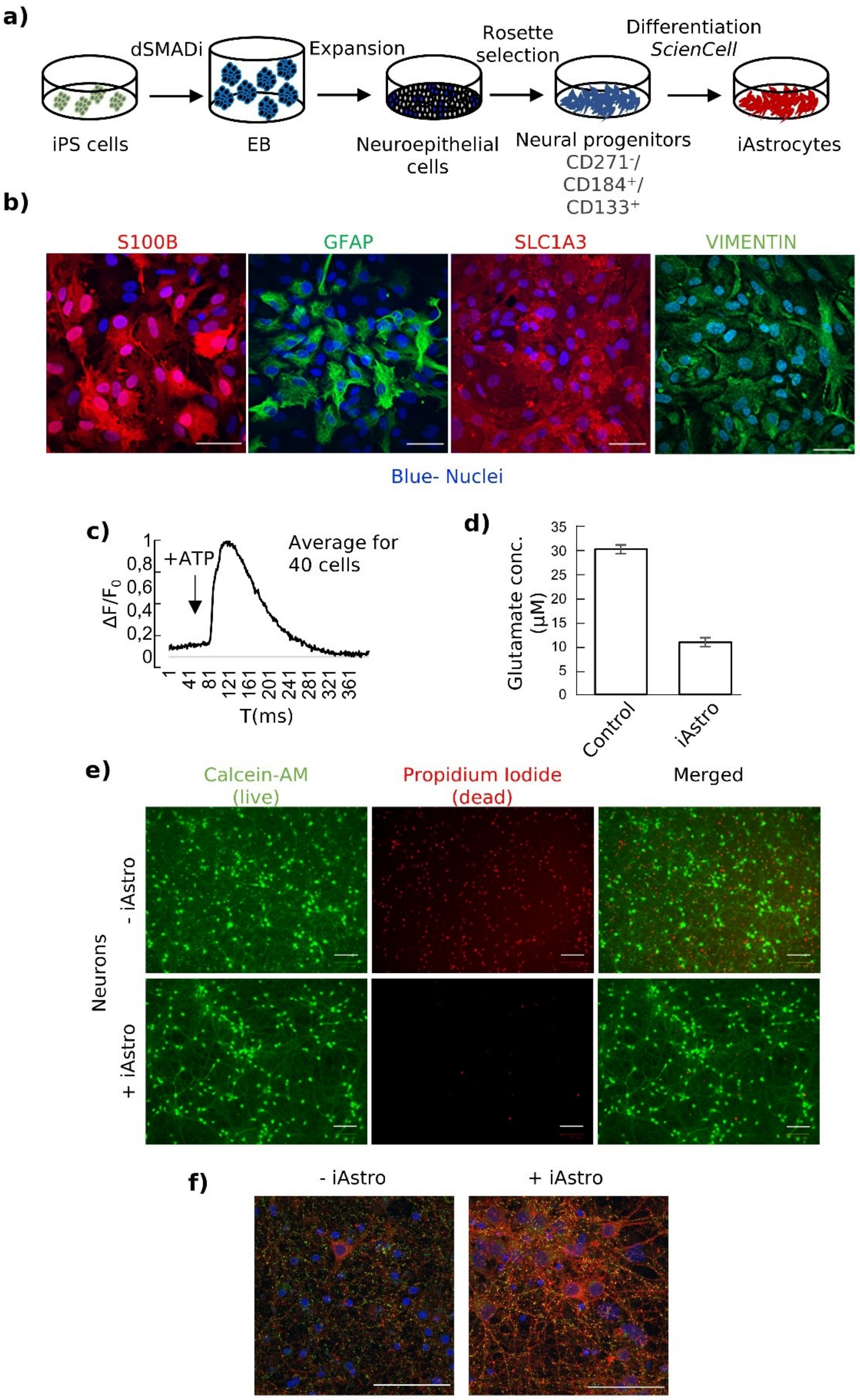


Figure 1. Generation and characterisation of iPS cell-derived astrocytes (iAstrocytes). a) iAstrocytes were derived from iPS cells as described in the Materials and Methods section and in Ciuba et al. (2025). In brief, dual SMAD inhibition (SMADi) was applied to generate neural progenitor (NP) cells. NP cells were subsequently immunopurified by flow cytometry-assisted cell sorting to enrich for CD271-negative and CD184 and CD133-double-positive cells. The purified NPCs were differentiated into iAstrocytes using *ScienCell* media. b) Representative

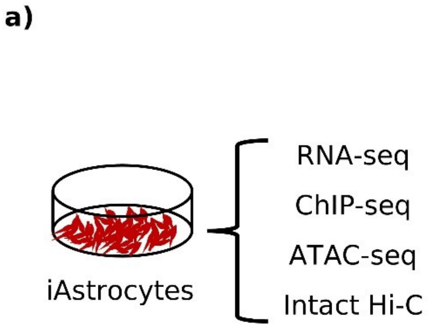
immunofluorescence staining for canonical astrocyte markers: S100β, GFAP, SLC1A3, and VIM. Blue – cell nuclei stained with Hoechst. c) Representative measurement of intracellular calcium dynamics in response to ATP stimulation. The time point of ATP addition is indicated with an arrow. d) Glutamate uptake of iAstrocytes. Glutamate concentration in the medium after one hour incubation with iAstrocytes was assessed with a commercially available kit. Control: a condition without iAstrocytes. Error bars indicate standard deviation (n=3). e) Representative assay assessing the effect of iAstrocytes on the survival of embryonic rat cortical neurons (E18). Green denotes live while red denotes dead cells. f) The iAstrocytes promote synapse formation of embryonic rat cortical neurons (E18). Acutely purified neurons were co-cultured in the presence or the absence of primate iAstrocytes for 10 days. The presence of pre- (Bassoon, green) and post-synaptic (Homer, red; inlet cartoon) markers was assessed in the cultures using immunofluorescence. Scale bar in panel e: 100 µm, other panels: 50 µm.

4.2. Characterization of astrocyte regulome

The transcriptional program defines and reflects the cell's identity. The "regulome" is the complete set of regulatory elements active in a given cell type and controlling its gene transcription[336]. As most of the genetic differences between individuals (including disease-related single-nucleotide polymorphisms (SNPs) and closely related species reside in the regulome, studying the biology of regulatory elements can help reveal the mechanisms underlying disease and evolutionary changes in gene expression [337,338]. However, the relationship between the sequence of regulatory elements and foetal astrocyte function in health and disease remains poorly understood. Likewise, while there are studies that identify the specific enhancers active in human astrocytes[339] and human-specific changes in the astrocyte transcriptome[246–249,312], the regulome alterations that drive the evolution of the foetal astrocyte transcriptome remain unexplored.

To address these important knowledge gaps, I aimed to identify the genetic links between brain disorders and traits, and regulatory elements active in foetal astrocytes. Furthermore, I attempted to define which type of human-specific alterations of astrocyte regulome drive corresponding changes in the transcriptome. Finally, I investigated whether human-specific changes in astrocyte regulome could illuminate the evolutionary origins of astrocyte function in the human brain and potential vulnerability to neurological disorders.

My work contributed to obtaining a detailed cartography of putative regulatory sequences in human, chimpanzee, and macaque astrocytes (Figure 2a). In particular, together with my colleagues Aleksandra Piotrowska and Dr. Katarzyna Ciuba, we mapped regions of open chromatin using ATAC-seq[276] and putative active elements by pulling down sequences associated with histone H3 lysine 27 acetylation (H3K27ac), an established active enhancer-associated mark[277,278]. To further distinguish active promoters, our lab conducted ChIP-seq for histone H3 lysine 4 trimethylation (H3K4me3), as this modification is a well-established mark of promoter elements in the genome[340,341]. Active enhancers frequently engage in physical interactions with their cognate promoters, and such contacts can be determined using chromatin conformation capture techniques. Together with dr. Anna Stroynowska-Czerwińska, a Postdoctoral fellow in the laboratory, we identified loops between regulatory elements using Intact Hi-C (protocol provided by prof. Erez Aiden-Lieberman, from the Baylor College of Medicine in Texas, USA). Finally, to examine the relationship between the regulome and transcriptional output, my colleagues in the group performed RNA-seq on iAstrocytes.



b)

Category	**Number of Peaks**
Total ATAC-seq peaks	150642
Total ChIP H3K27ac peaks	93590
Total ChIP H3K4me3 peaks	20787
Promoters (H3K4me3/TSS)	20592
Enhancers (ATAC-seq, not TSS/H3K4me3)	120100
Poised enhancers (Enhancers, not H3K27ac)	67265
Active enhancers (Enhancers, H3K27ac)	52835

Figure 2. Characterisation of the iAstrocyte regulome. a) The iAstrocyte model was used to profile gene expression (RNA-seq), histone modifications (ChIP-seq), regions of open chromatin (ATAC-seq), and chromatin conformation (Intact-C). b) Numbers of ATAC-seq and ChIP-seq peaks included in the analysis and their associated genomic features.

To generate the set of ATAC-seq peaks, sequencing data from two iAstrocyte lines (ELE30.2 and ELE10) and three sequencing libraries (one for ELE30.2 and two for ELE10) were analysed. ChIP-seq peaks were called from data obtained for the same two iAstrocyte lines, with one sequencing library per line. In total, 150,642 ATAC-seq peaks, 93,590 H3K27ac ChIP-seq peaks, and 20,787 H3K4me3 ChIP-seq peaks were identified (Figure 2b).

Using these data together with Ensembl gene annotations, I defined 20,592 promoters as ATAC-seq peaks intersecting the intervals ±1,000 bp around transcription start sites (TSS), and H3K4me3 peaks, and 120,100 distal regulatory elements (the remainder of ATAC-seq peaks). Based on the ChIP-seq data, 67,265 ATAC-seq peaks that did not overlap either H3K4me3 or H3K27ac peaks nor promoter intervals were defined as poised enhancers, and 52,835 ATAC-seq peaks that did not overlap H3K4me3 peaks but overlapped H3K27ac peaks were identified as active enhancers. These definitions formed the basis for subsequent analyses.

4.3. Chromosome conformation capture in human iAstrocytes

To link specific regulatory elements to their cognate promoters, we employed Intact Hi-C, which detects chromatin loops at high resolution. Integration of Intact Hi-C with ATAC-seq and H3K27ac ChIP-seq allowed precise mapping of regulatory landscapes for individual genes, including key astrocytic genes *SLC1A3* (encoding EAAT1) and *VIM* (encoding vimentin) (Figures 3 and 4).

Using the Juicer tool[318], I identified 34578loops at the resolution of 5kb. A significant percentage of poised (23.9%) and active (29.4%) enhancers were positioned at loop anchors (Figure 5 a, $p = 4.06*10^{-102}$, Fisher's exact test), indicating a relationship between transcriptional regulation by remote DNA elements and the formation of chromatin interactions. To determine whether these interactions were mediated by the architectural protein CTCF, I considered CTCF ChIP-seq data (obtained by Aleksandra Piotrowska). Notably, CTCF peaks were more often found to intersect putative poised (13.8%) than active enhancers (6.3%), further supporting the proposal that we are detecting genuine promoter-enhancer interactions implicated in the regulation of gene expression.

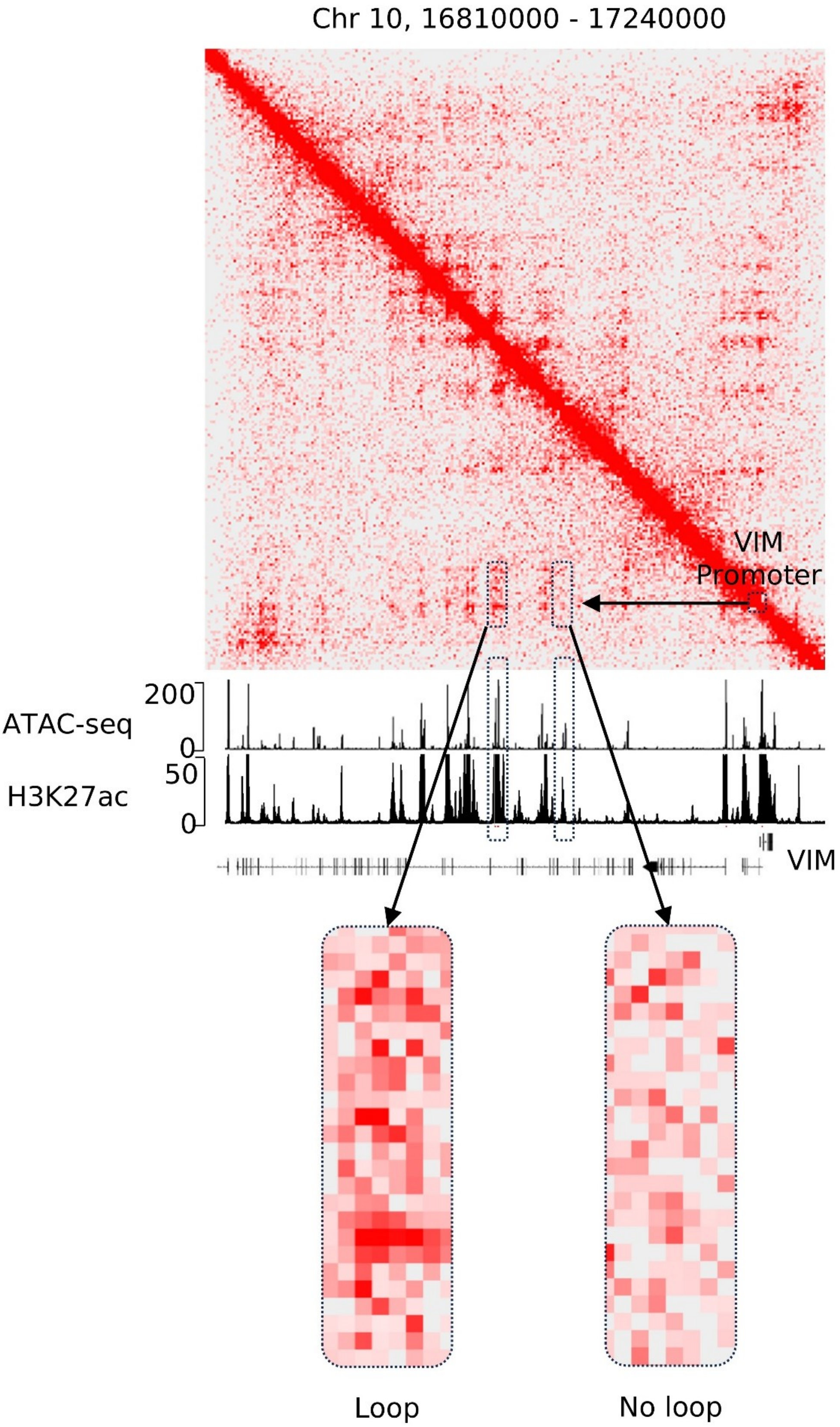


Figure 3. Intact Hi-C in iAstrocytes. Chromatin conformation at the Vimentin (VIM) locus. Boxes indicate example interactions. Visualisations generated using JuiceBox and the Integrated Genome Browser (IGB) tools.

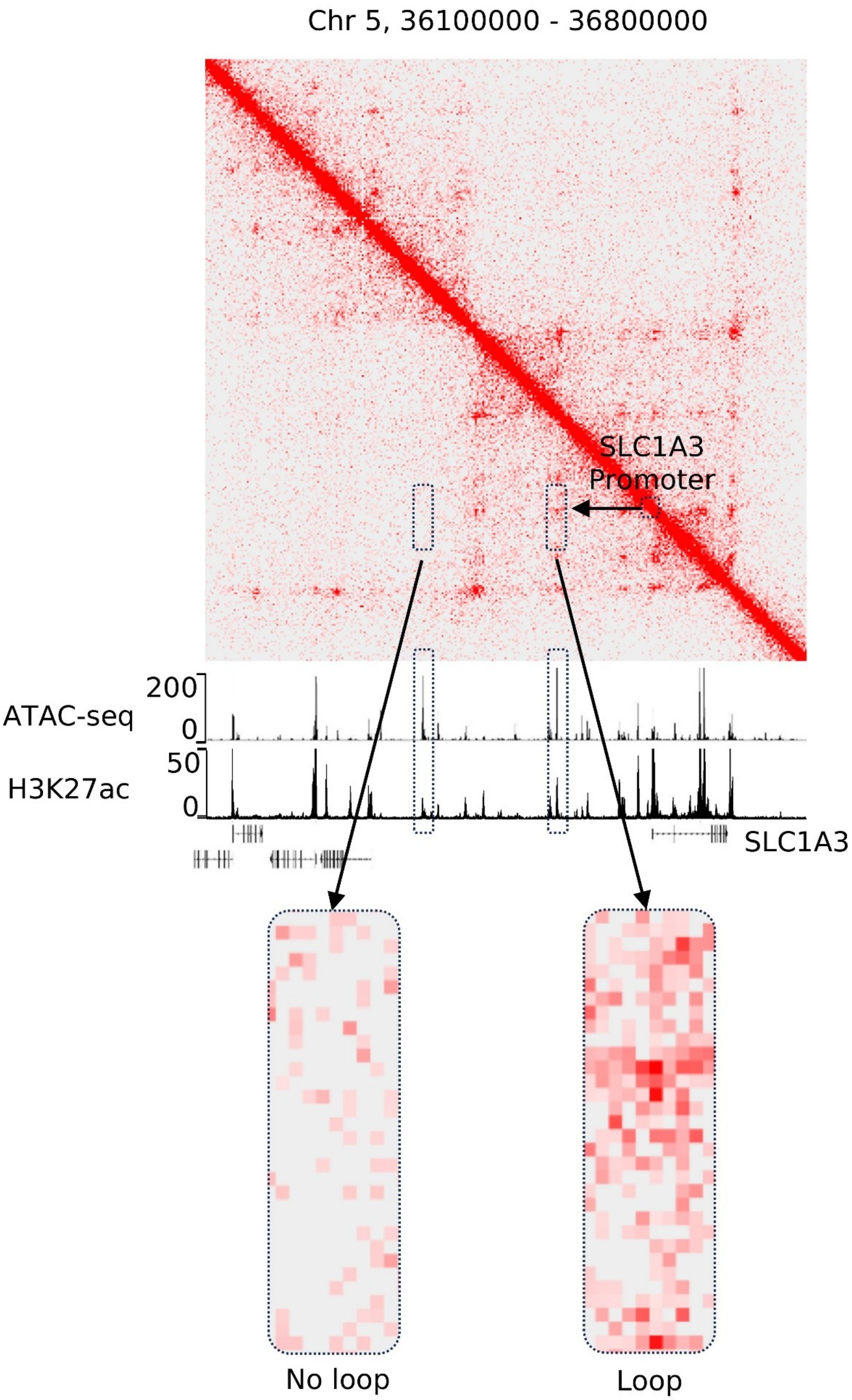


Figure 4. Chromatin conformation, openness and H3K27ac profiles at the SLC1A3 locus. Boxes indicate example interactions. Visualisations generated using JuiceBox and the Integrated Genome Browser (IGB) tools.

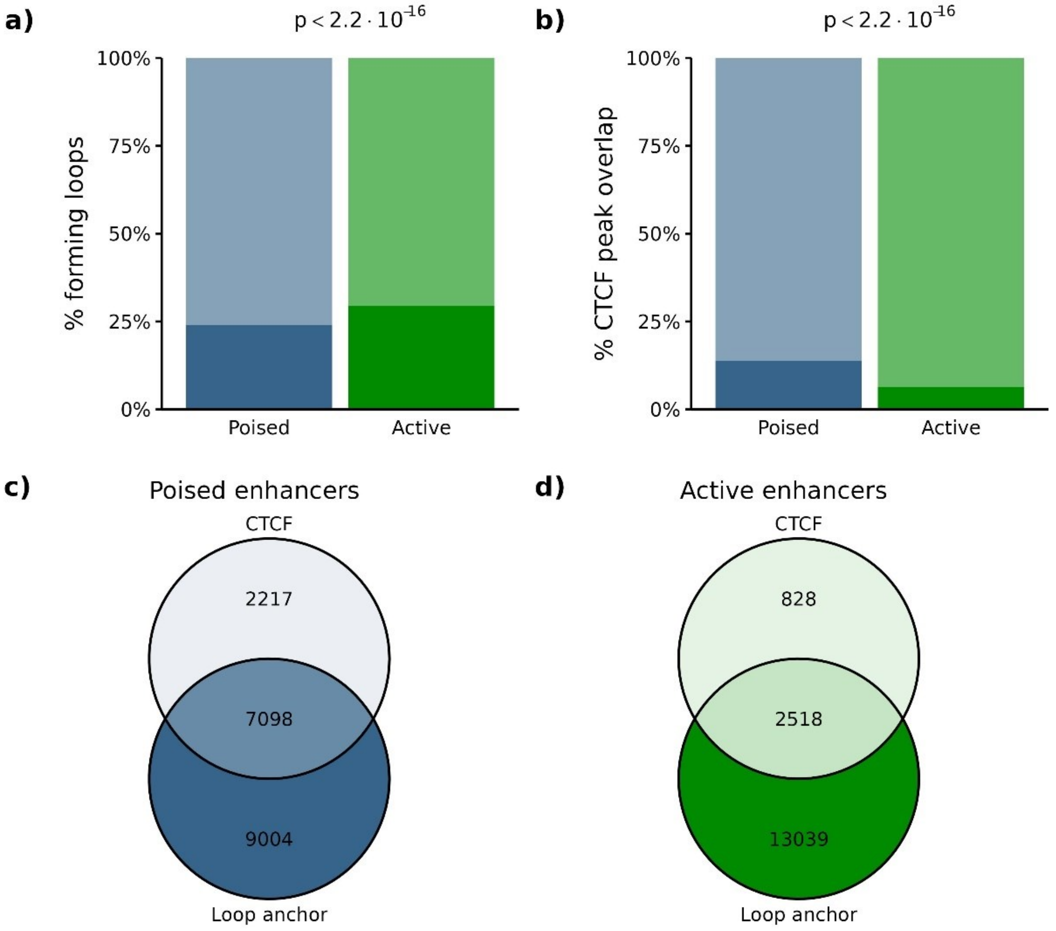


Figure 5. Loop formation and overlap with CTCF peaks. a) Fraction of poised and active enhancers that form loops. b) Fraction of poised and active enhancers that overlap CTCF peaks. Panels c) and d) indicate the overlap between CTCF peaks and loop anchors, in poised and active enhancers, respectively.

4.4. Transcription factor binding sites (TFBS) in active astrocyte enhancers

While the transcription factors (TF) that are key to neuronal cell fate commitment are increasingly better defined, those that shape foetal astrocyte regulomes remain less well understood[342–345]. One approach to address this question is to identify TFs whose binding motifs are overrepresented within enhancers. To this end, I considered all 52,835 active enhancers defined in the previous section and performed a scan using the FIMO tool from MEME suite[346,347] (Figure 6).

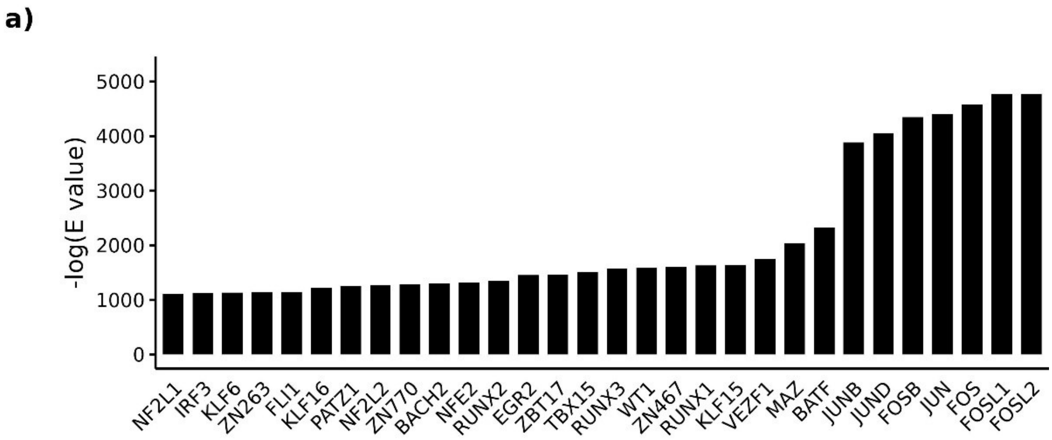


Figure 6. TFBS detected in HUMAN active enhancers using FIMO. Top 30 TFBS most enriched at enhancers active in HUMAN iAstrocytes. Motifs are rank-ordered, according to the -log10(E-value). E-value corresponds to the adjusted P-value obtained using the Fisher test when computing the motif enrichment in the assayed set of elements (enhancers) compared to the background (SEA tool in MEME suite[346,347].

Notably, TFBS of members of FOS and JUN TF families are the most enriched within active enhancers, consistent with previous observations in astrocytic precursors isolated from the foetal human brain[345]. FOS and JUN TFs together form the AP-1 transcription factor complex[348]. Interestingly, FOS activity in astrocytes has recently been shown to be crucial for memory formation[225,226]. This suggests that some of the key transcription factors influencing human cognition, including FOS, may preferentially regulate enhancers active in foetal astrocytes.

4.5. Disease-associated variants are enriched in the astrocyte regulome

Most of the trait- and disease-linked variants map to distal regulatory elements. Recognizing this, one can assess the contribution of specific cell types to shaping traits and disease susceptibility. To explore the relevance of foetal astrocytes in human brain diseases, I mapped SNPs retrieved from the EBI GWAS Catalogue (Sollis et al., 2023, accessed 21 November 2024) to enhancers identified in iAstrocytes (Figure 7). I observed a significant enrichment of both poised and active enhancers among all SNPs, compared to a set of randomly selected background regions (Figure 7a, for poised enhancers: 1.47-fold enrichment, $p = 3.19*10^{-51}$; for active enhancers: 1.58-fold enrichment, $p = 3.62*10^{-67}$, Fisher's exact test). These results indicate that the foetal astrocyte regulome is enriched for polymorphisms present in the human population and may thus contribute to interindividual variability.

To determine whether polymorphic sites within the astrocyte regulome are specifically overrepresented for variants that affect brain-related phenotypes, I examined the enrichment of all phenotype-associated SNPs from the GWAS-EBI database, including variants associated with neuropsychiatric disease. I focused on phenotypes that passed a Q-value threshold of 0.1 and were broadly related to brain biology, morphology or disease. This analysis revealed six brain structure-related phenotypes that were significantly enriched among SNPs overlapping active astrocyte enhancers (Figure 7b). Among these, "Vertex-wise cortical thickness", "Vertex-wise sulcal depth", and "Vertex-wise cortical surface area"[350,351] represent highly heritable measures of cortical patterning and folding. "Whole brain restricted isotropic diffusion (multivariate analysis)"[352], which showed more than 5-fold enrichment in active enhancers, measures diffusion of water within the brain, which allows for an estimation of proportions of cell bodies, axonal fibres, and interstitial fluids – thus broadly capturing brain microstructure[352]. The strong enrichment of this phenotype among single-nucleotide variants located in active enhancers highlights the importance of astrocytes for the organisation of brain structure. The observed association between the astrocyte regulome and water circulation in the brain is particularly intriguing, taking into account the most recent discoveries showing prominent movement of the astrocytic endfeet during glymphatic waste clearance during sleep[353]. Clearance of cerebrospinal fluid from the brain is disrupted in dementia[354], and is crucial for the removal of amyloid β, α-synuclein, and tau associated with AD[143,146]. Thus, the association of astrocyte regulome with water circulation in the brain suggests broader implications for understanding the astrocytic contribution to brain homeostasis and neurodegenerative disease.

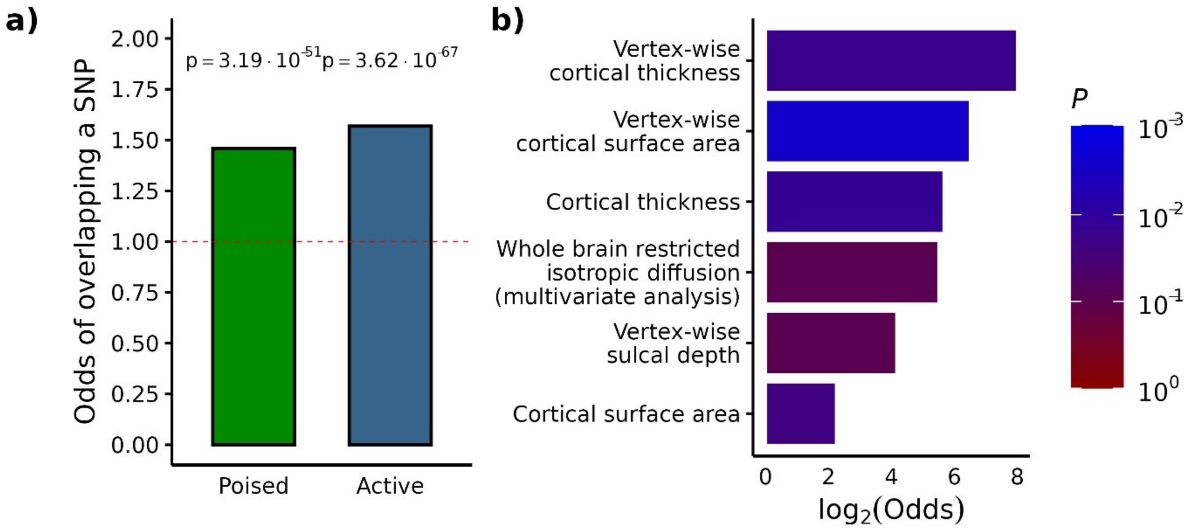


Figure 7. Astrocyte regulome is associated with disease. a) Enrichment of GWAS SNPs associated with any phenotype in active and poised enhancers relative to randomly selected genomic regions. b) GWAS SNPs associated with brain morphology phenotypes that are enriched in active enhancers.

The strong correlation between SNPs affecting those measures with the regulome of iAstrocytes is important, especially for foetal astrocytes. This is because previous studies have shown that cortical folding is largely determined prenatally, as evidenced by magnetic resonance imaging of the brain of unborn children[355] and prematurely born infants[356], as well as by measurements of sulcal depths in postmortem foetal and neonatal brain specimens[357]. Abnormalities in brain shape, including gyrification index, have been shown to persist into later life, in both children[358] and adults[359].

Emerging experimental evidence suggests that brain folding is, in large part, shaped by astrocytes. Studies in the macaque brain have demonstrated that increased astrocyte proliferation in the subventricular zone contributes to gyrification[360]. Moreover, experiments on ferrets, which possess brain folds, and mice, which do not, have shown that manipulation of fibroblast growth factor (FGF) signalling can modulate astrocyte proliferation and, in turn, induce cortical folding[361]. Thus, our iAstrocyte regulome analysis identifies the elements that may underlie the degree of cortical folding by regulating astrocyte biology.

Prenatally established features of brain shape, including gyrification index, not only persist into later life, in both children[358] and adults[359], but are also related to brain functions and disease susceptibility. In this respect, brain folding is related to IQ[359]. Furthermore, the Impairment of cortical folding due to premature birth, results in behavioural abnormalities, including lowered attention span[356], indicating that the depth and distribution of sulci are interconnected with the regulation of higher-level brain functions. Likewise, abnormalities in brain folding have been observed in disease[356,362,363]. including bipolar disorder[364], SCZ[365], ASD[366], and PD[367]. Collectively, these findings suggest that astrocytes, by influencing the development of brain morphology, may play a crucial role in shaping cognitive function.

Dihydropyrimidinase-Like 2 (DPYSL2, also known as Collapsin Response Mediator Protein 2, CRMP-2) is genetically linked with the structure of the brain connectome [368]. DPYSL2 is implicated in the regulation of neuronal plasticity, in part due to its functions in regulating the response to semaphorins[369] and in the regulation of axonal growth[370], neuronal migration[371], and axon pruning[372]. Genetic engineering with CRISPR-Cas9 in human iPS cells revealed that DPYSL2 is critical for the proper neuronal functions[373], indicating a conserved role of this factor in brain biology.

Furthermore, DPYSL2 was shown to be related to SCZ, both in the mouse model, where its deletion has been shown to result in SCZ-like symptoms[374], and in humans, where it is dysregulated in SCZ patient brains[375]. Hence, this gene is important for correct neuronal development and brain connectivity, and its dysregulation may potentially contribute to the diseases of the human mind.

Rs5029306 is associated with brain morphology variation[351], brain shape variation[376], cortical surface area and shape[351,377]; this SNP is located in close proximity to one of the promoters of DPYSL2. Notably, DPYSL2 is highly expressed in astrocytes[312], where it regulates the response to injury[378]. Our data show that DPYSL2 is regulated by the same promoter that is crucial for its functions in neurons[373]. Thus, the position of enhancer containing rs5029306 in proximity, and in the same contact loop as the promoter of DPYSL2, suggests that this enhancer may play a role in DPYSL2 regulation in astrocytes and, consequently, in regulating brain connectome and SCZ susceptibility.

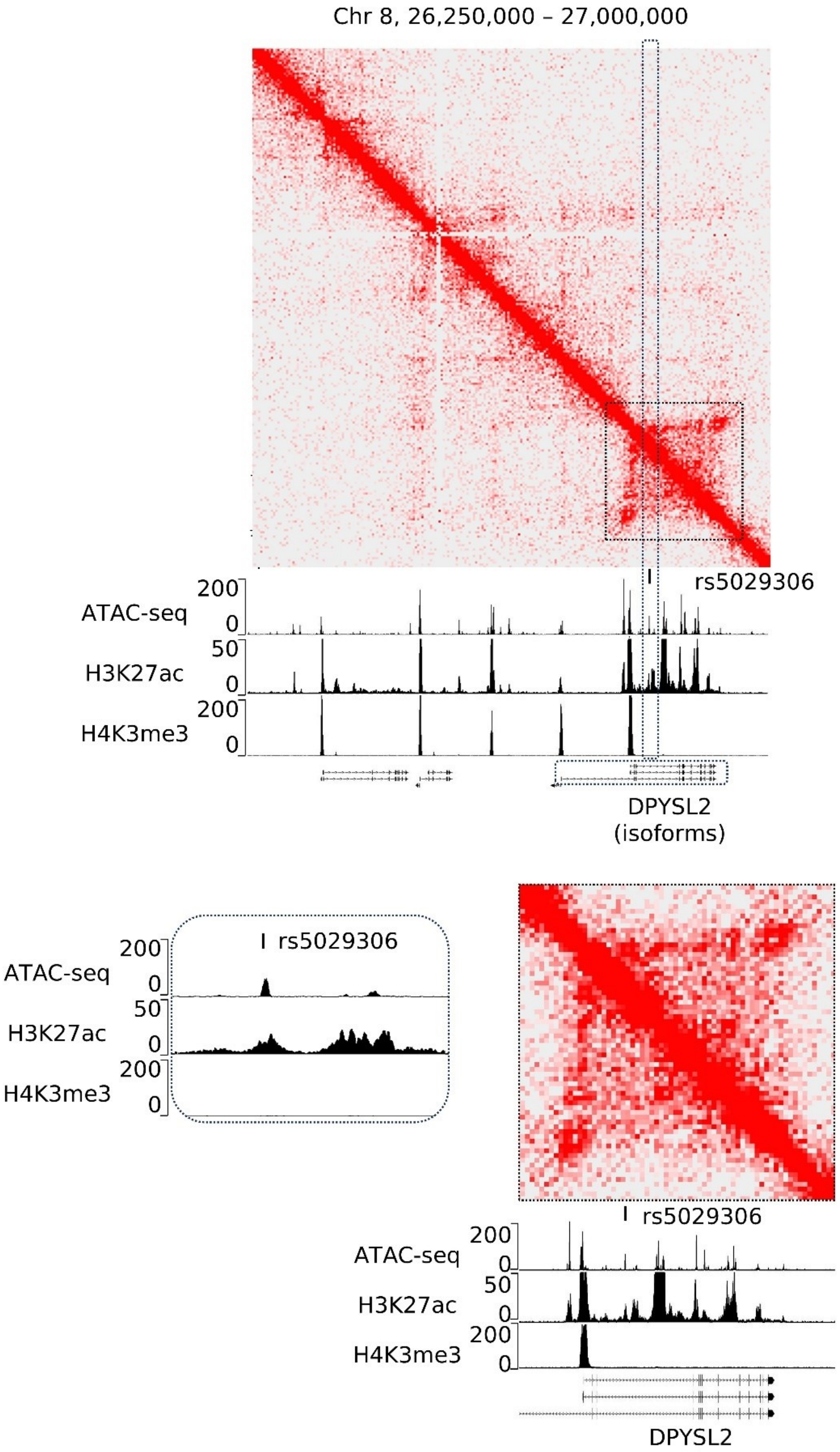


Figure 8. Genomic context of brain morphology-related SNP rs5029306. Visualisations generated using JuiceBox and Integrated Genome Browser (IGB). Local chromatin state around the enhancer is shown in the box in the lower panel.

Rs35904163 (Figure 9) is associated with the MRI parameter of whole brain restricted isotropic diffusion, a broad measure of brain structure[352]; rs35904163 is in proximity of three genes playing roles in the central nervous system: POU domain, class 3, transcription factor 2 (*POU3F2*), long non-coding RNA (lncRNA) *Pnky*, and F-box and Leucine-rich repeat-containing protein 4 (*FBXL4*). Each of those genes was shown to contribute to brain developmental behavioural control[379].

Deletion of the whole region (6q16.1) causes a developmental syndrome characterised by intellectual disability[380], indicating an important role of this locus in the regulation of brain physiology. Transcription factor POU3F2 (also called BRN2) is important for brain development and has been related to primate brain expansion, as demonstrated by knockout studies in the cynomolgus monkey (*Macaca fascicularis*)[381]. Deletion of *FBXL4* has been related to developmental delay in humans[382]. Small deletions encompassing *POU3F2* and *FBXL4* are related to intellectual disability in humans[380,383], and the activity of *POU3F2* is likely central to the correct regulatory network underlying higher-level brain functions[384–387]. Likewise, fluctuations in levels of transcript of the long non-coding RNA *Pnky* result in behavioural changes in a mouse model [388].

POU3F2 (BRN2) is one of the core “neurogenic” transcription factors used to convert fibroblasts into induced neurons (iNs), along with *ASCL1* and *MYT1L*[389–391]. *POU3F2* is expressed in radial glial progenitors and intermediate progenitor cells, helping balance self-renewal vs. neuronal differentiation. Notably, *POU3F2* is expressed primarily in astrocytes (the human protein atlas); and its overexpression is sufficient to promote neuronal conversion of astrocytes[392]. These data indicate the need for a tight regulation of *POU3F2* expression in astrocytes and suggest a possible role of *POU3F2* in the regulation of genes expressed in both neurons and astrocytes in glia. However, its role and regulatory network in astrocytes remain unknown (see below).

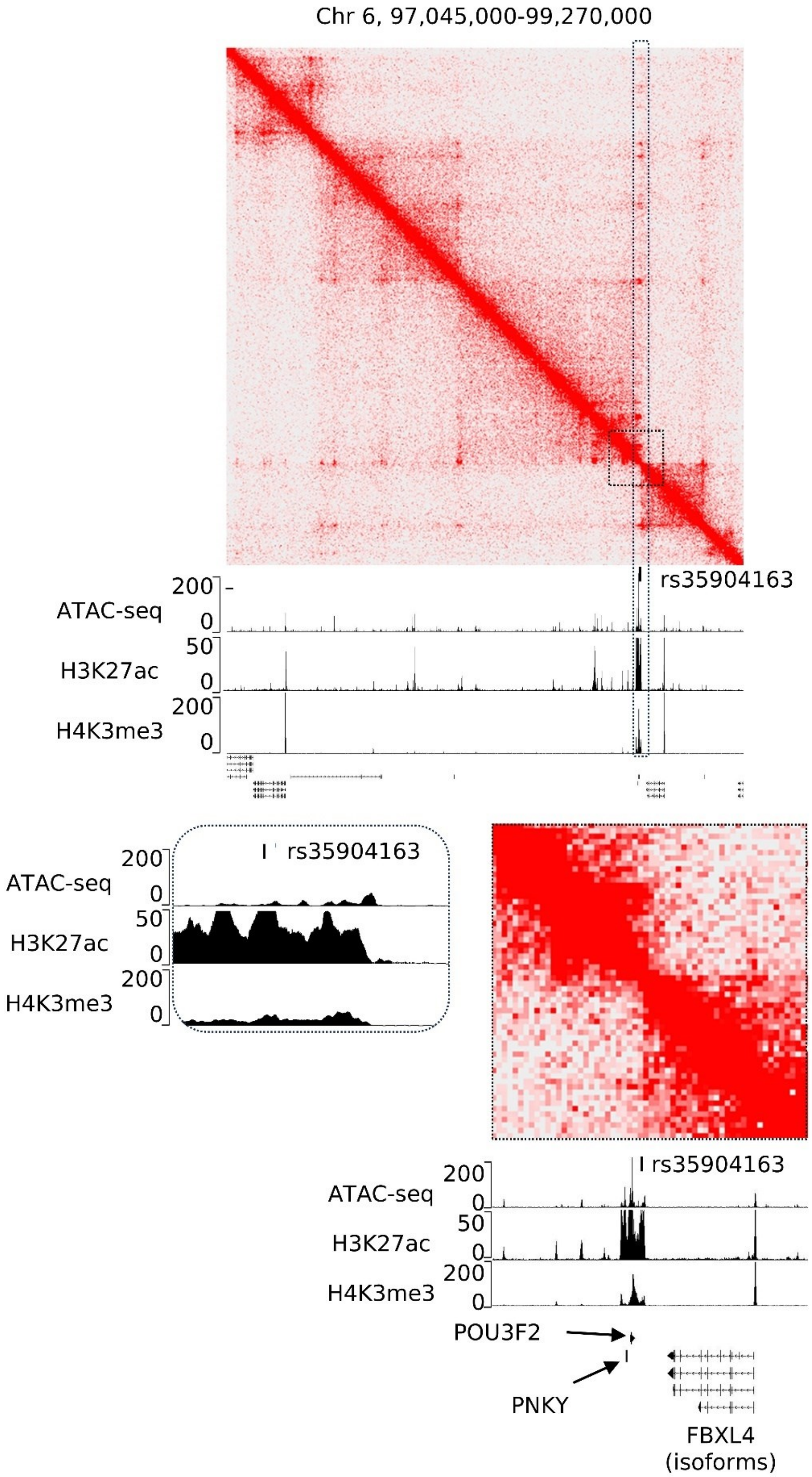


Figure 9. Genomic context of brain morphology-related SNP rs35904163. Visualisations generated using JuiceBox and Integrated Genome Browser (IGB). Local chromatin state around the enhancer is shown in the box in the lower panel.

4.5.1. Disease-associated variants and transcription factor binding sites

To fulfil their role of integrating signalling cues to properly regulate gene expression, functional enhancers feature a high number of TFBS motifs[280,393], which are disrupted by SNPs. Such disruptions may alter the dynamics of interactions between enhancers and transcription factors, thereby affecting their ability to modulate transcription. To investigate the effect of SNP variants on TFBS integrity, I derived a score to rank order SNPs according to their likelihood of disrupting a TFBS. Through this analysis, I hoped to identify key TFs related to astrocyte biology that could illuminate the mechanisms by which these cells contribute to higher-level brain functions.

I employed FIMO[324] to perform a genome-wide annotation of putative TFBS. Next, I considered the probability score of encountering any given nucleotide at a given position within the predicted TFBS, given its position-specific weight matrix (PWM, HOCOMOCO v.11 database [323]). This metric, hereafter referred to as "basepair criticality", ranges from 0.25, for a completely degenerate position within a TFBS, to 1, for a position in which the given nucleotide is critical for the TFBS, and modification of which would result in abolition of the predicted binding site of a TF (Figure 10).

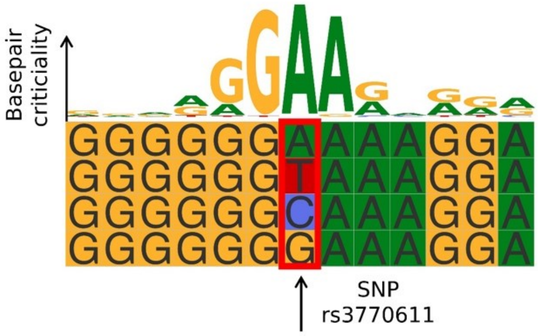


Figure 10. Determining the effect of SNP variants on TFBS in enhancers. An example of an ETS2 binding site (position chr2:169,286,977-169,286,991 that is disrupted by variants of SNP rs3770611 other than A. The TFBS logo is displayed in the upper part of the panel. The importance of each base in the TFBS is indicated by its height in the PWM. In the bottom panel, different SNP variants are shown. Only A variant would result in a predicted ETS2 binding site.

Then, I used this metric to assess the TFBS which may be disrupted by SNPs related to phenotypes describing measures of brain morphology, intellectual ability, and neuropsychiatric diseases, based on EBI GWAS Catalogue (Sollis et al., 2023, acessed on 21.11.2024). I focused on SNPs overlapping active enhancers (ATAC-seq/H3K27ac) detected in iAstrocyte regulome.

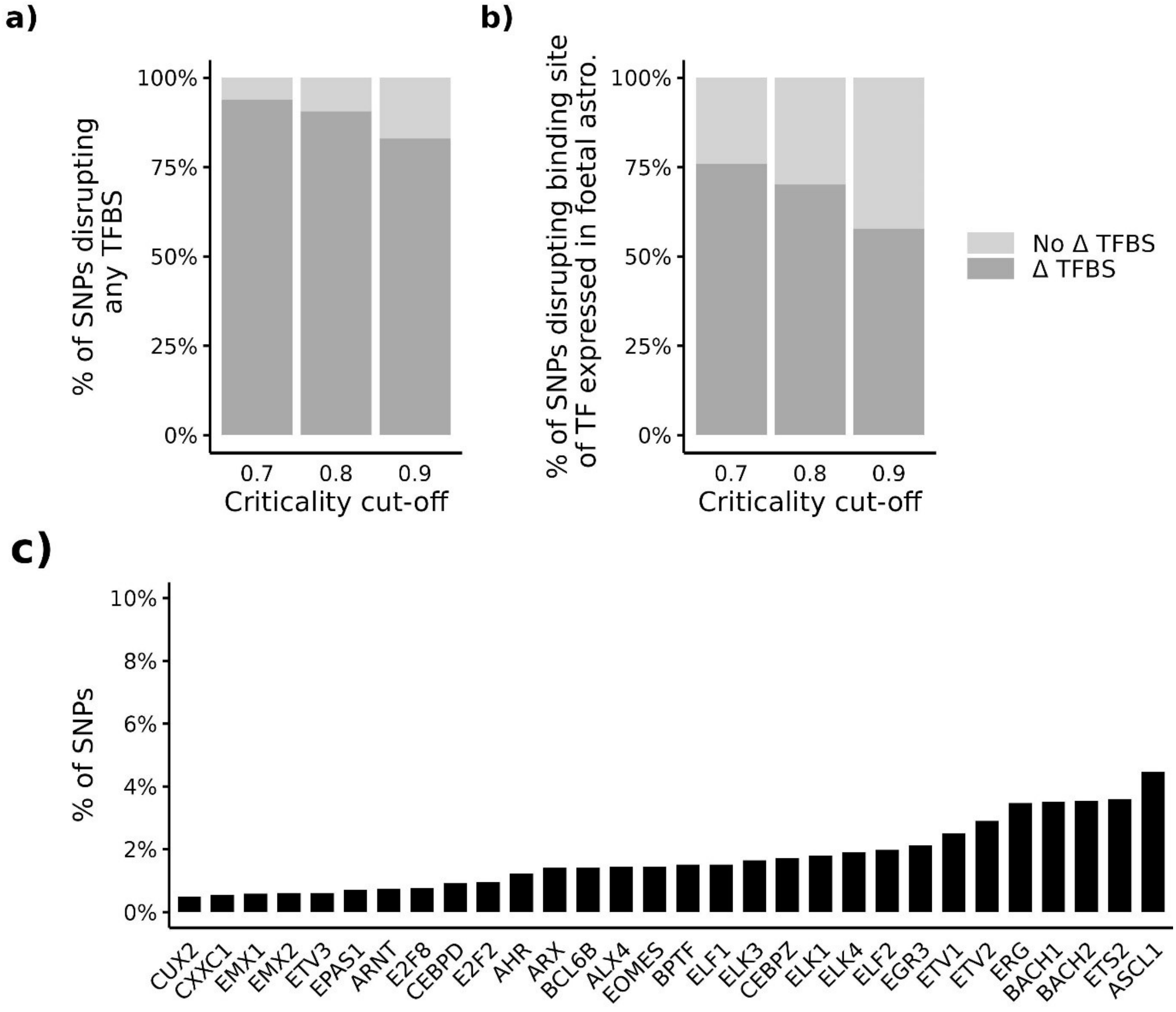


Figure 11. **Effect of SNPs in active astrocyte enhancers on TFBS.** a) Percentage of SNPs predicted to disrupt a TFBS, depending on basepair criticality threshold used. All possible TF were considered. Darker shade - the percentage of SNPs for which one variant affects a TFBS, lighter shade – the percentage of SNPs for which one variant affects a TFBS. b) Percentage of SNPs predicted to disrupt a TFBS, depending on basepair criticality threshold used. Only TFs expressed in foetal astrocytes (Zhang et al., (2016)) were considered. Darker shade - the percentage of genes that have a given feature in their loop domain, lighter shade - the percentage of genes that do not have a given feature in their loop domain. c) TFs expressed in foetal astrocytes that are more frequently disrupted by SNPs, by the percentage of SNPs in which they appear. Only TFBS that are significantly enriched among disrupted SNPs are shown. Basepair criticality threshold = 0.7, significance of enrichment determined using Fisher's exact test.

On one hand, this approach helped me identify interesting enhancer elements that could link the enhancer regulome to brain features (see below). On the other hand, it allowed me to explore a broader question- namely, which pathways active in foetal astrocytes might be instrumental for brain morphology, intelligence, or brain disorders. I addressed this question by assessing which TFs exhibit motif disruptions more frequently than expected (see Materials and Methods section). Enrichment was determined using Fisher's exact test against a set of randomly selected single-nucleotide positions from active enhancers ($p < 0.05$, odds ratio > 1). At a basepair criticality threshold of 0.7, I found that more than 93% of SNPs located in astrocyte enhancers potentially impact at least one TFBS, when basepairs with basepair criticality greater than 0.7 are considered. More than 75% of SNPs potentially disrupt a binding site of a factor that is expressed in foetal astrocytes (selected according to Zhang et al., (2016), Figures 11a and b). ASCL1 is the TF for which its binding sites were the most frequently disrupted by SNPs located in active foetal astrocyte enhancers. Interestingly, this factor has been shown to play a crucial role in the determination of astrocyte differentiation fate[394]. It suggests that SNP variants mapping to astrocyte regulome may indeed create or abolish binding sites for important astrocytic TFs.

I then considered SNPs related to phenotypes describing measures of brain morphology, intellectual ability, and neuropsychiatric diseases. I found that TFs whose binding sites are frequently disrupted by SNPs related to intellectual ability are often TFs that have been shown to affect brain development. Such genes include mothers against decapentaplegic homolog 1 (*SMAD1*) and retinoic acid (RA) receptor gamma and the related retinoic acid receptor-related orphan receptor alpha (*RORA*). *SMAD1* controls astrocyte fate commitment[395] and is essential for BMP2-induced astrocytic differentiation during development. Its absence leads to fewer astrocytes and impaired proliferation, as shown in both *in vitro* and *in vivo* models[396,397]. RORA senses RA, which is a molecule crucial for brain development. Alterations in RA level result in neurodegenerative disease[398], and its receptors are involved in the modulation of AD-like symptoms in a mouse model[399]. Specifically for astrocytes, RA was shown to mediate the expression of astrocytic glutamate transporter, and therefore affects crucial homeostatic functions of astrocytes[400]. Furthermore, mutations in *RORA* in humans result in developmental disability, cerebellar features, and a spectrum of myoclonic epilepsy[401], underlining the role of RA signalling in the regulation of brain functions. *FOSL1* and *FOSL2* genes, which are frequently affected by intelligence-related SNPs, are also enriched globally in iAstrocyte active enhancers (see Figure 6). Taking into account the very recent findings indicating an important role of astrocytic FOS in memory encoding [402], I conclude that FOS binding sites active in astrocytic enhancers may be critical for human cognitive ability.

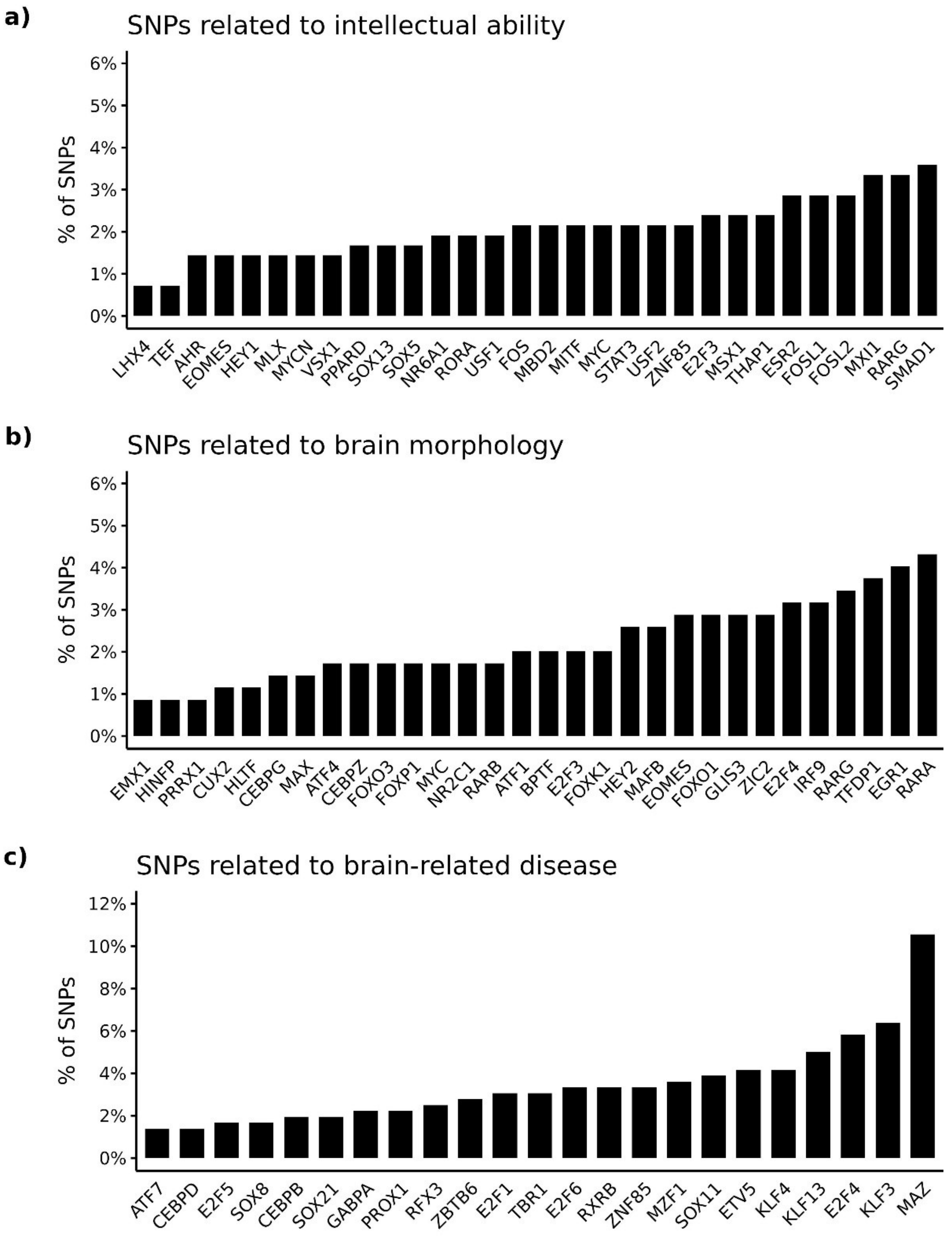


Figure 12. Effect of SNPs related to brain phenotypes (brain morphology, intellectual ability, and neuropsychiatric diseases) in active astrocyte enhancers on TFBS. TFs expressed in astrocytes that are disrupted by SNPs, by the percentage of SNPs in which they appear. Only TFBS that are significantly more frequently disrupted by each SNP category are shown (basepair criticality threshold = 0.7, enrichment determined using Fisher's exact test, p <0.05, odds ratio >1). a) SNPs related to intellectual ability. b) SNPs related to neuropsychiatric diseases. c) SNPs related to brain morphology.

Among TFBS that are more frequently disrupted by SNPs related to brain morphology (Figure 12b), RA receptors, including retinoic acid receptor alpha, beta, and gamma (RARA, RARB and RARG, respectively), stood out. As discussed above, RA is essential for proper brain development, and my analyses indicate an important role of RA signalling in astrocytes in the regulation of brain shape.

The second TFBS most commonly affected by SNPs related to brain morphology is EGR1 (Early Growth Response 1), which is required for Epidermal Growth Factor (EGF)-driven astrocyte proliferation[403].

Of note are also members of the forkhead box family (FOX)[404] - FOXO1, FOXK1, FOXP1, and FOXO3. FOXO1 plays a role in energy metabolism regulation through astrocytic signals in the hypothalamus[405]. The *FOXP1* gene is important for cortical development[406]. Likewise, its KO in mice causes a dramatic reduction of striatum size[407]. In humans, mutations in *FOXP1* result in an autosomal dominant syndrome which mental disability and ASD-like symptoms[408].

SNPs related to different aspects of brain disease, interestingly, most frequently affect binding sites of MAZ (Myc-associated zinc finger). MAZ has been shown to play a role in the regulation of gliogenesis, and its knockdown in mice was shown to inhibit it[409]. It has also been shown to neighbour binding sites for the well-defined insulator protein CCCTC-binding factor (CTCF)[410], recently shown to play a role in chromatin organisation, and thus, organism development[411]. Furthermore, this group of SNPs frequently affects Krüppel-like transcription factors (KLFs) – KLF4, KLF13, and KLF3. KLFs play diverse regulatory roles in cell growth and are highly expressed in the brain[412], but most of them do not have specific roles described in astrocytes; KLF4 is expressed in reactive astrocytes[413,414]. Notably, both KLF4 and MAZ belong to the USF family[1].

Interestingly, the TFBS that are frequently affected by SNPs from each of the three groups appear to have a multitude of effects in brain function – for example, RA receptors, KLFs, and FOX genes. It suggests that TFBS affected by the detected polymorphisms may have effects beyond the narrowly defined phenotypes associated with a given SNP.

Overall, these findings demonstrate that SNPs potentially affect important TFBS within astrocytic enhancers, and such changes may affect the binding of regulators with broad developmental and functional roles in the brain. This raises the question of whether similar mechanisms - through changes in enhancer architecture - may have also contributed to evolutionary modifications in the human regulome underlying brain development and function.

4.6. Astrocyte regulome is under evolutionary pressure

Alterations in enhancer landscape play critical roles in the development of complex traits[415,416]. The most studied group of human-specific regulatory changes is Human Accelerated Regions (HARs)[293]. HARs are genomic loci exhibiting rapid evolution in the human lineage relative to primates and are often located in the vicinity of genes expressed in the brain[269,271,417–421]. Therefore, HARs likely underlie a substantial share of the human-specific features.

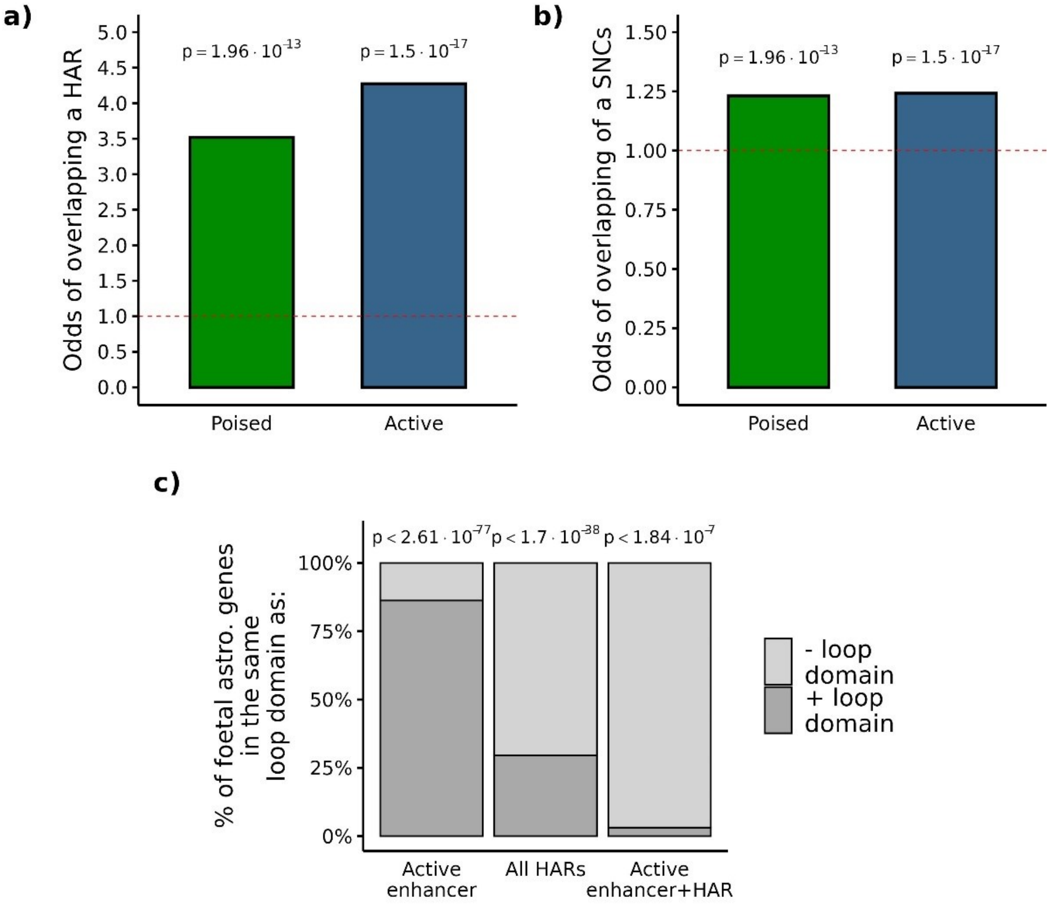


Figure 13. Astrocyte regulome bears marks of accelerated evolution. a) Enrichment of human accelerated regions (HARs) in active and poised enhancers compared to the background. b) Enrichment of single-nucleotide changes (SNCs) in active and poised astrocytic enhancers. c) Percentage of genes expressed in iAstrocytes, which are in the same loop domain as at least one active enhancer, at least one HAR, and at least one HAR overlapping an active enhancer. P-values shown for enrichment of astrocytic genes in loop domains containing a given feature, relative to all genes, calculated using Fisher's exact test.

To determine if the foetal astrocyte regulome is changed specifically in the human lineage, I assessed whether the putative distal regulatory elements mapped in the human iAstrocytes are enriched in HARs. I found a significant overrepresentation of HARs among putative poised and active enhancers in iAstrocytes (Figure 13a, for poised enhancers: 3.51-fold enrichment, $p = 1.96*10^{-13}$; for active enhancers: 4.27-fold enrichment, $p = 1.5*10^{-17}$, Fisher's exact test).

In order to test for even more recent evolutionary changes, I considered single-nucleotide changes (SNCs) that differ between humans and ancient hominids. I used the list of 321,820 human-specific sites defined by Prüfer et al., (2014): SNCs where at least one archaic human (Denisova or Neandertal) has at least one ancestral allele, and the derived allele is fixed or high-frequency in present-day humans. I mapped the original genomic coordinates, described in hg19 genome assembly to hg38 human genome assembly, which yielded 321,731 SNC sites. I observed enrichment for single-nucleotide changes (SNCs) fixed in the human lineage relative to the ancient hominids[422] (Figure 13b, for poised enhancers: 1.23-fold enrichment, $p = 1.65*10^{-16}$; for active enhancers: 1.24-fold enrichment, $p = 3.35*10^{-17}$, Fisher's exact test).

Altogether, these results show that the foetal astrocyte regulome has been a target of positive evolutionary pressure and suggest that such changes are important for the acquisition of human-specific features of the brain.

To map the genes that might be affected by HARs, I identified astrocytic genes that are in the same loop domain as HARs within active astrocytic enhancers. My analysis revealed a significant enrichment - around 3% of genes expressed in astrocytes are in the same loop domain as HARs intersecting an enhancer active in human astrocytes, which is less than expected by random (Figure 13c, $P = 1.8*10^{-7}$, Fisher's exact test). Thus, while astrocyte active enhancers are enriched in regions under evolutionary pressure, only a very specific subset of gene are potentially affect by those changes.

For instance, HARsv2_0056 at location chr1, 64125015-64125513[293], is located in the same loop domain as Receptor Tyrosine Kinase Like Orphan Receptor 1 (*ROR1*) gene; the HAR appears to form a loop with the *ROR1* promoter (Figure 14). ROR1 is expressed in astrocytes [312,423] and encodes a receptor for Wnt5a. *ROR1* has been shown to regulate the expression of genes related to fatty-acid metabolism in cultured astrocytes[424]. Notably, when overexpressed in cultured mouse astrocytes, *ROR1* was shown to induce a change in morphology, increasing the number of protrusions. The role of this enhancer in regulating the expansion of human astrocyte morphology is, however, unclear.

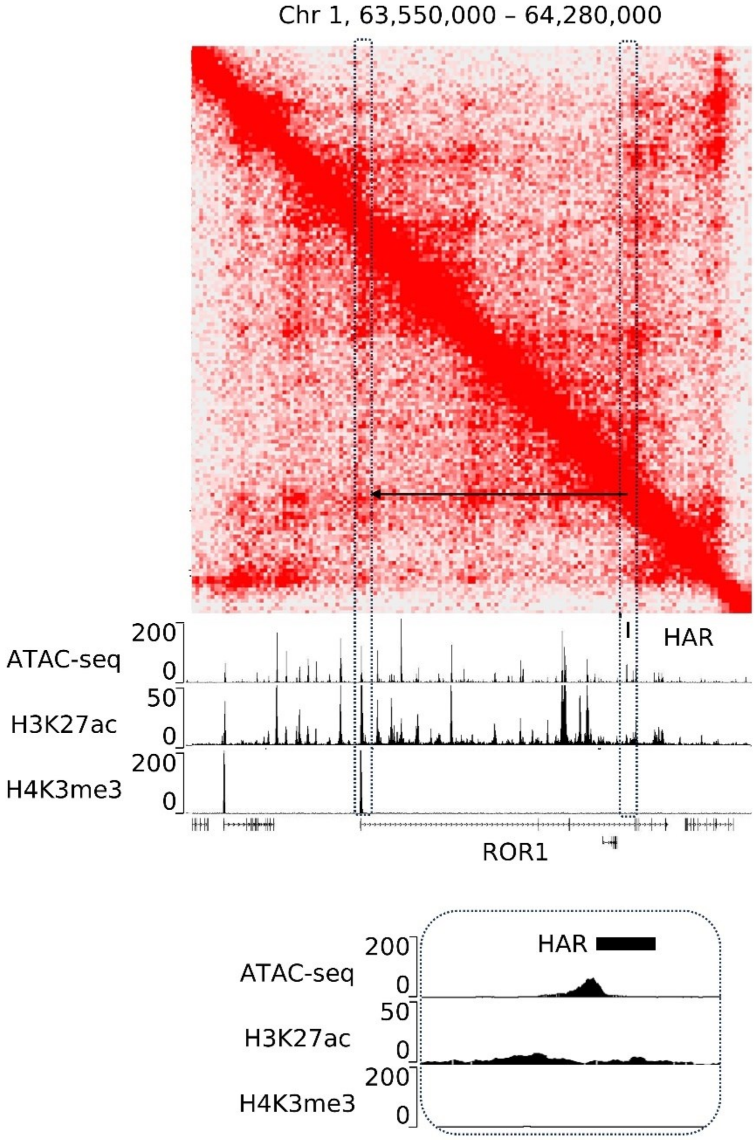


Figure 14. Genomic context of HARsv2_0056 located in proximity of *ROR1* gene (chr1, 64125015-64125513. Visualisations generated using JuiceBox and Integrated Genome Browser (IGB). Local chromatin state around the HAR is shown in the box in the lower panel.

Human-specific changes in the regulome may also act through more sophisticated regulation of gene expression, for example, by altering the timing of gene expression, or promoting the expression of a particular gene isoform. HARsv2_3052 (as defined by Girskis et al., (2021) at location chr9:116,377,285 - 116,377,731 may possibly act as an enhancer with such an effect. It appears to come into contact with an alternative promoter controlling a shorter isoform of astrotactin 2 promoter (*ASTN2*; Figure 15). *ASTN2*, together with astrotactin 1 (*ASTN1*), has a key role in glial-guided neuronal migration during brain development[425]. Interestingly, the short isoform, which appears to contact the putative active astrocytic enhancer overlapping the

HAR, is relatively more expressed in foetal than in adult brain[426]. *ASTN2*, in contrast to *ASTN1*, has a variable expression pattern in prenatal development[426]. Thus, the potential evolutionary effect of the human-specific changes in the enhancer that controls the expression of this isoform may be important at this stage of the brain development.

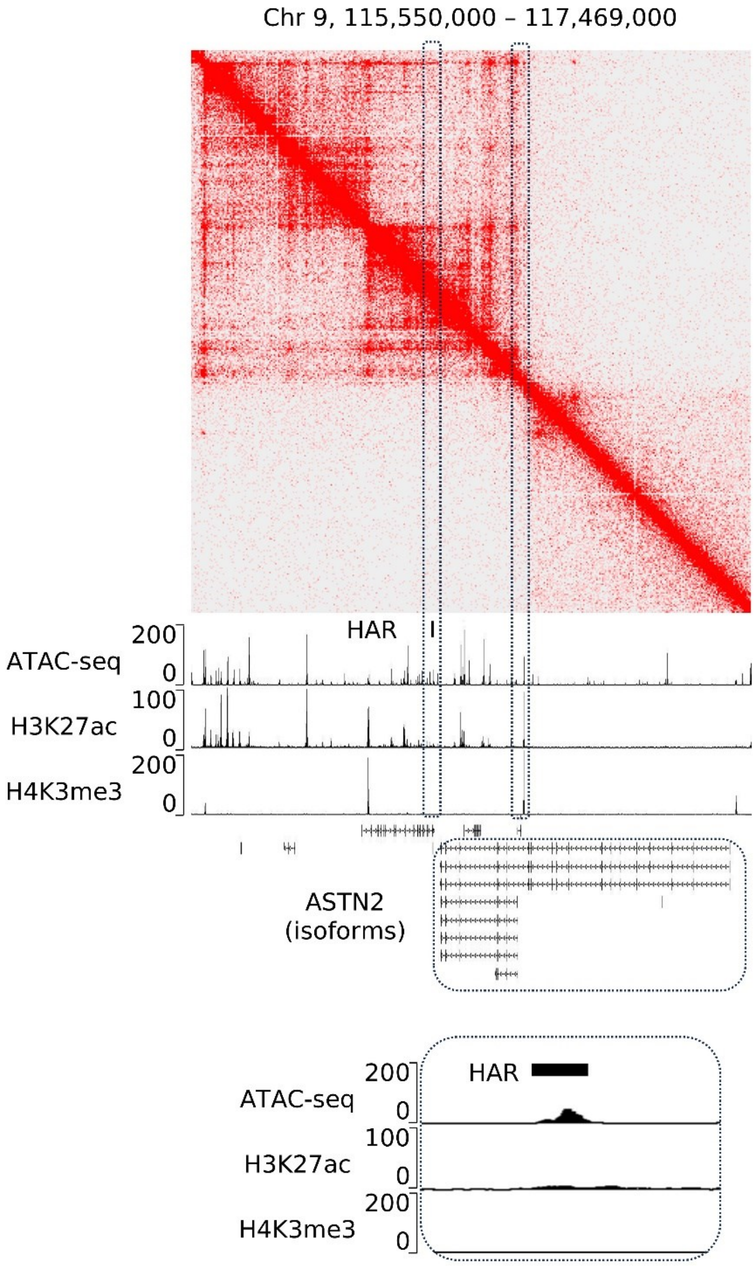


Figure 15. Genomic context of HARsv2_3052 located in proximity of *ASTN2* gene (chr9, 116,377,285 - 116,377,731) Visualisations generated using JuiceBox and Integrated Genome Browser (IGB). Local chromatin state around the HAR shown in the box in the lower panel.

To test the link between evolutionary changes and uniquely human cognitive abilities, I checked what genes are co-localised in the same loop domains as HARs that overlap active enhancers. Strikingly, the GO term analysis showed that the most enriched terms are related to brain development, cell proliferation and pattern specification ("Negative regulation of cell population proliferation", "Pattern specification process", "Central nervous system neuron differentiation", "Diencephalon development") (Figure 16). This suggests that the evolutionary change HARs which are active enhancers in astrocytes occurs at locations which can potentially influence expression genes crucial for brain development.

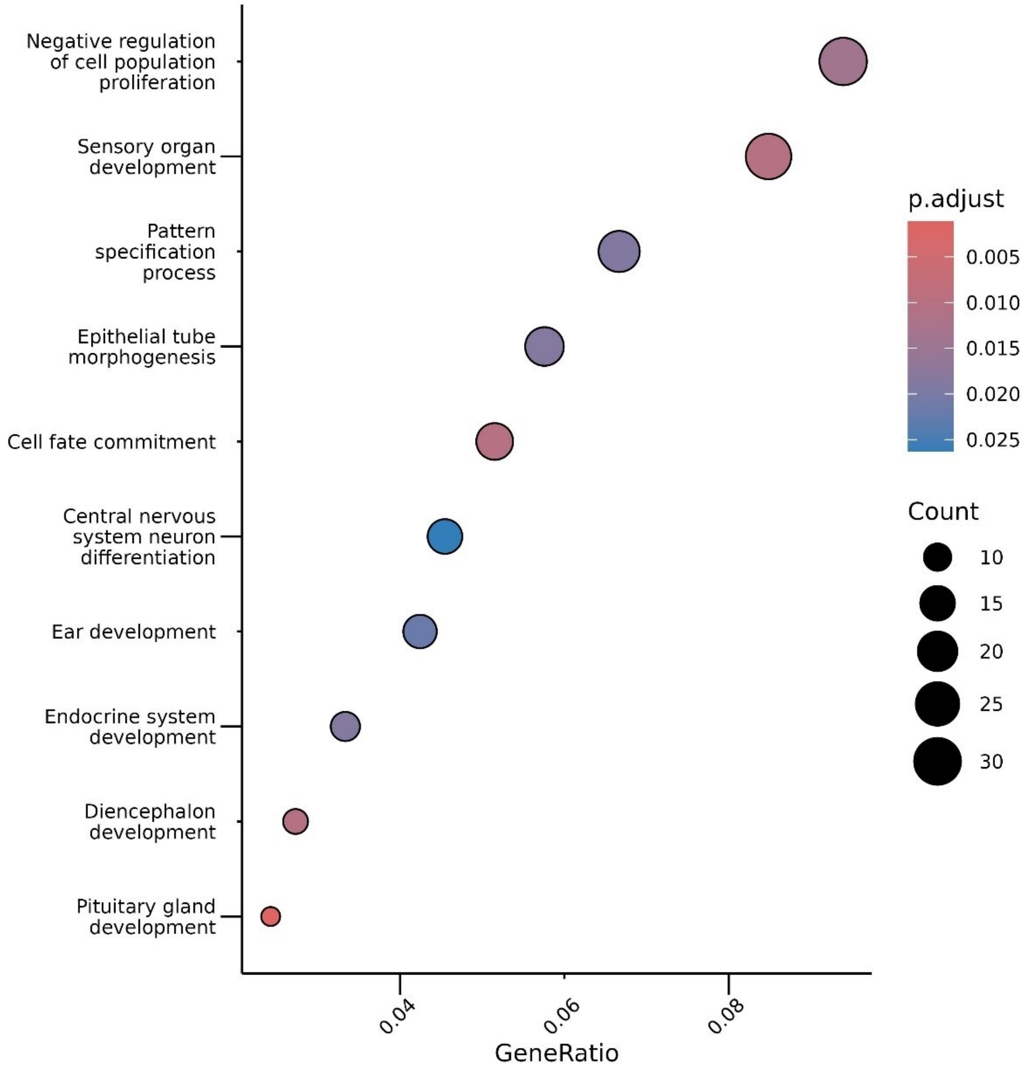


Figure 16. Enrichment in GO terms associated with genes co-localised in the same loop domain as enriched HARs overlapping active enhancers. Genes expressed in astrocytes (Zhang et al., (2016)) used as reference set. Enrichment determined using clusterProfiler [309] p 0.05, q-Value <0.05.

4.7. Evolutionary comparison of human and primate iAstrocytes

As shown above, the human-specific changes in the human astrocyte regulome have important implications for brain function. To identify enhancers that are potentially more active in human than NHP iAstrocytes, I compared regulomes from those species using our previously published data[286]. All iAstrocytes were profiled using the same genomic methods that had been applied to human cells (Figure 17a). Principal component analysis (PCA) of the open chromatin profile revealed a clear species-specific clustering of samples (Figure 17b). Notably, PCA of the regulome shows a more clearly species-specific separation than the transcriptome analysis.

I considered our published differential gene expression data, where we identified 1,730 genes significantly upregulated in human compared to chimpanzee, and 2,788 genes significantly upregulated in human compared to rhesus macaque (p-adj. < 0.1, Figure 17d-f). Merging the two lists, I found 677 up- and 486 down-regulated genes in human as compared to both chimpanzee and rhesus macaque (p-adj. < 0.01).

To identify enhancers driving this species-specific difference, I identified ATAC-seq peaks in human, chimpanzee and macaque iAstrocytes that were amenable to a liftOver between species (i.e., displaying at least 50% homology in each pairwise comparison) and compared the ATAC-seq signal at these elements in human, chimpanzee, and rhesus macaque iAstrocytes. I selected differentially open regions (DOR) using DESeq2 analysis[319] of ATAC-seq signals and identified 13,666 DORs that were differentially open in humans relative to the two NHPs (p-adj. for single comparisons < 0.1, p-adj. for double comparisons < 0.01, Figure 18a-c).

In order to mitigate the sequence homology cut-off that I applied during signal comparison, I also included 777 human ATAC-seq peaks annotated to the hg38 assembly that displayed low homology in chimpanzee and rhesus macaque and could not be properly compared using liftOver. I used ChIP-seq data for H3K4me3, H3K27ac, and TSS annotation to exclude regions potentially active in other species, and I compiled a list of 11,711 human-specific iAstrocyte enhancers, including 11,065 DORs and 646 regions with low homology between primates.

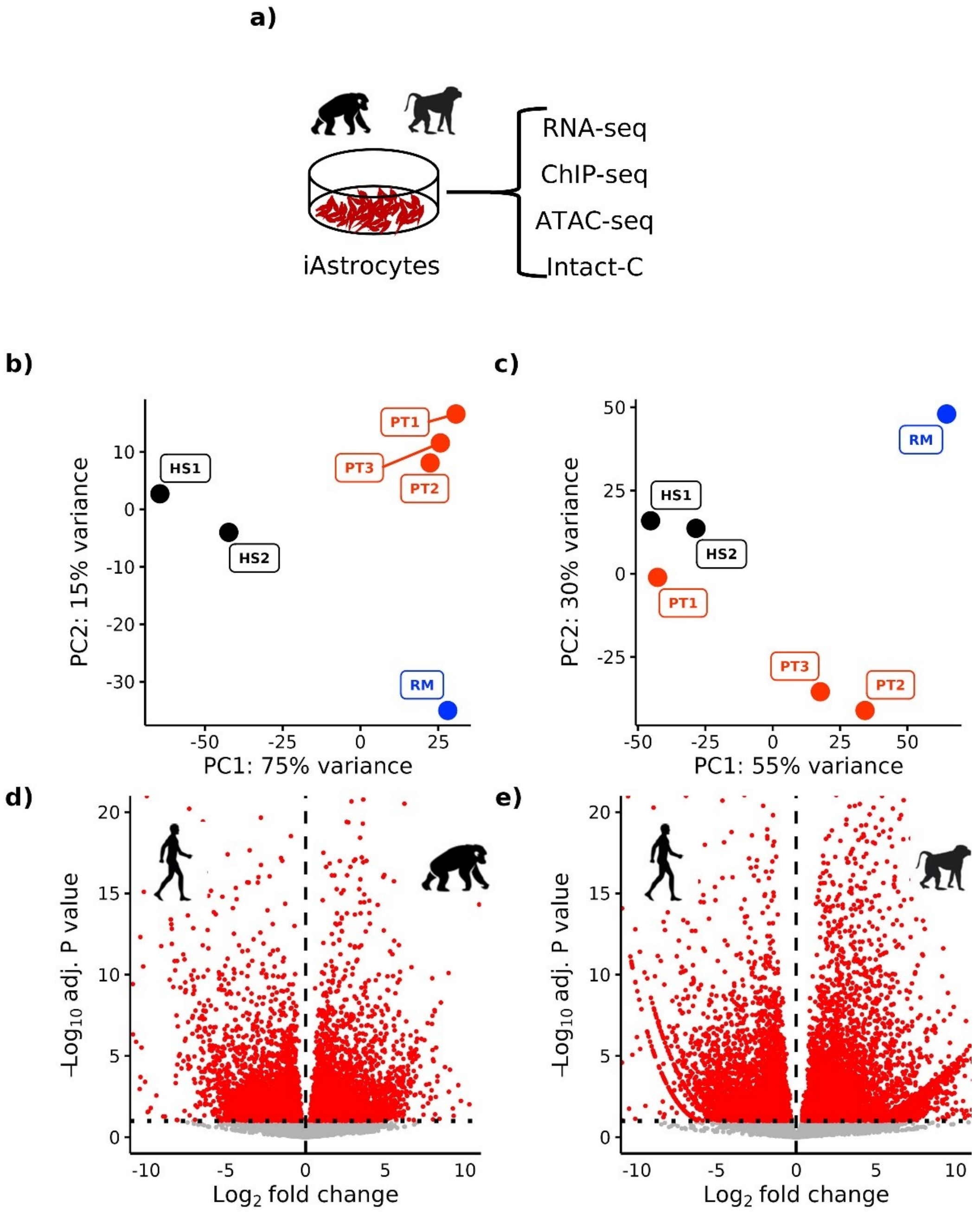
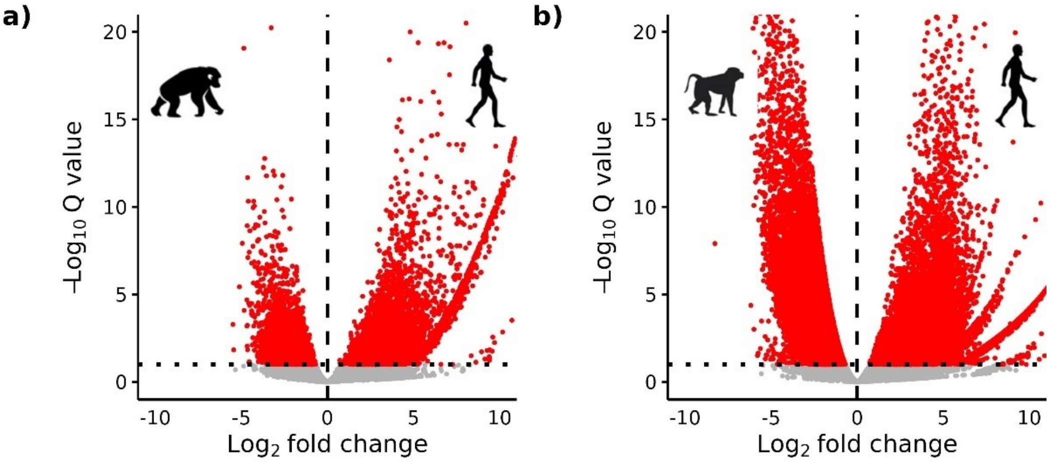


Figure 17. Comparison of regulomes and transcriptomes of HS, PT and RM. a) PT and RM iAstrocyte model was used to profile gene expression (RNA-seq), histone modifications (ChIP-seq), regions of open chromatin (ATAC-seq), and chromatin conformation (Intact-C). b) PCA analysis of ATAC-seq data from HS, PT and RM. c) PCA analysis of RNA-seq data from HS, PT and RM. d) Volcano plot of differentially expressed genes between HS and PT. e) Volcano plot of differentially expressed genes between HS and PT.



c)

Category	Number
DORs	13666
Non-homologous enhancers	777
HS-specific enhancers	11711
Enhancers linked to DEGs	3026
Enhancers not linked to DEGs	1231

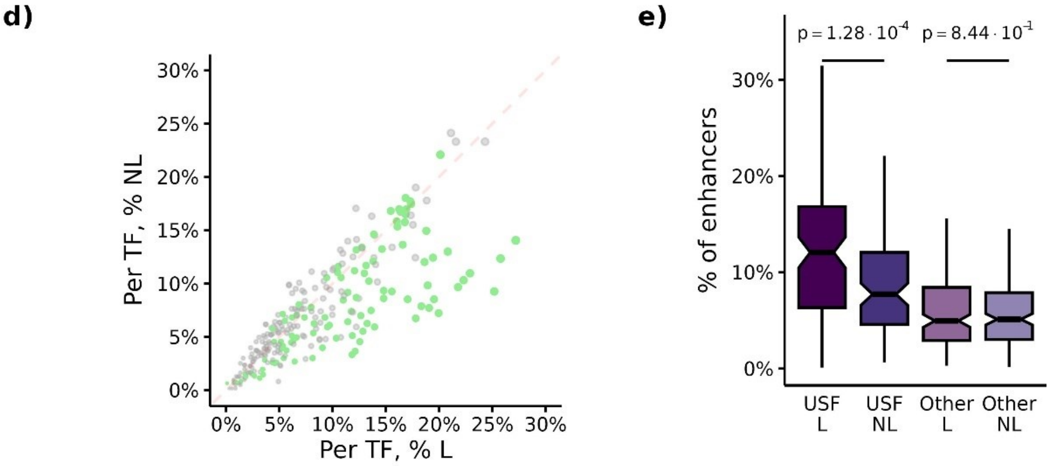


Figure 18. Selection of HS-specific enhancers affecting the expression of differentially expressed genes (DEGs). a) Volcano plot of "differential openness" (level of ATAC-seq), for comparisons between HS and PT, and HS and RM based on data aligned to the consensus genome. Red denotes a differentially open region (DOR). Final DORs were defined as regions with a congruent and significant ATAC-seq signal change in the two comparisons. b) Volcano plot of "differential openness" (level of ATAC-seq), for HS and RM based on data aligned to the consensus genome. c) Breakdown of DORs to distinct groups. d) Fraction of linked and non-linked enhancers that contain a given TFBS. Each dot represents a unique TFBS. The size of each dot corresponds to the number of linked enhancers that contain it. Green points correspond to USFs. f) Enrichment of USF TFBS in linked and non-linked enhancers, compared to other TFBS.

Next, to investigate how the human-specific changes in the astrocyte regulome relate to changes in gene expression, I identified human-specific enhancers located either proximally (< ±1,000,000 bp from the TSS) or distally (> ±1,000,000 bp from the TSS) to genes upregulated in human iAstrocytes. I detected 3,026 proximal human-specific enhancers (“linked” enhancers) and 1,231 distal human-specific enhancers (“non-linked” enhancers”) (Figure 18c).

The degree of “openness” of enhancer regions was recently shown to be modulated by the Universal Stripe Factor family of TFs [1]. This group comprises 288 factors and was identified by the laboratory of Rafael Casellas at the National Institutes of Health by studying TF co-occupancy at regions of open chromatin across different tissues, in both human and mouse. USFs impart accessibility to regulatory DNA and allow for the action of other, tissue-specific TFs. Therefore, I investigated whether USFs may be more frequently present in “linked” than “non-linked” enhancers. I calculated the fraction of linked and non-linked enhancers that contained predicted binding sites of USFs (Figure 18d). I found that linked enhancers contain binding sites of any given USF significantly more often than non-linked enhancers ($p = 1.28 * 10^{-4}$, one-sided Wilcoxon Rank Sum Test. No such difference is observable for other TFs (Figure 18e). This suggests that USFs may be especially important for human-specific enhancers proximal to genes overexpressed in human compared to NHP foetal astrocytes.

In a recent publication characterising the molecular signature of primate iAstrocytes[286], we have described this trend, showing that a gain for USF binding sites in enhancers linked to genes that are differentially expressed between human and NHP iAstrocytes. Furthermore, in that publication, we employed a convolutional neural network (CNN) that we trained on glioma and our iAstrocyte ATAC-seq datasets to predict whether evolutionary alterations in USFs binding sites located in enhancers result in a change in regulatory activity.

All these results point to an essential role of USF in the evolution of astrocytic enhancers and gene expression programs. Yet, the formal proof that the gain of USF binding sites was missing. I sought to address this issue using high-throughput reporter assays (see next).

4.8. Massively Parallel Reporter Assay - MPRA

Our work described in Ciuba et al., (2025) provides the initial evidence on the potential functionality of enhancers enriched in USFs for human-specific gene activity in iAstrocytes. Yet, the described machine-learning predictions required *in vivo* testing and validation. Therefore, I was interested to explore which TFBS are present in the most active enhancers in foetal astrocytes.

I also aimed to explore the manner in which SNPs affect the activity of enhancers, and thus, how changes in astrocyte regulome can exert an effect on human brain-related phenotypes, including disease states. Finally, I wanted to obtain a measure of enhancer activity in order to position the findings in a wider context of genomic organisation and see how enhancer activity relates to the broader chromatin architecture.

In order to obtain a measure of true *in vivo* activity of enhancers detected using genomic assays, I applied the Massively Parallel Reporter Assay (MPRA) technique, which allows for examination of the activity of enhancers in a high-throughput manner. I employed a modified protocol for episomal MPRA, described by the Sabeti lab[331] (Figure 20). This approach is based on a modified pGL4.23 plasmid backbone, has been shown to produce results consistent with other MPRA variants, including those relying on candidate fragments randomly integrated into the genome via lentiviral transduction[427,428].

For each enhancer, I designed a sequence that covered the central 200 basepairs centred around the ATAC-seq peak, as determined using MACS2 (see Materials and Methods). Each sequence was flanked by 15-basepair binding sites of primers, which were used to amplify the sequence.

I designed two libraries of regions that I wished to assay for enhancer activity (called candidate enhancer regions, CREs, later in the text). The library #1 comprised the following CREs (Figure 19a):

- 3026 “linked” enhancers, which are human-specific enhancers located within 1 Mb from promoters of upregulated DEGs expressed in iAstrocytes (Figure 18),
- 1231 “non-linked” enhancers, which do not have any promoters of DEGs (up/downregulated) within 1 Mb from promoters of DEGs, as determined by comparing both human against chimpanzee and rhesus macaque, both together and separately,
- 511 randomly chosen chimpanzee variants of “linked” enhancers,
- 301 enhancers surrounding genes crucial for astrocyte biology - GJA1, encoding connexin43, a crucial component of hemichannels which are essential for

communication between astrocytes through gap junctions[429]; *GFAP*, *SLC1A3* encoding EAAT1 and *VIM* encoding vimentin, which are astrocyte cell markers, thrombospondins 1, 2 and 3, which are astrocyte-secreted proteins crucial for functioning of the central nervous system[34], *JUN* and *FOS*, which are components of AP-1 general transcription factor, and was shown to be crucial for glial function[430,431] and memory formation[225,226] and *AQP4* encoding a crucial water channel located at astrocytic interface with brain-blood barrier[21,29],

- 511 Ensembl[432] SNP variants for 197 individual SNPs related to phenotypes describing measures of brain morphology (55 SNPs, 145 variants), intellectual ability (75 SNPs, 182 variants) and neuropsychiatric diseases (67 SNPs, 184 variants), based on EBI GWAS Catalogue (Sollis et al., 2023, acesssed on 21.12.2023), that overlap active enhancers that are also open in foetal astrocytes based.

As stated before, all CREs were centred around ATAC-seq peaks, except for SNP CREs, which were centred around the polymorphic site.

The library #2 comprised the following CREs (Figure 19b):

- 2387 ATAC-seq peaks selected based on Intact Hi-C loops surrounding the following genes: *VIM*, *GJA1*, *SLC1A3*, crucial elements of extracellular matrix produced by astrocytes: Ephrins: *EFGB2*, *EFNA1*, *EFNA4*, *EFNA5*[433] Glypicans: *GPC1*, *GPC3*, *GPC4*[32,33,434], *SEMA6A*, *THBS3*; *CTCF*, a protein crucial for genome organisation[307,435], MTSS2, which is more expressed in human than chimpanzee iAstrocytes and is involved in regulation of neural progenitor growth[436], *POU3F2*, which is a TF and a regulator of neurogenesis with direct effects on behaviours[392,437]; *CLU*, encoding clusterin, a secreted protein which affects glial physiology and alleviate some of AD symptoms[438–440]; *SHROOM3*, a gene regulating neural tube development[441], *AUTS2* which is a gene strongly associated with ASD, and bears strong marks of recent evolutionary change[442,443]; APOE, a gene encoding apolipoprotein E that is crucial in aetiology of AD[444,445]; *TGFB2* which encodes for TGF-β2, one of isoforms of transforming growth factor beta (TGF-β) family of regulators crucial for synaptogenesis [39] and plays a role in synaptogenesis[446–449]; *TCF7L2* which is a key transcriptional effector of the Wnt/β-catenin pathway, which affects behaviour[450]; *IGFBP7*, a regulator Wnt/β-catenin pathway correlated with astrocyte-derived tumours[451,452], *GLIS3*, a gene encoding a TF expressed in astrocytes[453], which correlates with Tau protein levels in AD[454]; *NR1H3*, *IGFBP2*, *LGALS1*, *RHOBTB3* and PALLD, which we identified as upregulated in human iAstrocytes, as compared to NHP; TEAD3 which encodes a transcription factor that we identified as regulator of evolutionary change of astrocyte size and shape[286];

- 789 chimpanzee sequences that are homologous to ATAC-seq peaks surrounding RHOBTB3, APOE, VIM, LGALS1, AUTS2, EFNA1, EFNA4, THBS3, IGFBP7, TEAD3 and CTCF
- 60 RM sequences that are homologous to ATAC-seq peaks surrounding APOE,
- 252 HAR regions[293], which completely overlapped ATAC-seq peaks in human iAstrocytes,
- 252 HAR sequences from chimpanzee, homologous to human HARs, selected for the library,
- 292 Ensembl[432] SNP variants for 108 SNPs that overlapped ATAC-seq peaks in proximity of selected genes, all phenotypes were included.
- 1538 fragments, representing a subset of 769 linked enhancers which differed from their homologous chimpanzee counterparts by binding sites of USF only, their variants modified with the chimpanzee version of USF binding sites (see section 4.8.4 for details of this approach).

Both libraries used identical negative and positive controls. For positive controls, I employed two groups of elements:

1. 58 regions previously validated for enhancer activity in the mouse forebrain (Vista Enhancer Browser[455]), and that overlap ATAC-seq peaks in foetal human astrocytes,
2. 144 regions showing the strongest ATAC-seq signal, from regions that:
    a. display the chromatin signature of active enhancers in our iAstrocytes;
    b. overlap with ATAC-seq peak in *ex vivo* cultured human foetal astrocytes;
    c. overlap with enhancers active in astrocytes as defined by ENCODE datasets[456–458] generated in the laboratory of Zhiping Weng (accession ENCSR021PEA, file: ENCFF878BLR);
    d. intersect active enhances defined by the 18-state ChromHMM model (Kellis lab, accession ENCSR421IQD, file: ENCFF256DNQ, terms: EnhA1, EnhA2).
3. Additionally, 200 bp fragments of the SV40 and CMV enhancers were included as positive controls (see Materials and Methods for sequences).

a)

Category	Number of CREs
Linked enhancers	3026
Non-linked enhancers	1231
PT variants of linked enhancers	511
Enhancers in proxomity of chosen genes	301
SNPs (individual variants)	511
Positive controls	204
Negative controls	199

b)

Category	Number of CREs
Enhancers in proximity of chosen genes	2387
PT enhancers in proximity of chosen genes	789
PT enhancers in proximity of APOE	60
HARs	252
HARs, PT variant	252
SNPs (individual variants)	292
Linked enhancers, HS variant	769
Linked enhancers, ΔUSF	769
Positive controls	204
Negative controls	199

Figure 19. Summary of compositions of libraries tested using MPRA. a) Composition of library #1. b) Composition of library #2.

The negative controls (N=199 regions) were chosen at random from genomic regions that did not overlap any marks of functional genomic regions available in the laboratory. This excluded ATAC-seq peaks, potential promoters, peaks of ChIP-seq for H3K27ac, H3K4me3, and CTCF, in both human and NHP (based on annotation to a consensus genome (see Materials and Methods for details). Furthermore, I excluded single tandem repeats, including imperfect repeats as annotated by the Tandem Repeats Finder (TRF) program[459] (retrieved from the *simpleRepeat* track in the UCSC Genome Browser[460]).

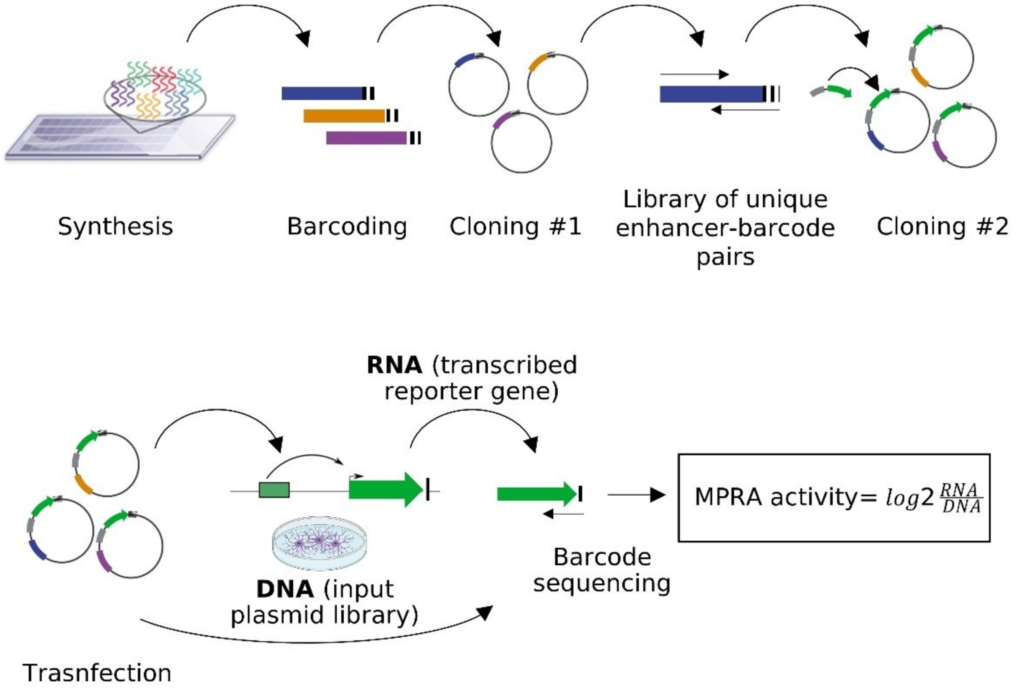
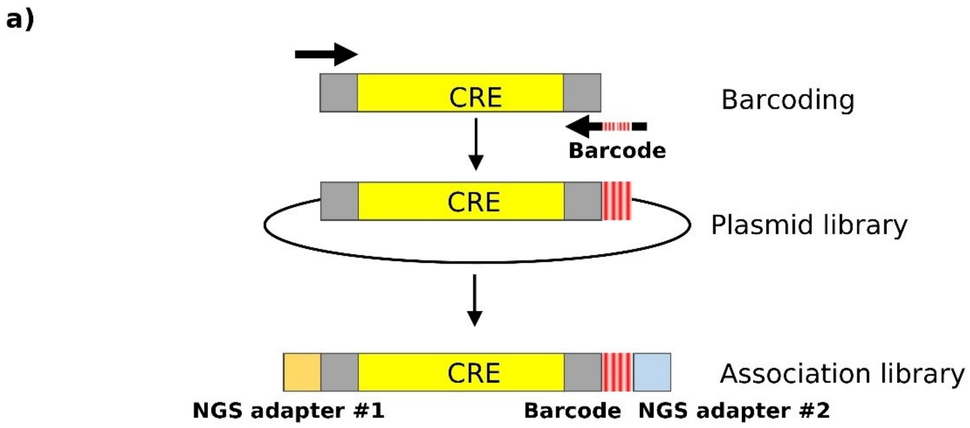


Figure 20. Graphical overview of the Massively Parallel Reporter Assay (MPRA) technique used. The experimental set-up was based on the protocol described by Tewhey et. al (2016). The workflow involved barcoding of a commercially-synthesized DNA library, two cloning steps, and two next-generation sequencing steps. First, cloning inserted the barcoded candidate enhancer sequence into the plasmid, and the second cloning inserted a minimal promoter-GFP construct between CRE and a barcode. The first sequencing step, where the plasmid DNA was sequenced produced a map of CRE-barcode associations. In the second experiment, I sequenced the GFP RNAs, which provided a readout of relative abundances of barcodes related to each enhancer. Comparison of abundances of barcodes in the transcribed reporter to the abundance of each barcode in the plasmid DNA library gave me a measure of activity of each CRE ("MPRA activity").

In MPRA, in brief, each CRE in the synthetic DNA library is linked to a unique molecular barcode and inserted into a reporter construct, which is then introduced into cells. Regulatory activity is quantified by sequencing barcodes in RNA, reflecting the transcriptional output driven by each CRE. Comparing RNA to DNA barcode counts provides a quantitative measure of enhancer activity across thousands of sequences in parallel (Figure 20)



b)

	Library #1	**Library #2**
Total CREs	5987	5950
CREs cloned	98.03 %	99.61 %
CREs not cloned	35 (0.58 %)	5 (0.08 %)
CREs barcodes<50	83 (1.39 %)	18 (0.3 %)
Average number of BCs	3607.89	31175.22

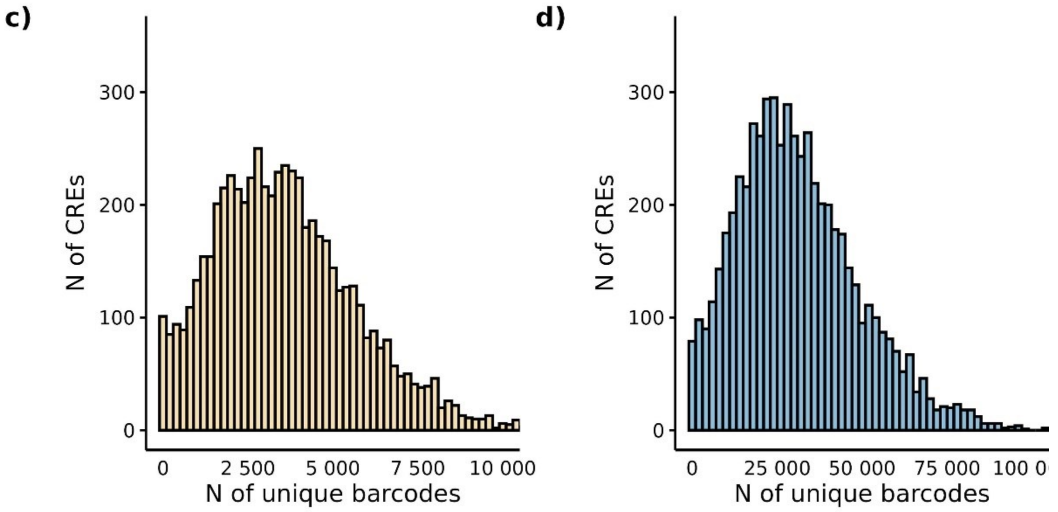


Figure 21. First step of preparation of the MPRA library. a) Schematic representation of the barcoding and cloning process. Each CRE was designed as a 230 basepair oligonucleotide, with 15bp overhangs for amplification. Each CRE is amplified using a PCR. The 3' primer contains a 19-basepair sequence of random bases, which yields multiple unique barcodes per CRE. The amplified library was then cloned using Gibson assembly into a plasmid backbone. b) Quantification of the first step of library preparation for libraries 1 and 2. Variability in the number of barcodes is in part due to differing sequencing depth. c) Distribution of the number of barcodes per CRE for libraries 1 and 2.

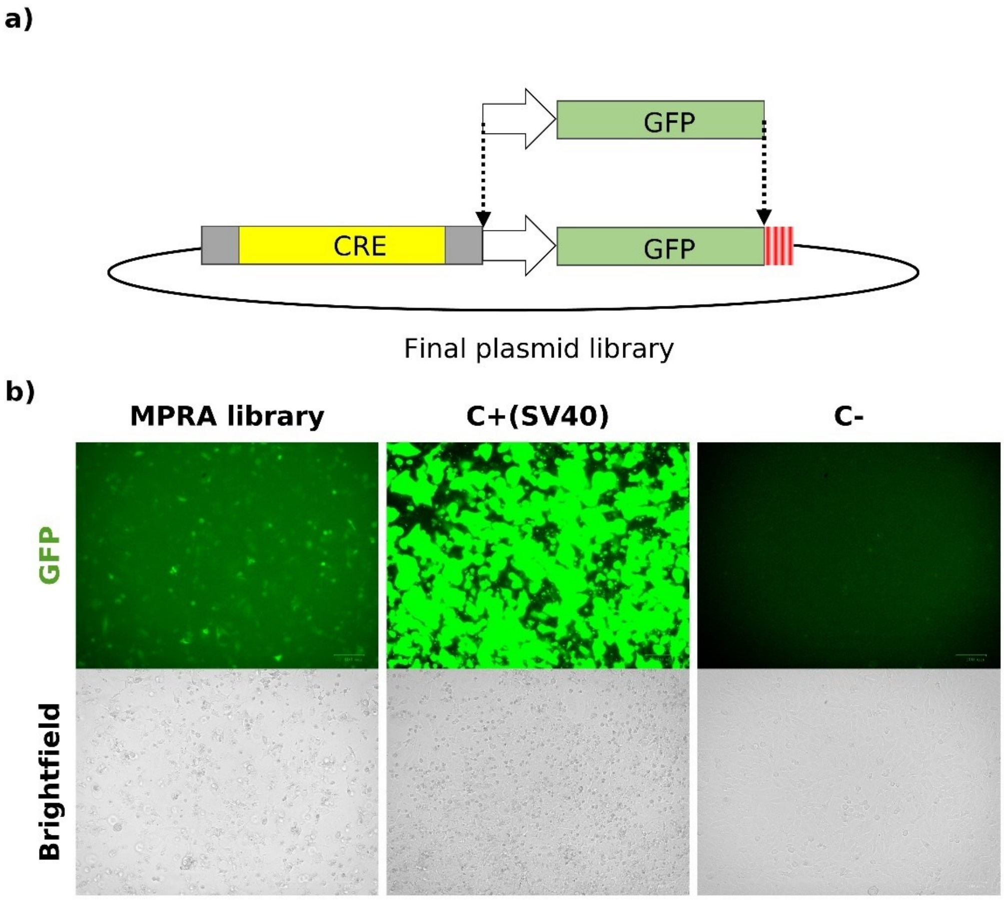


Figure 22. First step of the preparation of the MPRA library. a) Schematic representation of the second cloning step. Minimal promoter and GFP are cloned between CRE and barcode. b) Cultures of cells transfected with MPRA library (first panel), positive control (MPRA construct with a full-length SV40 enhancer, second panel), and non-transfected cells (third panel).

More specifically, for each CRE tested in MPRA, a 200 bp DNA sequence was chemically synthesised, and subsequently barcoded using a primer containing degenerate bases (Figure 20). This treatment added 19 random nucleotides to each construct. The resulting library was cloned into a destination plasmid via Gibson assembly. To establish an association between each tested fragment and unique barcodes generated through PCR, I prepared and sequenced an NGS library of the plasmid pool at this stage, which allowed me to read each enhancer together with its barcode. This readout formed an “association library” (Figure 21).

In the subsequent step, I introduced a minimal promoter-GFP cassette into the plasmid pool, cloning it between the enhancer and a barcode. This step generated the final plasmid, which was composed of CRE, followed by a minimal promoter and GFP with barcodes encoded in the 3’ untranslated regions (UTR) of the GFP mRNA (Figures 21 and 22a)

The assay was performed in a tractable glioblastoma line, LN-229 (ATCC catalogue number CRL-2611). This cell line, established from a female patient with right frontal parieto-occipital glioblastoma, was chosen because it is sex-matched to iAstrocyte cell lines used in this study, which are all female. Likewise, this line was highly transfectable, achieving 92% transfection efficiency in control experiments, as determined by flow cytometry. This was necessary to conduct the MPRA experiment.

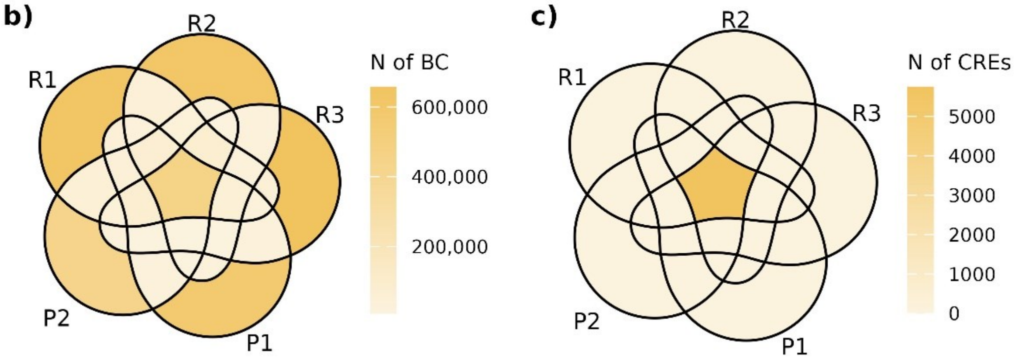
The pool of plasmids was transfected into the cell culture (Figure 22b), and after 48 hours from transfection, cells were collected and frozen at -80°C. Following RNA isolation and reverse transcription, I used the 3’ end of GFP mRNA to prepare an NGS library, which reflects the abundance of each barcode in the transcript pool (Figure 23). To normalise the readout, I prepared similar libraries from the input plasmid libraries used.

Quality checks are crucial in MPRA experiments. Each CRE must have a sufficient number of unique barcodes in the association library to avoid bias from overrepresented plasmids. Published reports recommend at least 20 barcodes per CRE[331]. In my libraries, I scored on average over 100 (experiment #1) and more than 70 (experiment #2, Figure 23a and c) unique barcodes per CRE, indicating highly complex and excellent quality preparations.

In both MPRA experiments, each replicate achieved >97% CRE coverage (Figure 23a, c). Furthermore, a large share of individual enhancers was specific to individual libraries, which further showed that the abundance of individual CREs is not due to the skewed distribution of barcoded plasmids during cloning (Figure 23b and e). These metrics confirm that the MPRA experiments provide a robust assessment of regulatory element activity, allowing me to proceed with testing enhancer activity in iAstrocyte regulomes.

a)

Sample	N CREs(fraction)	Avg. n BCs per CRE
R1	5839 (98 %)	132.76
R2	5813 (97 %)	103.85
R3	5808 (97 %)	105.74
P1	5797 (97 %)	105.15
P1	5828 (97 %)	116.02



d)

Sample	N CREs (fraction)	Avg. n BCs per CRE
R1	5793 (97 %)	77.94
R2	5789 (97 %)	77.51
R3	5775 (97 %)	77.51
P1	5805 (98 %)	88.43
P2	5813 (98 %)	91.28

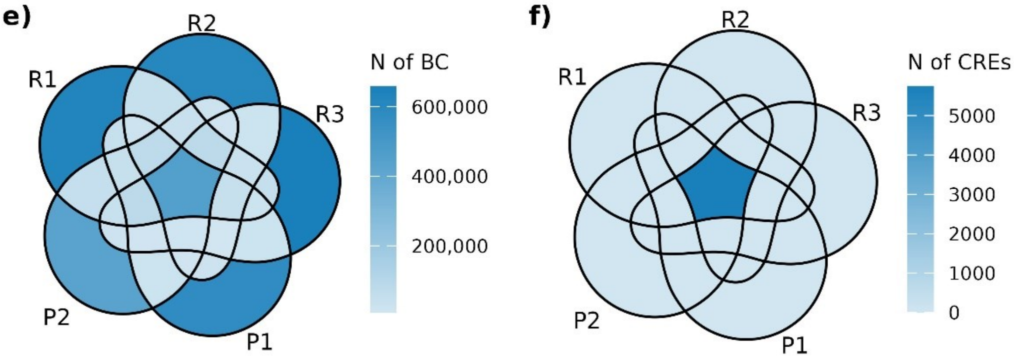


Figure 23. Quality assessment of experiments using MPRA library #1. a) Numbers and percentages of CREs, and the number of individual barcodes (BCs) read per each CRE in three experimental replicates (R1, R2, and R3) and input plasmid controls (P1, P2) of MPRA. b). Overlap between the individual BCs in the prepared libraries. c) Overlap between CREs present in each library. **MPRA experiment #2**: d) Numbers and percentages of CREs, and number of individual barcodes (BCs) read per each CRE in three experimental replicates (R1, R2 and R3) and input plasmid controls (P1, P2). e) Overlap between the individual BCs, shared between libraries and specific to them. f) Overlap between CREs present in each library.

4.8.1. Performance of MPRA

By integrating the sequencing data with the association library, I linked barcode abundances to specific enhancer sequences (Figure 23). Normalizing barcode counts to the input plasmid library allowed me to estimate the activity of each enhancer, which I refer to as **MPRA activity** throughout this manuscript.

I processed the MPRA data as described in the *Materials and Methods* section and analysed it using the DESeq2 package. Altogether, I estimated activity for 5,726 of the 5,987 (95.6%) CREs in the first library and 5,731 of the 5,950 (96.3%) CREs in the second library.

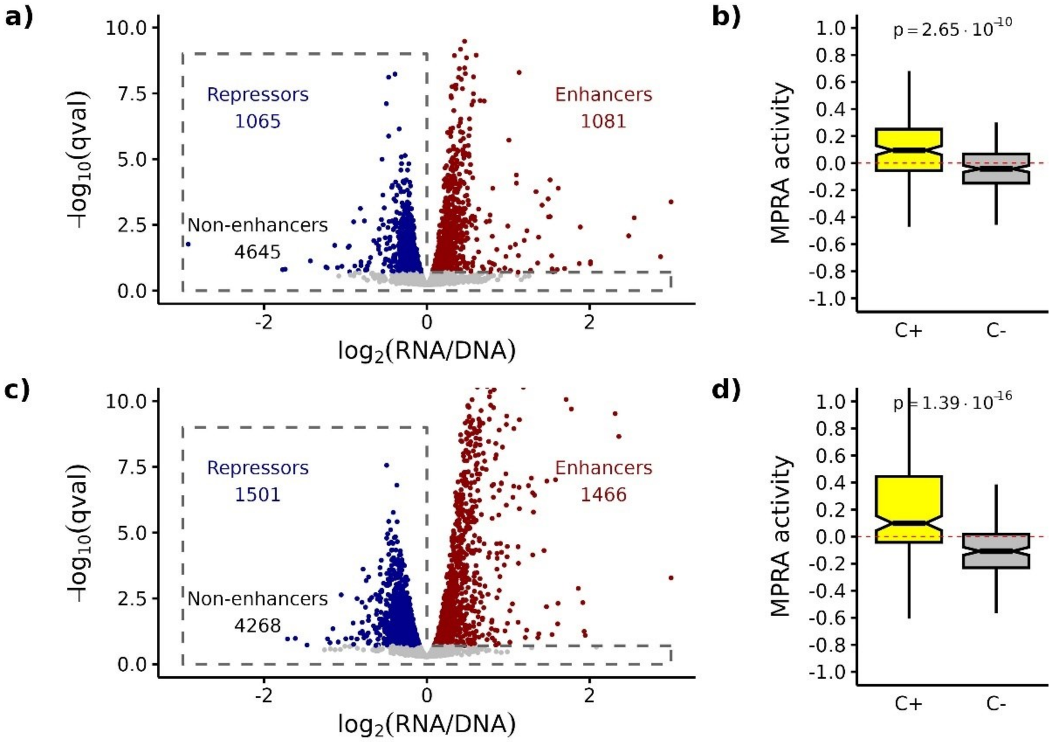


Figure 24. Summary of two MRPA screens performed. Volcano plots and comparisons of activities of positive and negative controls for screen number 1 (a) and screen number 2 (b).

I classified CREs with MPRA activity > 0 and Q-value < 0.2 as active enhancers ("**active MPRA enhancers**") and sequences with MPRA activity < 0 and Q-value < 0.2 as repressors ("**MPRA repressors**"). All CREs that are not active MPRA enhancers (the group also includes repressors) will be called "**non-enhancers"**. Using these criteria, I identified 1,081 active MPRA enhancers, 1,065 MPRA repressors, and 4,645 non-enhancers in the first screen, and 1,500 active MPRA enhancers, 1,466 MPRA repressors, and 4,267 non-enhancers in the second screen (Figure 24). This represents a high rate of enhancer discovery compared to previous studies.

4.8.2. General determinants of enhancer activity

As a first step of the analysis, I surveyed how the genomic assays and sequence determinants correlate with MPRA outputs.

Both levels of ATAC-seq and ChIP H3K27ac signal at the MPRA active enhancers are higher than non-enhancers (Figure 25a and b, p for ATAC-seq $=6,8*10^{-51}$, one-sided Wilcoxon Rank Sum Test, p for H3K27ac ChIP-seq $<2.82*10^{-32}$, two-sided Wilcoxon Rank Sum Test). However, the relationship between the strength of ATAC-seq and H3K27ac ChIP-seq signal and MPRA activity is not linear. The increase in any of those measures does not result in a proportional increase in MPRA activity (Figure 25c and e). The odds of a CRE being an active enhancer, increase with both ATAC-seq and H3K27ac signal only to a point (Figures 25d, f). This highlights the usefulness of experimental validation of predictions based on those assays.

I also analysed sequence features - CG/AT dinucleotide content and G4 quadruplexes [461] correlate with enhancer activity, as determined by MPRA. I found that active MPRA enhancers are decisively enriched in CG dinucleotides (Figure 26a, p = $4,74*10^{-83}$, two-sided Wilcoxon Rank Sum Test), while inactive sequences more commonly have a greater number of AT dinucleotides (p = $2.33*10^{-72}$, two-sided Wilcoxon Rank Sum Test). Greater CG content correlates with higher MPRA activity and enhancer probability, (Figure 26c and e) though not linearly. Conversely, higher AT content reduces MPRA activity and enhancer probability (Figure 26e and f).

G4 quadruplexes are more frequent in active MPRA enhancers (Figure 27a and b). While for both MPRA enhancers and CREs, which are not active MPRA enhancers, the vast majority of CREs do not have any G4 quadruplexes (more than 77% percent of non-enhancers, and more than 60% of MPRA active enhancers). Meanwhile CREs with a number of G4 quadruplexes above 4 have especially high MPRA activity and elevated odds of being an enhancer (Figure 27c and d).

In conclusion, while these correlations that I describe are not completely novel, they provide strong evidence of the successful MPRA experiment and useful background for TFBS-focused analyses.

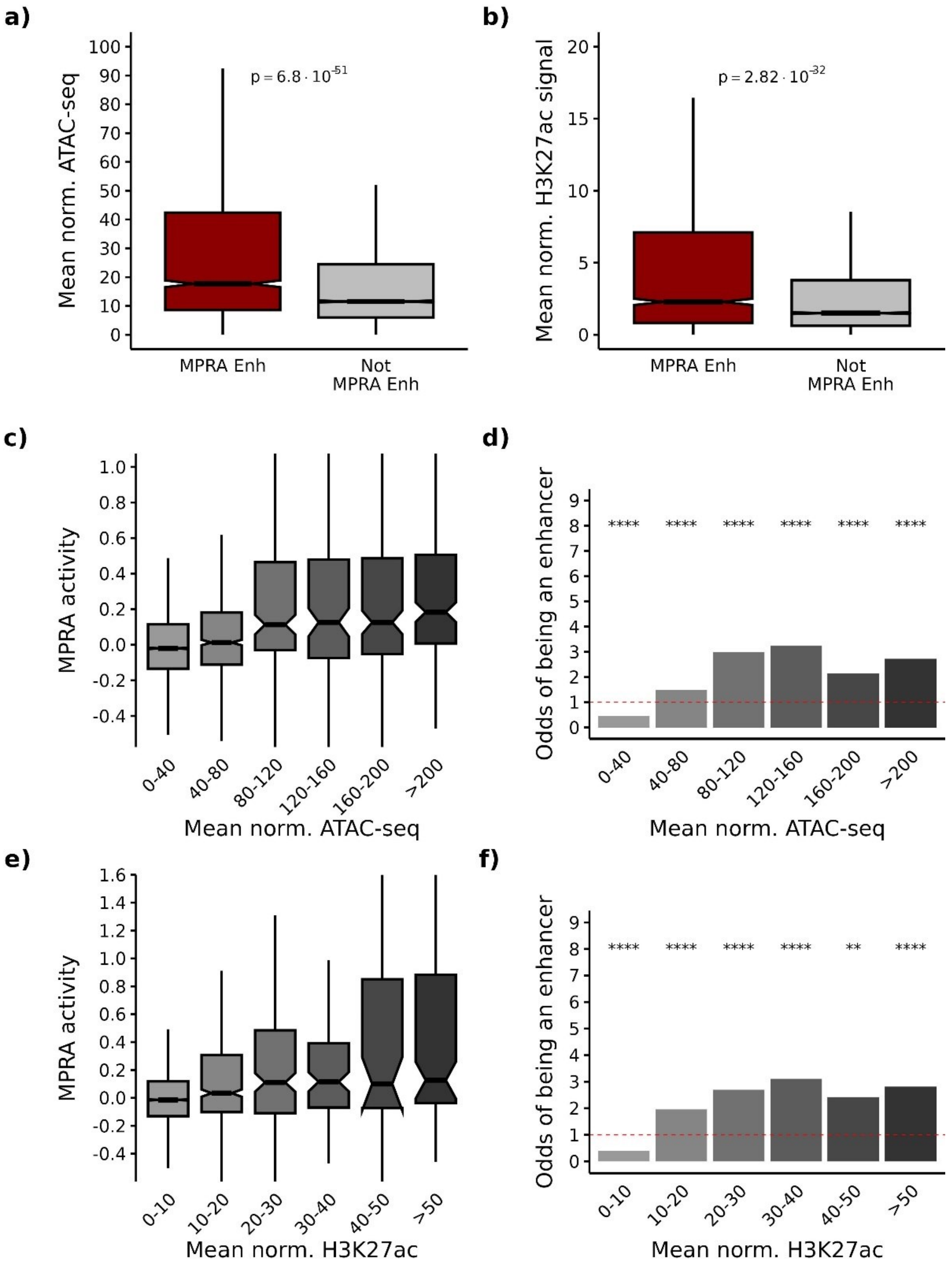


Figure 25. Correlation of enhancer activity with ATAC-seq and ChIP for H3K27ac signals. a) RPGC-normalised ATAC-seq signal in active MPRA enhancers versus CREs that are not active MPRA enhancers. b) RPGC-normalised H3K27ac ChIP-seq signal in active MPRA enhancers versus CREs that are not active MPRA enhancers. c) MPRA activity of CREs as a function of the average RPGC-normalised ATAC-seq signal at enhancer locations. d) Odds ratio of a CRE being an enhancer as a function of the average RPGC-normalised ATAC-seq signal. e) MPRA activity of CREs as a function of the average RPGC-normalised H3K27ac ChIP-seq signal at enhancer locations. f) Odds ratio of a CRE being an enhancer as a function of the average RPGC-normalised H3K27ac ChIP-seq signal. For clarity of display, p-values in panels d and f were converted to significance symbols using the following scheme: $p < 0.05$: "*", $p < 0.01$: "**", $p < 0.001$: "***", and $p < 0.0001$ "****"; non-significant values are labelled "ns."

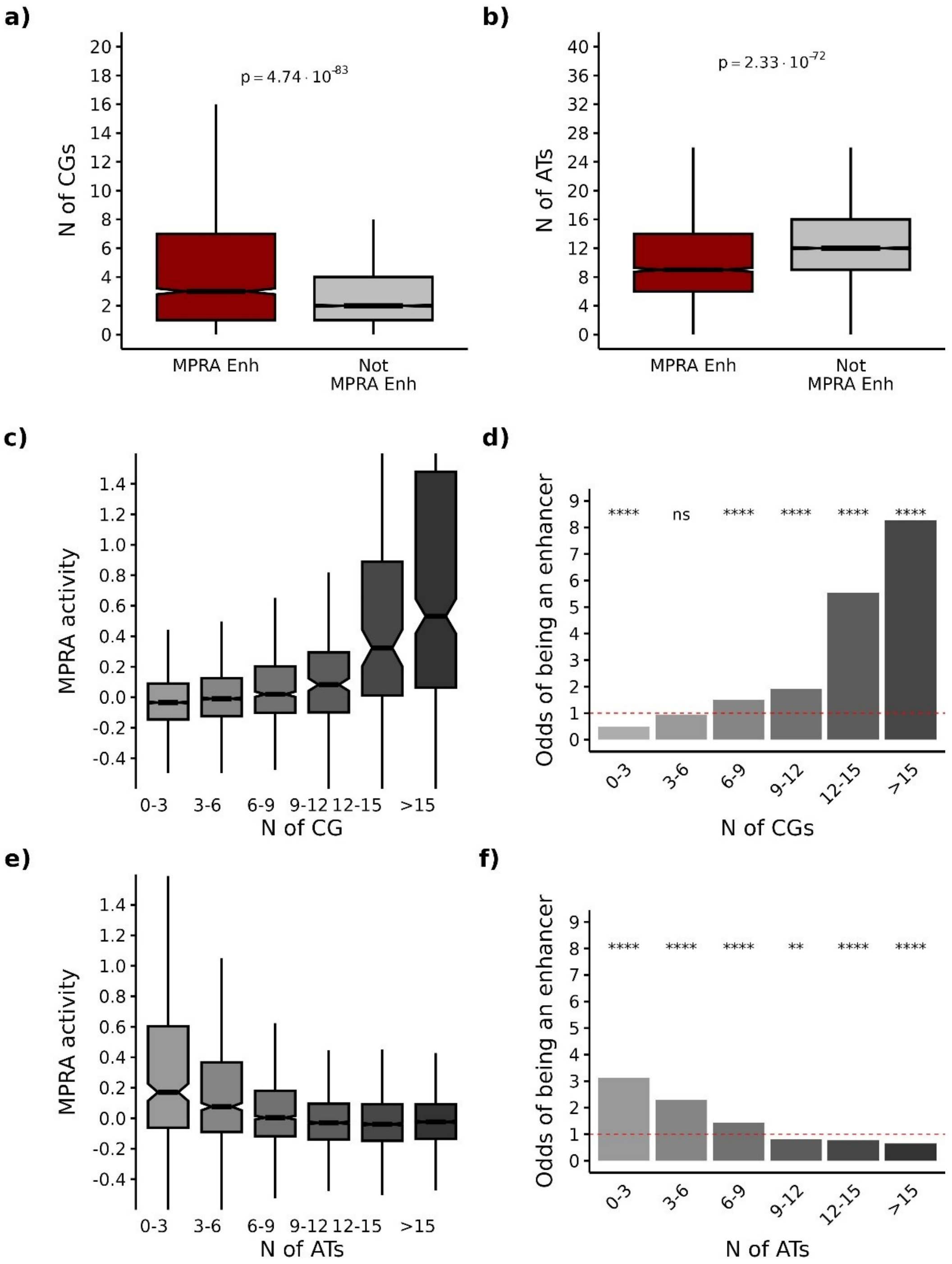


Figure 26. Correlation of enhancer activity with the number of CG and AT nucleotides. a) Number of CG dinucleotides versus CREs that are not active MPRA enhancers. b) Number of AT dinucleotides versus CREs that are not active MPRA enhancers. c) MPRA activity of a CRE as a function of the number of CG dinucleotides. d) Odds ratio of a CRE being an enhancer as a function of the number of CG dinucleotides. e) MPRA activity of a CRE as a function of the number of AT dinucleotides. f) Odds ratio of a CRE being an enhancer as a function of the number of AT dinucleotides. For clarity of display, p-values in panels d and f were converted to significance symbols using the following scheme: p < 0.05: "*", p < 0.01: "**", p < 0.001: "***", and p < 0.0001 "****"; non-significant values are labelled "ns."

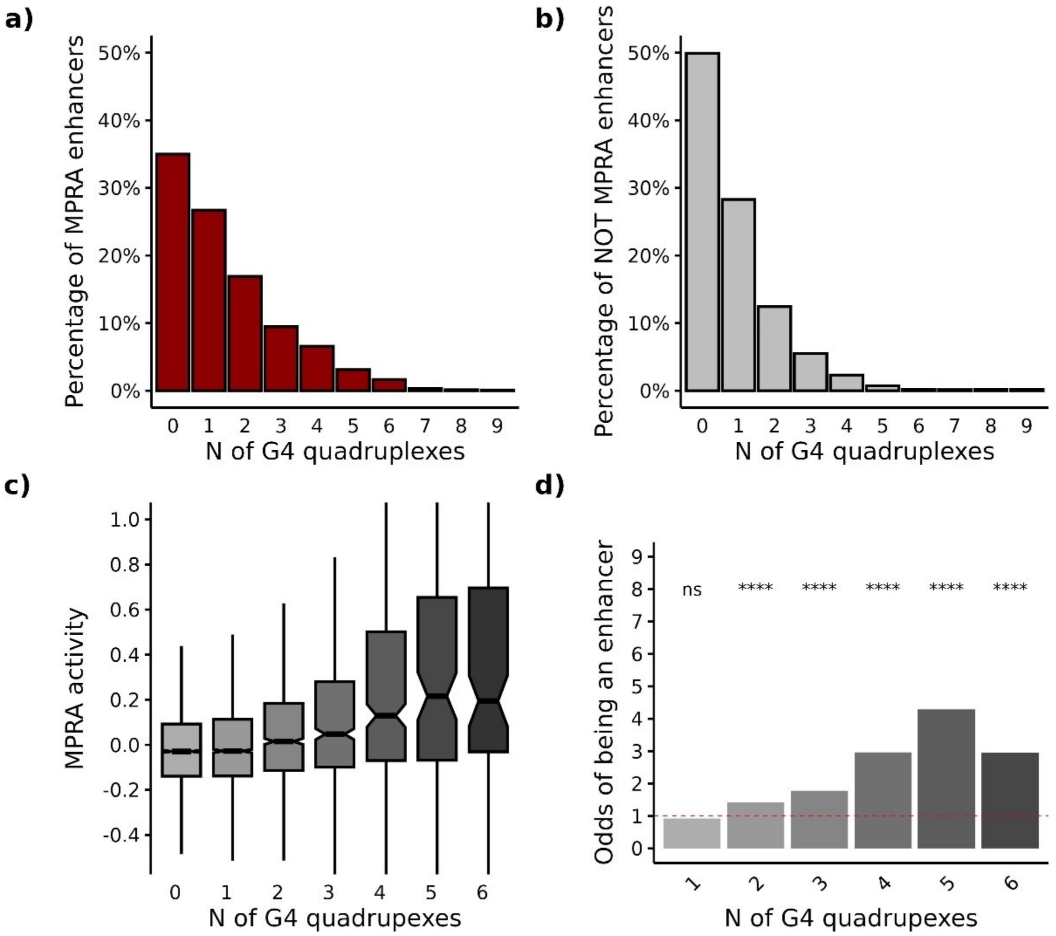


Figure 27. Correlation of enhancer activity with the number of G4 quadruplexes. a) Distribution of the number of G4 quadruplexes in MPRA active enhancers. b) Distribution of the number of G4 quadruplexes in CREs, which are not MPRA active enhancers. c) MPRA activity of a CRE as a function of the number of G4 quadruplexes. d) Odds ratio of a CRE being an enhancer as a function of the number of G4 quadruplexes. For clarity of display, p-values were converted to significance symbols using the following scheme: $p < 0.05$: "*", $p < 0.01$: "**", $p < 0.001$: "***", and $p < 0.0001$ "****"; non-significant values are labelled "ns."

Intact-C contact maps can be combined with MPRA data to generate complete characterisations of specific loci. Figures 28 and R29 illustrate examples of such regulatory landscapes, including both enhancers and repressors, within the contact domains of two astrocyte-related genes: *TEAD3* (chr6: 35,473,597-35,497,079, hg38 human genome assembly), a gene previously identified as a target of evolutionary change in astrocytes[286], and *VIM*, (chr10:17,228,241-17,237,593, hg human genome assembly) a key cytoskeletal component of astrocytes[462].

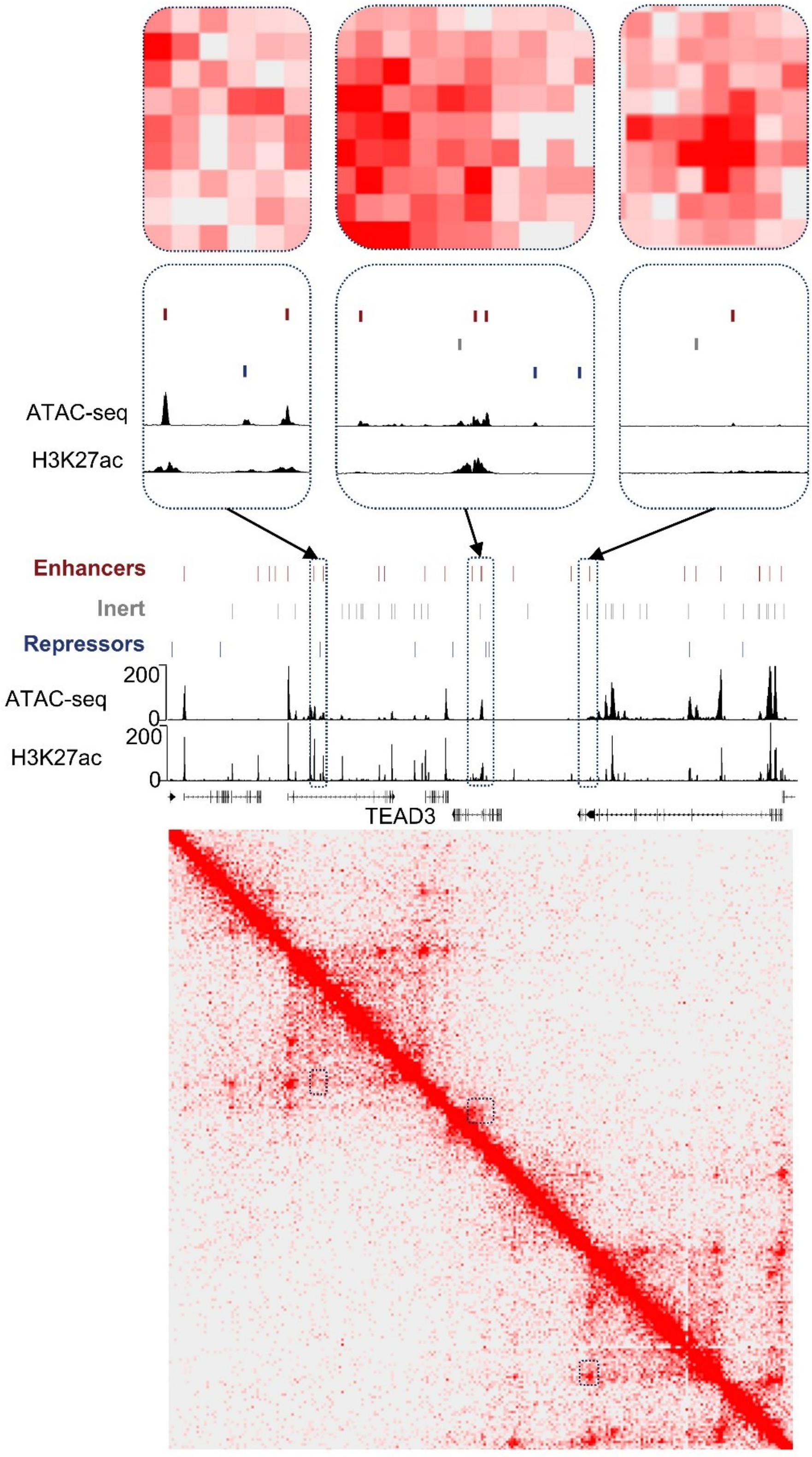


Figure 28. Integration of MPRA, ATAC-seq, and ChIP-seq for H3K27ac and Intact Hi-C to visualise the genomic landscape surrounding the TEAD3 gene. Enhancers, inert sequences, and repressors are marked above the line. Illustrative sites are shown in greater detail in boxes.

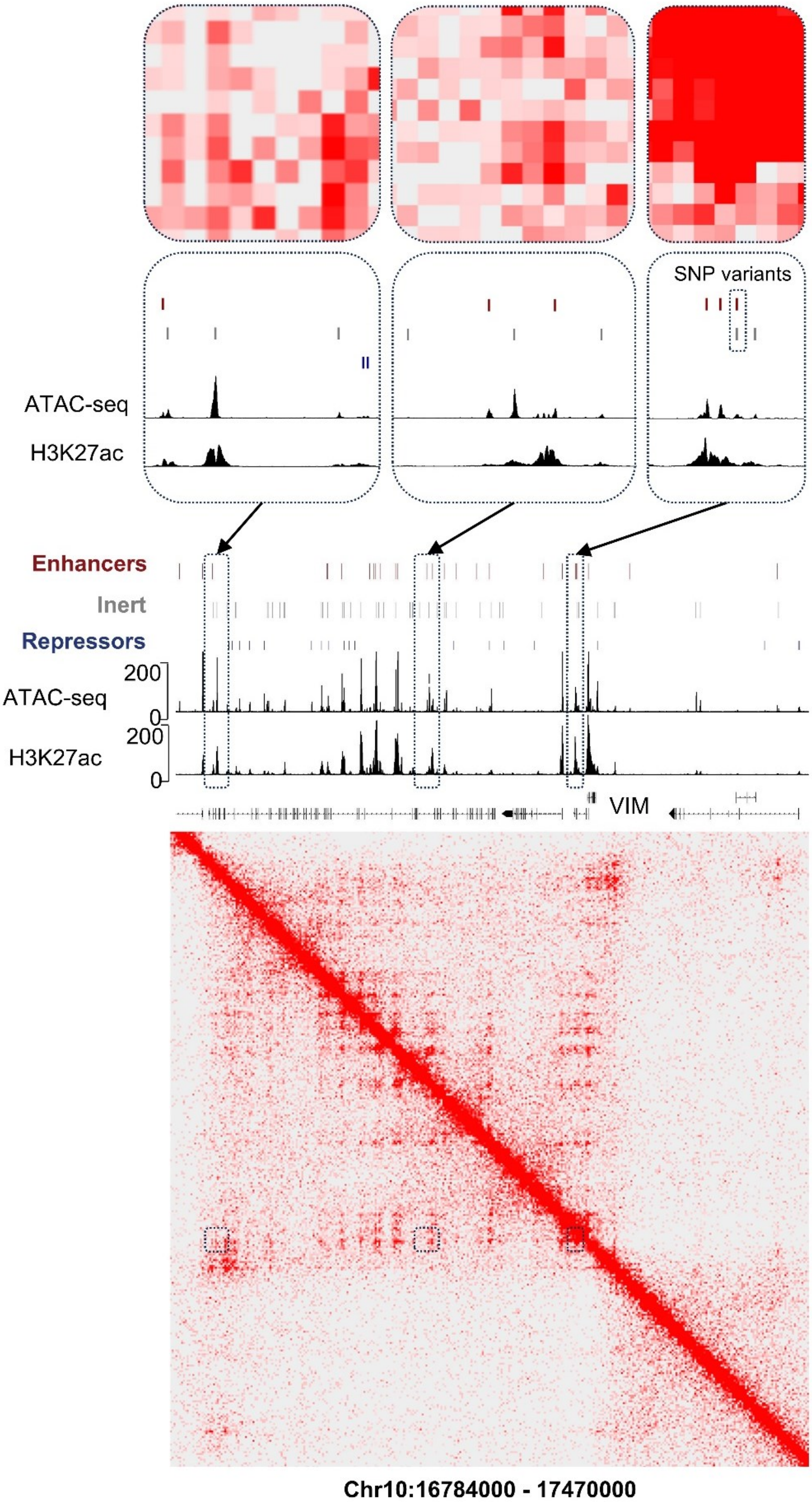


Figure 29. Integration of MPRA, ATAC-seq, and ChIP-seq for H3K27ac and Intact Hi-C to visualise the genomic landscape surrounding the VIM gene. Enhancers, inert sequences, and repressors are marked above the line. Illustrative sites are shown in greater detail in boxes.

I attempted to quantify whether the MPRA activity of active MPRA enhancers, inert sequences, and MPRA repressors is correlated with their positioning at chromatin loop anchors. I intersected genomic positions of those elements with loops called by Juicer on Intact Hi-C data. I found that MPRA activity of active MPRA enhancers correlates with the frequency of promoter contact (Figure 30a $p = 5.97*10^{-4}$, two-sided Wilcoxon Rank Sum Test). I did not observe this correlation for MPRA repressors and for regions that are not active MPRA enhancers or MPRA repressors (Figure 30a, inert sequences: $p = 9.18*10^{-2}$, repressors: $p = 6.77*10^{-1}$, two sided Wilcoxon Rank Sum Test).

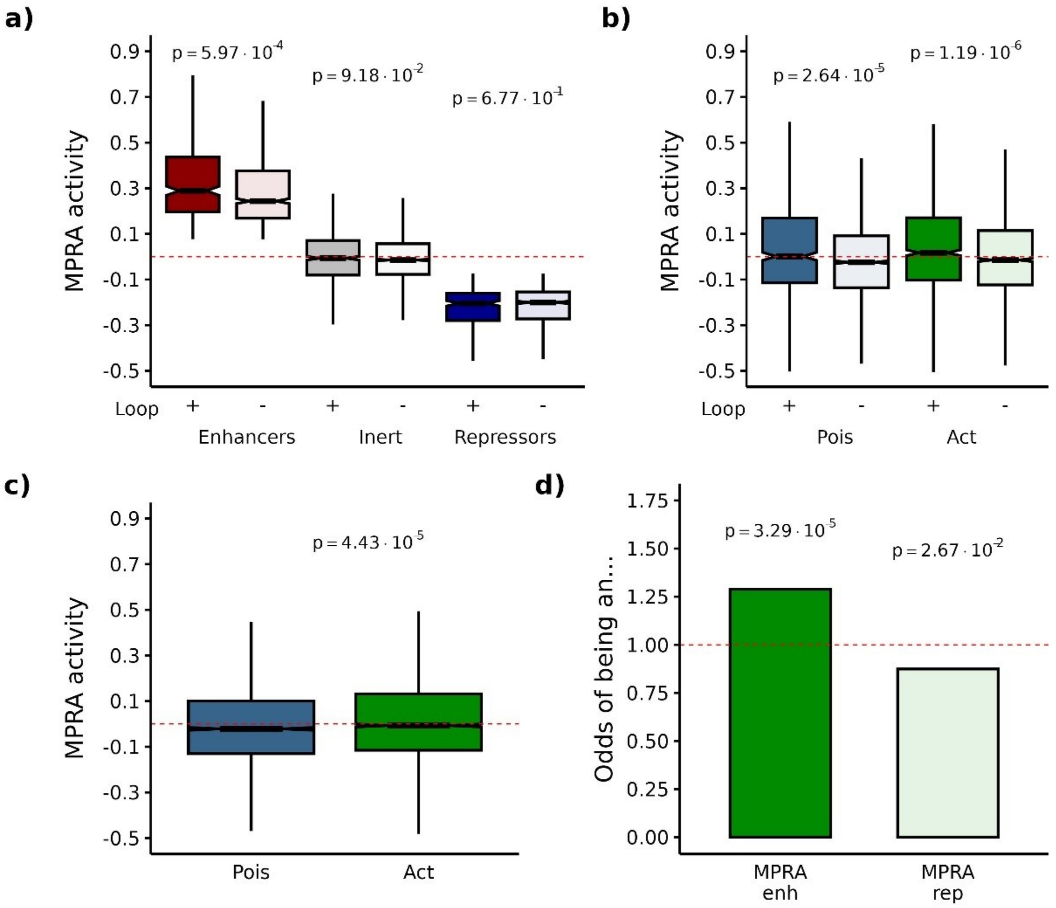


Figure 30. Correlations between MPRA signal and global enhancer annotations. a) Comparison of MPRA activity depending on Juicer-called loop contact, for MPRA enhancers, inert sequences, and MPRA repressors. b) MPRA activity of poised enhancers (ATAC-seq peaks that do not overlap H3K27ac peaks, "Pois") and active enhancers (ATAC-seq peaks that overlap H3K27ac peaks, "Act"), depending on loop formation. c) MPRA activity of poised enhancers and active enhancers. d) Odds of being an MPRA enhancer, or MPRA repressors, for active enhancers, relative to poised enhancers.

I found that active enhancers (ATAC-seq peaks that overlap H3K27ac) are more frequently active in the MRPA screen than poised enhancers (ATAC-seq peaks that do not overlap H3K27ac)(Figure 30c). CREs derived from active enhancers indeed display higher MPRA activity ($p = 4.43*10^{-5}$, two-sided Wilcoxon Rank Sum Test), which is in concordance with the standard predictions of enhancer activity[277]. Furthermore, MPRA activity is higher for both active and poised enhancers, which positioned at loop anchors, compared to those that are not (Figure 30b, $p = 1.19*10^{-6}$ and $2.64*10^{-5}$, respectively, two-sided Wilcoxon Rank Sum Test). These results validate MPRA as a tool for the prediction of genuine active enhancers and position my observations in a wider genomic context.

4.8.3. TFBS correlating with enhancer activity

Next, I asked if the MPRA readout can help discern the TFBS that determine astrocytic enhancer activity. I analysed the relationship between the presence of putative TFBS, inferred bioinformatically, and MPRA activity scores. I focused the analysis on TF that are expressed in foetal astrocytes, as described by Zhang et al., (2016).

I found that USFs are significantly enriched among the TFBS that are present in the most active enhancers relative to TFBS present in the least active enhancers (Figure 31, the right side of the panel shows quantification of the difference, $p = 3.68*10^{-4}$, Fisher's exact test). Furthermore, MPRA data revealed that USF binding sites occur more frequently in active enhancers than in repressors (Figure 31b). I did not detect similar enrichment for other TFBS. Hence, USF binding sites are frequently required for a sequence to function as an enhancer.

Notably, several of the TFs present in the most active enhancers belong to the FOS/JUN family, which together form the AP-1 transcription factor complex[348], which are also enriched among TFBS in active human astrocyte enhancers (See Figure 3). Those factors are also the factors most enriched among active enhancers identified sign genomic annotation (ATAC-seq/H3K27ac, Figure R3), which validates their activity. Very interestingly, FOS activity in astrocytes has recently been shown to be crucial for memory formation [225,226], which suggests that activity of enhancers mediated by those TFs in astrocytes, might be important for higher cognitive functions, including memory.

A number of TFs recognizing cAMP-response elements (CRE): CREB1, CREB3, ATF1, ATF3 and ATF7 are also among the TFs present in the most active enhancers. Cyclic AMP-Responsive Element-Binding Protein 1 (CREB1) has been linked to long-term memory formation[463,464]; CREB1 is instrumental in the regulation of engram cells[227]. CREB1 is also a known target of the CREBBP protein, which is dysfunctional in Rubinstein–Taybi syndrome[465], a model “epigenetic disorder”. Activating Transcription Factor 3 is a crucial regulator of astrocytic stress response and reactivity[466–468]. ATF3 also affects behaviour - ATF3 KO mice exhibit increased fear conditioning[469], and overexpression of ATF3 increases depression-like behaviour in mice[470].

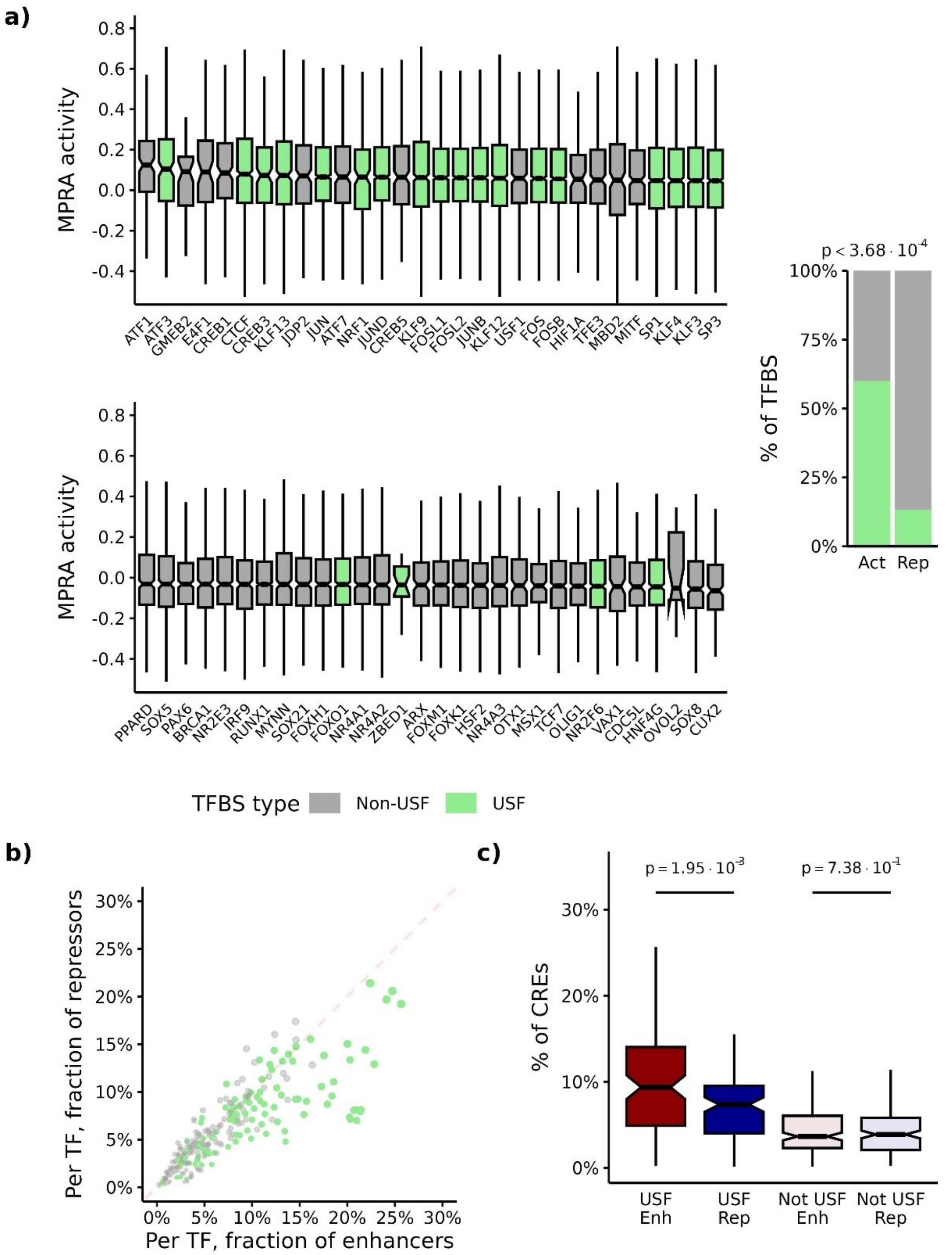


Figure 31. TFBS that determine enhancer activity. a) TFBS binding sites detected in sequences tested using MPRA, and activity of CREs containing those TFBS. 30 TFBS with the highest median enhancer activity are shown in the top panel, and 30 TFBS with the lowest median enhancer activity are shown in the bottom panel. USFs marked in green. Quantification of enrichment of USFS among TFBS associated with enhancer activity is shown in the barplot on the right. b) Fraction of MPRA enhancers and MPRA repressors that contain a given TFBS. Each dot represents a unique TFBS. The size of each dot corresponds to the number of linked enhancers that contain it. Green points correspond to USFs. **e)** Enrichment of USFs in MPRA enhancers and MPRA repressors. On the Y axis, the fraction of MPRA enhancers containing a given TFBS is shown.

Other notable TFs present in the most active enhancers are Methyl-CpG Binding Domain Protein 2 (MBD2), Hypoxia-inducible factor-1 (HIF1A) and Kruppel-like factor 13 (KLF13). MBD2 is an epigenetic regulator that binds methylated DNA, and its deletion in the mouse results in behavioural alterations resembling ASD[471]. HIF1A is a crucial regulator of responses to O_2 fluctuations[472], which controls levels of astrocyte-secreted anti-inflammatory factor HB-EGF[473]. *KLF13* is one of the genes affected by 15q13.3 microdeletion syndrome[474,475], which manifests in severe cognitive disability and symptoms of ASD[476]. The importance of those TFs for enhancer activity highlights the relevance of the astrocyte regulome in health and disease.

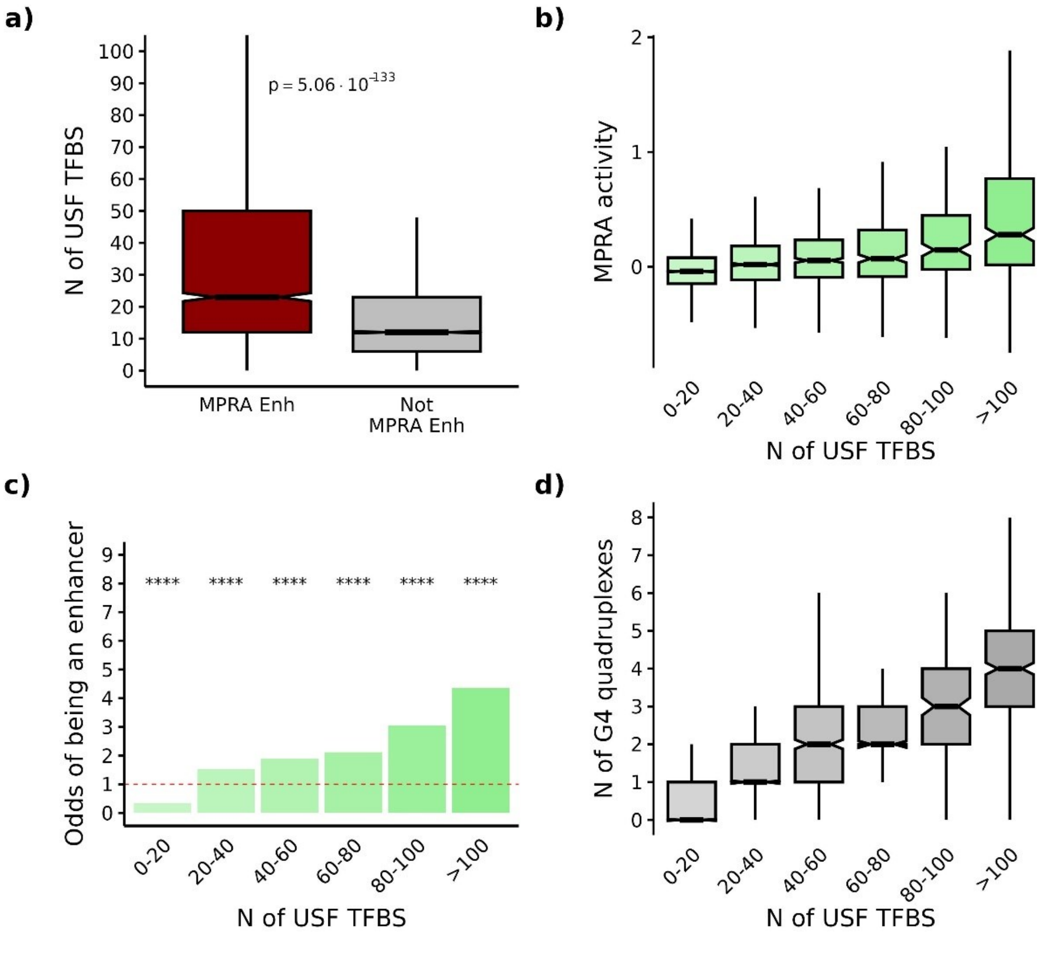


Figure 32. Correlation of enhancer activity with the number of USF TFBS. Only USF expressed in foetal astrocytes were considered. a) The number of USF TFBS in MPRA enhancers versus CREs that are not active MPRA enhancers. b) MPRA activity of a CRE as a function of the number of USF TFBS. c) Odds ratio of a CRE being an enhancer as a function of the number USF TFBS. d) Number of G4 quadruplexes as a function of the number USF TFBS.

Those observations suggest that the presence of USF TFBS correlates with MPRA activity. Indeed, MPRA active enhancers have more USF sites than non-enhancers (Figure 32a, p =

$5.06*10^{-133}$, two-sided Wilcoxon Rank Sum Test), the number of USF TFBS appears to linearly correlate with MPRA activity, and the odds of being an enhancer (Figure 32b and c). Number of USFs correlates with enhancer activity better than other analysed characteristics (ATAC-seq, ChIP for H3K27ac, GC/AT dinucleotides, and G4 quadruplexes. Figures 25, 26, 27). Interestingly, the number of USF sites correlates with the number of G4 quadruplexes, hinting that mechanisms of behind the influence of those two variables on enhancer activity might be connected. Taken together, these observations further sustain the essential role of USF in the astrocyte regulome. My data also argue for the important role of TFs that bind to CG-rich regions and the capacity of these sequences to form G4 quadruplexes in the regulation of enhancer action. Next, I sought to determine the role of gain in USF in evolutionary changes in enhancer activity.

4.8.4. Evolutionary differences in enhancer activity

To investigate the basis of increased enhancer activity in human astrocytes, I used MPRA to compare the activities of human CREs and their chimpanzee homologues, identified using the LiftOver tool. I tested a randomly selected subset of human-specific “linked” enhancers (Figure 18), enhancers near key astrocytic genes, and enhancers overlapping HARs, as described previously. I considered the human and chimpanzee versions of these elements in my assessment. In total, my MPRA libraries contained 1,546 human-chimpanzee CRE pairs; after accounting for losses in library diversity during cloning and missing human or chimpanzee, the final set for which I performed the analysis included 1,400 pairs.

Around 30% of tested human-chimpanzee CRE pairs differ by more than 50% of their nucleotides (Figure 33a), but, remarkably, I did not detect an obvious correlation between the degree of sequence divergence and MPRA activity change (Figure 33b). This indicates that even minor changes between human and chimpanzee variants can result in a change in MRA enhancer activity, and the changes must be caused by specific sequence alterations.

Next, I sought to address whether the differences in openness between human and NHP iAstrocytes could be ascribed to differences in cis-regulatory activity. To model this, I considered DORs with a stronger ATAC-seq signal in human than in NHP iAstrocytes. Among these, I identified active MPRA enhancers (Figure 24) and compared the MPRA signal of these elements to that of their chimpanzee sequence variants. Notably, human versions of these MPRA-validated elements show significantly higher activity than their chimpanzee counterparts (Figure 33b, $p = 9.1 \times 10^{-15}$, two-sided paired t-test). Hence, MPRA models evolutionary alterations in enhancer activity.

Then, I decided to test if the changes in enhancer activity between human and chimpanzee can be linked to the dynamics of USF. I found that in the human-chimpanzee element pairs assessed above, the human versions of these elements contain a substantially higher fraction of USF binding sites (Figure 33b, $p = 7.55 \times 10^{-3}$, two-sided paired t-test).

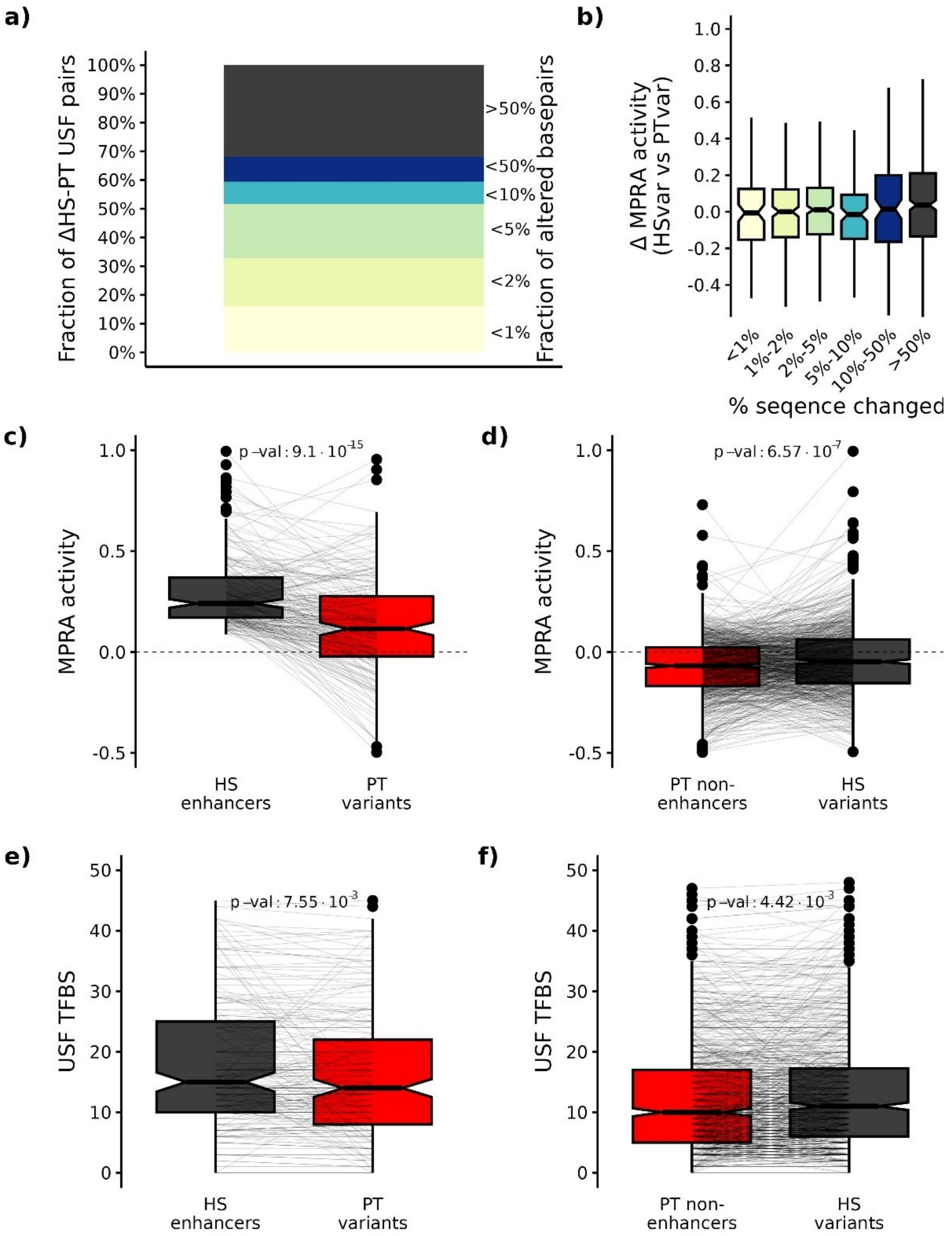


Figure 33. Differences in HS MPRA enhancers and their PT variants. a) Comparison of sequence divergence between HS CREs tested and their PT homologues. b) Comparison of interspecies change in MRPA activity and sequence divergence between HS and PT enhancers. c) Comparison of MPRA activity of HS active MPRA enhancers and their PT variants. d) Comparison of the fraction of USF binding sites among all detected TFBS sites between HS active MPRA enhancers and their PT variants. e) Comparison of MPRA activity of PT inactive CREs and their HS variants. f) Comparison of the fraction of USF binding sites among all detected TFBS sites between PT inactive CREs and their HS variants.

I also examined the question from the opposite perspective: I tested if for chimpanzee CREs are which are inactive, their human homologues is typically be more active. Indeed, human variants of inactive chimpanzee CREs display significantly higher MPRA activity (Figure 33d, $p = 6.57 \times 10^{-7}$, two-sided paired t-test). This activity increase also corresponds to a higher number of USF binding sites in human sequences (Figure 33f, $p = 4.42 \times 10^{-4}$, two-sided paired t-test).

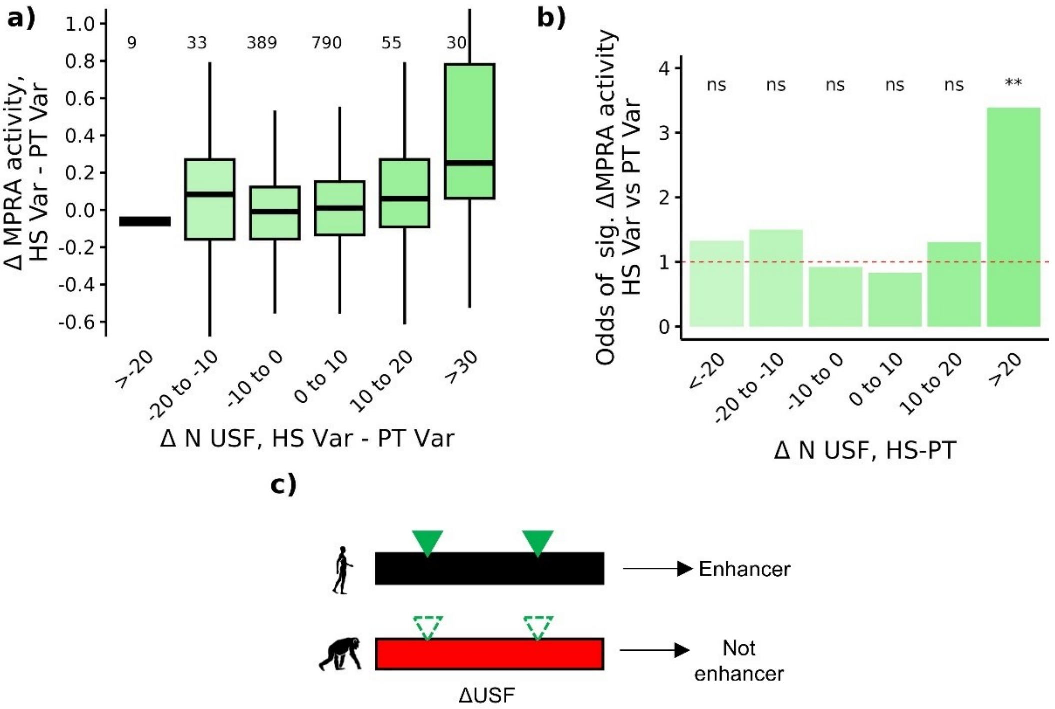


Figure 34. Variation in the number of stripe factors between HS and PT correlates with enhancer activity changes. a) Change in MPRA activity between HS and PT variants of enhancers (positive values indicate higher MPRA activity of HS variant, negative values indicate higher activity of PT variant), depending on the number of USF sites differing between enhancer pairs. Numbers above each category indicate the number of comparisons. b) Odds of MPRA activity of HS variant being higher than MPRA activity of PT variant, depending on the difference in the number of predicted USF sites. c) Schematic of the concept of the importance of USF binding sites of enhancer activity function of HS variants.

To investigate whether changes in MPRA activity are proportional to the number of USFs changed between human and chimpanzee variants, I analysed the magnitude of MPRA activity change between human and chimpanzee variants in the function of USF sequence changes (Δ MPRA activity; positive values indicate higher MPRA activity of the human variant, negative values indicate higher activity of the chimpanzee variant). I found that the greater the number of USF site gains in the assayed sequences, the higher the gain in MPRA activity (Figure 34a).

Similarly, human variants are more likely to be active MPRA enhancers than chimpanzee variants when they contain more USF sites (Figure 34b).

Interestingly, chimpanzee variants with more USF sites were not necessarily comparably more active than the human sequences that gain an equal number of USF. These observations indicate that additional human-specific changes occur that favour enhancer action.

Together with the observed evolutionary gain of USF sites in enhancers near human-specific genes, as we described in Ciuba et al., (2025), these results indicate that changes in USF binding underlie the evolutionary increase in enhancer activity in human-specific loci (Figure 34c). Although highly consistent, the extent of the effect suggests that gains in the USF binding sites are among several contributors to evolutionary changes in enhancer activity.

4.8.5. Verification of activity of linked and non-linked enhancers using MPRA

To assess whether enhancer activity differs between those linked to DEGs and those that are not, I included 3026 “linked” enhancers (within 1 Mb of a DEG) and 1231 “non-linked” enhancers (>1 Mb from a DEG) in one library tested using MPRA. Of those, I managed to determine activity for 2879 “linked” and 1180 “non-linked” enhancers.

Linked enhancers show significantly higher MPRA activity (Figure 35a, $p = 8.03*10^{-3}$, two-sided t-test), though the effect size is extremely small. These results echo our previous predictions.

Given the strong association between USF site number and enhancer activity, I compared USF binding site counts between the two groups and found that linked enhancers contain roughly twice as many USF sites (Figure 35b). This lower USF number in non-linked enhancers likely contributes to their reduced activity.

Crucially, most (71.6%) genes upregulated in human iAstrocytes relative to NHP have at least one nearby linked enhancer that is active in MPRA (Figure 35c)

Finally, as we stated in our recent publication[286], linked enhancers are significantly less conserved than non-linked enhancers. We have shown this using PhastCons conservation scores [477,478], and specifically the PhastCons30way metric, which is based on comparison of human to 27 primate species, and mouse (*Mus musculus*), dog (*Canis lupus familiaris*), and armadillo (*Dasypus novemcinctus*) (p of PhastCons signal 5bp bin in the middle of sequences $= 6.74*10^{-6,}$ t-test, Figure 35e).

Together, these findings suggest that linked enhancers contribute to the elevated expression of upregulated genes in humans relative to other species. Their higher activity likely stems from an increased number of USF binding sites, which may have been gained due to a recent greater rate of evolutionary change. Non-linked enhancers, on the other hand, have fewer USF binding sites, show weaker activity, and relatively greater evolutionary conservation.

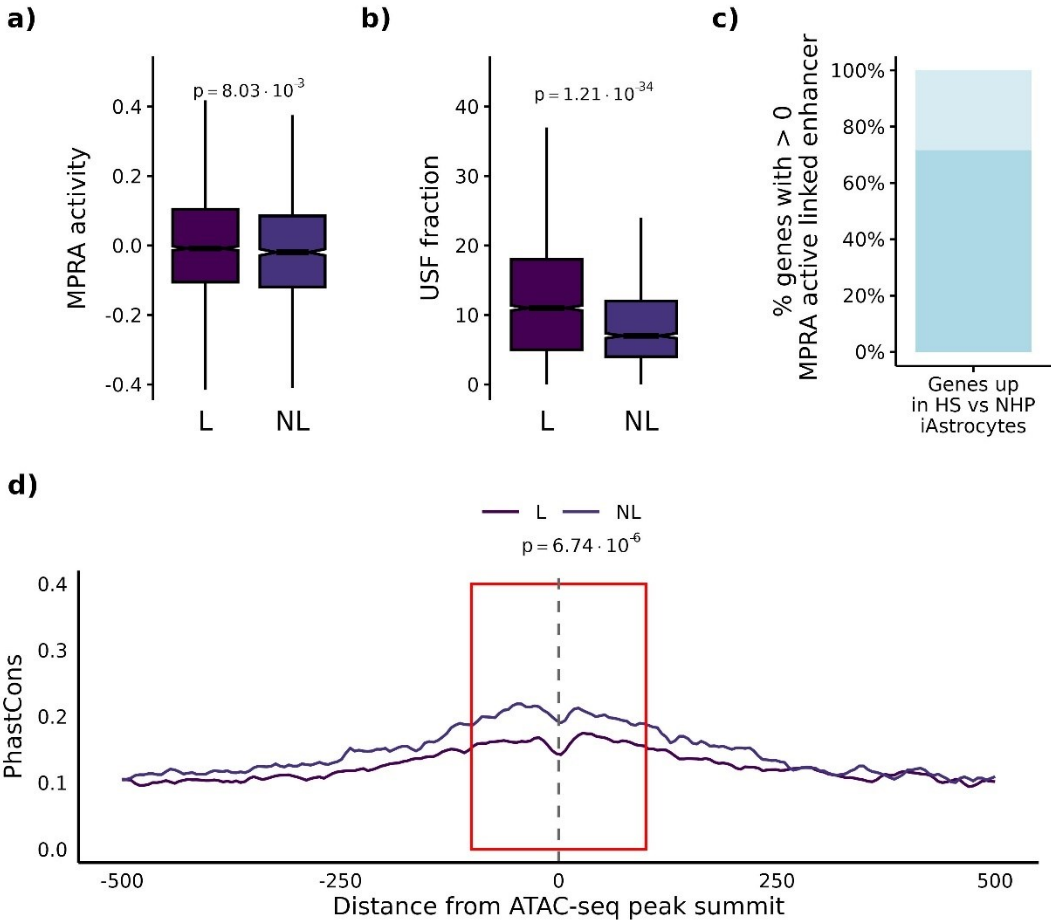


Figure 35. MPRA verification of the activity of "linked" and "non-linked" enhancers. a) Comparison of MPRA activity between "linked" and "non-linked" enhancers. b) Comparison of USF fraction in "linked" and "non-linked" enhancers. c) Percentage of gene upregulated in human iAstrocytes relative to NHP iAstrocytes that have at least one active linked enhancer in their proximity (+/- 1MB). d) Conservation score for linked and non-linked enhancers, represented by *phastCons* score. The central 200 basepair region that is tested in MPRA is marked in red.

To explicitly test this proposal, I identified 769 "linked" enhancers in which only USF binding sites were affected by human-chimpanzee sequence differences (Figure 36a). To test the functional impact of human-specific USF sites, I introduced chimpanzee sequences exclusively at USF binding sites in the human versions of CREs, generating modified sequences that differed only at predicted USF positions (Figure 36a).

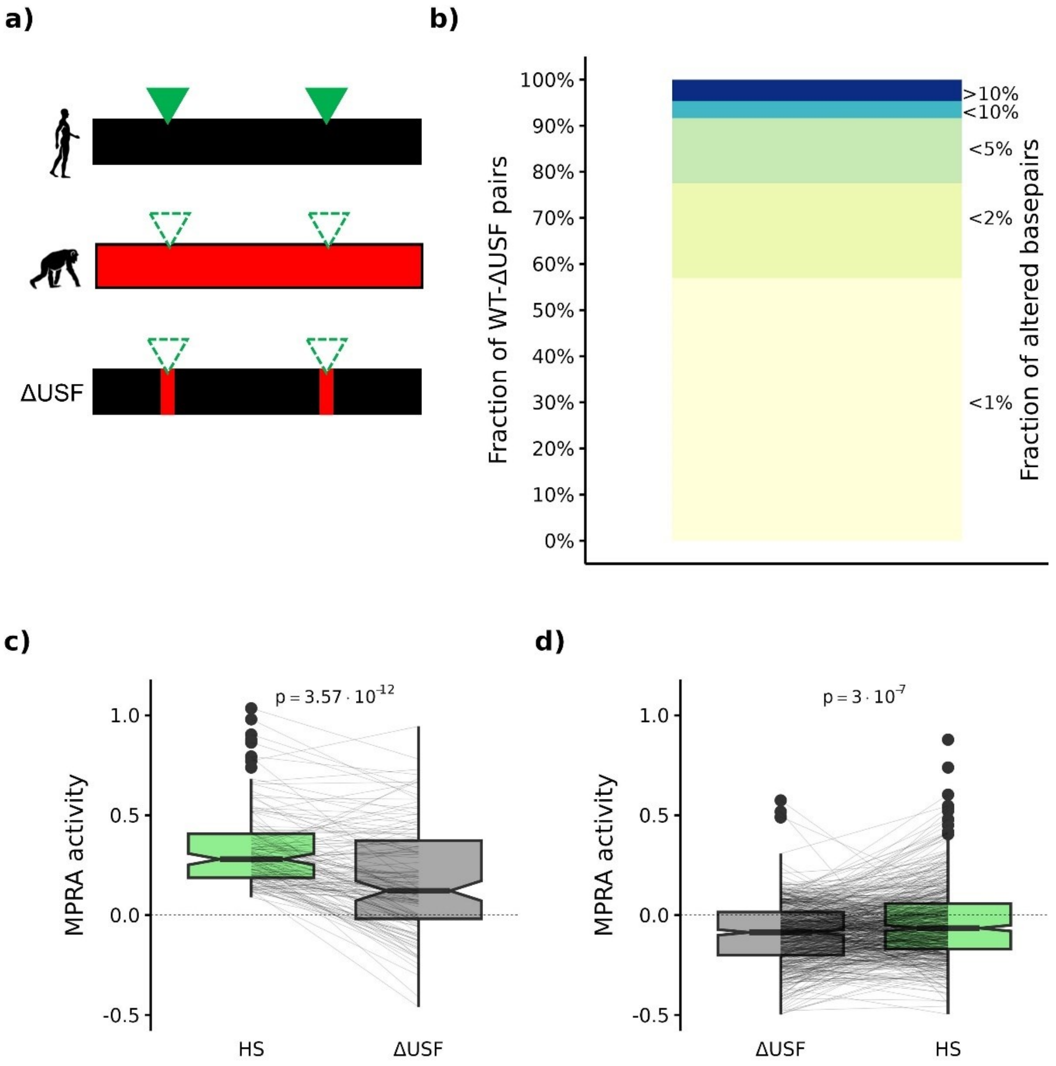


Figure 36. MPRA verification of the importance of USF for enhancer activity of HS-specific enhancers. a) Schematic representation of the design of oligonucleotides used to test the importance of USF binding sites for the activity of enhancers linked to DEGs. b) Distribution of the percentage of base pairs differing between WT and ΔUSF CRE sequences. c) Difference in activity between HS active MPRA enhancers and their ΔUSF counterparts. d) Difference in activity in ΔUSF CREs, which are not MPRA active enhancers, and their HS counterparts.

This approach allowed me to experimentally test the predictions of a convolutional neural network (CNN) developed by Ciuba et al., (2025), which was trained to predict enhancer activity based on regulome data from glioma and iAstrocyte lines. The machine-learning model predicted that replacing human USF motifs with chimpanzee versions would significantly reduce enhancer activity.

In the MPRA experiment, after cloning and processing, I compared 720 pairs of unmodified (“WT”) and modified (“ΔUSF”) enhancers. Among these pairs, over 90% differed by ≤5% of base pairs (i.e. ≤10 bp; Figure 36b). When predicted USF binding sites were disrupted in MPRA active human enhancers, the modified sequences showed a marked reduction in MPRA activity (Figure 36b, $p = 3.57 \times 10^{-12}$, two-sided paired t-test). Conversely, reconstitution of the human USF motif to ΔUSF CREs, which were not active MPRA enhancers, significantly increases their MPRA activity (Figure 36c, $p = 3 \times 10^{-7}$, two-sided paired t-test).

In summary, MPRA results demonstrate that USF binding sites are crucial determinants of enhancer activity and that even minor sequence alterations can significantly alter enhancer output. These findings validate our observations published in Ciuba et al., (2025), showing that evolutionary changes in USF motifs within enhancers near human-specific genes contribute to increased enhancer activity and may underlie unique human astrocytic traits.

4.9. Verification MPRA observations

4.9.1. CRISPRi perturbation of RHOBTB3 enhancer

The MPRA results demonstrate that USF binding sites within enhancers that are proximal to human-specific genes are essential for their activity. Rho-related BTB domain-containing protein 3 (RHOBTB3), is one of the genes differentially expressed between human and NHP iAstrocytes (RNA-seq $\log_2$ fold change in TPM-normalised expression: human vs chimpanzee = 1.75; human vs rhesus macaque = 2.10). RhoBTB3 is involved in intracellular vesicle trafficking[479,480], an aspect of astrocyte biology that differs between human and NHP astrocytes[286]. RhoBTB3 also functions as a component of the CULLIN3 (CUL3)-dependent E3 ubiquitin ligase complex, potentially influencing cell cycle dynamics[481–483]. It has also been shown to promote degradation of a crucial transcription factor HIFα[484]. Moreover, it has been suggested as an AD vulnerability gene[485].

To confirm that enhancer activity detected by MPRA is functionally relevant for gene regulation, I employed CRISPR interference (CRISPRi). This system uses a catalytically inactive Cas9 (dCas9) fused to the Krüppel-associated box (KRAB) domain, which induces heterochromatin formation, and the methyl-CpG-binding domain of MeCP2, which recruits histone deacetylases to repress transcription. Thus, this effector has a triple repressive effect on a targeted locus - dCas9 blocks steric interactions, KRAB reduces chromatin accessibility, and MeCP2 removes permissive histone acetylation marks.

I employed the system to perturb an enhancer upstream of RHOBTB3, named R1. Enhancer R1 contains a human-specific change in USF binding sites and was tested as part of the comparison of human WT and ΔUSF CREs. MPRA data show that the WT R1 enhancer is significantly more active than its ΔUSF variant (Figure 37a, $p = 5.73 \times 10^{-3}$, one-sided paired t-test). Furthermore, R1 enhancer lies in a region that forms an interaction with the promoter of RHOBTB3 (Figure 37b, upper panel), suggesting its contribution to human-specific upregulation of RHOBTB3. Both ATAC-seq and H3K27ac ChIP signals at R1 are increased in human compared to NHP iAstrocytes (Figure 37b, middle and lower panels), reinforcing the link between USF gain, enhancer activity, and RHOBTB3 expression.

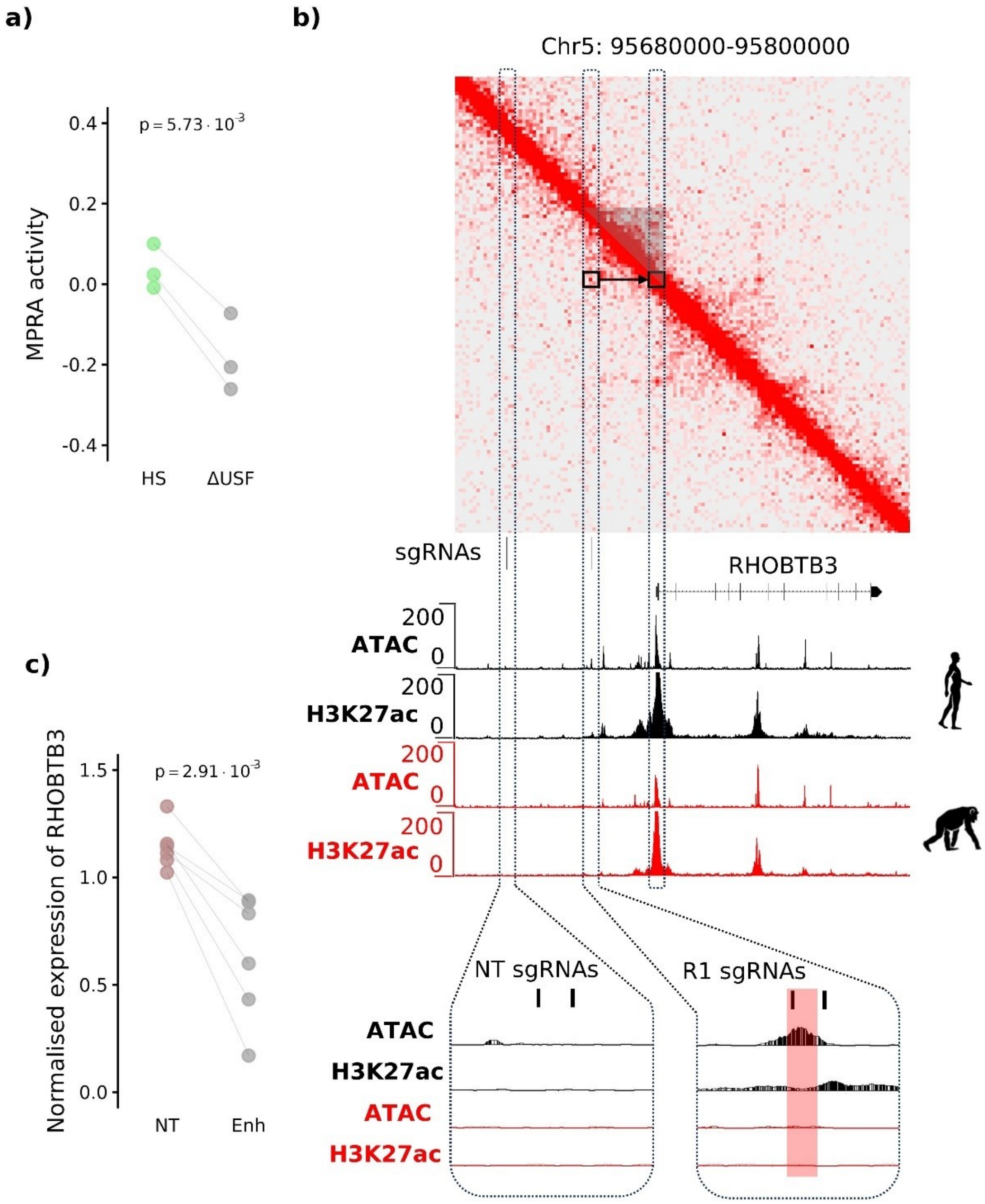


Figure 37. Perturbation of RhoBTB3 enhancer R1. a) Results of MPRA for WT HS variant and the variant with removed binding sites of USFs (ΔUSF). Each dot represents an individual replicate of the MPRA experiment. b) Design of CRISPRi perturbation of R1 enhancer. Intact Hi-C map of the region shown in the upper panel. The small contact domain formed upstream of the RhoBTB3 promoter is marked by a shaded grey triangle. The Intact Hi-C signal showing contact between the R1 enhancer and RhoBTb3 promoter is shown as solid boxes connected with an arrow. TPM-normalised ATAC-seq (ATAC) and ChIP for H3K27ac (H3K27ac) signal tracks for HS (black) and PT (red) are indicated below the Intact Hi-C map. Locations of the sgRNAs targeting the R1 enhancer and non-targeting control ("NT") are shown above genomic assay tracks and are [resented in greater detail in boxes in the lowest panel. c) Normalised expression of RhoBTB3 measured using qPCR, as an effect of perturbation of a control region (NT) or R1 enhancer (R1). Values normalised to Beta actin expression. Values for cells transfected with scrambled sgRNAs = 1.

I targeted the R1 region using two pairs of sgRNAs - one pair directed at the enhancer and another pair targeting a non-active genomic site that lacked active chromatin marks and did not form contact with the RhoBTB3 promoter (Figure 37b). A pair of non-targeting, scrambled sgRNAs served as a negative control.

To deliver the components of the CRISPRi system, I employed two types of viral particles: one encoding the repressor construct (dCas9–KRAB–MeCP2) and another carrying the sgRNA pair. After transduction, cells were selected with antibiotics to ensure enrichment for double-transduced populations. I then extracted RNA and performed RT–qPCR to quantify changes in RhoBTB3 expression following targeted epigenetic perturbation.

This experiment was conducted in two iAstrocyte lines (ELE30 and AG94)[302,303], each in triplicate (n = 6 per condition). Epigenetic repression of the R1 enhancer caused a significant reduction in RhoBTB3 expression relative to the non-targeting control (Figure 37c), confirming that R1 functions as a *bona fide* enhancer.

4.9.2. Luciferase assay tests

To elucidate the effects of individual sequence differences between the human and chimpanzee variants of the R1 enhancer, I used a luciferase reporter assay to assess the contribution of each specific human-chimpanzee difference to enhancer activity. I constructed luciferase assay plasmids (based on modified pGL4 plasmid from Promega), in which enhancer sequences were cloned upstream of a minimal promoter driving firefly luciferase expression. To maintain consistency with the MPRA setup, I used the same minimal promoter sequence from the MPRA plasmid (pMPRAv3:∆luc:∆XbaI). I employed a co-transfected Renilla luciferase reporter plasmid to normalise the readout. I performed this experiment in two human iAstrocyte lines (ELE30 and AG94) [302,303] and in duplicate for each line (4 samples per construct).

I considered the sequence of the ATAC-seq peak called by MACS2 (see Materials and Methods), which spans 315 bp. The two variants differ at four positions: 108 (G in human, A in chimpanzee), 211 (C insertion in chimpanzee), 293 (C in HS, A in chimpanzee), and 298 (A in human, C in chimpanzee) (Figure 38a). I generated luciferase constructs containing all possible combinations of human and chimpanzee variants at these positions (16 constructs in total). Direct comparison of human and chimpanzee versions of the full-length enhancer confirms observations from the MPRA assay – chimpanzee variant is significantly less active in luciferase assay than the human variant, although the effect size is not big (Figure 37b, p = 1.46 $*10^{-3}$, two-sided Wilcoxon Rank Sum test).

Surprisingly, the gradual introduction of chimpanzee sequence variants does not lead to a stepwise decrease in enhancer activity toward the level of the full-length chimpanzee enhancer (Figure 38c), which I expected to observe. Instead, certain combinations of human and chimpanzee variants in the enhancer sometimes result in activity significantly increased relative to the fully human version of the enhancer, as in the case of the PT1_HS2_HS3_HS4 version and PT1_HS2_HS3_PT4 version. Among the four positions, Site 3 appears to be the most functionally critical: four out of eight enhancer versions containing the PT sequence at Site 3 show significantly lower activity than the fully human version. To explore the possible basis for the observed variability in effects of human-chimpanzee sequence differences, I retrieved FIMO annotations of predicted TFBS for each site within the enhancer (Figures 39 and 40)

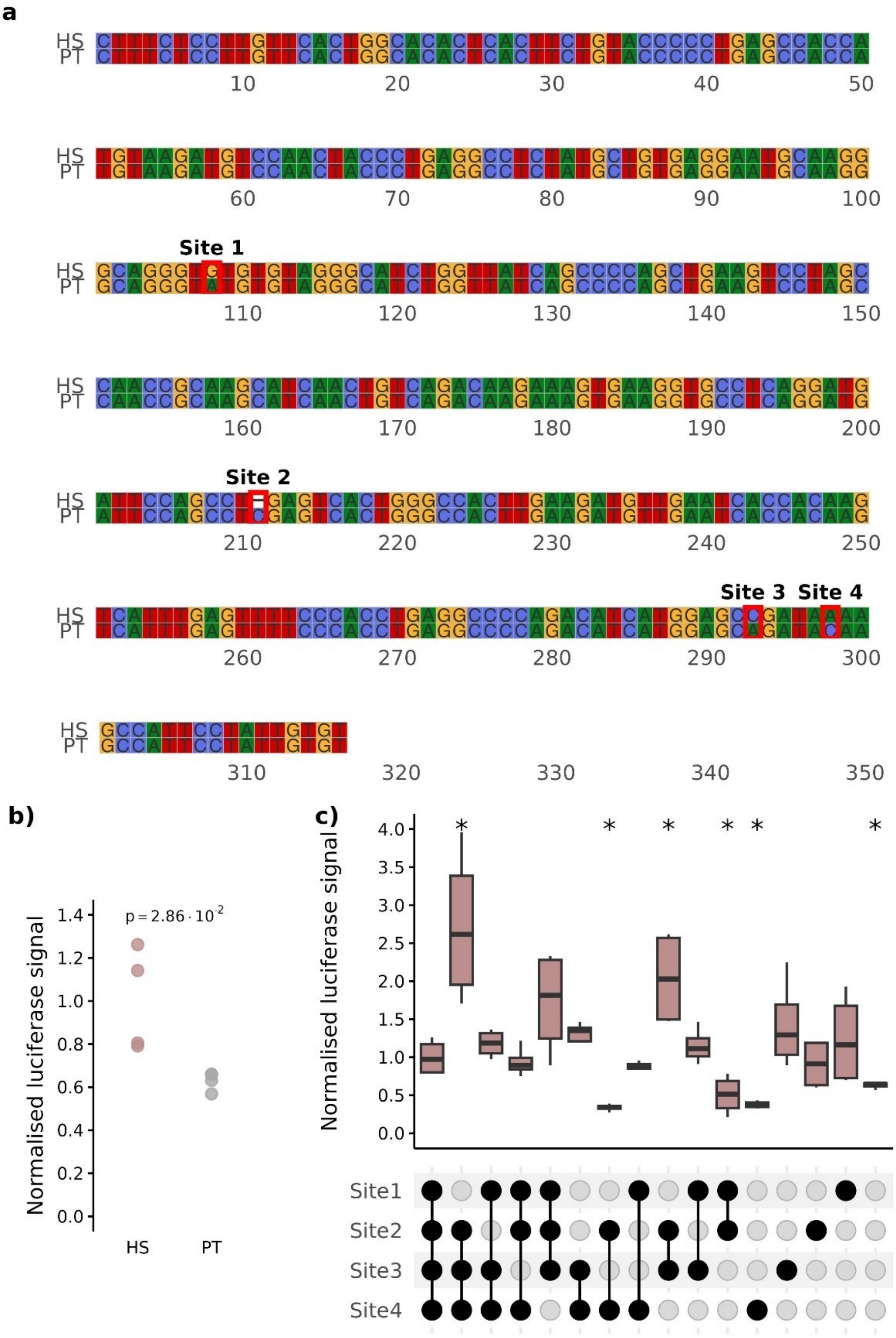


Figure 38. Dissection of the role of USF sites in the activity of the enhancer at the RhoBTB3 locus. a) Alignment of the HS and PT versions of the RhoBTB3 enhancer R1, with sequence differences highlighted in red. b) Comparison of luciferase signal between the HS and PT variants of enhancer R1. c) Functional dissection of the R1 enhancer. Black circles represent HS variants, while grey circles represent PT variants in each combination. First and last datapoints correspond to complete HS and PT variants, respectively. Significance determined using two-sided Wilcoxon Rank Sum Test. For clarity of display, p-values in panel c were converted to significance symbols

using the following scheme: p < 0.05: “*”, p < 0.01: “**”, p < 0.001: “***”, and p < 0.0001 “****”; non-significant values not labelled.

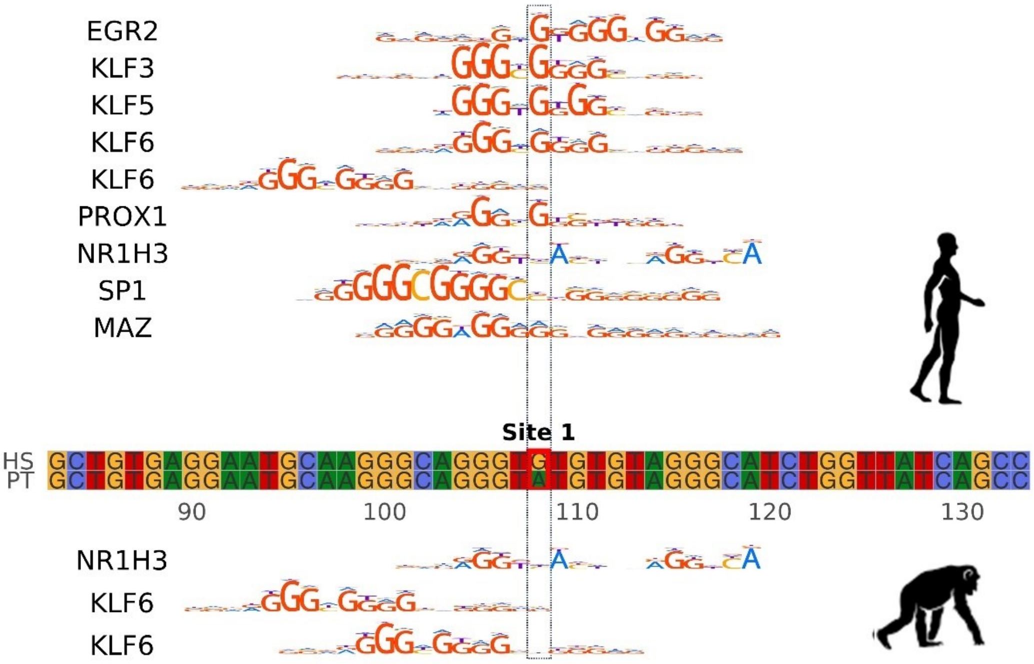


Figure 39. Effects of HS-PT sequence difference on predicted TFBS, R1 enhancer Site 1. Upper panel – TFBS predicted in HS using FIMO. Lower panel – TFBS predicted in PT using FIMO.

The PT-specific substitution at Site 1 affects the crucial basepairs in motifs of EGR2, KLF3, KLF5, one of KLF6 binding sites, PROX1, SP1, and MAZ. However, two out of three predicted KLF6 binding sites remain intact. Given that the introduction of the chimpanzee sequence at this site increases the enhancer activity (when all other sites retained the human sequence), it is plausible that one or more TFs predicted to bind the human variant exert a repressive effect.

The chimpanzee sequence at Site 2 has the greatest effect on predicted TFBS relative to the human variant. It abolishes nine predicted binding sites, including seven from the FOS/JUN family. Consistently, no version of R1 containing the chimpanzee sequence at Site 2 shows higher activity than the fully human variant (Figure 37c). Site 3, which also appears to decrease enhancer activity in the chimpanzee version, contains predicted THAP1, ZBTB6, and TBP binding sites. Interestingly, the human variant of Site 4 may exert a repressive effect. For example, in enhancer variants human, HS1_HS2_HS3_PT4, and PT1_HS2_HS3_PT4, introduction of the chimpanzee sequence at Site 4 increases activity, whereas in PT1_PT2_PT3_HS4 and chimpanzee variants, the construct containing the human sequence at Site 4 is less active than the fully chimpanzee version.

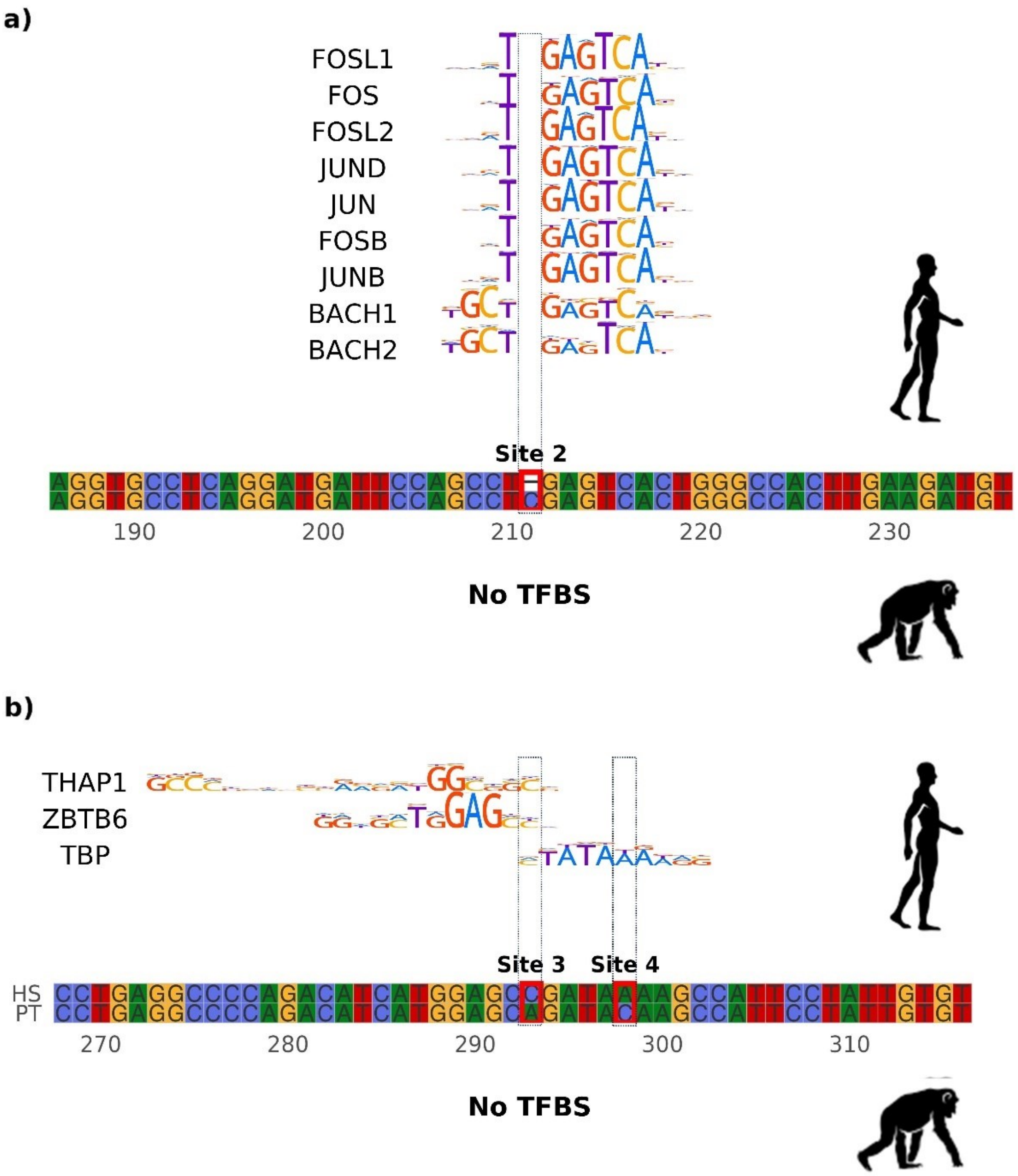


Figure 40. Effects of HS-PT sequence difference on predicted TFBS. a) R1 enhancer Site 2, b) R1 enhancer Site 3 and Site 4. Upper panels – TFBS predicted in HS using FIMO. Lower panel – TFBS predicted in PT using FIMO.

In summary, I confirmed the importance of human-specific gains in USF binding sites for the human-specific activity of enhancer R1. However, the results suggest a complex interplay between the specific arrangement of TFBS and the functional effects of individual sequence changes within this enhancer. These effects appear to be highly context-dependent, as shown by the luciferase assays. It is plausible that evolutionary selection has fine-tuned the human enhancer sequence to achieve an optimal level of activity. The activity of the human enhancer may therefore be lower than the maximal level achievable through alternative polymorphic combinations, reflecting the requirement for a moderate enhancer output at this genomic locus.

4.9.3. Characterisation of USF evolution in active enhancers

Based on the results presented above, changes in USF sites appear to be a key mechanism underlying differences in activity between human and chimpanzee enhancers. To explore evolutionary changes in USF sites across the whole genome, I examined whether USF sites within human enhancers - particularly those linked to genes differentially expressed in human compared with NHP iAstrocytes - bear marks of accelerated evolution. Since the MRPA experiment demonstrates that even a single-nucleotide change can significantly alter enhancer activity (Figure 36), I focused on the human-chimpanzee sequence differences. I identified the nucleotide positions within TFBS predicted in human ATAC-seq peaks that differ between the human hg38 and chimpanzee PanTro6 genome assemblies. I considered a nucleotide to be different between human and chimpanzee if the human nucleotide has undergone substitution, duplication, or deletion in the chimpanzee genome. I then calculated odds ratios of such a difference being present in USF TFBS compared to other TFBS (Figure 41a).

I analysed six enhancer groups: all human enhancers, poised ("Poised"), and active ("Active") enhancers (Figure 2), as well as human-specific enhancers (Figure 18c), subdivided into those not linked to up-regulated genes ("non-linked") and those linked to genes up-regulated in the human lineage ("linked"). This approach allowed me to assess in which of the enhancer categories the evolutionary pressure at USF sites is the strongest.

Across all groups, USF TFBS contain human-chimpanzee sequence differences significantly more often. The effect is stronger in poised (odds ratio = 2.2) than in active enhancers (odds ratio = 1.36), suggesting that human-chimpanzee sequence differences in USFs may preferentially affect regions that are inactive in astrocytes or function in other tissues, whereas USF sites in active enhancers in general are more evolutionarily constrained.

In contrast, the pattern is reversed between linked and non-linked enhancers. USF sites in linked enhancers show the greatest enrichment (odds ratio = 2.6), substantially greater than in non-linked enhancers (odds ratio = 1.78). This indicates that USF TFBS within linked enhancers have undergone especially strong selection, likely enhancing their activity and influencing the expression of adjacent genes.

In order to determine how more recent evolutionary pressures after the divergence of humans from other hominids affected USF binding sites, I performed the same analysis for SNCs, defined by Prüfer et al., (2014) (Figure 41b). Odds ratios for SNCs within USF motifs are lower than those observed for human-chimpanzee sequence differences, likely reflecting the smaller degree of evolutionary divergence between humans and hominids than between humans and chimpanzees. Notably, however, while “linked” enhancers are significantly enriched in SNCs located in USF TFBS (odds ratio = 1.32, p = 2.95*10^{-3}, Fisher’s exact test), “non-linked” enhancers are not (odds ratio = 1.15, p = 3.83*10^{-1}, Fisher’s exact test). This indicates that USF binding sites affecting enhancers linked to human-specific genes have not only undergone selection relative to chimpanzee, but also more recently, following human-hominid divergence.

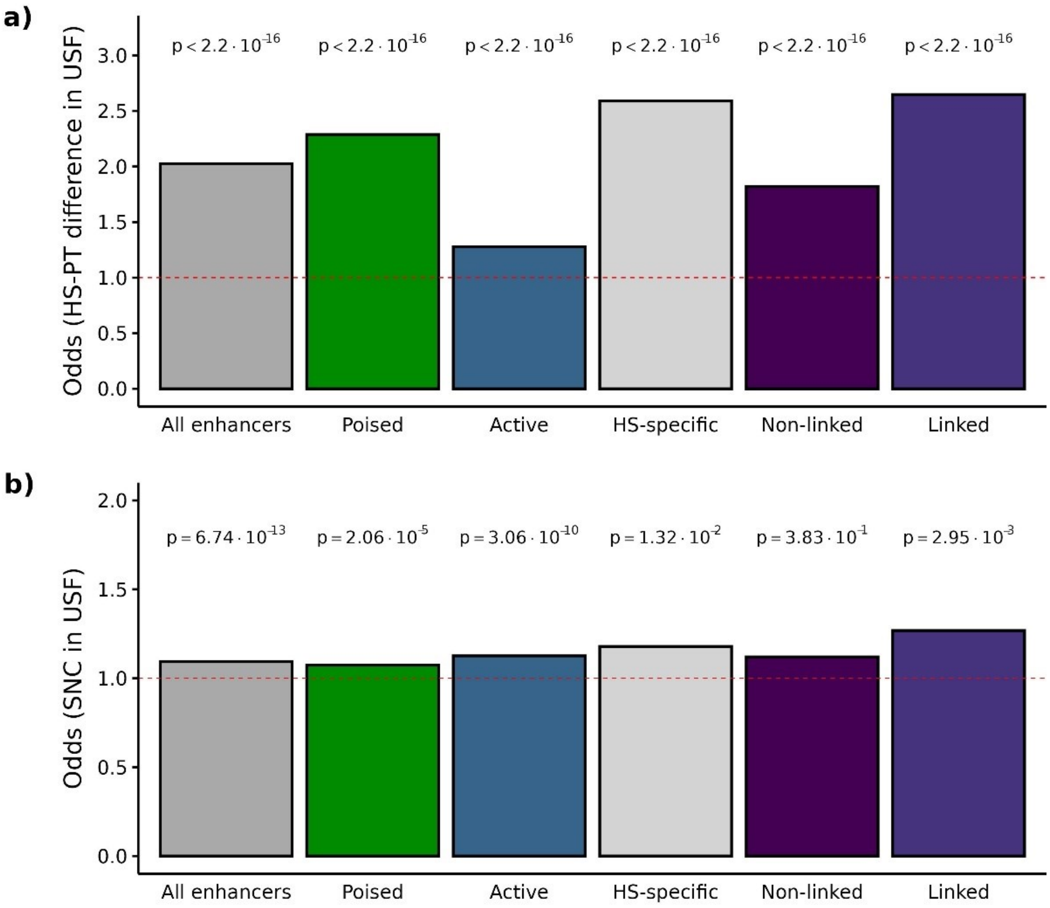


Figure 41. Evolutionary pressure at USF and non USF TFBS sites in HS enhancers. a) Odds ratios of HS-PT sequence differences being located in USF TFBS in different groups of enhancers. Calculations for all enhancers, poised enhancers and active enhancers performed on 10 million of randomly sampled positions from all TFBS. b) Odds ratios of SNCs differences being located in USF TFBS in different groups of enhancers.

I further explored the divergence in USF binding sites between species. I determined the degree of sequence divergence in TFBS located in astrocytic enhancers. For each TF, I calculated the average percentage of nucleotides that are different between human and chimpanzee, and between human and hominids. I then compared that measure between USF and non-USF binding sites. USF sites in poised enhancers contain more human-chimpanzee sequence differences than those in active enhancers (Figure 42a). However, USF motifs in linked enhancers show a markedly greater divergence, with an average of 30% of nucleotides differing between humans and chimpanzees in a given USF motif localised in a linked enhancer, indicating a strong evolutionary change resulting in the gain of USF sites in this enhancer group (Figure 42b; Ciuba et al., 2025). I performed the same analysis for SNCs. Although the overall fraction of SNCs within USF TFBS is small, particularly in poised and active enhancers, linked enhancers exhibit a clear enrichment of SNCs in USF motifs. These findings suggest that USF sites in linked enhancers appeared preferentially, even in the most recent phases of human evolution.

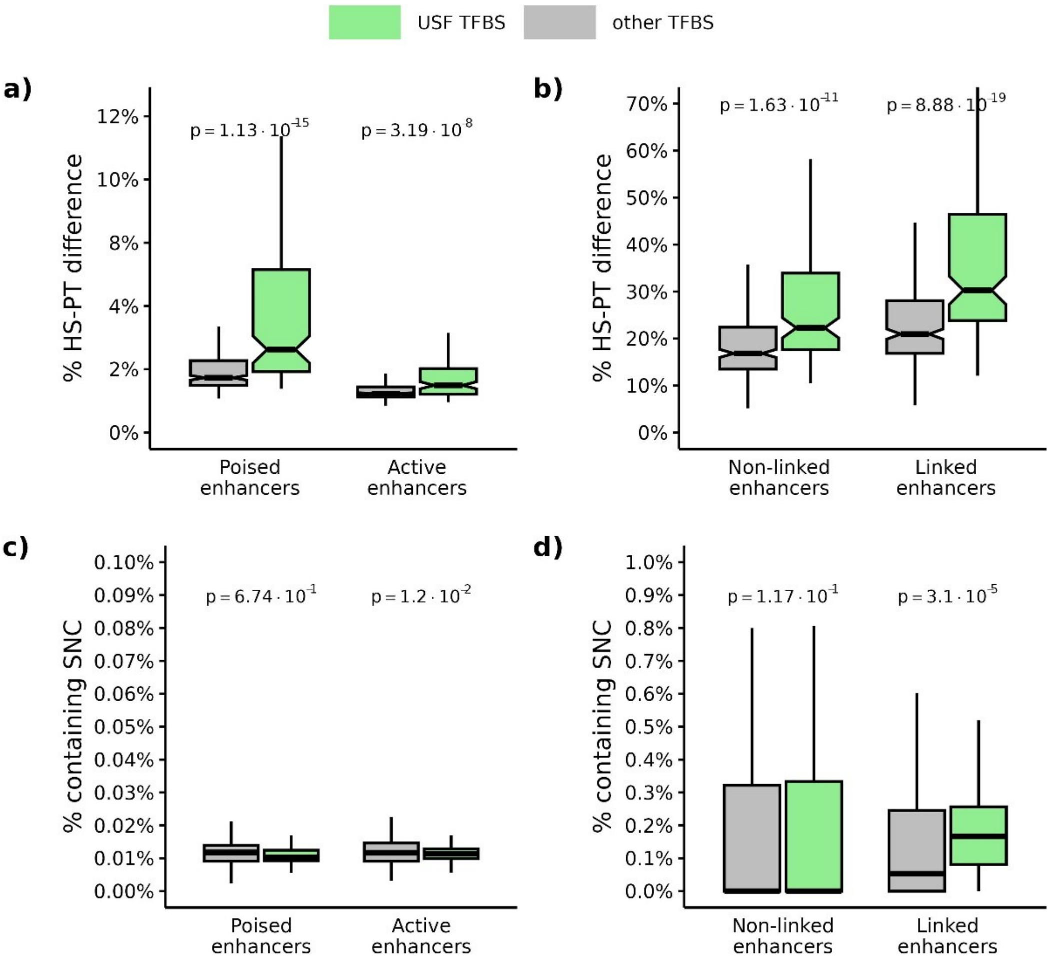
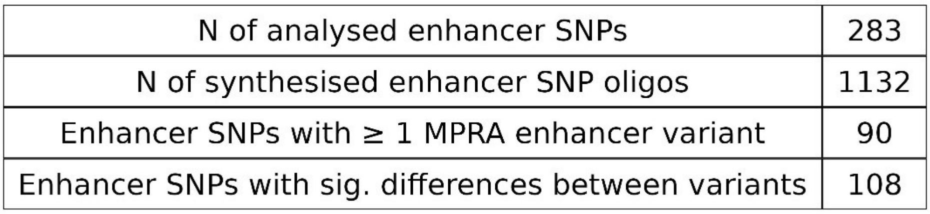


Figure 42. Evolutionary pressure at USF and non-USF binding sites in HS enhancers. Odds ratios of HS-PT sequence differences being located in USF TFBS in a) poised and active enhancers, b) linked and non-linked enhancers. Odds ratios of SNCs differences being located in USF TFBS in different groups of enhancers in c) poised and active enhancers, d) Linked and non-linked enhancers. Green - USF TFBS, grey – other TFBS. P determined using two-sided Wilcoxon Rank Sum Test.

4.10. SNP variants characterised using MPRA

SNPs located in the non-coding regions of the genome frequently modulate the activity of regulatory elements by altering transcription factor binding sites. Enhancers that contain SNPs likely play a key role in shaping traits and disease susceptibility by influencing transcriptional output [486]. To investigate which SNP variants might modulate the activity of foetal astrocyte enhancers, I assayed a number of SNP-containing enhancers using MPRA. Specifically, I considered four sets of SNP-containing enhancers. For the first three groups, I used the data on GWAS SNPs from EBI GWAS Catalogue[349] and grouped SNP-associated phenotypes into three categories: variants linked to intellectual ability, neuropsychiatric disorders, and morphological features of the brain. From those groups, I selected SNPs located in enhancers (ATAC-seq peaks) that were detected in both iAstrocytes and commercially obtained foetal astrocytes from the human brain. The fourth group of the tested SNPs, which were not filtered for specific phenotype associations, comprised variants located in enhancers surrounding genes known to play important roles in astrocyte biology. I used MPRA to test all possible variants described in the Ensembl database[487]. In total, I assayed 283 individual SNPs, for which I tested in total 1132 CRE variants.



N of analysed enhancer SNPs	283
N of synthesised enhancer SNP oligos	1132
Enhancer SNPs with ≥ 1 MPRA enhancer variant	90
Enhancer SNPs with sig. differences between variants	108

Figure 43. A summary of MPRA test of SNP activities.

Out of those SNPs, 90 contained one variant which is an active MPRA enhancer, and for 108 there was a significant difference in activity between any two given variants in a pairwise comparison (Figure 43).

4.10.1. SNP variants and USF TFBS disruption

For each SNP, I identified the most active variant (“maxVar”) and used it as a reference to determine significance in pairwise comparisons to its other sequence variants (“Var”). Building on previous analyses demonstrating the importance of USFs for enhancer activity, I first tested whether SNPs affecting USF binding sites exert stronger effects on MPRA activity than those impacting other TFBS. To address this, I compared the magnitude of MPRA activity change between SNP variant pairs in which there is a gain in the number of USF TFBS between maxVar and Var, and pairs in which the net number of USF TFBS remains constant (Figure 44a).

However, given that changes in USF sites appear to be a crucial mechanism shaping evolutionary changes in enhancer activity, I reasoned that only USF sites under evolutionary pressure might exert a modulatory effect on enhancer activity. To identify which USF sites overlapping SNPs show signs of accelerated evolution, I annotated each SNP with a phyloP score [488]. PhyloP provides a base-pair–level measure of sequence conservation, expressed as the negative logarithm of the *p*-value under the null hypothesis of no conservation. Positive phyloP values indicate sites under purifying (negative) selection, whereas negative values indicate accelerated evolution, or positive selection. I used the phyloP30way metric, which compares human sequences with 27 primate species, using mouse (*Mus musculus*), dog (*Canis lupus familiaris*), and armadillo (*Dasypus novemcinctus*) as outgroups.

When I analysed the effect of USF gain in SNPs under positive and negative selection separately, I found that for SNPs showing signatures of positive selection in primates, USF gain correlates significantly with increased MPRA activity (Figure 44b, $p = 3{,}55 * 10^{-2}$, two-sided Wilcoxon Rank Sum Test). In contrast, USF gain does not significantly alter MPRA activity for SNPs not under positive selection ($p = 6.28 * 10^{-1}$, two-sided Wilcoxon Rank Sum Test). This suggests that evolutionary changes in USF TFBS driven by positive selection may have a more substantial impact on enhancer modulation.

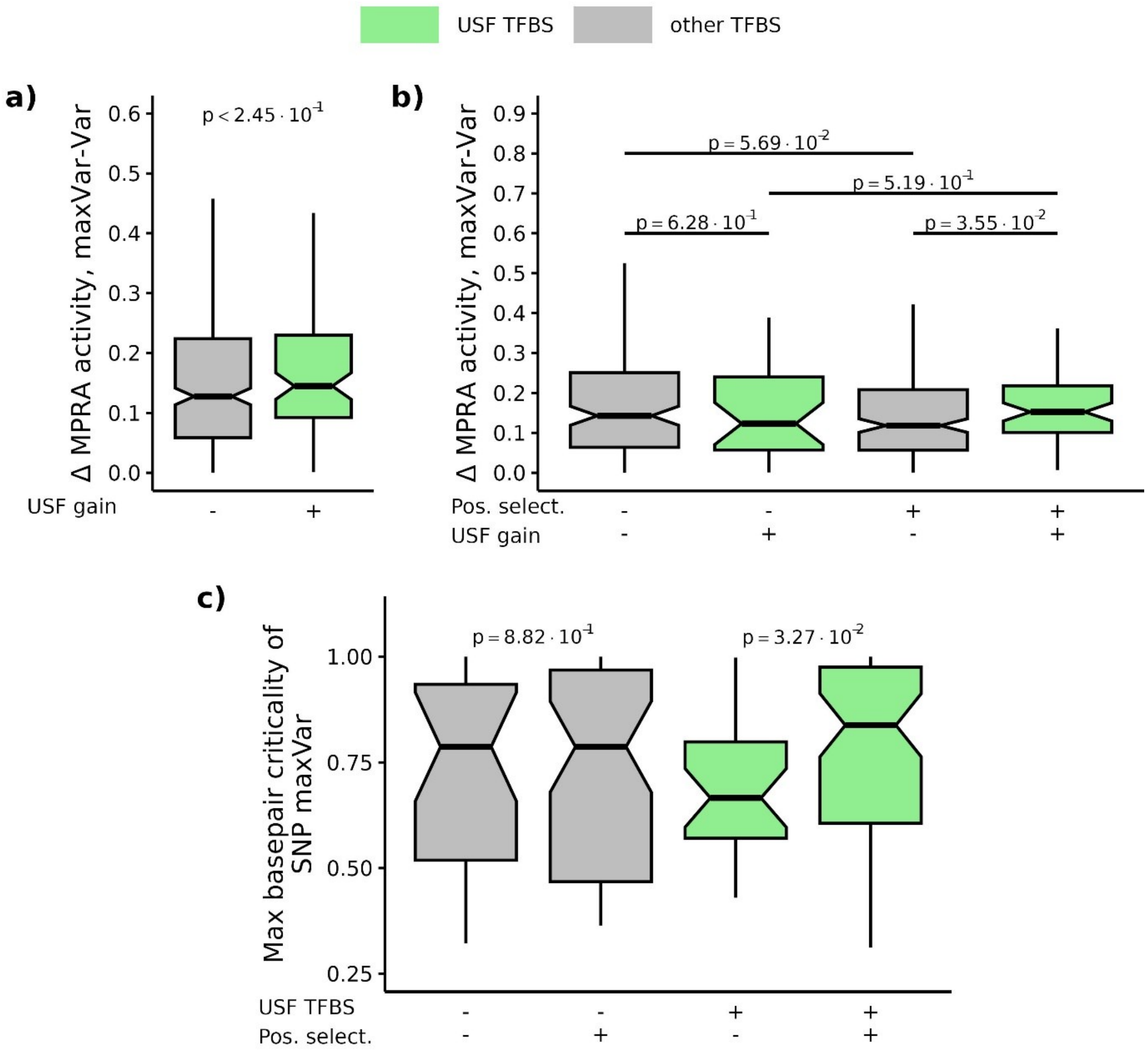


Figure 44. Positive selection of USFs binding sites in SNPs correlates with the gain of the MPRA activity. a) Comparison of the change in MPRA activity between the most active SNP variant ("maxVar") and other SNP variants ("Var"), analysed separately for changes that result and do not result in the gain of a USF binding site. b) Change in MPRA activity between the most active SNP variant and other SNP variants, depending on USF TFB gain and positive selection according to phyloP. c) Comparison of the importance of a nucleotide affected by SNP for TFBS ("basepair criticality") for non-USF and USF TFBS that are under positive or negative selection according to PhyloP score. Green - USF TFBS, grey – other TFBS.

To explore this question further, I assessed whether SNPs under different evolutionary pressures tend to affect more critical nucleotide positions within TFBS. I calculated the highest "basepair criticality" score (Figure 10) for each SNP variant and compared scores between sites under negative and positive selection, separately for USF and non-USF TFBS. Strikingly, while in non-USF TFBS the critical positions were not preferentially affected by either positive or negative selection (Figure 44c, $p = 8.82*10^{-1}$, two-sided Wilcoxon Rank Sum Test), USF

TFBS were more frequently positively selected for (Figure 44c, $p = 3.27*10^{-2}$, two-sided Wilcoxon Rank Sum Test).

In conclusion, in the analysed SNPs, USF binding sites tend to be under positive selection. In cases where USF TFBS arise as a result of positive selective pressure, such gains are associated with larger changes in enhancer activity than gains of non-USF TFBS. These findings reinforce the correlative observations from Ciuba et al., (2025) and the MPRA results presented previously. These observations provide more direct evidence of selective pressure favouring the gain of USF sites. Moreover, disruption of such selected sites through GWAS SNPs may contribute to altered enhancer function and, consequently, to human-specific health or disease traits.

4.10.2. SNP affecting the brain

To further investigate how the effects of SNP variants relate to phenotypic outcomes, I examined whether the groups of enhancers intersecting the SNPs show distinct patterns of evolutionary pressure. I compared phyloP scores between SNPs tested using MPRA that were associated with intellectual ability ("I"), neuropsychiatric disorders ("D"), and morphological features of the brain ("M"), and those linked to other phenotypes. The SNPs related to brain-associated traits feature lower phyloP scores and, therefore, are more evolutionarily variable than non-brain-related SNPs (Figure 45a).

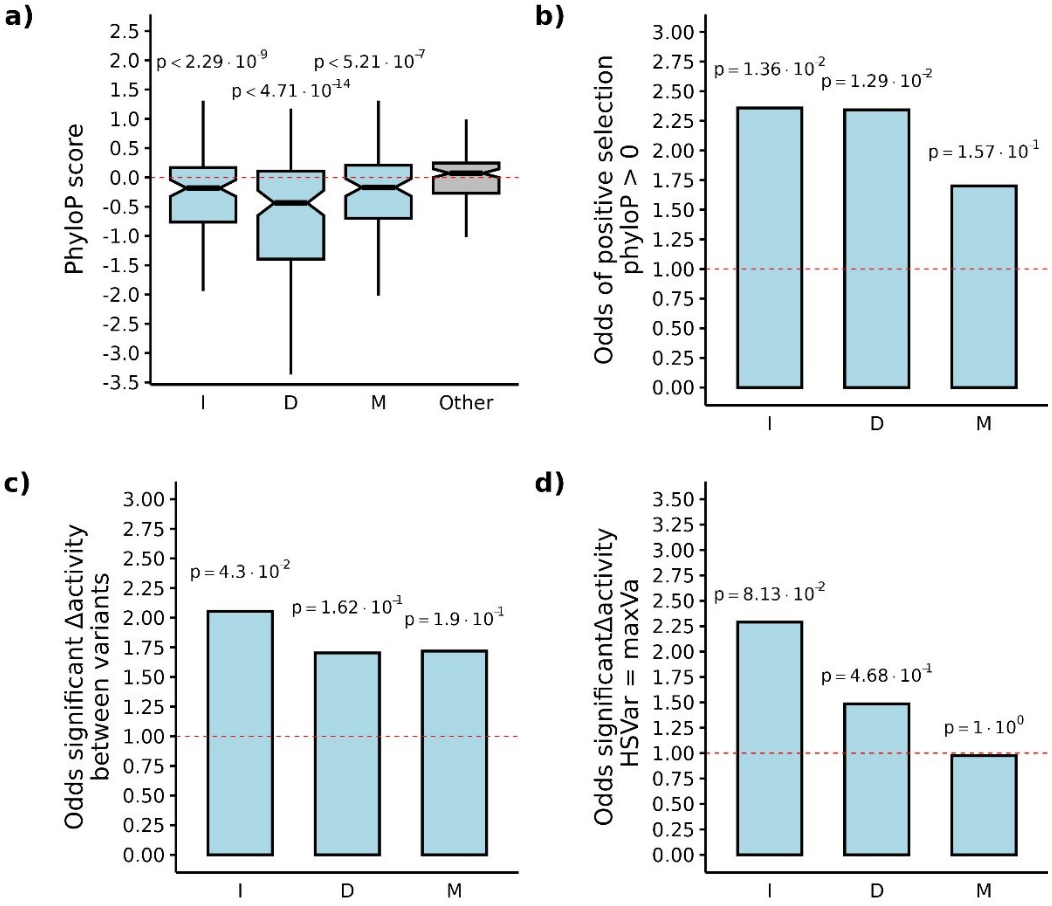


Figure 45. Analysis of SNPs depending on associated phenotype. a) PhyloP scores for different groups of SNPs tested using MPRA. SNPs related to intellectual ability ("I"), neuropsychiatric disorders ("D"), and morphological features of the brain ("M") and other phenotypes ("Other"). P values shown are relative to the "Other" category, determined using two-sided Wilcoxon Rank Sum Test. b) Odds of SNPs from different groups being under positive selection according to PhyloP scores. c) Odds of a significant change in activity between variants for SNPs from different groups. d) Odds that the human consensus variant is the most active variant for SNPs from different groups.

Moreover, SNPs associated with intellectual ability and those linked to neuropsychiatric disorders were significantly more likely to have undergone positive selection compared to non-brain-related SNPs (Figure 45a, odds ratios = 2.36 and 2.34, respectively, p = $1.36*10^{-2}$ and $1.29*10^{-2}$, respectively, Fisher's exact test). SNPs associated with intellectual ability were also more likely to have significantly different activity between variants (Figure 45c, odds ratio = 2.05, p = $4.3*10^{-2}$, Fisher's exact test). Notably, SNPs associated with intellectual ability are the only group that is enriched in polymorphisms that differ significantly in MPRA activity between variants (Figure 45c). Furthermore, only for SNPs associated with intellectual ability, in cases where the human consensus sequence variant ("HSVar") is also the most active variant ("MaxVar"), the difference from the other variant is more frequently statistically significant (Figure 45d, odds ratio = 2.05, p = $4.3*10^{-2}$, Fisher's exact test). These data suggest that enhancers harbouring SNPs impacting higher-level brain functions are particularly sensitive to sequence variation.

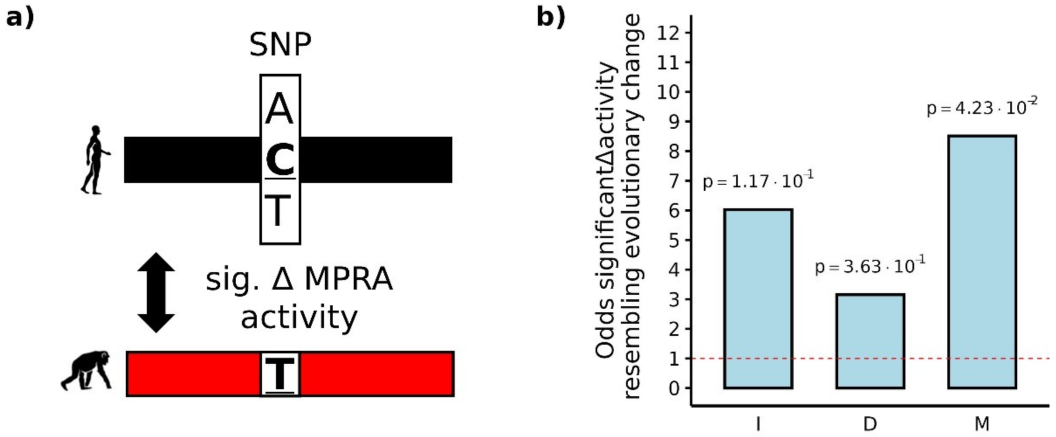


Figure 46. SNPs recreating evolutionary change. a) Visualisation of the concept: HS and PT consensus sequence variants at the site of the SNP differ (in this case, C is the consensus variant for HS and T is the consensus variant for PT). One of the possible HS variants (in the illustration, T) corresponds to the PT sequence. There is a significant difference in MPRA activity between CREs containing nucleotide variants corresponding to the consensus HS and PT sequences. b) Odds of significant change in activity between variants that resemble evolutionary change (the scenario described in panel a), for SNPs from different phenotype groups.

Some of the SNP variants tested in the MPRA effectively "recreate" evolutionary changes. In such cases, one variant matches the human consensus, while the other one corresponds to the chimpanzee consensus sequence. Such cases allow for direct comparison of evolutionary sequence modifications and their regulatory consequences. To identify important instances of such "evolutionary recapitulation," I focused on SNPs where the human consensus nucleotide differs from the chimpanzee consensus, and where the activity difference between the human-like and chimpanzee-like variants is significant (Figure 46a). Among the three brain-related phenotype groups, SNPs associated with brain morphology most frequently differed in activity between the human-like and chimpanzee-like variants (Figure 46b, odds ratio = 8.51, p = $4.2*10^{-2}$, Fisher's exact test). This suggests that, for these loci, the human variant may confer a regulatory state - either elevated or reduced enhancer activity - that is evolutionarily favored.

A number of SNPs identified in this study appear to recapitulate evolutionary changes. One particularly interesting example is rs10439607, which has been associated with variations in the volume of the pallidum, putamen, and caudate nucleus [489]. The G variant is present in most assayed humans, the variant A is the consensus in NHPs. In humans, the A variant of this SNP correlates with a reduction in the size of all of those brain areas.

Rs10439607 is located in intron of the BCL2L1 gene, which encodes Bcl-xL, which is a regulator of autophagy and an anti-apoptotic factor, protecting cells from oxidative stress [490]. Bcl-xL is essential for the survival of postnatal neurons[491], and is downregulated in response to increased β-amyloid levels. Overexpression of Bcl-xL in astrocyte-neuron co-cultures has been shown to reduce oxidative stress injury to neurons[492]. Mitochondrial Bcl-xL play a role in synaptogenesis by regulating autophagy[493]. Bcl-xL KO in mice leads to increased apoptosis in the neuronal population controlling higher-order brain functions[494]. Collectively, these findings indicate that Bcl-xL levels - including those produced by astrocytes - play a crucial role in neuronal survival, proliferation, and overall brain function. Fascinatingly, fluctuations in BCL2L1 expression have also been linked to seasonal changes in brain size in shrews[495], further emphasizing its evolutionary and physiological significance.

In MPRA experiments, the consensus human G variant exhibits significantly higher activity than the A variant (Figure 47a and b). Moreover, ATAC-seq and H3K27ac ChIP-seq profiles at the locus containing rs10439607 show higher chromatin accessibility and enhancer-associated histone acetylation in human relative to chimpanzee (Figure 47c and d). These findings suggest that the enhancer harbouring rs10439607 is more active in its human-like variant and may modulate the expression of BCL2L1.

Another notable example of an evolutionarily modified SNP is rs3770611, located in an intron of the LRP2 gene (Figure R51). This SNP has been associated with educational attainment in a study performed by Okbay et al., (2022) on more data from 3 million individuals. The A variant, which correlates positively with educational attainment, is also the consensus allele in human, whereas the G variant represents the consensus in all assayed NHPs.

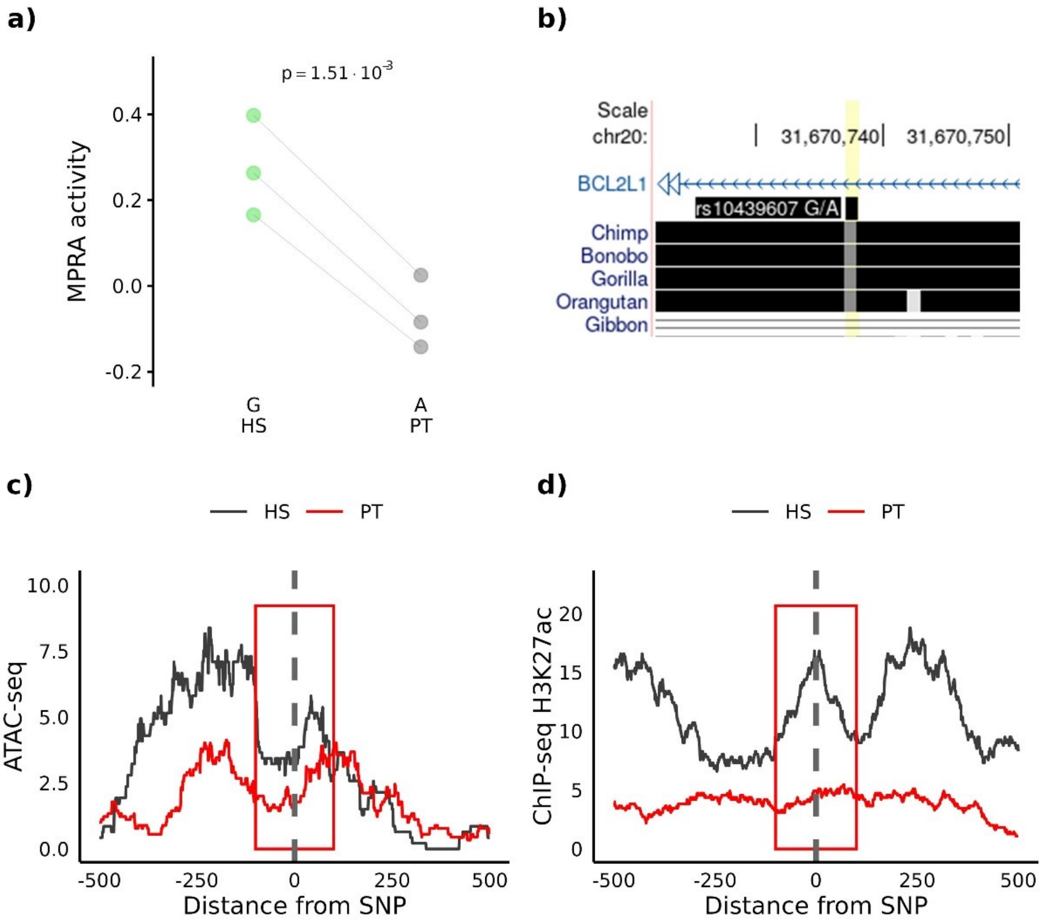


Figure 47. Characterisation of SNP evolutionary change of a SNP related to brain morphology, rs10439607. a) MPRA activity of the G variant, which is the consensus variant in HG38 HS genome, and the A variant, which is the consensus variant in PT6 genome assembly. b) UCSC genome browser snapshot of the region surrounding SNP rs10439607. c) ATAC-seq and d) ChIP-seq H3K27ac profile of HS and PT surrounding SNP rs10439607. Location of SNP indicated with a grey line. Location of the region assayed using MPRA is indicated by red frame.

rs3770611 is in proximity of, which LRP2 encodes a low-density lipoprotein receptor-related protein 2 (megalin), a large cell-surface glycoprotein[497], which is crucial for proper forebrain development[498–501]. Polymorphisms in the coding sequence of LRP2 have been associated with intellectual disability (Beydoun et al., 2012; Vasli et al., 2016). LRP2 is expressed in astrocytes and is involved in endocytosis, including uptake of Amyloid β[502] and albumin[503]. Thus, rs3770611 may represent another case in which the less active variant recapitulates the suboptimal regulatory state observed in chimpanzee.

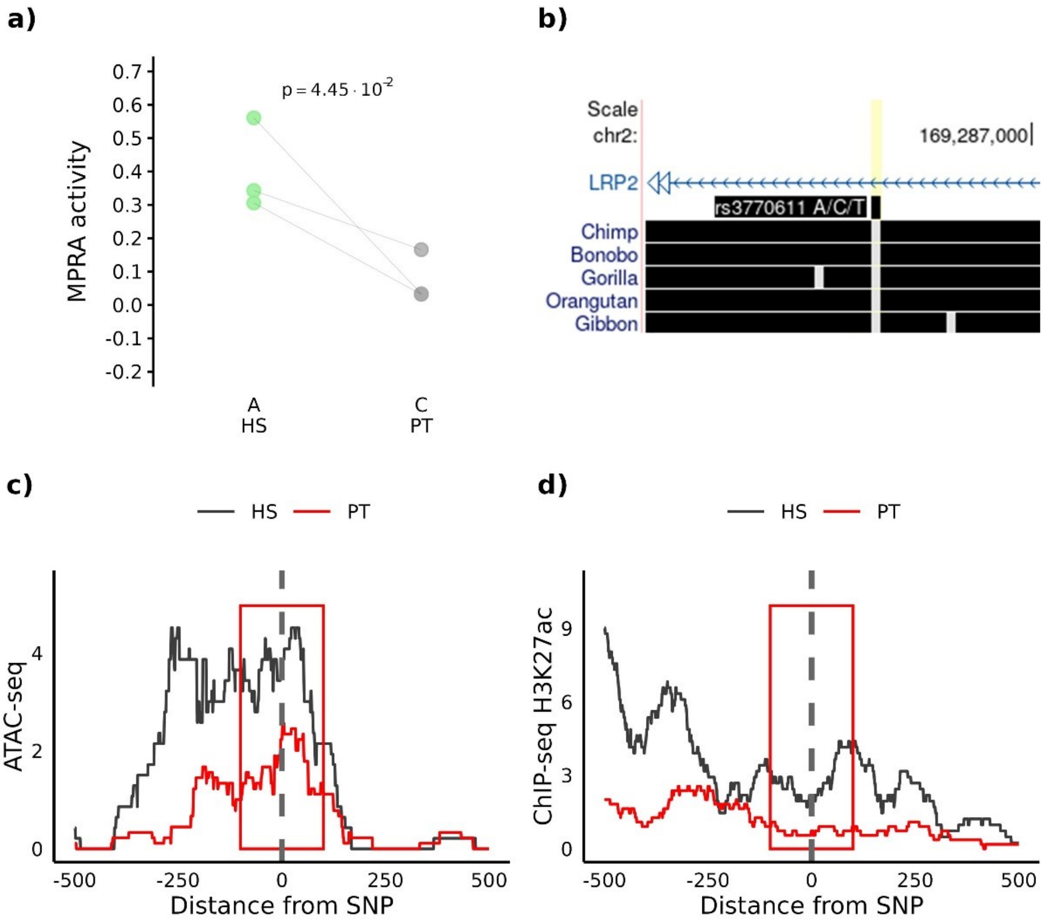


Figure 48. Characterization of SNP evolutionary change of a SNP related to intellectual ability, rs3770611. a) MPRA activity of the A variant, which is the consensus variant in the HG38 HS genome, and the C variant, which is the consensus variant in the PT6 genome assembly. b) UCSC genome browser snapshot of the region surrounding SNP rs3770611. c) ATAC-seq and d) ChIP-seq H3K27ac profile of HS and PT surrounding SNP rs3770611. Location of SNP indicated with a grey line. Location of the region assayed using MPRA is indicated by red frame.

Another example of a disease-associated SNP that underwent evolutionary change between human and NHP is rs439401, located near the APOE gene, one of the key loci implicated in AD[504]. The T variant, which is present in the human consensus genome, is positively associated with AD risk (Figure 49).

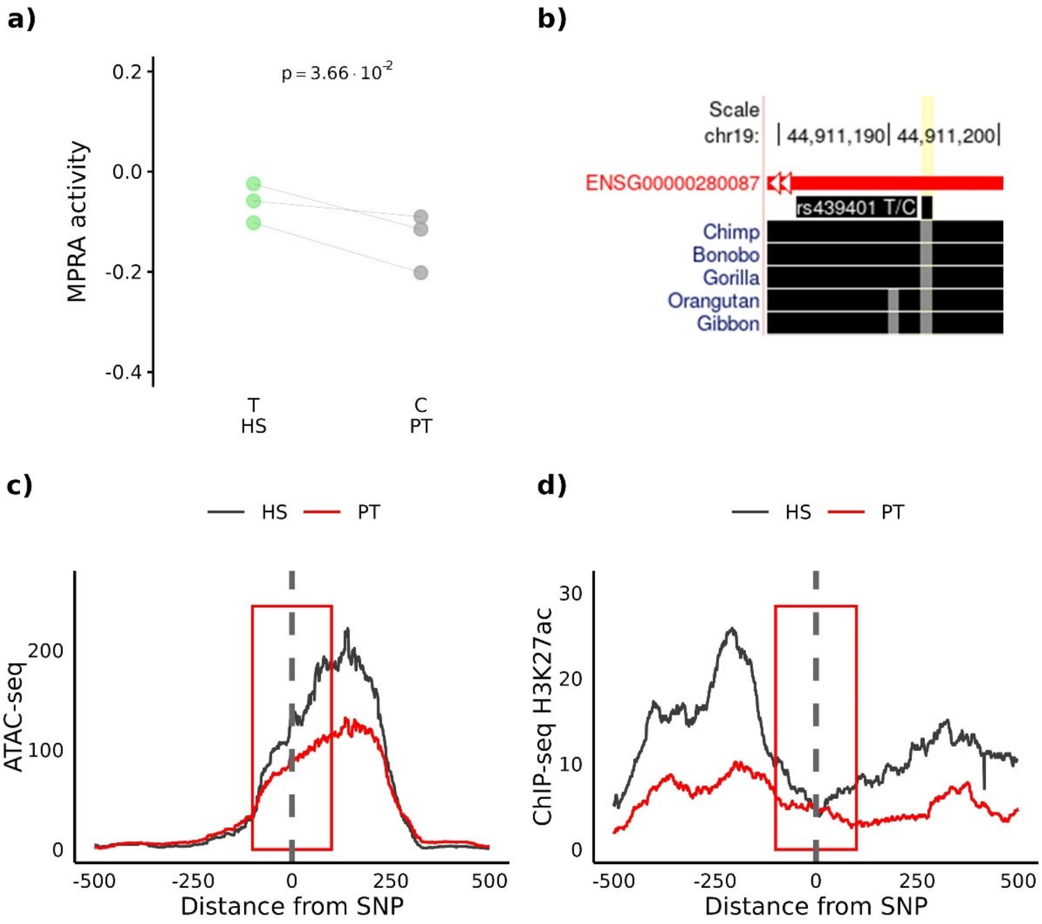


Figure 49. Characterization of SNP evolutionary change of a SNP related to brain morphology, rs439401. a) MPRA activity of T variant, which is the consensus variant in the HG38 HS genome, and C variant, which is the consensus variant in the PT6 genome assembly. b) UCSC genome browser snapshot of the region surrounding SNP rs439401. c) ATAC-seq and d) ChIP-seq H3K27ac profile of HS and PT surrounding SNP rs439401. Location of SNP indicated with a grey line. Location of the region assayed using MPRA is indicated by a red frame.

4.10.3. rs2473351 – CRISPRi perturbation

To verify whether an enhancer showing MPRA activity differences due to a SNP also affects endogenous gene regulation, I performed CRISPRi perturbation of the evolutionarily altered SNP rs2473351, which is associated with educational attainment. The more active T variant of this SNP, which is also the consensus allele in the human genome assembly, is positively associated with educational attainment[496], whereas the C variant, which is less active according to MPRA, is common among NHPs, except in orangutans, where the reference genome contains a G at this position (Figure 50).

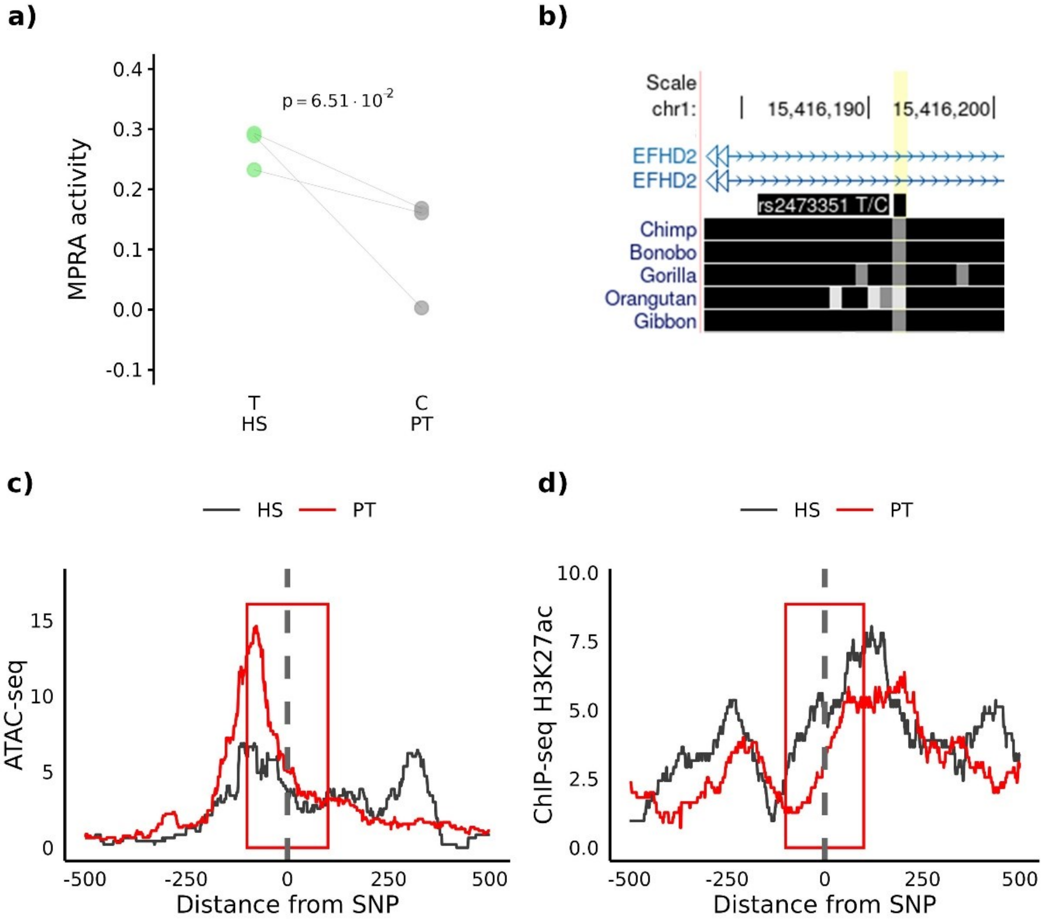


Figure 50. Characterization of SNP evolutionary change of a SNP related to educational attainment, rs2473351 a) MPRA activity of the T variant, which is the consensus variant in HG38 HS genome, and C variant, which is the consensus variant in PT6 genome assembly. b) UCSC genome browser snapshot of the region surrounding SNP rs2473351. c) ATAC-seq and d) ChIP-seq H3K27ac profile of HS and PT surrounding SNP rs2473351. Location of SNP indicated with a grey line. Location of the region assayed using MPRA is indicated by a red frame.

Very interestingly, EFHD2 was found to be associated with behavioural effects in mice, namely the KO effects in increased sensation-seeking and addictive behaviours[505] The same study has also shown that EFhd2 KO mice have reduced volumes of sensorimotor cortex and prefrontal cortex. Summary of the SNP is presented in figure 50.

rs2473351 is located in the first intron of EF-hand domain-containing protein D2 (EFHD2, also called Swiprosin-1,), a gene that encodes a calcium-binding protein[506]. Analysis of EFHD2 interactome from the mouse brain showed that it has region-specific roles in the brain and is potentially associated with multiple cellular and molecular processes[507], including cytoskeletal regulation, vesicle trafficking, and stress response. Efhd2 knockout (KO) mice show that the gene is required for adult neurogenesis, and negatively regulates neurite complexity[508]. Furthermore, Efhd2 KO mice exhibit increased sensation-seeking and addictive behaviours, along with reduced sensorimotor and prefrontal cortex volumes[505]. Efhd2 KO also leads to increased levels of pathological tau protein in the hippocampus, consistent with Efhd2's known association with tau in a mouse model[509], and its co-localization with pathological tau in AD brains[510]. In humans, EFHD2 expression is upregulated in SCZ, as shown by proteomic analyses[511]. EFHD2 is highly expressed in human foetal astrocytes, and the activity of this locus is markedly downregulated in adult astrocytes[312]. Notably, EFHD2 is more highly expressed in human astrocytes than in mouse astrocytes (Zhang et al., 2014, 2016b, data from https://brainrnaseq.org/), indicating that the gene is both dosage-sensitive and upregulated in the human lineage. In summary, EFHD2 is related to the regulation of brain functions.

To test the effect of the enhancer overlapping SNP rs2473351 I performed perturbation using the CRISPRi system, delivered using two types of viral particles, one containing repressor machinery (dCas9-KRAB-MeCP2) and another containing a plasmid carrying two sgRNAs targeting the enhancer or non-targeting control (Figure 51a). I performed the perturbation in two iAstrocyte lines (ELE30 and AG94) and in duplicate for each line (total 4 samples per condition). Cell cultures were incubated with antibiotics to select only double transduced cells. I then performed bulk RNA-seq to characterise the effects of the perturbation. However the effect on gene expression differed markedly by cell line (Figure R5b and c). (I found that CRISPRi silencing of the enhancer resulted in significant alteration of expression of more than 1000 genes (p-adj. < 0.05, analysed using the DESeq2 package). 810 genes were downregulated (log fold change vs scrambled <0), and 316 were upregulated (log fold change vs scrambled >0)

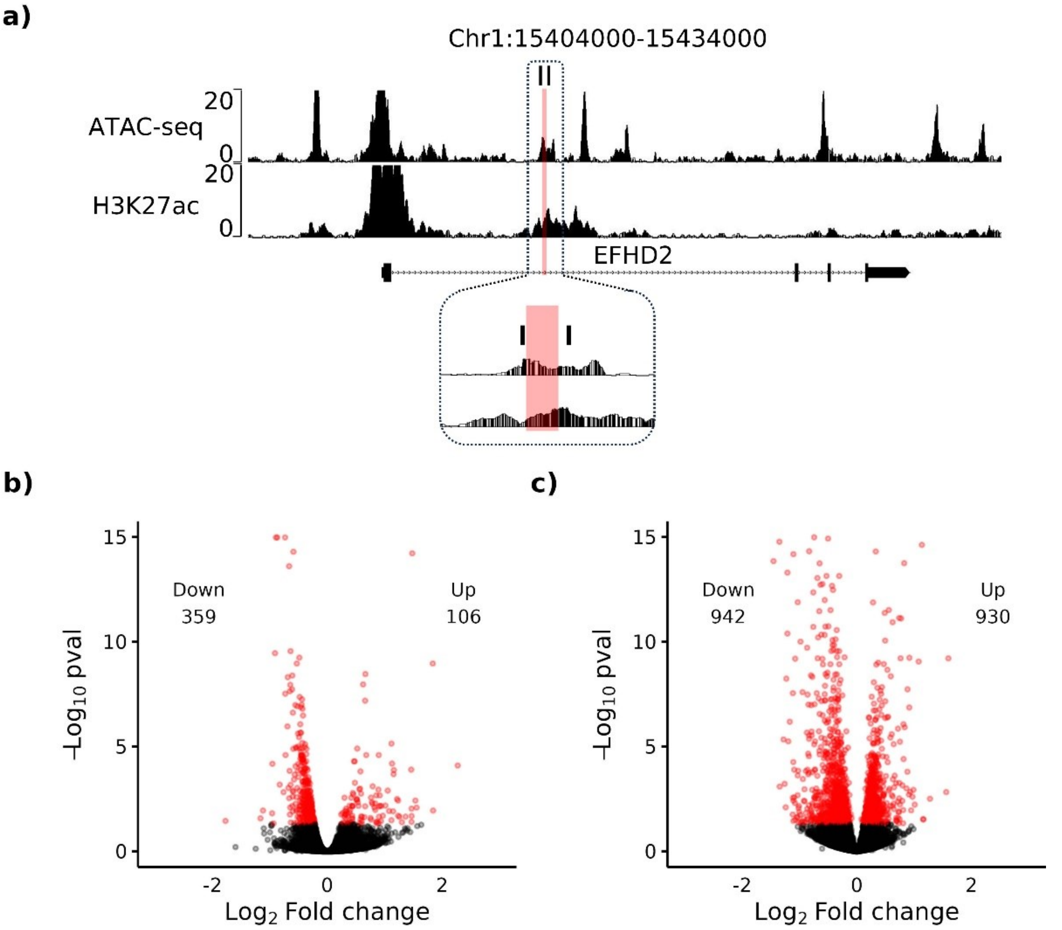


Figure 51. Perturbation of the enhancer intersecting SNP rs2473351 located in the first intron of *EFHD2*. a) Location of the targeted enhancer, ATAC-seq and ChIP H3K27ac tracks (TPM-normalised values); sgRNAs used for perturbation (two vertical black lines) and region tested using MPRA are highlighted in red. A zoomed-in view of the region is presented in the box below. SNP rs2473351 is located in the centre of the MPRA fragment, or in the middle of the region shaded in red. b) Volcano plot of differentially expressed genes from the experiment performed in ELE30 line. Genes with significantly altered expression are marked in red (p-adj. < 0.05, analysed using the DESeq2 package). c) Volcano plot of differentially expressed genes from the experiment performed in AG94 line. Genes with significantly altered expression are marked in red (p-adj. < 0.05, analysed using the DESeq2 package).

To determine the global effects of enhancer perturbation, I performed GO term enrichment analysis[513,514], on 91 genes which are congruently downregulated in experiments performed in both cell lines.

Biological processes affected by genes downregulated by the enhancer perturbation are predominantly associated with extracellular environment composition and organization (GO terms: “Extracellular matrix organization”, “Extracellular encapsulating structure organization”, “Extracellular structure organization”) (Figure 52).

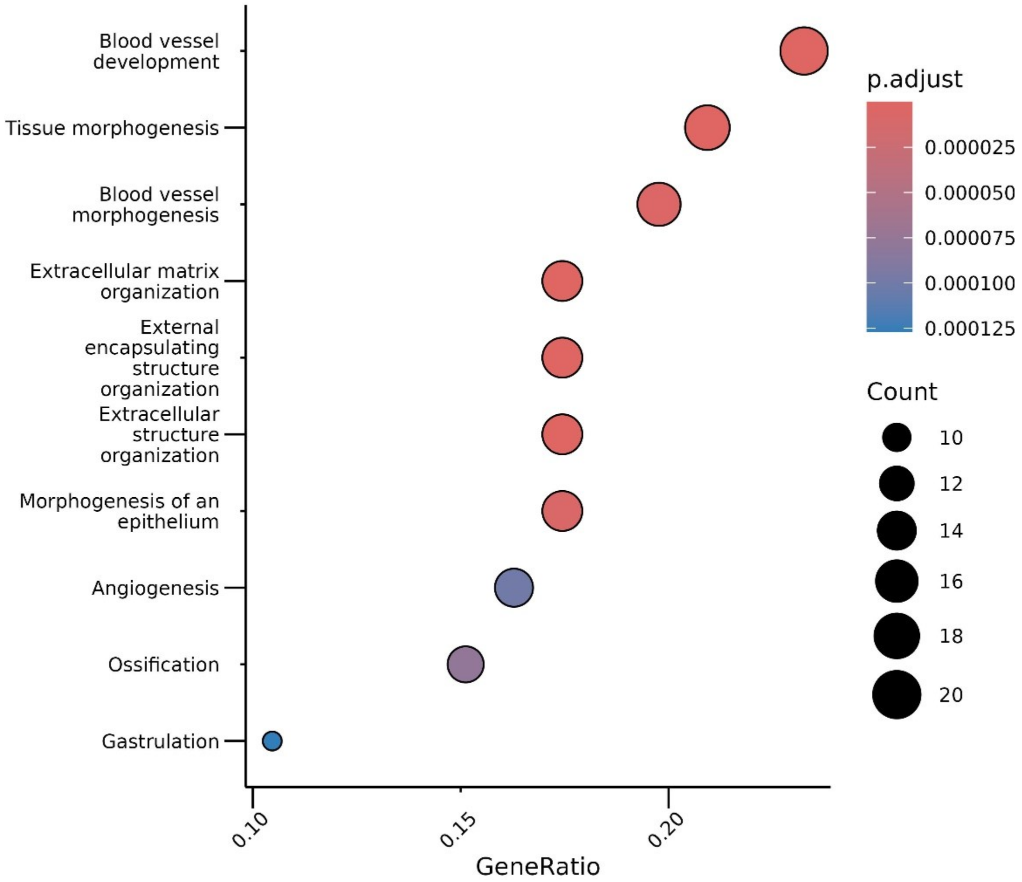


Figure 52. Functional associations of genes downregulated upon the silencing of the enhancer intersecting rs2473351. Enrichment in GO terms for biological processes significantly associated with downregulated genes.10 most strongly enriched terms are shown. p< 0.05, q-Value <0.05. b).

The GO terms that are affected by perturbation of the enhancer containing rs2473351 are consistent with previously described functions of EFHD2. An interactome analysis[507] demonstrated that EFHD2 interacts with components of the cytoskeleton. However, due to the great variability of effect in different cell lines, this result warrants further investigation.

5. Summary

Understanding how gene regulation shapes cellular identity and complex brain function is one of the most compelling questions at the intersection of neuroscience and genomics. Studying the regulome can reveal mechanisms that link noncoding DNA variation to the molecular and cellular processes underlying human cognition and neurological disease. Despite significant progress in mapping regulatory landscapes in brain cells, most of the studies focus on neurons (for example Song et al., (2019); Boulting et al., (2021); Tsitkov et al., (2024)), leaving the astrocyte regulome relatively unexplored. However, due to the crucial role of astrocytes in maintaining brain homeostasis and their involvement in neurological health and disease, a better understanding of their regulatory mechanisms is essential for clarifying their contributions to overall brain function.

This knowledge gap is especially apparent when studies of evolutionary changes in the brain regulome are considered. While the importance of changes in the regulome as the driver of evolutionary changes has long been recognised[265,266], most comparative studies on transcriptional and regulatory divergence between primate brains have focused predominantly on neurons[246–249,517]. Together with my colleagues, we aimed to address that gap in our recently published study, which we based on the iPSC-derived astrocytes (iAstrocytes) from human, chimpanzee, and rhesus macaque[286].

We employed an *in vitro* model that has the transcriptional signature of foetal astrocytes from the cerebral cortex at the 24th gestational week. In this period of brain development, the most intense gliogenesis occurs[518], and the formation of crucial connections between neurons commences[519–521]. Importantly, foetal astrocytes have a regulome distinct from other cell types[345,522] that likely shapes their functions during this critical developmental period. Studies also indicate that foetal astrocytes may play an especially important role in brain folding[523,524].

We characterised the transcriptomic and regulatory landscape of iAstrocytes using a combination of complementary genomic approaches. We employed RNA-seq to define the set of actively expressed genes, and ATAC-seq to identify regions of potentially active chromatin. To distinguish between promoter-and enhancer-associated regulatory elements, we performed ChIP-seq for histone modifications H3K27ac and H3K4me3, respectively. Finally, we leveraged the chromatin conformation map, generated using the most up-to-date protocol (Intact Hi-C), to integrate this data and infer genes possibly affected by the identified enhancers. To the best of my knowledge, such a comprehensive characterisation of the regulome in an *in vitro* foetal astrocyte model has not been performed previously. I validated this analysis by experimentally evaluating the regulatory activity of selected enhancer regions using a massively parallel reporter assay (MPRA).

Based on the genomic characterisation data, I detect the previously underappreciated aspects of astrocyte regulome. I show that human sequence variants associated with brain morphology are significantly enriched within human-specific regulatory regions.

Subsequently, I demonstrate that foetal astrocyte regulome, and particularly its active enhancers, is enriched for genomic changes which are human-specific relative to other primates, so-called human accelerated regions (HARs[293]), and single nucleotides that distinguish humans from other hominids (SNCs[422]). Fascinatingly, enhancers affected by greater evolutionary pressure vis-à-vis non-human primates appear to share chromatin interaction with genes affecting processes related to brain development.

I then place those findings in an evolutionary framework. I compare the human foetal astrocyte regulome with those of the most closely related species, chimpanzees and macaques. I take advantage of the unique iPSC-derived model of foetal iAstrocytes from those species, developed by my colleagues[286]. Through this comparative approach, in the published study, we identified a subset of enhancers that differ both in their chromatin accessibility (“openness”) and underlying DNA sequence between species. Notably, we found that enhancers associated with genes more highly expressed in human iAstrocytes were significantly enriched for binding motifs of the Universal Stripe Factors (USFs), a recently characterised family of transcription factors with broad regulatory influence[1]. This suggests that changes in USF-mediated enhancer regulation have contributed to the evolution of human-specific astrocyte gene expression profiles.

Using MPRA, I experimentally show that the number of predicted USF binding sites correlates most strongly with enhancer activity compared with other epigenetic and sequence features. By comparing the activity of human enhancers with versions in which USF motifs were substituted to match the chimpanzee sequence, I demonstrated that USF sites are required to sustain enhancer activity. This is the first time the importance of this class of transcription factors has been soundly verified experimentally. As the enhancers tested are proximal to genes more highly expressed in human iAstrocytes relative to non-human primate (NHP) iAstrocytes, these results confirm the proposal that evolutionary gains of USF motifs underlie increased expression of human astrocyte genes.

To illustrate the regulatory relevance of activity of human-specific enhancers, which feature an evolutionary difference in binding sites of USFs, I employed a state-of-the-art CRISPRi system to perturb the epigenetic state of a human-specific enhancer upstream of the RhoBTB3 gene. Follow-up analysis of single human–chimpanzee sequence differences within this enhancer revealed that certain evolutionary combinations of sites are sub-optimal for enhancer function, highlighting the fine-tuned state of regulatory architecture.

Finally, I apply MPRA to test whether single-nucleotide polymorphisms (SNPs) within active human enhancers modulate their activity. I identify 90 SNPs for which at least one pair of variants differs significantly in enhancer strength, providing mechanistic insight into how noncoding variants can influence gene regulation. Notably, several of these SNPs recapitulate evolutionary changes between chimpanzee and human genomes, producing phenotypes with potential implications for human disease. I demonstrate such an effect using the example of an enhancer within the EFHD1 gene, in which a SNP variant is similar to a chimpanzee sequence. CRISPRi-mediated silencing of this enhancer reveals widespread transcriptional effects, underscoring how both subtle evolutionary modifications and intra-species regulatory variation can exert broad, cell-wide consequences on gene expression.

6. Discussion

6.1. Link between astrocyte regulome and brain structure

The importance of the astrocyte regulome in brain function and disease is an area of active investigation. One of the aspects of brain development that appears to be crucially affected by astrocytes is brain size and organisation. Crucially, astrocytes contribute to prenatal brain development. While an older hypothesis posited that brain gyrification is driven by the expansion of the upper-layer neuronal precursors[525], I argue that my observations, and more recent evidence based on studying the rhesus macaque brain[360], suggest that astrocytic expansion is central to the process of the formation of brain folds. In this work, I reveal a genetic link between brain shape and regulatory elements active in foetal astrocytes, pointing to an important role of these cells in shaping brain morphology.

Indeed, gyrencephalic species, those with more highly folded cortices, tend to have a higher number of astrocytes than lissencephalic animals[526]. In line with this, in the rhesus macaque, the extent of gyrification correlates with increased gliogenesis[360]. During the development of the neocortex, a specific type of glia, namely the radial glia (RG) cells[527,528], provides the scaffolding for neuronal migration from the proliferative zone near the cerebral ventricle [524]. Notably, the timing of the peak of neurogenesis is discordant with the fastest progress of gyrification, which occurs later in prenatal development[529], indicating that the sole numbers of neurons cannot explain brain morphology.

In humans, the most intense gliogenesis occurs during a protracted period extending from the second trimester until after the first year after birth[518]. This developmental time window is modelled by iAstrocytes generated by me and my colleagues[286], which makes this cell model ideally suited to draw conclusions on the importance of foetal astrocytes. Analyses of the regulome iAstrocytes show that the putative enhancer elements active in human iAstrocytes impact brain morphology. It is, however, not clear which genes specifically are regulated by these elements, and how exactly astrocytes modify brain structure. Shinmyo et al., (2022) demonstrated in mice that alteration of astrocytic proliferation, through manipulation of FGF signalling, can alter the formation of brain folds. Thus, there is a possibility, supported by experimental evidence, that astrocytic numbers (and possibly their size), are impacted by genetic variants detected by GWAS, which collectively shape brain morphology.

Notably, AUTS2 is one of the genes highly expressed in astrocytes that is located in the vicinity of a putative enhancer that intersects a GWAS polymorphism related to the regulation of the cortical surface. AUTS2 is related to autism and the regulation of brain size[256]. Such a regulatory element may contribute to the regulation of AUTS2 in the human lineage, thereby impacting brain morphology. Another gene, more highly expressed in the human compared to NHP iAstrocytes[286], Lectin galactoside-binding soluble 1 (LGALS1, also referred to as Galectin-1), which is more highly expressed in the human compared to NHP iAstrocytes, is located in the genomic vicinity of an active (H3K27ac-positive) enhancer that intersects a GWAS SNP related to the regulation of the cortical surface. LGALS1 impacts intercellular communication, and by regulating neural cell interactions with the intercellular milieu, it may impact proliferation[530,531]. Experimental modulation of the expression of these genes on astrocyte-neuron interaction and brain morphology in humans would be needed to confirm those hypotheses.

Brain relies on the intricate organisation of connection between neurons for its function. The totality of such connections is referred to as “connectome”. On of the crucial development periods for development of the connectome is the second trimester of pregnancy [519,520]. Our iAstrocytes have transcriptional signature similar to foetal astrocytes from this gestational period[286]. The process of formation of the connectome affects cognitive ability and is highly sensitive to disruption, for example, during preterm birth. A recent study of over 1.4 million individuals demonstrated that preterm birth predicts lower intelligence scores and academic achievement [363]. Furthermore, disruptions in brain folding have been linked to an increased risk of multiple neurological disorders[39,359,363–366]. Impairment of cortical folding due to premature birth results in behavioural abnormalities, including lowered attention span[356]. These manifest as early as term-equivalent age (the age of a premature child corresponding full-term birth). Longitudinal studies indicate a higher risk of ADHD, ASD, and anxiety disorders in prematurely born children[362]. Furthermore, structural variations in the brain are strongly associated with both Schizophrenia and ASD[523,532–535], which share a closely related aetiology[536,537], including abnormal cortical folding[538,539] and connectivity[540].

My findings support the importance of astrocytes for brain morphology. It is through the modulation of the physical structure of the brain that astrocytes may influence normal and pathological brain functions. I suggest that astrocytes, apart from their crucial role in direct molecular interactions with neurons - such as regulation of synaptic activity or cell signalling - modulate brain structure and organization.

Another aspect identified in this work is the enrichment of the astrocyte regulome in variants associated with water movement. It aligns with the described role of astrocytes in regulating fluid circulation, including their contribution to waste clearance[353]. Water transport networks in the brain are formed the early second half of gestation[541]. Our iAstrocytes are transcriptionally similar to foetal astrocytes from this gestational period[286]. Efficient cerebrospinal fluid clearance is critical for brain health and is disrupted in dementia[354], and is crucial for the clearance of amyloid β, α-synuclein, and tau associated with AD[143,146]. These findings suggest that foetal astrocyte regulome may affect mechanisms of brain fluid regulation in humans. Therefore, my work suggests that activity of enhancers during in foetal astrocytes is one of the factors regulating astrocyte functions and normal and pathological brain development.

6.2. Importance of specific transcription factors for astrocyte regulome.

In this work, I identified a set of TFs potentially important for astrocyte function. I derived this list by two complementary approaches: firstly, by estimating the enrichment of TFBS in astrocyte enhancers and assessing the potential effects of SNPs on these binding sites; and secondly, by determining which TFBS best correlate with enhancer activity measured experimentally using MPRA. The combination of those two methods of discovery provides an especially strong support for the role of certain TFs for gene regulation in astrocytes.

Both analyses indicate that active astrocytic enhancers are particularly enriched in c-Fos binding sites. While the function of FOS has been well recognised in neurons, the contribution of this TF to astrocyte biology only begins to emerge. Previous genomic analysis of primary astrocyte precursors through ATAC-seq and ChIP-seq, coupled with promoter capture Hi-C, revealed enrichment of FOS binding motifs within enhancers featuring H3K27ac and forming loops with promoters[345]. Furthermore, c-Fos, is instrumental for cognitive function. Astrocytic activation in the hippocampus correlates with the induction of c-FOS protein and is required for memory formation[225,226]. Furthermore, spatial transcriptomics in mice has revealed that c-Fos in astrocytes expression is associated with engram formation[227]. A very recent study[402] has demonstrated that c-Fos expression in astrocytes is regulated via a mechanism specific to the cell type, distinct from neurons, and that astrocytes exhibit learning-dependent increases in adrenoreceptor expression, which mediate c-Fos activation. These results open an intriguing possibility that c-Fos could act over extended periods of time, beyond the framework of the immediate early gene activation pathway. Furthermore, since Nott et al., (2019), showed enrichment of SNPs related to character traits, including intelligence, in astrocytic enhancers, intellectual ability may be significantly affected by specific TFBS, such as FOS, in the astrocyte regulome.

In this work, I show that intelligence-related SNPs preferentially affect binding sites of FOSL1 and FOSL2, presenting a potential link between enhancers targeted by FOS in astrocytes and human cognitive capacity. Further molecular biology experiments should allow us to determine which genes rely on FOS activity in enhancers in the steady state of astrocytes, and whether FOS-enriched enhancers respond to unique cues, and how these responses are mechanistically linked to astrocytic support of higher-level brain functions.

My analysis of TFBS most frequently affected by SNPs related to brain morphology indicates that retinoic acid (RA) receptor binding to astrocyte enhancers may be important in determining brain surface area and folding. This observation agrees with the well-established role of RA in brain development, particularly in the prefrontal cortex[542,543]. Likewise, variations in RA levels have been associated with neurodevelopmental disorders, intellectual disability, and neurodegenerative diseases[398,399,401], and are known to affect glutamate transport, a critical function of astrocytes[400]. My work suggests that the astrocytic regulome may respond to RA in the developing brain.

Another important group of transcription factors (TFs), which is especially frequently affected by polymorphisms related to brain morphology in the human-specific regulome is the forkhead box (FOX) family of TFs. For instance, FOXP1, which according to my estimates may be affected by significantly enriched among factors affected by SNPs related to brain morphology, has been shown to play key roles in neuronal morphogenesis and migration during cortical development[544]. FOXP1 has also been implicated in the generation of RG, the main precursors of astrocytes[545,546]. In addition, FOXP1 impacts neuronal dendritic tree growth [547]. The literature on role of other forkhead factors in astrocytes is scarce. The potential involvement of this TF class in variants affecting astrocyte regulome-related brain morphology represents an intriguing avenue for future research.

In the MPRA assay, a notable group of TFs correlating with enhancer activity is those recognising cAMP-response elements (CRE): CREB1, CREB3, ATF1, and ATF3. CRE elements have been investigated by studying CREB1 and responses to cyclic AMP (cAMP). In astrocytes, cAMP signalling has been shown to enhance synaptic plasticity and memory formation[548], as well as the neuron-lactate shuttle, a crucial pathway for the provision of energy to neurons[549]. CREB1 regulates processes that are essential for astrocytes, such as glutamate transporter 1 (GLT-1) expression[550] and glycogen synthesis[551]. Notably, another CRE-binding TF, ATF3 is a critical regulator of astrocyte differentiation[342].

In summary, the binding of transcription factors to CRE has been long established as an important pathway for learning and memory[552–554]. Most of the studies on CRE show its function indirectly, and I believe it would be particularly interesting to further investigate the role of CRE and proteins that bind it specifically in astrocytes.

Finally, the precise molecular processes underlying memory storage remain largely unknown. The simplest explanation for the function of TFs from the FOS family and CRE-binding factors in memory is that they are activated by expression of the genes that drive molecular processes leading to memory formation. However, it is tempting to speculate that these factors may themselves be the drivers of yet unidentified pathways that are crucial for memory formation. Regardless, the role of FOS, JUN, and CRE-binding TFs in astrocytes represents a compelling avenue for future research.

6.3. Evolutionary changes affecting astrocyte regulome

This work aims to address the existing gap in research on evolution. Comparative studies of brain regulome focus on different biological materials and methods used, but invariably focus on neurons, leaving the problem of human astrocyte evolution largely unaddressed.

For example, a 2019 study by Kanton et al. compared the genomic and transcriptional landscapes of human and chimpanzee organoids using single-cell transcriptomics and genomics, focusing on neurons. Another transcriptomic study[249], focusing on neuronal subtypes, analysed dissected brain tissue and reported human-specific changes in expression of *FOXP2*, a gene associated with multiple neurological diseases.

Among those studies Khrameeva et al., (2020), have shown that human astrocytes exhibit the most divergent-specific gene expression signature across the compared brain regions and cell types based on analysis of acutely isolated brain slices from humans, chimpanzees, bonobos, and macaques. Jorstad et al., (2023) noted a high number of human-specific genes expressed in astrocytes, including synaptic components (e.g., ephrins, neuroligins) and extracellular matrix proteins supplied by astrocytes (e.g., hevin). In contrast, such enrichment was not observed in our study[286], which may reflect the fact that the expression of these genes could be expressed later in astrocyte development and also induced by neuronal signals absent in our *in vitro* model. A study Agoglia et al., (2021) employed a complex human–chimpanzee cell fusion model to explore regulatory network changes and highlighted a human-specific regulatory module including astrocyte-associated genes (*AQP4*, *HEPACAM*, *CLU*, *HEPN1*, *S100β*) as especially divergent. However, the complexity of this model makes it challenging to interpret these findings relative to studies performed on actual human and chimpanzee cells.

In this work, I identified signatures of positive evolutionary pressure shaping the human astrocyte regulome. This includes both changes relative to non-human primates (as indicated by the enrichment of active enhancers in human accelerated regions, HARs) and putative changes relative to other hominids (as suggested by enrichment of active enhancers in single-nucleotide changes, SNCs).

Genes can be influenced by enhancers located within the same topologically associated domain[283,556–560]. In this study, I am using this well-described phenomenon as a basis for determining genes potentially affected by evolutionary pressures specific to the human lineage. Previously reported study, which analysed association of HARs with astrocytic genes[248], based the associations on genomic proximity, whereas my analysis employs state-of-the-art Intact Hi-C chromatin contact maps for more refined enhancer–gene assignments.

It is very interesting to hypothesize that the association of evolutionary changes in the regulome with genes related to brain morphology may be specific to foetal astrocytes, which play a critical role during brain development. Indeed, human-specific shifts in the timing of prenatal brain development are particularly pronounced, even when compared to other hominids[561]. However, to further explore whether the results presented here are specific to foetal astrocytes, it would be informative to repeat the analysis using iAstrocytes differentiated toward a more mature state, for example, through co-culture with neurons.

6.4. General determinants of enhancer activity identified using Massively Parallel Reporter Assay

The results presented in this work allow us to survey the usefulness of sequence features other than TFBS, as well as genomic characterisations, for the prediction of activity of enhancers. This problem has been partially explored in previous studies, which explored the relationship between ATAC-seq, H3K27ac, and MPRA signals.

I found that the number of CG dinucleotides within enhancers correlates with MPRA activity, and the number of AT dinucleotides anticorrelates with it. Notably, Maricque et al., (2017) demonstrated that GC content correlates with higher enhancer activity in MPRA, and, interestingly, that the effect of dinucleotide distribution varies depending on the cell type in which the assay is performed. Since the exact dinucleotide distribution is correlated with a particular arrangement of nucleotides within TFBS, this trend reflects the presence and activity of different TFs in each cell type.

Guanine-rich stretches of DNA form tertiary structures termed G4 quadruplexes; their presence correlates with regulatory element openness and activity[563,564]. While MPRA studies have investigated the role of G-quadruplex (G4q) structures in promoters[565], no study has yet examined their effect on enhancer activity. Remarkably, inserting G4 quadruplexes distal to promoters has been shown to induce the formation of new enhancers[566], indicating the possible relevance of G4q for enhancer activity. I found that the presence of G4q appears to correlate with MPRA activity. Yet, how G4q formation impacts enhancer functions awaits further investigation.

In this work, I show that the H3K27ac signal is more correlated with the magnitude of MPRA activity, whereas a higher ATAC-seq signal primarily correlates with the binary probability of a CRE being active. Despite differences in experimental design and the set of enhancers studied, these findings are largely in agreement with Inoue et al. (2019) and Maricque et al., (2017). Previously, Inoue et al., (2019) tested approximately 3,000 candidate regulatory elements (CREs) using a lentiMPRA workflow in differentiating NP cells, and reported that the H3K27ac signal correlated with enhancer activity and transcriptional state more strongly than the ATAC-seq signal. Mauduit et al., (2021) analysed tiled fragments of H3K27ac and ATAC-seq peaks from regulatory elements in melanoma, and showed that, while active enhancers generally come from within ATAC-seq peaks, the mean signal of a given peak is not a good predictor of MPRA enhancer activity. ATAC-STARR-seq was described and applied by Hansen and Hodges, (2022) to directly interrogate the enhancer activity of fragments generated by ATAC-seq tagmentation. The study demonstrated that MPRA-like activity is strongest for sequences located at the centre of ATAC-seq peaks and correlates with H3K27ac signal. It is, however, important to note that STARR-seq is discordant with other reporter-based assays[428].

In conclusion, the relationships between MPRA signal, ATAC-seq accessibility, and H3K27ac enrichment observed in this work broadly align with previous studies, reinforcing the notion that enhancer activity is most strongly associated with open chromatin regions characterised by histone acetylation.

6.5. Universal Stripe Factors shaping evolutionary changes in astrocyte regulome

The question of how enhancers acquire their cis-regulatory potential in evolution is of fundamental significance for molecular biology. Single-enhancer changes have been shown to critically impact complex traits[415,416], but the broader mechanisms driving the gain of transactivatory potential in newly evolved enhancers remain largely underexplored[261,279]. Global analyses of pooled brain ChIP-seq data indicate that TFBS are under human-specific evolutionary pressure[569]; however, these studies have not identified specific TF families whose binding sites are preferentially affected. This may reflect the fact that such analyses not specific to a particular cell type.

The single most important discovery presented in this work is the experimental confirmation of the importance of USF binding sites for the evolution of the astrocyte regulome. Using MPRA, I demonstrate that the presence of USF binding sites is essential for enhancer activity and that even subtle sequence alterations affecting those sites in enhancers can modulate their function. Those results confirm the previously reported[286] correlation of the gain of USF binding sites in proximity of human-specific genes in foetal astrocytes, and validates a machine learning model predictions on their importance in enhancers.

In our published study and this work, we describe a genome-wide significance of USFs in astrocytes. The importance of only few of USFs for enhancer activity has been reported before[570]. Apart from our published study and this work, no study has explored the evolutionary significance of USF site gain. Furthermore, this work is the first to experimentally demonstrate the importance of a single TFBS class in an evolutionary context. MPRA results indicate that a majority of genes more expressed in humans than in NHPs have at least one human-specific enhancer active in MPRA assay in their proximity. Thus, “linked” enhancers may contribute to the increased expression of genes upregulated in humans relative to other species. This difference appears to be driven by a greater number of USF binding sites. I also show that differences between human and chimpanzee, and human and other hominids, occur more frequently in USF sites overlapping enhancers linked to genes upregulated in humans relative and within those enhancers, affect USF binding sites more frequently.

It is particularly striking that even minimal sequence alterations resulting in a loss or gain of a USF binding site can modify enhancer activity. This observation is especially significant in the light of the well-described phenomena of *de novo* enhancer emergence through only a few sequence changes[292], rapid enhancer turnover[571], and co-option of genomic regions with latent enhancer potential during evolution[572]. Thus, the emergence of new enhancers through the gain of USF binding sites might be a novel mechanism of enhancer evolution.

At the same time, it is well established that regulatory networks exhibit buffering capacity against alterations in single enhancers, conferring robustness to gene expression and phenotype[573]. How such buffering can be reconciled with the potential emergence of enhancers via small sequence changes that create a more accessible chromatin environment remains an open question. It is also worth considering that the rapid evolutionary changes observed between human and rhesus macaque, as reported by Li et al., (2023), may follow different dynamics and selective pressures than those inferred from comparisons across broader mammalian lineages.

6.6. Limitations of the study

The main experimental model employed in this study has several methodological limitations. Most notably, the MPRA tested the transcriptional activity of 200 bp fragments corresponding to the central regions of open chromatin peaks. This design constraint arises from the technical limitations of large-scale oligonucleotide synthesis. However, despite this simplification, both this and similar MPRA-based approaches have been shown to yield activity estimates that are broadly concordant with luciferase reporter assays and predictions derived from perturbation-based studies[293,331,428,562,574]. Nevertheless, it should be noted that some transcription factors preferentially bind at the edges of open chromatin regions[575]. As a result, such elements may be underrepresented in the MPRA measurements of enhancer activity and, consequently, in the inferred estimates of transcription factor binding site (TFBS) importance. This caveat should be considered when interpreting the TFBS–activity relationships reported in this work.

The MPRA protocol employed in this study is a variation of the method originally described by Tewhey et al. (2016) and is based on expression from a modified pGL4 plasmid. Consequently, this system provides an episomal readout, which may differ from how an enhancer behaves within its native chromatin context. The expression cassette consists of an enhancer fragment positioned upstream of a minimal promoter, followed by a reporter gene, with a unique barcode sequence located in the 3′ untranslated region (3′UTR) of the reporter. Several variants of the MPRA assay have been developed and systematically compared[428]. Among them, a notable version employs lentiviral delivery of the reporter cassette, referred to as lentiMPRA[427]. However, the same authors reported a high concordance (>85%) in enhancer activity predictions between episomal, pGL4-based MPRAs and lentiMPRA.

Another important limitation of this study lies in the cellular model used to express the MPRA library. The LN229 astrocytoma cell line was chosen due to its partial similarity to astrocytes, experimental tractability, ease of large-scale expansion, and high transfection efficiency. In contrast, I was unable to achieve sufficient transfection efficiency in iAstrocytes to perform an MPRA screen at the required scale. A similar strategy has been adopted in previous high-throughput reporter assay studies[293,331,576], which were nonetheless considered valid and informative within their respective experimental contexts.

A broader limitation of the astrocyte system used for enhancer characterisation and perturbation testing is its reliance on a monolayer culture that lacks the bidirectional signalling between neurons and astrocytes. This may restrict the detection of regulatory mechanisms involved in neuron-astrocyte cross-talk and could partially explain the absence of evolutionary changes in the regulome associated with genes controlling synapse formation, as reported by Jorstad et al. (2023). At the same time, analysing enhancers in isolation from neuronal influence may have the advantage of highlighting intrinsic astrocyte-specific regulatory changes, thereby revealing the most robust, independent of external cues.

7. Conclusions

Based on the data presented in this work, it can be concluded that:

- USFs are important for astrocytic enhancer activity, and their greater number correlates with greater enhancer activity.
- The iAstrocyte regulome is enriched in sites of accelerated evolutionary change specific to humans, relative to both non-human primates are other hominids.
- The iAstrocyte regulome is enriched in human polymorphisms associated with brain morphology, which places astrocytes at the heart of the regulation of brain size in evolution.
- Gain of USF binding sites in human-specific enhancers is essential for bolstering enhancer activity in evolution.

Furthermore, in this work, I experimentally tested the effects of human 283 polymorphisms on enhancer activity and identified 108 polymorphisms that significantly modulate enhancer activity in astrocytes.

References

1. Zhao, Y. *et al.* “Stripe” transcription factors provide accessibility to co-binding partners in mammalian genomes. *Mol. Cell* **82**, 3398-3411.e11 (2022).
2. Kandel, E. R., Koester, J., Mack, S. & Siegelbaum, S. *Principles of Neural Science*. (McGraw Hill, 2021).
3. Barres, B., Freeman, M. R. & Stevens, B. *Glia*. (Cold Spring Harbor Laboratory Press, 2015).
4. Halassa, M. M., Fellin, T., Takano, H., Dong, J. H. & Haydon, P. G. Synaptic islands defined by the territory of a single astrocyte. *J. Neurosci.* **27**, 6473–6477 (2007).
5. Somjen, G. G. Nervenkitt: Notes on the history of the concept of neuroglia. *Glia* **1**, 2–9 (1988).
6. von Bartheld, C. S., Bahney, J. & Herculano-Houzel, S. The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *J. Comp. Neurol.* **524**, 3865–3895 (2016).
7. Lindsay, R. M. Adult rat brain astrocytes support survival of both NGF-dependent and NGF-insensitive neurones. *Nature* **282**, 80–82 (1979).
8. Banker, G. A. Trophic Interactions Between Astroglial Cells and Hippocampal Neurons in Culture. *Science (80-.).* **209**, 809–810 (1980).
9. McCarthy, K. D. & de Vellis, J. Preparation of separate astroglial and oligodendroglial cell cultures from rat cerebral tissue. *J. Cell Biol.* **85**, 890–902 (1980).
10. Mennerick, S. & Zorumski, C. F. Glial contributions to excitatory neurotransmission in cultured hippocampal cells. *Nature* **368**, 59–62 (1994).
11. Parpura, V. *et al.* Glutamate-mediated astrocyte–neuron signalling. *Nature* **369**, 744–747 (1994).
12. Cornell-Bell, A. H., Finkbeiner, S. M., Cooper, M. S. & Smith, S. J. Glutamate Induces Calcium Waves in Cultured Astrocytes: Long-Range Glial Signaling. *Science (80-.).* **247**, 470–473 (1990).
13. Charles, A. C., Merrill, J. E., Dirksen, E. R. & Sandersont, M. J. Intercellular signaling in glial cells: Calcium waves and oscillations in response to mechanical stimulation and glutamate. *Neuron* **6**, 983–992 (1991).

14. Pfrieger, F. W. & Barres, B. A. Synaptic Efficacy Enhanced by Glial Cells in Vitro. *Science (80-.).* **277**, 9–12 (1997).

15. Ullian, E. M., Sapperstein, S. K., Christopherson, K. S. & Barres, B. A. Control of Synapse Number by Glia. *Science (80-.).* **291**, 657–661 (2001).

16. Nedergaard, M. Direct Signaling from Astrocytes to Neurons in Cultures of Mammalian Brain Cells. *Science (80-.).* **263**, 1768–1771 (1994).

17. Johnson, M. A., Weick, J. P., Pearce, R. A. & Zhang, S. C. Functional neural development from human embryonic stem cells: Accelerated synaptic activity via astrocyte coculture. *J. Neurosci.* **27**, 3069–3077 (2007).

18. Verkhratsky, A. & Nedergaard, M. Physiology of astroglia. *Physiol. Rev.* **98**, 239–389 (2018).

19. Verkhratsky, A., Nedergaard, M. & Hertz, L. Why are Astrocytes Important? *Neurochem. Res.* **40**, 389–401 (2015).

20. Schaeffer, S. & Iadecola, C. Revisiting the neurovascular unit. *Nat. Neurosci.* **24**, 1198–1209 (2021).

21. Szu, J. I. & Binder, D. K. The role of astrocytic aquaporin-4 in synaptic plasticity and learning and memory. *Front. Integr. Neurosci.* **10**, 1–16 (2016).

22. Pociūtė, A., Pivoriūnas, A. & Verkhratsky, A. Astrocytes dynamically regulate the blood-brain barrier in the healthy brain. *Neural Regen. Res.* **19**, 709–710 (2024).

23. Haddad-Tóvolli, R., Dragano, N. R. V., Ramalho, A. F. S. & Velloso, L. A. Development and function of the blood-brain barrier in the context of metabolic control. *Front. Neurosci.* **11**, 1–12 (2017).

24. Mathiisen, T. M., Lehre, K. P., Danbolt, N. C. & Ottersen, O. P. The perivascular astroglial sheath provides a complete covering of the brain microvessels: An electron microscopic 3D reconstruction. *Glia* **58**, 1094–1103 (2010).

25. Morgello, S., Uson, R. R., Schwartz, E. J. & Haber, R. S. The human blood-brain barrier glucose transporter (GLUT1) is a glucose transporter of gray matter astrocytes. *Glia* **14**, 43–54 (1995).

26. Pifferi, F., Laurent, B. & Plourde, M. Lipid Transport and Metabolism at the Blood-Brain Interface: Implications in Health and Disease. *Front. Physiol.* **12**, (2021).

27. Sun, J. & He, Z. Interactions between astrocytes and cerebral endothelial cells in

central nervous system diseases. *Biomed. Pharmacother.* **191**, 118516 (2025).

28. Bak, L. K. Astrocytes take the stage in a tale of signaling-metabolism coupling. *J. Biol. Chem.* **292**, 9439–9440 (2017).

29. Nagelhus, E. A. & Ottersen, O. P. Physiological roles of Aquaporin-4 in brain. *Physiol. Rev.* **93**, 1543–1562 (2013).

30. Christopherson, K. S. *et al.* Thrombospondins are astrocyte-secreted proteins that promote CNS synaptogenesis. *Cell* **120**, 421–433 (2005).

31. Farhy-Tselnicker, I. *et al.* Activity-dependent modulation of synapse-regulating genes in astrocytes. *Elife* **10**, 1–43 (2021).

32. Farhy-Tselnicker, I. *et al.* Astrocyte-Secreted Glypican 4 Regulates Release of Neuronal Pentraxin 1 from Axons to Induce Functional Synapse Formation. *Neuron* **96**, 428-445.e13 (2017).

33. Allen, N. J. *et al.* Astrocyte glypicans 4 and 6 promote formation of excitatory synapses via GluA1 AMPA receptors. *Nature* **486**, 410–414 (2012).

34. Risher, W. C. & Eroglu, C. Astrocytes and synaptogenesis. in *Synapse Development and Maturation* 55–75 (Elsevier, 2020). doi:10.1016/B978-0-12-823672-7.00003-X.

35. Risher, W. C. *et al.* Thrombospondin receptor α2δ-1 promotes synaptogenesis and spinogenesis via postsynaptic Rac1. *J. Cell Biol.* **217**, 3747–3765 (2018).

36. Eroglu, Ç. *et al.* Gabapentin Receptor α2δ-1 Is a Neuronal Thrombospondin Receptor Responsible for Excitatory CNS Synaptogenesis. *Cell* **139**, 380–392 (2009).

37. Blanco-Suarez, E., Liu, T.-F., Kopelevich, A. & Allen, N. J. Astrocyte-Secreted Chordin-like 1 Drives Synapse Maturation and Limits Plasticity by Increasing Synaptic GluA2 AMPA Receptors. *Neuron* **100**, 1116-1132.e13 (2018).

38. Beattie, E. C. *et al.* Control of synaptic strength by glial TNFα. *Science (80-.).* **295**, 2282–2285 (2002).

39. Diniz, L. P. *et al.* Astrocyte-induced synaptogenesis is mediated by transforming growth factor β signaling through modulation of d-serine levels in cerebral cortex neurons. *J. Biol. Chem.* **287**, 41432–41445 (2012).

40. Patel, M. R. & Weaver, A. M. Astrocyte-derived small extracellular vesicles promote synapse formation via fibulin-2-mediated TGF-β signaling. *Cell Rep.* **34**, 108829 (2021).

41. Hu, R., Cai, W. Q., Wu, X. G. & Yang, Z. Astrocyte-derived estrogen enhances synapse formation and synaptic transmission between cultured neonatal rat cortical neurons. *Neuroscience* **144**, 1229–1240 (2007).

42. Hama, H., Hara, C., Yamaguchi, K. & Miyawaki, A. PKC Signaling Mediates Global Enhancement of Excitatory Synaptogenesis in Neurons Triggered by Local Contact with Astrocytes. *Neuron* **41**, 405–415 (2004).

43. Garrett, A. M. & Weiner, J. A. Control of CNS synapse development by γ-protocadherin-mediated astrocyte-neuron contact. *J. Neurosci.* **29**, 11723–11731 (2009).

44. Chen, W. V. *et al.* Functional significance of isoform diversification in the protocadherin gamma gene cluster. *Neuron* **75**, 402–409 (2012).

45. Steffen, D. M. *et al.* The γ-Protocadherins Interact Physically and Functionally with Neuroligin-2 to Negatively Regulate Inhibitory Synapse Density and Are Required for Normal Social Interaction. *Mol. Neurobiol.* **58**, 2574–2589 (2021).

46. Molumby, M. J. *et al.* γ-Protocadherins Interact with Neuroligin-1 and Negatively Regulate Dendritic Spine Morphogenesis. *Cell Rep.* **18**, 2702–2714 (2017).

47. Stogsdill, J. A. *et al.* Astrocytic neuroligins control astrocyte morphogenesis and synaptogenesis. *Nature* **551**, 192–197 (2017).

48. Illes, P., Burnstock, G. & Tang, Y. Astroglia-Derived ATP Modulates CNS Neuronal Circuits. *Trends Neurosci.* **42**, 885–898 (2019).

49. Lezmy, J. How astrocytic ATP shapes neuronal activity and brain circuits. *Curr. Opin. Neurobiol.* **79**, 102685 (2023).

50. Suadicani, S. O., Brosnan, C. F. & Scemes, E. P2X7 receptors mediate ATP release and amplification of astrocytic intercellular Ca2+ signaling. *J. Neurosci.* **26**, 1378–1385 (2006).

51. Li, H. *et al.* Astrocytes release ATP/ADP and glutamate in flashes via vesicular exocytosis. *Mol. Psychiatry* **30**, 2475–2489 (2025).

52. Xiong, Y. *et al.* Stretch-induced Ca2+ independent ATP release in hippocampal astrocytes. *J. Physiol.* **596**, 1931–1947 (2018).

53. Rothstein, J. D. *et al.* Knockout of glutamate transporters reveals a major role for astroglial transport in excitotoxicity and clearance of glutamate. *Neuron* **16**, 675–686 (1996).

54. Henneberger, C., Papouin, T., Oliet, S. H. R. & Rusakov, D. A. Long-term potentiation depends on release of d-serine from astrocytes. *Nature* **463**, 232–236 (2010).

55. Liu, J., Feng, X., Wang, Y., Xia, X. & Zheng, J. C. Astrocytes: GABAceptive and GABAergic Cells in the Brain. *Front. Cell. Neurosci.* **16**, 1–12 (2022).

56. Sasi, M., Vignoli, B., Canossa, M. & Blum, R. Neurobiology of local and intercellular BDNF signaling. *Pflugers Arch. Eur. J. Physiol.* **469**, 593–610 (2017).

57. De Pins, B. *et al.* Conditional BDNF delivery from astrocytes rescues memory deficits, spine density, and synaptic properties in the 5xFAD mouse model of alzheimer disease. *J. Neurosci.* **39**, 2441–2458 (2019).

58. Webb, T. R. *et al.* X-linked megalocornea caused by mutations in CHRDL1 identifies an essential role for ventroptin in anterior segment development. *Am. J. Hum. Genet.* **90**, 247–259 (2012).

59. Bosworth, A. P. *et al.* Astrocyte glypican 5 regulates synapse maturation and stabilization. *Cell Rep.* **44**, 115374 (2025).

60. Kucukdereli, H. *et al.* Control of excitatory CNS synaptogenesis by astrocyte-secreted proteins hevin and SPARC. *Proc. Natl. Acad. Sci. U. S. A.* **108**, (2011).

61. Singh, S. K. *et al.* Astrocytes Assemble Thalamocortical Synapses by Bridging NRX1α and NL1 via Hevin. *Cell* **164**, 183–196 (2016).

62. Araque, A., Parpura, V., Sanzgiri, R. P. & Haydon, P. G. Tripartite synapses: Glia, the unacknowledged partner. *Trends Neurosci.* **22**, 208–215 (1999).

63. Durkee, C. A. & Araque, A. Diversity and Specificity of Astrocyte–neuron Communication. *Neuroscience* **396**, 73–78 (2019).

64. Noriega-Prieto, J. A. & Araque, A. Sensing and Regulating Synaptic Activity by Astrocytes at Tripartite Synapse. *Neurochem. Res.* **46**, 2580–2585 (2021).

65. Perea, G., Navarrete, M. & Araque, A. Tripartite synapses: astrocytes process and control synaptic information. *Trends Neurosci.* **32**, 421–431 (2009).

66. Pathak, D. & Sriram, K. Neuron-astrocyte omnidirectional signaling in neurological health and disease. *Front. Mol. Neurosci.* **16**, 1–27 (2023).

67. Hillen, A. E. J., Burbach, J. P. H. & Hol, E. M. Cell adhesion and matricellular support by astrocytes of the tripartite synapse. *Prog. Neurobiol.* **165–167**, 66–86 (2018).

68. Shen, F. S. *et al.* Emerging evidence of context-dependent synapse elimination by

phagocytes in the CNS. *J. Leukoc. Biol.* **116**, 511–522 (2024).

69. Stevens, B. *et al.* The Classical Complement Cascade Mediates CNS Synapse Elimination. *Cell* **131**, 1164–1178 (2007).

70. Vainchtein, I. D. *et al.* Astrocyte-derived interleukin-33 promotes microglial synapse engulfment and neural circuit development. *Science (80-.).* **359**, 1269–1273 (2018).

71. Wang, Y. *et al.* Astrocyte-secreted IL-33 mediates homeostatic synaptic plasticity in the adult hippocampus. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–12 (2021).

72. Faust, T. E. *et al.* Microglia-astrocyte crosstalk regulates synapse remodeling via Wnt signaling. *Cell* **188**, 5212-5230.e21 (2025).

73. Chung, W. S. *et al.* Astrocytes mediate synapse elimination through MEGF10 and MERTK pathways. *Nature* **504**, 394–400 (2013).

74. Yang, J. *et al.* Astrocytes contribute to synapse elimination Via type 2 inositol 1,4,5-trisphosphate receptor-dependent release of ATP. *Elife* **5**, 1–17 (2016).

75. Koeppen, J. *et al.* Functional consequences of synapse remodeling following astrocyte-specific regulation of ephrin-B1 in the adult hippocampus. *J. Neurosci.* **38**, 5710–5726 (2018).

76. Araque, A. *et al.* Gliotransmitters travel in time and space. *Neuron* **81**, 728–739 (2014).

77. Wolosker, H. *The Neurobiology of D-Serine Signaling. Advances in Pharmacology* vol. 82 (Elsevier Inc., 2018).

78. Andersen, J. V. & Schousboe, A. Milestone Review: Metabolic dynamics of glutamate and GABA mediated neurotransmission — The essential roles of astrocytes. *J. Neurochem.* **166**, 109–137 (2023).

79. Pajarillo, E., Rizor, A., Lee, J., Aschner, M. & Lee, E. The role of astrocytic glutamate transporters GLT-1 and GLAST in neurological disorders: Potential targets for neurotherapeutics. *Neuropharmacology* **161**, 107559 (2019).

80. Robinson, M. B. & Jackson, J. G. Astroglial glutamate transporters coordinate excitatory signaling and brain energetics. *Neurochem. Int.* **98**, 56–71 (2016).

81. Dong, X. X., Wang, Y. & Qin, Z. H. Molecular mechanisms of excitotoxicity and their relevance to pathogenesis of neurodegenerative diseases. *Acta Pharmacol. Sin.* **30**, 379–387 (2009).

82. Waniewski, R. A. & Martin, D. L. Exogenous Glutamate Is Metabolized to Glutamine and Exported by Rat Primary Astrocyte Cultures. *J. Neurochem.* **47**, 304–313 (1986).

83. Rose, C. F., Verkhratsky, A. & Parpura, V. Astrocyte glutamine synthetase: Pivotal in health and disease. *Biochem. Soc. Trans.* **41**, 1518–1524 (2013).

84. Zhang, D., Hua, Z. & Li, Z. The role of glutamate and glutamine metabolism and related transporters in nerve cells. *CNS Neurosci. Ther.* **30**, 1–11 (2024).

85. McKenna, M. C. *et al.* Glutamate oxidation in astrocytes: Roles of glutamate dehydrogenase and aminotransferases. *J. Neurosci. Res.* **94**, 1561–1571 (2016).

86. McKenna, M. C. Glutamate pays its own way in astrocytes. *Front. Endocrinol. (Lausanne).* **4**, 1–6 (2013).

87. Chatton, J. Y., Magistretti, P. J. & Barros, L. F. Sodium signaling and astrocyte energy metabolism. *Glia* **64**, 1667–1676 (2016).

88. Allaman, I., Bélanger, M. & Magistretti, P. J. Astrocyte-neuron metabolic relationships: For better and for worse. *Trends Neurosci.* **34**, 76–87 (2011).

89. Romanos, J., Benke, D., Saab, A. S., Zeilhofer, H. U. & Santello, M. Differences in glutamate uptake between cortical regions impact neuronal NMDA receptor activation. *Commun. Biol.* **2**, (2019).

90. Andersen, J. V, Schousboe, A. & Verkhratsky, A. Astrocyte energy and neurotransmitter metabolism in Alzheimer ' s disease : Integration of the glutamate/ GABA-glutamine cycle. *Prog. Neurobiol.* **217**, 0–3 (2022).

91. Rothstein, J. D., Kammen, M. Van, Levey, A. I., Martin, L. J. & Kuncl, R. W. Selective loss of glial glutamate transporter GLT-1 in amyotrophic lateral sclerosis. *Ann. Neurol.* **38**, 73–84 (1995).

92. Binder, D. K. & Auser, C. S. Functional Changes in Astroglial Cells in Epilepsy. *Glia* **54**, 358–368 (2006).

93. Rodrigues, R. J., Almeida, T., Richardson, P. J., Oliveira, C. R. & Cunha, R. A. Dual presynaptic control by ATP of glutamate release via facilitatory P2X1, P2X2/3, and P2X3 and inhibitory P2Y 1, P2Y2, and/or P2Y4 receptors in the rat hippocampus. *J. Neurosci.* **25**, 6286–6295 (2005).

94. Lalo, U., Palygin, O., Verkhratsky, A., Grant, S. G. N. & Pankratov, Y. ATP from synaptic terminals and astrocytes regulates NMDA receptors and synaptic plasticity through PSD-95 multi-protein complex. *Sci. Rep.* **6**, 1–20 (2016).

95. Boué-Grabot, E. & Pankratov, Y. Modulation of Central Synapses by Astrocyte-Released ATP and Postsynaptic P2X Receptors. *Neural Plast.* **2017**, (2017).

96. Pougnet, J. T. *et al.* ATP P2X receptors downregulate AMPA receptor trafficking and postsynaptic efficacy in hippocampal neurons. *Neuron* **83**, 417–430 (2014).

97. Tan, Z. *et al.* Glia-derived ATP inversely regulates excitability of pyramidal and CCK-positive neurons. *Nat. Commun.* **8**, (2017).

98. Matos, M. *et al.* Astrocytes detect and upregulate transmission at inhibitory synapses of somatostatin interneurons onto pyramidal cells. *Nat. Commun.* **9**, (2018).

99. Boddum, K. *et al.* Astrocytic GABA transporter activity modulates excitatory neurotransmission. *Nat. Commun.* **7**, 1–10 (2016).

100. Martin-Fernandez, M. *et al.* Synapse-specific astrocyte gating of amygdala-related behavior. *Nat. Neurosci.* **20**, 1540–1548 (2017).

101. Corkrum, M. *et al.* Dopamine-Evoked Synaptic Regulation in the Nucleus Accumbens Requires Astrocyte Activity. *Neuron* **105**, 1036-1047.e5 (2020).

102. Cao, X. *et al.* Astrocyte-derived ATP modulates depressive-like behaviors. *Nat. Med.* **19**, 773–777 (2013).

103. Wolosker, H., Balu, D. T. & Coyle, J. T. The Rise and Fall of the D-Serine-Mediated Gliotransmission Hypothesis. *Trends Neurosci.* **39**, 712–721 (2016).

104. Beltrán-Castillo, S. *et al.* D-serine released by astrocytes in brainstem regulates breathing response to CO2 levels. *Nat. Commun.* **8**, 1–13 (2017).

105. Benneyworth, M. A., Li, Y., Basu, A. C., Bolshakov, V. Y. & Coyle, J. T. Cell selective conditional null mutations of serine racemase demonstrate a predominate localization in cortical glutamatergic neurons. *Cell. Mol. Neurobiol.* **32**, 613–624 (2012).

106. Langlais, V. C. *et al.* Astrocytic EphB3 receptors regulate D-serine-gated synaptic plasticity and memory. *Prog. Neurobiol.* **248**, (2025).

107. Soghomonian, J. J. & Martin, D. L. Two isoforms of glutamate decarboxylase: Why? *Trends Pharmacol. Sci.* **19**, 500–505 (1998).

108. Zhou, Y. & Danbolt, N. C. GABA and glutamate transporters in brain. *Front. Endocrinol. (Lausanne).* **4**, 1–14 (2013).

109. Duarte, J. M. N. & Gruetter, R. Glutamatergic and GABAergic energy metabolism measured in the rat brain by 13C NMR spectroscopy at 14.1 T. *J. Neurochem.* **126**,

579–590 (2013).

110. Andersen, J. V. *et al.* Extensive astrocyte metabolism of γ-aminobutyric acid (GABA) sustains glutamine synthesis in the mammalian cerebral cortex. *Glia* **68**, 2601–2612 (2020).

111. Andersen, J. V. The Glutamate/GABA-Glutamine Cycle: Insights, Updates, and Advances. *J. Neurochem.* **169**, (2025).

112. Wójtowicz, A. M., Dvorzhak, A., Semtner, M. & Grantyn, R. Reduced tonic inhibition in striatal output neurons from Huntington mice due to loss of astrocytic GABA release through GAT-3. *Front. Neural Circuits* **7**, 1–12 (2013).

113. Pirttimaki, T., Parri, H. R. & Crunelli, V. Astrocytic GABA transporter GAT-1 dysfunction in experimental absence seizures. *J. Physiol.* **591**, 823–833 (2013).

114. Woo, J. *et al.* Control of motor coordination by astrocytic tonic GABA release through modulation of excitation/inhibition balance in cerebellum. *Proc. Natl. Acad. Sci.* **115**, 5004–5009 (2018).

115. Mederos, S. *et al.* GABAergic signaling to astrocytes in the prefrontal cortex sustains goal-directed behaviors. *Nat. Neurosci.* **24**, 82–92 (2021).

116. Barca-mayo, O. *et al.* Astrocyte deletion of Bmal1 alters daily locomotor activity and cognitive functions via GABA signalling. *Nat. Commun.* **8**, (2017).

117. Lia, A. *et al.* Calcium Signals in Astrocyte Microdomains, a Decade of Great Advances. *Front. Cell. Neurosci.* **15**, 1–7 (2021).

118. Denizot, A., Arizono, M., Nägerl, U. V., Berry, H. & De Schutter, E. Control of Ca2+ signals by astrocyte nanoscale morphology at tripartite synapses. *Glia* **70**, 2378–2391 (2022).

119. Cahill, M. K. *et al.* Network-level encoding of local neurotransmitters in cortical astrocytes. *Nature* **629**, 146–153 (2024).

120. Kuga, N., Sasaki, T., Takahara, Y., Matsuki, N. & Ikegaya, Y. Large-scale calcium waves traveling through astrocytic networks in vivo. *J. Neurosci.* **31**, 2607–2614 (2011).

121. Polykretis, I., Ivanov, V. & Michmizos, K. P. A Neural-Astrocytic Network Architecture: Astrocytic calcium waves modulate synchronous neuronal activity. *ACM Int. Conf. Proceeding Ser.* (2018) doi:10.1145/3229884.3229890.

122. Goenaga, J., Araque, A., Kofuji, P. & Herrera Moro Chao, D. Calcium signaling in astrocytes and gliotransmitter release. *Front. Synaptic Neurosci.* **15**, 1–11 (2023).

123. Semyanov, A. Spatiotemporal pattern of calcium activity in astrocytic network. *Cell Calcium* **78**, 15–25 (2019).

124. Benoit, L. *et al.* Astrocytes functionally integrate multiple synapses via specialized leaflet domains. *Cell* 6453–6472 (2025) doi:10.1016/j.cell.2025.08.036.

125. Smith, S. Chapter 10: Do astrocytes process neural information? in *Neuronal-astrocytic interactions; implications for normal and pathological CNS function* vol. 94 119–136 (1992).

126. Wang, X., Bukoreshtliev, N. V. & Gerdes, H. H. Developing Neurons Form Transient Nanotubes Facilitating Electrical Coupling and Calcium Signaling with Distant Astrocytes. *PLoS One* **7**, 1–9 (2012).

127. Zhou, C. *et al.* Tunneling nanotubes: The transport highway for astrocyte-neuron communication in the central nervous system. *Brain Res. Bull.* **209**, 110921 (2024).

128. Khattar, K. E., Safi, J., Rodriguez, A. M. & Vignais, M. L. Intercellular Communication in the Brain through Tunneling Nanotubes. *Cancers (Basel).* **14**, 1–22 (2022).

129. Chen, J. & Cao, J. Astrocyte-to-neuron transportation of enhanced green fluorescent protein in cerebral cortex requires F-actin dependent tunneling nanotubes. *Sci. Rep.* **11**, 1–11 (2021).

130. Allaman, I. & Magistretti, P. J. Brain Energy Metabolism. in *Fundamental Neuroscience* 261–284 (Elsevier, 2013). doi:10.1016/B978-0-12-385870-2.00012-3.

131. Qutub, A. A. & Hunt, C. A. Glucose transport to the brain: A systems model. *Brain Res. Rev.* **49**, 595–617 (2005).

132. Zhang, W. *et al.* Fatty acid transporting proteins: Roles in brain development, aging, and stroke. *Prostaglandins Leukot. Essent. Fat. Acids* **136**, 35–45 (2018).

133. Schiera, G. *et al.* Synergistic effects of neurons and astrocytes on the differentiation of brain capillary endothelial cells in culture. *J. Cell. Mol. Med.* **7**, 165–70 (2003).

134. Schiera, G. *et al.* Permeability properties of a three-cell type in vitro model of blood-brain barrier. *J. Cell. Mol. Med.* **9**, 373–9 (2005).

135. Proia, P. *et al.* Astrocytes shed extracellular vesicles that contain fibroblast growth factor-2 and vascular endothelial growth factor. *Int. J. Mol. Med.* **21**, 63–7 (2008).

136. Fu, J. *et al.* Astrocyte-Derived TGFβ1 Facilitates Blood-Brain Barrier Function via Non-Canonical Hedgehog Signaling in Brain Microvascular Endothelial Cells. *Brain Sci.* **11**, (2021).

137. Siegenthaler, J. A., Sohet, F. & Daneman, R. 'Sealing off the CNS': Cellular and molecular regulation of blood-brain barriergenesis. *Curr. Opin. Neurobiol.* **23**, 1057–1064 (2013).

138. Wosik, K. *et al.* Angiotensin II controls occludin function and is required for blood-brain barrier maintenance: Relevance to multiple sclerosis. *J. Neurosci.* **27**, 9032–9042 (2007).

139. Siddharthan, V., Kim, Y. V., Liu, S. & Kim, K. S. Human astrocytes/astrocyte-conditioned medium and shear stress enhance the barrier properties of human brain microvascular endothelial cells. *Brain Res.* **1147**, 39–50 (2007).

140. Heithoff, B. P. *et al.* Astrocytes are necessary for blood–brain barrier maintenance in the adult mouse brain. *Glia* **69**, 436–472 (2021).

141. Guérit, S. *et al.* Astrocyte-derived Wnt growth factors are required for endothelial blood-brain barrier maintenance. *Prog. Neurobiol.* **199**, (2021).

142. Yao, L. L. *et al.* Astrocytic neogenin/netrin-1 pathway promotes blood vessel homeostasis and function in mouse cortex. *J. Clin. Invest.* **130**, 6490–6509 (2020).

143. Iliff, J. J. *et al.* A Paravascular Pathway Facilitates CSF Flow Through the Brain Parenchyma and the Clearance of Interstitial Solutes, Including Amyloid b. *Sci. Transl. Med.* **4**, 147ra111 (2012).

144. Jessen, N. A., Munk, A. S. F., Lundgaard, I. & Nedergaard, M. The Glymphatic System: A Beginner's Guide. *Neurochem. Res.* **40**, 2583–2599 (2015).

145. Xie, L. *et al.* Sleep drives metabolite clearance from the adult brain. *Science (80-.).* **342**, 373–377 (2013).

146. Rasmussen, M. K., Mestre, H. & Nedergaard, M. Fluid transport in the brain. *Physiol. Rev.* **102**, 1025–1151 (2022).

147. Mächler, P. *et al.* In Vivo Evidence for a Lactate Gradient from Astrocytes to Neurons. *Cell Metab.* **23**, 94–102 (2016).

148. Pfrieger, F. W. & Ungerer, N. Cholesterol metabolism in neurons and astrocytes. *Prog. Lipid Res.* **50**, 357–371 (2011).

149. van Deijk, A. L. F. *et al.* Astrocyte lipid metabolism is critical for synapse development and function in vivo. *Glia* **65**, 670–682 (2017).

150. Ioannou, M. S. *et al.* Neuron-Astrocyte Metabolic Coupling Protects against Activity-Induced Fatty Acid Toxicity. *Cell* **177**, 1522-1535.e14 (2019).

151. Liu, L. *et al.* Glial lipid droplets and ROS induced by mitochondrial defects promote neurodegeneration. *Cell* **160**, 177–190 (2015).

152. Gibbs, M. E., Anderson, D. G. & Hertz, L. Inhibition of glycogenolysis in astrocytes interrupts memory consolidation in young chickens. *Glia* **54**, 214–222 (2006).

153. Hertz, L., O'Dowd, B. S., Ng, K. T. & Gibbs, M. E. Reciprocal changes in forebrain contents of glycogen and of glutamate/glutamine during early memory consolidation in the day-old chick. *Brain Res.* **994**, 226–233 (2003).

154. Gentry, M. S., Guinovart, J. J., Minassian, B. A., Roach, P. J. & Serratosa, J. M. Lafora disease offers a unique window into neuronal glycogen metabolism. *J. Biol. Chem.* **293**, 7117–7125 (2018).

155. Hertz, L., Lovatt, D., Goldman, S. A. & Nedergaard, M. Adrenoceptors in brain: Cellular gene expression and effects on astrocytic metabolism and [Ca2+]i. *Neurochem. Int.* **57**, 411–420 (2010).

156. Müller, M. S., Fox, R., Schousboe, A., Waagepetersen, H. S. & Bak, L. K. Astrocyte glycogenolysis is triggered by store-operated calcium entry and provides metabolic energy for cellular calcium homeostasis. *Glia* **62**, 526–534 (2014).

157. Pierre, K. & Pellerin, L. Monocarboxylate transporters in the central nervous system: Distribution, regulation and function. *J. Neurochem.* **94**, 1–14 (2005).

158. Dienel, G. A. The metabolic trinity, glucose–glycogen–lactate, links astrocytes and neurons in brain energetics, signaling, memory, and gene expression. *Neurosci. Lett.* **637**, 18–25 (2017).

159. Bonvento, G. & Bolaños, J. P. Astrocyte-neuron metabolic cooperation shapes brain activity. *Cell Metab.* **33**, 1546–1564 (2021).

160. Beard, E., Lengacher, S., Dias, S., Magistretti, P. J. & Finsterwald, C. Astrocytes as Key Regulators of Brain Energy Metabolism: New Therapeutic Perspectives. *Front. Physiol.* **12**, (2022).

161. Herrero-Mendez, A. *et al.* The bioenergetic and antioxidant status of neurons is controlled by continuous degradation of a key glycolytic enzyme by APC/C-Cdh1. *Nat.*

Cell Biol. **11**, 747–752 (2009).

162. Bolaños, J. P., Almeida, A. & Moncada, S. Glycolysis: a bioenergetic or a survival pathway? *Trends Biochem. Sci.* **35**, 145–149 (2010).

163. Suzuki, A. *et al.* Astrocyte-neuron lactate transport is required for long-term memory formation. *Cell* **144**, 810–823 (2011).

164. Yang, C. *et al.* Rewiring Neuronal Glycerolipid Metabolism Determines the Extent of Axon Regeneration. *Neuron* **105**, 276-292.e5 (2020).

165. Liu, L., MacKenzie, K. R., Putluri, N., Maletić-Savatić, M. & Bellen, H. J. The Glia-Neuron Lactate Shuttle and Elevated ROS Promote Lipid Synthesis in Neurons and Lipid Droplet Accumulation in Glia via APOE/D. *Cell Metab.* **26**, 719-737.e6 (2017).

166. Ferris, H. A. *et al.* Loss of astrocyte cholesterol synthesis disrupts neuronal function and alters whole-body metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 1189–1194 (2017).

167. Mauch, D. H. *et al.* CNS synaptogenesis promoted by glia-derived cholesterol. *Science (80-.).* **294**, 1354–1357 (2001).

168. Saher, G. *et al.* High cholesterol level is essential for myelin membrane growth. *Nat. Neurosci.* **8**, 468–475 (2005).

169. Nanjundaiah, S., Chidambaram, H., Chandrashekar, M. & Chinnathambi, S. Role of Microglia in Regulating Cholesterol and Tau Pathology in Alzheimer's Disease. *Cell. Mol. Neurobiol.* **41**, 651–668 (2021).

170. Getz, G. S. & Reardon, C. A. Apoprotein E as a lipid transport and signaling protein in the blood, liver, and artery wall. *J. Lipid Res.* **50**, S156–S161 (2009).

171. Talantova, M. *et al.* Aβ induces astrocytic glutamate release, extrasynaptic NMDA receptor activation, and synaptic loss. *Proc. Natl. Acad. Sci. U. S. A.* **110**, (2013).

172. Rama Rao, K. V., Curtis, K. M., Johnstone, J. T. & Norenberg, M. D. Amyloid-β inhibits thrombospondin 1 release from cultured astrocytes: Effects on synaptic protein expression. *J. Neuropathol. Exp. Neurol.* **72**, 735–744 (2013).

173. Ludwin, S. K., Rao, V. T. S., Moore, C. S. & Antel, J. P. Astrocytes in multiple sclerosis. *Mult. Scler.* **22**, 1114–1124 (2016).

174. Streubel-Gallasch, L. *et al.* Parkinson's Disease–Associated LRRK2 Interferes with Astrocyte-Mediated Alpha-Synuclein Clearance. *Mol. Neurobiol.* **58**, 3119–3140

(2021).

175. Yun, S. P. *et al.* Block of A1 astrocyte conversion by microglia is neuroprotective in models of Parkinson's disease. *Nat. Med.* **24**, 931–938 (2018).

176. Giehrl-Schwab, J. *et al.* Parkinson's disease motor symptoms rescue by CRISPRa-reprogramming astrocytes into GABAergic neurons. *EMBO Mol. Med.* **14**, 1–20 (2022).

177. Vezzani, A. *et al.* Astrocytes in the initiation and progression of epilepsy. *Nat. Rev. Neurol.* **18**, 707–722 (2022).

178. Chen, Z. P. *et al.* *Lipid-accumulated reactive astrocytes promote disease progression in epilepsy*. *Nature Neuroscience* vol. 26 (Springer US, 2023).

179. Ma, T. M. *et al.* Pathogenic disruption of DISC1-serine racemase binding elicits schizophrenia-like behavior via D-serine depletion. *Mol. Psychiatry* **18**, 557–567 (2013).

180. Tanahashi, S., Yamamura, S., Nakagawa, M., Motomura, E. & Okada, M. Clozapine, but not haloperidol, enhances glial d -serine and L-glutamate release in rat frontal cortex and primary cultured astrocytes. *Br. J. Pharmacol.* **165**, 1543–1555 (2012).

181. Rial, D. *et al.* Depression as a glial-based synaptic dysfunction. *Front. Cell. Neurosci.* **9**, 1–11 (2016).

182. Duarte, Y. *et al.* The role of astrocytes in depression, its prevention, and treatment by targeting astroglial gliotransmitter release. *Proc. Natl. Acad. Sci.* **121**, 1–12 (2024).

183. Sofroniew, M. V. Astrogliosis. *Cold Spring Harb. Perspect. Biol.* **7**, a020420 (2015).

184. Liddelow, S. A. *et al.* Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017).

185. Escartin, C. *et al.* Reactive astrocyte nomenclature, definitions, and future directions. *Nat. Neurosci.* **24**, 312–325 (2021).

186. Verkhratsky, A. *et al.* Astrocytes in human central nervous system diseases: a frontier for new therapies. *Signal Transduct. Target. Ther.* **8**, (2023).

187. Doméné, A. *et al.* Expression of Phenotypic Astrocyte Marker Is Increased in a Transgenic Mouse Model of Alzheimer's Disease versus Age-Matched Controls: A Presymptomatic Stage Study. *Int. J. Alzheimers. Dis.* **2016**, (2016).

188. González-Reyes, R. E., Nava-Mesa, M. O., Vargas-Sánchez, K., Ariza-Salamanca, D.

& Mora-Muñoz, L. Involvement of astrocytes in Alzheimer's disease from a neuroinflammatory and oxidative stress perspective. *Front. Mol. Neurosci.* **10**, 1–20 (2017).

189. Abramov, A. Y., Canevari, L. & Duchen, M. R. β-Amyloid Peptides Induce Mitochondrial Dysfunction and Oxidative Stress in Astrocytes and Death of Neurons through Activation of NADPH Oxidase. *J. Neurosci.* **24**, 565–575 (2004).

190. Matos, M. *et al.* Astrocytic adenosine A2A receptors control the amyloid-β peptide-induced decrease of glutamate uptake. *J. Alzheimer's Dis.* **31**, 555–567 (2012).

191. Vincent, A. J., Gasperini, R., Foa, L. & Small, D. H. Astrocytes in Alzheimer's disease: Emerging roles in calcium dysregulation and synaptic plasticity. *J. Alzheimer's Dis.* **22**, 699–714 (2010).

192. Dai, D. L., Li, M. & Lee, E. B. Human Alzheimer's disease reactive astrocytes exhibit a loss of homeostastic gene expression. *Acta Neuropathol. Commun.* **11**, 1–19 (2023).

193. Dejanovic, B. *et al.* Complement C1q-dependent excitatory and inhibitory synapse elimination by astrocytes and microglia in Alzheimer's disease mouse models. *Nat. Aging* **2**, 837–850 (2022).

194. Cabezas, R. *et al.* Astrocytic modulation of blood brain barrier: Perspectives on Parkinson's disease. *Front. Cell. Neurosci.* **8**, 1–11 (2014).

195. Saijo, K. *et al.* A Nurr1/CoREST Pathway in Microglia and Astrocytes Protects Dopaminergic Neurons from Inflammation-Induced Death. *Cell* **137**, 47–59 (2009).

196. Phatnani, H. & Maniatis, T. Astrocytes in neurodegenerative disease. *Cold Spring Harb. Perspect. Biol.* **7**, 1–18 (2015).

197. Ting Lau, L. & Cheung-hoi, A. Y. Astrocytes Produce and Release Interleukin-1, Interleukin-6, Tumor Necrosis Factor Alpha and Interferon-Gamma Following Traumatic and Metabolic Injury. *J. Neurotrauma* **18**, 351–359 (2001).

198. Leal, M. C., Casabona, J. C., Puntel, M. & Pitossi, F. Interleukin-1beta and TNF-alpha: Reliable targets for protective therapies in Parkinson's disease? *Front. Cell. Neurosci.* **7**, 1–10 (2013).

199. Yun, S. P. *et al.* Block of A1 astrocyte conversion by microglia is neuroprotective in models of Parkinson's disease. *Nat. Med.* **24**, 931–938 (2018).

200. Absinta, M. *et al.* A lymphocyte–microglia–astrocyte axis in chronic active multiple sclerosis. *Nature* **597**, 709–714 (2021).

201. Horng, S. *et al.* Astrocytic tight junctions control inflammatory CNS lesion pathogenesis. *J. Clin. Invest.* **127**, 3136–3151 (2017).

202. Kadowaki, A. *et al.* CLEC16A in astrocytes promotes mitophagy and limits pathology in a multiple sclerosis mouse model. *Nat. Neurosci.* **28**, 470–486 (2025).

203. Jin, L. *et al.* Astrocytic SARM1 promotes neuroinflammation and axonal demyelination in experimental autoimmune encephalomyelitis through inhibiting GDNF signaling. *Cell Death Dis.* **13**, (2022).

204. Lee, H. S. *et al.* Astrocytes contribute to gamma oscillations and recognition memory. *Proc. Natl. Acad. Sci. U. S. A.* **111**, (2014).

205. Lima, A. *et al.* Astrocyte pathology in the prefrontal cortex impairs the cognitive function of rats. *Mol. Psychiatry* **19**, 834–841 (2014).

206. Jaffe, A. E. *et al.* Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat. Neurosci.* **21**, 1117–1125 (2018).

207. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science (80-.).* **362**, (2018).

208. Bernstein, H. G. *et al.* Glial cell deficits are a key feature of schizophrenia: implications for neuronal circuit maintenance and histological differentiation from classical neurodegeneration. *Mol. Psychiatry* **30**, 1102–1116 (2025).

209. Trindade, P. *et al.* Induced pluripotent stem cell-derived astrocytes from patients with schizophrenia exhibit an inflammatory phenotype that affects vascularization. *Mol. Psychiatry* **28**, 871–882 (2023).

210. Magee, J. C. & Grienberger, C. Synaptic Plasticity Forms and Functions. *Annu. Rev. Neurosci.* **43**, 95–117 (2020).

211. Allen, N. J. & Lyons, D. A. Glia as architects of central nervous system formation and function. *Science (80-.).* **362**, 181–185 (2018).

212. Santello, M., Toni, N. & Volterra, A. Astrocyte function from information processing to cognition and cognitive impairment. *Nat. Neurosci.* **22**, 154–166 (2019).

213. Lüscher, C. & Malenka, R. C. NMDA Receptor-Dependent Long-Term Potentiation and Long-Term Depression (LTP / LTD). *Cold Spring Harb. Perspect. Biol.* **4**, 1–16 (2012).

214. Greger, I. H., Watson, J. F. & Cull-Candy, S. G. Structural and Functional Architecture of AMPA-Type Glutamate Receptors and Their Auxiliary Proteins. *Neuron* **94**, 713–730 (2017).

215. Descalzi, G., Gao, V., Steinman, M. Q., Suzuki, A. & Alberini, C. M. Lactate from astrocytes fuels learning-induced mRNA translation in excitatory and inhibitory neurons. *Commun. Biol.* **2**, (2019).

216. Inglebert, Y., Aljadeff, J., Brunel, N. & Debanne, D. Synaptic plasticity rules with physiological calcium levels. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 33639–33648 (2020).

217. Lisman, J. A mechanism for the Hebb and the anti-Hebb processes underlying learning and memory. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 9574–9578 (1989).

218. Morquette, P. *et al.* An astrocyte-dependent mechanism for neuronal rhythmogenesis. *Nat. Neurosci.* **18**, 844–854 (2015).

219. Zhang, Z. *et al.* The Appropriate Marker for Astrocytes: Comparing the Distribution and Expression of Three Astrocytic Markers in Different Mouse Cerebral Regions. *Biomed Res. Int.* **2019**, (2019).

220. Nishiyama, H., Knöpfel, T., Endo, S. & Itohara, S. Glial protein S100B modulates long-term neuronal synaptic plasticity. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 4037–4042 (2002).

221. Brockett, A. T. *et al.* Evidence supporting a role for astrocytes in the regulation of cognitive flexibility and neuronal oscillations through the Ca2+ binding protein S100β. *PLoS One* **13**, 1–19 (2018).

222. Abreu, D. S. *et al.* Astrocytes control hippocampal synaptic plasticity through the vesicular-dependent release of D-serine. *Front. Cell. Neurosci.* **17**, (2023).

223. Chen, A. B. *et al.* Norepinephrine changes behavioral state through astroglial purinergic signaling. *Science (80-.).* **388**, 769–775 (2025).

224. Refaeli, R., Kreisel, T., Yaish, T. R., Groysman, M. & Goshen, I. Astrocytes control recent and remote memory strength by affecting the recruitment of the CA1→ACC projection to engrams. *Cell Rep.* **43**, 113943 (2024).

225. Adamsky, A. *et al.* Astrocytic Activation Generates De Novo Neuronal Potentiation and Memory Enhancement. *Cell* **174**, 59-71.e14 (2018).

226. Kol, A. *et al.* Astrocytes contribute to remote memory formation by modulating hippocampal–cortical communication during learning. *Nat. Neurosci.* **23**, 1229–1239

(2020).

227. Sun, W. *et al.* Spatial transcriptomics reveal neuron–astrocyte synergy in long-term memory. *Nature* **627**, 374–381 (2024).

228. Yang, Y. & Jackson, R. Astrocyte identity: evolutionary perspectives on astrocyte functions and heterogeneity. *Curr. Opin. Neurobiol.* **56**, 40–46 (2019).

229. Falcone, C. Evolution of astrocytes: From invertebrates to vertebrates. *Frontiers in Cell and Developmental Biology* vol. 10 (2022).

230. Verkhratsky, A., Arranz, A. M., Ciuba, K. & Pękowska, A. Evolution of neuroglia. *Ann. N. Y. Acad. Sci.* **1518**, 120–130 (2022).

231. Kálmán, M., Matuz, V., Sebők, O. M. & Lőrincz, D. Evolutionary Modifications Are Moderate in the Astroglial System of Actinopterygii as Revealed by GFAP Immunohistochemistry. *Front. Neuroanat.* **15**, 1–21 (2021).

232. Morizet, D., Foucher, I., Alunni, A. & Bally-Cuif, L. Reconstruction of macroglia and adult neurogenesis evolution through cross-species single-cell transcriptomic analyses. *Nat. Commun.* **15**, 1–15 (2024).

233. Lamanna, F. *et al.* A lamprey neural cell type atlas illuminates the origins of the vertebrate brain. *Nat. Ecol. Evol.* **7**, 1714–1728 (2023).

234. Grupp, L., Wolburg, H. & Mack, A. F. Astroglial structures in the zebrafish brain. *J. Comp. Neurol.* **518**, 4277–4287 (2010).

235. Baumgart, E. V., Barbosa, J. S., Bally-cuif, L., Götz, M. & Ninkovic, J. Stab wound injury of the zebrafish telencephalon: A model for comparative analysis of reactive gliosis. *Glia* **60**, 343–357 (2012).

236. Ari, C. & Kálmán, M. Evolutionary changes of astroglia in Elasmobranchii comparing to Amniotes: A study based on three immunohistochemical markers (GFAP, S-100, and glutamine synthetase). *Brain. Behav. Evol.* **71**, 305–324 (2008).

237. Cano, A. C. S. S., Santos, D. & Beltrão-Braga, P. C. B. *Astrocyte-Neuron Interactions in Health and Disease. Advances in Neurobiology* vol. 39 (Springer Nature Switzerland, 2024).

238. King, J. S. A comparative investigation of neuroglia in representative vertebrates: A silver carbonate study. *J. Morphol.* **119**, 435–465 (1966).

239. Lambeth, L. & Blunt, M. J. Electron-microscopic study of monotreme neuroglia. *Cells*

Tissues Organs **93**, 115–125 (1975).

240. López-Hidalgo, M., Hoover, W. B. & Schummers, J. Spatial organization of astrocytes in ferret visual cortex. *J. Comp. Neurol.* **524**, 3561–3576 (2016).

241. Falcone, C. *et al.* Cortical Interlaminar Astrocytes Are Generated Prenatally, Mature Postnatally, and Express Unique Markers in Human and Nonhuman Primates. *Cereb. Cortex* **31**, 379–395 (2021).

242. Colombo, J. A. & Reisin, H. D. Interlaminar astroglia of the cerebral cortex: A marker of the primate brain. *Brain Res.* **1006**, 126–131 (2004).

243. Oberheim, N. A. *et al.* Uniquely hominid features of adult human astrocytes. *J. Neurosci.* **29**, 3276–3287 (2009).

244. Oberheim Bush, N. A. & Nedergaard, M. Do Evolutionary Changes in Astrocytes Contribute to the Computational Power of the Hominid Brain? *Neurochem. Res.* **42**, 2577–2587 (2017).

245. Falcone, C. & Martínez-Cerdeño, V. Astrocyte evolution and human specificity. *Neural Regen. Res.* **18**, 131–132 (2023).

246. Khrameeva, E. *et al.* Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Res.* **30**, 776–789 (2020).

247. Kanton, S. *et al. Organoid single-cell genomic atlas uncovers human-specific features of brain development*. *Nature* vol. 574 (2019).

248. Jorstad, N. L. *et al.* Comparative transcriptomics reveals human-specific cortical features. *Science (80-.).* **382**, (2023).

249. Ma, S. *et al.* Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science (80-.).* **377**, (2022).

250. Han, X. *et al.* Forebrain engraftment by human glial progenitor cells enhances synaptic plasticity and learning in adult mice. *Cell Stem Cell* **12**, 342–353 (2013).

251. Enard, W. Comparative genomics of brain size evolution. *Front. Hum. Neurosci.* **8**, 1–5 (2014).

252. Florio, M. *et al.* Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science (80-.).* **347**, 1465–1470 (2015).

253. Fiddes, I. T. *et al.* Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis. *Cell* **173**, 1356-1369.e22 (2018).

254. Fossati, M. *et al.* SRGAP2 and Its Human-Specific Paralog Co-Regulate the Development of Excitatory and Inhibitory Synapses. *Neuron* **91**, 356–369 (2016).

255. Ju, X. C. *et al.* The hominoid-specific gene TBC1D3 promotes generation of basal neural progenitors and induces cortical folding in mice. *Elife* **5**, 1–25 (2016).

256. Duński, E. & Pękowska, A. Keeping the balance: Trade-offs between human brain evolution, autism, and schizophrenia. *Front. Genet.* **13**, 1–14 (2022).

257. Espinós, A., Fernández-Ortuño, E., Negri, E. & Borrell, V. Evolution of genetic mechanisms regulating cortical neurogenesis. *Dev. Neurobiol.* **82**, 428–453 (2022).

258. Bae, B. Il, Jayaraman, D. & Walsh, C. A. Genetic changes shaping the human brain. *Dev. Cell* **32**, 423–434 (2015).

259. Liu, J., Mosti, F. & Silver, D. L. Human brain evolution: Emerging roles for regulatory DNA and RNA. *Curr. Opin. Neurobiol.* **71**, 170–177 (2021).

260. Staes, N. *et al.* FOXP2 variation in great ape populations offers insight into the evolution of communication skills. *Sci. Rep.* **7**, 1–10 (2017).

261. Prescott, S. L. *et al.* Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp Neural Crest. *Cell* **163**, 68–83 (2015).

262. Wang, Y. *et al.* Accelerated evolution of an Lhx2 enhancer shapes mammalian social hierarchies. *Cell Res.* **30**, 408–420 (2020).

263. Dumas, G., Malesys, S. & Bourgeron, T. Systematic detection of brain protein-coding genes under positive selection during primate evolution and their roles in cognition. *Genome Res.* **31**, 484–496 (2021).

264. King, M.-C. & Wilson, A. C. Evolution at Two Levels in Humans and Chimpanzees. *Science (80-.).* **188**, 107–116 (1975).

265. Rubinstein, M. & de Souza, F. S. J. Evolution of transcriptional enhancers and animal diversity. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, (2013).

266. Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).

267. Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science (80-.).* **328**, 1036–1040 (2010).

268. Konopka, G. *et al.* Human-Specific Transcriptional Networks in the Brain. *Neuron* **75**, 601–617 (2012).

269. Bush, E. C. & Lahn, B. T. A genome-wide screen for noncoding elements important in primate evolution. *BMC Evol. Biol.* **8**, 1–10 (2008).

270. Prabhakar, S., Noonan, J. P., Pääbo, S. & Rubin, E. M. Accelerated evolution of conserved noncoding sequences in humans. *Science (80-.).* **314**, 786 (2006).

271. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).

272. Mangan, R. J. *et al.* Adaptive sequence divergence forged new neurodevelopmental enhancers in humans. *Cell* **185**, 4587-4603.e23 (2022).

273. Reilly, S. K. *et al.* Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science (80-.).* **347**, 1155–1159 (2015).

274. Chatterjee, S. & Ahituv, N. Gene Regulatory Elements, Major Drivers of Human Disease. *Annu. Rev. Genomics Hum. Genet.* **18**, 45–63 (2017).

275. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482–485 (2000).

276. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

277. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci.* **107**, 21931–21936 (2010).

278. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–285 (2011).

279. Buecker, C. & Wysocka, J. Enhancers as information integration hubs in development: Lessons from genomics. *Trends Genet.* **28**, 276–284 (2012).

280. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170–1187 (2016).

281. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science (80-.).* **326**, 289–293 (2009).

282. Pombo, A. & Dillon, N. Three-dimensional genome architecture: Players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **16**, 245–257 (2015).

283. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

284. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre `. *Nature* 381–385 (2012) doi:10.1038/nature11049.

285. Krijger, P. H. L. & De Laat, W. Regulation of disease-associated gene expression in the 3D genome. *Nat. Rev. Mol. Cell Biol.* **17**, 771–782 (2016).

286. Ciuba, K. *et al.* Molecular signature of primate astrocytes reveals pathways and regulatory changes contributing to human brain evolution. *Cell Stem Cell* **32**, 426-444.e14 (2025).

287. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).

288. Rowe, R. G. & Daley, G. Q. Induced pluripotent stem cells in disease modelling and drug discovery. *Nat. Rev. Genet.* **20**, 377–388 (2019).

289. CITES Secretariat. *CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) handbook*. (2001).

290. Kampmann, M. CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine. *ACS Chem. Biol.* **13**, 406–416 (2018).

291. Leng, K. *et al.* CRISPRi screens in human iPSC-derived astrocytes elucidate regulators of distinct inflammatory reactive states. *Nat. Neurosci.* **25**, 1528–1542 (2022).

292. Li, S., Hannenhalli, S. & Ovcharenko, I. De novo human brain enhancers created by single-nucleotide mutations. *Sci. Adv.* **9**, (2023).

293. Girskis, K. M. *et al.* Rewiring of human neurodevelopmental gene regulatory programs by human accelerated regions. *Neuron* **109**, 3239-3251.e7 (2021).

294. Inoue, F., Kreimer, A., Ashuach, T., Ahituv, N. & Yosef, N. Identification and Massively Parallel Characterization of Regulatory Elements Driving Neural Induction. *Cell Stem Cell* **25**, 713-727.e10 (2019).

295. Whalen, S. *et al.* Machine learning dissection of human accelerated regions in primate neurodevelopment. *Neuron* **111**, 1–17 (2023).

296. Qi, L. S. *et al.* Resource Repurposing CRISPR as an RNA-Guided Platform for

Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).

297. Margolin, J. F. *et al.* Krüppel-associated boxes are potent transcriptional repression domains. *Proc. Natl. Acad. Sci.* **91**, 4509–4513 (1994).

298. Yeo, N. C. *et al.* An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat. Methods* **15**, 611–616 (2018).

299. Xu, J. *et al.* Subtype-specific 3D genome alteration in acute myeloid leukaemia Check for updates. *Nature* **611**, 1–12 (2022).

300. Moonen, J. *et al.* KLF4 recruits SWI/SNF to increase chromatin accessibility and reprogram the endothelial enhancer landscape under laminar shear stress. *Nat. Commun.* **13**, 4941 (2022).

301. Zhang, N. *et al.* Muscle progenitor specification and myogenic differentiation are associated with changes in chromatin topology. *Nat. Commun.* **11**, (2020).

302. Liszewska, E. *et al.* Establishment of two hiPSC lines (IIMCBi001-A and IIMCBi002-A) from dermal fibroblasts of healthy donors and characterization of their cell cycle. *Stem Cell Res.* **52**, 1–5 (2021).

303. Brennand, K. J. *et al.* Modelling schizophrenia using human induced pluripotent stem cells. *Nature* **473**, 221–225 (2011).

304. Horlbeck, M. A. *et al.* Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* **5**, 1–20 (2016).

305. Concordet, J. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018).

306. Yao, D. *et al.* Multicenter integrated analysis of noncoding CRISPRi screens. *Nat. Methods* **21**, 723–734 (2024).

307. Pękowska, A. *et al.* Gain of CTCF-Anchored Chromatin Loops Marks the Exit from Naive Pluripotency. *Cell Syst.* **7**, 482-495.e10 (2018).

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

309. Xu, S. *et al.* Using clusterProfiler to characterize multiomics data. *Nat. Protoc.* **19**, 3292–3320 (2024).

310. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).

311. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

312. Zhang, Y. *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89**, 37–53 (2016).

313. TCW, J. *et al.* An Efficient Platform for Astrocyte Differentiation from Human Induced Pluripotent Stem Cells. *Stem Cell Reports* **9**, 600–614 (2017).

314. Krueger, F. *et al.* FelixKrueger/TrimGalore: v0.6.10 - add default decompression path (0.6.10). (2023) doi:https://doi.org/10.5281/zenodo.7598955.

315. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, (2009).

316. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

317. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, (2008).

318. Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* **3**, 95–98 (2016).

319. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

320. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).

321. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

322. Vian, L. *et al.* The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* **173**, 1165-1178.e20 (2018).

323. Kulakovskiy, I. V. *et al.* HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018).

324. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

325. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

326. He, Z. *et al.* Comprehensive transcriptome analysis of neocortical layers in humans, chimpanzees and macaques. *Nat. Neurosci.* **20**, 886–895 (2017).

327. He, Z., Bammann, H., Han, D., Xie, G. & Khaitovich, P. Conserved expression of lincRNA during human and macaque prefrontal cortex development and maturation. *Rna* **20**, 1103–1111 (2014).

328. Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).

329. Magoč, T. & Salzberg, S. L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

330. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. **00**, 1–3 (2013).

331. Tewhey, R. *et al.* Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell* **165**, 1519–1529 (2016).

332. Hon, J., Martínek, T., Zendulka, J. & Lexa, M. pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. *Bioinformatics* **33**, 3373–3379 (2017).

333. Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C. & Hochreiter, S. Msa: An R package for multiple sequence alignment. *Bioinformatics* **31**, 3997–3999 (2015).

334. Álvarez-Viejo, M. CD271 as a marker to identify mesenchymal stem cells from diverse sources before culture. *World J. Stem Cells* **7**, 470 (2015).

335. Yuan, S. H. *et al.* Cell-surface marker signatures for the Isolation of neural stem cells, glia and neurons derived from human pluripotent stem cells. *PLoS One* **6**, (2011).

336. Townsley, K. G., Brennand, K. J. & Huckins, L. M. Massively parallel techniques for cataloguing the regulome of the human brain. *Nat. Neurosci.* **23**, 1509–1521 (2020).

337. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science (80-.).* **337**, 1190–1195 (2012).

338. Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *Am. J. Hum. Genet.* **102**, 717–730 (2018).

339. Nott, A. *et al.* Brain cell type–specific enhancer–promoter interactome maps and

disease-risk association. *Science (80-.).* **366**, 1134–1139 (2019).

340. Liang, G. *et al.* Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7357–7362 (2004).

341. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).

342. Tiwari, N. *et al.* Stage-Specific Transcription Factors Drive Astrogliogenesis by Remodeling Gene Regulatory Landscapes. *Cell Stem Cell* **23**, 557-571.e8 (2018).

343. Malik, N. *et al.* Comparison of the gene expression profiles of human fetal cortical astrocytes with pluripotent stem cell derived neural stem cells identifies human astrocyte markers and signaling pathways and transcription factors active in human astrocytes. *PLoS One* **9**, 1–16 (2014).

344. Quist, E. *et al.* Transcription factor-based direct conversion of human fibroblasts to functional astrocytes. *Stem Cell Reports* **17**, 1620–1635 (2022).

345. Song, M. *et al.* Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat. Genet.* **51**, 1252–1262 (2019).

346. Bailey, T. L. & Grant, C. E. SEA: Simple Enrichment Analysis of motifs. *bioRxiv* 2021.08.23.457422 (2021).

347. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).

348. Madrigal, P. AP-1 Takes Centre Stage in Enhancer Chromatin Dynamics. *Trends Cell Biol.* **28**, 509–511 (2018).

349. Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* **51**, D977–D985 (2023).

350. van der Meer, D. *et al.* Understanding the genetic determinants of the brain with MOSTest. *Nat. Commun.* **11**, 1–9 (2020).

351. van der Meer, D. *et al.* van der Meer et al The genetic architecture of human cortical folding. *Sci. Adv* **7**, 9446 (2021).

352. Fan, C. C. *et al.* Multivariate genome-wide association study on tissue-sensitive diffusion metrics highlights pathways that shape the human brain. *Nat. Commun.* **13**,

2423 (2022).

353. Bojarskaite, L. *et al.* Sleep cycle-dependent vascular dynamics in male mice and the predicted effects on perivascular cerebrospinal fluid flow and solute transport. *Nat. Commun.* **14**, (2023).

354. Ringstad, G. *et al.* Brain-wide glymphatic enhancement and clearance in humans assessed with MRI. *JCI insight* **3**, (2018).

355. Garel, C. *et al.* Fetal cerebral cortex: Normal gestational landmarks identified using prenatal MR imaging. *Am. J. Neuroradiol.* **22**, 184–189 (2001).

356. Dubois, J. *et al.* Primary cortical folding in the human newborn: An early marker of later functional development. *Brain* **131**, 2028–2041 (2008).

357. Nishikuni, K. & Ribas, G. C. Study of fetal and postnatal morphological development of the brain sulci: Laboratory investigation. *J. Neurosurg. Pediatr.* **11**, 1–11 (2013).

358. Zhang, Y. *et al.* Cortical structural abnormalities in very preterm children at 7years of age. *Neuroimage* **109**, 469–479 (2015).

359. Papini, C. *et al.* Altered Cortical Gyrification in Adults Who Were Born Very Preterm and Its Associations With Cognition and Mental Health. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **5**, 640–650 (2020).

360. Rash, B. G. *et al.* Gliogenesis in the outer subventricular zone promotes enlargement and gyrification of the primate cerebrum. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7089–7094 (2019).

361. Shinmyo, Y. *et al.* Localized astrogenesis regulates gyrification of the cerebral cortex. *Sci. Adv.* **8**, 1–14 (2022).

362. Fitzallen, G. C., Taylor, H. G. & Bora, S. What Do We Know About the Preterm Behavioral Phenotype? A Narrative Review. *Front. Psychiatry* **11**, 1–9 (2020).

363. Sejer, E. P. F. *et al.* Preterm birth and subsequent intelligence and academic performance in youth: A cohort study. *Acta Obstet. Gynecol. Scand.* **103**, 850–861 (2024).

364. Penttilä, J. *et al.* Cortical folding in patients with bipolar disorder or unipolar depression. *J. Psychiatry Neurosci.* **34**, 127–135 (2009).

365. Cachia, A. *et al.* Cortical folding abnormalities in schizophrenia patients with resistant auditory hallucinations. *Neuroimage* **39**, 927–935 (2008).

366. Nordahl, C. W. *et al.* Cortical folding abnormalities in autism revealed by surface-based morphometry. *J. Neurosci.* **27**, 11725–11735 (2007).

367. Pereira, J. B. *et al.* Assessment of cortical degeneration in patients with Parkinson's disease by voxel-based morphometry, cortical folding, and cortical thickness. *Hum. Brain Mapp.* **33**, 2521–2534 (2012).

368. Wainberg, M. *et al.* Genetic architecture of the structural connectome. *Nat. Commun.* 1–20 (2024) doi:10.1038/s41467-024-46023-2.

369. Goshima, Y., Nakamura, F., Strittmatter, P. & Strittmatter, S. M. Collapsin-induced growth cone collapse mediated by an intracellular protein related to UNC-33. *Nature* **376**, 509–514 (1995).

370. Yoshimura, T. *et al.* GSK-3β Regulates Phosphorylation of CRMP-2 and Neuronal Polarity. *Cell* **120**, 137–149 (2005).

371. Yamazaki, Y. *et al.* Requirement of CRMP2 Phosphorylation in Neuronal Migration of Developing Mouse Cerebral Cortex and Hippocampus and Redundant Roles of CRMP1 and CRMP4. *Cereb. Cortex* **32**, 520–527 (2022).

372. Ziak, J. *et al.* CRMP 2 mediates Sema3F-dependent axon pruning and dendritic spine remodeling . *EMBO Rep.* **21**, 1–22 (2020).

373. Feuer, K. L., Peng, X., Yovo, C. K. & Avramopoulos, D. DPYSL2/CRMP2 isoform B knockout in human iPSC-derived glutamatergic neurons confirms its role in mTOR signaling and neurodevelopmental disorders. *Mol. Psychiatry* **28**, 4353–4362 (2023).

374. Zhang, H. *et al.* Brain-specific Crmp2 deletion leads to neuronal development deficits and behavioural impairments in mice. *Nat. Commun.* **7**, (2016).

375. Izumi, R. *et al.* Dysregulation of DPYSL2 expression by mTOR signaling in schizophrenia: Multi-level study of postmortem brain. *Neurosci. Res.* **175**, 73–81 (2022).

376. Naqvi, S. *et al.* Shared heritability of human face and brain shape. *Nat. Genet.* **53**, 830–839 (2021).

377. Shadrin, A. A. *et al.* Vertex-wise multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology. *Neuroimage* **244**, 118603 (2021).

378. Li, W. *et al.* Suppression of Glial Activation in Tau Transgenic Mice Through Inhibition of CRMP2 Phosphorylation: a Morphometric Analysis. *NeuroMolecular Med.* **27**, 70

(2025).

379. Andersen, R. E. *et al.* Short Article The Long Noncoding RNA Pnky Is a Trans -acting Regulator of Cortical Development In Vivo Short Article The Long Noncoding RNA Pnky Is a Trans -acting Regulator of Cortical Development In Vivo. *Dev. Cell* **49**, 632-642.e7 (2019).

380. Kasher, P. R. *et al.* Small 6q16 . 1 Deletions Encompassing POU3F2 Cause Susceptibility to Obesity and Variable Developmental Delay with Intellectual Disability. *Am. J. Hum. Genet.* **98**, 363–372 (2016).

381. Zhu, X. *et al.* BRN2 as a key gene drives the early primate telencephalon development. *Sci. Adv.* **8**, 1–16 (2022).

382. Saini, N., Vijayasree, V., Nandury, E. C., Dalal, A. & Aggarwal, S. Prenatal phenotype of FBXL4-associated encephalomyopathic mitochondrial DNA depletion syndrome-13. *Prenat. Diagn.* **42**, 1682–1685 (2022).

383. Westphal, M., Hänsel, M., Hamel, W., Kunzmann, R. & Hölzel, F. Karyotype analyses of 20 human glioma cell lines. *Acta Neurochir. (Wien).* **126**, 17–26 (1994).

384. Ding, C. *et al.* Transcription factor POU3F2 regulates TRIM8 expression contributing to cellular functions implicated in schizophrenia. *Mol. Psychiatry* **26**, 3444–3460 (2020).

385. Nasu, M., Anan, K., Abe, Y., Kozuki, N. & Matsushima, A. Biochemical and Biophysical Research Communications Reduced home cage and social activity in Pou3f2 ⊿ mice. *Biochem. Biophys. Res. Commun.* **523**, 411–415 (2020).

386. Pearl, J. R. *et al.* Genome-Scale Transcriptional Regulatory Network Models of Psychiatric and Neurodegenerative Article Genome-Scale Transcriptional Regulatory Network Models of Psychiatric and Neurodegenerative Disorders. *Cell Syst.* **8**, 122-135.e7 (2019).

387. Chen, C. *et al.* The transcription factor POU3F2 regulates a gene coexpression network in brain tissue from patients with psychiatric disorders. *Sci. Transl. Med.* **10**, 1–11 (2018).

388. Saha, P. *et al.* Sex-specific role for the long noncoding RNA Pnky in mouse behavior. *Nat. Commun.* **15**, (2024).

389. Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).

390. Pang, Z. P. *et al.* transcription factors. *Nature* **476**, 3–7 (2011).

391. Dominguez, M. H., Ayoub, A. E. & Rakic, P. POU-III Transcription Factors (Brn1, Brn2, and Oct6) Influence Neurogenesis, Molecular Identity, and Migratory Destination of Upper-Layer Cells of the Cerebral Cortex. *Cereb. Cortex* **23**, 2632–2643 (2013).

392. Hashizume, K., Yamanaka, M. & Ueda, S. POU3F2 participates in cognitive function and adult hippocampal neurogenesis via mammalian-characteristic amino acid repeats. *Genes, Brain Behav.* **17**, 118–125 (2018).

393. Jindal, G. A. & Farley, E. K. Enhancer grammar in development, evolution, and disease: dependencies and interplay. *Dev. Cell* **56**, 575–587 (2021).

394. Vue, T. Y., Kim, E. J., Parras, C. M., Guillemot, F. & Johnson, J. E. Ascl1 controls the number and distribution of astrocytes and oligodendrocytes in the gray matter and white matter of the spinal cord. *Development* **141**, 3721–3731 (2014).

395. Stipursky, J. & Gomes, F. C. A. TGF-β1/SMAD signaling induces astrocyte fate commitment in vitro : Implications for radial glia development. *Glia* **55**, 1023–1033 (2007).

396. Huang, Z. *et al.* Neogenin promotes BMP2 activation of YAP and smad1 and enhances astrocytic differentiation in developing mouse neocortex. *J. Neurosci.* **36**, 5833–5849 (2016).

397. Huang, Z. *et al.* YAP stabilizes SMAD1 and promotes BMP2-induced neocortical astrocytic differentiation. *Dev.* **143**, 2398–2409 (2016).

398. Pouso, M. R. & Cairrao, E. Effect of retinoic acid on the neurovascular unit: A review. *Brain Res. Bull.* **184**, 34–45 (2022).

399. Biyong, E. F. *et al.* Role of Retinoid X Receptors (RXRs) and dietary vitamin A in Alzheimer's disease: Evidence from clinicopathological and preclinical studies. *Neurobiol. Dis.* **161**, (2021).

400. Chan, T. J., Her, L. S., Liaw, H. J., Chen, M. C. & Tzeng, S. F. Retinoic acid mediates the expression of glutamate transporter-1 in rat astrocytes through genomic RXR action and non-genomic protein kinase C signaling pathway. *J. Neurochem.* **121**, 537–550 (2012).

401. Talarico, M. *et al.* RORA-neurodevelopmental disorder: A unique triad of developmental disabilities, cerebellar anomalies, and myoclonic seizures. *Genet. Med.* **27**, (2025).

402. Dewa, K.-I. *et al.* The astrocytic ensemble acts as a multiday trace to stabilize memory. *Nature* (2025) doi:10.1038/s41586-025-09619-2.

403. Mayer, S. I., Rössler, O. G., Endo, T., Charnay, P. & Thiel, G. Epidermal-growth-factor-induced proliferation of astrocytes requires Egr transcription factors. *J. Cell Sci.* **122**, 3340–3350 (2009).

404. Golson, M. L. & Kaestner, K. H. Fox transcription factors: From development to disease. *Dev.* **143**, 4558–4570 (2016).

405. Doan, K. Van *et al.* Astrocytic FoxO1 in the hypothalamus regulates metabolic homeostasis by coordinating neuropeptide Y neuron activity. *Glia* **71**, 2735–2752 (2023).

406. Usui, N. *et al.* Foxp1 regulation of neonatal vocalizations via cortical development. *Genes Dev.* **31**, 2039–2055 (2017).

407. Bacon, C. *et al.* Brain-specific Foxp1 deletion impairs neuronal development and causes autistic-like behaviour. *Mol. Psychiatry* **20**, 632–639 (2015).

408. Trelles, M. P. *et al.* Individuals with FOXP1 syndrome present with a complex neurobehavioral profile with high rates of ADHD, anxiety, repetitive behaviors, and sensory symptoms. *Mol. Autism* **12**, 1–15 (2021).

409. Liu, B. *et al.* MAZ mediates the cross-talk between CT-1 and NOTCH1 signaling during gliogenesis. *Nat. Publ. Gr.* 1–12 (2016) doi:10.1038/srep21534.

410. Xiao, T., Li, X. & Felsenfeld, G. The Myc-associated zinc finger protein epigenetically controls expression of interferon-γ-stimulated genes by recruiting STAT1 to chromatin. *Proc. Natl. Acad. Sci.* **121**, 2017 (2024).

411. Ortabozkoyun, H. *et al. CRISPR and biochemical screens identify MAZ as a cofactor in CTCF-mediated insulation at Hox clusters*. *Nature Genetics* vol. 54 (Springer US, 2022).

412. Yin, K. J., Hamblin, M., Fan, Y., Zhang, J. & Chen, Y. E. Krüpple-like factors in the central nervous system: novel mediators in Stroke. *Metab. Brain Dis.* **30**, 401–410 (2015).

413. Wang, C. & Li, L. The critical role of KLF4 in regulating the activation of A1/A2 reactive astrocytes following ischemic stroke. *J. Neuroinflammation* **20**, 1–12 (2023).

414. Park, J. H. *et al.* Induction of Krüppel-like factor 4 expression in reactive astrocytes following ischemic injury in vitro and in vivo. *Histochem. Cell Biol.* **141**, 33–42 (2014).

415. Kvon, E. Z. *et al.* Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* **167**, 633-642.e11 (2016).

416. Chan, Y. F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a pitxl enhancer. *Science (80-.).* **327**, 302–305 (2010).

417. Bird, C. P. *et al.* Fast-evolving noncoding sequences in the human genome. *Genome Biol.* **8**, 1–12 (2007).

418. Gittelman, R. M. *et al.* Comprehensive identification and analysis of human accelerated regulatory DNA. *Genome Res.* **25**, 1245–1255 (2015).

419. Hubisz, M. J. & Pollard, K. S. Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. *Curr. Opin. Genet. Dev.* **29**, 15–21 (2014).

420. Pollard, K. S. *et al.* An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006).

421. Pollard, K. S. *et al.* Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* **2**, 1599–1611 (2006).

422. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).

423. Paganoni, S., Anderson, K. L. & Ferreira, A. Differential subcellular localization of Ror tyrosine kinase receptors in cultured astrocytes. *Glia* **46**, 456–466 (2004).

424. Tanaka, Y., Minami, Y. & Endo, M. Ror1 promotes PPARα-mediated fatty acid metabolism in astrocytes. *Genes to Cells* **28**, 307–318 (2023).

425. Wilson, P. M., Fryer, R. H., Fang, Y. & Hatten, M. E. Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration. *J. Neurosci.* **30**, 8529–8540 (2010).

426. Lionel, A. C. *et al.* Disruption of the ASTN2/TRIM32 locus at 9q33.1 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. *Hum. Mol. Genet.* **23**, 2752–2768 (2014).

427. Gordon, M. G. *et al.* lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements. *Nat. Protoc.* **15**, 2387–2412 (2020).

428. Klein, J. C. *et al.* A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nat. Methods* **17**, 1083–1091 (2020).

429. Wang, H. *et al.* Connexin43 and Its Regulation of Astrocyte Gap Junction Function: Influencing Depression Progression by Mediating Electrical and Chemical Signals. *CNS Neurosci. Ther.* **31**, (2025).

430. Cruz-Mendoza, F., Jauregui-Huerta, F., Luquin, S., Aguilar-Delgadillo, A. & García-Estrada, J. Immediate Early Gene c-fos in the Brain: Focus on Glial Cells. *Brain Sci.* **12**, (2022).

431. Byrns, C. N., Saikumar, J. & Bonini, N. M. Glial AP1 is activated with aging and accelerated by traumatic brain injury. *Nat. Aging* **1**, 585–597 (2021).

432. Martin, F. J. *et al.* Ensembl 2023. *Nucleic Acids Res.* **51**, D933–D941 (2023).

433. Rasool, D. & Jahani-Asl, A. Master regulators of neurogenesis: the dynamic roles of Ephrin receptors across diverse cellular niches. *Transl. Psychiatry* **14**, 1–12 (2024).

434. Jen, Y. H. L., Musacchio, M. & Lander, A. D. Glypican-1 controls brain size through regulation of fibroblast growth factor signaling in early neurogenesis. *Neural Dev.* **4**, 1–19 (2009).

435. Ong, C. T. & Corces, V. G. CTCF: An architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15**, 234–246 (2014).

436. Carabalona, A. *et al.* Identification of novel human microcephaly-linked protein Mtss2 that mediates cortical progenitor cell division and corticogenesis through Nedd9-RhoA. *Elife* **13**, 1–23 (2025).

437. Nasu, M. *et al.* Deficient maternal behavior in multiparous Pou3f2⊿ mice is associated with an impaired exploratory activity. *Behav. Brain Res.* **427**, 113846 (2022).

438. Chen, C. *et al.* Astrocyte-derived clusterin disrupts glial physiology to obstruct remyelination in mouse models of demyelinating diseases. *Nat. Commun.* **15**, (2024).

439. Chen, F. *et al.* Clusterin secreted from astrocyte promotes excitatory synaptic transmission and ameliorates Alzheimer's disease neuropathology. *Mol. Neurodegener.* **16**, 1–16 (2021).

440. Lish, A. M. *et al.* CLU alleviates Alzheimer's disease-relevant processes by modulating astrocyte reactivity and microglia-dependent synaptic density. *Neuron* **113**, 1925-1946.e11 (2025).

441. Takla, T. N. *et al.* A Shared Pathogenic Mechanism for Valproic Acid and SHROOM3 Knockout in a Brain Organoid Model of Neural Tube Defects. *Cells* **12**, (2023).

442. Oksenberg, N. & Ahituv, N. The role of AUTS2 in neurodevelopment and human evolution. *Trends Genet.* **29**, 600–608 (2013).

443. Oksenberg, N., Stevison, L., Wall, J. D. & Ahituv, N. Function and Regulation of AUTS2, a Gene Implicated in Autism and Human Evolution. *PLoS Genet.* **9**, (2013).

444. Preman, P. *et al.* APOE from astrocytes restores Alzheimer's Aβ-pathology and DAM-like responses in APOE deficient microglia. *EMBO Mol. Med.* **16**, 3113–3141 (2024).

445. Golden, L. R. *et al.* APOE4 to APOE2 'Switching' in Astrocytes Alters the Cerebral Transcriptome and Decreases AD Neuropathology at Different Points on the AD Pathophysiological Timeline . *Alzheimer's Dement.* **20**, 1–2 (2024).

446. Saad, B. *et al.* Astrocyte-derived TGF-β2 and NGF differentially regulate neural recognition molecule expression by cultured astrocytes. *J. Cell Biol.* **115**, 473–484 (1991).

447. Fuchshofer, R., Birke, M., Welge-Lussen, U., Kook, D. & Lütjen-Drecoll, E. Transforming growth factor-β2 modulated extracellular matrix component expression in cultured human optic nerve head astrocytes. *Investig. Ophthalmol. Vis. Sci.* **46**, 568–578 (2005).

448. Fukushima, T., Liu, R. & Byrne, J. H. Transforming growth factor-β2 modulates synaptic efficacy and plasticity and induces phosphorylation of CREB in hippocampal neurons. *Hippocampus* **17**, 5–9 (2007).

449. Heupel, K. *et al.* Loss of transforming growth factor-beta 2 leads to impairment of central synapse function. *Neural Dev.* **3**, 1–16 (2008).

450. Szewczyk, L. M. *et al.* Astrocytic β-catenin signaling via TCF7L2 regulates synapse development and social behavior. *Mol. Psychiatry* **29**, 57–73 (2024).

451. Zhao, L. *et al.* IGFBP7+ subpopulation and IGFBP7 risk score in astrocytoma: insights from scRNA-Seq and bulk RNA-Seq. *Front. Immunol.* **15**, 1–24 (2024).

452. Li, N., Han, J., Tang, J. & Ying, Y. IGFBP-7 inhibits the differentiation of oligodendrocyte precursor cells via regulation of Wnt/β-Catenin signaling. *J. Cell. Biochem.* **119**, 4742–4750 (2018).

453. Sjöstedt, E. *et al.* An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science (80-.).* **367**, (2020).

454. Cruchaga, C. *et al.* GWAS of cerebrospinal fluid tau levels identifies risk variants for alzheimer's disease. *Neuron* **78**, 256–268 (2013).

455. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser - A database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, 88–92 (2007).

456. Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* **48**, D882–D889 (2020).

457. Hitz, B. C. *et al.* Analysis Pipelines. (2023).

458. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

459. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).

460. Kuhn, R. M., Haussler, D. & James Kent, W. The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–161 (2013).

461. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **21**, 459–474 (2020).

462. Chen, K. Z. *et al.* Vimentin as a potential target for diverse nervous system diseases. *Neural Regeneration Research* vol. 18 969–975 (2023).

463. Josselyn, S. A. *et al.* Long-term memory is facilitated by cAMP response element-binding protein overexpression in the amygdala. *J. Neurosci.* **21**, 2404–2412 (2001).

464. Bourtchuladze, R. *et al.* Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. *Cell* **79**, 59–68 (1994).

465. Pérez-Grijalba, V. *et al.* New insights into genetic variant spectrum and genotype–phenotype correlations of Rubinstein-Taybi syndrome in 39 CREBBP- positive patients. *Mol. Genet. Genomic Med.* **7**, 1–13 (2019).

466. Neve, L. D., Savage, A. A., Koke, J. R. & García, D. M. Activating transcription factor 3 and reactive astrocytes following optic nerve injury in zebrafish. *Comp. Biochem. Physiol. - C Toxicol. Pharmacol.* **155**, 213–218 (2012).

467. Liu, S. *et al.* The Dual Roles of Activating Transcription Factor 3 (ATF3) in Inflammation, Apoptosis, Ferroptosis, and Pathogen Infection Responses. *Int. J. Mol. Sci.* **25**, (2024).

468. Kim, K. H., Jeong, J. Y., Surh, Y. J. & Kim, K. W. Expression of stress-response ATF3 is mediated by Nrf2 in astrocytes. *Nucleic Acids Res.* **38**, 48–59 (2009).

469. Pai, C. S. *et al.* The activating transcription factor 3 (Atf3) homozygous knockout mice exhibit enhanced conditioned fear and down regulation of hippocampal GELSOLIN. *Front. Mol. Neurosci.* **11**, 1–14 (2018).

470. Green, T. A. *et al.* Induction of activating transcription factors (ATFs) ATF2, ATF3, and ATF4 in the nucleus accumbens and their regulation of emotional behavior. *J. Neurosci.* **28**, 2025–2032 (2008).

471. Lax, E. *et al.* Methyl-CpG binding domain 2 (Mbd2) is an epigenetic regulator of autism-risk genes and cognition. *Transl. Psychiatry* **13**, (2023).

472. Mojsilovic-Petrovic, J. *et al.* Hypoxia-inducible factor-1 (HIF-1) is involved in the regulation of hypoxia-stimulated expression of monocyte chemoattractant protein-1 (MCP-1/CCL2) and MCP-5 (Ccl12) in astrocytes. *J. Neuroinflammation* **4**, 1–15 (2007).

473. Linnerbauer, M. *et al.* The astrocyte-produced growth factor HB-EGF limits autoimmune CNS pathology. *Nat. Immunol.* **25**, 432–447 (2024).

474. Rees, K. A. *et al.* Molecular, physiological and behavioral characterization of the heterozygous Df[h15q13]/+ mouse model associated with the human 15q13.3 microdeletion syndrome. *Brain Res.* **1746**, 147024 (2020).

475. Malwade, S. *et al.* Identification of Vulnerable Interneuron Subtypes in 15q13.3 Microdeletion Syndrome Using Single-Cell Transcriptomics. *Biol. Psychiatry* **91**, 727–739 (2022).

476. Ziats, M. N. *et al.* The complex behavioral phenotype of 15q13.3 microdeletion syndrome. *Genet. Med.* **18**, 1111–1118 (2016).

477. Arneson, A. & Ernst, J. Systematic discovery of conservation states for single-nucleotide annotation of the human genome. *Commun. Biol.* **2**, 248 (2019).

478. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).

479. Long, M., Kranjc, T., Mysior, M. M. & Simpson, J. C. RNA Interference Screening Identifies Novel Roles for RhoBTB1 and RhoBTB3 in Membrane Trafficking Events in Mammalian Cells. *Cells* **9**, (2020).

480. Ji, W. & Rivero, F. Atypical Rho GTPases of the RhoBTB Subfamily: Roles in Vesicle Trafficking and Tumorigenesis. *Cells* **5**, 28 (2016).

481. Lu, A. & Pfeffer, S. R. Golgi-associated RhoBTB3 targets cyclin E for ubiquitylation

and promotes cell cycle progression. *J. Cell Biol.* **203**, 233–250 (2013).

482. Schenková, K., Lutz, J., Kopp, M., Ramos, S. & Rivero, F. MUF1/leucine-rich repeat containing 41 (LRRC41), a substrate of RhoBTB-dependent cullin 3 ubiquitin ligase complexes, is a predominantly nuclear dimeric protein. *J. Mol. Biol.* **422**, 659–673 (2012).

483. Berthold, J. *et al.* Characterization of RhoBTB-dependent Cul3 ubiquitin ligase complexes - Evidence for an autoregulatory mechanism. *Exp. Cell Res.* **314**, 3453–3465 (2008).

484. Zhang, C. S. *et al.* RHOBTB3 promotes proteasomal degradation of HIFα through facilitating hydroxylation and suppresses the Warburg effect. *Cell Res.* **25**, 1025–1042 (2015).

485. Miller, J. A., Woltjer, R. L., Goodenbour, J. M., Horvath, S. & Geschwind, D. H. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Med.* **5**, (2013).

486. Corradin, O. & Scacheri, P. C. Enhancer variants: Evaluating functions in common disease. *Genome Med.* **6**, 1–14 (2014).

487. Birney, E. *et al.* An Overview of Ensembl. *Genome Res.* **14**, 925–928 (2004).

488. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).

489. Satizabal, C. L. *et al.* Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat. Genet.* **51**, 1624–1636 (2019).

490. Nakamura, M. *et al.* Ser14 phosphorylation of Bcl-xL mediates compensatory cardiac hypertrophy in male mice. *Nat. Commun.* **14**, (2023).

491. Parsadanian, A. S., Cheng, Y., Keller-Peck, C. R., Holtzman, D. M. & Snider, W. D. Bcl-x(L) is an antiapoptotic regulator for postnatal CNS neurons. *J. Neurosci.* **18**, 1009–1019 (1998).

492. Xu, L., Lee, J. E. & Giffard, R. G. Overexpression of bcl-2, bcl-x(L) or hsp70 in murine cortical astrocytes reduces injury of co-cultured neurons. *Neurosci. Lett.* **277**, 193–197 (1999).

493. Nguyen, T. T. M. *et al.* Mitochondrial Bcl-xL promotes brain synaptogenesis by controlling non-lethal caspase activation. *iScience* **26**, (2023).

494. Nakamura, A. *et al.* Bcl-xL is essential for the survival and function of differentiated neurons in the cortex that control complex behaviors. *J. Neurosci.* **36**, 5448–5461 (2016).

495. Thomas, W. R. *et al.* Seasonal and comparative evidence of adaptive gene expression in mammalian brain size plasticity. *Elife* **13**, (2025).

496. Okbay, A. *et al.* Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat. Genet.* **54**, 437–449 (2022).

497. Goto, S. *et al.* Cryo-EM structures elucidate the multiligand receptor nature of megalin. *Proc. Natl. Acad. Sci.* **121**, 2017 (2024).

498. Spoelgen, R. *et al.* LRP2/megalin is required for patterning of the ventral telencephalon. *Development* **132**, 405–414 (2005).

499. Christ, A. *et al.* LRP2 Is an Auxiliary SHH Receptor Required to Condition the Forebrain Ventral Midline for Inductive Signals. *Dev. Cell* **22**, 268–278 (2012).

500. Rebekah Prasoona, K., Sunitha, T., Srinadh, B., Muni Kumari, T. & Jyothy, A. LRP2 gene variants and their haplotypes strongly influence the risk of developing neural tube defects in the fetus: a family-triad study from South India. *Metab. Brain Dis.* **33**, 1343–1352 (2018).

501. Kowalczyk, I. *et al.* Neural tube closure requires the endocytic receptor Lrp2 and its functional interaction with intracellular scaffolds. *Dev.* **148**, (2021).

502. Takano-Kawabe, K., Matoba, K., Nakamura, Y. & Moriyama, M. Low Density Lipoprotein Receptor-related Protein 2 Expression and Function in Cultured Astrocytes and Microglia. *Neurochem. Res.* **49**, 199–211 (2024).

503. Bento-Abreu, A. *et al.* Megalin is a receptor for albumin in astrocytes and is required for the synthesis of the neurotrophic factor oleic acid. *J. Neurochem.* **106**, 1149–1159 (2008).

504. Dorey, E. *et al.* Apolipoprotein E Isoforms Differentially Regulate Alzheimer's Disease and Amyloid-β-Induced Inflammatory Response in vivo and in vitro. *J. Alzheimer's Dis.* **57**, 1265–1279 (2017).

505. Mielenz, D. *et al.* EFhd2/Swiprosin-1 is a common genetic determinator for sensation-seeking/low anxiety and alcohol addiction. *Mol. Psychiatry* **23**, 1303–1319 (2018).

506. Dütting, S., Brachs, S. & Mielenz, D. Fraternal twins: Swiprosin-1/EFhd2 and

Swiprosin-2/EFhd1, two homologous EF-hand containing calcium binding adaptor proteins with distinct functions. *Cell Commun. Signal.* **9**, 1–14 (2011).

507. Soliman, A. S. *et al.* EFhd2 brain interactome reveals its association with different cellular and molecular processes. *J. Neurochem.* **159**, 992–1007 (2021).

508. Regensburger, M. *et al.* Impact of Swiprosin-1/Efhd2 on Adult Hippocampal Neurogenesis. *Stem Cell Reports* **10**, 347–355 (2018).

509. Vega, I. E. *et al.* A novel calcium-binding protein is associated with tau proteins in tauopathy. *J. Neurochem.* **106**, 96–106 (2008).

510. Ferrer-Acosta, Y. *et al.* EFhd2 is a novel amyloid protein associated with pathological tau in Alzheimer's disease. *J. Neurochem.* **125**, 921–931 (2013).

511. Martins-De-Souza, D. *et al.* Prefrontal cortex shotgun proteome analysis reveals altered calcium homeostasis and immune system imbalance in schizophrenia. *Eur. Arch. Psychiatry Clin. Neurosci.* **259**, 151–163 (2009).

512. Zhang, Y. *et al.* An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. *J. Neurosci.* **34**, 11929–11947 (2014).

513. Aleksander, S. A. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, 1–14 (2023).

514. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

515. Tsitkov, S. *et al.* Disease related changes in ATAC-seq of iPSC-derived motor neuron lines from ALS patients and controls. *Nat. Commun.* **15**, 1–15 (2024).

516. Boulting, G. L. *et al.* Activity-dependent regulome of human GABAergic neurons reveals new patterns of gene regulation and neurological disease heritability. *Nat. Neurosci.* **24**, 437–448 (2021).

517. Vermunt, M. W. *et al.* Large-Scale Identification of Coregulated Enhancer Networks in the Adult Human Brain. *Cell Rep.* **9**, 767–779 (2014).

518. Wang, L. *et al. Molecular and cellular dynamics of the developing human neocortex*. *Nature* (2025). doi:10.1038/s41586-024-08351-7.

519. Turk, E. *et al.* Functional connectome of the fetal brain. *J. Neurosci.* **39**, 9716–9724 (2019).

520. Ji, L. *et al.* Trajectories of human brain functional connectome maturation across the birth transition. *PLoS Biol.* **22**, 1–19 (2024).

521. Keunen, K., Counsell, S. J. & Benders, M. J. N. L. The emergence of functional architecture during early brain development. *Neuroimage* **160**, 2–14 (2017).

522. Lattke, M. *et al.* Extensive transcriptional and chromatin changes underlie astrocyte maturation in vivo and in culture. *Nat. Commun.* **12**, (2021).

523. Lee, P. H. *et al.* Partitioning heritability analysis reveals a shared genetic basis of brain anatomy and schizophrenia. *Mol. Psychiatry* **21**, 1680–1689 (2016).

524. Bystron, I., Blakemore, C. & Rakic, P. Development of the human cerebral cortex: Boulder Committee revisited. *Nat. Rev. Neurosci.* **9**, 110–122 (2008).

525. Kriegstein, A., Noctor, S. & Martínez-Cerdeño, V. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. *Nat. Rev. Neurosci.* **7**, 883–890 (2006).

526. Herculano-Houzel, S. The glia/neuron ratio: How it varies uniformly across brain structures and species and what that means for brain physiology and evolution. *Glia* **62**, 1377–1391 (2014).

527. Rakic, P., Ayoub, A. E., Breunig, J. J. & Dominguez, M. H. Decision by division: making cortical maps. *Trends Neurosci.* **32**, 291–301 (2009).

528. Ge, W. P. & Jia, J. M. Local production of astrocytes in the cerebral cortex. *Neuroscience* **323**, 3–9 (2016).

529. Moffat, A. & Schuurmans, C. The Control of Cortical Folding: Multiple Mechanisms, Multiple Models. *Neuroscientist* **30**, 704–722 (2024).

530. Imaizumi, Y. *et al.* Galectin-1 is expressed in early-type neural progenitor cells and down-regulates neurogenesis in the adult hippocampus. *Mol. Brain* **4**, 7 (2011).

531. Sakaguchi, M. *et al.* Regulation of adult neural progenitor cells by Galectin-1/beta1 Integrin interaction. *J. Neurochem.* **113**, 1516–1524 (2010).

532. Smeland, O. B. *et al.* Genetic Overlap between Schizophrenia and Volumes of Hippocampus, Putamen, and Intracranial Volume Indicates Shared Molecular Genetic Mechanisms. *Schizophr. Bull.* **44**, 854–864 (2018).

533. Cheng, W. *et al.* Genetic Association Between Schizophrenia and Cortical Brain Surface Area and Thickness. *JAMA Psychiatry* **78**, 1020 (2021).

534. Stauffer, E. M. *et al.* Grey and white matter microstructure is associated with polygenic risk for schizophrenia. *Mol. Psychiatry* **26**, 7709–7718 (2021).

535. Stauffer, E. M. *et al.* The genetic relationships between brain structure and schizophrenia. *Nat. Commun.* **14**, (2023).

536. Kushima, I. *et al.* Comparative Analyses of Copy-Number Variation in Autism Spectrum Disorder and Schizophrenia Reveal Etiological Overlap and Biological Insights Article Comparative Analyses of Copy-Number Variation in Autism Spectrum Disorder and Schizophrenia Reveal Etiol. *Cell Rep.* 2838–2856 (2018) doi:10.1016/j.celrep.2018.08.022.

537. Rees, E. *et al.* Schizophrenia, autism spectrum disorders and developmental disorders share specific disruptive coding mutations. *Nat. Commun.* **12**, 1–9 (2021).

538. Ni, H. C., Lin, H. Y., Chen, Y. C., Tseng, W. Y. I. & Gau, S. S. F. Boys with autism spectrum disorder have distinct cortical folding patterns underpinning impaired self-regulation: a surface-based morphometry study. *Brain Imaging Behav.* **14**, 2464–2476 (2020).

539. Ning, M., Li, C., Gao, L. & Fan, J. Core-Symptom-Defined Cortical Gyrification Differences in Autism Spectrum Disorder. *Front. Psychiatry* **12**, 1–12 (2021).

540. Di Martino, A. *et al.* The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* **19**, 659–667 (2014).

541. Holst, C. B., Brøchner, C. B., Vitting-Seerup, K. & Møllgård, K. Astrogliogenesis in human fetal brain: complex spatiotemporal immunoreactivity patterns of GFAP, S100, AQP4 and YKL-40. *J. Anat.* **235**, 590–615 (2019).

542. Maden, M. Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nat. Rev. Neurosci.* **8**, 755–765 (2007).

543. Shibata, M. *et al.* Regulation of prefrontal patterning and connectivity by retinoic acid. *Nature* **598**, 483–488 (2021).

544. Li, X. *et al.* FOXP1 regulates cortical radial migration and neuronal morphogenesis in developing cerebral cortex. *PLoS One* **10**, 1–19 (2015).

545. Pearson, C. A. *et al.* Foxp1 Regulates Neural Stem Cell Self-Renewal and Bias Toward Deep Layer Cortical Fates. *Cell Rep.* **30**, 1964-1981.e3 (2020).

546. Park, S. H. E., Kulkarni, A. & Konopka, G. FOXP1 orchestrates neurogenesis in

human cortical basal radial glial cells. *PLoS Biol.* **21**, 1–26 (2023).

547. Rocca, D. L., Wilkinson, K. A. & Henley, J. M. SUMOylation of FOXP1 regulates transcriptional repression via CtBP1 to drive dendritic morphogenesis. *Sci. Rep.* **7**, 1–12 (2017).

548. Zhou, Z. *et al.* Astrocytic cAMP modulates memory via synaptic plasticity. *Proc. Natl. Acad. Sci.* **118**, 1–10 (2021).

549. Zhou, Z., Ikegaya, Y. & Koyama, R. The Astrocytic cAMP Pathway in Health and Disease. *Int. J. Mol. Sci.* **20**, 779 (2019).

550. Karki, P. *et al.* cAMP Response Element-binding Protein (CREB) and Nuclear Factor □ B Mediate the Tamoxifen-induced Up-regulation of Glutamate Transporter 1 (GLT-1) in Rat Astrocytes *. *J. Biol. Chem.* **288**, 28975–28986 (2013).

551. Lim, W. L. *et al.* CREB - regulated transcription during glycogen synthesis in astrocytes. *Sci. Rep.* 1–18 (2024) doi:10.1038/s41598-024-67976-w.

552. Impey, S. *et al.* Stimulation of cAMP response element (CRE)-mediated transcription during contextual learning. *Nat. Neurosci.* **1**, 595–601 (1998).

553. Impey, S. *et al.* Induction of CRE-mediated gene expression by stimuli that generate long-lasting ltp in area ca1 of the hippocampus. *Neuron* **16**, 973–982 (1996).

554. Athos, J., Impey, S., Pineda, V. V, Chen, X. & Storm, D. R. Hippocampal CRE-mediated gene expression is required for contextual memory formation. *Nat. Neurosci.* **5**, 1119–1120 (2002).

555. Agoglia, R. M. *et al.* Primate cell fusion disentangles gene regulatory divergence in neurodevelopment. *Nature* **592**, 421–427 (2021).

556. Pachano, T., Haro, E. & Rada-Iglesias, A. Enhancer-gene specificity in development and disease. *Dev.* **149**, (2022).

557. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).

558. Symmons, O. *et al.* Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).

559. Zhan, Y. *et al.* Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome Res.* **27**, 479–490 (2017).

560. Sun, F. *et al.* Promoter-Enhancer Communication Occurs Primarily within Insulated Neighborhoods. *Mol. Cell* **73**, 250-263.e5 (2019).

561. Gómez-Robles, A., Nicolaou, C., Smaers, J. B. & Sherwood, C. C. The evolution of human altriciality and brain development in comparative context. *Nat. Ecol. Evol.* **8**, 133–146 (2024).

562. Maricque, B. B., Dougherty, J. D. & Cohen, B. A. A genome-integrated massively parallel reporter assay reveals DNA sequence determinants of cis-regulatory activity in neural cells. *Nucleic Acids Res.* **45**, e16 (2017).

563. Esnault, C. *et al.* G4access identifies G-quadruplexes and their associations with open chromatin and imprinting control regions. *Nat. Genet.* **55**, 1359–1369 (2023).

564. Esnault, C. *et al. G-quadruplexes are promoter elements controlling nucleosome exclusion and RNA polymerase II pausing*. *Nature Genetics* vol. 57 (Springer US, 2025).

565. Georgakopoulos-Soares, I. *et al.* High-throughput characterization of the role of non-B DNA motifs on promoter function. *Cell Genomics* **2**, 100111 (2022).

566. Roy, S. S. *et al.* Artificially inserted strong promoter containing multiple G-quadruplexes induces long-range chromatin modification. *Elife* **13**, 1–20 (2024).

567. Mauduit, D. *et al.* Analysis of long and short enhancers in melanoma cell states. *Elife* **10**, 1–30 (2021).

568. Hansen, T. J. & Hodges, E. ATAC-STARR-seq reveals transcription factor-bound activators and silencers across the chromatin accessible human genome. *Genome Res.* 1529–1541 (2022) doi:10.1101/gr.276766.122.

569. Zhang, X., Fang, B. & Huang, Y. F. Transcription factor binding sites are frequently under accelerated evolution in primates. *Nat. Commun.* **14**, (2023).

570. Deng, C. *et al.* Massively parallel characterization of regulatory elements in the developing human cortex. *Science (80-.).* **384**, 868–876 (2024).

571. Castelijns, B. *et al.* Recently Evolved Enhancers Emerge with High Interindividual Variability and Less Frequently Associate with Disease. *Cell Rep.* **31**, 107799 (2020).

572. Emera, D., Yin, J., Reilly, S. K., Gockley, J. & Noonan, J. P. Origin and evolution of developmental enhancers in the mammalian neocortex. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E2617–E2626 (2016).

573. Kvon, E. Z., Waymack, R., Gad, M. & Wunderlich, Z. Enhancer redundancy in development and disease. *Nat. Rev. Genet.* **22**, 324–336 (2021).

574. McAfee, J. C. *et al.* Focus on your locus with a massively parallel reporter assay. *J. Neurodev. Disord.* **14**, 1–14 (2022).

575. Grossman, S. R. *et al.* Positional specificity of different transcription factor classes within enhancers. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E7222–E7230 (2018).

576. Agarwal, V. *et al. Massively parallel characterization of transcriptional regulatory elements*. *Nature* vol. 639 (Springer US, 2025).

List of publications

Publications which were a part of presented work

Ciuba, K., Piotrowska, A., Chaudhury, D., Dehingia, B., **Duński, E.**, Behr, R., et al. (2025). *Molecular signature of primate astrocytes reveals pathways and regulatory changes contributing to human brain evolution*. Cell Stem Cell 32, 426-444.e14. doi: 10.1016/j.stem.2024.12.011.

The manuscript describing results from Massively Parallel Assay (MPRA) described below (Duński et. al) is currently in preparation.

Other publications by the author

Eryk Duński, Aleksandra Pękowska, *Keeping the balance: Trade-offs between human brain evolution, autism, and schizophrenia* (**2022**), Frontiers in Genetics, Sec. Epigenomics and Epigenetics Volume 13 – 2022 doi: 10.3389/fgene.2022.1009390

Nethaji J Gallage, Kirsten Jorgensen, Christian Janfelt, Agnieszka J Z Nielsen, Thomas Naake, **Eryk Duński**, Lene Dalsten, Michel Grisoni, Birger Lindberg Moller, *The Intracellular Localization of the Vanillin Biosynthetic Machinery in Pods of Vanilla planifolia* (**2018**), Plant and Cell Physiology, Volume 59, Issue 2, February 2018, Pages 304–318, doi: doi.org/10.1093/pcp/pcx185

Patents:

Henry John Lamble, Christopher Egan, David Lloyd, Eryk Dunski (2018), *Nucleic acid detection method.* European Patent No. EP 3325651B1, European Patent Office

Henry John Lamble, Christopher Egan, David Lloyd, Eryk Dunski (2018), *Nucleic acid detection method.* U.S. Patent No. US 11390909B2. Washington, DC: U.S. Patent and Trademark Office.