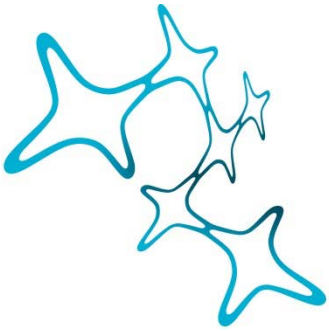
Identification of Functional Genetic Variants Modulating the Stress Response

Signe Penner-Goeke



Graduate School of
Systemic Neurosciences
LMU Munich

Dissertation der
Graduate School of Systemic Neurosciences der
Ludwig-Maximilians-Universität München
February 8th, 2021

Supervisor:
Prof. Dr. Dr. med. univ. Elisabeth Binder
Department of Translational Research in Psychiatry
Max Planck Institute of Psychiatry

First Reviewer: Prof. Dr. Dr. med. univ. Elisabeth Binder
Second Reviewer: Dr. Sebastiaan Meijsing
External Reviewer: Prof. Dr. Onno Meijer

Date of Defense: May 14th, 2021

Table of Contents

List of Abbreviations and Terms

MDD	Major depressive disorder
SCZ	Schizophrenia
OR	Odds Ratio
CNV	Copy number variations
ASD	Autism spectrum disorder
GWAS	Genome wide association studies
OCD	Obsessive compulsive disorder
PTSD	Post-traumatic stress disorder
BPD	Bipolar disorder
ADHD	Attention deficit hyperactivity disorder
AN	Anorexia nervosa
TS	Tourette's syndrome
LD	Linkage disequilibrium
RE	Regulatory element
TF	Transcription factor
(e)QTL	(Expression) Quantitative trait locus
MPRA	Massively parallel enhancer assays
STARR-seq	Self-transcribing active regulatory region sequencing
ENCODE	Encyclopedia of DNA Elements
REMC	Roadmap Epigenomics Mapping Consortium
ORF	Open reading frame
GxE	Gene by environment
HPA	Hypothalamus pituitary adrenal
PRS	Polygenic risk score
SAM	Sympathetic adrenomedullary
CRH	Corticotrophin-releasing hormone
ACTH	Adrenocorticotrophic hormone
GC	Glucocorticoid
MR	Mineralocorticoid receptor

GR	Glucocorticoid receptor
Dex	Dexamethasone
GRE	Glucocorticoid response elements
Veh	Vehicle (ethanol) control
MR	Mendelian randomization
DRE	STARR-seq identified dex responsive regulatory element
UMI	Unique molecular identifier
WB	Western Blot
Variant-DRE	STARR-seq identified dex-responsive regulatory element containing a variant driving allele dependent activity
non-functional GR-eQTL	GR-eQTLs not identified as being functional with STARR-seq
IPSC	Induced pluripotent stem cells
iNeurons	Neurons derived from induced pluripotent stem cells

Abstract

Stress and exposure to stressful life events are some of the strongest risk factors for the development of psychiatric disorders.[1] While it is known that genetic factors play an important role in one's response to stress, the genetic variants driving differences in the stress response have yet to be determined. Using an expression trait quantitative loci (eQTL) approach, previous research from our group identified over 3600 genetic variants associated with altered transcriptional activity upon activation of the glucocorticoid receptor (GR),[2] a key transcription factor in the stress response system. These variants (GR-eQTLs) were enriched in GR binding sites and putative regulatory elements, and predicted case-control status for major depressive disorder, as well as amygdala activity in an emotional reactivity task. However, due to linkage disequilibrium, the functional variants driving these transcriptional changes remain unidentified.

In order to identify which of the GR-eQTLs had a functional effect on GR mediated transcription, we performed STARR-sequencing, an approach allowing thousands of DNA sequences to be assessed for regulatory element activity in parallel. This allowed us to determine the effect of each GR-eQTL on transcriptional activity at baseline and after GR activation with dexamethasone (dex) in GR18 and U138MG cells. STARR-seq revealed that 1220 variants were located in active regulatory elements, either at baseline or after dex. Of these, 547 variants show differential transcriptional activity after dex treatment (DREs). These were enriched in GR-binding sites determined by ChIP-seq, as well as chromatin loop anchor points. Next, DREs with allele dependent activity at baseline and after dex were identified. We found 165 DREs to show allele dependent activity at baseline, and 172 DREs after dex. Both the baseline and dex functional variants showed a significant overlap with variants nominally associated with psychiatric disorders from the latest psychiatric cross disorder meta-analysis GWAS.[3] Furthermore, using Mendelian Randomization, nineteen and eleven DREs with allele-dependent activity in the dex and veh condition, respectively, were identified to have a putative causal effect on psychiatric disorders. These results suggest that the loci modulating the transcriptomic response to stress are putatively causally associated with psychopathology.

1. Introduction

1.1. An overview of the role of genetics in psychiatry

1.1.1. Genetics of psychiatric disorders

Psychiatric disorders are a primary cause of disability worldwide, currently comprising 12% of all disability adjusted life years.[4] Despite their significant social and economic costs, the molecular mechanisms underlying psychiatric disorders remain poorly understood. It is well established that risk for psychiatric disorders are partially mediated by genetic factors, with sibling heritability estimates ranging from 0.3 for major depressive disorder (MDD) to over 0.8 for schizophrenia (SCZ).[5] For example, a meta-analysis on the heritability of MDD quantified this increased risk as an odds ratio (OR) of 2.84 (95% CI = 2.31-3.49) for first-degree relatives of subjects with MDD.[6] In more recent years, advances made in high-throughput sequencing technology have enabled researchers to move beyond family based heritability studies of psychiatric disorders to identify disease-associated susceptibility loci using methods such as whole exome sequencing, microarrays, and whole-genome sequencing. These advances have allowed both small effect size common variants and larger effect size rare variants to be detected. While a number of larger effect size loci have been detected in some psychiatric disorders, such as copy number variations (CNV) in autism spectrum disorder (ASD)[7, 8] and SCZ,[9] these findings are rare, and few or no robust large effect size loci have been identified for the majority of psychiatric disorders.[10] On the other hand, genome-wide associations studies (GWAS) have been successful in identifying genetic loci with small effect sizes involved in many psychiatric disorders. GWAS identify genetic loci associated with a specific phenotype by genotyping both a large cohort of individuals displaying the phenotype, as well as a control cohort. By comparing the cases to controls, genetic markers associated with a trait can be identified, even when a single genetic locus has a small effect. What has become clear from these studies is that in most cases, psychiatric disorders are polygenic disorders, with multiple loci being involved, but each of small effect. Single disease GWAS have identified between zero (obsessive-compulsive disorder (OCD) and post-traumatic stress disorder (PTSD)) to 155 (SCZ) loci associated with psychiatric disorders.[11] Furthermore, there is evidence that some genetic loci have pleiotropic effects, and thus play a role in multiple psychiatric disorders. A meta-analysis of eight psychiatric traits (bipolar disorder (BPD), MDD, attention deficit hyperactivity disorder (ADHD), anorexia nervosa (AN), OCD,

SCZ, Tourette's Syndrome (TS), and ASD) found 109 loci that were associated with more than one of these disorders,[3] suggesting that there may be underlying genetic mechanisms common across disorders. Furthermore, of the 109 loci, 33 were novel, likely due to the increased sample size enabled by the meta-analysis approach, which allows for loci with smaller effect sizes to be detected. As sample sizes for GWAS increase, the number of associated loci and their effect sizes are projected to increase.

1.1.2. Using functional genomic approaches in psychiatry

While GWAS and other high-throughput techniques have enabled genetic variants involved in various psychiatric disorders to be identified, their functional role in the underlying biological processes governing these disorders remains elusive. The majority of GWAS loci are located within non-coding genomic regions, far from genes, and exhibit high linkage disequilibrium (LD).[12,13] This LD, which refers to the non-random association between two alleles located at different loci, makes it difficult to identify the variants that have a causal role in disease, and which variants are only associated with the causal variant. Better functional fine-mapping of GWAS loci is required in order to pinpoint which variants play a functional role, and how they contribute to psychiatric disorders.

Although the majority of GWAS loci are located in non-coding regions, they can still exert a functional effect on gene expression via modulation of regulatory element (RE) activity. REs, such enhancers and silencers, are genomic regions that regulate the expression of one or occasionally, multiple genes. They are located within non-coding regions, and can regulate nearby (<1Mb) (cis-REs) or distant (trans-REs) genes.[14] Enhancers are REs that typically increase transcription at their target gene(s), which can be located as far as millions of base pairs away.[15] Enhancers generally exert their effect by binding transcription factors (TFs), thereby promoting chromatin looping that brings the enhancer into proximity with its target gene. Though certain chromatin states and histone marks are associated with enhancer regions,[16,17] there is no consensus enhancer sequence, making their identification challenging. Therefore, the first step in identifying functional GWAS loci located in non-coding regions is to determine whether they are located within REs and if so, the target genes which they regulate. To this end, various approaches, both experimental and computational, can be employed and integrated. Xiao et al. has proposed a

framework in which risk variants located in non-coding regions of the genome can be functionally prioritized by integrating various genomic datasets and experimental methods.[18]

According to Xiao et al., the first step in this functional prioritization is to determine which genes are associated with the genetic variants by combining genotype and transcriptomic data. With these two datasets and a sufficient sample size, an expression quantitative trait loci (eQTL) analysis can be performed. eQTLs are genetic variants that affect, to differing degrees, the expression of a transcript.[19] This approach prioritizes genetic variants that have an effect on gene expression, and maps the genetic variant to (a) specific gene(s). By mapping the variants to specific genes, downstream gene-level analyses can be performed which may provide insight into the biological processes and pathways underlying the GWAS trait. Given the cell-specificity of gene expression, brain specific transcriptomic data is especially important when investigating the pathology of psychiatric disorders. In more recent years, consortia have generated brain-specific eQTL databases, which map genetic variants to changes in gene expression in the brain.[20, 21, 22] Although eQTL analyses link genetic variants with changes in gene expression, it does not address the challenge posed by LD.

The output from GWAS and eQTL analyses provide only indirect associations which are not sufficient to confirm whether a genetic variant is indeed located within a RE and whether it is functional. For this, one must directly assess the activity of the putative REs using reporter assays. Traditional reporter assays allow for the activity of only one putative RE to be assessed at a time. For example, in a luciferase assay, the putative RE modulates the activity of the luciferase gene, whose product emits light in the presence of a substrate that can be quantified when expressed. Therefore, the amount of light emitted is proportional to the strength of the RE.[23] However, these low-throughout approaches are not compatible with probing thousands of putative REs. Fortunately, in recent years, next-generation sequencing techniques have been used to develop high throughput assays that allow thousands of putative REs to be assessed in parallel. To date, two different approaches have been developed; massively parallel enhancer assays (MPRA) and self-transcribing active regulatory region sequencing (STARR-seq). Both approaches involve synthesizing the putative REs and cloning these oligonucleotide fragments in parallel into a plasmid that contains a basal promoter and an open reading frame (ORF). In MPRA, each fragment in the oligonucleotide fragment library contains a unique barcode which is located upstream of a polyadenylated tail and downstream of the ORF, so that the barcode is transcribed. The library is

transfected into the cells of interest, mRNA is isolated, and cDNA synthesized. The barcodes are sequenced before transfection to determine the composition of the input oligonucleotide fragment library, and to associate each putative RE with its respective barcodes. The difference between the enrichment of the barcodes pre- and post-transfection is proportional to the strength of the RE. In principle STARR-seq is very similar except no barcodes are used. Instead, the oligonucleotide fragments are cloned downstream of a minimal promoter meaning the mRNA transcribed is the sequence of the RE itself. Rather than using barcodes, the mRNA is quantified, as it contains the sequence of the RE. Therefore, the strength of a single RE is proportional to its enrichment pre- vs. post-transfection, as strong REs will induce transcription of mRNA fragments containing its own sequence when transfected into cells.[24] Both these high throughput methods allow the allele effects on RE activity to be tested, but also the effects of different conditions such as hypoxia or a drug treatment.[25,26] To date, few studies have employed these high throughout techniques to assess psychiatric disorder GWAS variants. A study in 2019 assessed 1049 variants that were in LD with 64 SCZ associated GWAS loci and identified 53 and 148 functional REs with allele-dependent activity in a neuroblastoma and chronic myelogenous leukemia lymphoblast cell line, respectively.[27] This provides support that these high throughput methods are capable of both identifying functional REs and detecting the effects of single psychiatric disorder associated variants on transcriptional activity.

While high-throughput functional assays allow for functional REs to be identified, they need to be mapped to downstream effects to gain a better understanding of how they regulate gene expression, thereby providing insight into the mechanisms by which they exert their effects. To gain better insight into the regulatory mechanisms of these variants, other approaches are required. Fortunately, there exists publicly available datasets and corresponding computational tools that can aid with this task. These tools allow for genetic variants to be functionally annotated in silico and across various tissues types. Since REs are associated with certain features, such as specific histone marks, open chromatin states, and TF binding, these marks can be used to predict the location of REs within different tissues and cell types. Current large datasets include the Encyclopedia of DNA Elements (ENCODE) which contains nearly 5000 human datasets,[28] and the Roadmap Epigenomics Mapping Consortium (REMC). Various tools that integrate data from these publicly available datasets have been developed to functionally annotate the genome, such as ChromHMM, a tool that uses a multivariate Hidden Markov Model to predict chromatin states

and identify REs across the genome.[29] However, as useful as these computational tools are, they are still limited in the cell types and tissues that they probe, and include only a subset of TFs and histone marks. Furthermore, co-localization approaches can only indicate which putative mechanisms may be involved, but cannot test them directly. Therefore, additional experimental assays, such as ATAC-seq and ChIP-seq, using the cell types and genomic features of interest are often necessary. With these datasets and a sufficient sample size, QTL analyses such as ChIP-QTLs or chromatin-QTLs can be performed, providing better mechanistic insight into how genetic variants may exert their effects, such as by disrupting TF binding or modulating chromatin states. However, to move beyond association studies, which are provided by QTL and co-localization approaches, direct approaches are required, such as SNP-editing of isogenic cell lines.

While GWAS provide valuable information on genetic loci implicated in psychiatric disorders, alone they provide little information on functionality. While eQTL analyses, in silico functional modelling, and other experimental approaches can help prioritize putatively functional variants and even bring us closer to identifying the implicated genes and putative mechanisms involved in these disorders, they only provide associations between risk variants and a molecular phenotype. To validate these associations, direct assessment of the risk variants is required. Integration of the above-mentioned datasets, as well as harnessing the potential of recent genome editing technology will not only provide new insights into which variants play a functional role, but also illuminate the biological pathways underlying these diseases, which could be exploited to develop better diagnostic tools and identify novel therapeutic targets.

1.1.3. Gene x environment interactions in psychiatric disorders

While there exists strong evidence from both family-based heritability studies and GWAS that genetic risk factors play an integral role in psychiatric disorders, it is also clear from these studies that other factors also mediate the risk for these disorders. Indeed, it has been well established that environmental factors, such as pre- and postnatal environment, socioeconomic status, and stress,[30,31,32] contribute to risk for psychiatric disorders. For example, a meta-analysis of 26 studies found that childhood trauma, especially neglect and emotional abuse, was significantly associated with MDD in adulthood (OR 2.78 and 2.75, respectively).[33] However, these two factors, genetics and environment, are not independent, and they have been shown to interact. Gene by environment (GxE) interactions refer to the influence of environmental and

genetic factors on a measured phenotype, in this case, psychiatric disorders. However, in GxE interactions, the magnitude of the environmental influence on the phenotype depends on the genotype of the individual. Many examples of GxE interactions exist within medicine, such as the interaction between alcohol exposure and genetic variants regulating N-glycosylation genes, which predicts the severity of fetal alcohol disorder,[34] or the interaction between diet and genetic variants in type 2 diabetes.[35] While GxE interactions have been studied and replicated in a number of somatic disorders, the results from GxE interaction studies in psychiatry have not been so conclusive. For example, many studies that have focused on the interaction between the serotonin receptor transporter gene, 5-HTTLPR, and stressful live events have found a significant effect of the GxE interaction on various psychopathologies.[36,37,38] However, results from other studies have failed to replicate these findings, such as two recent metanalyses focusing on the effect of this interaction on depression which have reported negative results.[39] These inconsistent results may be the result of differences in study design, including how stressful life events and the outcome variable, in this case MDD, are defined and diagnosed.

However, there are some GxE interactions in psychiatry that have been well-studied and appear to be more robust. One such well-studied GxE interaction is the interaction between a genetic variant (rs1360780) located in an enhancer region within FK506 binding protein 5 gene (FKBP5) and childhood trauma. FKBP5 is an important mediator of the hypothalamus pituitary adrenal (HPA) axis and is involved in a negative feedback loop to terminate the stress response. In one study where exposure to physical childhood abuse was found to be associated with MDD, the association was significantly elevated in exposed individuals carrying the risk allele (T/T) compared to exposed individuals carrying the reference allele (C/C or C/T).[40] In another study, this variant, along with three others in FKBP5 was shown to interact with child abuse to predict symptom severity of post-traumatic stress disorder (PTSD).[41] This interaction between childhood trauma and the T/T genotype, has also been studied on a mechanistic level. Klengel et al.[42] found that methylation of the locus was associated with childhood trauma in carriers of the risk allele, but not those with childhood trauma carrying the reference allele. Using a series of experiments, the authors propose the following model: the risk allele of rs1360780 causes differential interactions between the enhancer and the transcription start site upon GR activation induced via child abuse, resulting in transcriptional induction of FKBP5. Persistent overexpression of FKBP5 results in an impaired termination of the stress response, including prolonged GR activation. The

genotype dependent changes in chromatin structure, along with prolonged GR activity, cause DNA demethylation at CpGs located within and proximal to GREs. During certain developmental periods, the demethylation is stable, and thus results in long-term transcriptional changes of FKBP5 upon GR activation. GR-dependent changes in methylation of CpGs in FKBP5 after acute GR activation has since been validated.[43,44]

As psychiatric disorders typically involve many genetic variants each contributing a small effect, it is reasonable to assume that when studying GxE interactions, more insight might be gained by considering multiple risk variants by employing a polygene x environment interaction model, rather than using a single candidate gene. Polygenic risk scores (PRS), for example, consider multiple risk variants and represent the additive effect of the individual variants, which are typically weighed by their effect sizes to generate a single score. In polygene x environment interactions, a PRS can operationalize the integration of multiple genetic loci with environmental variables. Indeed, a few studies have examined the interaction between PRS for depression and childhood trauma which, as with the candidate gene studies, yielded conflicting results[45,46,47,] likely due in part to inconsistencies in defining the exposure and outcome phenotype, and insufficient power. In general, the results from both polygene and gene x environment studies in psychiatry highlight the need to carefully define the environmental exposures and clinical phenotypes to include, and to ensure the study has sufficient power.

1.1.4. Stress as a model for polygene x environment interactions

The challenge of defining environmental factors and outcome phenotypes in psychiatry can be partly circumvented by using more reductionistic models and studying well-characterized endophenotypes. The findings enabled by these models can then be integrated with findings from other datasets, such as clinical cohort data. Endophenotypes are intermediate traits with a genetic component more directly related to the underlying biological processes than the clinical phenotype itself.[48] Using endophenotypes generally increases statistical power, as endpoint clinical phenotypes are likely a result of multiple aberrant biological processes converging, which is less pronounced in intermediate traits.[49] Endophenotypes are especially useful in psychiatry, where diagnoses are mainly based on symptom clusters assessed by a clinician, and relatively little on biological or genetic markers. As stress is a strong risk factor for multiple psychiatric disorders, I have chosen to study the transcriptional response to stress, and how common genetic variants

mediate this response. In this case, the environmental factor, stress, is induced pharmacologically, allowing for the environmental factor to be precisely and uniformly applied. The endophenotype, the transcriptomic response to stress, is quantifiable, and is driven by molecular mechanisms that can be investigated. These factors allow for polygene x E interactions to be investigated with sufficient power, even with a smaller sample size.

1.2. Stress

1.2.1. The stress response - Hypothalamus Pituitary Adrenal Axis

The stress response is a highly dynamic process that helps humans assess and adapt to the environment, deal with stressful events and threats, and to anticipate future challenging situations. Exposure to stress results in physiological changes that are mediated by two main systems; the hypothalamic-pituitary-adrenal (HPA) axis, which is a neuroendocrine system, and the autonomic sympathetic adrenomedullary system (SAM). The SAM system responds immediately (within seconds) and results in the rapid release of catecholates which prepares the body for the "flight or fight" response.[50] The HPA axis, on the other hand, is somewhat slower and culminates with the release of glucocorticoids (GCs) that have both non-genomic and genomic effects,[51] with the latter being the focus of this thesis.

As the name implies, the HPA axis starts with the hypothalamus. Within minutes of being exposed to stress, the hypophysiotropic neurons of the periventricular nucleus of the hypothalamus are activated and synthesize corticotrophin-releasing hormone (CRH). CRH is released into the hypophyseal portal system, and transported to the pituitary gland. In the anterior pituitary gland, CRH binds to CRHR receptor 1, inducing the release of adrenocorticotrophic hormone (ACTH), which enters the circulatory system. ACTH largely targets the melanocortin type 2 receptors in the adrenal cortex, resulting in GC synthesis and release.[52] Cortisol is the most abundant circulating natural GC in humans and controls a number of physiological processes. The lipophilic nature of GCs allow them to translocate across the brain blood barrier and cell membranes without additional carriers, enabling them to access every organ system in the body.[53] Within a cell, they exert their effect by binding two nuclear hormone receptors, the mineralocorticoid and glucocorticoid receptors (MR, GR, respectively). MR has a greater affinity for cortisol, meaning that at basal cortisol levels, it is nearly saturated.[54] Conversely, GR is not significantly occupied in

basal conditions, but rather only binds GCs upon activation of the HPA axis or during cortisol peaks in the circadian cycle.[54,55] This means that GR plays a more dynamic role in responding to environmental stressors or threats than MR. In experimental models, the activity of GR, and not MR, can be assessed by applying a ligand with a much high affinity for GR compared to MR. Dexamethasone (dex), a synthetic GC, is a potent ligand of GR that has a high affinity for GR but a very low affinity for MR,[56] making it an appropriate ligand for studying GR.

1.2.2. The Glucocorticoid Receptor: Function and binding

GR is encoded by the nuclear receptor subfamily 3, group C, member 1 (NR3C1) gene, located on chromosome 5. In the absence of ligand, GR resides in the cytosol, existing as a complex with a number of proteins, including FKBP5. Upon activation by GCs, GR dimerizes and translocates from the cytoplasm to the nucleus. Within the nucleus, the GR-complex interacts with glucocorticoid response elements (GREs), two hexameric DNA sequences separated by three base pairs, to modulate the expression of a number of genes, many of which are involved in the stress response. For example, GR upregulates FKBP5, which is involved in the termination of the stress response, thus creating a negative feedback loop.[52] However, GCs regulate the expression of genes involved in many other integral physiological processes, including those involved with immune, metabolic, developmental, and cognitive functions.[51]

Binding of GR does not occur at all GREs. Accordingly, the presence of GREs alone cannot predict GR-complex occupancy. It has generally been hypothesized that additional factors such as histone post-translational modifications, and chromatin accessibility influence GR-complex occupancy at GREs.[57] Although the majority (95%) of GR binding has been reported to occur in regions of accessible chromatin prior to dex treatment, for a subset of GR binding sites, the presence of dex results in chromatin remodeling, making the chromatin landscape conducive to GR binding.[58] Furthermore, GR does not exclusively bind to GREs, but can tether to other TFs,[59] complexifying the model of GR binding further. GR binding has been reported to occur predominantly in regions containing active enhancer marks and in regions distal to its target genes.[60] In line with this, GR has been found to be involved in chromatin looping, bringing the enhancer in spatial proximity with the promoter of the target gene. Interestingly, the majority of the chromatin loops have been reported to be pre-established prior to dex treatment, and are involved in both GR-mediated transcriptional activation and repression.[61] Overall, the exact

mechanisms governing GR binding remains elusive. What is clear, is that a multitude of factors, including chromatin and TF landscapes, play an important role.

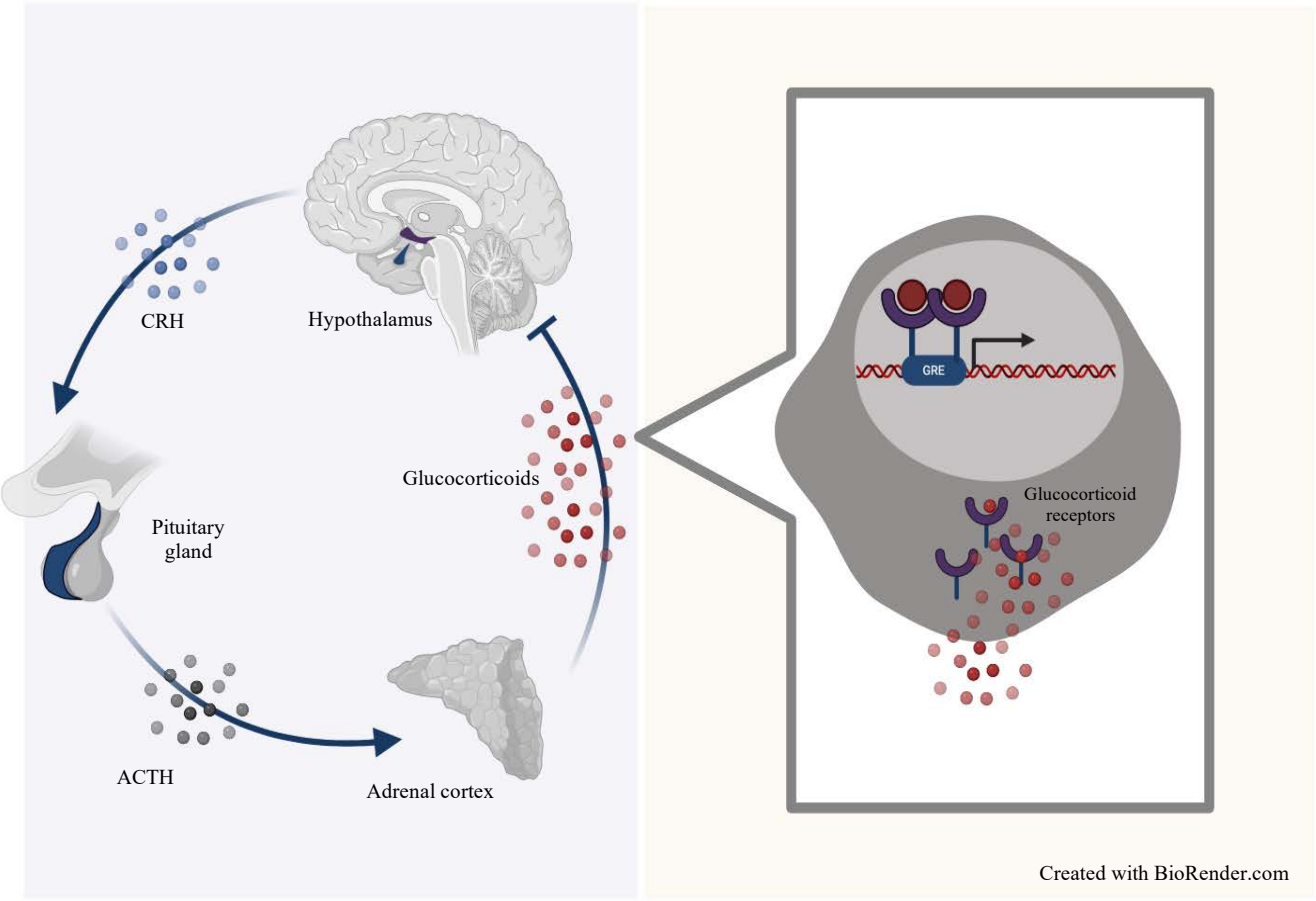


Figure 1. Overview of the main constituents of the hypothalamus pituitary adrenal axis (HPA axis) and activity of the glucocorticoid receptor (GR). A) The HPA axis is activated in response to an environmental stressor which stimulates hypophysiotropic neurons in the PVN to secrete CRH into the hypophyseal portal vessels. CRH is transported to the pituitary gland where it binds to receptors, triggering the secretion of ACTH. ACTH exits the CNS and binds to the adrenal cortex, leading to the secretion of GCs into circulatory system. GCs bind to GRs (and MRs) which modulate the expression many target genes, some of which are involved in a negative feedback loop of the HPA axis. B) At the cellular level, GR resides in the cytoplasm as part of a multiprotein complex. Once bound to its ligand, GR dimerizes and translocates into the nucleus. GR binds to DNA, acting as a TF, and regulates the expression of many genes.

1.2.3. Evidence for role of stress in psychiatric disease

Exposure to stress and stressful life events is one of the most robust risk factors for psychiatric disorders.[1] Ample evidence has implicated dysregulation of the HPA axis in stress related psychiatric disorders, such as MDD and PTSD. One of the most robust findings is hyperactivity of the HPA axis in MDD. Consequently, increased CRH transmission[62] and elevated basal and stimulated cortisol release has been observed,[63,64, 65] although not in all cohorts of MDD subjects.[66,67] Accordingly, some evidence exists that the negative feedback mechanisms responsible for terminating the stress response is blunted,[64] but again, this has not been replicated in other cohorts.[68] In PTSD, cortisol levels have generally been observed to be reduced,[69,70] likely due to an enhanced feedback mechanism,[71] but as with MDD, these findings are not consistent.[72] Aberrant HPA axis functioning in other psychiatric disorders, including BPD and SCZ, have also been observed, further highlighting the cross-diagnosis relevance of studying dysregulation the HPA axis as an endophenotype in psychiatry.

1.2.4. Genetic variants modulating the stress response

In this thesis, the functional role of genetic variants moderating the molecular stress response is investigated. This builds upon the work of a previous study that identified genetic variants mediating the transcriptomic response to stress.[73] This study employed a stimulated eQTL approach, meaning genetic variants associated with differential gene expression were identified at baseline but also after activation of GR. For this eQTL study, 93 controls and 71 subjects with MDD were treated with 1.5 mg of oral dex. Blood was drawn at baseline (18:00) and three hours post administration (21:00). RNA was extracted from the whole blood samples and gene expression for these two conditions was measured using a microarray. Genotyping was performed on all individuals. By integrating the genotype and transcriptomic data, eQTLs could be mapped at baseline and post-dex treatment, allowing genetic variants influencing gene expression at baseline and importantly, after dex to be identified. In total, 3662 variants were identified, each modulating the expression of a target gene upon dex stimulation (GR-eQTLs). These variants were enriched in GR binding sites and in genomic regions defined by an in silico analysis as being enhancers, consistent with previous findings on GR binding patterns at enhancers.[60] Interestingly, they were found to be especially enriched in enhancer regions in fetal brain. Furthermore, the GR-eQTLs were associated with inappropriate amygdala activity in an emotional reactivity task, and

were enriched for variants associated with MDD and SCZ, highlighting their relevance in psychopathology.

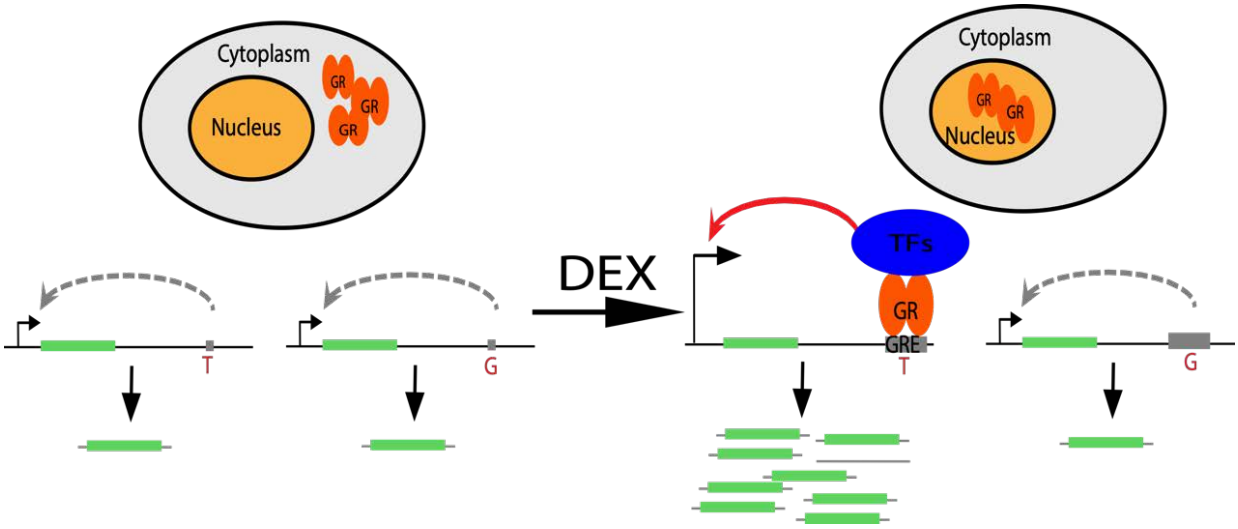


Figure 2. Schematic of a representative GR-eQTL. Prior to dex, basal transcription is low, therefore the variant exerts little or no significant effect. However, upon dex treatment, transcription is induced by GR binding. The variant, by perturbing the GRE, causes transcriptional repression. Therefore, only after dex, is allele-dependent transcriptional activity detectable.

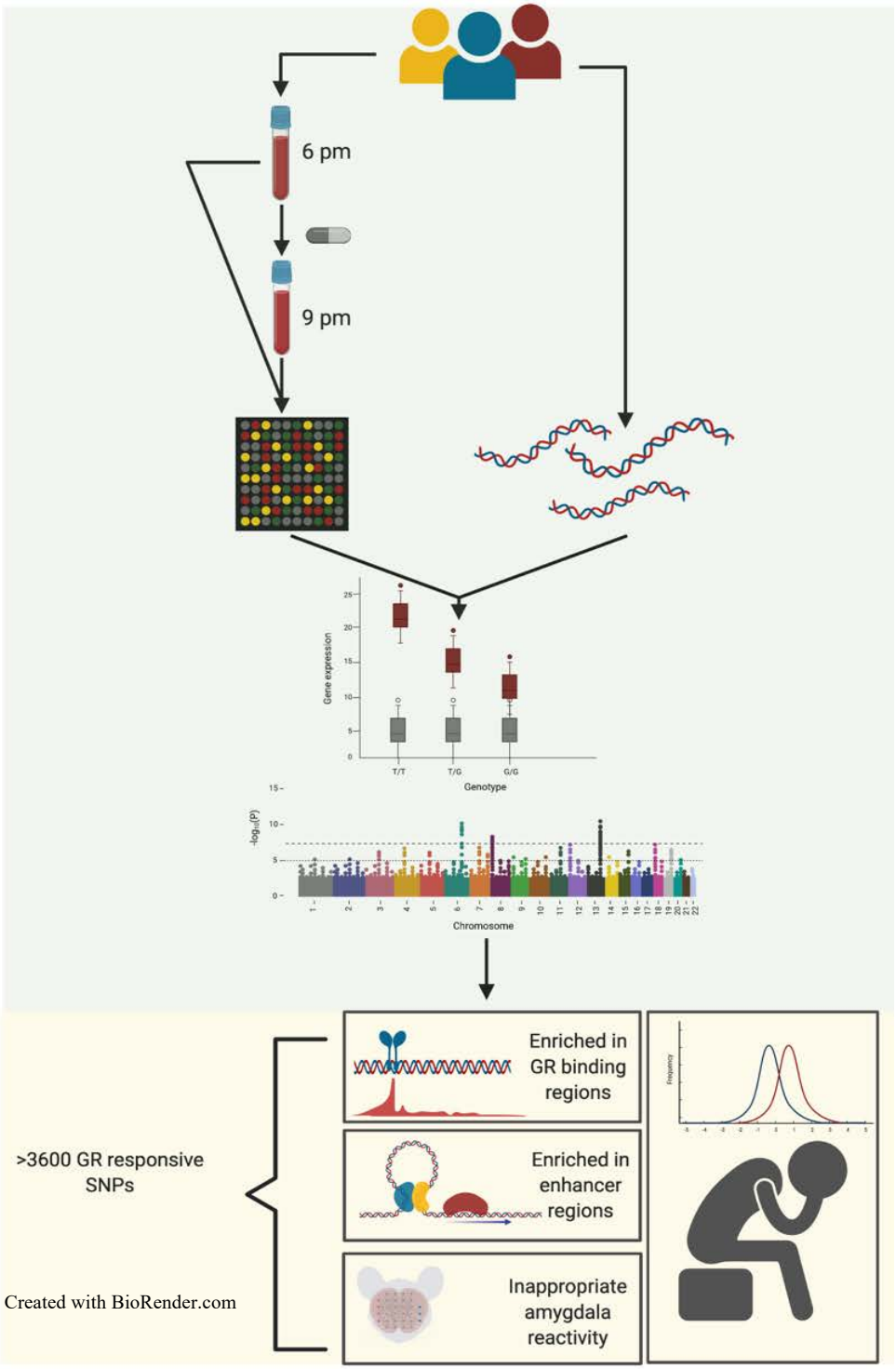


Figure 3. Schematic summarizing the study identifying GR-eQTLs. Genome-wide expression data was generated from peripheral blood from 160 male MDD cases and controls at baseline and 3 hours post-dex administration. These data were integrated with genotyping data from these individuals to identify 3662 GR-eQTLs. These GR-eQTLs were enriched in GR binding sites and putative enhancers, and associated with inappropriate amygdala reactivity. Furthermore, they were enriched in variants associated with MDD and SCZ.

1.3. Technological background for functional genomics

1.3.1. STARR-seq

STARR-seq was employed to test which GR-eQTLs identified by Arloth et al. were located within REs and if they were, the effect of the GR-eQTLs on RE activity. To this end, each of the GR-eQTLs, along with a 100bp flanking region on either side, was cloned into a STARR-seq vector.

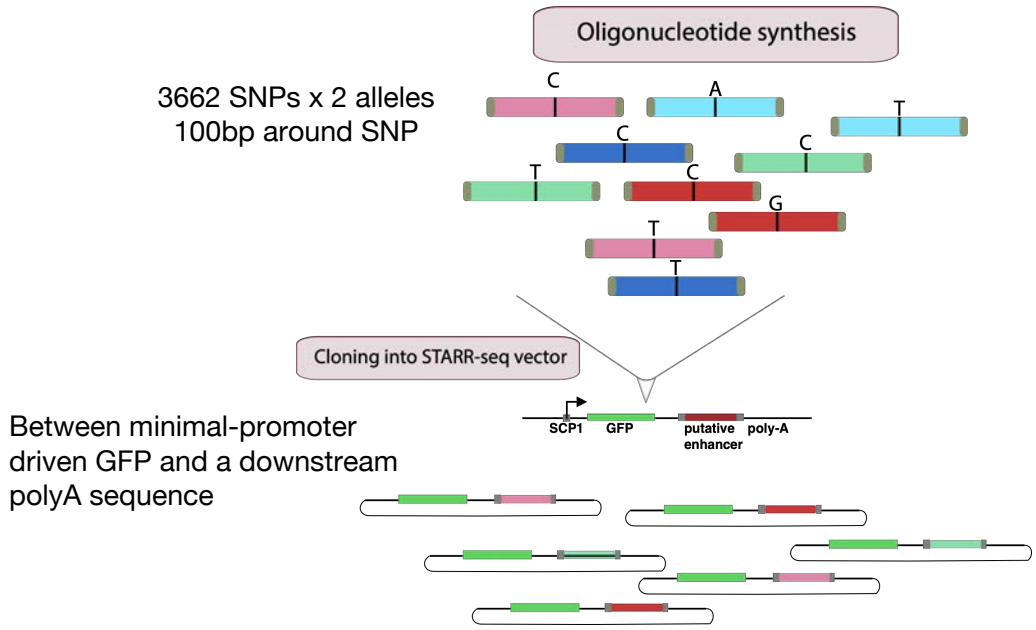


Figure 4. Schematic illustrating the STARR-seq fragment generation and generation of the plasmid library.

The activity of each fragments harboring the GR-eQTLs were tested at baseline and upon stimulation with dex. If the GR-eQTL was indeed dex-responsive and functional, the following was observed; a transcriptional activation or repression upon stimulation with dex, and a difference in RE activity between the fragments containing the two alleles. Since in STARR-seq the strength of the REs is reflected by the abundance of mRNA molecules containing its own sequence, the strength of the RE and the effect of the GR-eQTL could be quantified using sequencing, with higher abundant transcripts reflecting stronger REs.

1.3.2. Mendelian Randomization

Due to the law of independent segregation which states that alleles are randomly distributed at conception independent of environmental confounders, one can use a statistical framework, termed Mendelian Randomization (MR), to estimate putatively causal relationships between two variables; the instrument and outcome, using a third variable, the exposure. Although MR has typically been used in the field of epidemiology, in recent years, MR has been applied in the field of functional genomics to determine causal relationships between genetic variants and complex traits.[74,75] In these studies, genetic variants (the instrumental variable) are used to estimate the causal effect of a molecular QTL (exposure variable), such as eQTLs, on a complex trait (outcome variable). However, in order for the estimates to be valid, the instrumental variable must meet the following conditions:

(1) The instrument must be associated with the exposure.

(2) The instrument must not be associated with confounding variables of the exposure-outcome relationship.

(3) Instrument must only be associated with the outcome through the exposure.

The third condition, also known as the "no pleiotropy" rule, has been especially problematic as it is difficult to exhaustively determine how a genetic variant may exert its effect, and to rule out additional pathways not mediated by the exposure. Fortunately sensitivity tests have been developed and can be applied to model pleiotropy.[76]

Although MR has been most conventionally employed in the field of epidemiology to estimate the effect of a risk factor on an outcome using genetic variants that associate with the risk factor,[77] as mentioned, in recent years it has also been applied to estimate causal effects of molecular QTLs on complex traits. For example, a study in 2019 used summary statistics from multiple GWAS of complex traits as well as multiple eQTL databases and identified over 2300 genes causally associated with one or more phenotypes. In this study, 36% of the MR significant genes did not reach genome-wide level significance in the GWAS, highlighting the ability of this approach to identify novel genes or molecular phenotypes causally associated with diseases, even when they are undetectable by more conventional approaches, such as GWAS, due to power limitations. Within the field of psychiatry, the few studies that have implemented MR have uncovered novel causal relationships. For example, a study in 2020 identified a causal association between increased signaling of an inflammatory marker, IL-6 and suicidality,[78] and another MR

study found glycemic traits were causally associated with AN.[79] However, to my knowledge, no MR study has used molecular QTLs on exclusively psychiatric phenotypes. Given that the stress response is highly implicated in psychiatric disorders, we used MR to investigate whether the stress-modulating functional genetic variants were causally involved in psychopathology. This framework is especially advantageous as the direct and reductionistic setup of the STARR-seq (i.e. directly testing the effect of a variant on RE activity) decreases the confounding effect of pleiotropy and ensures the causal estimates are valid.

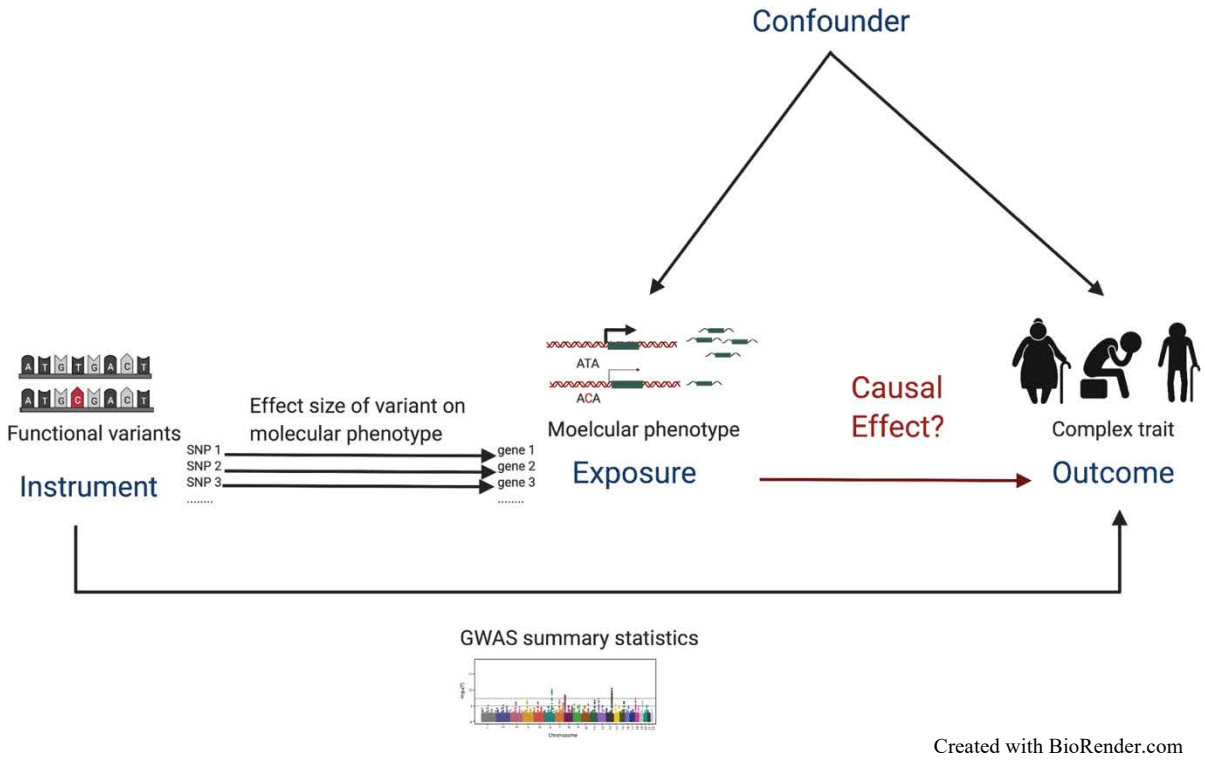


Figure 5. Overview of Mendelian Randomization. Using information about the effect between the instrument and outcome (from GWAs) and the instrument and exposure, the causal exposure-outcome relationship can be assessed, reducing the effect of confounders.

2. Objectives of research

The overarching aim of this work was to functionally fine-map and characterize REs involved in the genetic moderation of the molecular stress response and to better understand how they relate to psychopathology.

2.1. Identification of functional variants involved in the transcriptomic response to stress

In the first part of this work, I employ a high throughout reporter assay to test 3662 variants located in putative REs in order to identify those that are functional. Firstly, I identify dex responsive REs (DREs). From these DREs, I determine which of these DREs harbor functional genetic variants.

2.2. Characterization of regulatory elements modulating the stress response

The second objective of this work was to characterize the REs by integrating GR ChIP-seq, chromatin conformation assays, and in silico modelling to better understand the genomic and epigenomic landscape surrounding the stress-moderating functional variants.

2.3. Causal relationship between stress-mediating variants and psychiatric traits

The first objectives focus primarily on identifying functional variants and understanding the mechanisms by which the stress response is mediated. However, given the strong relationship between stress exposure and psychiatric disorders, I aimed to determine whether the same genetic variants involved moderating the transcriptomic stress response are also causally involved in psychopathology. To this end, I use GWAS identified psychiatric risk loci to perform enrichments analyses. To estimate a causal effect, I employ two-sample MR.

3. Material and Methods

3.1. Material

3.1.1. Cell lines

- U138MG cells (German Collection of Microorganisms and Cell Cultures GmbH, ACC 291)
- Lymphoblastoid cell line from B-Lymphocyte (Coriell Institute, GM18516)
- GR18 cells (gift from Dr, Sebastiaan Meijsing, Max Planck Institute of Molecular Genetics)

3.1.2. Bacterial Cells

- MegaX DH10B T1R Electrocompetent Cells (Invitrogen, C6400-03)
- Stellar Competent E. coli HST08 Cells (Takara Bio, 636763)
- One Shot ccdB Survival 2T1 (Invitrogen, A10460)

3.1.3. Plasmids (see Table S2)

- pSTARR-seq_human backbone (gift from Dr. Sebastiaan Meijsing, Addgene #71509)
- pSTARR-seq_human with CMV (gift from Dr. Sebastiaan Meijsing, see Table S2)
- pSTARR-seq_human scr sequence (gift from Dr. Sebastiaan Meijsing, Addgene #71509)
- pSTARR-seq_human with FKBP5 binding site (gift from Dr. Sebastiaan Meijsing, Addgene #71509)

3.1.4. Antibodies

- Anti-glucocorticoid receptor (Proteintech, 24050-1-AP)
- Anti-histone H3 (Cell Signaling, 4499T)
- Anti-GAPDH (Millipore CB1001)
- Anti-beta actin (Cell Signaling, 3700)
- HRP-conjugated secondary antibody anti-rabbit (Cell Signaling, 7074)

3.1.5. Restriction Enzymes

- SalI-HF (New England Biolabs, R3138S)
- AgeI-HF (New England Biolabs, R3138S)
- HindII-HF (New England Biolabs, R3104)
- EcoRI-HF (New England Biolabs, R3101)

3.1.6. Reagents and other materials

- Minimum Essential Medium Eagle (Gibco, 31095-029)
- Sodium pyruvate (ThermoFisher, 11360070)
- Antibiotic-antimycotic (ThermoFisher, 15240-062)
- RMPI medium (Biochrom, FG 1385)
- Fetal bovine serum (Thermo Fisher, 10270-106)
- Dexamethasone (Sigma-Aldrich, D4902,) resuspended in 100% ethanol
- Ethanol (VWR, 1.00983.1011)
- Cell Fractionation Kit (Cell Signaling, 9038)
- Protease inhibitor cocktail (Sigma-Aldrich, P2714)
- Qiazol Lysis reagent (Qiagen, 79306)
- RNeasy Mini Kit (Qiagen, 74104)
- UltraPure™ Phenol:Chloroform:Isoamylalcohol 25:24:1 (ThermoFisher, 15593031)
- QuantiTect Reverse Transcriptase kit (Qiagen, 20531)
- Methanol-free 16% Formaldehyde Ampules (ThermoFisher, 28908)
- ECL Western Blotting Substrate (Pierce, Pier32106)
- Phosphate Buffered Saline tablets (Merck, P4417)
- Qubit Protein Assay kit (ThermoFisher, Q33212)
- 12-230 kDa Jess or Wes Separation Module, 8 x 25 capillary cartridges (ProteinSimple, SM-W004)
- Anti-Mouse Detection Module (ProteinSimple, DM-002)
- Anti-Rabbit Detection Module (ProteinSimple, DM-001)
- Protan 0.45um nitrocellulose Cytiva, 15259794)

- Taqman Fast Advanced Master Mix (4444964, ThermoFisher)
- TapeStation d5000 Reagents (Agilent Technologies, 5067-5589)
- TapeStation d5000 ScreenTape (Agilent Technologies, 5067-5588)
- Dynabeads M-280 Sheep Anti-Rabbit IgG, (Thermofisher, 11203D)
- Bovine Serum Albumin (Sigma Aldrich, 05482-25G)
- Powdered Milk (Roth, 68514-61-4)
- LiCl (Roth, 3739)
- TRIS (Roth, 77-86-1)
- Igepal CA-630 (Sigma Aldrich, I8896)
- Sodium deoxycholate (Sigma Aldrich, D6750)
- Sodium bicarbonate (Sigma Aldrich, S5761-500G)
- Sodium n-Dodecyl Sulfate, 20% Solution (Sigma Aldrich, 05030)
- RNAse A (Merck Millipore, 70856)
- Proteinase K (Merck Millipore, 70663)
- QIAquick PCR Purification Kit (Qiagen, 28104)
- Qubit dsDNA HS Assaykit (Invitrogen, Q32851)
- LightCycler® 480 SYBR Green I Master (Roche, 04707516001)
- NEB Next Ultra II Library Preparation kit (New England Biolabs, E7645L)
- NEBNext Multiplex Oligos for Illumina Index Primers Set 1-3, 96 Index Primers (E7335S, E7500S, E7710S, E6609S, New England Biolabs)
- KAPA Library Quantification Kits (Roche, 07960140001)
- llumina Free Adapter Blocking Reagent (Illumina, 20024145)
- Novaseq 6000 S1 Reagent V1.5 (300 cycles) (Illumina, 20028317)
- In-Fusion HD Cloning Plus kit (Takara Bio, 638910)
- Kappa HiFi HotStart ReadyMix (Roche, KK2601)
- Agencourt AMPure XP beads (Beckman Coulter, 10453438)
- Sodium chloride (Sigma Aldrich, S7653)
- Bactotryptone (Thermo Fisher, 211701)
- Yeast extract (Merck Millipore, 70161)

- NucleoBond Xtra Maxi (Macherey Nagel, 740414.50)
- MiSeq Micro Reagent Kit V2 (300 cycles) (Illumina, MS-102-2002)
- Cell Line Nucleofector Kit (Amaxa, VCA-1003
- Lipofectamine 3000 Transfection Reagent (ThermoFisher, L3000015)
- Opti-MEM I Reduced Serum Media (Gibco, 31985062)
- RNAeasy Midi Kit (Qiagen, 75144
- RNA ScreenTape (Agilent, 5067-5576)
- RNA ScreenTape Ladder (Agilent, 5067-5577)
- RNA ScreenTape Sample Buffer(Agilent, 5067-5578)
- Dynabeads mRNA Purification Kit (ThermoFisher, 61006)
- TurboDNase (ThermoFisher, AM2238)
- RNAClean XP (Beckman Coulter, A63987)
- SuperScript III Reverse Transcriptase (ThermoFisher, 18080051)
- Noveseq 6000 SP Reagent Kit v1.5 (300 cycles) (Illumina, 20028400)
- PeqGold Agarose (VWR, 732-2789)
- RotiGelStain (Carl Roth, 3865.1)
- LB Broth (Lennox) (Roth, X964.1)
- LB Agar (Thermo Fisher, 22700025)
- Bromophenol Blue Sodium Salt (Sigma Aldrich, 114405)
- Glycerol (Roth, 56-81-5)
- Betamercaptoethanol (Roth, 60-24-2)
- Q5 High Fidelity 2X MasterMix (New Engand Biolabs, M0492)
- EvaGreen 20X (Biotium, 31000-T)

3.1.7. Instruments

- Bioruptor Pico (Diagenode)
- Nanodrop 2000 spectrophotometer (NanoDrop Technologies)
- Luna Automated Cell Counter (Logos, Biosystems)
- LightCycler 480 Instrument II (Roche, Mannheim)
- Tapestation 2200 (Agilent Technologies)

- Qubit 1.0 Fluorometer (Invitrogen)
- NovaSeq 6000 (Illumina)
- Illumina MiSeq (Illumina)
- Amaxa Nucleofector IIB (Lonza Bioscience, AAB-1001)
- ProteinSimple Wes automated Western Blotting machine (ProteinSimple)
- ChemiDoc MP system (Biorad, 17082)
- LifeEco Thermal Cycler (Bioer, BYQ6078)
- Gene Pulser Xcell (Biorad, 1652660)
- E-Box VX2 Complete Imaging System (Vilber Lourmat)

3.2. Methods

3.2.1. Cell culture

U138MG cells were purchased from the German Collection of Microorganisms and Cell Cultures GmbH. Cells were cultured in Minimum Essential Medium Eagle supplemented with 10% fetal bovine serum, 1% sodium pyruvate and 1% antibiotic-antimycotic, and were passged every 5-7 days. GR18 cells were provided as a gift from Dr. Sebastiaan Meijsing from the Max Planck Institute of Molecular Genetics and were cultured in Dulbecco's Modified Eagle Medium with high glucose supplemented with 10% fetal bovine serum and 1% antibiotic-antimycotic, and were passaged every 2-3 days. LCLs were cultured in RMPI media supplemented with 10% fetal bovine serum and 1% antibiotic-antimycotic, and passaged every 3-4 days. All cells were incubated at 37°C and 5% CO_2.

3.2.2. Pharmacological activation of GR

To activate GR, cells were treated with dex. For all experiments involving dex, cells were treated with either 100nM dex or 0.001% ethanol as a vehicle control (veh) for 4 hours, unless stated otherwise.

3.2.3. Characterization of GR in model cell lines

3.2.3.1. Fractionation

To assess the translocation of GR upon ligand binding, dex and veh treated GR18 and U138MG cells were split into three distinctive fractions; whole cell, nuclear/cytoskeleton, and cytoplasm using a cell fractionation kit (Cell Signaling). The fractionation was performed according to the manufacturer's instructions except for the following two changes: from the resuspended cell pellet, 100uL was removed for the whole cell fraction and centrifuged at 350g at 4°C and for 5 minutes. The resulting pellet was resuspended in 100uL RIPA buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% Igepal, 0.5% sodium deoxycholate, and 0.1% SDS)
with 1x protease inhibitor and constituted the whole cell fraction. Secondly, the sonification steps were performed on the Bioruptor Pico using 5 cycles (30s on/30s off). Fractions were quantified using the Qubit Protein Assay Kit on the Qubit 1.0 Fluorometer.

3.2.3.2. Analysis of subcellular protein fractions

To assess the translocation of GR into the nucleus upon stimulation with dex, GR expression in the subcellular fractions were analyzed on the ProteinSimple Wes automated Western Blot system according to the manufacturer's instructions, using 0.5ug of protein per sample and the following antibody dilutions: GR (1:100), H3 (1:100), GAPDH (1:200), and beta-Actin (1:50). The appropriate secondary antibody modules were purchased from ProteinSimple (anti-mouse and anti-rabbit) and used according to the protocol.

3.2.3.3. Quantification of GR target gene expression

Two million of both GR18 and U138MG cells were treated with dex or veh. After four hours, the cells were collected and pelleted by centrifugation at 2000rpm for 5 minutes at 4 °C. Cells pellets were lysed in 750μL of Qiazol lysis reagent and stored at -80°C until RNA was extracted. Total RNA was extracted using the RNeasy mini kit. Briefly, samples were thawed on ice before 100μL of Phenol-Chloroform-Isoamyl mixture was added. Samples were shaken for 15 seconds, incubated for 2-3 minutes, and centrifuged for 15 minutes at 15 000g at 4°C for phase separation. The aqueous phase was extracted and washed with 70% ethanol before being transferred to a spin column. Columns were washed with RWI and RPE, according to the protocol. RNA was eluted in 30μL of RNase free water, and quantified using the Nanodrop 2000 spectrophotometer.

Using the Quantitect Reverse Transcriptase kit, 1ug of RNA was used to synthesize cDNA according to the manufacturer's protocol. To confirm the activation of GR, the induction of three GR target genes after treatment with dex was quantified. To this end, qPCR was performed in technical triplicates using gene specific IDT Taqman probes for FKBP5, SGK1, TSC22D3 and the housekeeping gene, YWHAZ (see Table S1, primers 1-4). Each reaction contained 2uL of diluted cDNA (1:10 dilution), 5uL of Taqman Fast Advanced Master Mix, 1ul of the gene specific probe, 2uL of water. The reaction was amplified on the LightCycle 480 with the cycling conditions summarized in Table 1.

Table 1. Cycling conditions for PrimeTime gene expression analysis LightCycler 480 Instrument II (Roche, Mannheim, Germany)

UNG Incubation	50 ˚C	2 mins	
Polymerase activation	95˚C	20s	
Denature	60°C	15s	
Anneal/Elongation	72°C	30s	65 cycles

The fold change of FKBP5 expression in the dex condition compared to the veh control was calculated using the following formula:

$$2^{(\text{FKBP5 dex - FKBP5 veh)- (YWHAZ dex - YWHAZ veh)}}$$

3.2.4. GR binding in model cell lines

3.2.4.1. Validation of the GR antibody for ChIP-seq

As outlined by the ENCODE TF ChIP-sequencing guidelines[80], the antibody used for the GR-ChIP was validated using two independent methods in the LCL cell line. As a primary validation method, the specificity of the antibody was assessed by Western Blotting. As a secondary validation method, a motif analysis was performed on data generated from a pilot ChIP-sequencing experiment using the same GR antibody.

3.2.4.1.1. Western Blotting

Frozen LCL cells pellets (2 million cells) were lysed in 100uL of RIPA buffer containing protease inhibitor. Lysates were sonified for 5 cycles (30s on/30s off) on the Bioruptor Pico. 4x loading buffer (250 mm Tris–HCl (pH 6.8), 8% SDS, 0.2% bromophenol blue, 40% glycerol, 20% β-mercaptoethanol) was added to the protein sample (10ug), for a final 1x concentration. The samples were heated at 95°C for 5 minutes. Proteins were separated on a 10% acrylamide gel for 65 minutes at 140V in 1x running buffer (25 mM Tris-HCl, 0.1% SDS, 190mM glycine) and electrotransferred onto a 0.45um nitrocellulose membrane using the Mini Trans-Blot Electrophoretic Transfer Cell and 1x transfer buffer (90% 1x running buffer, 10% methanol) for 90 minutes at 100V. Membranes were blocked in 5% low-fat milk in TBST buffer (137mM NaCl,

3mM KCl, 25M Tris-HCl, 0.1% Tween-20) for 1 hour at room temperature, rinsed three times for 10 minutes with TBST, and incubated with primary antibodies overnight (1:2000 GR, diluted in TBST, 1:10000 GAPDH) at 4°C. The next day, the blot was rinsed three times for 10 minutes with TBST and incubated with the secondary antibodies for 1 hour at room temperature. The blot was rinsed an additional three times for 10 minutes with TBST before being probed with 1mL of ECL Western Blotting Substrate according to the manufacturer's instructions. Membranes were visualized using the ChemiDoc MP system.

3.2.4.1.2. Motif analyses

To validate the GR antibody, a pilot GR ChIP-sequencing dataset was generated using LCLs. Data was processed as described Section 3.2.4.5. A motif analysis was performed on peaks (FDR <0.05), using the web tool "MEME-ChIP Suite" (meme-suite.org/tools/meme-chip). The classic analysis mode using the vertebrate motif library with standard settings was applied. A motif analysis on the STARR-seq DREs was performed using the online tool, "TRAP (Transcription Factor Affinity Prediction)". The "multiple DNA sequence analysis" was performed using the transfac 2010 vertebrate database and human promoters for the background model.

3.2.4.2. GR ChIP

The GR-ChIP in the U138MG cells was performed in biological triplicates. Each triplicate was stimulated independently. After the dex stimulation, 20 million cells per condition were collected and crosslinked on a rotating platform for 10 minutes in 1% formaldehyde at room temperature. Crosslinking was quenched by adding cold glycine to a final concentration of 0.125M for 5 minutes on the rotating platform. Cells were centrifuged for 5 minutes at 400g at 4°C and washed twice with cold PBS. The cell pellet was snap frozen in liquid nitrogen and stored at -80°C until ChIP was performed.

Once the triplicate samples were collected, the crosslinked cell pellets were thawed on ice and chromatin was sheared according to the NEXSON protocol.[81] Briefly, nuclei were extracted in cell lysis buffer (0.01M Tris-HCl, 0.01M NaCl, 0.2% Igepal, and 1x protease inhibitor) by sonicating for 10 cycles (5s on/30 off) on the Bioruptor Pico Sonifier. Nuclei were pelleted by centrifugation at 2000rpm for 5 minutes at 4°C and washed once with cell lysis buffer. Nuclei

pellets were resuspended in 1mL of shearing buffer (0.01M Tris-HCl, 0.1% SDS, and 0.001M EDTA, 1x protease inhibitor). Extracted nuclei were quantified using the Luna Automated Cell Counter and 10 million nuclei were suspended in 600μL of shearing buffer. Nuclei were sonicated in 3x 200μL aliquots in the Bioruptor Pico Sonifier for 20 x 2 cycles (30s on/30s off). To assess size distribution, 10uL of sheared sample was reverse crosslinked at 95°C for 15 minutes with NaCl at a final concentration of 0.2M. Sheared chromatin samples were treated for 15 minutes with RNAse A at 37°C. To asses size distribution, reverse cross-linked samples were run on the Tapestation 2200 with the DNA d5000 kit. If necessary, chromatin samples were sheared for an additional 1-12 cycles, to ensure that the majority of fragments were 100-300bp. Chromatin was stored overnight at 4°C. For each IP sample, 200uL of Dynabeads M-280 Sheep anti-rabbit IgG beads were washed twice with cold bovine serum albumin (5mg/mL) and blocked overnight on a rotator at 4°C. The following day, 5μg of anti-GR antibody was incubated with the blocked beads for 2 hours at 4°C. Chromatin samples were cleared by centrifugation for 15 minutes, 13000rpm at 4°C. 10μL of each sample was extracted and stored at -20°C to use as an input control. The antibody-bead conjugate was washed twice with cold PBS/bovine serum albumin before 500μL of sample was added. Samples were rotated for 1 hour at room temperature, followed by 1 hour at 4°C. Samples were washed 5 times with LiCL buffer (0.1M Tris (pH 7.5), 0.5M LiCl, 1% NP-40, 1% sodium deoxycholate) for 3 minutes and once with TE buffer (0.01mM Tris-HCl (pH 7.5), 0.0001M EDTA) for 1 minute at 4°C. Samples were eluted in 200uL elution buffer (1% SDS, 0.1M $NaHCO_3$) and placed on a 65°C heating block for 30 minutes, shaking at 1200rpm. To reverse the crosslinking, the supernatant and input controls were treated with NaCl at a final concentration of 0.2M overnight on a heating block at 65°C. Samples were digested with 0.050μg/μL of RNAse A for 30 minutes at 37°C, followed by 0.2μg/μL of Proteinase K for 1 hour at 55°C. Samples were purified using a PCR Cleanup Kit and eluted in 50μL of EB buffer. Samples were quantified using the Qubit high sensitivity dsDNA Assay kit and stored at -80°C until qPCR or library preparation.

3.2.4.3. qPCR to quantify GR ChIP enrichment

Using the ChIP and input control samples, qPCR was performed to assess the efficiency of the ChIP before continuing with sequencing. To this end, two primers were used. One primer, the positive primer, was chosen to amplify a known binding site of the GR (FKBP5 intron 5). The

other primer, the negative control, amplified an intergenic region where the GR has not been shown to bind (TULP-FKBP5). Each qPCR reactions contained 1uL of the ChIP or input samples, 5uL LightCycler 480 SYBR Green I Master, 0.5uL of both the forward and reverse 10uM primers (Table S1, primers 5-8), and 3uL of water. qPCR was performed on the LightCycler 480 Instrument II, using the cycling conditions summarized in Table 2.

Table 2. Cycling conditions for ChIP-qPCR on the LightCycler 480 Instrument II

Preincubation	95°C	5 mins	
Denaturation	95°C	10s	
Annealing	60°C	10s	
Elongation	72°C	10s	45 cycles
	95°C	5s	
Melting Curve	60°C	1 min	
	97°C	-	
Cooling	40°C	10s	

To calculate ChIP efficiency, the percent input for both dex and veh samples was calculated using the following formula, where Ct represents the cycle threshold for detection. Since only 1% of the input control was collected, the Ct value was corrected.

$$100 \times 2^{\wedge \Delta Ct(Ct\ Input\ -6.64\ -\ Ct\ ChIP)}$$

Then, the resulting percent inputs from the dex and veh treated samples for both the positive and negative primers were compared.

3.2.4.4. Library generation

To generate ChIP-libraries for sequencing, 1ng of each ChIP sample and the corresponding input control were used. Libraries were generated using the NEB Next Ultra II Library Preparation kit according to the manufacturer's protocol, with the following changes: adaptors were diluted 1:40 with Tris-HCl (pH 8.0) and the resulting adaptor ligated DNA was not size selected. PCR was performed using NEBNext Multiplex Oligos for Illumina (Index Primers Set 1-3) using 15

cycles. After the final elution, individual libraries were quantified using the Qubit high sensitivity dsDNA Assay Kit and the molarity assessed using the Bioanalyzer DNA High Sensitivity Assay. Individual libraries were multiplexed based on the molarity of the fragments between 200-1000bp. The multiplexed library was then quantified with the KAPA Library Quantification Kit according to the manufacturer's protocol using the LightCycler 480 Instrument II Finally, the multiplex library was treated with 30uL of the Illumina Free Adaptor Blocking Reagent according to the manufacturer's protocol.

3.2.4.5. Sequencing and data analysis

The U138MG GR-ChIP libraries were sequenced at the Sequencing Core facility of the Max Planck Institute of Molecular Genetics, Berlin on the Illumina NovaSeq 6000 platform. Paired-end sequencing reads of 50bp were generated using an S1 flowcell. The LCL GR-ChIP libraries were sequenced on the Illumina HiSeq 4000 platform with 50 bp paired-end sequencing. For both libraries, paired end reads were trimmed for adapters using cutadapt. Afterwards, sequenced reads were aligned to hg19 (LCLs) and hg38 (U138MG) with Bowtie2. Duplicate reads mapping to the same genomic positions were filtered using Picard (MarkDuplicates). Peak-calling was performed with MACS2 (q-value cutoff <0.05) on the remaining non-redundant reads. Blacklist regions were removed using bedtools and the ENCODE blacklist regions for the appropriate assembly (hg19 or hg38). Peaks on mitochondrial genes and unassigned contigs were removed. For the U138MG ChIP-seq data, only peaks that were present in two (of three) replicates were included in the consensus peak set.

3.2.4.6. ChIP-seq for GR18 cells

GR ChIP-seq sequencing data for the GR18 cell lines treated with 1uM dex for 90 minutes were publicly available from the sequencing read archive (Replicate 1: SRA accession SRX256867/SRX256891;[82] Replicate 2: ArrayExpress accession E-MTAB-9616[83]). Sequencing data were processed as described previously and were provided by the research group of Dr. Sebstiaan Meijsing (Max Planck Institute of Molecular Genetics, Berlin Germany).[84] Briefly, reads were mapped to hg19 with Bowtie2 and duplicates were removed using Picard (MarkDuplicates). Reads with filtered with SAMTools to only include reads with a mapping

quality of >10. Bigwig files were generated using command bamCoverage from deepTools and normalized with RPKM. Paired end reads were trimmed for adapters using cutadapt. Peak-calling was performed with MACS2 (q-value cutoff <0.01) on the remaining non-redundant reads. Blacklist regions were removed using bedtools and the ENCODE blacklist regions for hg19. Peaks on mitochondrial genes and unassigned contigs were removed.

3.2.5. Hi-C analysis

Processed Hi-C data (interaction frequencies, insulator regions, and TAD boundaries) which was performed on thymidine synchronized U2OS (parental GR18 cells) cells were provided by the research group of Dr. Martin Hetzer (Salk Institute of Biological Studies, University of California San Diego). Data were acquired and processed as described previously.[85] Only data generated from Hi-C experiments performed 360-minutes post-thymidine treatment were analyzed to ensure that the 3D structure had completely recovered from the treatment. A consensus interaction set was generated by creating a union set of regions from both biological replicates using the R package LoopRig 0.1.1.

3.2.6. STARR-seq

3.2.6.1. Generation of the STARR-seq input library

DNA fragments for integration into the human STARR-seq plasmid were designed by using the hg19 genomic coordinates of the GR-SNPs[73] and extending them by 100bp up and down stream to generate 201bp fragments. On either side of the fragment, sequences compatible with the P5 and P7 Illumina adaptors (Figure 6, red) and In-Fusion HD Cloning Plus kit (Figure 6, orange) were added. Additionally,
the following positive and negative controls were included; known positive dex responsive REs defined by previous STARR-seq experiments[86] located near H3K27me ChIP peaks, randomly selected genomic regions and random sequences generated using a Markov Model. The controls fragments were also 201bp in size.

Figure 6. In-Fusion HD cloning (orange) and Illumina (red) compatible fragment design for integration of the GR-SNPs into the human STARR-seq plasmid.

An oligo pool, containing all the fragments and controls, was ordered from Twist Biosciences and resuspended in low TE for a final concentration of 10ng/uL. 10ng of the oligo pool was amplified using the Kappa HiFi HotStart ReadyMix PCR kit and primers complementary to the synthesized fragments (Table 1, primers 9-10) using the following program:

Table 3. Cycling conditions for STARR-seq plasmid library amplification

Initial denaturation	95°C	3 mins	
Denaturation	98°C	20s	
Annealing	60°C	15s	12 cycles
Elongation	72°C	30s	
Final Elongation	72°C	1 min	

The PCR product was purified using the NucleoSpin Gel and PCR purification kit (Macherey-Nagel) and eluted in 30uL of TE buffer. The purified oligo pool was then cloned into the gel-purified STARR-seq vector which was linearized with SalI-HF and AgeI-HF at 37°C for 3 hours. Cloning was performed according to the In-Fusion HD Cloning Plus kit protocol. Briefly, ten parallel reactions were performed, each containing 100ng of linearized STARR-seq vector, 25ng of the oligo pool, and 2uL 5x In-Fusion HD Enzyme Premix, in a total reaction volume of 10uL. The reactions were heated to 50°C for 15 and immediately placed on ice. Reactions were pooled before being purified using a 1.8:1 ratio of Agencourt AMPure XP beads and eluted in 25uL of water.

For the bacterial transformation, ten parallel reactions were performed each containing 20ul MegaX DH10B T1R Electrocompetent Cells and 2uL of the purified STARR-seq oligo plasmid pool. Bacterial cells were electroporated using 1.5V on the Gene Pulser Xcell. Five transformants were pooled and each pool was added to 500mL of LB medium (10g/L sodium chloride, 10g/L

bactotryptone, 5g/L yeast extract) containing 100ug/mL of ampicillin and cultured overnight at 37°C and shaking at 190rpm. The next morning, the cultures were evenly divided into three reactions and the STARR-seq plasmid pool was isolated using the NucleoBond Xtra Maxi kit according to the manufacturer's instructions using an elution volume of 600uL. The supernatants were pooled, and the pool was quantified using the Nanodrop 2000 spectrophotometer. The size distribution was assessed using the Bioanalyzer DNA High Sensitivity Kit.

3.2.6.2. Preparation of the STARR-seq input plasmid library

To determine the representation of the individual fragments within the input library, the plasmid library was sequenced. To prepare the plasmid library for sequencing, five parallel PCR reactions were performed using 5ng of the plasmid pool, 25uL of Kapa Hifi HotStart ReadyMix, 1.5uL of each of two primers (Table S1, 11-12) complementary to the STARR plasmid, and enough water for a total reaction volume of 50uL.

Table 4. PCR conditions to amplify the STARR-seq plasmid DNA pool

Initial denaturation	98˚C	45s	
Denaturation	98˚C	15s	
Annealing	65°C	30s	8 cycles
Elongation	72°C	30s	
Cooling	4°C	forever	

The PCR reactions were combined and purified using Ampure XP beads with a 1:1 ratio of PCR reaction volume to bead volume, and eluted in 50uL of water. The concentration was measured using the Qubit dsDNA HS Assaykit. To make the plasmid pool compatible for Illumina sequencing, five parallel PCR reactions were performed each containing 5ng of purified plasmid DNA, 1.5uL of NEB Universal Primer, 1.5uL of NEBNext Oligo #5, 25uL of Kapa HiFi HotStart ReadyMix, and enough water for a total volume of 50uL, using the following amplification conditions:

Table 5. PCR Conditions for adaptor ligation to the amplified STARR-seq DNA plasmid pool

Initial denaturation	98˚C	45s	
Denaturation	98˚C	15s	
Annealing	65°C	30s	10 cycles
Elongation	72°C	70s	
Cooling	4°C	forever	

The PCR reactions were combined and purified using Ampure XP beads with a 1:1 ratio of PCR reaction volume to bead volume, and eluted in 50uL of water.

3.2.6.3. Sequencing of the STARR-seq input library

The resultant plasmid library was sequenced on the Illumina MiSeq platform. Since the SNPs of interest were located directly in the center of the fragments, 150bp paired end sequencing was performed, using the MiSeq Micro Reagent Kit V2 (300 cycles). Reads were assessed using FastQC and trimmed using the tool, Cutadapt according to the adaptor sequence and quality of the read. The sequences of the ordered fragments were used as the reference for alignment, which was performed using the tool bwa. Since the effect of a single variant was to be assessed in the STARR-seq assay, only reads that perfectly matched the reference sequences were included. The reads per fragments were summed and the distribution was assessed to ensure that the majority of the fragments were represented and no single fragments was majorly overrepresented. A count table was generated and used for downstream analyses.

3.2.6.4. Transfection of GR18 cells with the STARR-seq input library

GR18 cells were transfected with the input STARR-seq library using Amaxa Nucleofector II Kit V and program X-001. Five million cells were transfected with 5ug of input STARR library and transferred to a 15cm cell culture plate, for a total of 6 plates (3 dex, 3 veh). The next morning, the cells were treated with dex or the veh control.

3.2.6.5. Transfection of U138MG cells with the STARR-seq input library

U138MG cells were transfected with the STARR input library using Lipofectamine 3000. 24 hours before transfection, 5 million cells were plated in a 15cm cell culture plate, for a total of

6 plates. The next day, each plate was transfected with 15ug STARR-seq plasmid library according to the manufacturer's protocol for a 6-well plate and 5uL of Lipofectamine, but all volumes were scaled by a factor of 15, as 15-cm plates were used. Eighteen hours following the transfection, 3 plates were treated with dex and 3 plates with a veh control.

3.2.6.6. RNA isolation and STARR-seq library preparation

After 4 hours of treatment, the cells were collected in 2ml of RLT buffer by scraping, followed by homogenization by vigorous pipetting. Total RNA was isolated using the RNAeasy Midi Kit, as per the manufacturer's instructions, with the following changes: no DNA digest was performed on the columns and RNA was eluted twice in 150uL of RNAse free water. Extracted RNA was quantified using the Nanodrop 2000 Spectrophotometer and the RNA integrity number was assessed on the Tapestation 2200 using the RNA ScreenTape Assay. To isolate the mRNA, the Dynabeads mRNA Purification Kit was used according to the manufacturer's instructions with 75ug of total RNA used as input. The mRNA was eluted in 45uL of RNAse-free water. The mRNA was treated with 1uL of TurboDNase (2U/uL) in 5uL of the supplied 10x reaction buffer for 30 minutes at 37°C. Following the treatment, RNA was purified using RNAClean XP magnetic beads with a 1:1.8 ratio of mRNA volume to bead volume and according to the manufacturer's protocol. Purified mRNA was eluted in 30uL of RNAse free-water and quantified on the Nanodrop 2000 spectrophotometer.

To generate cDNA, SuperScript III Reverse Transcriptase was used. For each of the mRNA samples, the mRNA was split into 3-5 identical reactions (depending on quantity of mRNA available in a given experiment). In the first step of cDNA synthesis, each reaction contained 500ng of mRNA, 1uL of a primer with a random unique molecular identifier (UMI, 2uM) (Table S1, primer 13), 2uL of a dNTP mix (2mM), and enough RNAse free water was added for the reaction for a total volume of 14uL. The reaction was heated to 65°C for 5 minutes, before being placed on ice. In the second step, the 14uL of the reaction from the first step was combined with 5uL of the 5x first-strand buffer, 1uL of 0.1M DTT, and 1uL of Superscript III reverse transcriptase. This reaction was heated to 50°C for 60 minutes, followed by 70°C for 15 minutes.

Reactions for the same sample were then pooled and treated with 1uL of RNAse A (10mg/mL) per 100uL of reaction volume for 30 minutes at 37°C. Following the treatment, cDNA was purified using RNAClean XP magnetic beads with a 1:1.8 ratio of cDNA volume to bead

volume according to the manufacturer's protocol. Purified cDNA was eluted in 30uL of RNAse free-water and quantified on the Nanodrop 2000 Spectrophotometer.

To generate the STARR-seq libraries, 25ng of the cDNA was used per reaction. Therefore depending on the cDNA concentration, each sample was divided into 3-5 reactions. Each reaction contained 25ng cDNA, 1.5uL of one of a unique index (Table S1, 15-26), 1.5uL of a universal adaptor (Table S1, primer 14), 25uL of KAPA Hifi HotStart Ready Mix, and enough water for a final volume of 50uL. Table 6 summarizes the amplification conditions, which were performed in an LifeEco Thermal Cycler.

Table 6. PCR conditions for STARR-seq library amplification

Initial denaturation	98˚C	45s	
Denaturation	98˚C	15s	
Annealing	65°C	30s	17 cycles
Elongation	72°C	30s	
Hold	4°C	∞	

PCR reactions for an individual sample were combined and purified using a 1:1 ratio of pooled PCR reaction volume to Agencourt AMPure XP beads according to the manufacturer's instructions. Libraries were eluted in 25uL of water and quantified using the Bioanalyzer DNA High Sensitivity kit. Individual libraries were multiplexed based on the molarity of the fragments at the expected size (350 bp). The multiplexed library was then quantified with the KAPA Library Quantification Kit according to the manufacturer's protocol using the LightCycler 480 Instrument II. Finally, the multiplexed library was treated with 30uL of the Illumina Free Adaptor Blocking Reagent according to the manufacturer's protocol.

3.2.6.7. Sequencing and data analysis

The STARR libraries were sequenced at the Sequencing Core facility of the Max Planck Institute of Molecular Genetics, Berlin on the Illumina NovaSeq 6000 platform. Paired-end sequencing reads of 150bp were generated using the Noveseq 6000 SP Reagent Kit v1.5 (300 cycles). Sequencing quality was assessed using FastQC. The forward and reverse reads were

stitched together using the tool, FLASH 1.2.11. Only stitched reads that had exactly 99 overlapping base pairs (Figure 7) were included.

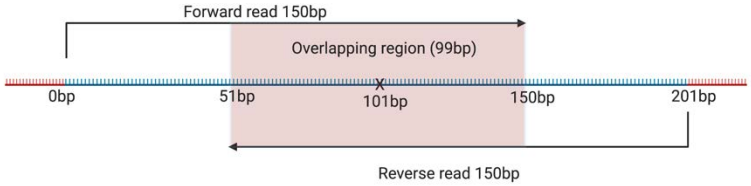


Figure 7. Schematic of read stitching for STARR-seq sequencing reads.

After the forward and reverse reads were stitched together, sequences that contained the same UMI were deduplicated to prevent amplification bias that may have been introduced during the PCR library amplification step. The remaining reads were aligned to the reference sequences (i.e., ordered sequences from Twist Biosciences) and the reads per fragment were summed to generate a count table. Again, only reads that perfectly matched the sequences of the reference fragments were included.

Count tables were imported into RStudio 3.6.2. To assess the correlation between the replicates, the Pearson correlation coefficient on the log-transformed counts was calculated. The fragments were filtered for a minimum read count of ten. For the downstream analysis, we used the R package MPRAnalyze 1.5.1, an R package specifically designed to analyze massively parallel enhancer screen data. It uses both RNA count tables and DNA counts from the DNA plasmid library for normalization. Fragments were normalized for sequencing depth using the total sum scaling method (TSS), using the function estimateDepthFactors. The normalized count tables were then used to identify which of the fragments were active REs. To this end, the 155 negative control fragments were used to generate a null model. Then, using a generalized linear model, the transcriptional activity of each fragment was quantified based on this null distribution using the function analyzeQuantification and the likelihood ratio test. Next, the REs with differential activity after dex were identified using the command analyzeComparative. All normalized fragments were included in the model, but only those identified as being both active REs and showing significantly different (FDR <0.1) activity between dex and veh were considered to be dex responsive REs (DREs). Lastly, regulatory elements with allele-dependent activity (variant-DREs) were identified using the MPRAnalyze command analyzeComparative. Since allele effect on DRE activity was to be estimated in each condition, a new model was required where the reference and alternative

alleles were compared independently in the veh and dex conditions. The results were filtered post-hoc to include only those considered DREs in at least one allele. For this analysis, the Wald test was performed and an FDR cutoff of <0.1 was implemented.

3.2.7. Validation of STARR-seq results

To validate the results from the STARR-sequencing, the activity of individual fragments was tested to determine whether they showed the same direction of effects when tested individually as was observed in the STARR-seq data. In total, twelve fragments were tested in the GR18 cell lines; two dex inducible control REs, two STARR-seq identified DREs, two negative controls (no/little transcriptional activity) and six STARR-seq identified variant-DREs (three putative regulatory elements x two alleles).

3.2.7.1. Cloning

The twelve individual fragments were ordered as "Gene Fragments" from Twist Biosciences. Identical to the fragments ordered for the STARR-seq experiment, the region of interest of each fragment was 201bp long. On either side of the fragment, sequences compatible with the P5 and P7 Illumina adaptors (Figure 7, red) and In-Fusion HD Cloning kit (Figure 7, orange) were added, as well as an additional 21 bp adaptor required for synthesis by Twist Biosciences. The individual fragments were resuspended in low TE buffer to a final concentration of 10ng/uL. Using the 1x Kapa HiFi HotStart polymerase, along with 1uL of 25mM $MgCl_2$, 5uL of 5x Kapa HiFi HotStart 5X Buffer, 1.5uL of 10mM dNTPs, and 0.75uL of the both the 10uM forward and reverse primers complementary to the adaptors (Table 1, 27-28), 10ng of the individual fragment was amplified using the following conditions on the LifeEco Thermal Cycler.

Table 7. PCR conditions for amplification of individual putative RE DNA fragments

Initial denaturation	95°C	3 mins	
Denaturation	98°C	20s	
Annealing	60°C	15s	18 cycles
Elongation	72°C	30s	
Final Elongation	72°C	1 min	

The amplified fragments were run on a 2% agarose gel stained with RotiGelStain at 80V for 60 minutes. The ~300bp band was excised and purified using the gel purification kit supplied in the In-Fusion HD Cloning Kit, according to the manufacturer's protocol and eluted in 30uL of water. The purified fragments were quantified on the Nanodrop 2000 Spectrophotometer. The purified PCR products were then cloned into the STARR-seq vector which was linearized with SalI-HF and AgeI-HF (NEB) at 37°C for 3 hours and gel purified. Cloning was performed using the In-Fusion Cloning Kit according to the manufacturer's protocol, with the following changes: a 10:1 ratio of linearized plasmid: insert fragment was used, corresponding to 100ng of the linearized plasmid and 10ng of the fragment. For the bacterial transformation, 2.5uL of the cloning reaction were added to 50uL of Stellar Competent cells and incubated on ice for 30 minutes. After the 30 minutes, the cells were shocked by incubation at 42°C for 45s. Immediately following the heat shock, cells were placed on ice for 2 minutes before 450uL of prewarmed SOC media were added. The transformed cells recovered in the SOC media by shaking at 160rpm for 1 hour at 37°C. Following this recovery, 20uL of the culture were streaked onto agar plates containing 100ug/mL of ampicillin, and the plates were incubated overnight at 37°C.

The following day, single colonies were selected and placed in 3mL of LB medium containing 100ug/mL of ampicillin for 8 hours at 37°C, shaking at 160rpm. After 8 hours, 200uL of the culture were added to 200mL of LB Lennox medium containing 100ug/mL of ampicillin and cultured overnight at 37°C, shaking at 190rpm. The next morning, the plasmids were isolated using the NucleoBond Xtra Maxi kit according to the manufacturer's instructions with an elution volume of 500uL and quantified using the Nanodrop. For cloning quality control, 500ng of the cloned product were digested with 1) no restriction enzyme 2) HindIII and 3) HindIII and EcoRI for 3 hours at 37°C and the products were run on a 1% agarose gel for 1 hour at 80V. If bands of the expected size were visible, the cloning products were sequenced using the Eurofins DNA Sanger sequencing service, using a single primer (Table S1, Primer 35).

3.2.7.2. Transfection with the individual fragments and RNA/DNA extraction

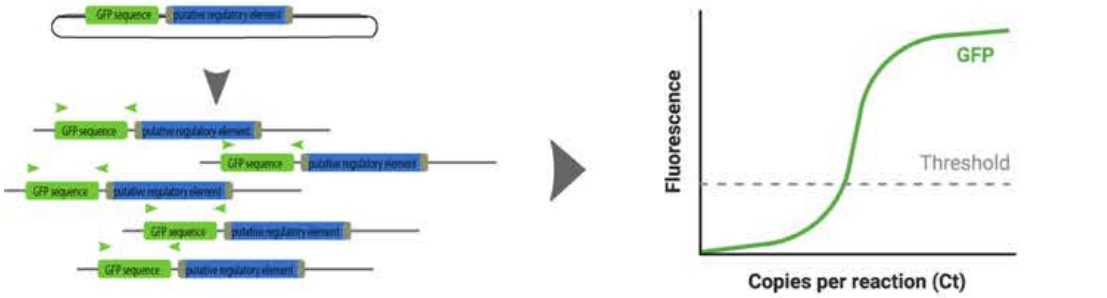
Two million GR18 cells were transfected with 2ug of the STARR-seq vector containing the individual fragment using the Nucleofector IIB, Kit V and program X-001. In addition to the 12 transfections with the STARR-seq plasmids containing the individual fragments, the cells were also transfected with three control constructs; a positive control plasmid containing the CMV

enhancer, a negative control plasmid containing a scrambled CMV enhancer sequence, and a dex-inducible enhancer located in a GR binding site in FKBP5 (see Table S2).

Sixteen to twenty hours following the transfection, the cells were stimulated with dex or the veh (1 well dex; 1 well veh) for 4 hours. The cells were lysed with RLT Plus buffer and RNA and DNA were extracted using the AllPrep RNA/DNA Extraction Kit according the instructions for adherent cell lines. Extracted nucleic was quantified on the Nanodrop 2000 spectrophotometer. From the extracted RNA, 1ug was reverse transcribed to cDNA using the Quantitect Reverse Transcriptase kit following the manufacturer's instructions except two gene specific primers, GFP and Rpl19 (Table S1, 29-30), were used rather than the random hexamers. Since the fragments were cloned downstream of a GFP sequence, the transcriptional activity of the putative REs could be quantified by measuring mRNA containing the GFP sequence (Figure 8). RPL19 was included as a housekeeping gene for normalization purposes. The cDNA was diluted 1:3 (GFP) and 1:10 (RPL19) with water before being used in qPCR.

3.2.7.3. qPCR

Quantitative PCR for RPL19 and GFP was performed using 2.5uL of the diluted cDNA, 5uL of Q5 High-Fidelity 2x MaterMix, 1uL of both the 2uM gene specific forward and reverse primers, and 0.5uL of 20X Evagreen Dye. The following amplification conditions were used on the LightCycler 480 Instrument II.



Created with BioRender.com

Figure 8. Schematic of quantification method for activity of single putative regulatory elements using qPCR.

Table 8. qPCR cycling conditions for the quantification of individual putative RE activity

Initial denaturation	95 ˚C	30s	
Denaturation	98˚C	10s	
Annealing	60°C	30s	45 cycles
Elongation	72°C	20s	
Cooling	37°C	30s	
Hold	4°C	∞	

For each of the primers, qPCR reactions were done in technical triplicates. The experiment was performed in biological triplicates.

To quantify the RE activity after dex, the following formula for fold change was used:

$$2^{\wedge\Delta\Delta Ct}$$

Where ΔΔCt was equal to (Ct GFP dex – Ct GFP veh) – (Ct Rpl9 dex – Ct Rpl9 veh). The ΔΔCt values for the dex inducible controls and DREs were compared to the negative controls and tested for significance using an unpaired student t-test. For the variant-DRE analysis, the ΔΔCt for the reference was compared to the alternative. An unpaired student t-test was used to test whether the allele effect was significant.

3.2.8. Enrichment analysis for variants associated with psychiatric traits

Variant-DREs were overlapped with variants found to be associated (p-value <0.05) with psychiatric traits in a GWAS meta-analysis.[3] To determine whether these variant-DREs were enriched for psychiatric trait variants, an enrichment analysis was performed. To this end, the GR-eQTLs not identified as being allele-dependent by STARR-seq were permutated into sets (of the same size as the variant-DRE set) and overlapped with the psychiatric trait variants. The number of overlapping variants were counted. This was performed 1000 times to generate a null distribution. An enrichment p-value was calculated using the following formula:

$$\text{p-value} = (r+1) / (1000+1)$$

where r is the number of times the background GR-eQTL set contains the same or more overlapping variants than the number observed in the set of interest.

3.2.9. Annotation of DREs and variant-DREs with ChromHMM states

STARR-seq identified DREs and variant-DREs were annotated using the R package HaploReg 4.0.2 using the core 15-state model. Fifteen different broad tissue categories were included to assess whether these variants had regulatory activity in other tissues. To assess whether these variants were active REs in tissues more relevant to psychiatric disorders, a sub-analysis was performed, which functionally annotated these variants in all available brain regions.

3.2.10. Mendelian Randomization

To determine whether there was a causal effect of the differential DRE activity (driven by the variant-DREs) on psychiatric traits, a two-sample MR approach was employed using the R package "TwoSampleMR". The instrumental variable was set as all variant-DREs and the log-fold change between the two alleles was used as the beta estimate. The exposure variables were the corresponding DREs showing allele-dependent activity, and the GWAS summary statistics for the PGC cross-disorder variants were used to define the outcome variables. A FDR threshold of <0.05 was used to define differential DRE activity with a significant causal effect on psychiatric traits.

4. Results

In this dissertation, I identified dex responsive REs (DREs) using the high throughout reporter assay, STARR-seq. From these DREs, I determined those displaying allele dependent activity in the veh and dex conditions (variant-DREs). Furthermore, I integrated the STARR-seq data with functional genomic data to better understand the genomic and epigenomic context in which these variant-DREs reside, using both experimental and in silico approaches. Lastly, I explore the relationship between these variant-DREs and psychiatric disorders.

4.1. Characterization of GR function in GR18 and U138MG cells

In order to identify functional genetic variants modulating the transcriptional response to stress, it was first necessary to confirm the presence of a functional GR; i.e., to ensure that GR translocated from the cytosol into the nucleus upon ligand binding, where it could modulate transcription via DNA binding. It was especially important to characterize GR activity in both of the cell lines used for STARR-seq and other downstream experiments, as both the expression of GR,[87] and its translocation efficiency are cell type specific.[88]

4.1.1. Dex treatment induces nuclear translocation of GR in GR18 and U138MG cells

To confirm GR translocation into the nucleus after activation, both GR18 and U138MG cells were stimulated with dex and separated into cytoplasmic and nuclear fractions. Purity of the fractions was confirmed on the ProteinSimple, an automated Western Blotting system, using known cytoplasmic and nuclear markers, GAPDH and H3, respectively. GR translocation was assessed by quantifying GR expression in each of the fractions at baseline (veh) and after dex. In the U138MG and GR18 cells, there was a 3.8 and 2.4 fold increase of GR in the nuclear fraction after dex when compared to veh, whereas in the cytoplasm there was 0.21 and 0.47 fold change for the U138MG and GR18 cells, respectively.

A B

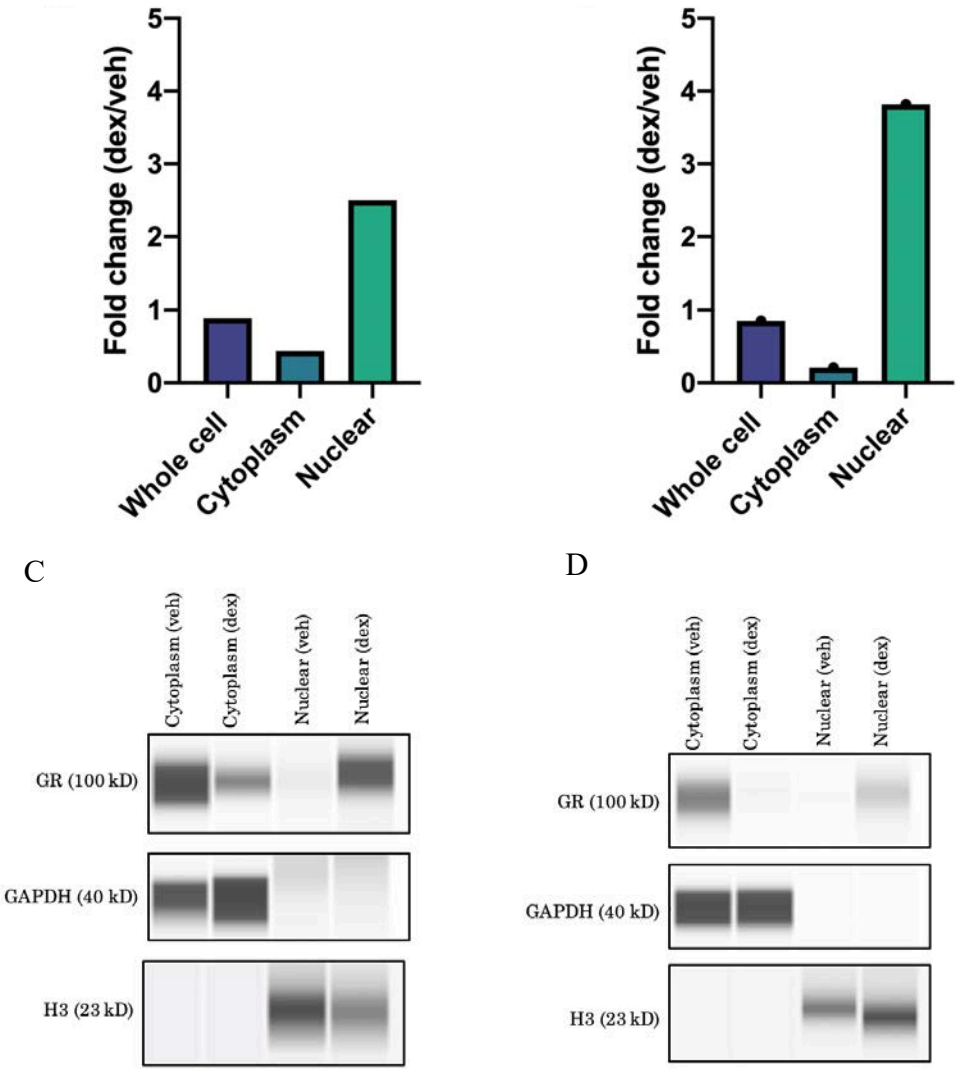


Figure 9. Subcellular fractionation of GR18 and U138MG cells. The fold change of GR expression in the dex condition compared to the veh condition for the whole cell, cytoplasmic and nuclear fractions normalized to actin, H3, and GADPH, respectively in A) GR18 and B) U138MG cells. ProteinSimple blots for the subcellular fractions in C) GR18 and D) U138MG cells.

4.1.2. Dex treatment induces a transcriptional response at GR target genes in GR18 and U138MG cells

To ensure GR was able to induce a transcriptional response in GR18 and U138MG cells upon dex treatment, I quantified the expression of three known canonical GR target genes using gene specific TaqMan qPCR probes. The experiment was performed in biological triplicates. The induction after dex treatment was calculated and compared to a housekeeping gene, YWHAZ, which is not dex responsive. All three genes were induced upon dex treatment in both cell lines. In the GR18 cells, FKBP5, SGK1, and TSC22D3 had mean ΔΔCt values of -1.4 (SD = 0.2), -1.2 (SD = 0.6), and -3.5 (SD = 0.5), corresponding to mean fold changes of 3.7, 3.1, and 15.5 when comparing expression of dex treated to veh treated cells. In the U138MG cells, the induction was weaker for SGK1 and TSC22D3 compared to the GR18 cells, but all three genes were still significantly induced by dex, with mean ΔΔCt values of -1.7 (SD = 0.5), -0.9 (SD = 0.2), and -2.8 (SD = 0.2), corresponding to mean fold changes of 3.3, 3.1, and 7.0 for FKBP5, SGK1, and TSC22D3, respectively.

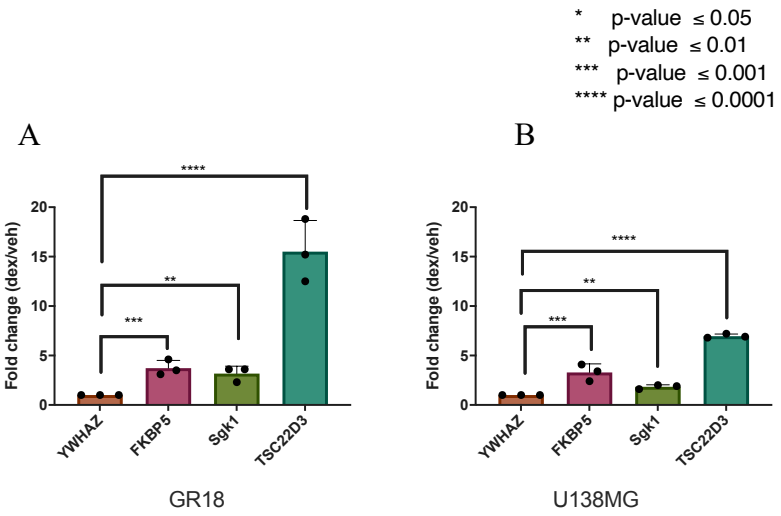


Figure 10. Induction of three dex responsive genes in GR18 (A) and U138MG (B) cell lines using quantitative PCR. Ct values were compared to YWHAZ. Cells were treated for 4 hours with 100nM of dex or 0.001% ethanol (veh). Dex had a significant effect on gene expression for all genes tested in A) GR18 cells ($F(3,6) = 82.71$, $P < 0.0001$) B) in U138MG cells ($F(3,6) = 115.7$, $P < 0.0001$) using a two-way ANOVA analysis.

4.2. Quality control of the synthetic STARR-seq input library

To detect functional variants located within REs regulating the stress response, the 3662 regions containing the eQTLs previously identified to modulate the stress response were tested using STARR-seq. For each of the 3662 eQTLs, two 200bp fragments encompassing the eQTL and a flanking region (100bp up- and downstream of each eQTL) were synthesized, one fragment harboring the reference allele at position 101, and the other the alternative. Additionally, 131 random genomic regions, and 59 known dex-inducible REs were synthesized as negative and positive controls, respectively. All fragments were cloned in parallel into a STARR-seq vector to create a DNA input library. To determine the efficiency of this cloning and the composition of the DNA input library, it was sequenced with 150bp paired-end reads. Since STARR-seq was used to determine the effect of a single genetic variant, only reads that perfectly matched the sequence of the ordered fragments were considered. Nearly 4M of the 5.3M reads met this criterium, which provided sufficient coverage of the input library. One potential issue in high throughput reporter assays is generating a highly biased input library due to factors such as differences in cloning efficiency between fragments or PCR bias, which results in insufficient sequence representation, or a library highly biased to single fragments.[89] Fortunately, the distribution of the fragments in the input library was normally distributed and did not majorly favor a few highly expressed fragments (Figure 11). Furthermore, 99.7% of the fragments were represented. The mean coverage per fragment was 523 reads, with a range from 2-5020.

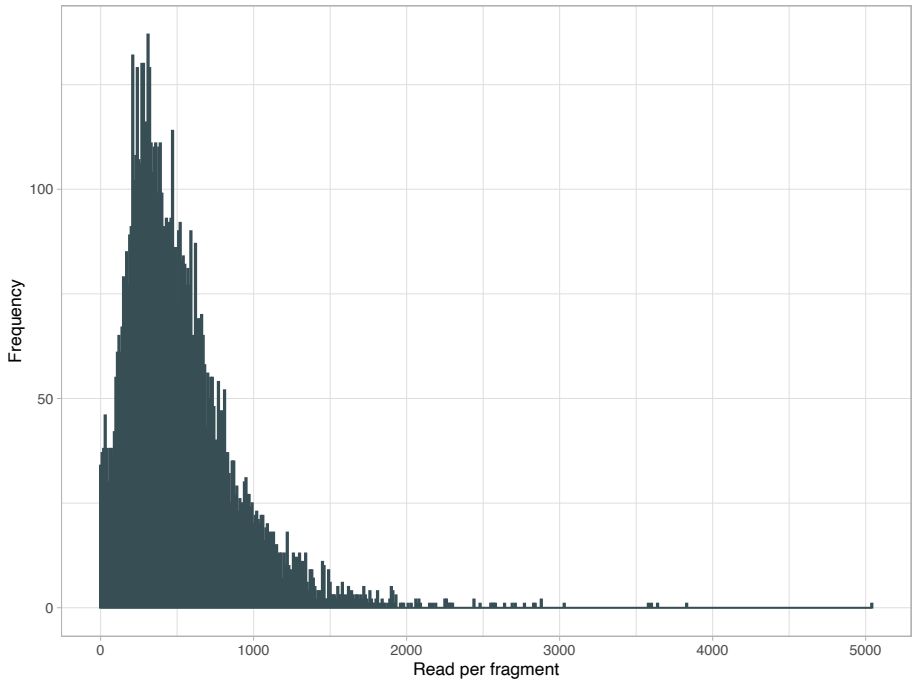


Figure 11. Distribution of DNA fragments in the STARR-seq DNA input library, quantified by 150bp paired-end DNA sequencing.

4.3. STARR-seq in GR18 cells

4.3.1. Quality control of STARR-seq in GR18 cells

After having assessed the quality and composition of the input library, the STARR-seq experiment was performed in biological triplicates. After transfection of the DNA input library, cells were treated with dex or a veh for 4 hours, after which RNA was collected. Libraries were prepared, pooled, and sequenced as described (see Methods). After sequencing, the forward and reverse reads were stitched together to generate one sequence per read. Consistent with the analysis of the DNA input library, only reads perfectly aligning to the synthesized sequences were included. Additionally, a unique molecular identifier (UMI) added in the library amplification step allowed PCR duplicates to be removed during the preprocessing, an amendment to the original STARR-seq protocol that eliminates PCR amplification bias which decreases accuracy.[89] Table 9 summarizes the number of surviving reads after read filtering, stitching, and PCR deduplication per sample. Note that despite having sufficient reads, the number of remaining reads after deduplication for replicate 1 dex (R1_dex) is considerably lower than for replicate 2 (R2_dex) and

replicate 3 (R3_dex), indicating a high percentage of PCR duplicates. Due to this low read count, which was accompanied by a low correlation to R2_dex and R3_dex, (Pearson's $R^2 < 0.8$), which were highly correlated to one another (Figure 12B), R1 (both dex and veh) was excluded from the downstream analyses.

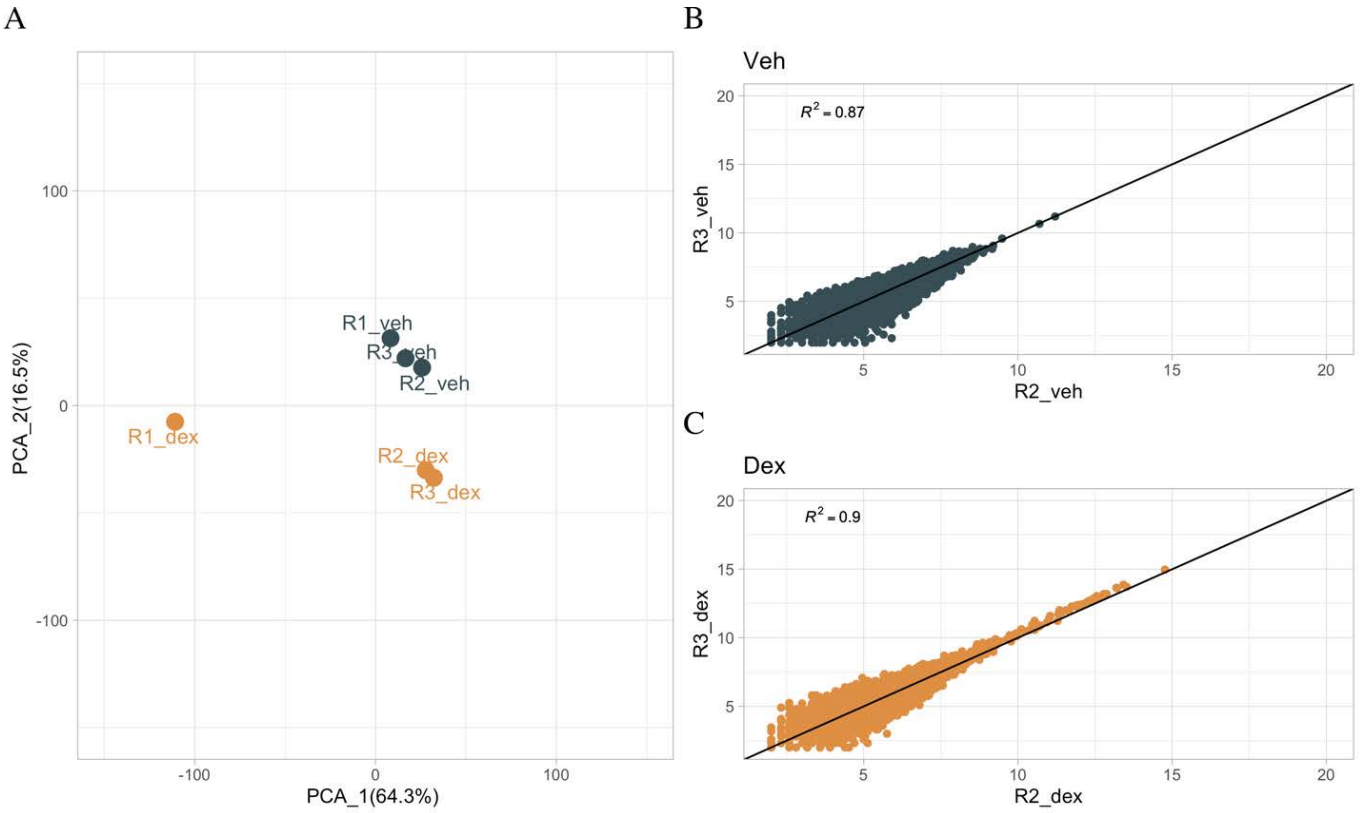


Figure 12. Quality control of GR18 STARR-seq. A) Principal component analysis of the three biological replicates for dex and veh. B) Pearson's R^2 between log-transformed counts of veh treated R2 and R3 C) Pearson's R^2 between log-transformed counts of dex treated R2 and R3.

Table 9. Number of reads prior and post-deduplication for each of the six STARR-seq samples.

Sample	Stitched Reads	Reads remaining post PCR deduplication
R1_dex	24.2	0.39M
R2_dex	22.5	0.81M
R3_dex	40.1	0.90M
R1_veh	22.7	0.49M
R2_veh	27.2	0.54M
R3_veh	21.9	0.51M

4.3.2. STARR-seq identifies active REs in GR18 cells

To identify active REs, the R package MPRAnalyze was employed. MPRAnalyze estimates putative RE activity by using the negative controls fragments (n = 131) included in the STARR-seq library which have no or minimal RE activity. These negative controls are leveraged to generate a null distribution, allowing an estimate of transcriptional activity, termed α, to be generated. Fragments are classified as active or non-active based on the median absolute deviation (MAD) p-values, a measure similar to z-scores but more robust to outliers.[90] Importantly, MPRAnalyze considers the abundance of each fragment in the input DNA library and corrects for this in the model. Overall, 991 (27%) and 1023 (28%) of the putative RE fragments were active in the veh and dex conditions, respectively, and 1131 in the union of dex and veh. A fragment was considered active if it had a MAD FDR <0.1 in at least one of the alleles (reference or alternative). All (100%) of the dex-inducible positive controls were active in the dex condition, but only 53% were active in the veh, confirming the ability of the STARR-seq to identify DREs. Only 5% of the negative controls were classified as active in both the veh and dex conditions.

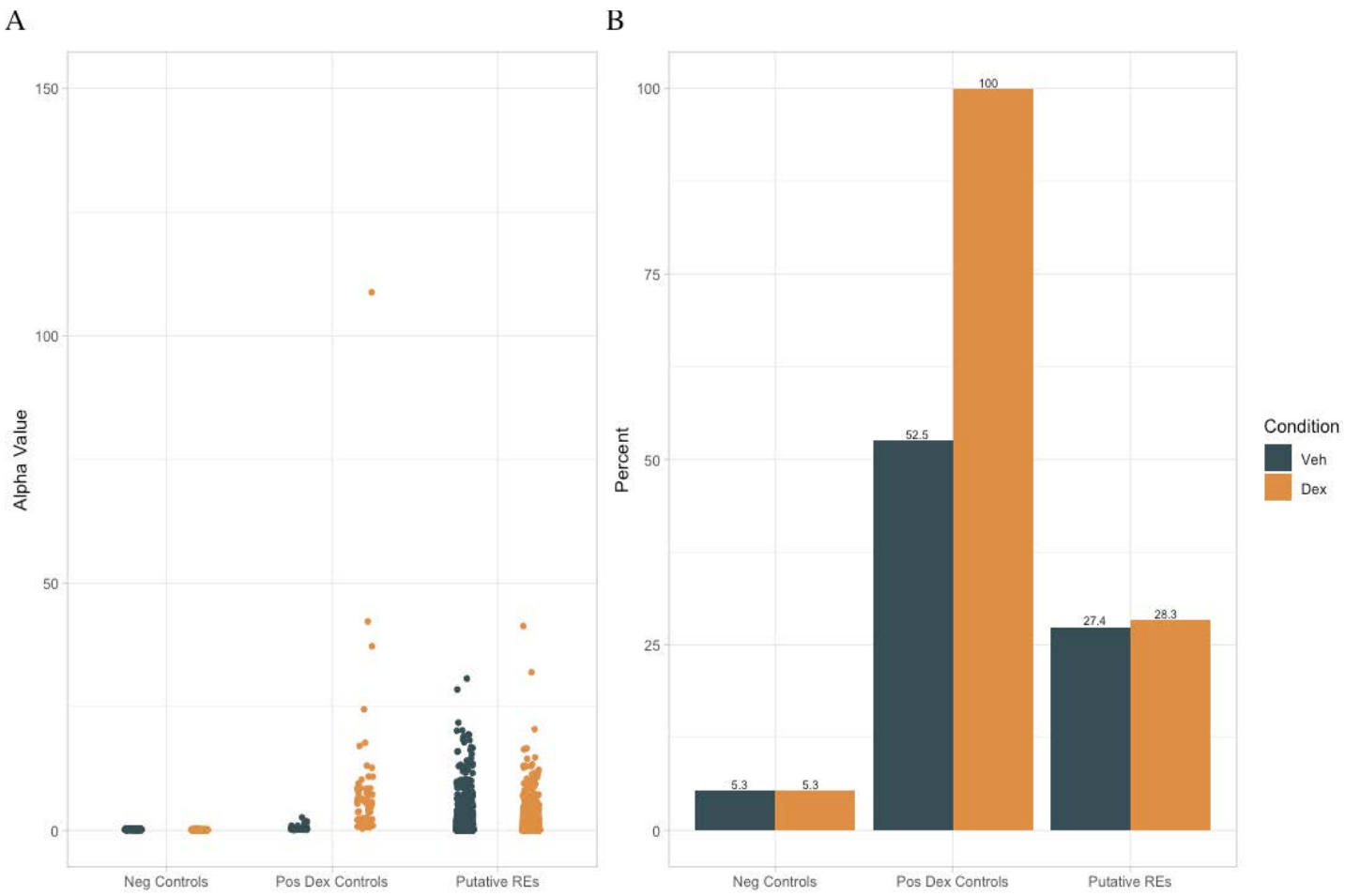


Figure 13. Active REs in GR18 cells. A) Alpha estimates for each negative control, dex-inducible positive control, and putative RE fragments. A higher alpha is associated with higher transcriptional activity. B) Percentage of negative controls, dex-inducible positive controls, and putative RE fragments classified as active (MAD FDR < 0.1).

4.3.3. STARR-seq identifies DREs in GR18 cells

Given our aim of identifying functional variants modulating the transcriptional response to stress, we next needed to identify which of the active REs were responsive to dex. To this end, we compared the activity of a single fragment in the dex and veh conditions using the analyzeComparative function from MPRAnalyze, which identifies REs that are differentially active between conditions. In total, 508 (14.1%) REs were dex-responsive (DREs). An active RE was considered a DRE if it showed significant (FDR <0.1) differential activity upon dex treatment in either of the alleles. Although dex caused both transcriptional activation (52% of DREs) and repression (48% of the DREs) the magnitude of change was nearly twice as great in DREs exhibiting transcriptional activation compared to those exhibiting transcriptional repression, with mean $\log_2$ fold changes of 0.61 [0.013, 4.09] and -0.34 [-1.22, -0.010], respectively. All (100%) dex inducible positive controls were classified as dex-responsive, and 5% of the negative controls.

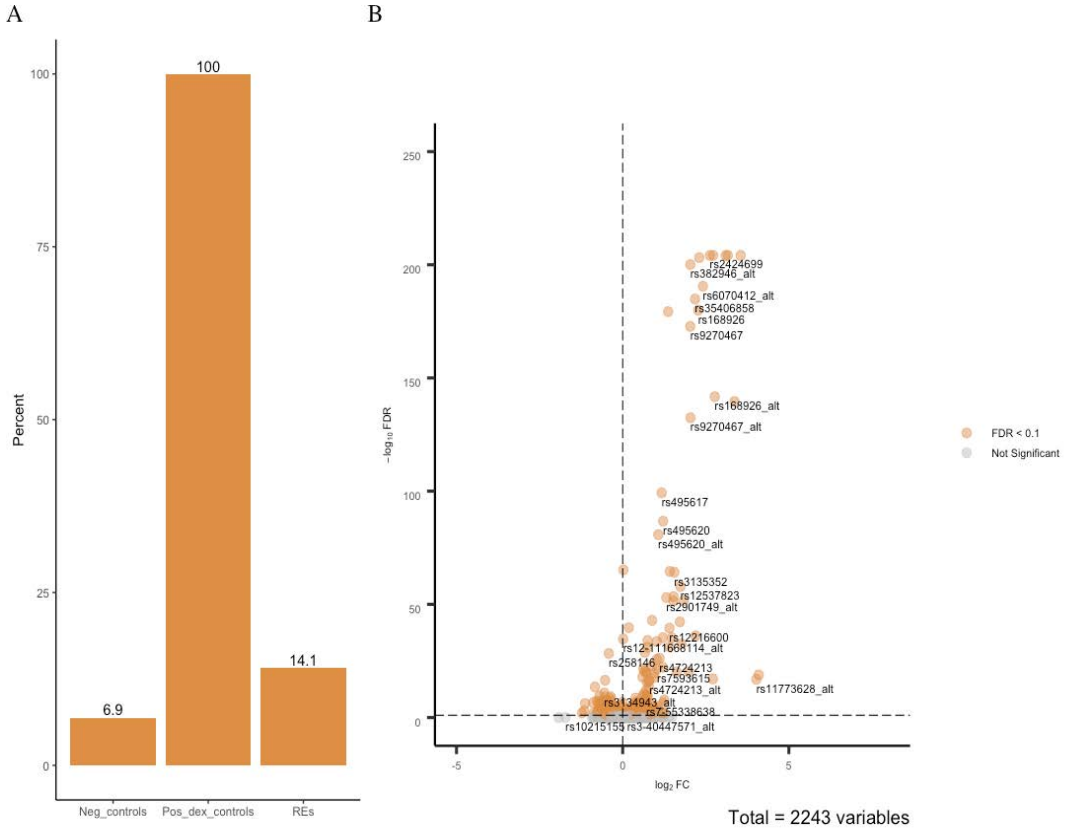


Figure 14. Dex-responsive REs in GR18 cells. A) Percentage of negative controls, dex-inducible positive controls, and putative REs classified as dex-responsive (FDR < 0.1). B) Volcano plot showing the $\log_2$ fold change (x-axis) and statistical significance (y-axis; -log10 FDR) of all active REs. Those is orange show significant dex responsiveness (n=508).

4.3.4. A subset of the DREs display allele dependent activity in GR18 cells

Finally, from the DREs (i.e. active REs showing significant dex-responsiveness) we identified those exhibiting allele-dependent activity. Again, we used the analyzeComparative function from MPRAnalyze. This required a new model to be computed where the activity of the REs containing the reference allele are compared to the same REs harboring the alternative allele. This analysis was performed for both the veh and dex conditions, generating a list of DREs displaying allele-dependent activity at baseline (veh variant-DREs) and after dex stimulation (dex variant-DREs). Although all fragments were included in this model, the results were filtered post-hoc to include only DREs. Of the 508 dex-REs, 154 were veh variant-REs and 164 dex variant-DREs. 92 of these RE-variants overlapped in the dex and veh condition, whereas 73 exclusively showed allele-dependent activity in the dex condition and 62 exclusively in veh. There was no significant difference between the magnitude of the differential activity the veh variant-REs and the dex variant-DREs.

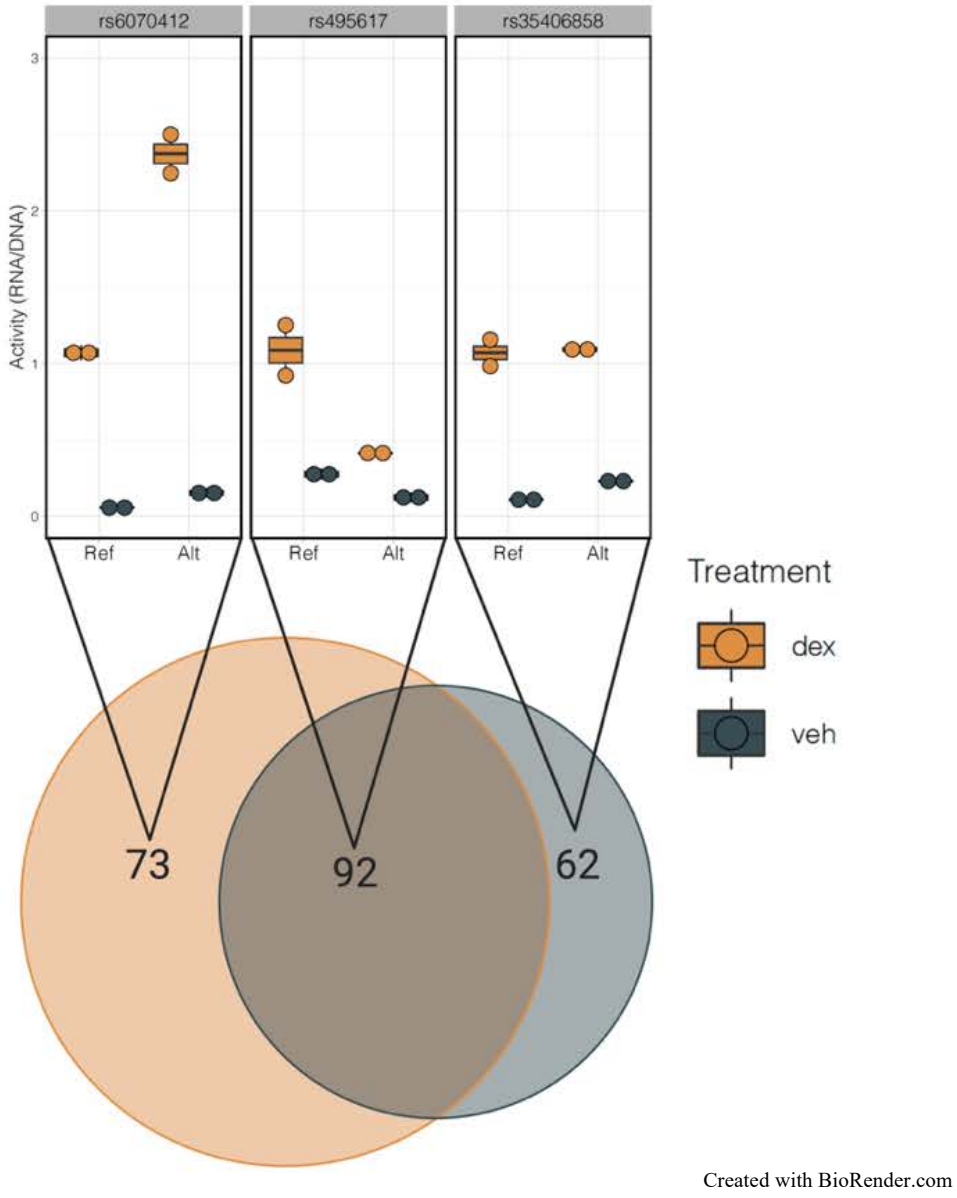


Figure 15. Allele-dependent DREs in GR18 cells. 73 DREs exhibited allele-dependent activity exclusively in the dex condition and 62 exclusively in the veh condition. 92 DREs with allele-dependent activity both in the veh and dex conditions were identified. A representative example of each of these three types of variant-DREs are displayed. Activity of the enhancers are displayed as RNA/DNA ratios.

4.4. STARR-seq in U138MG cells

Given the cell-type specificity of GR activity[91,92] we performed STARR-seq in a second cell line, U138MG cells, a glioblastoma cell line. The same input DNA library (containing the GR-eQTLs) was used, as the aim of identifying functional variants modulating the stress response was the same. Initially, a neuroblastoma cell line (SH-SY5Y) was selected as the second cell line for the STARR-seq, however due to an insufficient response to dex (1.3 fold change of FKBP5 control enhancer activity after dex), likely caused by a low expression of GR[87], it was not an appropriate model for the STARR-seq. U138MG cells, on the other hand, have a high endogenous expression of GR[87], which translocates into the nucleus upon dex treatment, as displayed in Section 4.1.

4.4.1. Quality control of STARR-seq data in U138MG cells

The STARR-seq was performed in biological duplicates and processed using the same analysis pipeline as the GR18 cells. Due to practical considerations, the library was sequenced to a greater depth than the GR18 STARR-seq library with 3.7 to 6.9 million reads remaining after preprocessing (Table 10). Biological duplicates were highly correlated with one another (Pearson correlation coefficient > 0.99, Figure 16B). However, unlike in the GR18 cell lines, the Pearson correlation coefficient between dex and veh samples were also very high (>0.97), indicating the dex treatment had a lesser global effect on the samples compared to what was observed in the GR18 cells. This is also demonstrated by the principal component analysis which displays no clustering of replicates into distinct treatment groups which was observed in the GR18 cells.

Table 10. Number of reads prior and post-deduplication for each of the six STARR-seq samples.

Sample	Stitched Reads	Reads remaining post PCR deduplication
R1_dex	32.7	3.7M
R2_dex	43.6	5.7M
R1_veh	34.1	6.9M
R2_veh	32.2	2.7M

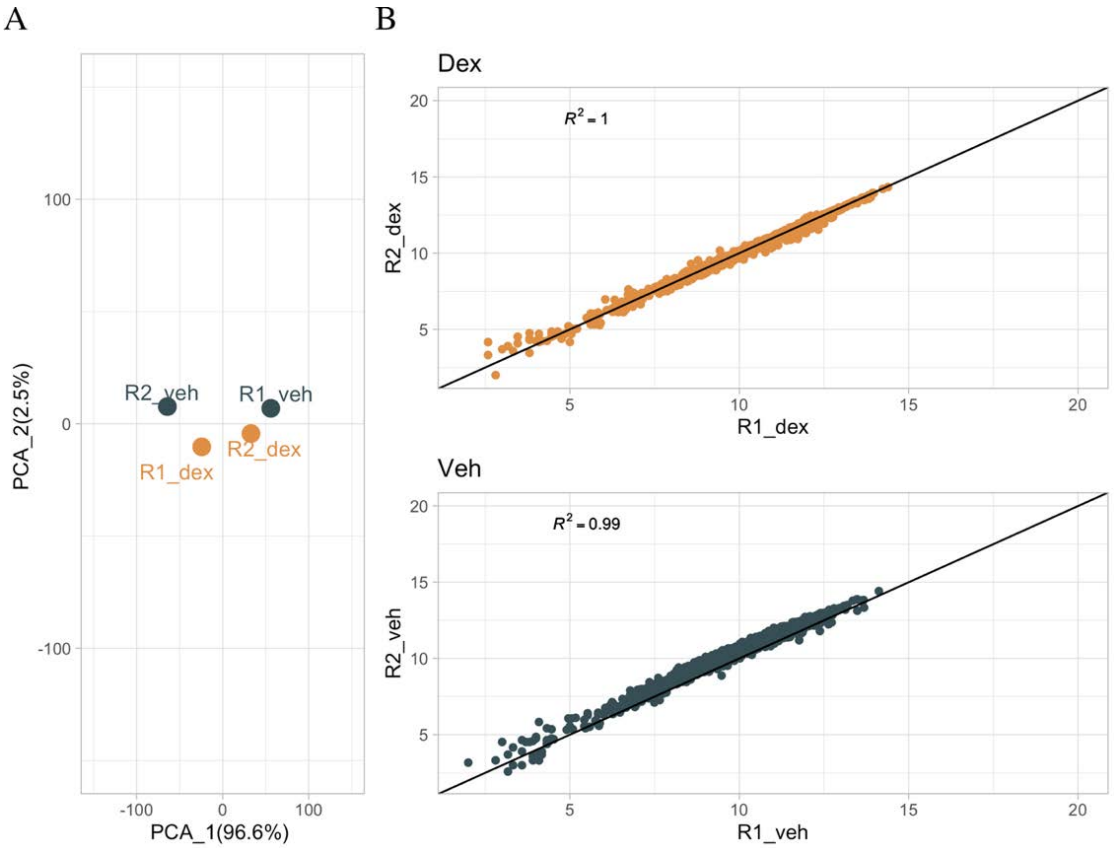


Figure 16. Quality control of U138MG STARR-seq. A) Principal component analysis of the two biological replicates for dex and veh. B) Pearson's R^2 between log-transformed counts of veh treated R1 and R2 C) Pearson's R^2 between log-transformed counts of dex treated R1 and R2.

4.4.2. STARR-seq identifies active REs in U138MG cells

As with the GR18 cells, the first step was to identify active REs using the negative controls to generate a null distribution. Using the same function as that used in the GR18 cells, 870 (23.9%) and 872 (23.9%) of the putative REs were classified as being active in the dex and veh condition, respectively. The union set of the active REs was 894 indicating that the majority of the active REs were active in both conditions. 62.7% of the dex-inducible controls were active in the dex condition, and 37.3% in the veh. This was likely due to the fact that these controls were selected by overlapping the most dex-responsive REs identified from previous STARR-seq experiments with regions flanked by histone marks indicative of active enhancers, both of which were done in GR18 cells, and not U138G cells. Few negative controls (1.5% for both dex and veh) were classified as active, as expected.

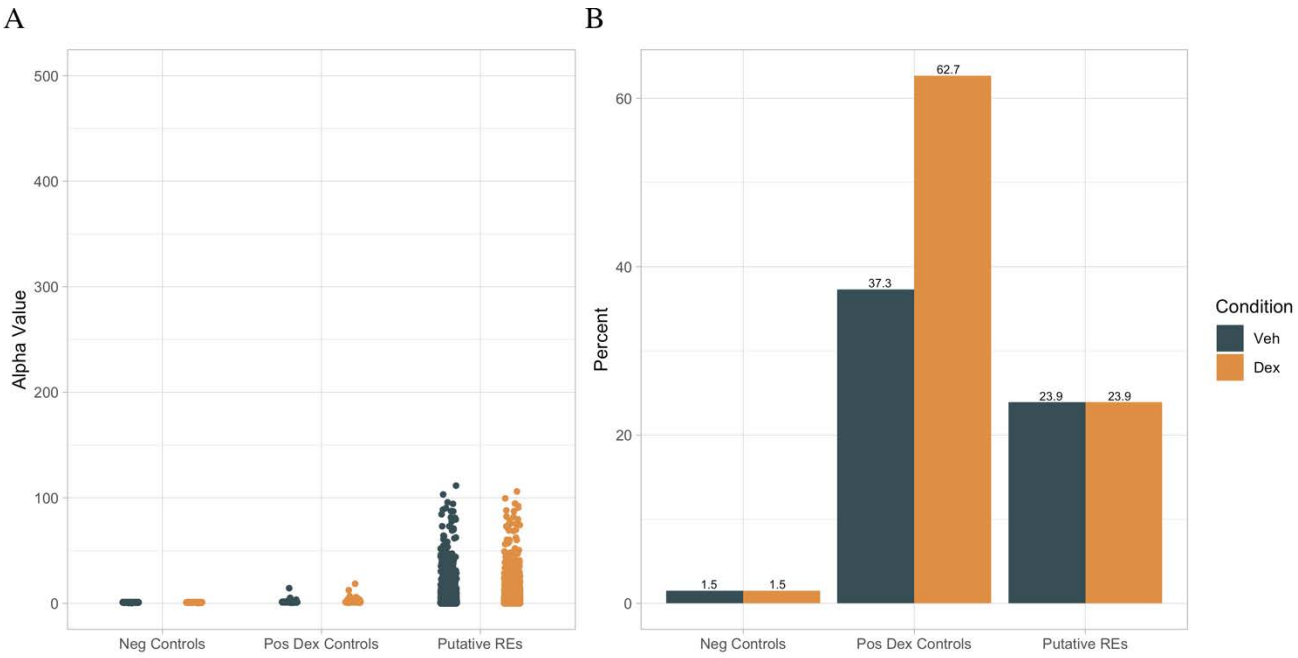


Figure 17. Active REs in U138MG cells. A) Alpha estimates for each negative control, dex-inducible positive control, and putative RE fragment. B) Percentage of negative controls, dex-inducible positive controls, and putative RE fragments classified as active (FDR < 0.1).

4.4.3. STARR-seq identifies DREs in U138MG cells

From all putative REs, only 66 were both active and displayed significant dex-responsiveness (DREs)(FDR< 0.1), corresponding to 1.8% of all fragments. Only 44.1% of the dex-inducible controls were dex responsive, compared to 100% in the GR18 cells, which is again likely due to the cell-type restricted selection method of these controls (see section 4.3.2). Again, the magnitude of the $\log_2$ fold change in DREs showing transcriptional activation upon dex treatment was over twice those displaying transcriptional repression upon dex treatment (0.3 [0.085, 1.6] and -0.14 [-0.40, -0.04], respectively).

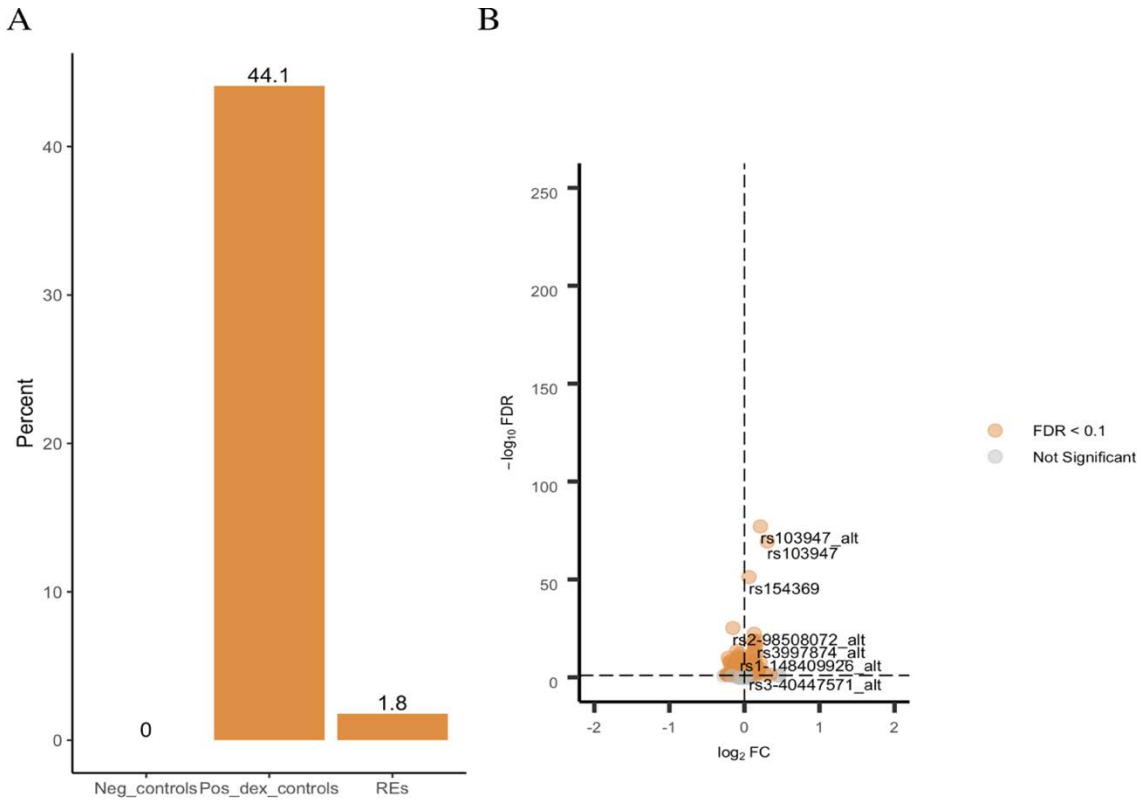


Figure 18. Dex-responsive REs in U138MG cells. A) Percentage of negative controls, dex-inducible positive controls, and RE fragments classified as dex-responsive (FDR < 0.1). B) Volcano plot showing the $\log_2$ fold change (x-axis) and statistical significance (y-axis; -log10 FDR) of all active REs. Those is orange show significant dex responsiveness. In total, 66 active REs were dex-responsive.

4.4.4. A subset of the DREs display allele dependent activity in U138MG cells

The 66 DREs were tested to determine which were variant-DREs (i.e. showing allele-dependent activity). Two of the 66 DREs were veh variant-DREs and ten were dex variant-DREs. Both of the veh variant-DREs overlapped with dex variant-REs, meaning there were no variant-DREs exclusive to the veh condition. Again, no significant difference in the magnitude of the differential allele-dependent activity was observed between the dex variants-DREs and the veh variant-DREs.

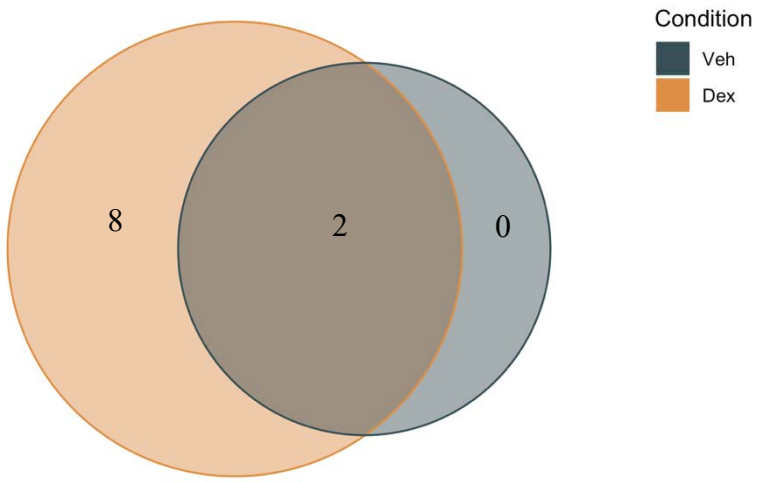


Figure 19. Overlapping variant-DREs in U138MG cells. Ten DREs displayed allele dependent activity in the dex condition (dex variant-DREs), and only two in veh (veh variant-DREs).

4.5. Validation of individual STARR-seq identified REs in GR18 cells

To validate the findings from the STARR-seq, individual fragments were tested for activity with qPCR in GR18 cells. To measure RE activity, the individual fragments were cloned into the STARR-seq vector, downstream of a GFP sequence. GR18 cells were transfected with individual constructs and treated with dex or a veh control for 4 hours. RNA was extracted and cDNA was synthesized using a primer specific for the GFP sequence, and a primer specific to a housekeeping gene, RPL19, for normalization. Since stronger RE induced transcription of the GFP sequence, RE strength could be measured by GFP quantification via qPCR. In total, two negative controls, two positive DRE controls, two STARR-seq identified DREs and three dex variant-DREs (two alleles per RE) were tested (n=12). The experiment was performed in biological triplicates. The induction after dex was quantified and compared to the negative controls, which were not dex-inducible

(average $\log_2$ fold change of 1.0 (SD = 0.15) and 0.9 (SD= 0.03)). All DREs were significantly induced by dex treatment, as compared to the negative controls (paired t-test on the ΔΔCt values), with mean ΔΔCt values of -2.7 (SD=0.4), -3.7 (SD=2.4), -2.7 (SD = 0.3), -5.6 (SD = 0.3) for the two DRE controls and two STARR-seq identified DREs, respectively, corresponding to fold changes of 30, 14, 3, and 33 when comparing activity of dex to veh treated cells. For the dex variant-DREs, the $\log_2$ fold changes in activity between the dex and veh conditions were quantified for the reference and alternative allele fragments. The induction after dex treatment in the variant-DRE containing the reference allele was compared to that harboring the alternative allele. The dex variant-DREs rs382964 and 9525228 showed significant differences in dex responsiveness between the two alleles, both in the direction observed in the STARR-seq data. The third dex-variant, rs9274181 showed no significant difference. It should be noted that the difference to be detected in this variant-DRE was minimal (dex/veh fold change of 1.1 and 1.2 for reference and alternative alleles, respectively) meaning that more power may be required to detect such small differences. Lastly, the magnitude of the $\log_2$ fold-changes of all individual REs (including controls) from the validation assays were compared to the magnitudes observed in the STARR-seq data. The $\log_2$ fold changes calculated by each method were significantly correlated, with a Pearson Correlation Coefficient of 0.98. Overall, the validation experiments in the GR18 cell lines indicated that the results from the STARR-seq were robust, with the majority of the constructs displaying activity patterns congruent with the direction and even relative magnitude of the STARR-seq data. One variant-DRE (rs9274181) did not show a significant allele effect, which may be due to the small effect size attempted to be detected.

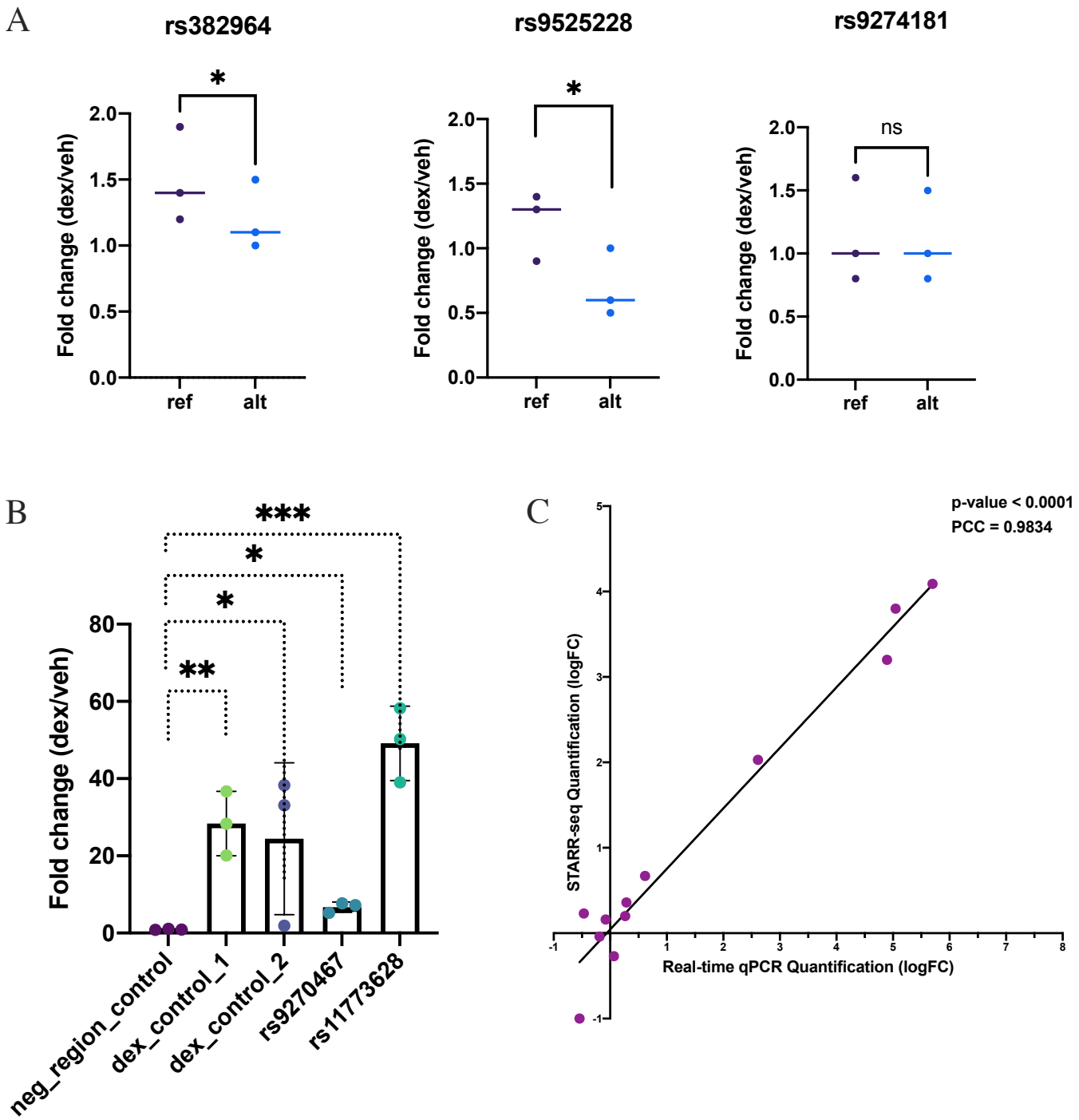


Figure 20. Validation of STARR-seq results using qPCR. Individual fragments were tested for activity in dex and veh conditions. A) The activity of three allele dependent REs (variant-DREs) were quantified. Two of the three variant-DREs displayed significant allele dependent differences in the dex condition. B) The dex-induced activity for two DRE controls and two STARR-seq identified DREs were tested. All showed significant differences in dex responsiveness compared to negative controls in the direction observed in the STARR-seq. C) The correlation between the magnitude of $\log_2$ fold changes observed in the qPCR validation was compared to those observed in the STARR-seq data. *P < 0.05; **P < 0.01; ***P< 0.001. Paired t-test, multiple comparison test.

4.6. A subset of STARR-seq functional REs overlap between GR18 and U138MG cells

Given the cell-specific activity of GR, the results from the STARR-seq in the GR18 cells were compared to those in the U138MG cells. Firstly, all active REs were compared between the two cell lines. Interestingly, the majority of the active REs were overlapping, with only 25% and 5% being exclusive to the GR18 and U138MG cells, respectively. Next, the DREs were compared between the two cell lines. Of the 66 DREs in the U138MG cells, 41% were overlapping with GR18 DREs, indicating that, to a certain extent, the functional effect of the DREs are not exclusively cell-type specific, but a subset may act as functional DREs across tissues. Lastly, the overlap of the variant-DREs between the GR18 and U138MG cells were compared. As expected, due to the overall limited number of variant-DREs in the U138MG cells (2 and 10 for veh and dex variant-DREs, respectively), there were only two overlapping dex variant-DREs and no overlapping veh variant-DREs between the GR18 and U138MG cells.

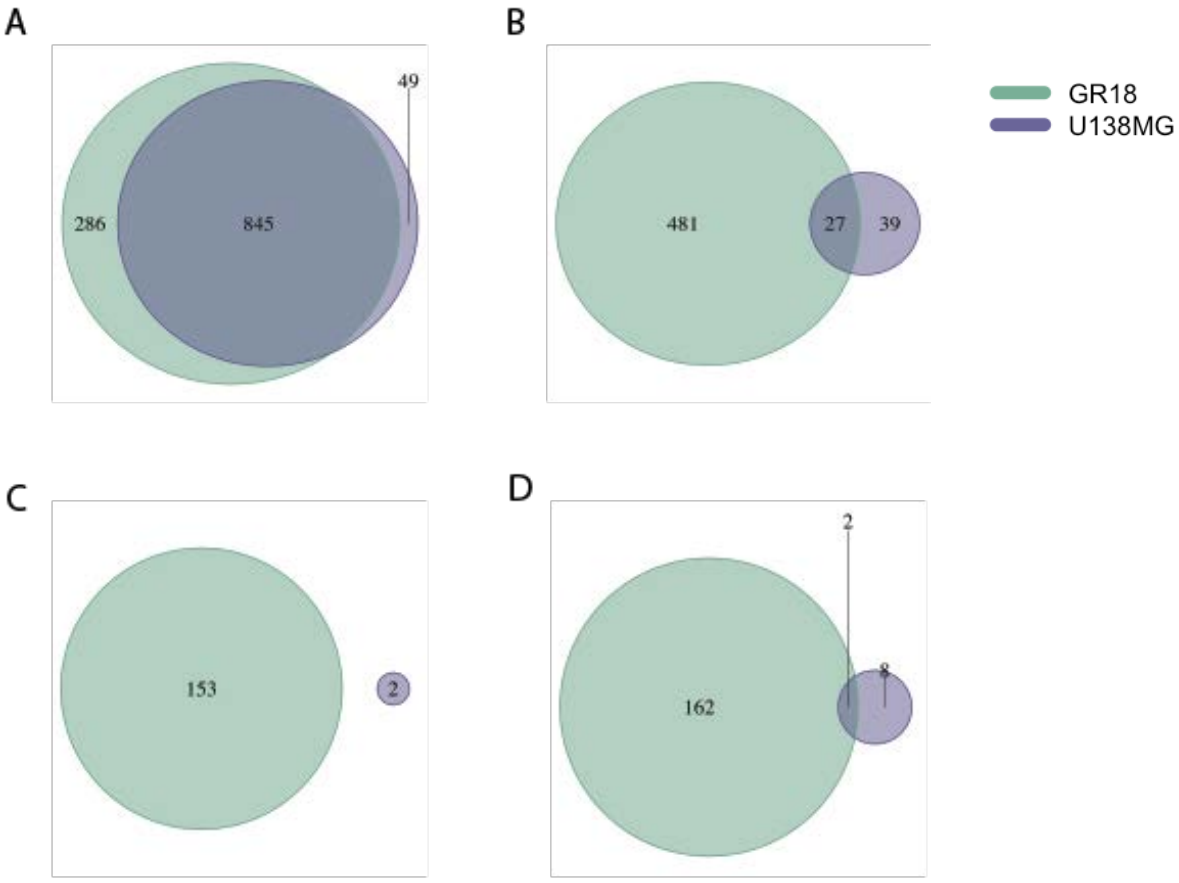


Figure 21. Overlapping functional elements identified by STARR-seq in GR18 and U138MG cells. A) The number of overlapping active REs in the GR18 and U138MG cells. The majority of the identified active REs overlapped between the two cell lines. B) Overlapping dex responsive REs (DREs) between GR18 and U138MG cells. Although far fewer DREs were identified in the U138MG cells, 41% overlapped with those identified in GR18 cells. C) Overlapping DREs displaying allele dependent activity in the veh condition (veh variant-DREs, C) and in the dex condition (dex variant-DREs, D). Due to the limited number of DREs identified in the U138MG, there were few overlapping variant-DREs.

4.7. Functional annotation of STARR-seq DREs and variant-DREs

Having identified DREs and variant-DREs, I next sought to better understand the epigenetic and TF landscapes surrounding these functional elements, and compare these landscapes to the GR-eQTLs not identified by STARR-seq as being functional (non-functional GR-eQTLs). Note, that this GR-eQTL background set was already found to be enriched for various functional annotations, such as TF binding sites, disease-associated loci, and chromatin states associated with enhancer states. We used this highly enriched GR-eQTL background set to determine whether the STARR-seq, a direct experimental approach, could improve the functional fine-mapping when compared to the eQTL and in silico approaches.

4.7.1. The GR ChIP antibody fulfills ENCODE antibody characterization criteria

We first sought to determine whether the DREs were enriched in binding sites of GR. However, in order to fulfill the ENCODE ChIP-seq guidelines, we validated the GR used for ChIP using two different methods. For the primary validation, we performed immunoblotting on whole cell-lysate from the LCL cell line, GM18516. The following criteria outlined by ENCODE were used to assess whether the antibody was suitable for ChIP-seq: 1) The major band must be within 20% of the size predicted by the size of the coding region 2) The major band corresponds to >50% of all bands on the gel.[80] Both criteria were met as the Western Blot clearly showed one single band at the expected size (100kD) (Figure 22A). For the secondary validation method, we performed a motif analysis on pilot ChIP-seq data from the dex-treated LCL cell line. Compared to an IgG control, 2578 peaks were identified. A motif analysis using MEME Suite's MEME-ChIP was performed, an online tool that scans ChIP-seq data and identifies enriched TF motifs.[93] In the GR ChIP-seq pilot data, the GRE motif was significantly enriched (FDR < 0.0038)(Figure 22C). Furthermore, manual scanning of the peaks tracks displayed strong ChIP-seq peaks at canonical GR binding sites (Figure 22B).

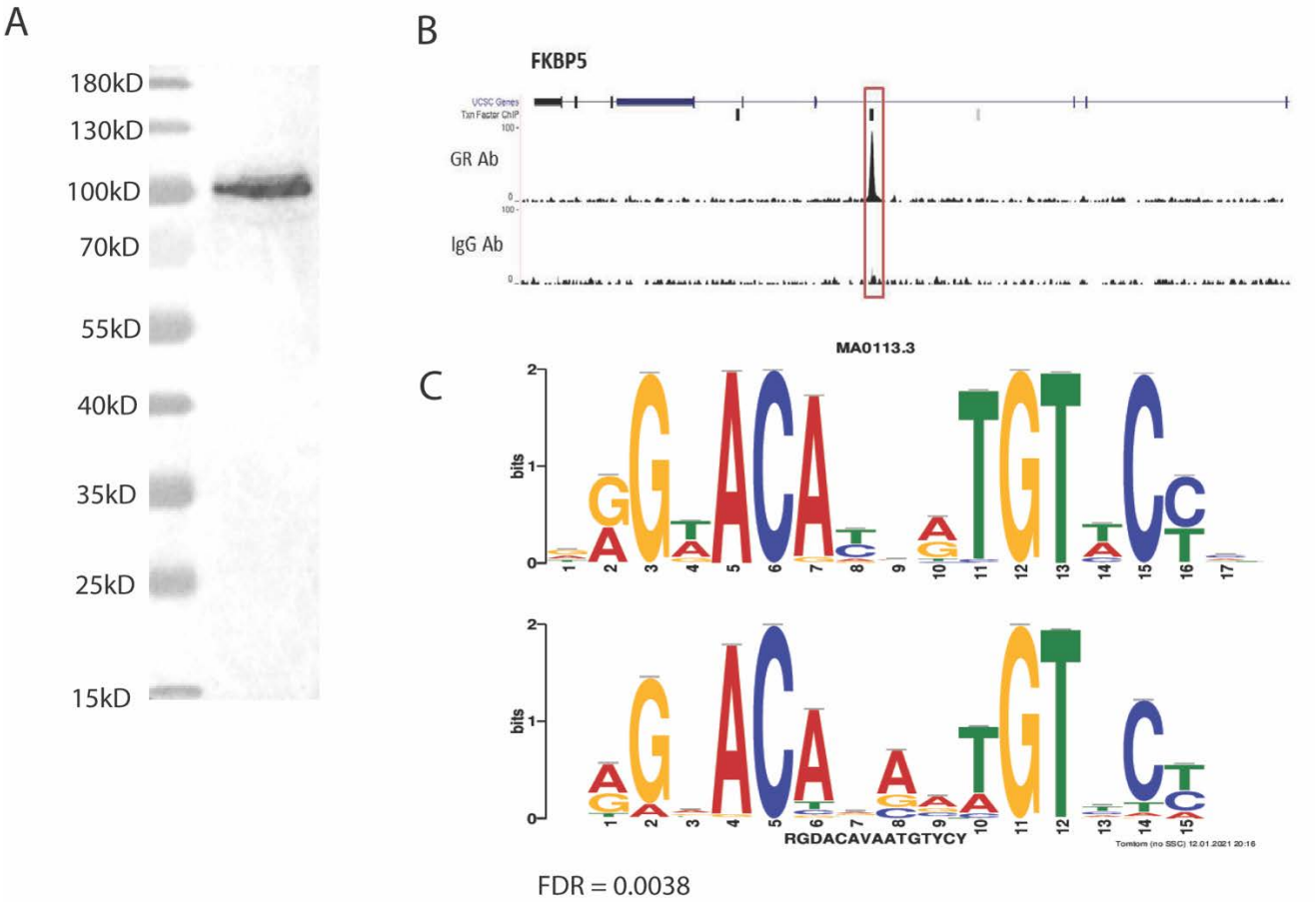


Figure 22. Validation of GR Antibody for ChIP seq. A) The Western Blot using the ChIP-seq GR antibody and whole cell lysate from LCLs resulted in a single band at the expected size. B) Example of a ChIP-seq peak in a known binding site of GR located in FKBP5. C) A motif analysis revealed the GRE motif as being highly enriched within the ChIP-seq data generated.

4.7.2. DREs are enriched in GR binding sites

Firstly, we wanted to determine whether the DREs were enriched within in or near binding sites of the GR, as bound GR is known to recruit co-activators and repressors, as well as other coregulators, that ultimately influence transcription at GR-target genes.[94],[95] Indeed, in both the GR18 and in the U138MG cells, the DREs were enriched in binding sites of GR after dex treatment. In the U138MG cells, 2 (of 66) dex-REs overlapped with a GR binding site. To establish whether this was significant, a null distribution was generated by permutating the non-functional GR-eQTLs into sets of the same size (66), and counting the overlap with GR ChIP peaks. This was performed 1000 times (without replacement). In the null distribution, the mean number of GR-eQTLs overlapping with a GR-ChIP peak was 0.4, meaning the dex-REs had an enrichment

of over 4-fold in GR ChIP binding sites, but this was not significant (permutation p-value 0.1). For the GR18 cells, 23 dex-REs overlapped with GR ChIP peaks, representing a 1.6-fold enrichment (permutation p-value = 0.04).

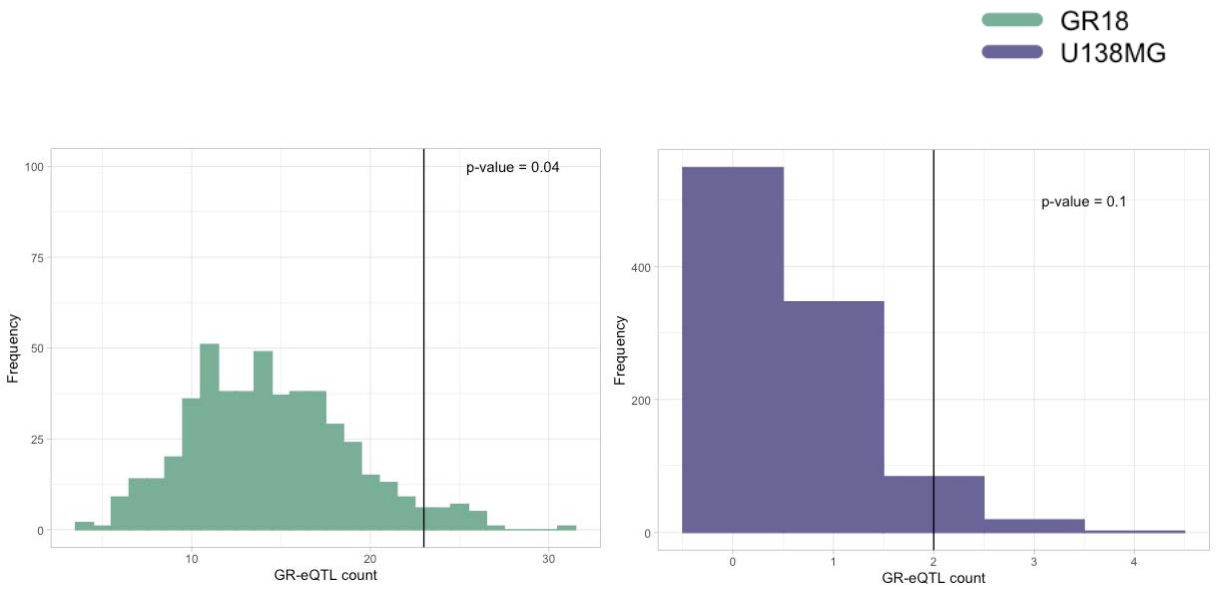


Figure 23. DREs are enriched in GR binding sites in GR18 but not U138MG cells. A) Distribution of overlapping non-functional GR-eQTLs and GR binding sites in GR18 (green) and U138MG (purple) cells with vertical lines denoting the number of observed overlaps between the DREs and GR binding sites.

4.7.3. Dex variant-DREs are enriched in GR binding sites in GR18 cells

Next we assessed whether the variant-DREs were enriched in GR binding sites, as it is well-established that genetic variants can exert a functional effect by disrupting TF binding sites, thus modulating TF binding affinity.[96],[97] This has been observed with GR, specifically with variants located within or directly adjacent to GREs, resulting in transcriptional changes.[98] To determine whether this may represent a mechanism by which the variant-DREs exert their transcriptional effects, they were colocalized with GR ChIP-seq binding data from GR18 and U138MG cells. Indeed, dex variant-DREs were enriched in GR binding sites, when compared to a randomly selected GR-eQTL background set of the same size (without replacement), with a 5.5 fold-enrichment (permutation p-value 0.008). Though the veh variant-DREs were also enriched (2.9 fold), this was not significant (p-value = 0.15). In the U138MG cells, none of the variant-DREs in

neither the dex nor veh condition overlapped with a binding site of GR. Given the few dex and veh variant-DREs (10 and 2, respectively), this was not surprising.

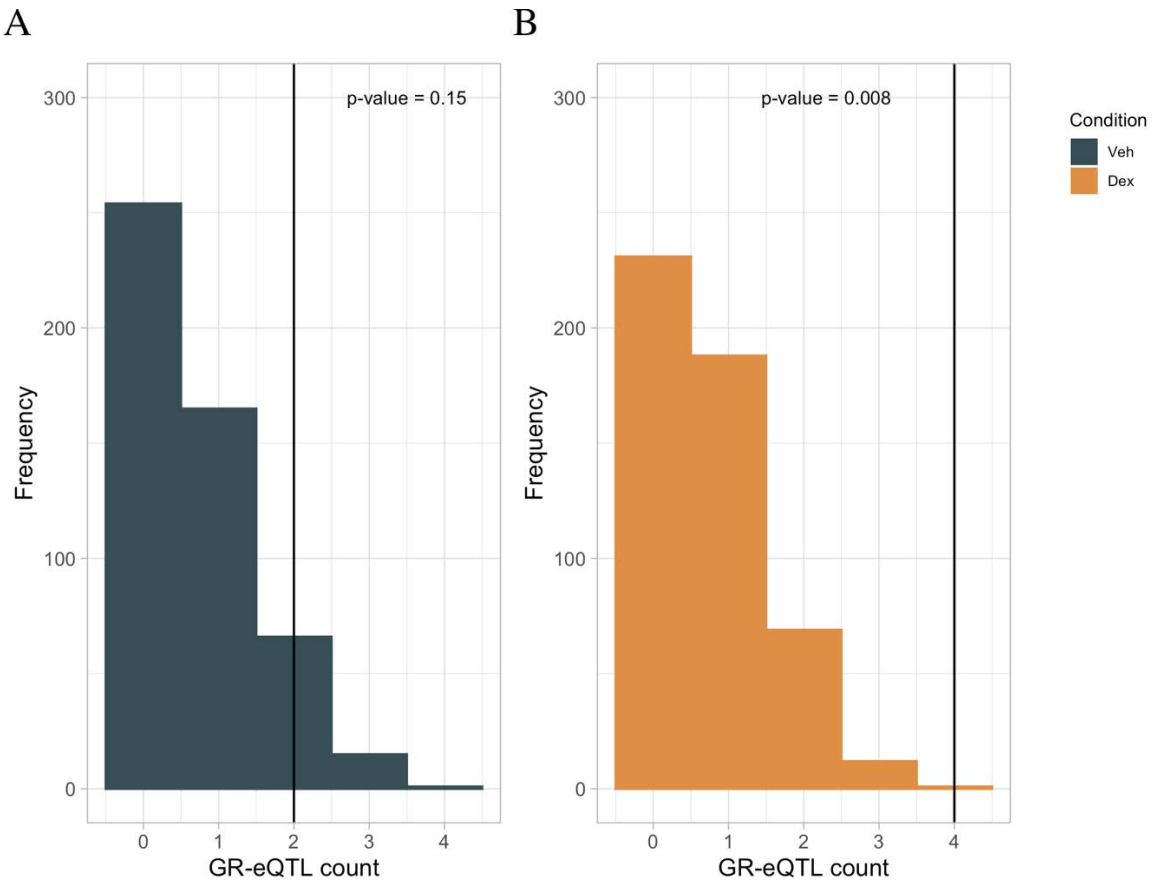


Figure 24. Dex variant-DREs are enriched in GR binding sites. A) Two veh variant-DREs were located in GR binding sites, which is not significant over the background B) Four dex variant-DREs were located GR binding sites, which is a significant enrichment over the background.

4.7.4. DREs are enriched in chromatin loops in GR18 cells

Previous data has shown that GR binding regions are involved in chromatin looping to promoter regions of genes both upregulated and downregulated by dex.[61] Interestingly, these the majority of these interactions already primarily exist at baseline (i.e. prior to dex), but the strength of the interactions are modified by dex, likely due to GR binding, which recruits co-activators, resulting in transcriptional activation. The picture is less clear for GR-mediated transcriptional repression, but recent evidence has also implicated chromatin looping involving in dex induced transcriptional repression.[61] Therefore, since chromatin interactions are involved in modulating transcriptional activity in response to dex, I examined whether the DREs were enriched in anchor points of chromatin loops.. To this end, we analyzed Hi-C data, a chromatin conformation assay that captures chromatin interactions across the genome, which was available in the parental GR18 cell lines, U2OS. The two available biological replicates were merged to create a union set. Using this union set, an enrichment analysis was performed using randomly chosen non-functional GR-eQTLs for the null distribution. Of the GR-REs, ~10% were located within anchor regions of chromatin loops in GR18 cells. This was significantly enriched over the non-functional GR-eQTLs, where ~5% of regions were located within anchor points (fold enrichment = 1.86, p-value = 0.002).

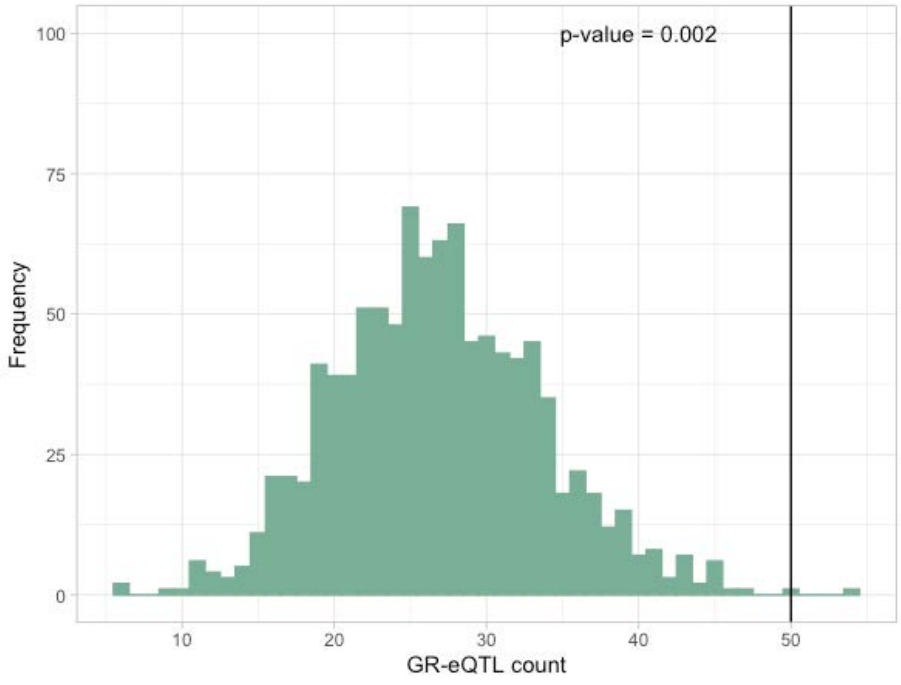


Figure 25. DREs are enriched in regions involved in chromatin looping.

Next we determined whether the variant-DREs were enriched in the chromatin anchor regions, as this would also represent a mechanism by which these functional variants could exert their effect, namely by interfering with chromatin interaction frequency at these sites and thus alter transcriptional activity. We found that both the veh and dex variant-DREs were highly enriched in these interaction regions, with a fold enrichment of 2.8 and 2.6 for veh and dex variant-DREs, respectively. This translates to ~15% of the variant-REs being located within these regions, considerably higher than the ~5% for the GR-eQTLs and even high than the 10% for the DREs.

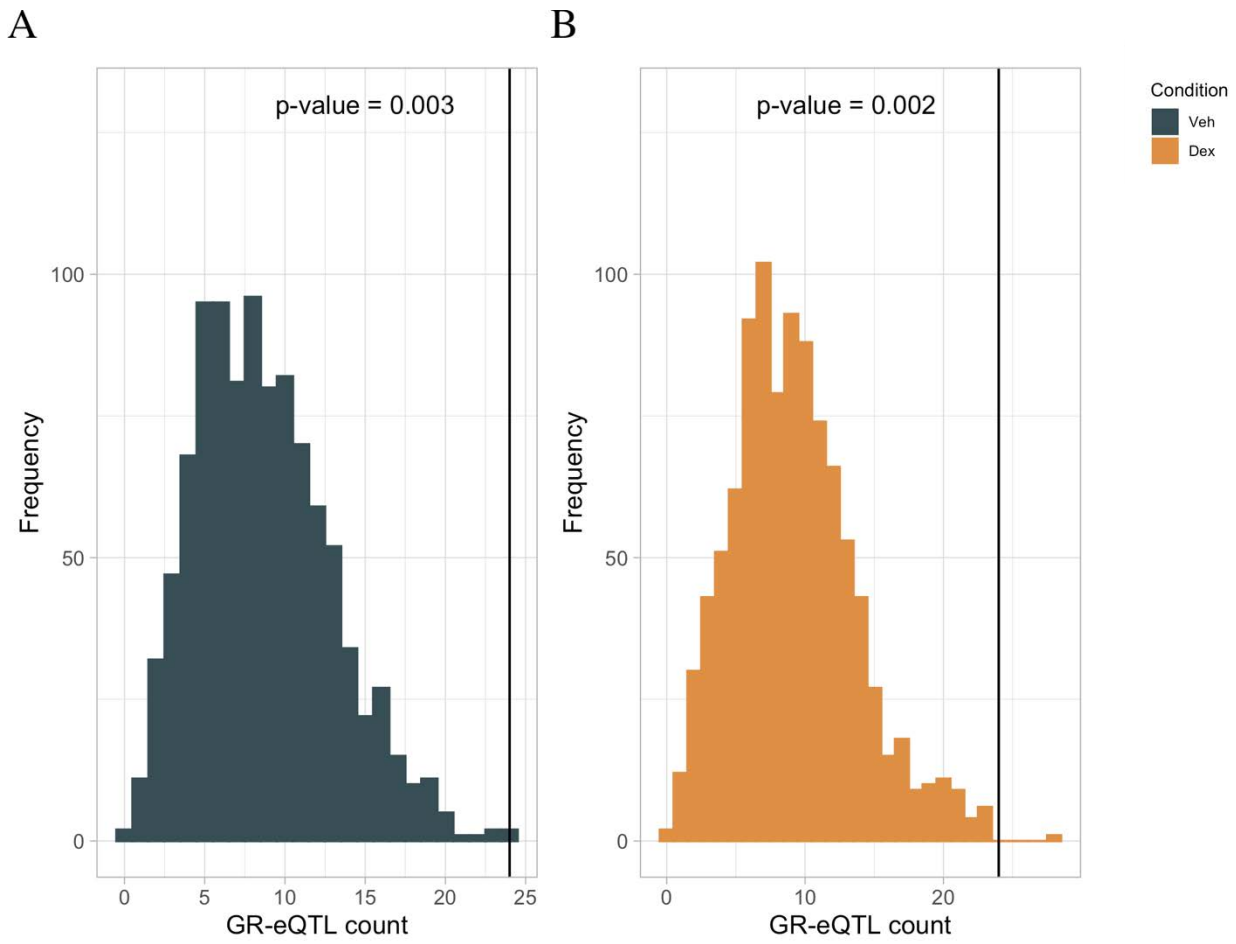


Figure 26. Variant-DREs are enriched in regions involved in chromatin looping. Both GR18 veh variant-DREs (A) and dex variant-REs are (B) are enriched in anchor points of chromatin loops in GR18 cells.

4.7.5. DREs are enriched in enhancer regions across tissues

Given that both GR activity[99] and REs[100],[101] are highly cell-type specific, I performed an in silico analysis to annotate the DREs and variant-DREs across different cell types to determine whether they were predicted to be functional in other tissues. Using HaploReg 4.0.2, an R package that integrates multiple levels of publicly available functional data to annotate the noncoding genome across multiple tissues, I assessed whether the dex and veh variant-DREs were enriched within predicted enhancer regions across tissues. Firstly, since the GR-eQTLS were discovered in blood samples and given my interest in psychiatric disorders, I annotated each variant-DRE across all available blood and brain tissues (Figure 27A). Visual scanning reveals that many of the variant-REs colocalized with active regulatory regions in brain and blood tissue. I then assessed broader tissue categories and quantified this enrichment, using sets of randomly chosen eQTLs from the GTEx dataset[20] to generate a null distribution (n permutations = 100). Interestingly, both the veh and dex variant-DREs were enriched in active enhancer regions across all 15 tissues tested (Figure 27B), indicating that the variant-DREs identified within the cell line models have regulatory potential across tissues. Then, I quantified the enrichment across all available specific brain regions to assess whether there were areas of the brain that were more highly enriched. From the eight regions, dex and veh variant-DREs were both significantly (permutation FDR <0.05) enriched in two regions (angular gyrus and middle hippocampus), and dex variant-DREs were exclusively significantly (permutation FDR <0.05) enriched in a further two, the cingulate gyrus, an area which has been implicated in psychopathology[102] and the germinal matrix, a region responsible for neural and glia cell production during development.[103]

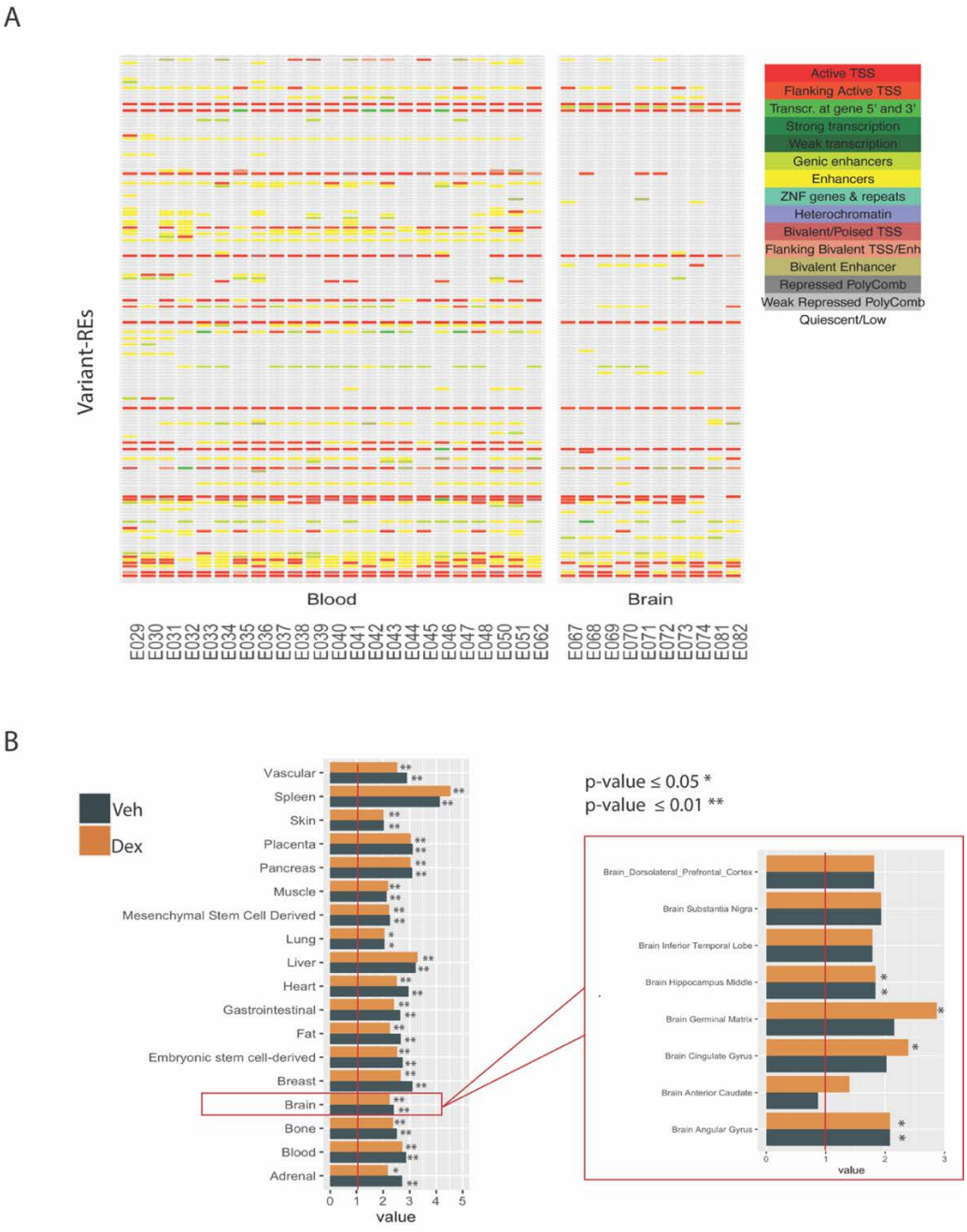


Figure 27. Variant-REs are located within active regulatory regions across tissues. A) Predicted chromatin state for all variant-REs (both dex and veh in GR18 and U138MG, each horizontal line on y-axis = 1 variant-DRE) across all available blood and brain tissues (x-axis = tissues EID, see Table S3 for descriptions of each EID). B) Enrichment of variant-DREs predicted to be located within enhancer regions across 15 broad tissue classes. C) Enrichment of veh and dex variant-DREs predicted to be located within enhancer regions across specific brain regions. * FDR < 0.05; ** FDR< 0.01; permutation FDR.

4.8. DRE regulated transcripts are enriched in neuronal terms

The DREs and variant-DREs were mapped to the 320 transcripts identified in the GR-eQTL analysis.[2] The 547 DREs from the union the DREs in the GR18 and U138MG cells were found to regulate 122 of the 320 transcripts. To better understand pathways that the DREs may regulate, a functional enrichment analysis on these target transcripts was performed using the Reactome Biological Processes Enrichment Analysis Tool.[104] As expected, terms related to stress were highly enriched, such as "cellular response to stress". Other more general cellular processes, such as rRNA processing and translation were also enriched. However, despite the GR18 and U138MG cells being non-neuronal, the most significantly enriched term was "regulation of expression of ROBO and SLITs", proteins integral to axon guidance during neural development.[105] An additional three brain related terms; signaling by ROBO receptors, axon guidance, and nervous system development were within the top 15 terms (Figure 28). The variant-DREs were also mapped to the transcripts, with 59 and 61 transcripts being regulated by the veh and dex variant-DREs, respectively. Table 11 summarizes the number of DREs and veh and dex-variant DREs regulating the transcripts. The transcripts not regulated by any DRE are not included.

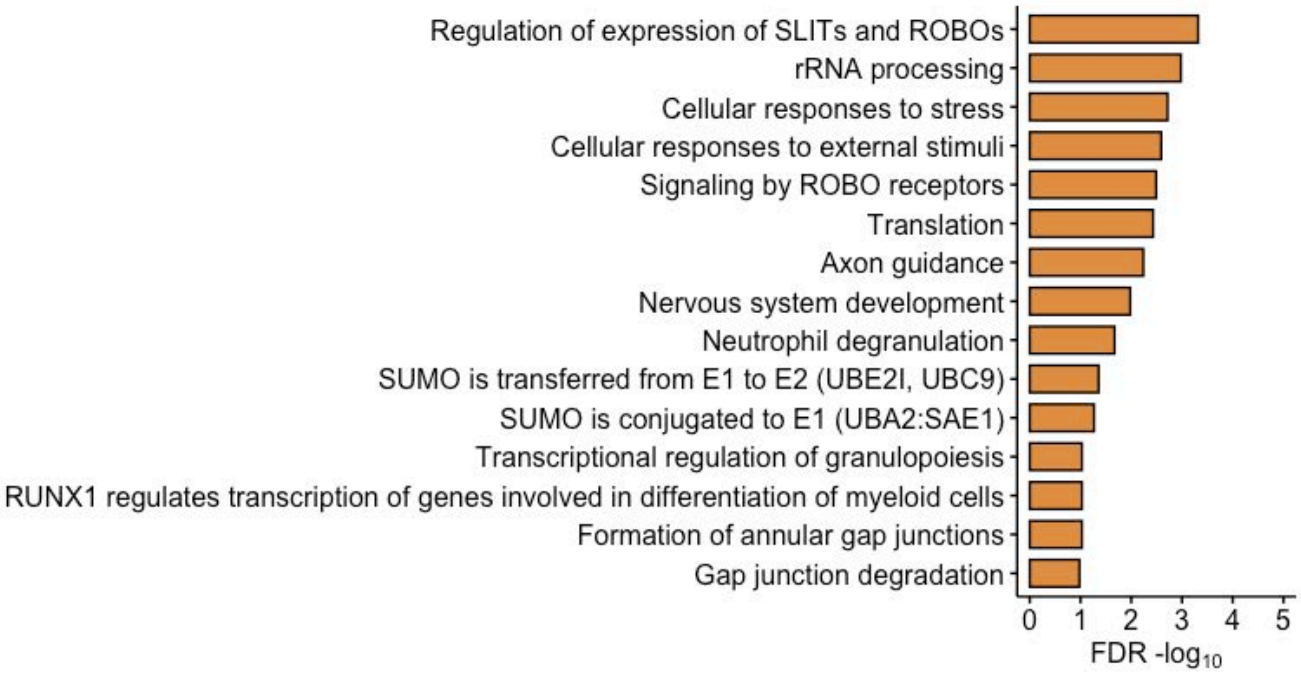


Figure 28. Top 15 reactome terms enriched in the DRE target transcripts.

Table 11. Number of DREs or variant-DREs regulating a predicted transcript. Variant-transcript pairs were derived from the GR-eQTL analysis from Arloth et al.[2]

Transcript	DRE	Dex-variant	Veh-variant	Transcript	DRE	Dex-variant	Veh-variant
ABHD5	2	1	0	TMEM176A	3	1	0
ACPL2	3	2	0	TPST1	46	18	16
ACTR2	1	1	1	ZAP70	18	8	8
ACYP1	2	1	2	ZNF641	12	2	3
AK024143	1	1	0	ANXA11	4	0	1
AK026751	10	2	0	C20orf3	3	0	1
ANXA2	1	1	0	CD96	1	0	1
ATP5F1	3	1	0	CKLF	2	0	1
C19orf35	2	1	1	CLDN14	1	0	1
C7orf44	23	10	5	CR617556	1	0	1
CD47	1	1	0	DPM2	3	0	2
CDC16	9	2	2	ERV3	13	0	2
CLEC4C	7	2	2	GNLY	2	0	1
COX5A	3	1	0	KDM3B	3	0	2
CR1	4	1	1	MTMR15	1	0	1
CST7	13	2	5	RBM4	1	0	1
CYBRD1	3	1	1	RPL7A	1	0	1
DDOST	2	2	1	SLC7A7	2	0	1
DNM2	4	2	2	ST6GAL0C4	3	0	2
F2RL1	2	1	1	SULF2	1	0	1
FOXC1	1	1	0	WDR1	2	0	1
FZD2	3	2	3	AI445566	1	0	0
GART	1	1	1	AI970822	1	0	0
GPX1	1	1	0	AL137655	1	0	0
HIGD1A	5	2	2	ALPL	1	0	0
HIST2H2AA3,HIS T2H2AA4	23	8	7	BTNL3	1	0	0
HLA-DRB4	25	5	6	CD74	2	0	0
HLA-DRB5	57	17	12	CLC	1	0	0
IL27RA	1	1	0	CORO1A	1	0	0
IMP3	2	1	0	COX17	1	0	0
KLF4	3	2	1	EXTL3	1	0	0
LAMP1	9	2	2	FAM117B	1	0	0
LCN2	1	1	0	FYN	1	0	0
LGALS2	2	1	0	GMIP	3	0	0
MAK	5	2	1	GRB10	1	0	0
MAP3K2	4	2	1	GTPBP8	1	0	0
MFN2	11	2	2	HACL1	1	0	0
MRPL24	11	6	4	HIST2H3D	1	0	0
NUAK2	1	1	0	KRT72	1	0	0
OAS2	5	2	0	LOC401497	1	0	0
PAAF1	1	1	1	LSM5	1	0	0
PAICS	5	3	3	MAGEL2	1	0	0
PAM	42	17	15	MAGT1	1	0	0
PHACTR1	6	2	2	MPZL1	1	0	0
PPM1A	2	1	1	MSRB2	3	0	0

PRKAA1	1	1	0	PARP8	1	0	0
PRKCB	4	2	1	POLR2I	2	0	0
RBM38	10	6	6	PPID	1	0	0
RBP7	2	2	1	PTEN	2	0	0
RFX5	13	3	4	RPL23AP64	1	0	0
RNASE6	2	1	1	RPL35A	2	0	0
RNF149	2	1	0	RPLP1	2	0	0
RPL14	17	3	7	SEC14L1	1	0	0
RPL26	4	2	0	SLC25A37	1	0	0
RPS2	1	1	0	STAMBPL1	1	0	0
RPS25	8	3	2	TIMM23	1	0	0
SIRPB1	5	2	3	TOB1	1	0	0
SLCO3A1	1	1	1	TTC31	1	0	0
SUMO2	1	1	0	UBE2G2	6	0	0
TBX1	2	1	0	UBR2	1	0	0
TCIRG1	1	1	1	UBTD1	1	0	0

4.9. Variant-DREs are enriched for psychiatric disease associated variants

So far, I have identified functional REs modulating the transcriptomic response to stress and have demonstrated that they are enriched in GR binding sites, chromatin loops, and enhancer regions across tissues compared to those non-functional GR-eQTLs. Given the association between dysregulation of the HPA axis and psychiatric disorders, we next sought to identify whether the variant-DREs were also variants associated with psychiatric disorders. To this end, we overlapped the variant-DREs with variants nominally associated with psychiatric disorders, from a cross-disorder GWAS meta-analysis of eight psychiatric traits (BPD, MDD, ADHD, AN, OCD, SCZ, TS, ASD).[3] As a background, the non-functional GR-eQTLs were permutated into 1000 SNP sets of the size as the number of variant-DREs, and those which overlapped with psychiatric disorder-associated variants were counted to generate a null distribution. Note the GR-eQTLs themselves were highly enriched for variants associated with MDD and SCZ[2], two of the disorders included in the meta-analysis. Of the 165 unique (union set of the U138MG and GR18) veh variant-DREs, 30 (19.3%) were associated with psychiatric traits, which was not a significant enrichment over the background GR-eQTLs. On the other hand, the dex variant-DREs were significantly enriched in variants associated with psychiatric disorders (permutation p-value = 0.008), with 39 (22.6%) overlapping variants, translating to a 1.78 fold enrichment over the GR-eQTLs, which themselves were significantly enriched for variants nominally associated with MDD and SCZ (fold enrichment values of 1.29 and 1.22 for SCZ and MDD, respectively).[73] Both the

veh and dex variant-DREs were significantly enriched in variants associated with psychiatric disorders compared to randomly selected eQTLs from the GTEx database (fold enrichments of 1.5 and 1.9 for veh and dex variant-DREs, respectively). The psychiatric disorder variants overlapping with the variant-DREs were then separated into two groups; those present both veh and dex variant-DREs, those exclusive to dex variant-DREs, and those exclusive to veh variant-REs (Table 12).

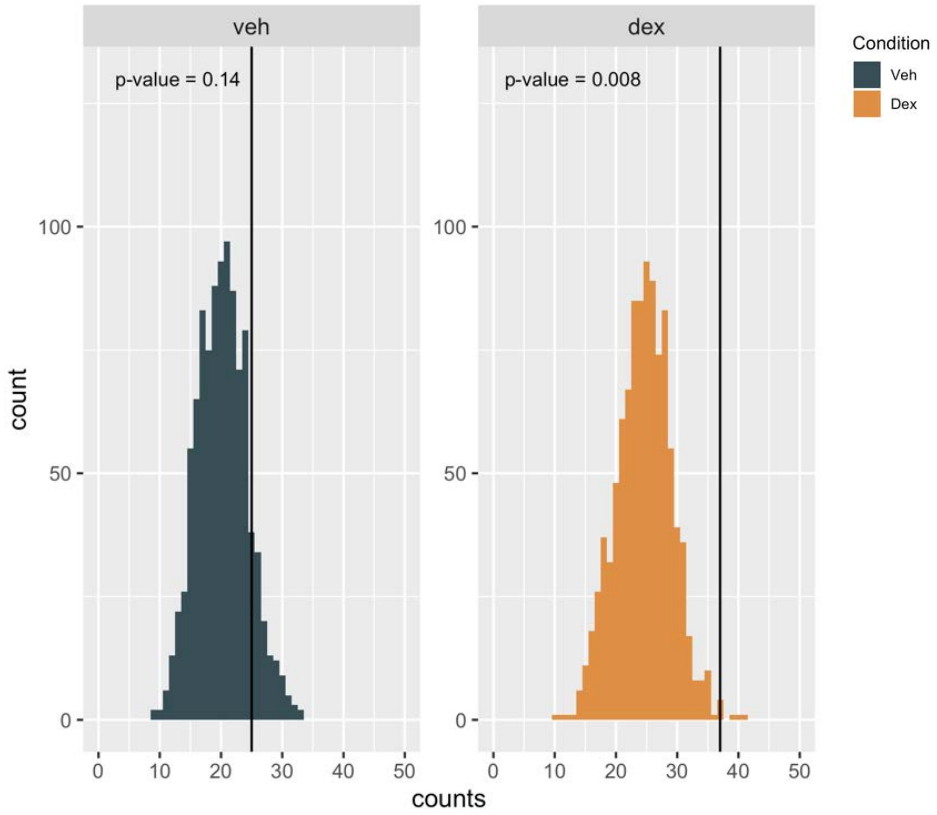


Figure 29. Dex variant-DREs are enriched for loci associated with psychiatric disorders. The consensus set of veh and dex variant-DREs from GR18 and U138MG cells were overlapped with variants associated with psychiatric traits. Veh variant-DREs were not enriched over the background non-functional GR-eQTLs. In contrast, the dex variant-DREs were significantly enriched.

Table 12. Variant-DREs overlapping with variants associated with psychiatric traits. Cross disorder meta-analysis beta and p-values were derived from the meta-GWAS summary statistics.[3] Genes refers to transcripts regulated by the variant-DRE, as defined by a GR-eQTL analysis.[2] Associated disorder refers to those disorders with a high confidence association to the variant within the meta-GWAS.

Variant-DRE	Beta-value	P-Value	Alt	Ref	Gene	Associated disorder
Exclusive Veh Variant-REs						
rs11788797	-0.02	1.66E-04	G	C	ST6GALNAC4	SCZ,BIP,MDD,AUT
rs11788797	-0.02	1.66E-04	G	C	DPM2	SCZ,BIP,MDD,AUT
rs2263658	0.07	1.94E-03	C	T	SIRPB1	ADD
rs258145	-0.09	4.50E-02	G	A	PAM	TS
rs34928543	0.06	1.04E-02	G	C	HLA-DRB4	SCZ
rs447531	0.1	6.70E-03	C	T	PAM	TS
rs45585631	0.03	5.64E-05	C	T	DPM2	SCZ,BIP,MDD,AUT
rs45585631	0.03	5.64E-05	C	T	ST6GALNAC4	SCZ,BIP,MDD,AUT
rs4850875	-0.05	1.51E-02	G	A	ZAP70	BIP,OCD
rs4851136	-0.05	1.41E-02	T	A	ZAP70	BIP,OCD
rs55900716	0.05	1.86E-02	G	A	ZAP70	BIP,OCD
Exclusive Dex Variant-REs						
rs11901568	-0.05	1.83E-02	C	T	ZAP70	BIP,OCD
rs2278212	-0.05	1.54E-02	C	T	ZAP70	BIP,OCD
rs35288741	-0.04	7.80E-03	G	A	NUAK2	SCZ,TS,OCD
rs35375092	0.04	5.36E-03	G	A	GPX1	BIP,ADD
rs3754888	-0.05	1.35E-02	C	T	ZAP70	BIP,OCD
rs3754893	0.05	2.34E-02	T	A	ZAP70	BIP,OCD
rs3826440	0.03	1.46E-03	C	T	RPL26	MDD,AUT,ADD,TS,OCD
rs382964	-0.11	4.51E-03	C	T	PAM	TS
rs4768230	0.1	4.11E-02	G	A	AK026751	ANO
rs56125600	0.08	1.28E-11	C	T	HIST2H2AA3,HIST2H2AA4	SCZ,BIP,OCD
rs6771365	-0.03	2.24E-04	T	A	ABHD5	SCZ,MDD,ADD,TS,OCD
rs6888677	0.1	7.73E-03	G	C	PAM	TS
rs7194275	-0.1	1.98E-02	C	T	RPS2	ANO,OCD
rs809972	0.02	4.31E-02	G	T	RBP7	MDD,AUT,ADD,TS
rs9270858	-0.04	1.67E-03	C	T	HLA-DRB5	SCZ,AUT
rs9271076	-0.04	2.99E-03	G	A	HLA-DRB5	SCZ,AUT
rs9271171	-0.06	6.93E-04	C	T	HLA-DRB5	SCZ
rs9271687	0.05	2.94E-02	G	A	HLA-DRB5	SCZ
rs9288853	0.06	7.21E-03	C	T	CD47	ADD,TS
rs9890920	0.03	9.43E-04	C	T	RPL26	MDD,AUT,ADD,TS,OCD
Both veh and dex REs						
rs11205362	-0.08	4.02E-08	G	A	HIST2H2AA3,HIST2H2AA4	SCZ,BIP,OCD
rs11588837	0.09	1.23E-09	G	A	HIST2H2AA3,HIST2H2AA4	SCZ,BIP,OCD
rs11589922	-0.09	1.16E-11	C	A	HIST2H2AA3,HIST2H2AA4	SCZ,BIP,OCD
rs12193156	-0.08	2.98E-02	G	A	PHACTR1	TS,OCD
rs168926	-0.09	4.04E-02	C	T	PAM	TS
rs2243603	0.06	3.88E-03	G	C	SIRPB1	ADD,TS
rs2422604	0.07	3.39E-03	G	A	SIRPB1	ADD
rs258146	0.09	4.05E-02	T	A	PAM	TS
rs369099	0.11	4.98E-03	G	A	PAM	TS
rs3754884	0.05	1.84E-02	C	T	ZAP70	BIP,OCD
rs3754885	0.05	1.81E-02	G	T	ZAP70	BIP,OCD
rs3769743	0.05	1.36E-02	C	T	ZAP70	BIP,OCD
rs419099	0.11	4.46E-03	G	A	PAM	TS
rs521199	0.02	2.10E-02	G	A	MAK	BIP,MDD,ADD,TS,ANO,OCD
rs9270678	-0.06	2.12E-04	T	A	HLA-DRB5	SCZ

4.10. Variant-DREs are causally associated with psychiatric disorders

Since the variant-DREs were enriched in psychiatric disorder associated variants, we employed MR, a statistical framework (see Section 1.3.2) that estimates the causal effect of an exposure variable on an outcome variable. In this case, MR allowed the causal effect of the differential DRE activity (driven by the variants-DREs) on psychiatric disorders to be estimated. In total, 21 variant-DREs were found to be significantly (FDR < 0.05) putatively causally associated with the psychiatric traits. Eleven veh variant-DREs were significantly associated, and had an OR between 0.78 and 1.29. From the dex variant-REs, 19 were significantly causally associated with psychiatric disorders with an OR ranging from 0.78 to 1.52. Ten of these 19 were exclusive dex variant-DREs. Additionally, MR was performed on the variant-DREs using the summary statistics from GWAS of extreme height[106] and fracture risk[107]. Only one and zero variant-DREs were found to be significantly causally associated with these traits, respectively, a marked decrease from the 21 variant-DREs implicated in psychiatric disorders. It should be noted that both these GWAS were somewhat smaller than that for the psychiatric cross disorders, with 37,857 cases and 227,116 controls (fracture) and 183,727 (height) compared to the 232,964 cases and 494,162 controls used in the psychiatric disorder meta-GWAS.

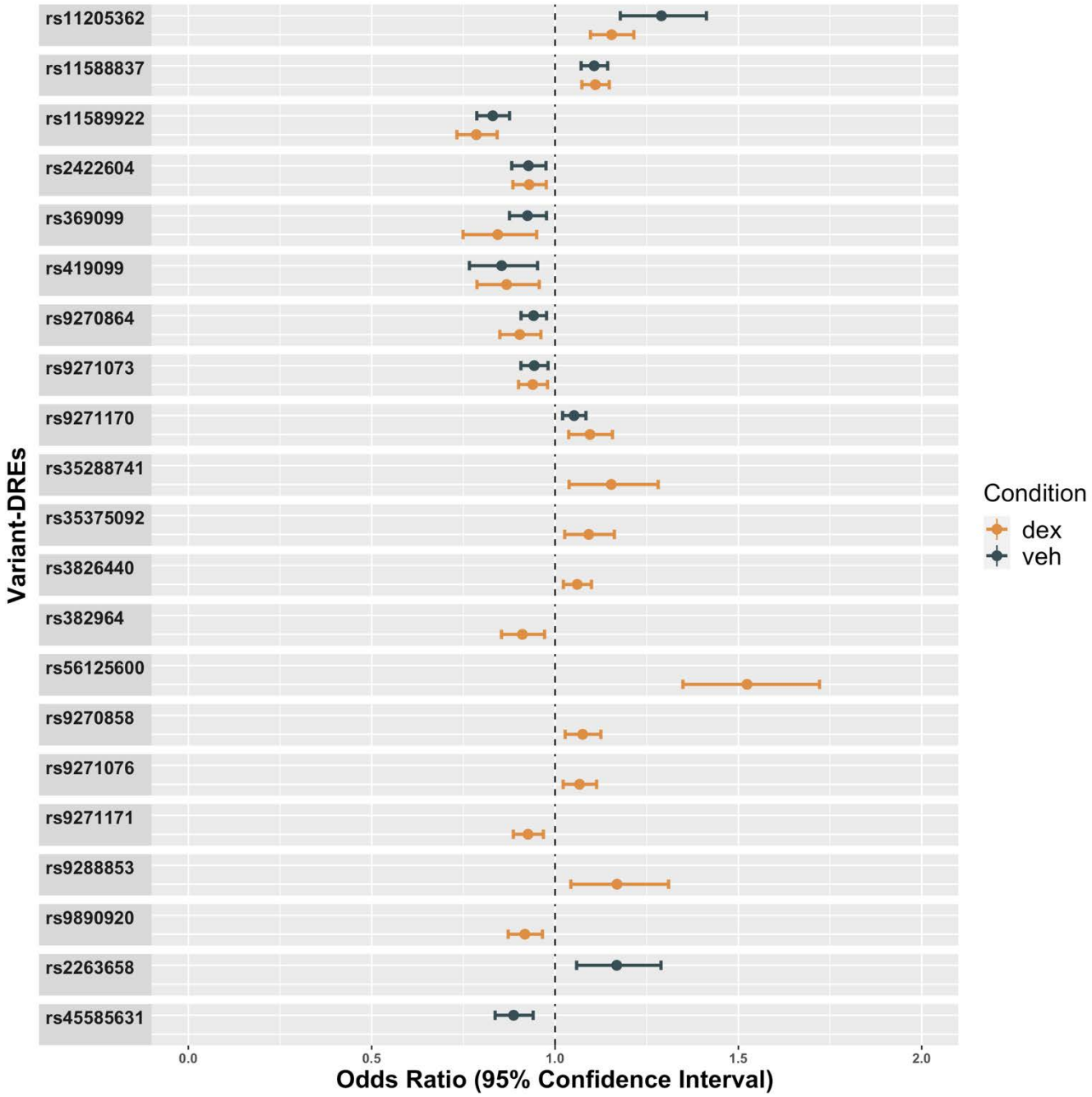


Figure 30. Mendelian Randomization results that estimate the causal effect of allele-driven differential RE activity on psychiatric disorders. Nine veh variant-DREs had a significant effect on psychiatric disorders whereas 19 of the dex variant-DREs had a significant effect.

5. Discussion

One of the strongest environmental risk factors for psychiatric disorders is exposure to stress and stressful life events.[108] Yet it is clear that not every individual exposed to these environments develops a psychiatric disorder, indicating that other factors are modulating the risk.[109] Evidence suggests that genetic variants interact with environmental risk factors in a complex interplay, termed G x E interactions. Yet, while evidence for an interaction between genetic predisposition and stress exists, the mechanisms governing this interaction remain elusive. Examining the HPA axis and its molecular mediators, specifically GR, can provide a deeper understanding of how genetic variants and stress converge in the regulation - and dysregulation of this system.

A study in 2015 by Arloth et al. used a stimulated GR-eQTL approach to identify over 3600 genetic variants that modulate the transcriptomic response to stress. However, this eQTL approach did not allow direct identification of individual functional genetic variants, but only associations between genetic variants in LD and transcripts responsive to treatment with dex. In this thesis, I followed-up on this study and tested all GR-eQTLs for their regulatory function using a high-throughput reporter assay in two cell lines. This identified 547 DREs mediating the GR-effects. Furthermore, by directly testing all SNP sequences, I determined that ~ 25% of these DREs display allele-dependent activity (variant-DREs). Experimental and in silico approaches revealed that these DREs were located within genomic regions enriched in TFBSes, including GREs, and chromatin loop anchor points. Lastly, I explored the relationship between the variant-DREs and psychiatric disorders and found evidence that the genomic loci modulating the transcriptomic response to stress have a putatively causal effect on psychiatric disorders.

5.1. Using STARR-seq to improve the resolution of functional fine-mapping and to identify putative mechanisms of regulation

In order to transition from identifying associations between variants and a phenotype (e.g. the stress response) to actually understanding the function of the variants, they need to be functional annotated and ultimately directly tested for activity. In this thesis, I started with a set of GR-eQTLs identified by a stimulated eQTL analysis. Using STARR-seq to directly test the functionality of each GR-eQTL SNPs in two cell lines, I determined that 165 and 172 variants located within the DREs display allele-dependent activity in the veh and dex conditions, respectively (variant-DREs).

Although the exact mechanisms by which the variant-DREs exert their effects remain elusive, I propose below putative mechanisms by which the variant-DREs may exert their effects, given the genomic and epigenomic context in which they reside.

5.1.1. Features associated with activating vs. repressive DREs

To date, a number of studies have used STARR-seq to investigate DREs and to better understand dex mediated transcriptional patterns.[110, 86, 83, 26] Interestingly, one of these studies reported a lack of DREs that were repressive in response to dex treatment,[110] and another study was unable to validate the STARR-seq identified repressive DREs.[86] However, in my STARR-seq experiment, I showed that nearly half (~45%) of the identified DREs are repressive. To ensure this was a robust finding, I included one repressive DRE in the validation experiment using qPCR, and was able to confirm the finding. Furthermore, my findings on repressive-DREs are in line with the observations Johnson et al. made in a recent study that investigated genome-wide DREs using STARR-seq. They found that 62% of the DREs were inductive, while 38% were repressive in response to dex.[26] Comparison of the inductive vs. repressed DREs in this study revealed that the inductive DREs displayed a 1.2 - 1.5-fold greater response to dex compared to the repressive DREs. Interestingly, I also observe a greater magnitude of responsiveness in the inductive DREs compared to the repressive DREs, with a 1.2 greater fold change in the inductive DREs in the GR18 cells, and 1.1 in the U138MG.

Given that both my STARR-seq data and those of Johnson et al. provide evidence for repressive DREs, it is likely a robust finding. Furthermore, this is congruent with multiple RNA-seq and expression microarray datasets that have demonstrated that dex treatment results in both upregulation and downregulation at target genes,[111,112,113] providing further evidence that repressive DREs exist. One possible reason that may explain either 1) the lack of repressive DREs identified by STARR-seq in some studies, or 2) the relatively small magnitude of fold-change identified in repressive DREs, is that the design of STARR-seq is better suited to detect inductive REs rather than repressive REs. Since STARR-seq uses a minimal promotor, there is low basal activity. Therefore, the repressive potential of STARR-seq REs may be limited, as baseline activity is already low and repressing it much further is impossible. This could result in a decreased ability to detect repressive REs but also limit the magnitude of repression possible to detect. Therefore, it cannot be determined whether the reduced magnitude of repression observed in my results are the

result of technical limitations of the STARR-seq method, or whether it is truly specific to GR-mediated repression, and thereby biological meaningful. Providing support for the latter are the results from the colocalization of the STARR-seq data with other functional datasets, as well as conclusions made by other studies. Together, these data lend support that GR-mediated repression and induction are governed by unique mechanisms, both of which are discussed below.

5.1.2. Putative mechanisms governing inductive DREs

Various models of GR-mediated transcriptional induction have proposed. Most involve GR binding at canonical GREs, although it has been reported that a large proportion (42%) of GR binding occurs at alternative sequences.[114] In this study, I demonstrated that the DREs are enriched in binding sites of GR, especially when considering only inductive DREs. However, despite this enrichment, only a small proportion of the DREs (~6%) overlapped with GR bindings sites in GR18 cells, suggesting other GR-mediated mechanisms must be involved. This may also, in part, be a result of the cell-type specificity of GR binding, which is largely dependent on underlying chromatin accessibility. Since STARR-seq is episomal and therefore independent of chromatin accessibility, it may have identified DREs located within chromatin contexts not conducive to GR binding in the GR18 cells, but conducive in other cell types. Supporting this hypothesis is that the majority (72%) of the DREs contained a GRE, indicating GR binding might be possible in another cellular context. Another possibility is that GR does not bind directly to the DNA, but rather interacts with a co-factor, which would not be detected using classical ChIP crosslinking approaches.[115] In fact, this mode of GR binding has even been reported to be more common than direct DNA binding.[116, 117]

Also thought to be involved in GR-mediated activation, are chromatin loops. In 2018 D'Ippolito et al. showed that the vast majority of GR target genes are involved in chromatin loops that are pre-established prior to dex treatment.[61] Dex, by activating GR, increases the strength of the interactions, either by binding GR directly to the DNA or tethering to another TF, thereby activating transcription. Using HiC data from the parental GR18 cells (U2OS) without dex treatment, I showed that the DREs are enriched in chromatin loop anchor points which is consistent with this model of GR-mediated activation. In fact, the number of DREs that are located within these anchors is nearly double the number of DREs located within GR binding sites, supporting

the hypothesis that both GR directly bound to DNA, and tethered GR are involved in chromatin looping.

5.1.3. Putative mechanisms governing repressive DREs

The mechanisms of GR-mediated repression remain more elusive, and numerous putative mechanisms have been postulated, including negative GREs,[118] and tethering to repressive TFs.[119] However, in recent years, the majority of these mechanisms have been disputed, and no consensus mechanism has been identified. The so-called negative GREs have been found nearly as frequently at activated compared to repressed genes,[115] and while tethering likely plays a role in GR-mediated repression, binding partners exclusive to repression have not been identified,[115,120] making both of these unlikely widespread mechanisms of GR-mediated repression.

In my results, a comparison of inductive vs. repressive DREs revealed that repressive DREs, unlike inductive DREs, were not enriched in GR binding sites. In fact, in the GR18 cells, over twice the number of inductive DREs were located within or proximal to a GR binding site compared to repressive DREs. This is consistent with what was reported by Johnston et al., who found a depletion of GR binding sites within repressive DREs.[26] Moreover, in my data, the repressive DREs, unlike the inductive DREs, were not enriched in the GR binding motif, an observation also made by Johnson et al. who performed this analysis in a different cell line. This suggests that the lack of GR binding at repressive-DREs is not specific to the GR18 cells, but rather represents a broader pattern of dex-mediated repression. However, similar to GR-mediated induction, GR could still exert its effects by tethering to other TFs, which has been reported. Further investigations to determine the tethering partners of GR need to be performed to better understand how the depletion of GR binding can still result in GR-mediated repression.

Despite the lack of GR binding at repressive DREs, I found that the repressive DREs were enriched in chromatin anchor points at a level comparable to that in inductive DREs. This is consistent with a study that reported that both genes induced and repressed by dex are involved in chromatin looping to DREs. In this study, they found that repressed genes loop to transcriptionally silent components, unlike induced genes, which loop to transcriptionally active components.[61] However, no concrete mechanism for this repressive looping was proposed in this study. One mechanism proposed by a recent study suggested that tethered GR at repressive DREs competes with inductive REs located within the chromatin compartment for CBP/Ep300. In this model, the

depletion of Ep300 impairs the inductive activity of the RE, thereby repressing its target gene.[121] If this is indeed a mechanism of GR-mediated repression, then chromatin interactions, which would bring the competing REs into proximity with one another, may increase the sequestering ability of the repressive DRE. However, this putative mechanism needs to be further studied.

5.1.4. Putative mechanisms of variant-DREs

Given these putative models for both of inductive and repressive DREs, the mechanisms by which the variant-DREs function can be postulated. It is well-established that variants within TFBSes can modulate the affinity of TF binding to DNA.[97,96] Specific to GR, previous studies have demonstrated that single base pair variations within GREs,[122] or regions directly flanking this sequence[86] can modulate GR's activity by inducing minor conformational changes in DNA structure, influencing the strength and persistence of GR binding. This mechanism would likely be more prominent in inductive DREs, as I, and others, show that repressive DREs are not enriched in GR binding sites nor GRE motifs. For the sites where GR is not directly bound to the DNA, the same principal could apply (i.e., by changing the binding affinity of a TF) but involve the disruption of DNA binding of another TF, and therefore, alter the ability of GR to tether. This is supported by the results of the motif search on variant-DREs which shows known tethering partners of GR (AP-1,[123],[117] CEBPB,[124] STAT5[125]) as being highly enriched in the variant-DREs. In both (directly-bound or tethered GR), changes in GR binding affinity could modulate the strength of chromatin looping, resulting in differential DRE activity, either by increasing or decreasing the frequency of interactions. In the case of repressive DREs, increased interactions would result in transcriptional repression, and in the case of inductive-DREs, in transcriptional induction.

For the veh variant-DREs, a number of the principles outlined above may still apply, given that the majority of GR-regulated genes are basally expressed, and dex only either up or downregulates their expression.[111] Therefore, TFBS disruption by TFs involved in basal expression remains a possible mechanism by which the veh variant-DREs could exert their effect. Likewise, since the large majority of chromatin loops involving GR target genes are formed prior to dex, the veh variant-DREs might perturb interaction frequencies, leading to differential gene expression in the veh condition.

5.2. The value of identifying small effect size variants

The mean dex-induced fold change of the DREs was 1.4 [-2.3, 17.0] and approximately 90% of the DREs had fold-changes of less than two. This is in line with what has been observed in a previous STARR-seq experiment investigating DREs, which found that 84% of the identified DREs had a less than two-fold change in activity upon dex treatment.[110] The effect size of the variant-DREs was also modest, (mean fold change 1.5 [1.12, 5.8]), which is consistent with the effects observed in other parallel enhancer screen (STARR-seq and MPRA).[126,127] Such modest effects are not surprising given that complex traits, such as the stress response, are typically modulated by many common variants located within noncoding genomic regions. Moreover, psychiatric and mental health related phenotypes have been found to be especially polygenic with effect sizes smaller than many other complex traits.[128] Yet, loci with small effect sizes do not necessarily make them trivial. For example, identifying disease-associated loci, even when they have small effects, can help identify novel molecules and biological pathways that can provide insight into disease mechanisms and identify new therapeutic targets, especially when multiple, small effect-size loci converge onto common biological pathways or molecules.

Furthermore, in common traits, individuals typically harbor multiple common variants, each contributing a small effect.[129] Therefore, the additive effects of small-effect size variants need to be considered. The additive effects of common variants can be captured by calculating PRSes which provide quantitative estimates of the genetic burden of risk loci, and therefore risk for disease. For example, individuals with a SCZ PRS in the top decile, have a threefold increased risk for developing psychosis.[130] By generating a PRS, the additive effect of the variant-DREs could be summarized to generate a "stress sensitivity score". In fact, in the GR-eQTL study, a PRS for stress sensitivity was calculated. This score was associated with dysregulation of amygdala reactivity, risk for MDD,[2] and the hemodynamic response to stress in a second neuroimaging study.[131] Given that STARR-seq filtered the GR-eQTLS to identify only those that were functional which yielded a subset more enriched in variants associated with psychiatric disorders, a PRS approach using only variant-DREs would likely provide even more predictive power and insight into the systems they regulate. Especially interesting would be to generate a PRS including only those variant-DREs recognized by MR as having a significant putatively causal effect on psychiatric disorders and weighing each variant-DRE by the OR.

5.3. Choosing appropriate models for high-throughput functional reporter assays for understand neuropsychiatric disorders

A recent review on functional genomics approaches used in neuropsychiatric disorders highlighted the need to functionally annotate GWAS variants using high-throughput functional assays, such as MPRA or STARR-seq. However, the review also highlighted the need to interrogate these variants in neuropsychiatric disorder relevant cellular models. Specifically, reprogramming of peripheral cells into induced-pluripotent stem cells (IPSCs) and then differentiating them into specific neuronal or astrocytic subtypes specifically relevant to the disease of interest was proposed as the optimal approach.[132] Given that RE activity and GR binding is cell-type specific,[58] this approach would be superior to using the cell lines used in my STARR-seq experiments in order to better understand the regulation of stress in a neuronal context. Indeed, I did assess the suitability of IPSC-derived neurons (iNeurons) for STARR-seq however, they proved not to be an appropriate model for stress research. Specifically, the iNeurons, which were generated by a collaborating research group, did not show a robust response to dex, with no significantly differentially expressed genes identified by RNA-seq after dex treatment. To validate these findings, qPCR was performed on GR target genes, and only FKBP5 showed a minimal response, with a 1.5-fold increase in FKBP5 expression after dex, which was far less than the cell lines ultimately chosen, which showed a ~4 fold increase (see Figure S1). Although this lack of response is surprising, this is consistent with another study that identified no differential expressed genes after dex treatment in primary mouse neurons compared to astrocytes[133], and a study of the expression of FKBP5 in human iPSC-derived iNeurons which found a minimal (1.23 fold) induction upon dex stimulation.[134] Furthermore, immunohistochemistry and WB analyses showed no evidence of GR translocating into the nucleus upon dex treatment in our iNeurons. Together, these data indicate that the iNeurons generated were not an appropriate model for studying the stress response. Yet, only one iNeuron subtype (glutamatergic forebrain iNeurons) was assessed, and other specific subtypes of iNeurons may elicit a sufficient response to GR, given GR's critical role in certain brain regions, such as the hippocampus,[135] and amygdala,[136] although even within specific brain regions, there is evidence for a neuronal subtype specific response to dex.[137] Furthermore, single-cell RNA-seq data from brain organoids exposed to dex demonstrate a cell-type specific response to dex.[92] Therefore, before employing STARR-seq in iNeurons as suggested

by Townsley et al., an appropriate iNeuron model showing a sufficient transcriptomic response to dex needs to be identified.

Despite the advantages enabled by performing functional assays in iNeurons, I would argue that especially when performing episomal, rather than genome integrated, high-throughput methods, using models that are not disease-relevant can still provide valuable information. This is especially true for my STARR-seq experiments which focus on GR. Given that GR binding events are highly dependent on chromatin accessibility, performing episomal STARR-seq in disease-relevant models may not have provided much additional information, given that chromatin conformation has a defining role in determining DRE activity. Therefore, I would argue it is imperative that either an genome integrated version of STARR-seq or MPRA is performed in the appropriate neuronal model, or that episomal STARR-seq data is integrated with functional data from disease-relevant tissues to predict whether the STARR-seq identified features were located within regions conducive to activity. Such integration approaches are highlighted by Townsley who argues that loci identified by high-throughput reporter assays need to be contextualized by integrating these loci with data on the chromatin and epigenomic landscapes surrounding the loci. In fact, a study is highlighted that contextualizes SCZ MPRA loci identified in non-neuronal cell lines, by performing RNA-seq and ATAC-seq on three different iNeurons subtypes derived from 20 individuals heterozygous at the risk loci of interest. This allowed the loci to be understood in an endogenous context that was relevant to the diseases of interest.[138] However, contextualizing variants identified by high-throughput reporter assays does not necessarily need to involve iNeurons.

Using HaploReg, I annotated the DREs and variant-DREs to the 15 ChromHMM chromatin states and found that they were highly enriched in regions predicted to be active enhancers across all 15 tissues I assessed. I further analyzed all individual brain regions available from HaploReg. From the eight regions, dex and veh variant-DREs were both significantly enriched in two regions (angular gyrus and middle hippocampus), and dex variant-REs were exclusively enriched in a further two, including the cingulate gyrus and the germinal matrix. Both the hippocampus and cingulate gyrus, which are components of the limbic system, have been implicated in various psychiatric disorders, such as decreases in hippocampal volume in MDD,[139] and hypoactivity of the cingulate gyrus in SCZ.[140] Further integration of the variant-DREs could be performed, for

example using functional data from post-mortem brains, brain organoids, or dex-responsive iNeurons.

One interesting finding from the motif search on the variant-DREs was that the second most highly enriched motif in the dex variant-DREs was OTX. OTX is family of TFs that are essential in brain development.[141] Moreover, OTX2 has been shown to play a critical role in the transcriptional programming of stress. In one study, a knockdown of Otx2 in mice was shown to mimic early life stress by increasing stress vulnerability, whereas the opposite effect was seen when it was overexpressed. This led the authors to conclude that OTX2 is the mediator of long-term transcriptional programming in the ventral tegmental area that encodes stress susceptibility.[142] The fact that this motif was highly enriched in the dex variant-DREs not only provides more insight into how the dex variant-DREs may function, but also lends support that the STARR-seq findings from cell lines can identify brain-relevant biological processes, in this case, long-term encoding of stress sensitivity.

5.4. Using a networks approach to better understand variant-DRE effects

The 3662 GR-eQTLs that were identified by Arloth et al.[2] regulated the expression of 320 transcripts. STARR-seq identified DREs that harbored GR-eQTLs associated with the expression of 122 of these transcripts, meaning that nearly 40% of the transcripts were represented. Interestingly, a pathway enrichment analysis on these 122 transcripts revealed nervous system development and axon guidance as two of the significantly enriched terms. Considering the initial GR-eQTL analysis was performed in blood and the STARR-seq in two cell lines, having enriched neuronal development terms strengthens my suggestion that the STARR-seq approach, even when performed in non-neuronal models, can provide valuable insight into diseases that affect the brain. Nervous system development was especially interesting, as there is evidence that exposure to stress, especially during development, can cause long-lasting effects on neuronal function and lifelong susceptibility to stress.[142] A study that assessed 65 GR-eQTLs-regulated transcripts associated with MDD and SCZ from the Arloth et al. study, found them to form an interconnected GR responsive co-expression network in mice brains.[143] Exposure to social stress or a positive environment was found to significantly alter the network structure in a brain-region and developmental stage specific manner, suggesting distinct early-life environments may promote differential network formation, especially in the hypothalamus and hippocampus, two regions

crucial to the stress response. Given this tight co-expression network, differential expression of single transcripts could cause perturbations that extend beyond the differential expression of a single gene, but rather shift the entire stress response network. It is then plausible that the variant-DREs that modulate the expression of these transcripts could cause shifts in the brain's response to stress. Indeed, the veh and dex variant-DREs harbored GR-eQTLs associated with 26% and 33% of these 65 transcripts, respectively. Since only the GR-eQTL associated MDD and SCZ genes were included in this network analysis, further investigation including all variant-DRE regulated transcripts associated with psychiatric disorders (from the meta-analysis GWAS) is warranted to determine whether additional transcripts are involved in this stress co-expression network.

While a better understanding of the variant-DRE regulated transcripts involved in the stress network is crucial, to confirm the role of the variant-DREs in a stress responsive network, and moreover, to link them to stress-related behaviors and phenotypes, direct approaches are required. For example, SNP-editing or RE knockouts using Crispr/Cas in iNeurons or brain organoids could be used to determine whether variant-REs could cause neuronal phenotypes known to be associated with stress, such as reduced spine density or altered spine morphology.[144],[145] Recent methodological advances makes these strategies possible in humanized animal models,[146],[147] which would shed light onto the effects of the variant-DREs on stress-related behavior, psychiatric-relevant phenotypes, and neurocircuitry.

5.5. Using massively parallel reporter assays in psychiatry

GWAS in psychiatry have identified genetics variants associated with multiple psychiatric disorders. Due to efforts such as the Psychiatric Genetics Consortium, higher powered studies are being conducted, enabling variants with smaller effect sizes to be identified.[11] Yet, in order for the potential of these findings to be fully harnessed, follow-up studies investigating the functionality of these variants need to be performed in order to both further prioritize variants for follow-up experiments, and to better understand the biological pathways and mechanisms governing these disorders. Although psychiatry-specific resources exist that allow for functional annotation of these variants, such as PyschEncode,[148] very few studies have employed parallel reporter assays to directly test the effect of genetic variants on regulatory activity. A study from Myint et al. which tested 1049 loci associated with SCZ and 30 loci associated with Alzheimer's disease, found that

148 of the variants displayed allele-dependent effects in two cell lines.[27] A second study which performed a GWAS on ASD followed by MPRA, identified a functional locus on DDHD2, which colocalized with a prenatal and adult brain eQTL.[149] While this thesis does not use STARR-seq to directly test GWAS identified genetic variants, it does test variants that are associated with the stress response, which is an important endophenotype in psychiatry. This approach, to map variants to endophenotypes, rather than directly to disorders, is a bottom-up approach that has advantages over directly searching for associations between variants and diseases, especially when studying psychiatric disorders, which rely on diagnostic tools mainly based on symptoms, and not on biological or genetic markers. Therefore, mapping variants to endophenotypes is an important first step in gaining a better understanding of the underlying systems involved in psychopathology. Mechanistic insights generated from this approach could then be harnessed and applied to understand the psychiatric disorders themselves. In this thesis, I focused on understand the genetic variants modulating the transcriptomic stress response, an important endophenotype in psychiatry. In fact, dysregulation of the stress response on the transcriptomic level has been implicated in various psychiatric disorders, meaning it may act as a cross-diagnostic endophenotype. For example, in MDD, a study by Menke et al. found differences in GR-stimulated gene expression in patients with MDD. Using these GR-stimulated gene expression profiles as a classifier for case vs. control status resulted in 77% of individuals being correctly classified, as opposed to only 41.6% correctly classified using baseline gene expression.[150] Another study on GR-stimulated gene expression identified unique expression profiles in PTSD patients vs. controls after stimulation of peripheral blood mononuclear cells with dex.[151] Consistent with these results, the functional variants identified by STARR-seq that modulate the transcriptomic stress response were highly enriched for variants identified in a GWAS meta-analysis of eight psychiatric traits. Together, this suggests that a better understanding of the stress response, an endophenotype, and the mechanisms by which it is governed, may provide valuable insight not only relevant to single psychiatric disorders, but more broadly applicable to general psychopathology.

5.6. Moving from association to causality and the implications for psychiatry

The fact that the variant-DREs were enriched in variants associated with psychiatric disorders provided indication that the genomic loci modulating the transcriptomic response to stress were also those involved in psychopathology. However, the enrichment analysis was limited as it could only

reveal an association between the variant-DREs and psychiatric disorders. I wanted to move in the direction of causality to probe whether there might be a causal relationship between the differential enhancer activity driven by the variant-DREs and psychiatric disorders. Therefore, I employed MR. This revealed that variant-DREs are not only associated with psychiatric disorders, but also putatively play a causal role in psychopathology. A large subset of both the veh (7%) and dex (11%) variant-DREs were found to be putatively causally associated to psychiatric traits. Interestingly, a larger proportion of the dex variant-DREs were found to have a significant effect on the disorders. Comparing these results to those from an MR study that estimated the causal effect of differential gene expression on 43 complex traits,[75] the proportion of significant exposure variables in my analysis is notably higher than in all 43 traits assessed. This may be partially be due to the fact that in my MR analysis only functional variants were included, whereas in the above-mentioned study all eQTLs available from a large database were included, many of which were likely non- functional. To my knowledge, this is the first study that has implemented a high throughput functional assays to select instrumental variables for MR. This may be a useful approach to better detect exposures with small effects, which is especially relevant when studying psychiatric disorders.

5.7. Conclusions and future directions

Psychiatric disorders are complex disorders with a very high burden of disease, affecting nearly 30% of the population within a lifetime.[152] To date, diagnostic tools rely on clinical assessment of symptom clusters, and many individuals respond insufficiently to available treatments. In order to develop better diagnostic tools, a better understanding of the genetics and molecular hallmarks of these disorders is required. In order to develop new treatments, the implicated pathways and networks need to be identified. While this work has fine-mapped functional variants modulating the stress response, some of which are putatively causally implicated in psychopathology, follow-up studies are required to gain a better mechanistic understanding of these variants. However, to my knowledge, this is the first study that has implicated the same genomic loci involved in modulating the stress response in being putatively causally associated with psychiatric disorders. This knowledge could already be used to support interventions and preventative strategies targeted to vulnerable populations, such as those with a high burden of stressful life events.

6. References

1. Harkness, K. L., Hayden, E. P. & Lopez-Duran, N. L. Introduction: Stress sensitivity and stress sensitization in psychopathology: An introduction to the special section. J. Abnorm. Psychol. (2015) doi:10.1037/abn0000041.
2. Arloth, J. et al. Genetic Differences in the Immediate Transcriptome Response to Stress Predict Risk-Related Brain Function and Psychiatric Disorders. Neuron (2015) doi:10.1016/j.neuron.2015.05.034.
3. Lee, P. H. et al. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell (2019) doi:10.1016/j.cell.2019.11.020.
4. Vigo, D. V., Kestel, D., Pendakur, K., Thornicroft, G. & Atun, R. Disease burden and government spending on mental, neurological, and substance use disorders, and self-harm: cross-sectional, ecological study of health system response in the Americas. Lancet Public Heal. (2019) doi:10.1016/S2468-2667(18)30203-2.
5. Pettersson, E. et al. Genetic influences on eight psychiatric disorders based on family data of 4 408 646 full and half-siblings, and genetic data of 333 748 cases and controls. Psychol. Med. (2019) doi:10.1017/S0033291718002039.
6. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: Review and meta-analysis. Am. J. Psychiatry 157, 1552–1562 (2000).
7. Jacquemont, M. L. et al. Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. J. Med. Genet. (2006) doi:10.1136/jmg.2006.043166.
8. Sebat, J. et al. Strong association of de novo copy number mutations with autism. Science (80-.). (2007) doi:10.1126/science.1138659.
9. Levinson, D. F. et al. Copy number variants in schizophrenia: Confirmation of five previous finding sand new evidence for 3q29 microdeletions and VIPR2 duplications. Am. J. Psychiatry (2011) doi:10.1176/appi.ajp.2010.10060876.
10. Geschwind, D. H. & Flint, J. Genetics and genomics of psychiatric disease. Science (2015) doi:10.1126/science.aaa8954.
11. Sullivan, P. F. et al. Psychiatric genomics: An update and an Agenda. Am. J. Psychiatry (2018) doi:10.1176/appi.ajp.2017.17030283.

12. Barešić, A., Nash, A. J., Dahoun, T., Howes, O. & Lenhard, B. Understanding the genetics of neuropsychiatric disorders: the potential role of genomic regulatory blocks. Molecular Psychiatry (2020) doi:10.1038/s41380-019-0518-x.
13. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. Genome Res. (2012) doi:10.1101/gr.136127.111.
14. Westra, H. J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. (2013) doi:10.1038/ng.2756.
15. Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. & Bejerano, G. Enhancers: Five essential questions. Nature Reviews Genetics (2013) doi:10.1038/nrg3458.
16. Visel, A. et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature (2009) doi:10.1038/nature07730.
17. Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. (2007) doi:10.1038/ng1966.
18. Xiao, X., Chang, H. & Li, M. Molecular mechanisms underlying noncoding risk variations in psychiatric genetic studies. Molecular Psychiatry (2017) doi:10.1038/mp.2016.241.
19. Sun, W. & Hu, Y. eQTL Mapping Using RNA-seq Data. Stat. Biosci. (2013) doi:10.1007/s12561-012-9068-3.
20. Lonsdale, J. et al. The Genotype-Tissue Expression (GTEx) project. Nature Genetics (2013) doi:10.1038/ng.2653.
21. Ramasamy, A. et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat. Neurosci. (2014) doi:10.1038/nn.3801.
22. Colantuoni, C. et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature (2011) doi:10.1038/nature10524.
23. Brasier, A. R., Tate, J. E. & Habener, J. F. Optimized use of the firefly luciferase assay as a reporter gene in mammalian cell lines. Biotechniques (1989).
24. Spicuglia, S., Santiago-Algarra, D., Dao, L. T. M., Pradel, L. & España, A. Recent advances in high-throughput approaches to dissect enhancer function. F1000Research (2017) doi:10.12688/f1000research.11581.1.
25. Kamps-Hughes, N., Preston, J. L., Randel, M. A. & Johnson, E. A. Genome-wide

identification of hypoxia-induced enhancer regions. PeerJ (2015) doi:10.7717/peerj.1527.

26. Johnson, G. D. et al. Human genome-wide measurement of drug-responsive regulatory activity. Nat. Commun. (2018) doi:10.1038/s41467-018-07607-x.
27. Myint, L. et al. A screen of 1,049 schizophrenia and 30 Alzheimer's-associated variants for regulatory potential. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. (2020) doi:10.1002/ajmg.b.32761.
28. Abascal, F. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature (2020) doi:10.1038/s41586-020-2493-4.
29. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. Nat. Protoc. (2017) doi:10.1038/nprot.2017.124.
30. Davis, J. et al. A review of vulnerability and risks for schizophrenia: Beyond the two hit hypothesis. Neuroscience and Biobehavioral Reviews (2016) doi:10.1016/j.neubiorev.2016.03.017.
31. Rai, D. et al. Parental socioeconomic status and risk of offspring autism spectrum disorders in a swedish population-based study. J. Am. Acad. Child Adolesc. Psychiatry (2012) doi:10.1016/j.jaac.2012.02.012.
32. Kendler, K. S., Karkowski, L. M. & Prescott, C. A. Causal relationship between stressful life events and the onset of major depression. Am. J. Psychiatry 156, 837–841 (1999).
33. Mandelli, L., Petrelli, C. & Serretti, A. The role of specific early trauma in adult depression: A meta-analysis of published literature. Childhood trauma and adult depression. European Psychiatry (2015) doi:10.1016/j.eurpsy.2015.04.007.
34. De La Morena-Barrio, M. E. et al. Genetic predisposition to fetal alcohol syndrome: Association with congenital disorders of N-glycosylation. Pediatr. Res. (2018) doi:10.1038/pr.2017.201.
35. Cornelis, M. C. & Hu, F. B. Gene-environment interactions in the development of type 2 diabetes: Recent progress and continuing challenges. Annual Review of Nutrition (2012) doi:10.1146/annurev-nutr-071811-150648.
36. Caspi, A. et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science 301, 386–389 (2003).
37. Van Der Meer, D. et al. The serotonin transporter gene polymorphism 5-HTTLPR moderates the effects of stress on attention-deficit/hyperactivity disorder. J. Child Psychol.

Psychiatry Allied Discip. (2014) doi:10.1111/jcpp.12240.

38. Stoltenberg, S. F., Anderson, C., Nag, P. & Anagnopoulos, C. Association between the serotonin transporter triallelic genotype and eating problems is moderated by the experience of childhood trauma in women. Int. J. Eat. Disord. (2012) doi:10.1002/eat.20976.
39. Munafò, M. R. et al. 5-HTTLPR genotype and anxiety-related personality traits: A meta-analysis and new data. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. (2009) doi:10.1002/ajmg.b.30808.
40. Appel, K. et al. Moderation of adult depression by a polymorphism in the fkbp5 gene and childhood physical abuse in the general population. Neuropsychopharmacology (2011) doi:10.1038/npp.2011.81.
41. Binder, E. B. et al. Association of FKBP5 polymorphisms and childhood abuse with risk of posttraumatic stress disorder symptoms in adults. JAMA - J. Am. Med. Assoc. (2008) doi:10.1001/jama.299.11.1291.
42. Klengel, T. et al. Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. Nat Neurosci 16, 33–41 (2013).
43. Wiechmann, T. et al. Identification of dynamic glucocorticoid-induced methylation changes at the FKBP5 locus. Clin. Epigenetics (2019) doi:10.1186/s13148-019-0682-5.
44. Risch, N. et al. Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: A meta-analysis. JAMA - Journal of the American Medical Association (2009) doi:10.1001/jama.2009.878.
45. Mullins, N. et al. Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. Psychol. Med. (2016) doi:10.1017/S0033291715002172.
46. Peyrot, W. J. et al. Effect of polygenic risk scores on depression in childhood trauma. Br. J. Psychiatry (2014) doi:10.1192/bjp.bp.113.143081.
47. Musliner, K. L. et al. Polygenic risk, stressful life events and depressive symptoms in older adults: A polygenic score analysis. Psychol. Med. (2015) doi:10.1017/S0033291714002839.
48. Gould, T. D. & Gottesman, I. I. Psychiatric endophenotypes and the development of valid animal models. Genes, Brain and Behavior (2006) doi:10.1111/j.1601-183X.2005.00186.x.

49. Iacono, W. G., Malone, S. M. & Vrieze, S. I. Endophenotype best practices. Int. J. Psychophysiol. (2017) doi:10.1016/j.ijpsycho.2016.07.516.
50. Ulrich-Lai, Y. M. & Herman, J. P. Neural regulation of endocrine and autonomic stress responses. Nature Reviews Neuroscience (2009) doi:10.1038/nrn2647.
51. Joëls, M. & Baram, T. Z. The neuro-symphony of stress. Nature Reviews Neuroscience (2009) doi:10.1038/nrn2632.
52. Tsigos, C. & Chrousos, G. P. Hypothalamic-pituitary-adrenal axis, neuroendocrine factors and stress. in Journal of Psychosomatic Research vol. 53 865–871 (2002).
53. Kadmiel, M. & Cidlowski, J. A. Glucocorticoid receptor signaling in health and disease. Trends in Pharmacological Sciences (2013) doi:10.1016/j.tips.2013.07.003.
54. Reul, J. M. H. M. & De Kloet, E. R. Two receptor systems for corticosterone in rat brain: Microdistribution and differential occupation. Endocrinology (1985) doi:10.1210/endo-117-6-2505.
55. Spencer, R. L., Miller, A. H., Moday, H., Stein, M. & Mc Ewen, B. S. Diurnal differences in basal and acute stress levels of type I and type II adrenal steroid receptor activation in neural and immune tissues. Endocrinology (1993) doi:10.1210/endo.133.5.8404640.
56. Liu, D. et al. A practical guide to the monitoring and management of the complications of systemic corticosteroid therapy. Allergy, Asthma and Clinical Immunology (2013) doi:10.1186/1710-1492-9-30.
57. John, S. et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nat. Genet. 43, 264–8 (2011).
58. John, S. et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nature Genetics (2011) doi:10.1038/ng.759.
59. Oh, K. S. et al. Anti-Inflammatory Chromatinscape Suggests Alternative Mechanisms of Glucocorticoid Receptor Action. Immunity (2017) doi:10.1016/j.immuni.2017.07.012.
60. McDowell, I. C. et al. Glucocorticoid receptor recruits to enhancers and drives activation by motif-directed binding. Genome Res. (2018) doi:10.1101/gr.233346.117.
61. D'Ippolito, A. M. et al. Pre-established Chromatin Interactions Mediate the Genomic Response to Glucocorticoids. Cell Syst. (2018) doi:10.1016/j.cels.2018.06.007.
62. Holsboer, F. The corticosteroid receptor hypothesis of depression. Neuropsychopharmacology 23, 477–501 (2000).

63. Bhagwagar, Z., Hafizi, S. & Cowen, P. J. Increased salivary cortisol after waking in depression. Psychopharmacology (Berl). (2005) doi:10.1007/s00213-005-0062-z.
64. Burke, H. M., Davis, M. C., Otte, C. & Mohr, D. C. Depression and cortisol responses to psychological stress: A meta-analysis. Psychoneuroendocrinology (2005) doi:10.1016/j.psyneuen.2005.02.010.
65. Vreeburg, S. A. et al. Major depressive disorder and hypothalamic-pituitary-adrenal axis activity: Results from a large cohort study. Arch. Gen. Psychiatry (2009) doi:10.1001/archgenpsychiatry.2009.50.
66. Posener, J. A. et al. 24-Hour monitoring of cortisol and corticotropin secretion in psychotic and nonpsychotic major depression. Arch. Gen. Psychiatry (2000) doi:10.1001/archpsyc.57.8.755.
67. Strickland, P. L. et al. Bio-social origins of depression in the community. Br. J. Psychiatry (2002) doi:10.1192/bjp.180.2.168.
68. Posener, J. A., DeBattista, C., Williams, G. H. & Schatzberg, A. F. Cortisol feedback during the HPA quiescent period in patients with major depression. Am. J. Psychiatry (2001) doi:10.1176/appi.ajp.158.12.2083.
69. Yehuda, R. Status of glucocorticoid alterations in post-traumatic stress disorder. in Annals of the New York Academy of Sciences (2009). doi:10.1111/j.1749-6632.2009.04979.x.
70. Olff, M., Güzelcan, Y., de Vries, G. J., Assies, J. & Gersons, B. P. R. HPA- and HPT-axis alterations in chronic posttraumatic stress disorder. Psychoneuroendocrinology (2006) doi:10.1016/j.psyneuen.2006.09.003.
71. Yehuda, R., Golier, J. A., Yang, R. K. & Tischler, L. Enhanced sensitivity to glucocorticoids in peripheral mononuclear leukocytes in posttraumatic stress disorder. Biol. Psychiatry (2004) doi:10.1016/j.biopsych.2004.02.010.
72. Meewisse, M. L., Reitsma, J. B., De Vries, G. J., Gersons, B. P. R. & Olff, M. Cortisol and post-traumatic stress disorder in adults: Systematic review and meta-analysis. in British Journal of Psychiatry (2007). doi:10.1192/bjp.bp.106.024877.
73. Arloth, J. et al. Genetic Differences in the Immediate Transcriptome Response to Stress Predict Risk-Related Brain Function and Psychiatric Disorders. Neuron 86, 1189–1202 (2015).
74. Zhao, T., Hu, Y., Zang, T. & Wang, Y. Integrate GWAS, eQTL, and mQTL Data to

Identify Alzheimer's Disease-Related Genes. Front. Genet. (2019) doi:10.3389/fgene.2019.01021.

75. Porcu, E. et al. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. Nat. Commun. (2019) doi:10.1038/s41467-019-10936-0.
76. Hemani, G. et al. The MR-base platform supports systematic causal inference across the human phenome. Elife (2018) doi:10.7554/eLife.34408.
77. Smith, G. D. & Ebrahim, S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology (2003) doi:10.1093/ije/dyg070.
78. Kappelmann, N. et al. Dissecting the Association Between Inflammation, Metabolic Dysregulation, and Specific Depressive Symptoms. JAMA Psychiatry (2020) doi:10.1001/jamapsychiatry.2020.3436.
79. Adams, D. M., Reay, W. R., Geaghan, M. P. & Cairns, M. J. Investigation of glycaemic traits in psychiatric disorders using Mendelian randomisation revealed a causal relationship with anorexia nervosa. Neuropsychopharmacology (2020) doi:10.1038/s41386-020-00847-w.
80. Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Research (2012) doi:10.1101/gr.136184.111.
81. Arrigoni, L. et al. Standardizing chromatin research: A simple and universal method for ChIP-seq. Nucleic Acids Res. (2015) doi:10.1093/nar/gkv1495.
82. Schiller, B. J., Chodankar, R., Watson, L. C., Stallcup, M. R. & Yamamoto, K. R. Glucocorticoid receptor binds half sites as a monomer and regulates specific target genes. Genome Biol. (2014) doi:10.1186/s13059-014-0418-y.
83. Borschiwer, M., Bothe, M., Kibar, G., Fuchs, A. & Schöne, S. Androgen and glucocorticoid receptor direct distinct transcriptional programs by receptor-specific and shared DNA binding sites. 1–47 (2020).
84. Bothe, M., Buschow, R. & Meijsing, S. H. Glucocorticoid receptor activation induces gene-specific transcriptional memory and universally reversible changes in chromatin accessibility. bioRxiv (2021).
85. Kang, H. et al. Dynamic regulation of histone modifications and long-range chromosomal

interactions during postmitotic transcriptional reactivation. Genes Dev. (2020) doi:10.1101/GAD.335794.119.

86. Schöne, S. et al. Synthetic STARR-seq reveals how DNA shape and sequence modulate transcriptional output and noise. PLoS Genet. (2018) doi:10.1371/journal.pgen.1007793.
87. Uhlén, M. et al. Tissue-based map of the human proteome. Science (80-.). (2015) doi:10.1126/science.1260419.
88. Sarabdjitsingh, R. A., Meijer, O. C., Schaaf, M. J. M. & de Kloet, E. R. Subregion-specific differences in translocation patterns of mineralocorticoid and glucocorticoid receptors in rat hippocampus. Brain Res. (2009) doi:10.1016/j.brainres.2008.10.048.
89. Neumayr, C., Pagani, M., Stark, A. & Arnold, C. D. STARR-seq and UMI-STARR-seq: Assessing Enhancer Activities for Genome-Wide-, High-, and Low-Complexity Candidate Libraries. Curr. Protoc. Mol. Biol. (2019) doi:10.1002/cpmb.105.
90. Ashuach, T. et al. MPRAnalyze: Statistical framework for massively parallel reporter assays. Genome Biol. (2019) doi:10.1186/s13059-019-1787-z.
91. Love, M. I. et al. Role of the chromatin landscape and sequence in determining cell type-specific genomic glucocorticoid receptor binding and gene regulation. Nucleic Acids Res. (2017) doi:10.1093/nar/gkw1163.
92. Cruceanu, C. et al. Cell-type specific impact of glucocorticoid receptor activation on the developing brain. bioRxiv (2020) doi:10.1101/2020.01.09.897868.
93. Machanick, P. & Bailey, T. L. MEME-ChIP: Motif analysis of large DNA datasets. Bioinformatics (2011) doi:10.1093/bioinformatics/btr189.
94. Biddie, S. C. & Hager, G. L. Glucocorticoid receptor dynamics and gene regulation. Stress (2009) doi:10.1080/10253890802506409.
95. Weikum, E. R., Knuesel, M. T., Ortlund, E. A. & Yamamoto, K. R. Glucocorticoid receptor control of transcription: Precision and plasticity via allostery. Nature Reviews Molecular Cell Biology (2017) doi:10.1038/nrm.2016.152.
96. Deplancke, B., Alpern, D. & Gardeux, V. The Genetics of Transcription Factor DNA Binding Variation. Cell (2016) doi:10.1016/j.cell.2016.07.012.
97. Johnston, A. D., Simões-Pires, C. A., Thompson, T. V., Suzuki, M. & Greally, J. M. Functional genetic variants can mediate their regulatory effects through alteration of transcription factor binding. Nat. Commun. (2019) doi:10.1038/s41467-019-11412-5.

98. Schöne, S. et al. Sequences flanking the core-binding site modulate glucocorticoid receptor structure and activity. Nat. Commun. (2016) doi:10.1038/ncomms12621.
99. Quatrini, L. & Ugolini, S. New insights into the cell- and tissue-specificity of glucocorticoid actions. Cellular and Molecular Immunology (2020) doi:10.1038/s41423-020-00526-2.
100. Dong, X. et al. Modeling gene expression using chromatin features in various cellular contexts. Genome Biol. (2012) doi:10.1186/gb-2012-13-9-r53.
101. Karlić, R., Chung, H. R., Lasserre, J., Vlahoviček, K. & Vingron, M. Histone modification levels are predictive for gene expression. Proc. Natl. Acad. Sci. U. S. A. (2010) doi:10.1073/pnas.0909344107.
102. Stevens, F. L., Hurley, R. A. & Taber, K. H. Anterior cingulate cortex: Unique role in cognition and emotion. J. Neuropsychiatry Clin. Neurosci. (2011) doi:10.1176/jnp.23.2.jnp121.
103. Nadarajah, B. & Parnavelas, J. G. Modes of neuronal migration in the developing cerebral cortex. Nature Reviews Neuroscience (2002) doi:10.1038/nrn845.
104. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. (2005) doi:10.1073/pnas.0506580102.
105. Blockus, H. & Chédotal, A. Slit-robo signaling. Dev. (2016) doi:10.1242/dev.132829.
106. Allen, H. L. et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature (2010) doi:10.1038/nature09410.
107. Trajanoska, K. et al. Assessment of the genetic and clinical determinants of fracture risk: Genome wide association and mendelian randomisation study. BMJ (2018) doi:10.1136/bmj.k3225.
108. Kendler, K. S., Karkowski, L. M. & Prescott, C. A. Causal relationship between stressful life events and the onset of major depression. Am. J. Psychiatry (1999) doi:10.1176/ajp.156.6.837.
109. Caspi, A. & Moffitt, T. E. Gene-environment interactions in psychiatry: Joining forces with neuroscience. Nature Reviews Neuroscience (2006) doi:10.1038/nrn1925.
110. Vockley, C. M. et al. Direct GR Binding Sites Potentiate Clusters of TF Binding across the Human Genome. Cell (2016) doi:10.1016/j.cell.2016.07.049.

111. Reddy, T. E. et al. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. Genome Res. (2009) doi:10.1101/gr.097022.109.
112. Severinova, E. et al. Glucocorticoid Receptor-Binding and Transcriptome Signature in Cardiomyocytes. J. Am. Heart Assoc. (2019) doi:10.1161/JAHA.118.011484.
113. Murani, E., Trakooljul, N., Hadlich, F., Ponsuksili, S. & Wimmers, K. Transcriptome responses to dexamethasone depending on dose and glucocorticoid receptor sensitivity in the liver. Front. Genet. (2019) doi:10.3389/fgene.2019.00559.
114. Polman, J. A. E. et al. A genome-wide signature of glucocorticoid receptor binding in neuronal PC12 cells. BMC Neurosci. (2012) doi:10.1186/1471-2202-13-118.
115. Uhlenhaut, N. H. et al. Insights into Negative Regulation by the Glucocorticoid Receptor from Genome-wide Profiling of Inflammatory Cistromes. Mol. Cell (2013) doi:10.1016/j.molcel.2012.10.013.
116. Chandler, V. L., Maler, B. A. & Yamamoto, K. R. DNA sequences bound specifically by glucocorticoid receptor in vitro render a heterologous promoter hormone responsive in vivo. Cell (1983) doi:10.1016/0092-8674(83)90430-0.
117. Gertz, J. et al. Distinct properties of cell-type-specific and shared transcription factor binding sites. Mol. Cell (2013) doi:10.1016/j.molcel.2013.08.037.
118. Dostert, A. & Heinzel, T. Negative glucocorticoid receptor response elements and their role in glucocorticoid action. Curr. Pharm. Des. (2004) doi:10.2174/1381612043383601.
119. Miner, J. N. & Yamamoto, K. R. Regulatory crosstalk at composite response elements. Trends Biochem. Sci. (1991) doi:10.1016/0968-0004(91)90168-U.
120. Biddie, S. C. et al. Transcription Factor AP1 Potentiates Chromatin Accessibility and Glucocorticoid Receptor Binding. Mol. Cell 43, 145–155 (2011).
121. Portuguez, A. S. et al. Gene activation and repression by the glucocorticoid receptor are mediated by sequestering Ep300 and two modes of chromatin binding. bioRxiv (2019) doi:10.1101/764480.
122. Meijsing, S. H. et al. DNA Binding Site Sequence Structure and Activity. Science (80-.). (2009).
123. De Bosscher, K., Vanden Berghe, W. & Haegeman, G. The interplay between the glucocorticoid receptor and nuclear factor-κB or activator protein-1: Molecular

mechanisms for gene repression. Endocrine Reviews (2003) doi:10.1210/er.2002-0006.

124. Grøntved, L. et al. C/EBP maintains chromatin accessibility in liver and facilitates glucocorticoid receptor recruitment to steroid response elements. EMBO J. (2013) doi:10.1038/emboj.2013.106.
125. Petta, I. et al. The Interactome of the Glucocorticoid Receptor and Its Influence on the Actions of Glucocorticoids in Combatting Inflammatory and Infectious Diseases. Microbiol. Mol. Biol. Rev. (2016) doi:10.1128/mmbr.00064-15.
126. Choi, J. et al. Massively parallel reporter assays of melanoma risk variants identify MX2 as a gene promoting melanoma. Nat. Commun. (2020) doi:10.1038/s41467-020-16590-1.
127. Vockley, C. M. et al. Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. Genome Res. (2015) doi:10.1101/gr.190090.115.
128. Zhang, Y., Qi, G., Park, J. H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. Nat. Genet. (2018) doi:10.1038/s41588-018-0193-x.
129. Momozawa, Y. & Mizukami, K. Unique roles of rare variants in the genetics of complex diseases in humans. Journal of Human Genetics (2021) doi:10.1038/s10038-020-00845-2.
130. Calafato, M. S. et al. Use of schizophrenia and bipolar disorder polygenic risk scores to identify psychotic disorders. Br. J. Psychiatry (2018) doi:10.1192/bjp.2018.89.
131. Elbau, I. G. et al. The brain's hemodynamic response function rapidly changes under acute psychosocial stress in association with genetic and endocrine stress response markers. Proc. Natl. Acad. Sci. U. S. A. (2018) doi:10.1073/pnas.1804340115.
132. Townsley, K. G., Brennand, K. J. & Huckins, L. M. Massively parallel techniques for cataloguing the regulome of the human brain. Nature Neuroscience (2020) doi:10.1038/s41593-020-00740-1.
133. Piechota, M. et al. Transcriptional signatures of steroid hormones in the striatal neurons and astrocytes. BMC Neurosci. (2017) doi:10.1186/s12868-017-0352-5.
134. Lieberman, R., Kranzler, H. R., Levine, E. S. & Covault, J. Examining FKBP5 mRNA expression in human iPSC-derived neural cells. Psychiatry Res. (2017) doi:10.1016/j.psychres.2016.11.027.
135. Fitzsimons, C. P. et al. Knockdown of the glucocorticoid receptor alters functional

integration of newborn neurons in the adult hippocampus and impairs fear-motivated behavior. Mol. Psychiatry (2013) doi:10.1038/mp.2012.123.

136. Kolber, B. J. et al. Central amygdala glucocorticoid receptor action promotes fear-associated CRH activation and conditioning. Proc. Natl. Acad. Sci. U. S. A. (2008) doi:10.1073/pnas.0803216105.
137. Hartmann, J. et al. Forebrain glutamatergic, but not GABAergic, neurons mediate anxiogenic effects of the glucocorticoid receptor. Mol. Psychiatry (2017) doi:10.1038/mp.2016.87.
138. Zhang, S. et al. Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. Science (80-.). (2020) doi:10.1126/science.aay3983.
139. Schmaal, L. et al. Subcortical brain alterations in major depressive disorder: Findings from the ENIGMA Major Depressive Disorder working group. Mol. Psychiatry (2016) doi:10.1038/mp.2015.69.
140. Bersani, F. S. et al. Cingulate Cortex in Schizophrenia: Its relation with negative symptoms and psychotic onset. A review study. Eur. Rev. Med. Pharmacol. Sci. (2014) doi:10.1111/j.1601-5215.2011.00640.x.
141. Acampora, D., Postiglione, M. P., Avantaggiato, V., Di Bonito, M. & Simeone, A. The role of Otx and Otp genes in brain development. Int. J. Dev. Biol. (2000) doi:10.1387/ijdb.11061431.
142. Peña, C. J. et al. Early life stress confers lifelong stress susceptibility in mice via ventral tegmental area OTX2. Science (80-.). (2017) doi:10.1126/science.aan4491.
143. Zimmermann, C. A. et al. Stress dynamically regulates co-expression networks of glucocorticoid receptor-dependent MDD and SCZ risk genes. Transl. Psychiatry (2019) doi:10.1038/s41398-019-0373-1.
144. Radley, J. J., Anderson, R. M., Hamilton, B. A., Alcock, J. A. & Romig-Martin, S. A. Chronic stress-induced alterations of dendritic spine subtypes predict functional decrements in an hypothalamo-pituitary-adrenal-inhibitory prefrontal circuit. J. Neurosci. (2013) doi:10.1523/JNEUROSCI.0287-13.2013.
145. Sebastian, V., Estil, J. B., Chen, D., Schrott, L. M. & Serrano, P. A. Acute Physiological Stress Promotes Clustering of Synaptic Markers and Alters Spine Morphology in the Hippocampus. PLoS One (2013) doi:10.1371/journal.pone.0079077.

146. Abe, S. et al. Modification of single-nucleotide polymorphism in a fully humanized CYP3A mouse by genome editing technology. Sci. Rep. (2017) doi:10.1038/s41598-017-15033-0.
147. Lee, H., Yoon, D. E. & Kim, K. Genome editing methods in animal models. Animal Cells and Systems (2020) doi:10.1080/19768354.2020.1726462.
148. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. Science (80-.). (2018) doi:10.1126/science.aat8464.
149. Matoba, N. et al. Common genetic risk variants identified in the SPARK cohort support DDHD2 as a candidate risk gene for autism. Transl. Psychiatry (2020) doi:10.1038/s41398-020-00953-9.
150. Menke, A. et al. Dexamethasone stimulated gene expression in peripheral blood is a sensitive marker for glucocorticoid receptor resistance in depressed patients. Neuropsychopharmacology (2012) doi:10.1038/npp.2011.331.
151. Breen, M. S. et al. Differential transcriptional response following glucocorticoid activation in cultured blood immune cells: a novel approach to PTSD biomarker development. Transl. Psychiatry (2019) doi:10.1038/s41398-019-0539-x.
152. Steel, Z. et al. The global prevalence of common mental disorders: A systematic review and meta-analysis 1980-2013. Int. J. Epidemiol. (2014) doi:10.1093/ije/dyu038.

7. Supplementary Tables

Table S1. Primer information and sequences (when available) used in this thesis.

IDT TaqMan Probes for analyzing GR activation		
ID	Description	IDT ID
1	FKBP5	Hs.PT.58.20523859
2	YWHAZ	Hs.PT.39a.22214858
3	SGK1	Hs.PT.58.19153459.gs
4	TSC22D3	Hs.PT.58.4331913
GR ChIP-qPCR Primers		
ID	Description	Sequence (5'-3')
5	Intergenic (FKBP5_TULP)_neg forward	CGGCCCAGACTTGACTTTCA
6	Intergenic (FKBP5_TULP) neg reverse	CCTTGCCCAGCCCCAAATAA
7	FKBP5 (intron 5) positive forward	ACTGCCCTAGAGCAATTTTGTTT
8	FKBP5 (intron 5) positive reverse	TGTCAGCACATCGAGTTCATGT
STARR plasmid DNA amplification primers		
9	plasmid_pool_1	TAGAGCATGCACCGGACACTCTTTCCCTACACGAC
10	plasmid_pool_2	GGCCGAATTCGTCGAGTGACTGGAGTTCAGACGTG
STARR-seq plasmid DNA library Illumina		
11	Plasmid_library_Illumina_1	TATCATGTCTGCTCGAAGCGG
12	Plasmid_library_Illumina_2	GGATTTGATATTCACCTGGC
STARR-seq library preparation		
13	STARR-seq for RT with barcode	CAAGCAGAAGACGGCATACGAGATnnnnnnnnGTGACTGGAGTTCAGACGTGTG CTCTTCCGATCT
14	Universal primer	CAAGCAGAAGACGGCATACGA
15	2nd_strand-i501	AATGATACGGCGACCACCGAGATCTACACTAGATCGCACACTCTTTCCCTAC ACGACGCTC
16	2nd_strand-i502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACA CGACGCTC
17	2nd_strand-i503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACA CGACGCTC
18	2nd_strand-i504	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTAC ACGACGCTC

19	2nd_strand-i505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTAC ACGACGCTC
20	2nd_strand-i506	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTAC ACGACGCTC
21	2nd_strand-i507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTAC ACGACGCTC
22	2nd_strand-i508	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTACACTCTTTCCCTAC ACGACGCTC
23	2nd_strand-i517	AATGATACGGCGACCACCGAGATCTACACGCGTAAGAACACTCTTTCCCTAC ACGACGCTC
24	2nd_strand-i510	AATGATACGGCGACCACCGAGATCTACACCGTCTAATACACTCTTTCCCTAC ACGACGCTC
25	2nd_strand-i511	AATGATACGGCGACCACCGAGATCTACACTCTCTCCGACACCTTTCCCTACAC GACGCTC
26	2nd_strand-i518	AATGATACGGCGACCACCGAGATCTACACCTATTAAGACACTCTTTCCCTAC ACGACGCTC
Primers for STARR-seq Validation		
27	Gene fragment amplification forward	TAGAGCATGCACCGGACACTCTTTCCCTACACGAC
28	Gene fragment amplification reverse	GGCCGAATTCGTCGAGTGACTGGAGTTCAGACGTG
29	mRNA STARR-seq	CAAACTCATCAATGTATCTTATCATG
30	mRNA_Rpl19	GAGGCCAGTATGTACAGACAAAGTGG
31	hRPL19 forward	ATGTATCACAGCCTGTACCTG
32	hRPL19 reverse	TTCTTGGTCTCTTCCTCCTTG
33	GFP forward	GGCCAGCTGTTGGGGTGTC
34	GFP reverse	TTGGGACAACTCCAGTGAAGA
Sanger sequencing of individual STARR-seq fragments		
35	STARR_sanger	GCATTCTAGTTGTGGTTTGTCCA

Table S2. Regulatory regions of individual STARR-seq constructs.

ID	Description	Reference	Sequence
STARR-FKBP5	STARR-seq vector with known dex-inducible binding site on FKBP5 Fragment inserted between Age1 and Sal1	[98]	TAGAGCATGCACCGGACACTCTTTCCCT ACACGACGCTCTTCCGATCTAGCGAAA GAACATCCTGTGCCCGTCGCCGGATGC GCGCGGGGGGAGGAGGTAGTCCGGGG AGGGAGAGCAAGACCGGGGGAGGCCC GGGACGGGGAAAGGCGCGGCCTCCTTC CCCCCGCCGCAACCTCCTCCTCTGGGGG CGCTGGCCCCCTCTGCTCCCCGCGCCTT AGGCTGAGCTACCCGGAGGCCCCAGGA TCTGGTTCCTGGGCAAGATCGGAAGAG CACACGTCTGAACTCCAGTCACTCGAC GAATTCGGCC
STARR-CMV	STARR-seq vector with CMV enhancer that is constitutively active Fragment inserted between Age1 and Sal1	[98]	TAGAGCATGCACCGGACACTCTTTCCCT ACACGACGCTCTTCCGATCTTCAATATT GGCCATTAGCCATATTATTCATTGGTTA TATAGCATAAATCAATATTGGCTATTGG CCATTGCATACGTTGTATCTATATCATA ATATGTACATTTATATTGGCTCATGTCC AATATGACCGCCATGTTGGCATTGATTA TTGACTAGTTATTAATAGTAATCAATTA CGGGGTCATTAGTTCATAGCCCATATAT GGAGTTCCGCGTTACATAACTTACGGTA AATGGCCCGCCTGGCTGACCGCCCAAC GACCCCCGCCCATTGACGTCAATAATG ACGTATGTTCCCATAGTAACGCCAATA GGGACTTTCCATTGACGTCAATGGGTG GAGTATTTACGGTAAACTGCCCACTTGG CAGTACATCAAGTGTATCATATGCCAA GTCCGCCCCCTATTGACGTCAATGACGG TAAATGGCCCGCCTGGCATTATGCCCA GTACATGACCTTACGGGACTTTCCTACT TGGCAGTACATCTACGTATTAGTCATCG CTATTACCATGGTGATGCGGTTTTGGCA GTACACCAATGGGCGTGGATAGCGGTT TGACTCACGGGGATTTCCAAGTCTCCAC CCCATTGACGTCAATGGGAGTTTGTTTT GGCACCAAAATCAACGGGACTTTCCAA AATGTCGTAACAACTAGATCGGAAGAG CACACGTCTGAACTCCAGTCACTCGAC GAATTCGGCC
STARR-scr	STARR-seq vector with scrambled version of a fragment of known dex-inducible binding site on FKBP5 Insert fragment between Age1 and Sal1	[98]	TAGAGCATGCACCGGACACTCTTTCCCT ACACGACGCTCTTCCGATCTCAGCGAA AGAAACTCCGTTGCCCGTCGCTAGATC GGAAGAGCACACGTCTGAACTCCAGTC ACTCGACGAATTCGGCC

Table S3. Descriptions for EIDs.

EID	Tissue	Description
E062	Blood & T-cell	Primary mononuclear cells from peripheral blood
E034	Blood & T-cell	Primary T cells from peripheral blood
E045	Blood & T-cell	Primary T cells effector/memory enriched from peripheral blood
E033	Blood & T-cell	Primary T cells from cord blood
E044	Blood & T-cell	Primary T regulatory cells from peripheral blood
E043	Blood & T-cell	Primary T helper cells from peripheral blood
E039	Blood & T-cell	Primary T helper naive cells from peripheral blood
E041	Blood & T-cell	Primary T helper cells PMA-I stimulated
E042	Blood & T-cell	Primary T helper 17 cells PMA-I stimulated
E040	Blood & T-cell	Primary T helper memory cells from peripheral blood 1
E037	Blood & T-cell	Primary T helper memory cells from peripheral blood 2
E048	Blood & T-cell	Primary T CD8+ memory cells from peripheral blood
E038	Blood & T-cell	Primary T helper naive cells from peripheral blood
E047	Blood & T-cell	Primary T CD8+ naive cells from peripheral blood
E029	HSC & B-cell	Primary monocytes from peripheral blood
E031	HSC & B-cell	Primary B cells from cord blood
E035	HSC & B-cell	Primary hematopoietic stem cells
E051	HSC & B-cell	Primary hematopoietic stem cells G-CSF-mobilized Male
E050	HSC & B-cell	Primary hematopoietic stem cells G-CSF-mobilized Female
E036	HSC & B-cell	Primary hematopoietic stem cells short term culture
E032	HSC & B-cell	Primary B cells from peripheral blood
E046	HSC & B-cell	Primary Natural Killer cells from peripheral blood
E030	HSC & B-cell	Primary neutrophils from peripheral blood
E071	Brain	Brain Hippocampus Middle
E074	Brain	Brain Substantia Nigra
E068	Brain	Brain Anterior Caudate
E069	Brain	Brain Cingulate Gyrus
E072	Brain	Brain Inferior Temporal Lobe
E067	Brain	Brain Angular Gyrus
E073	Brain	Brain_Dorsolateral_Prefrontal_Cortex
E070	Brain	Brain Germinal Matrix
E082	Brain	Fetal Brain Female
E081	Brain	Fetal Brain Male

8. Supplementary Figures

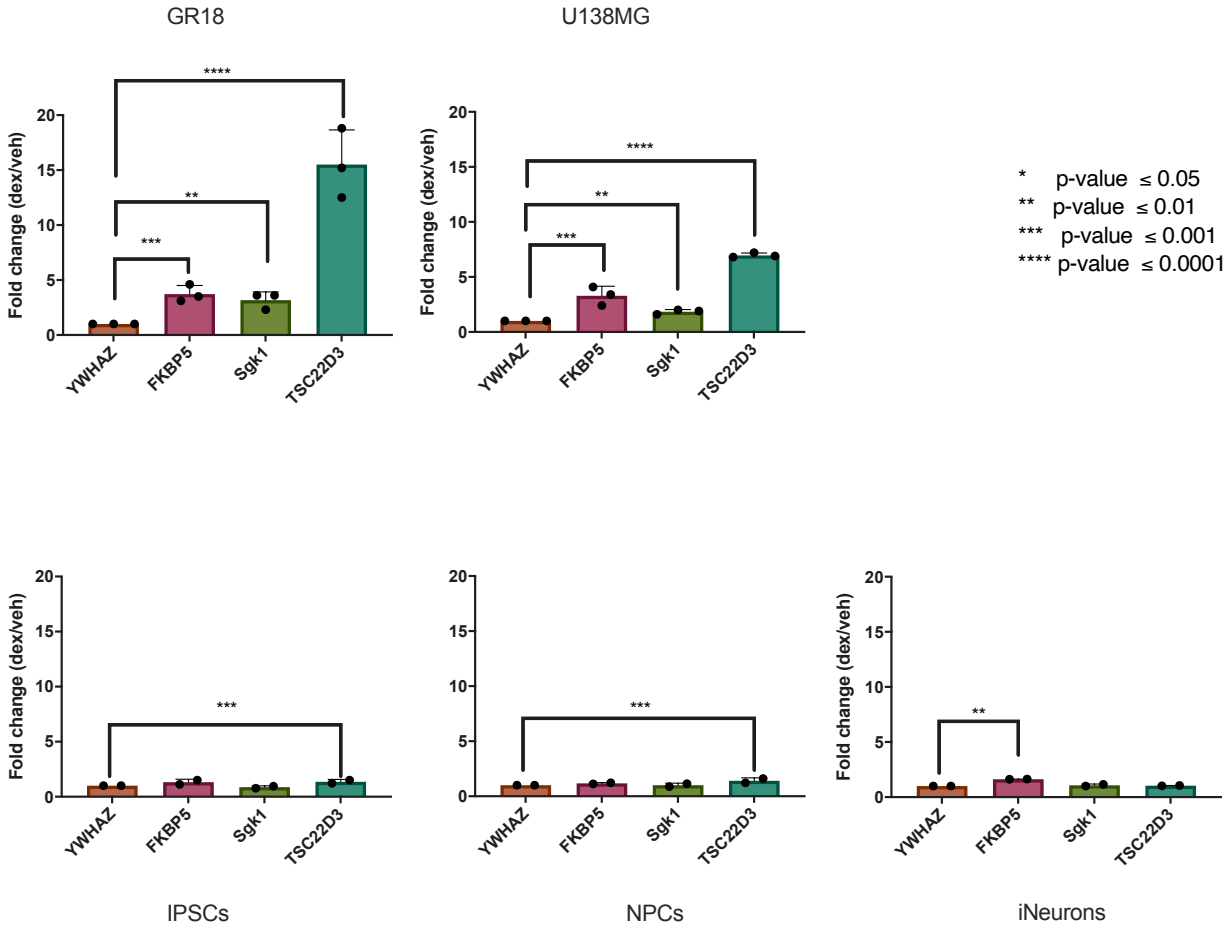


Figure S1. Transcriptional response to dex in GR target genes. Fold changes of expression for three GR-target genes after dex treatment in two different cell lines, GR18 and U138MG. Fold changes of expression for IPSCs, and neural precursor cells (NPCs) and iNeurons derived from IPSCs from two individuals are displayed below.

9. Acknowledgements

I would first like to thank my TAC members, Chris Turck, Nadine Provencal, and Sebastiaan Meijsing, for supporting me throughout my PhD, providing me with constructive comments and feedback, and for fruitful discussions. I would especially like to thank Elisabeth for her guidance and mentorship over the past six years. I have appreciated your insightful comments, support, and enthusiasm. In moments of frustration and discouragement, I knew that speaking to you would always make me feel excited about research again. I am also extremely grateful for your compassion and kindness throughout my PhD.

Ich möchte mich an dieser Stelle auch an Maik, Susi, Bärbel, Jose, und Laura für alle eure technische Hilfe bedanken. Ich bin sehr dankbar, dass Ihr mich mit eurer Expertise stets unterstützt habt – Sowohl bei den RNA und ChIP-library preps als auch bei den Western Blots und beim Deutsch lernen. I would also like to thank the rest of the Binder lab for their support and help, as well as Christine, my fellow parallel enhancer screen partner– and good friend. Marta, thank you for bearing with me and all my qPCR problems, and trying to understand the mysteries of the lab with me. I appreciate our frequent meetings dearly. I am especially grateful for all the office supplies you most graciously lent to me. I am also extremely grateful to Monika, who has helped me with various organizational tasks, but also for her more general support over the years.

To my "foster" lab in Berlin – I'm grateful for all the help I received for the STARR-seq. I could not have felt more at home and welcomed for the few months I was in Berlin. Your generosity with your time and data are very much appreciated. I would especially like to thank Melissa for taking so much time to help me set up the STARR-seq. Thank you Sebastiaan, for patiently answering all of my questions, inviting me into your group, and providing me with valuable feedback and ideas for project.

To my friends in Canada; especially Jacquelyn, Nathalie and Mariette - the phone calls, letters and motivational postcards I received helped me get through these last years. Leaving you in 2015 was not an easy decision, but I am so grateful to have you by my side (even from 7000km abroad). To Linnea and Michi, the first year of GSN would have not been possible without you. I think very often about our time in K3. Michi, I am glad you stayed in Munich at least a little longer, as your friendship and support has been invaluable. Knowing I could always run upstairs if I was having a bad day for a coffee or a glass of wine with you was invaluable.

An die Schwestern Simone, Nella, Annette, und Anja: Danke für all das, was ihr für mich gemacht habt. Ich habe mich immer sehr geliebt und zugehört gefühlt. Ihr seid tatsächlich meine "Muttern" in Deutschland.

Silvia, your friendship is something I value more than you can know. The years in the lab (and out) would not have been same without you. You can always make me laugh, even on the worst, most frustrating days. Thank you for always listening to me and teaching me how to properly eat pesto.

Peter; danke. Danke für dein Geduld, deine Hilfe, und dass du immer an meiner Seite bist.

To my lovely sisters and my parents. Thank you for everything you have done for me. To mum and dad; thank you for instilling in me a curiosity and passion for science from a young age. Thank you for dragging me to science museums and teaching me about biology by taking us camping, hiking, and gardening (even if I complained at the time). Kirsten and Lara; living away from you has been so difficult, but I am so grateful to have you lovely sisters. Thank you all for supporting me unconditionally.

10. Curriculum Vitae

Education

2015/10 - present	Doctor of Philosophy in Systemic Neurosciences Max Planck Institute of Psychiatry Graduate School of Systemic Neurosciences Munich, Germany
2013/5 - 2015/6	Bachelor of Science, Biology, University of Winnipeg Winnipeg, Manitoba
2012/9 - 2013/4	Canadian Mennonite University Winnipeg, Manitoba

Recognitions and Awards

2019/11-2021/11	Add-On Fellowship for Interdisciplinary Life Science (12 500€) Joachim Herz Foundation
2018/05-2021/04	Foreign Doctoral Award (105 000$) Canadian Institutes of Health Research

List of publications

Penner-Goeke S, Binder EB. (2019) Epigenetics and Depression. Dialogues of Clinical Neuroscience. 21(4): 397-405.

Penner-Goeke S, Lichtensztejn Z, Neufeld M, Ali J, Altman A, Nachtigal M, McManus K. (2017). The temporal dynamics of chromosome instability in ovarian cancer cell lines and primary patient samples. PLoS Genetics. 13(4).

Cisyk A, Penner-Goeke S, Lichtensztejn Z, Nugent Z, Wightman RH, Singh H, McManus K. (2015). Characterizing the prevalence of chromosome instability in interval colorectal cancer. Neoplasia. 17(3): 306-316