
Characterizing the epigenetic response to stress using next generation sequencing

Dissertation an der Fakultät für Biologie
der Ludwig-Maximilians-Universität München



Simone Röh-Karamihalev
München, 2023

1. Gutachter: PD Dr. Mathias V. Schmidt
2. Gutachter: Professor Dr. Wolfgang Enard

Datum der Abgabe: 15.05.2023
Mündliche Prüfung: 21.12.2023

Table of Contents

Curriculum Vitae	1
List of Publications	3
Abbreviations	7
Declaration of contribution	9
Summary	11
Introduction	13
Epigenetics	13
Histone modification	13
DNA methylation	14
Epigenetic regulation and the brain	15
Hypothalamic–pituitary–adrenal axis	16
FKBP5	19
Next generation sequencing	22
The evolution of NGS	22
The technology in NGS.....	23
The applications of NGS.....	25
Results	31
Publication I.....	31
Publication II.....	43
Publication III.....	55
Discussion	71
Sequencing on the SOLiD 5500xl System – in-depth characterization of the GC bias	71
HAM-TBS: high-accuracy methylation measurements via targeted bisulfite sequencing.....	73
Identification of dynamic glucocorticoid-induced methylation changes at the FKBP5 locus	77
Conclusion	80
References	81
Acknowledgements.....	89
Eidesstattliche Erklärung.....	91

List of Publications

Years 2015-present

Penner-Goeke, S., Bothe, M., Kappelmann, N., Kreitmaier, P., Kaya, E., Pöhlchen, D., Kühnel, A., Czamara, D., group, B. working, Glaser, L. V., **Roeh, S.**, Ködel, M., Monteserin-Garcia, J., Rummel, C., Arloth-Knauer, J., Diener-Hölzl, L., Woelfel, B., Sauer, S., Riesenberger, S., Ziller, M. J., Labeur, M., Meijnsing, S. H. & Binder, E. B. Assessment of glucocorticoid-induced enhancer activity of eSNP regions using STARR-seq reveals novel molecular mechanisms in psychiatric disorders. *Medrxiv* 2022.05.18.22275090 (2022). doi:10.1101/2022.05.18.22275090

Cruceanu, C., Dony, L., Krontira, A. C., Fischer, D. S., **Roeh, S.**, Giaimo, R. D., Kyrousi, C., Kaspar, L., Arloth, J., Czamara, D., Gerstner, N., Martinelli, S., Wehner, S., Breen, M. S., Koedel, M., Sauer, S., Sportelli, V., Rex-Haffner, M., Cappello, S., Theis, F. J. & Binder, E. B. Cell-Type-Specific Impact of Glucocorticoid Receptor Activation on the Developing Brain: A Cerebral Organoid Study. *Am J Psychiat* 179, 375–387 (2022).

Gerstner, N., Krontira, A. C., Cruceanu, C., **Roeh, S.**, Pütz, B., Sauer, S., Rex-Haffner, M., Schmidt, M. V., Binder, E. B. & Knauer-Arloth, J. DiffBrainNet: Differential analyses add new insights into the response to glucocorticoids at the level of genes, networks and brain regions. *Neurobiology Stress* 21, 100496 (2022).

Womersley, J. S., **Roeh, S.**, Martin, L., Ahmed-Leitao, F., Sauer, S., Rex-Haffner, M., Hemmings, S. M. J., Binder, E. B. & Seedat, S. FKBP5 intron 7 methylation is associated with higher anxiety proneness and smaller right thalamus volume in adolescents. *Brain Struct Funct* 227, 2809–2820 (2022).

Lopez, J. P., Lücken, M. D., Brivio, E., Karamihalev, S., Kos, A., Donno, C. D., Benjamin, A., Yang, H., Dick, A. L. W., Stoffel, R., Flachskamm, C., Ressler, A., **Roeh, S.**, Huettl, R.-E., Parl, A., Eggert, C., Novak, B., Yan, Y., Yeoh, K., Holzapfel, M., Hauger, B., Harbich, D., Schmid, B., Giaimo, R. D., Turck, C. W., Schmidt, M. V., Deussing, J. M., Eder, M., Dine, J., Theis, F. J. & Chen, A. Ketamine exerts its sustained antidepressant effects via cell-type-specific regulation of *Kcnq2*. *Neuron* 110, 2283-2298.e9 (2022).

Krontira, A. C., Cruceanu, C., Kyrousi, C., Dony, L., Link, M.-H., Kappelmann, N., Pöhlchen, D., **Roeh, S.**, Sportelli, V., Wölfel, B., Ködel, M., Sauer, S., Monika-Rex-Haffner, Labeur, M., Cappello, S. & Binder, E. B. Temporal regulation of ZBTB16 expression by glucocorticoids alters human cortical neurogenesis. *Biorxiv* 2022.08.21.504700 (2022). doi:10.1101/2022.08.21.504700

Lopez, J. P., Brivio, E., Santambrogio, A., Donno, C. D., Kos, A., Peters, M., Rost, N., Czamara, D., Brückl, T. M., **Roeh, S.**, Pöhlmann, M. L., Engelhardt, C.,

Ressle, A., Stoffel, R., Tontsch, A., Villamizar, J. M., Reincke, M., Riester, A., Sbiera, S., Fassnacht, M., Mayberg, H. S., Craighead, W. E., Dunlop, B. W., Nemeroff, C. B., Schmidt, M. V., Binder, E. B., Theis, F. J., Beuschlein, F., Andoniadou, C. L. & Chen, A. Single-cell molecular profiling of all three components of the HPA axis reveals adrenal ABCB1 as a regulator of stress adaptation. *Sci Adv* 7, eabe4497 (2021).

Häusl, A. S., Brix, L. M., Hartmann, J., Pöhlmann, M. L., Lopez, J.-P., Menegaz, D., Brivio, E., Engelhardt, C., **Roeh, S.**, Bajaj, T., Rudolph, L., Stoffel, R., Hafner, K., Goss, H. M., Reul, J. M. H. M., Deussing, J. M., Eder, M., Ressler, K. J., Gassen, N. C., Chen, A. & Schmidt, M. V. The co-chaperone Fkbp5 shapes the acute stress response in the paraventricular nucleus of the hypothalamus of male mice. *Mol Psychiatr* 26, 3060–3076 (2021).

Dick, A. L., Zhao, Q., Crossin, R., Baker-Andresen, D., Li, X., Edson, J., **Roeh, S.**, Marshall, V., Bredy, T. W., Lawrence, A. J. & others. Adolescent chronic intermittent toluene inhalation dynamically regulates the transcriptome and neuronal methylome within the rat medial prefrontal cortex. *Addiction Biology* e12937 (2020).

Provençal, N., Arloth, J., Cattaneo, A., Anacker, C., Cattane, N., Wiechmann, T., **Röh, S.**, Ködel, M., Klengel, T., Czamara, D., Müller, N. S., Lahti, J., team, P., Räikkönen, K., Pariante, C. M. & Binder, E. B. Glucocorticoid exposure during hippocampal neurogenesis primes future stress response by inducing changes in DNA methylation. *Proc National Acad Sci* 117, 23280–23285 (2020).

Dournes, C., Dine, J., Lopez, J.-P., Brivio, E., Anderzhanova, E., **Roeh, S.**, Kuehne, C., Holzapfel, M., Huettl, R.-E., Stoffel, R., Tietze, L., Eggert, C., Schieven, M., Jakovcevski, M., Deussing, J. M. & Chen, A. Hypothalamic glucocorticoid receptor in CRF neurons is essential for HPA axis habituation to repeated stressor. *bioRxiv* (2020). doi:10.1101/2020.11.30.402024

Zannas, A. S., Jia, M., Hafner, K., Baumert, J., Wiechmann, T., Pape, J. C., Arloth, J., Kodel, M., Martinelli, S., Roitman, M. & **others**. Epigenetic upregulation of FKBP5 by aging and stress contributes to NF-κB--driven inflammation and cardiovascular risk. *Proceedings of the National Academy of Sciences* 116, 11370--11379 (2019).

Wiechmann, T., Roh, S., Sauer, S., Czamara, D., Arloth, J., Kodel, M., Beintner, M., Knop, L., Menke, A., Binder, E. B. & others. Identification of dynamic glucocorticoid-induced methylation changes at the FKBP5 locus. *Clinical epigenetics* 11, 1--14 (2019).

Forkosh, O., Karamihalev, S., **Roeh, S.**, Alon, U., Anpilov, S., Touma, C., Nussbaumer, M., Flachskamm, C., Kaplick, P. M., Shemesh, Y. & others. Identity domains capture individual differences from across the behavioral repertoire. *Nature Neuroscience* 22, 2023--2028 (2019).

Prestel, M., Prell-Schicker, C., Webb, T., Malik, R., Lindner, B., Ziesch, N., Rex-Haffner, M., **Roeh, S.**, Viturawong, T., Lehm, M. & others. The atherosclerosis risk

variant rs2107595 mediates allele-specific transcriptional regulation of HDAC9 via E2F3 and Rb1. *Stroke* 50, 2651--2660 (2019).

Roeh, S., Wiechmann, T., Sauer, S., Kodol, M., Binder, E. B. & Provencal, N. HAM-TBS: high-accuracy methylation measurements via targeted bisulfite sequencing. *Epigenetics & chromatin* 11, 1--10 (2018).

Schroeder, M., Jakovcevski, M., Polacheck, T., Drori, Y., Luoni, A., **Roeh, S.**, Zaugg, J., Ben-Dor, S., Albrecht, C. & Chen, A. Placental miR-340 mediates vulnerability to activity based anorexia in mice. *Nature communications* 9, 1--15 (2018).

Schroeder, M., Jakovcevski, M., Polacheck, T., Drori, Y., Ben-Dor, S., **Roeh, S.** & Chen, A. Sex dependent impact of gestational stress on predisposition to eating disorders and metabolic disease. *Molecular metabolism* 17, 1--16 (2018).

Engel, M., Eggert, C., Kaplick, P. M., Eder, M., **Roeh, S.**, Tietze, L., Namendorf, C., Arloth, J., Weber, P., Rex-Haffner, M. & others. The role of m6A/m-RNA methylation in stress response regulation. *Neuron* 99, 389--403 (2018).

Roeh, S., Weber, P., Rex-Haffner, M., Deussing, J. M., Binder, E. B. & Jakovcevski, M. Sequencing on the SOLiD 5500xl System--in-depth characterization of the GC bias. *Nucleus* 8, 370--380 (2017).

Grunewald, L., Landaas, E. T., Geissler, J., Weber, H., Quast, C., **Roeh, S.**, Schartner, C., Lesch, K.-P., Romanos, M., Kittel-Schneider, S. & others. Functional impact of an ADHD-associated DIRAS2 promoter polymorphism. *Neuropsychopharmacology* 41, 3025--3031 (2016).

Zannas, A. S., Arloth, J., Carrillo-Roa, T., Iurato, S., **Roeh, S.**, Ressler, K. J., Nemeroff, C. B., Smith, A. K., Bradley, B., Heim, C. & others. Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling. *Genome biology* 16, 1--12 (2015).

Arloth, J., Bader, D. M., **Roeh, S.** & Altmann, A. Re-Annotator: Annotation pipeline for microarray probe sequences. *PLoS One* 10, e0139516 (2015).

Abbreviations

A	adenine
ACTH	adrenocorticotrophic hormones
C	cytosine nucleotide
CpG	cytosine nucleotide followed by guanine nucleotide
GC	guanine cytosine nucleotides
ChIP	chromatin immunoprecipitation
CRF	corticotrophin releasing factor
DEX	dexamethasone
DNA	deoxyribonucleic acid
ELS	early life stress
G	guanine nucleotide
GR	glucocorticoid receptor
GRE	glucocorticoid response element
GWAS	genome-wide association studies
GxE	gene by environment
HAM-TBS	high-accuracy methylation measurements via targeted bisulfite sequencing
HDAC	histone lysine deacetylases
HPA	hypothalamic-pituitary-adrenals
MeDIP	methylated DNA immunoprecipitation
mRNA	messenger RNA
NGS	next generation sequencing
PCR	polymerase chain reaction
PTSD	post-traumatic stress disorder
PVN	paraventricular nucleus
RNA	ribonucleic acid
SNP	single nucleotide polymorphism
T	thymine nucleotide
TAD	topologicaaly associating domain
TBS	targeted bisulfite sequencing
U	uracil nucleotide
WGBS	whole genome bisulfite sequencing
5mC	5-methylcytosine

Declaration of contribution

Publication I

S. Röh, Weber, P., Rex-Haffner, M., Deussing, J. M., Binder, E. B., & Jakovcevski, M. (2017). Sequencing on the SOLiD 5500xl System - in-depth characterization of the GC bias. *Nucleus*, 8(4), 370--380.

SR, PW, JD, EB and JM contributed to experimental design. MRH, MJ performed wet lab work. **SR** performed the data analyses. **SR** and MJ prepared the manuscript.

Publication II

S. Röh*, T. Wiechmann*, S. Sauer, M. Kodel, E. B. Binder and N. Provencal (2018). "HAM-TBS: high-accuracy methylation measurements via targeted bisulfite sequencing." *Epigenetics Chromatin* 11(1): 39.

TW, **SR**, NP, EBB contributed to experimental design. TW, SS, MK performed wet lab work. SR performed the data analyses. **SR**, TW, NP, EBB prepared the manuscript.

Publication III

T. Wiechmann, **S. Röh**, S. Sauer, D. Czamara, J. Arloth, M. Ködel, M. Beintner, L. Knop, A. Menke, E. B. Binder and N. Provencal (2019) Identification of dynamic glucocorticoid- induced methylation changes at the FKBP5 locus. *Clinical Epigenetics* 11:83.

TW, SS, MK, MB performed wet lab work. **SR** performed sequencing data processing. JA performed processing of 450K data. TW, DC and NP analyzed and illustrated the data. MB, LK and AM collected and processed study samples. TW, NP, EBB prepared the manuscript. NP, EBB conceptualized and supervised the study.

Summary

Stress-related disorders have complex etiologies and high incidences, with mood disorders expected to become the leading cause of disability in the coming years. The mechanisms leading to resilience or susceptibility are not well understood and involve both genetic and environmental aspects, especially exposure to stress and adversity, integrated in the form of epigenetic changes such as DNA methylation or histone modification. In recent years, next generation sequencing (NGS) has become a powerful tool for epigenetic profiling. This thesis aims to improve on existing, as well as designing new NGS methods with increased sensitivity and accuracy for improved detection of small epigenetic changes even in composite tissues such as blood. We apply these methods to study DNA methylation changes in the context of stress by genotype interaction within the FKBP5 gene locus.

In the scope of minimizing the bias and maximizing the sensitivity of NGS technology, the first part of this work represents an in-depth characterization and reduction of GC bias on the SOLiD sequencers relative to the currently dominant Illumina sequencing technology.

Subsequently, we present a new method, termed HAM-TBS, which enables highly sensitive assessment of methylation levels of a target set of CpG sites in studies with cohort-level sample sizes. Using HAM-TBS, we can tailor the set of CpGs to fit the biological question at hand. To our knowledge, this is the most sensitive method to detect methylation levels in NGS data to date allowing us to resolve small changes. With it, we supply a fully tested panel of amplicons targeting selected CpGs in functionally relevant regions within the FKBP5 locus including glucocorticoid receptor elements, CTCF binding sites, topologically associating domain boundaries and the transcription start site.

Finally, we apply HAM-TBS using the established amplicon panel to assess dynamic methylation changes in the FKBP5 gene locus following a stress challenge. We expose a cohort of healthy individuals to dexamethasone (DEX), a synthetic glucocorticoid activating the stress reactive hypothalamic-pituitary-adrenal axis. We found dynamic and lasting methylation changes in blood as well as a genotype-dependent response following DEX exposure.

This work demonstrates the potential of NGS technologies to aid in improving our understanding of epigenetics in the context of the stress response system.

Introduction

Epigenetics

The human genome comprises roughly 3 billion base pairs of deoxyribonucleic acid (DNA) harboring estimated 30.000 genes forming an average of 3 proteins each (“NIH National Human Genome Research Institute,” n.d.; Piovesan et al., 2019). Almost every cell in our bodies is equipped with the same genetic blueprint, yet presents with vastly different morphologies and functions. This is enabled by tailored differences in gene activity and expression modulated by epigenetic mechanisms. The term “*epigenetics*” was first introduced in the 1940s by developmental biologist C. H. Waddington as “the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being” (Goldberg, Allis, & Bernstein, 2007; Waddington, 2012).

Epigenetic modifications to the DNA can influence gene expression and the phenotype without altering the underlying genetic sequence. Such dynamic epigenetic regulatory mechanisms involve, among others, DNA methylation and post-translational modification of histones by e.g. methylation or acetylation. However, it is important to note that multiple epigenetic mechanisms can often act simultaneously in a coordinated fashion (Moore, Le, & Fan, 2013).

Histone modification

In the cell, DNA is packaged in chromosomes made of chromatin (Figure 1). Chromatin refers to the complex of nucleosomal subunits, each consisting of a histone octamer core with 147bp of DNA wrapped around it (Luger, Mäder, Richmond, Sargent, & Richmond, 1997). A histone octamer consists of 4 different histone proteins, namely H2A, H2B, H3 and H4 each represented by 2 copies. Chromatin can be more or less condensed with the extreme states named euchromatin (open) and heterochromatin (closed). By regulating access to the packaged DNA the present state of chromatin can repress or permit the transcription of the respective genes. Hence, histone modification provides a mechanism of regulation that does not require a change to the underlying sequence. Post-translational histone modifications occur at specific amino acid residues on the N-terminal tails and are often transient and reversible. They comprise – among others – acetylation and methylation primarily appearing on specific lysines of H3 and H4 (Renthal & Nestler, 2008). Most of the specific

modifications are either associated with open (active transcription) or closed (silenced transcription) chromatin (Kouzarides, 2007). Of note, while histone acetylation is usually connected to open chromatin, the majority of histone methylation marks are associated with repression of transcription. The enzymatic machinery to modify specific marks is provided by histone lysine deacetylases (HDAC and sirtuines), acetyltransferases, methyltransferases and lastly demethylases.

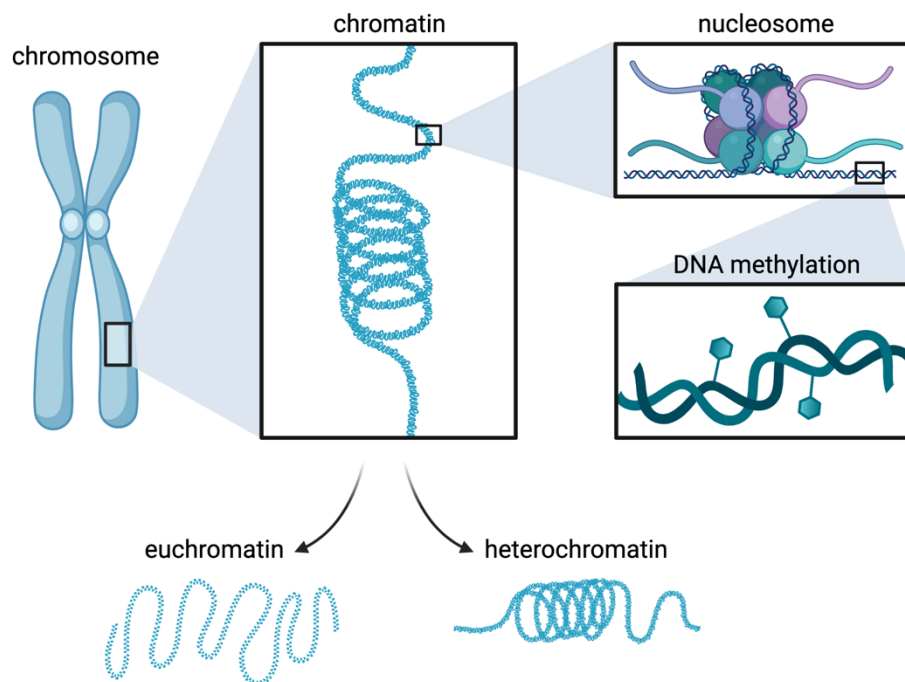


Figure 1 – Visualization of genomic organizational makeup. Created with BioRender.com

DNA methylation

DNA methylation is the covalent addition of a methyl group to a cytosine nucleotide referred to as 5-methylcytosine (5mC) (Figure 1). This is a chemically stable mark occurring almost exclusively in cytosine-guanine (CpG) context (Lister et al., 2009) with 70-80% in methylated state (Bird, 2002). Roughly 28 million CpG sites are present in the human genome (Babenko, Chadaeva, & Orlov, 2017) mostly randomly distributed with less than 10% located in clusters of high CpG density called CpG islands (Smith & Meissner, 2013). They are associated with approximately 70% of proximal promoters and predominantly uniformly unmethylated facilitating transcription. Moreover, even CpG islands not located in proximity to known promoters are associated with transcriptional activation. Silencing their strong promoter activity requires dense CpG methylation or

polycomb recruitment (Deaton & Bird, 2011). However, methylation levels are higher within the gene body and in intergenic regions outside of CpG islands.

5mC has a long been exclusively associated with transcriptional repression due to its role in processes like genomic imprinting and X chromosome inactivation in females (Bird, 2002; Edwards, Yarychivska, Boulard, & Bestor, 2017). The inhibitory function of DNA methylation on gene expression is facilitated by two main mechanisms. First, DNA methylation can simply obstruct sequence recognition and binding of transcription factors to the DNA and therefore hinder their enhancer activity. And second, 5mC is recognized by both methyl-CpG-binding proteins that draw on co-repressor molecules like histone deacetylases to modify the surrounding chromatin state as well as some zinc-finger domain proteins binding methylated CpG sites and thereby silencing transcription (Klose & Bird, 2006; Moore et al., 2013). However, in some instances the opposite, a transcription enhancing effect of DNA methylation, has also been observed and emphasizes the complex nature of epigenetic modifications (Ball et al., 2009; Harris et al., 2018).

Epigenetic regulation and the brain

Epigenetics enable cell type diversity, however the importance of a plastic epigenome extends far beyond that. Throughout life, we are exposed to a plethora of environmental factors. Among these are our lifestyle choices, exercise, diet, and socioeconomic status. A lifetime of positive, negative or even traumatic experiences leave a footprint on the epigenome. Epigenetic mechanisms supply a tool to our organism to adapt by dynamically encoding and integrating information from these environmental exposures to facilitate changes in gene expression.

This is particularly important in the context of the brain. Neuroplasticity or brain plasticity refers to the brain's ability to change and adapt in response to changes in the environment or the body. It can form new neural pathways, modify existing connections, and even reorganize its structure. Brain plasticity is at its highest during childhood, when the brain is still developing, and decreases with age. However, the brain remains capable of plasticity throughout adulthood, which is essential for normal functioning including learning and memory as well as the regulation of mood, emotion, and behavior (O'Donnell & Meaney, 2020). Much of this plasticity is made possible by epigenetics.

A further complication is the interaction of epigenetic processes and genomic status. This is enabled by single nucleotide polymorphisms (SNPs) where the substitution of a single nucleotide is giving rise to more than one phenotype. Depending on an individual's genotype, the possibility for epigenetic encoding following an environmental cue may differ as e.g. CpG sites may be disrupted or transcription factor binding sites modified. Therefore, the same environmental cue may have a different epigenetic manifestation and have different consequences for gene regulation. This interplay between different genotypes affecting gene expression and the environment is termed gene by environment (GxE) interaction.

Environmental factors such as exposure to stressful events early in life interact with genetic factors to shape the risk for developing stress-related psychiatric disorders (Caspi, Hariri, Holmes, Uher, & Moffitt, 2010; Caspi & Moffitt, 2006; Heim et al., 2009; Xie et al., 2010). However, it is not well understood how such differences in the psychiatric risk profile are reflected in individual physiology. A relevant point of focus is the hypothalamic–pituitary–adrenal (HPA) axis - an essential part of the body's neuroendocrine response to stress.

Hypothalamic–pituitary–adrenal axis

Our dynamic environments create the need for continuous adaptation throughout life. Environmental perturbations may be thought of as stressors and the reaction to them as the organism's stress response – a series of biological reactions with the aim to restore homeostasis.

To this end, our bodies draw on a set of complex physical, emotional and behavioral mechanisms. At the core of the stress response system sits the hypothalamic–pituitary–adrenal axis, a neuroendocrine system known as the *HPA axis* (Figure 2).

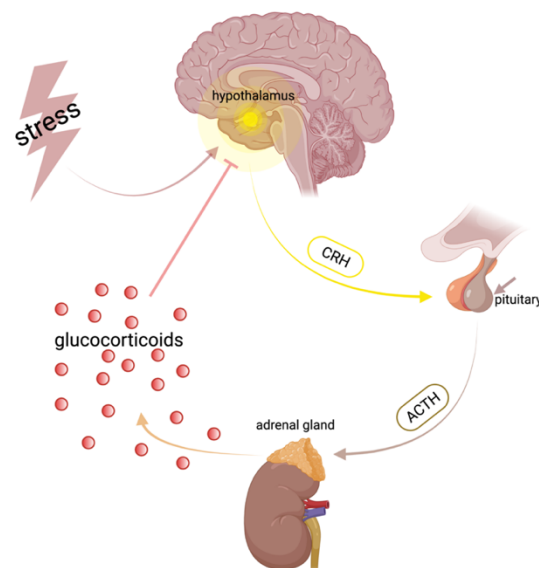


Figure 2 - Schematic depiction of the HPA axis. Created with BioRender.com

In brief, the HPA axis is a physiological cascade with broad involvement in e.g. circadian rhythm, exercise and stress. It is activated when neurons of the paraventricular nucleus (PVN) of the hypothalamus secrete the neuropeptide corticotrophin releasing factor (CRF). This stimulates the anterior pituitary to release adrenocorticotrophic hormones (ACTH) reaching the adrenal cortex via the blood stream. The resulting secretion of glucocorticoids (cortisol in humans and corticosterone in rodents) provides negative feedback to the HPA axis via both the glucocorticoid receptor (GR) and mineralocorticoid receptor (MR) expressed in the brain and peripheral tissues (Dick & Provencal, 2018). For example, neurons in the hippocampal formation can act on CRF-positive cells in the PVN to reduce overall HPA tone (Cole, Montgomery, Bale, & Thompson, 2022).

The negative feedback property of the HPA axis via the GR is crucial for a healthy stress response. The underlying molecular mechanism (Figure 3) involves, among others, FKBP51 encoded by the *FKBP5* gene. It is a co-chaperone that binds to the GR complex, inhibiting GR translocation from the cytoplasm to the nucleus. In the presence of cortisol, FKBP51 is exchanged for FKBP52, and together with other co-chaperones facilitates GR translocation. This enables GR to directly repress or enhance transcription, modulate transcription by interaction with other transcription factors and importantly, inhibit the HPA axis (Zannas, Wiechmann, Gassen, & Binder, 2016). *FKBP5* expression is regulated by GR directly, forming an ultra-short negative feedback loop in addition to changes in FKBP51 levels impacting several other molecular pathways (e.g., NF- κ B signaling and autophagy).

In the context of the stress-response system, GxE interactions have been identified for some key players of the HPA axis as *CRH* and *CRHR1*, but foremost and importantly, the *FKBP5* gene (Binder et al., 2008; Buttenschøn et al., 2017; Gerritsen et al., 2017; Z. Liu et al., 2013; Zimmermann et al., 2011).

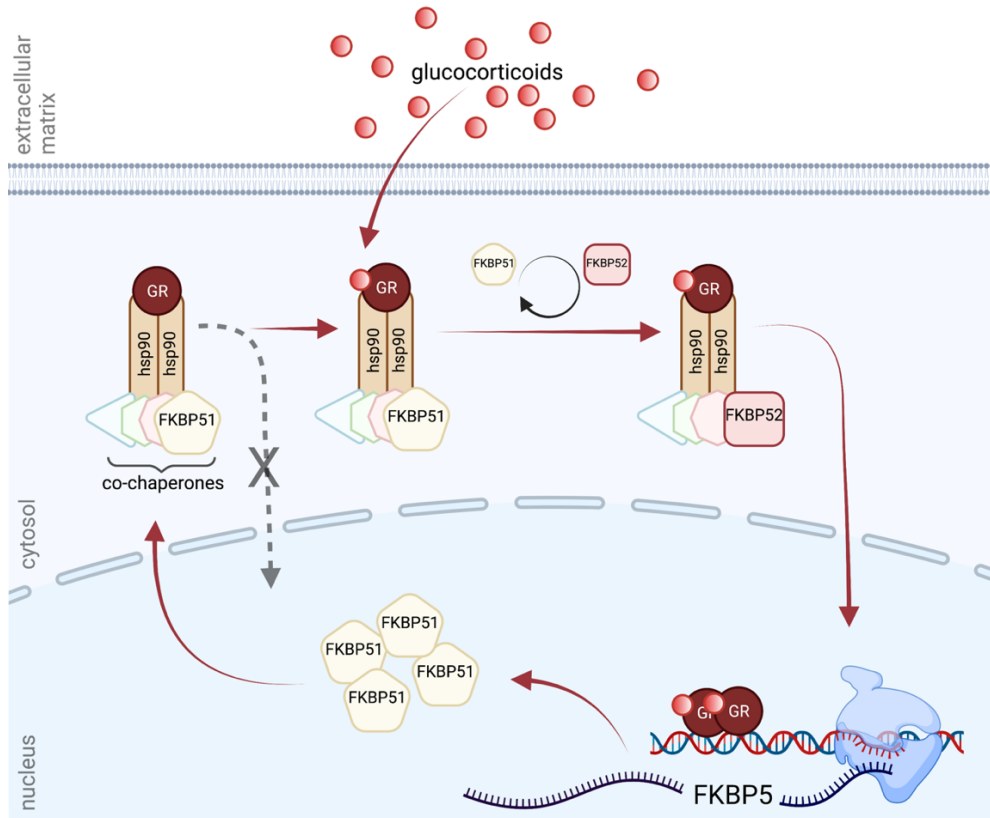


Figure 3 - Ultrashort negative feedback loop. FKBP51 bound to the glucocorticoid receptor complex reduces glucocorticoid binding to the GR and inhibits translocation to the nucleus. Upon glucocorticoid binding FKBP51 is exchanged for FKBP52 (Misiak et al., 2020) facilitating translocation. In the nucleus, among other functions, GR acts as transcription factor activating the transcription of FKBP5. Higher amounts of FKBP5 are translated and can occupy the glucocorticoid complex. Translocation is delayed.

Created with BioRender.com

FKBP5

FK506-binding proteins (FKBPs) are a family of immunophilins found in many organisms, in humans, at least 15 FKBP5s have been identified thus far (Rulten et al., 2006; Somarelli, Lee, Skolnick, & Herrera, 2008). Among them is FKBP51, encoded by the *FKBP5* gene, an approximately 180 kilobase pair long region on chromosome 6 with a total of 13 exons. It is expressed in many tissues of the human body with moderate expression levels across brain regions (Figure). While *FKBP5* plays a role in association with some somatic diseases as e.g. cancer (Romano et al., 2013), here, the focus will be on the involvement in the field of psychiatry.

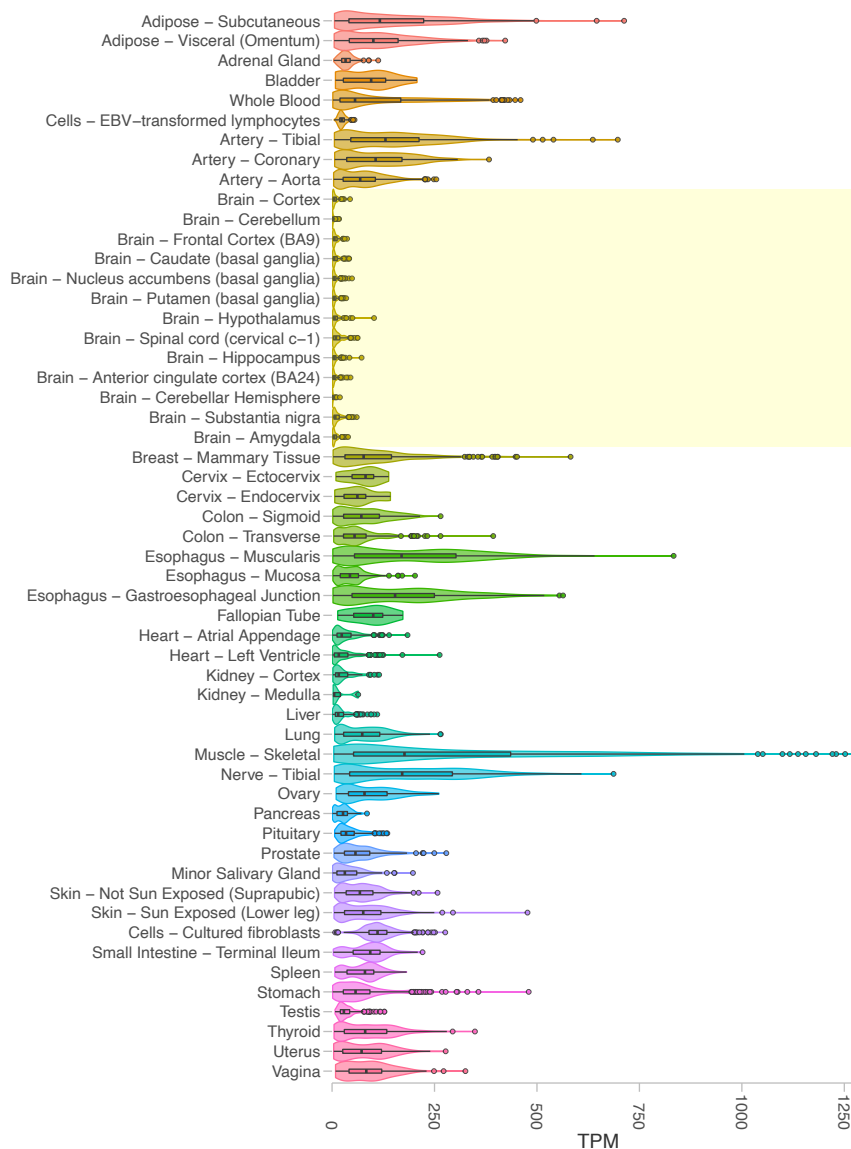


Figure 4 - GTEX (release V8) FKBP5 expression data across tissues. Brain tissue highlight in yellow.

FKBP5 regulation

FKBP5 expression levels are tightly coupled to glucocorticoid signaling and exposure in humans.

Firstly, FKBP5 is directly regulated by the GR in its capacity as transcription factor via glucocorticoid response elements (GREs). These elements occur genome-wide, are 15 base pairs long and allow for direct binding of homo- or heterodimers of GR to the DNA (Mifsud & Reul, 2016). *FKBP5* harbors GREs located both in distal intronic regions (2, 5 and 7) as well as upstream of the promotor (Paakinaho, Makkonen, Jääskeläinen, & Palvimo, 2010). Upon GR binding to the GREs, the DNA forms a three-dimensional loop structure with the core promotor site and directly influences *FKBP5* transcription. The interplay of *FKBP5*/FKBP51 and the GR is bi-directional and forms a ultra-short negative feedback loop. In turn to GR promoting the expression of *FKBP5*, FKBP51 presence and binding to the GR complex inhibits its translocation to the nucleus. This mechanism facilitates the termination of the stress response and therefore is crucial to the regulation of the HPA axis.

In addition, a haplotype comprising up to 18 SNPs spans the entire FKBP5 locus. It is often tagged by the functional SNP rs1360780 located close to a GRE in intron 2 of the gene body. This variant has been investigated in several studies and found to be associated with elevated *FKBP5* expression levels, higher risk for stress-related psychiatric disorders and faster response to antidepressant treatment (Binder et al., 2008, 2004; Zannas et al., 2016). The SNP presents with 2 alleles, the T and the C allele with 35% of people carrying the risk associated T allele (Siva, 2008). In deciphering the molecular mechanism, Klengel and colleagues showed that the regulatory three-dimensional loop structure is influenced by this polymorphism with the T allele better supporting a loop structure connecting the GRE and the TSS and facilitating the expression of FKBP5 as opposed to the C allele.

Lastly, a SNP can moderate the occurrence of epigenetic marks in its proximity. This has been shown for rs1360780 with regard to the interaction with early life stress on adult PTSD (Binder et al., 2008; Klengel et al., 2013).

FKBP5 in psychopathology

Studies consistently indicate a connection between FKBP5 genotype and aberrant expression levels in human stress-related psychopathology (Binder et al., 2008, 2004; Koenen et al., 2005; Matosin, Halldorsdottir, & Binder, 2018). In addition, preclinical work in mice has associated FKBP5 expression with HPA reactivity (Häusl et al., 2021), stress coping behavior, and anxiety-like behavior, especially in the amygdala and hippocampus (Zannas et al., 2016). However,

even in healthy individuals FKBP5 levels are not constant. Expression is regulated during development and consistently increases throughout life (Blair et al., 2013; Matosin et al., 2021; Weickert, Webster, Boerrigter, & Sinclair, 2016). Altered expression trajectories across the lifespan are not well understood and it is conceivable that multiple genetic and epigenetic factors accumulate in some individuals. Given the involvement of FKBP5 in HPA axis regulation, the gene is a likely medium for converting adverse life events into increased susceptibility to stress-related disorders later on. In fact, studies on early life stress have shown FKBP5 haplotype-dependent (rs1360780 functional polymorphism) interactions of ELS on adult PTSD, suicide, and major depressive disorder (Appel et al., 2011; Binder et al., 2008; Koenen et al., 2005; Roy, Gorodetsky, Yuan, Goldman, & Enoch, 2010). The presence of a vulnerable phase suggest an epigenetic regulatory mechanism which (Klengel et al., 2013) interrogated. In line with studies showing long-lasting DNA methylation changes following early life stress (McGowan et al., 2009; Murgatroyd et al., 2009), childhood trauma-induced demethylation was found in one GRE of FKBP5 specific to carriers of the risk allele regarding rs1360780.

This body of work leaves many open questions regarding the role of dynamic methylation at the FKBP5 gene and its dependence on haplotype and stress.

Next generation sequencing

The evolution of NGS

The structure of deoxyribonucleic acid (DNA) was famously discovered in the 1950's, less than 70 years ago, by the physical chemist Rosalind Franklin and the molecular biologists Watson and Crick (Watson & Crick, 1953). This laid grounds for a burst of technological developments enabling great advancements in the understanding of our genetic makeup and the complex involvement of genetics in health and disease.

The act of reading the genetic code, the sequence of bases it is made up of, A (adenine), C (cytosine), G (guanine) and T (thymine), is known as sequencing. First methods were introduced in the 1970's (Maxam & Gilbert, 1977; Sanger et al., 1977) incorporating different approaches allowing for single base pair (bp) resolution of 100-1000 bp sequences by displaying the information via gel electrophoresis. Among these methods is Sanger sequencing, a rather laborious method that led to the first fully sequenced genome belonging to a bacteriophage at a length of ~5000 nucleotides. The next big step toward higher scalability and the sequencing of larger genomes was taken with the development of the shotgun sequencing approach where a longer sequence is fragmented, sequenced, and reconstructed from the resulting reads. The combination of Sanger and shotgun-sequencing was sufficient to initiate and carry out the next big milestone - the Human Genome Project. The undertaking started in 1990 as an international research project with the aim of assembling the full sequence of the human genome within 15 years funded with 3 billion dollars. The project was completed after 13 years, delivering 3.1 giga bp of assembled genomic sequence. Since then, a significant amount of development has gone into sequencing technology, for instance enabling the 1000 genomes project (Siva, 2008) that was started in 2008. The aim was to build an extensive catalogue of genetic variation by sequencing over 1000 genomes from donors around the world. By then, the throughput had increased and cost per genome had dropped to approximately 100,000 dollars ("NIH National Human Genome Research Institute," n.d.). Today, it is possible to sequence a complete genome at under

1.000 dollars, opening the door to a completely different scale of data generation and knowledge discovery.

The technology in NGS

To date, several approaches have been developed to perform sequencing. Two aspects are particularly important to measure and compare by. First, the read length that can be sequenced. The majority of NGS data produced today still uses the shotgun sequencing technique and has a rather short read length of around 100 - 300 bp. While for many applications this is sufficient, it can be of great advantage to produce longer reads especially in regard to e.g. novel genome assembly or gene isoform detection. Second, the production of multiple sequences in parallel, increasing sequencing depth. While for smaller genomes, e.g. microbial genomes, lower read numbers can be sufficient, when working with genomes of the size of the human or mouse genome at 3.1 and 2.7 giga bp respectively, a higher throughput may be needed in order to achieve sufficient coverage. Both aspects of length and throughput are still hard to combine and machines are typically forced to specialize in one over the other. While long-read approaches like nanopore sequencing by Oxford Nanopore Technologies or Single-Molecule Real-Time Sequencing by Pacific Biosciences do exist and may play the bigger role in the future, the largest research market share today is held by short-read sequencing technologies, which are also the methods used in the work described in the following.

Illumina sequencing

Over time, Illumina established themselves as the leader in the field of sequencing technology. This is in part attributed to the fact that their machines produce the highest yield, meaning the most bp and reads in the shortest time, which led to the high number of sequencers in circulation. For instance, their newest machine is able to produce up to 3000 Gb in ~44 hours. This is more than 80 times the amount of data produced by the Human Genome Project over a time period of 13 years. This level of throughput can be achieved by high parallelization as Illumina machines produce millions to billions of reads simultaneously. Briefly, there are some important steps in the process of constructing an Illumina sequencing library and ultimately sequencing it on one of their machines (Figure 5). First, still following the principle of shotgun sequencing, Illumina requires the input material of interest to be provided in small fragments usually around 200-300 bp. Second, each fragment is equipped with adapter sequences providing

compatibility with the Illumina technology for sequencing. Parallelization is achieved by binding millions of fragments on a solid surface called a flow cell. Here, the adapter sequences find complementary oligos to bind to on the surface. Next, the bridge polymerase chain reaction (PCR) is used to locally amplify each fragment to form small clusters. This is used to enhance the fluorescent signal each fragment is emitting during the subsequent sequencing. Illumina uses a sequencing-by-synthesis approach, performing a PCR incorporating nucleotides with base-specific fluorescent markers and subsequent imaging during each cycle. The sequence of each read can be deciphered by the colors emitted by the clusters for all cluster / fragments in parallel.

Sequencing by Oligonucleotide Ligation and Detection (SOLiD) sequencing

The main competitor to Illumina's high-throughput short-read sequencing technologies was developed by Life Technologies in 2006 and eventually discontinued in 2016. The SOLiD 5500xl sequencer was able to produce up to 45 Gb of short-reads a day, and has contributed a substantial amount of data to science since its development. The workflows of SOLiD and Illumina sequencing differ in some noteworthy points (Figure 5), though both approaches achieve parallelization by incorporation of the shotgun sequencing and subsequently adding adapter sequences for amplification and machine compatibility. Complementary to the bridge PCR, the SOLiD system implements a PCR performed in an emulsion where each droplet ideally contains one bead and one fragment which is then amplified on the bead surface. These beads are then transferred to a flow cell and ready for sequencing. Hence, each bead corresponds to one Illumina generated cluster. SOLiD sequencing relies on sequencing-by-ligation, with a 2-nucleotide encoding system.

It is of high value to understand the effects the separate steps have on the data generation as well as the individual strengths and weaknesses of the approaches. This aids to advance the development of sequencing technology as well as to choose the best suited environment to run an experiment and interpreting the generated data in this regard. While some work is published on the assessment of specific sequencing machines and their performance, this is usually restricted to Illumina. To aid in this task, in one arm of this thesis the

performance of the SOLiD is evaluated and compared to the Illumina sequencing approach with a focus on the GC bias.

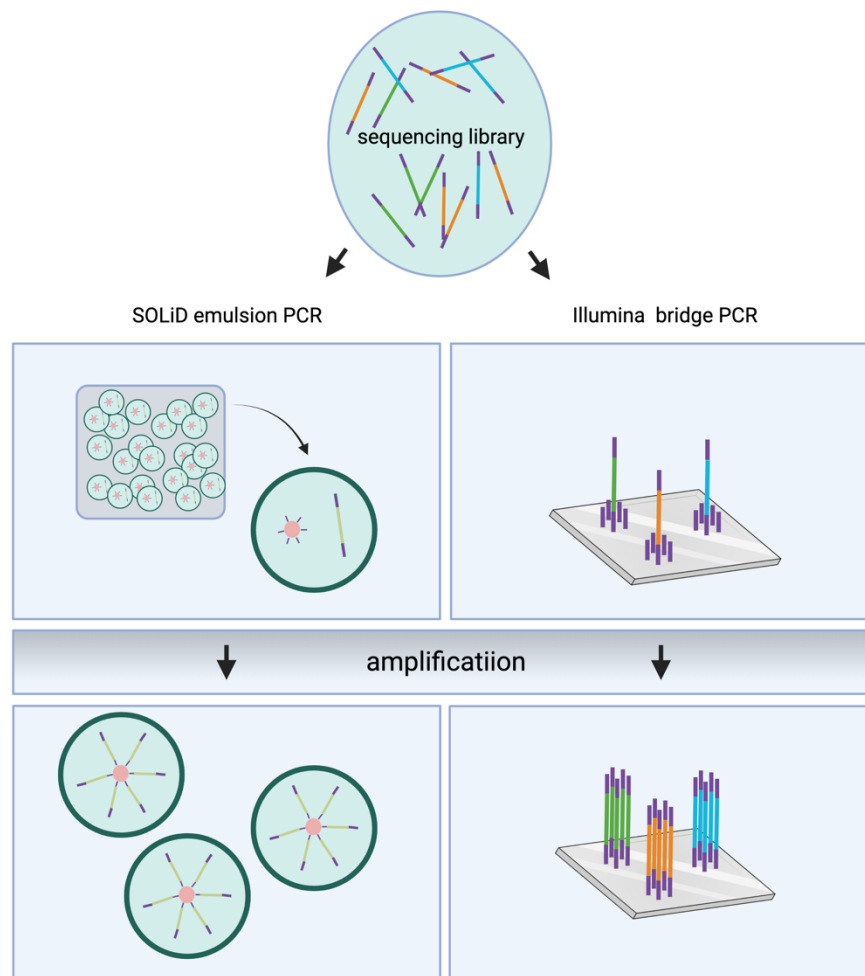


Figure 2 - Side by side depiction of amplification step for both the Illumina and Life Technologies sequencing approaches. Created with BioRender.com

The applications of NGS

NGS is a versatile and scalable set of methods with countless applications. It is rapidly developing and can be flexibly tailored to a wide variety of scientific questions. This is owed to the fact that practically any sufficiently numbered set of short fragments of DNA or ribonucleic acid (RNA) can be turned into a library ready for sequencing. In consequence, among numerous other things, NGS can be used to interrogate the expression of genes (RNA sequencing) as much as binding sites of proteins to DNA (chromatin immunoprecipitation (ChIP) sequencing), methylation states (methylated DNA immunoprecipitation

sequencing, bisulfite sequencing), microRNA expression, whole genome resequencing and targeted amplicon-based approaches.

There are several potential ways to structure or categorize applications. For instance, the vast majority can be divided in either count-based or sequence-based. Count-based applications describe approaches where the information is contained on a per-read basis as e.g. ChIP sequencing or RNA sequencing. More specifically, e.g. in RNA sequencing, reads originate from the messenger RNA of genes and expression levels are quantified by analyzing the number of reads produced by each detected gene. Sequencing depth for this kind of application is counted in the number of reads produced by the sequencer. In contrast, sequence-based applications are focused on the actual base content and can benefit from longer reads as each individual base may hold information. Sequencing depth can be measured in bases for these applications. Broadly used applications of the latter type are genome resequencing, exome resequencing, genome assembly or whole genome bisulfite sequencing (WGBS). In the context of this thesis, some applications will be discussed in more detail.

ChIP sequencing

Briefly, this is a method to determine the interaction sites between proteins and DNA, e.g. the binding sites of a transcription factor or the location or presence of a specific histone modification. In most but not all cases, these protein-DNA interactions are first preserved by fixation with formaldehyde. While at this point part of the input sample is commonly kept as the input control, the immunoprecipitation is performed with a specific antibody targeting a protein of interest and isolating the attached DNA to it. This DNA and the input control are eventually converted into a DNA library and sequenced. Lastly, the sequencing reads can be mapped back to the reference genome and by comparison of the enriched sites of IP sample over the input control give indication on the genomic position that the protein of interest was bound to. Noteworthy but in the context of this thesis not discussed in more detail, antibodies exist that can recognize DNA methylation. An IP performed in this way is a special case of ChIP sequencing termed methylated DNA immunoprecipitation (MeDIP) sequencing and allows to estimate genome wide methylation profiles.

Bisulfite sequencing

There are multiple methods to assess methylation levels both genome wide and/or at single base resolution. A long-standing low throughput approach is pyrosequencing, allowing for quantification of any CpG site of interest. However, the current genome-wide gold standard is the Illumina Infinium MethylationEPIC array (EPIC array) which is able to quantify the methylation of ~850,000 CpGs

(Zhou, Laird, & Shen, 2017) of the roughly 28 million CpGs present in the human genome (Lövkvist, Dodd, Sneppen, & Haerter, 2016). While the EPIC array-based method has high accuracy, it is restricted by a preselected set of target CpG sites that may not be sufficient depending on the scientific question at hand. NGS offers alternative approaches to measure methylation profiles and CpGs genome wide or of a individually selected set via amplicon approaches.

To measure methylation via the EPIC array and NGS alike, when offering single base resolution an initial conversion step is used. Briefly, the DNA is treated with bisulfite, leading to a conversion of cytosines to uracil in unmethylated state – methylated cytosines (both 5mC and 5hmC) are protected from this conversion. The EPIC array provides two alternative oligo probes depending on the methylation state for each CpG site and extracts the ratio of binding by emitted light signal reflecting the percentage of methylation. In NGS, the process is significantly different. A converted uracil will be read as a thymine and when compared to the reference genome will present as a mutation in contrast to the known cytosine. Hence, the methylation ratio is inferred by comparing the number of mutated thymines versus cytosines at the location of the CpG. Therefore, the readout in NGS is not directly a ratio it needs to be calculated from the individual reads. Importantly, the number of reads underlying the quantification of each site influences the resolution and accuracy at which methylation can be measured. A site that was covered by e.g. 4 reads indicating methylated and 1 read indicating unmethylated state would result in a readout of 80% methylation level but the resolution by only 5 reads is very low. A site quantified by e.g. 1000 reads showing 80% methylation has to be evaluated accordingly as resolution and accuracy here are much higher. In short, the readout in NGS consist of both the methylation ratio as well as the coverage that was used to infer it. This aspect is regarded during the processing and analysis of bisulfite converted NGS data.

The choice of technology largely depends on purpose and cost. While WGBS at high coverage would be ideal and useful in a broad variety of studies, it is remains cost intensive – even more so when applied to large sample sizes – and therefore rarely realizable. For instance, the cost of running an EPIC array versus a shallow WGBS at an average coverage of 10 is still more than 3 fold (price can vary per lab / conditions). Currently, WGBS approaches can often require coverages around 5x to 15x or higher depending on the aim (Ziller, Hansen, Meissner, & Aryee, 2015). To increase robustness of the methylation assessment, in WGBS the differential analysis is often spanned across regions classified as differentially methylated regions (DMRs) at the expense of single CpG resolution. This method can detect larger changes in methylation it lacks the resolution to reliably detect minor changes of methylation in the range of only few

percent (~1-5%). Importantly, this can be a crucial aspect when working with cell type specific effects measured in bulk tissue. This is of particular relevance to psychiatric research where, due to the limitation in access to brain tissue, blood is often used as proxy.

Another recent advancement in the analysis of cell type specific effects are single-cell sequencing technologies. Most commonly they enable quantification of gene expression in single cells, but ATAC sequencing and DNA methylation sequencing are also promising applications. While this new technology is tremendously valuable and key to numerous successful approaches requiring cell type specificity (H. Liu et al., 2021), it has clear limitations in scalability (Nichols et al., 2022). Firstly, each sample can comprise thousands of cells and data acquisition needs to match this demand as each cell usually requires thousands reads. Hence, this method is very expensive per sample with respect to single-cell library preparation and sequencing demands, rendering a study design using cohort data largely impractical. In addition, these methods start from very little input material and rely on excessive amplification. This leads to rather low sensitivity of the per cell data, e.g. in RNA sequencing lowly expressed genes will not be detectable below a certain expression level. However, power is increased by the number of cells available per cell type which is advantageous when working with common cell types. In single-cell DNA methylation sequencing single CpG resolution is possible. However, since each cell only has 3 states per CpG site (both alleles methylated, both alleles unmethylated, or one allele methylated) sensitivity and number of covered CpGs depends even more on the number of cells per cell type. Hence, regional aggregation of the data is common.

Overall, amplicon approaches are indispensable for assessing CpGs at high resolution and in contexts not covered by the EPIC array. These approaches restrict data generation to a set of selected target regions <500 bp in size, hence, making it both feasible and affordable to sequence these regions at very high coverage (> 1000 reads per CpG) in larger sample sizes and allow for an in-depth analysis resolving methylation changes of only small magnitude.

Only a handful of approaches exist today that aid in the task to measure methylation in sites specifically selected to match the biological question at hand with high sensitivity (Bernstein, Kameswaran, Lay, Sheaffer, & Kaestner, 2015; Masser, Berg, & Freeman, 2013; Masser, Stanford, & Freeman, 2015). All perform bisulfite conversion, amplify the target regions via PCR followed by an NGS library preparation. However, several steps differ and can affect accuracy, e.g. the bisulfite conversion, throughput and the specifics of the implemented library preparation methods. In this thesis, we present a cost-effective, medium to high-throughput target approach optimized to achieve – to our knowledge - the highest

accuracy to measure methylation levels in NGS data available to date, termed HAM-TBS. With this we provide a valuable tool to be able to decipher methylation changes as low as 1%, a resolution necessary when resolving cell type specific effects from complex tissues like blood. In addition, we provides a highly useful and thoroughly tested amplicon panel targeting key regions within the FKBP5 gene. This gene is of high interest in field of psychiatric research today.

Results

Publication I

Sequencing on the SOLiD 5500xl System – in-depth characterization of the GC bias

Authors:

Simone Röh¹, Peter Weber¹, Monika Rex-Haffner¹, Jan M. Deussing², Elisabeth B. Binder¹, and Mira Jakovcevski²

¹Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany.

²Department of Stress Neurobiology and Neurogenetics, Max Planck Institute of Psychiatry, Munich, Germany

Publication details:

Published in NUCLEUS VOL. 8, NO. 4, 370–380 on July 7th, 2017

<https://doi.org/10.1080/19491034.2017.1320461>

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, reproduction in any medium, provided the original work is properly cited.

METHODS



Sequencing on the SOLiD 5500xl System – in-depth characterization of the GC bias

Simone Roeh^a, Peter Weber^a, Monika Rex-Haffner^a, Jan M. Deussing^b, Elisabeth B. Binder^{a,c}, and Mira Jakovcevski^{a,b}

^aDepartment of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany; ^bDepartment of Stress Neurobiology and Neurogenetics, Munich, Germany; ^cDepartment of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA

ABSTRACT

Different types of sequencing biases have been described and subsequently improved for a variety of sequencing systems, mostly focusing on the widely used Illumina systems. Similar studies are missing for the SOLiD 5500xl system, a sequencer which produced many data sets available to researchers today. Describing and understanding the bias is important to accurately interpret and integrate these published data in various ongoing research projects. We report a particularly strong GC bias for this sequencing system when analyzing a defined gDNA mix of 5 microbes with a wide range of different GC contents (20–72%) when comparing to the expected distribution and Illumina MiSeq data from the same DNA pool. Since we observed this bias already under PCR-free conditions, changing the PCR conditions during library preparation – a common strategy to handle bias in the Illumina system – was not relevant. Source of the bias appeared to be an uneven heat distribution during the SOLiD emulsion PCR (ePCR) – for enrichment of libraries prior loading – since ePCR in either small pouches or in 96-well plates improved the GC bias.

Sequencing of chromatin immunoprecipitated DNA (ChIP-seq) is a common approach in epigenetics. ChIP-seq of the mixed source histone mark H3K9ac (acetyl Histone H3 lysine 9), typically found on promoter regions and on gene bodies, including CpG islands, performed on a SOLiD 5500xl machine, resulted in major loss of reads at GC rich loci (GC content $\geq 62\%$), not explained by low sequencing depth. This was improved with adaptations of the ePCR.

ARTICLE HISTORY

Received 10 February 2017
Revised 7 April 2017
Accepted 12 April 2017

KEYWORDS


chromatin immunoprecipitation (ChIP); CpG island; emulsion polymerase chain reaction (ePCR); GC bias; H3K9ac; microbial genomic DNA; next generation sequencing (NGS); PCR-free library preparation; sequencing depth; upscale PCR

Introduction

Next Generation Sequencing (NGS) techniques have become increasingly popular methods, since they are powerful tools for genomics and epigenomics research.^{1,2} Several different NGS systems have emerged in parallel using various approaches of library generation and performing of the actual sequencing procedure. Besides the currently most popular Illumina sequencing technology, the SOLiD System (Applied Biosystems) has been among the most frequently used NGS platforms. Data generated by both approaches is a highly valuable source for meta-analyses of any type. However, the specific weaknesses and strengths of the sequencing technique used, have to be taken into account for meaningful interpretation of the data.

One of the common problems of NGS techniques is the under- or over-representation of GC or AT rich sequences.^{3–5} These biases are often generated during library preparation, mostly when libraries undergo an upscale polymerase chain reaction (PCR), while PCR-free libraries are believed to be sequenced with significantly less to almost no bias.^{6,7} However, often an upscale PCR cannot be avoided for many applications from single cell sequencing to ChIP-seq (sequencing of DNA obtained from chromatin immunoprecipitation), due to low amounts of starting material.^{3,8} Obviously, concerning each specific application, the weakness of a particular NGS technology might have important consequences on the quality of particular data sets. For example, among the most frequently used techniques in epigenetics is genome-wide ChIP-

CONTACT Mira Jakovcevski  mira_jakovcevski@psych.mpg.de  Max Planck Institute of Psychiatry, Kraepelinstr. 2-10, 80804 Munich, Germany.

 Supplemental data for this article can be accessed on the [publisher's website](#).

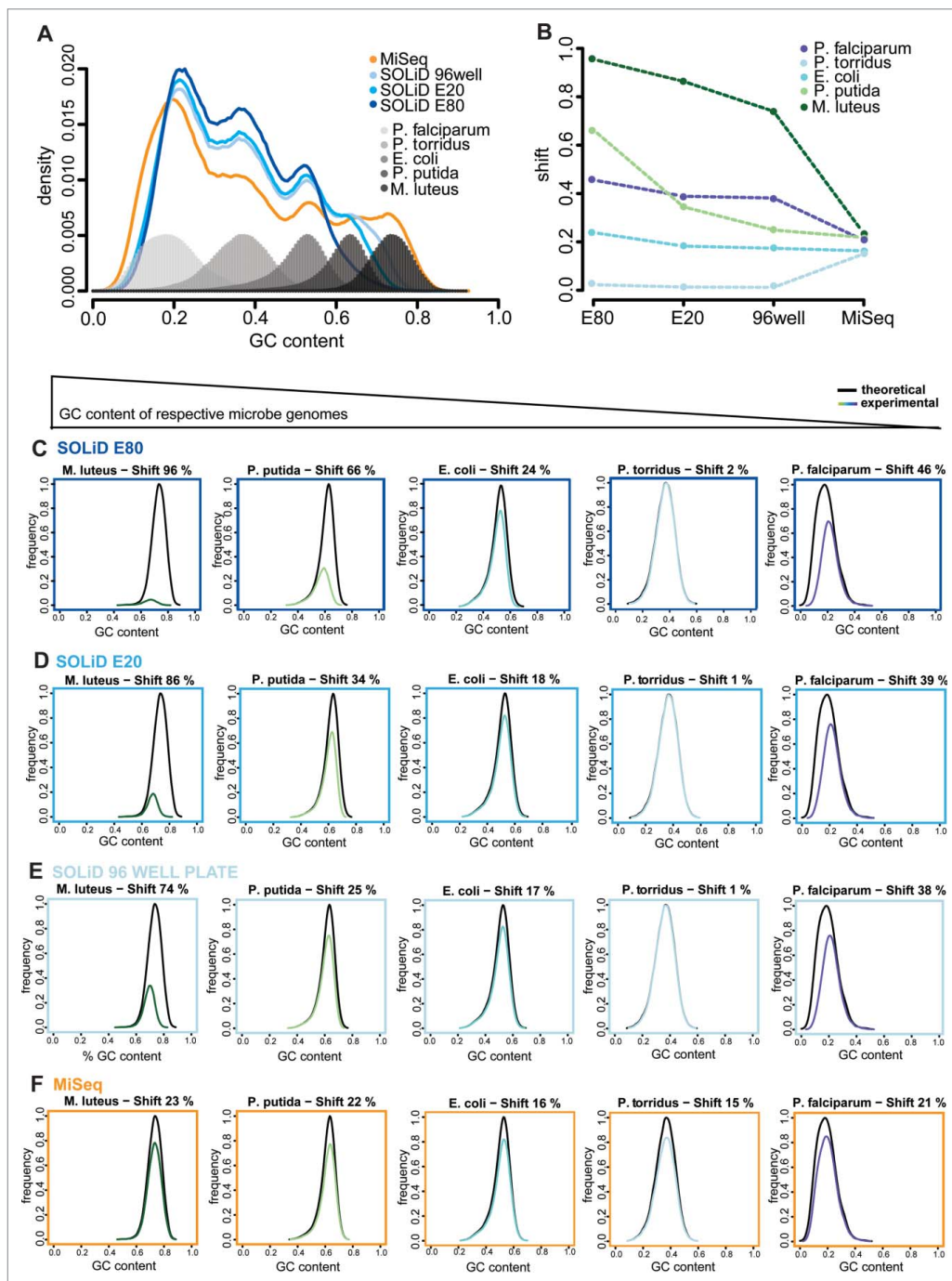


Figure 1. Characterization and improvement strategy of the GC sequencing bias on SOLiD 5500xl machines. (A) Theoretical distribution of reads (increasing GC content from left to right, visualized in a light-gray to dark-gray color gradient) overlaid with the actual (experimental) number of reads in the entire microbial gDNA mixture in relation to %GC content per fragment under different ePCR conditions, i.e., E80 ePCR pouch (dark-blue line), E20 ePCR pouch (medium-blue line), 96 well plate (contents of an E20 ePCR pouch distributed into 96 well plates, light-blue line), in comparison to an Illumina library (orange line). (B) % of shift of genomic content for the different microbes, plotted in relation to ePCR conditions and compared with Illumina sequencing results, *Plasmodium falciparum* (20% GC, violet line), *Picrophilus torridus* (36% GC, light-blue line), *Escherichia coli* (51% GC, cyan line), *Pseudomonas putida* (62% GC, light-green line) and *Micrococcus luteus* (70% GC, dark-green line). (C-F) GC sequencing bias for the individual microbes. Panel shows theoretical (black lines) and experimental (colored lines) read frequencies for each of the microbes under the different sequencing conditions. (From left to right) highest to lowest GC content: *M. luteus* (dark green), *P. putida* (light green), *E. coli* (cyan), *P. torridus* (light blue) and *P. falciparum* (violet). (From top to bottom) (C) SOLiD E80 ePCR pouch (blue), (D) SOLiD E20 ePCR pouch (cyan), (E) SOLiD 96 well plate (light blue) and (F) Illumina MiSeq (orange).

seq of transcription factors or histone marks. Histone modifications are present in a wide range of genomic regions, depending on the particular modification and on the investigated cell type.^{9,10} The range from point source, via broad source to mixed source distributions among histone marks is associated with a distribution over a vast range of genomic elements that can be potentially challenging for sequencing.⁹ Especially, for ChIP-seq data or DNA methylation studies, quality depends on the amount of bias introduced during library preparation or sequencing, since both types of marks have been reported to be associated with challenging GC rich CpG islands.^{11,12}

So far, GC and AT biases have been well addressed and successfully reduced for the Illumina technologies by adapting PCR protocols, including changes to denaturation times and ramp rates, and using optimized polymerases for library amplification.^{3,13} In contrast, comparable approaches are missing for the SOLiD system. Thus, we were interested to investigate how strong the sequencing bias in the SOLiD system might be and to which extent this would affect ChIP-seq data. With the idea to characterize and improve the potential bias for the SOLiD machine and make it more suitable for applications such as ChIP-seq, we initially tested PCR-free library preparations for potential GC and other biases on this system.

In the present study, we observed a very strong GC bias in SOLiD sequencing data that exceeded previous estimations of the tentative sequencing bias in this system. We identified the emulsion PCR (ePCR) of the SOLiD sequencing system as the major source of GC bias. In response to that, we show approaches to reduce this bias by changing the ePCR conditions accordingly. Likewise, ChIP-seq of acetylated Histone H3 lysine 9 (H3K9ac), a mixed source histone mark, on a SOLiD 5500xl machine resulted in data almost void of reads at GC rich loci (GC content higher than 62%) while these regions were well covered by an Illumina sequencing system. ChIP-seq data on H3K9ac on the SOLiD was improved by optimized ePCR conditions and sufficient sequencing depth.

Results

Characterization of GC bias on the SOLiD 5500xl system

First, we aimed to characterize and potentially improve the sequencing bias of the SOLiD 5500xl

machine. While there are several previous studies on how to counteract sequencing biases for the Illumina system, i.e. mostly by adaptations to the library construction protocols, no such attempts have been reported for the SOLiD library preparation workflow. Typically, these studies utilize mixes of small microbial genomes (1.5 to 24 MB, as opposed to human 3.2 GB^{3,6,13}). Such microbial gDNA mixes contain microorganisms with different GC content covering the whole spectrum of high, low and medium genomic GC content, allowing the characterization of the full range of bias of a protocol and its potential improvement through applied modifications.

In this study, we mixed 5 microbes with genomic GC contents of 20%, 36%, 51%, 62% and 72% that would allow for a good characterization of the sequencing performance. We build a PCR-free library from this gDNA mixture using the standard protocol and chemistry. Against the general view that the sequencing bias in a library should be minimized in the absence of an upscale PCR, but in line with our hypothesis, a strong bias was already present in the PCR-free library. Both, regions with very high AT content and to an even stronger extent those with high GC content were underrepresented in the PCR-free SOLiD sequencing data (Fig. 1A–C). Comparing our SOLiD data to the expected GC distribution (calculated based on genomic sequence of the pooled microbes), we observed a shift in distribution for all microorganisms – with the strongest shifts for organisms with high GC and high AT and moderate to negligible shifts in organisms with moderate GC/AT content (see Table 1). Subsequently, the exact same microbial gDNA mixture was used to build a PCR-free library for the Illumina sequencing system (i.e., MiSeq) using standard Illumina reagents. Indeed, this library had a reduced sequencing bias when compared with the expected GC distribution and to the data of the PCR-free library sequenced on the SOLiD

Table 1. Summary of shift for the different microbes and ePCR conditions and for Illumina bridge PCR.

Microbe	GC content	SOLiD E80*	SOLiD E20*	SOLiD 96*	Illumina
<i>Plasmodium falciparum</i>	20%	46%	39%	38%	21%
<i>Picrophilus torridus</i>	36%	2%	1%	1%	15%
<i>Escherichia coli</i>	51%	24%	18%	17%	16%
<i>Pseudomonas putida</i>	62%	66%	34%	25%	22%
<i>Micrococcus luteus</i>	72%	96%	86%	74%	23%

*ePCR performed in an E80 or E20 pouch or in 96 well plates.

5500xl system (Fig. 1A, B and F). The shift for the same gDNA mix, sequenced on the Illumina MiSeq, was comparable low to moderate for all microbe genomes (Table 1). Interestingly, for *Picrophilus torridus* (with a moderate to low GC content of 36%), sequencing bias was significantly lower in the SOLiD library than for the Illumina library (2% versus 15% shift, Fig. 1A–C and F).

Since the SOLiD PCR-free library already presented with a strong bias, attempts to improve it by changing the upscale PCR (PCR protocol or polymerase) did not seem useful. The data rather suggested that the sequencing bias is largely introduced during the SOLiD emulsion PCR (ePCR), the equivalent to the Illumina bridge PCR (needed for cluster generation). Thus, we aimed to address this hypothesis by changing the ePCR conditions.

Improving the SOLiD GC bias with modifications to the emulsion PCR (ePCR)

Typically, the ePCR is performed in pouches of different volumes (from 10 mL to 120 mL, depending on the amount of libraries to be sequenced) in a bead amplifier, before loading the libraries onto the sequencer. Since we initially used a larger pouch of 80 mL for cluster generation/amplification of the libraries from the microbial gDNA mix, we hypothesized that a potentially uneven heat distribution in the pouch may have contributed to the observed bias.

To test this, we performed the emulsion PCR of the exact same microbial gDNA mix library, used throughout this study, in a 20 mL pouch and in 96 well plates. Indeed, both approaches improved the sequencing quality (Fig. 1A, B and D, E). Sequencing data with prior ePCR in a 20 mL pouch showed a reduction of the shift for all genomes compared with the 80 mL pouch ePCR condition, with most pronounced improvements for the GC rich genomes. The 96 well plate condition improved the SOLiD sequencing data further. We observed a reduction of the shift for all genomes compared with the 80 mL and 20 mL (except for the well represented *Picrophilus torridus* genome) pouch ePCR conditions (Fig. 1A–E, Table 1).

H3K9ac ChIP-seq on the SOLiD 5500xl sequencer is prone to GC bias

To determine to what extent this bias may hamper typical applications such as histone ChIP-seq, we first

sequenced H3K9ac ChIP-DNA under standard conditions (ePCR in a 80 mL pouch). We deliberately selected a mixed source histone mark to get a better idea on how the bias might affect different genomic elements.

Sequencing of H3K9ac ChIP-DNA (prepared from adult mouse hippocampus) on the SOLiD 5500xl sequencer showed a strong enrichment compared with input controls (Fig. 2A, blue vs. gray tracks), indicative of reliable performance of the ChIP procedure. In line with published data on H3K9ac ChIP-seq, in e.g. embryonic stem cell (ESC) nuclei,¹⁴ we find H3K9ac signals not only restricted to the transcription start site (TSS) of genes (Fig. 2B), which is typical for mixed source distribution marks.⁹ We observed H3K9ac occupancies mostly on intronic regions, followed by intergenic regions, exons and promoter regions. In addition, we found occupancies in 5'UTRs, CpG islands, 3'UTRs and down-stream of the TSS.

Yet, as expected, we observed the strong GC bias in H3K9ac ChIP-seq as well. Similar to the GC bias present in our data from the microbial mix gDNA, respective regions with high and low GC content were most strongly affected. We noticed a complete absence of coverage for some loci, such as CpG islands with >62% GC content. Specifically, we observed a good coverage and highly elevated ChIP signal for the shore regions of CpG islands and a sudden decrease or complete loss of coverage for the islands themselves which typically have a higher GC content than their shores. These loci missing in the ChIP-seq data had no coverage in the input samples, strongly indicative of a sequence and not histone-mark dependent loss of coverage (Fig. 2A and D).

The peaks with the loss of coverage on the SOLiD sequencer were much better preserved when sequenced with the Illumina system MiSeq (Fig. 2A and D, orange tracks). Overall, compared with the H3K9ac ChIP-seq data generated with the Illumina sequencing technology, the data from the SOLiD 5500xl showed a loss of reads at several genomic elements, including exons and 5'UTRs. Specifically, CpG islands had a much lower coverage which is explained by the fact that loci with high GC contents (above 62%) were extremely underrepresented when sequenced on a SOLiD 5500xl system. In contrast, promoters, introns, 3'UTRs, and the shores of CpG

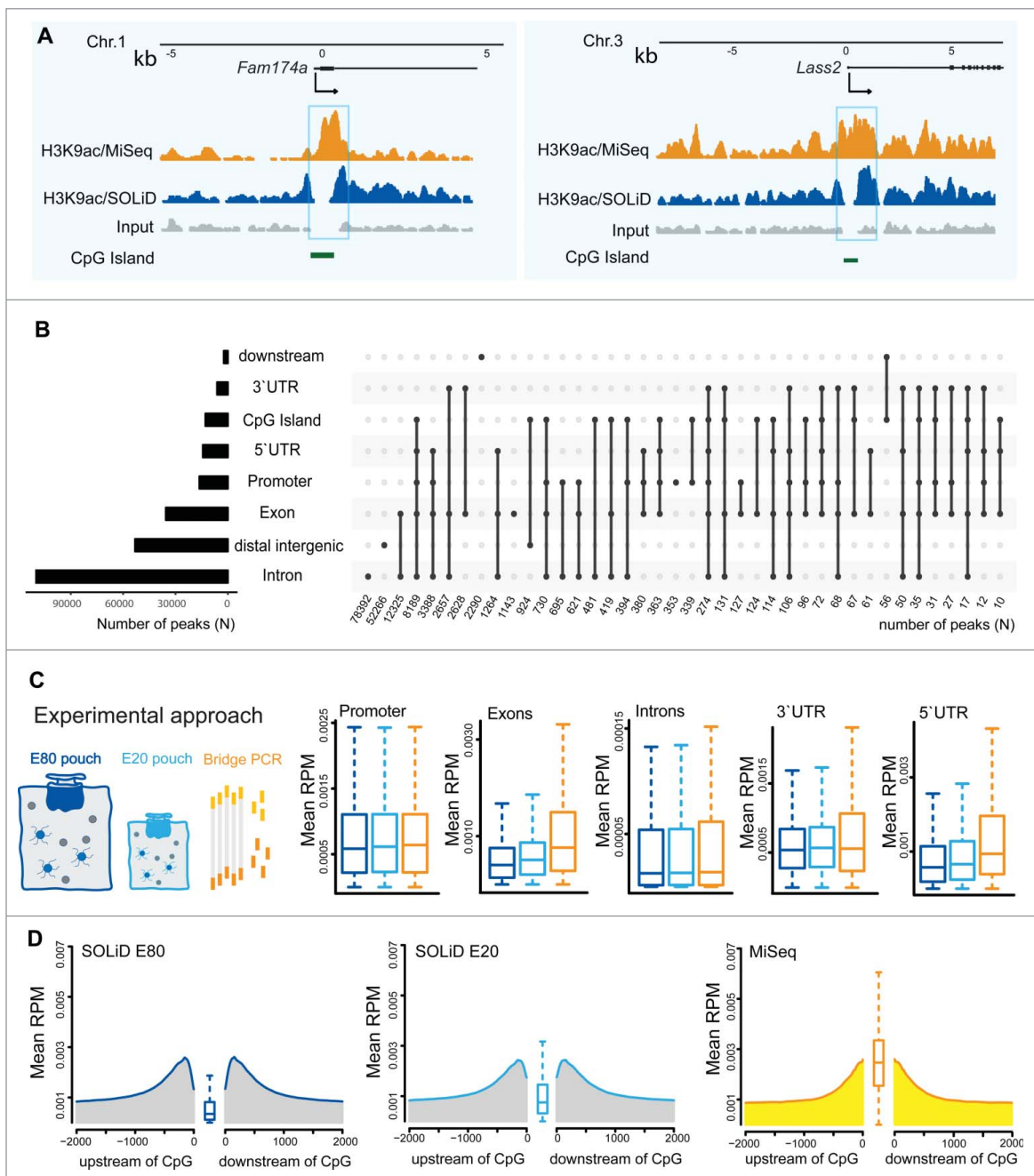


Figure 2. H3K9ac ChIP-seq on SOLiD 5500xl vs. Illumina MiSeq. (A) Representative UCSC genome browser screenshots from H3K9ac ChIP-DNA sequenced on a SOLiD 5500xl (blue tracks) and an Illumina MiSeq (orange tracks). Note enrichment of ChIP-DNA tracks over input controls (gray tracks). Light blue boxes indicate gaps in the SOLiD 5500xl sequencing tracks, as compared with the Illumina sequencing tracks, typically, over CpG islands (green bars). (B) (Left) graph shows genomic elements with occupancy with the H3K9ac mark, plotted with corresponding number of peaks. (Right) dotted grid displays distribution of peaks over these genomic elements, including the information if peaks cover multiple genomic elements. (C) (Left side of the panel) Cartoon of experimental design for H3K9ac ChIP-seq with ePCR in 80 mL and 20 mL pouches (SOLiD) and with bridge PCR (Illumina). (Right side of the panel) Box-plots show sequencing coverage for H3K9ac on the SOLiD 5500xl (20 mL pouch, cyan and 80 mL pouch, blue) and on an Illumina platform (MiSeq, orange) for promoters, exons, introns, 3'UTRs, and 5'UTRs. (D) CpG island plots showcase coverage for the island and its shores, continuously 2kb up and downstream from the CpG island. Coverage is expressed in reads per million (RPM).

islands were covered to comparable extent by both sequencing systems (Fig. 2C, D).

Next, we were interested if our observation was common and generalizable. Therefore, we re-analyzed publically available H3K9ac ChIP-seq data (i.e., on mouse ESCs) from both sequencers. As expected, the Illumina and SOLiD sequencing data differed in their coverage comparable to our results. Likewise, we observed a loss of reads for promoters, exons, 5'UTRs, and the most pronounced loss of coverage for GpG islands (Fig. S1), suggesting that this underrepresentation, likely due to the GC bias in the SOLiD 5500xl system, can be generalized for data sets generated with this method.

Improving H3K9ac ChIP-seq data

Given that our modifications to the SOLiD ePCR improved the GC sequencing bias for the microbial gDNA mixture, we tested whether this would sufficiently improve our data on H3K9ac ChIP-DNA. Indeed, compared with the library, processed for ePCR in a big E80 pouch (80 mL), using a smaller E20 pouch (20 mL) in this enrichment step improved the data significantly. We primarily gained reads for the previously severely underrepresented GpG islands (Fig. 2D).

Consideration on sequencing depth for H3K9ac

Finally, we wondered if the lack of coverage could be partially attributed to insufficient sequencing depth which is also relevant for the interpretation of other published H3K9ac-seq data sets generated on the SOLiD 5500xl. In particular, this question is of high interest when sequencing or analyzing a histone mark with mixed source distribution such as H3K9ac which requires a higher sequencing depth than point source marks such as H3K4me3 to reach sufficient coverage.⁹

In the present study we sequenced with a depth of 40 Million reads per H3K9ac sample which exceeds the recommendation by ENCODE and those commonly found in the literature.^{9,15} Our sequencing depth reached saturation, indicating that we were sequencing in a sufficient range (Fig. S2B).

To test to which extent lower read numbers would affect H3K9ac data we applied first an *in silico* random down-sampling approach. Read numbers of the H3K9ac sequencing track were randomly removed, down to 20 Million, 10 Million and 5 Million reads.

As expected, with every down-sampling step we lost actual coverage, affecting even the highly covered promoter regions at a sequencing depth as low as 5 Million reads. Also, signals on the gene body were lost with 20 Million reads. To evaluate this finding experimentally, we specifically tested latter loci that lost their coverage from 20 Million reads on by qPCR for their relative level of occupancy. These regions were compared with loci on the same gene that had H3K9ac signals under all levels of coverage, for a positive control, and to loci on the gene body and intergenic loci (in proximity to this gene) which did not show occupancies with 40 Million reads, as putative 'negative' controls. As predicted, amplification for the target loci was comparable to the positive control (*Hap1*) or even exceeded the value of the positive control locus (*Meis2*), suggesting that a sequencing depth of below 40 Million reads/per sample for H3K9ac ChIP-seq will lead to significant loss of data. Furthermore, we detected moderate qPCR signals for loci on the gene body and to some extent in intergenic regions (e.g., for *Hap1*), clearly indicating that H3K9ac ChIP-seq can benefit from increasing the sequencing depth beyond 40 Million reads per library (Fig. S2C). These intergenic regions might be highly relevant for studies into genetic and epigenetic mechanisms of neuropsychiatric disease mechanisms, since e.g., relevant GWAS risk loci often map to intergenic regions.¹⁶

Discussion

This study provides an in-depth characterization of the sequencing bias of the SOLiD 5500xl sequencer that was previously not well described and delivers profound insights into addressing biases in this system. We report that data generated on the SOLiD 5500xl sequencer is strongly impacted by an extensive GC bias. In contrast to theoretical assumptions, this bias was present in a PCR-free protocol and thus cannot be improved by changing the library preparation conditions such as changing the PCR-polymerase or the ramp rate of the thermocycler during upscale PCR as previously applied with good success for Illumina libraries.^{3,13}

The presence of the bias in the PCR-free protocol led us to identify the emulsion-PCR step (ePCR, the SOLiD equivalent to the Illumina bridge PCR, required for cluster generation) as the major source of GC bias. In fact, even heat distribution in the emulsion

reaction cocktail has been reported as essential for ePCR applications.¹⁷ Thus, we aimed to improve the GC under-representation by changing the ePCR conditions. And indeed, both ways to improve the uniformity of thermal conditions, i.e., using a smaller ePCR-pouch and distributing the contents of a small ePCR-pouch into 96 well plates for running the ePCR reduced the GC bias significantly. Yet, our attempts to improve the sequencing performance for GC rich loci did not reach the level of on an Illumina machine, particularly at very high GC contents. The difference between the Illumina library and the SOLiD library enriched in an E20-pouch and a 96 well plate, respectively, was 12% and 3% for the *Pseudomonas putida* (62% GC content) genome, but 63% and 51% for *Micrococcus luteus* (72% GC content). It is noteworthy though that the observed bias was minimal for the *P.torridus* microbe (GC content of 36%) under all sequencing conditions on the SOLiD platform and far better than the shift observed on the Illumina system, suggesting the SOLiD as the ideal platform of choice for researchers dealing with organisms in a similar range of GC content.

It is conceivable that other factors than the evenness of heat distribution may be additional sources of the observed GC bias in the ePCR, such as the polymerase used or the ePCR protocol itself. While we already reduced the number of cycles in the ePCR by one third and a further reduction may lead to an incomplete enrichment of beads, additional approaches might consist of changes to ramp rate, denaturation and annealing length and time and temperatures, and the polymerase used for ePCR.

In a next step we demonstrate to which extent actual common sequencing applications are affected by the SOLiD sequencing bias and if they can be rescued sufficiently. We chose histone ChIP-seq for the mixed source distribution mark H3K9ac (Fig. S2A), as an example of an application that requires good coverage of multiple genomic elements such as introns, exons, promoter regions and CpG islands. We observed that H3K9ac ChIP-seq on a SOLiD 5500xl machine led to an extensive loss of coverage in regions with high GC content (> 62%) in comparison with samples run on an Illumina sequencer. In line with our data on the microbial gDNA mix, this bias was again most pronounced in GC rich loci. The bias was not due to an overall bad sequencing quality of H3K9ac ChIP-seq, which actually exceeded ENCODE

standards and recommendations by recent literature (of 20–40 Million reads for ChIP-seq) in sequencing depth.^{9,15} Notably, sequencing depth reached saturation (Fig. S2B), and thus our experiments were clearly performed with *bona fide* sequencing depth for a mixed source distribution histone mark such as H3K9ac. Moreover, the bias was not associated with bad quality of our ChIP-DNA, e.g., caused by failure of the ChIP procedure or over-shearing of DNA, since our tracks showed strong enrichment of histone acetylation signals, the GC bias occurred in the input samples as well (Fig. 2A), and the phenomenon was also present in published data sets of other groups (Fig. S1,¹⁴). Interestingly, the latter data sets presented with a significant, but less pronounced bias than our ChIP-seq data. This might be due to the wider distribution of H3K9ac in adult brain (our study), e.g., we observed broader and higher peaks than in embryonic stem cells. Notwithstanding, our bias may be exceedingly strong, because we sequenced sufficiently deep and thus more peaks (on the Illumina sequencer) or the absence of them (in the SOLiD system) were detected. This view is supported by our down-sampling data in conjunction with ChIP-qPCR, indicating that 5 million reads hardly suffice to detect even all major reads at the promoter regions and a sequencing depth of 20 million reads may still miss loci with H3K9ac signal on the gene body.

In conclusion, Illumina platforms seem to be better suited for applications that require sequencing of GC rich loci such as histone ChIP-seq and presumable DNA methylation studies with focus on CpG islands. Comparable sequencing coverage in the Illumina and SOLiD systems can be reached at balanced levels of GC content (e.g., *E. coli*, GC content 51%), and better coverage in the SOLiD sequencing libraries can be obtained for lower GC contents (e.g., *P. torridus*, GC content 36%), with and without improved ePCR.

Above all, it is essential to know the extent of the GC bias in the SOLiD 5500xl system for better interpretation of data previously generated on this machine. These data can still be valuable for diverse meta-analysis and may be improved by pooling samples during analysis to reach higher coverage and detect all potential peaks in ChIP samples. Findings from this study are also relevant for the interpretation of data, obtained with other sequencing systems such as Polony Sequencing, Roche 454 or Ion Torrent using ePCR for enrichment.

Material and methods

gDNA isolation and preparation (microbial gDNA)

All microbes were grown in appropriate media under specific conditions required for each of the organisms and thereafter pelleted and stored at -20°C until use. gDNA was isolated using Tris-HCl lysis buffer (1% SDS; pH 8.0) and TE-buffered Roti-Phenol/Chloroform/Isoamyl alcohol (25:24:1, pH 7.5–8.0, Carl Roth, #A156.2). Samples were precipitated with ethanol and purified with QIAquick spin columns (Qiagen, #28104). Contaminating RNA (which could interfere with the quantification of gDNA) was digested with RNase A. Aliquots of each sample were run on an agarose gel to check size and purity of the respective gDNA. Concentration of samples was measured with the Qubit Fluorometer (Life Technologies, #Q33216). Thereafter, the 5 different microbes were pooled in the following proportion: *Picrophilus torridus*: 100%, *Plasmodium falciparum*: 140%, *Pseudomonas putida*: 60%, *Escherichia coli*: 80% and *Micrococcus luteus*: 100%. Samples were then sheared to a size of 150 bps using a water bath Covaris ultrasonicator (#S220).

Library preparation and sequencing

1/ Library preparation for SOLiD sequencing

Libraries were built from ChIP DNA, input DNA or gDNA (microbial gDNA mixture) using the fragment library preparation kit for 5500 series SOLiD systems (Applied Biosystems by Life Technologies, #4464412) according to the manufacturer's instructions. ChIP and input libraries were amplified on a Gene Amp PCR System 9700 (Applied Biosystems, #N8050200) with standard settings (ramp rate $5^{\circ}\text{C}/\text{s}$) and SOLiD standard reagents (AmpliTaq Gold polymerase, Applied Biosystems, #N8080241). The library from microbial gDNA was prepared amplification free. All libraries were quantified on a Qubit Fluorometer and checked for expected size on a Bioanalyzer (Agilent Technologies, #G2939AA).

For bead emulsion PCR (ePCR) the SOLiD EZ Bead E80 System (Applied Biosystems #4472999) was initially used. Alternatively, E20 System pouches (Applied Biosystems, #4453094) were used instead. Pre-set cycling conditions, used on the bead amplifier (ePCR, Applied Biosystems, #4448419) were as follows: 95°C for 350 sec, 60°C for 60 sec and 75°C

for 75 sec, followed by 60 cycles of 96°C for 65 sec, 60°C for 60 sec and 75°C for 75 sec, and a final step of 75°C for 420 sec, 50°C for 120 sec and 30°C for 12 sec, each. In a third approach the content of a E20 bag was distributed to 96 well plates and run on a thermocycler (ABI 9700) using the following conditions: initial denaturation (95°C for 5 min), followed by 40 cycles of denaturation (93°C for 15 sec), annealing (62°C for 30 sec) and extension (72°C for 75 sec), followed by final heating (72°C for 7 min). ePCR was cleaned up and enriched to a concentration of approximately 1.5 million beads/ μL , according to the color scale provided in the manufactures manual. Thereafter samples were loaded onto the SOLiD 5500 xl machine (Applied Biosystems, # 4460730).

2/ Library preparation for Illumina sequencing

Libraries were prepared using either the Illumina TruSeq ChIP Sample Preparation kit for ChIP libraries (Illumina, #IP-202–1012) or the TruSeq DNA PCR-free sample preparation kit (Illumina, #FC-121–3001) for library preparation from the microbial gDNA mixture. The ChIP DNA library was amplified by PCR (Biorad T100 Thermal Cycler, #1861096). The following PCR conditions were used: denaturation for 1 min at 95°C , 15 cycles of denaturation (at 95°C for 50 sec), annealing (at 65°C for 1 min) and extension (at 72°C for 30 sec), followed by final heating (72°C for 10 min). Amplified libraries were size selected on a 2% agarose gel and purified using the QIAquick Gel Extraction Kit (Qiagen). Libraries were quantified on a Qubit fluorometer and checked for correct size distribution on a Bioanalyzer. Samples were diluted to a concentration of 4 nM, denatured in 1 N NaOH and 16 pM library, with 5% PhiX control (Illumina, #FC-110–3001) spiked in, were loaded on a MiSeq machine (Illumina, #SY-410–1003) equipped with MiSeq Reagent Kit v3 (Illumina, #MS-102–3001) sequencing chemistry.

Chromatin immunoprecipitation (ChIP)

One mouse hippocampus per ChIP reaction was dounced in 400 μL MNase digestion/ nuclei permeabilization buffer and digested with MNase (Sigma Aldrich, #N3755) to obtain mononucleosomal DNA (~ 150 bps). ChIP was performed with 4 μL anti H3K9ac polyclonal rabbit antibody (Millipore, #ABE18). DNA from input was extracted in parallel.

A small aliquot of the input samples was checked on agarose gels for size distribution of sheared chromatin.¹⁸

ChIP qPCR

For quantification by qPCR H3K9ac ChIP DNA was diluted 1:5 with elution buffer (Qiagen) and amplified on a Light Cycler 2.0 (Roche diagnostics, #03531414001) using QuantiFast SYBR Green Kit Master Mix (Qiagen, #204054). The following primers were used: *Meis2* (NC_000068.7), positive control locus (present under all sequencing depth settings), fwd 5'-TCGGTCAATATGCGTGTGGT-3', rev 5'-CTGCCCCATGCTTGTGTTTC-3', target locus (present only in the deepest sequencing condition), fwd 5'-GGGCTCTTCAGAATGGCACT-3' rev 5'-CAAATGAATGGGGTGGGGG-3', negative control locus on gene body (present in none of the sequencing conditions), fwd 5'-AAATGTCACCCAGGGACACC-3', rev 5'-AACCTTTCAGGCTGGAGTT-3' and a negative control in an intergenic locus (present in none of the sequencing conditions), fwd 5'-AACAGTGGGGTCTGCTGATG-3', rev 5'-GGACAGCAAACGCTAGACCT-3'; *Hap1*, (NC_000077.6) positive control locus, fwd 5'-GGGGTGACCGTTGATCAGTT-3', rev 5'-CCTATCTCGTCACCACTGGC-3', target locus, fwd 5'-GGTGGTGGA AAGGTGGA ACT-3', rev 5'-TCCCGCATTGGGCACTATTT-3', locus on gene body, fwd 5'-CGCAGGGTCAGTGATGAACT-3', rev 5'-TGTTGGGGTGGAAATGTCTC-3', and intergenic locus, fwd 5'-ATTGTTGTGCTAGCCAGCCT-3', rev 5'-TACCTGACCCAGGATGGTG-3'. PCR reactions (10 μ L final volume) were run in duplicates with of 1.5 μ M of specific primers and 2 μ L of ChIP-DNA or input DNA, respectively. ChIP Cts for each sample and primer were normalized by the Ct for the input DNA. Amplification levels are presented relative to the levels for the positive control locus which was set to 1.

Bioinformatic analysis

Read quality was checked using FastQC tool.¹⁹ Adapters were trimmed using cutadapt.²⁰ SOLiD reads were processed with in house scripts for compatibility with the aligner. The alignment to mm9 and the microbial mix genomes was done using either BFAST v0.7.0a²¹ for color space alignment, or BWA.aln v.0.7.10²² for nucleotide space with standard parameters. Only

uniquely mapping reads were accepted. To estimate fragment size in each library, we used MaSC.²³ Reads were elongated to the estimated size using BEADS.²⁴ To calculate the GC distribution, BEADS was used again. Power calculation for H3K9ac was performed with ChIP-Seq Statistical Power (CSSP) analysis in R using the standard settings with two- and fourfold enrichment as parameters.²⁵

Construction of the theoretical GC distribution

For the construction of the theoretical GC distribution of the microbial genome pool, we used the reference genomes (*M. luteus* NCTC2665, *P. putida* KT2440, *E. coli* K12-DH10B, *P. torridus* DSM 9790 and *P. falciparum* Pf3D7_v2.1.5), shredded them to 75 bp to match our sequencing reads using BEADS, and aligned them against themselves to account for mappability issues. The resulting alignment was elongated to 150 bp (to match the shearing of our library) and the GC distribution was calculated again using BEADS.

The estimation of the shift between the theoretical and actually observed GC distribution is made possible by excluding any amplification procedures in our library preparation (before the emulsion PCR). In this type of library, each fragment of the original gDNA mix can either be sequenced once or fail to cluster and drop out. Therefore, we know that the resulting GC distribution after sequencing will not surmount the theoretical distribution at any GC level and we can fit it within using the loess function in R.²⁶ To calculate the shift between the fitted observed GC distribution and the theoretical distribution, we subtracted the areas under the curve ($\frac{AUC_{theoretical} - AUC_{observed}}{AUC_{theoretical}}$) using R.

This shift between the 2 distributions is used as a proxy for the magnitude of the bias in the analysis.

Accession numbers

gDNA and ChIP-seq data have been deposited in the Sequence Read Archive (SRA) under the BioProject accession number PRJNA380045.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Acknowledgment

We thank Dr Angel Angelov (Technical University, Munich) for kindly providing us with *Picrophilus torridus*

gDNA, Prof Dr Marc Bramkamp and Dr Karin Schubert (Ludwig-Maximilians-University, Munich) who friendly supplied us with *Micrococcus luteus* and *Pseudomonas putida* pellets. Likewise, we thank Dr Berens-Riha (Ludwig-Maximilian-University, Munich) for providing *Plasmodium falciparum* gDNA. We are grateful to Dr Tobias Spielmann and Florian Kruse (Bernhard Nocht Institute, Hamburg) for kindly preparing large amounts of *Plasmodium falciparum* gDNA for us. Also, we thank Dr Lutz Wiehlmann (Hannover Medical School, Hannover) for technical support.

Funding

This work was supported by a Marie Curie International Incoming Fellowship within the 7th European Community Framework Program from the European Commission under Grant #332297 (to MJ) and in part by a NARSAD Young Investigator Grant from the Brain and Behavior Research Foundation under Grant #22809 (to MJ). Dr Mira Jakovcevski is an “Attias Family Foundation Investigator.”

References

- [1] Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. *Cell* 2013; 155(1):27-38; PMID:24074859; <https://doi.org/10.1016/j.cell.2013.09.006>
- [2] PsychENCODE Consortium, Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, Crawford GE, Jaffe AE, Pinto D, Dracheva S, et al. The Psychencode Project. *Nat Neurosci* 2015; 18(12):1707-12; PMID:26605881; <https://doi.org/10.1038/nn.4156>
- [3] Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* 2011; 12(2):R18; PMID:21338519; <https://doi.org/10.1186/gb-2011-12-2-r18>
- [4] Rieber N, Zapatka M, Lasitschka B, Jones D, Northcott P, Hutter B, Jäger N, Kool M, Taylor M, Lichter P, et al. Coverage bias and sensitivity of variant calling for four whole-genome sequencing technologies. *PLoS One* 2013; 8(6):e66621; PMID:23776689; <https://doi.org/10.1371/journal.pone.0066621>
- [5] Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB. Characterizing and measuring bias in sequence data. *Genome Biol* 2013; 14(5):R51; PMID:23718773; <https://doi.org/10.1186/gb-2013-14-5-r51>
- [6] Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G + C)-biased genomes. *Nat Methods* 2009; 6(4):291-5; PMID:19287394; <https://doi.org/10.1038/nmeth.1311>
- [7] Huptas C, Scherer S, Wenning M. Optimized Illumina PCR-free library preparation for bacterial whole genome sequencing and analysis of factors influencing de novo assembly. *BMC Res Notes* 2016; 9:269; PMID:27176120; <https://doi.org/10.1186/s13104-016-2072-9>
- [8] Kebschull JM, Zador AM. Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Res* 2015; 43(21):e143; PMID:26187991
- [9] Sims D, Sudbery I, Illott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 2014; 15(2):121-32; PMID:24434847; <https://doi.org/10.1038/nrg3642>
- [10] Jakovcevski M, Akbarian S, Di Benedetto B. Pharmacological modulation of astrocytes and the role of cell type-specific histone modifications for the treatment of mood disorders. *Curr Opin Pharmacol* 2016; 26:61-6; PMID:26515273; <https://doi.org/10.1016/j.coph.2015.10.002>
- [11] Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D’Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 2010; 466(7303):253-7; PMID:20613842; <https://doi.org/10.1038/nature09165>
- [12] Roh TY, Cuddapah S, Zhao K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev* 2005; 19(5):542-52; PMID:15706033; <https://doi.org/10.1101/gad.1272505>
- [13] Oyola SO, Otto TD, Gu Y, Maslen G, Manske M, Campino S, Turner DJ, Macinnis B, Kwiatkowski DP, Swerdlow HP, et al. Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC Genom* 2012; 13:1; <https://doi.org/10.1186/1471-2164-13-1>
- [14] Hezroni H, Tzchori I, Davidi A, Mattout A, Biran A, Nissim-Rafinia M, Westphal H, Meshorer E. H3K9 histone acetylation predicts pluripotency and reprogramming capacity of ES cells. *Nucleus* 2011; 2(4):300-9; PMID:21941115; <https://doi.org/10.4161/nucl.2.4.16767>
- [15] Nakato R, Shirahige K. Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. *Brief Bioinform* 2016; 2016:1-12
- [16] Reddy AS, O’Brien D, Pisat N, Weichselbaum CT, Sakers K, Lisci M, Dalal JS, Dougherty JD. A Comprehensive analysis of cell Type-Specific nuclear RNA from neurons and glia of the brain. *Biol Psychiatry* 2017; 81(3):252-64; PMID:27113499; <https://doi.org/10.1016/j.biopsych.2016.02.021>
- [17] Castellanos-Rizaldos E, Milbury CA, Makrigiorgos GM. Enrichment of mutations in multiple DNA sequences using COLD-PCR in Emulsion. *PLoS One* 2012; 7(12):e51362; PMID:23236486; <https://doi.org/10.1371/journal.pone.0051362>
- [18] Jakovcevski M, Ruan H, Shen EY, Dincer A, Javidfar B, Ma Q, Peter CJ, Cheung I, Mitchell AC, Jiang Y, et al. Neuronal Kmt2a/Mll1 histone methyltransferase is essential for

- prefrontal synaptic plasticity and working memory. *J Neurosci* 2015; 35(13):5097-108; PMID:25834037; <https://doi.org/10.1523/JNEUROSCI.3004-14.2015>
- [19] Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- [20] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMB Net J* 2011; <https://doi.org/10.14806/ej.17.1.200>
- [21] Homer N, Merriman B, Nelson SF. BFAST: An Alignment Tool for Large Scale Genome Resequencing. *PLoS One* 2009; 4(11):e7767; PMID:19907642; <https://doi.org/10.1371/journal.pone.0007767>
- [22] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 2009; 25(14):1754-60; PMID:19451168; <https://doi.org/10.1093/bioinformatics/btp324>
- [23] Ramachandran P, Palidwor GA, Porter CJ, Perkins TJ. MaSC: mappability-sensitive cross-correlation for estimating mean fragment length of single-end short-read sequencing data. *Bioinformatics* 2013; 29(4):444-50; PMID:23300135; <https://doi.org/10.1093/bioinformatics/btt001>
- [24] Cheung MS, Down TA, Latorre I, Ahringer J. Systematic bias in high-throughput sequencing data and its correction by BEADS. *Nucleic Acids Res* 2011; 39(15):e103; PMID:21646344; <https://doi.org/10.1093/nar/gkr425>
- [25] Zuo C, Keleş S. A statistical framework for power calculations in ChIP-seq experiments. *Bioinformatics* 2014; 30(6):753-60; PMID:23665773; <https://doi.org/10.1093/bioinformatics/btt200>
- [26] Statistical Research. Random Statistics and Data Science. That's Smooth; 2013 Oct 10 [assessed 2016 Jan 05]. <https://statistical-research.com/index.php/2013/10/10/thats-smooth/>

Publication II

HAM-TBS: high-accuracy methylation measurements via targeted bisulfite sequencing

Authors:

Simone Röh*¹, Tobias Wiechmann*¹, Susann Sauer¹, Maik Ködel¹, Elisabeth B. Binder^{1,2} and Nadine Provençal^{1,3,4}

* shared first authorship

¹ Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany.

² Department of Psychiatry and Behavioral Sciences, Emory University Medical School, Atlanta, GA 30329, USA.

³ Faculty of Health Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

⁴ Healthy Starts Theme, BC Children's Hospital Research Institute, Vancouver, BC V5Z 4H4, Canada.

Publication details:

Published in Epigenetics & Chromatin 11:39 on July 4th, 2018

<https://doi.org/10.1186/s13072-018-0209-x>

© The Author(s) 2018. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

METHODOLOGY

Open Access



HAM-TBS: high-accuracy methylation measurements via targeted bisulfite sequencing

Simone Roeh^{1†}, Tobias Wiechmann^{1†}, Susann Sauer¹, Maik Ködel¹, Elisabeth B. Binder^{1,2} and Nadine Provençal^{1,3,4*}

Abstract

Background: The ability to accurately and efficiently measure DNA methylation is critical to advance the understanding of this epigenetic mechanism and its contribution to common diseases. Here, we present a highly accurate method to measure methylation using bisulfite sequencing (termed HAM-TBS). This novel method is able to assess DNA methylation in multiple samples with high accuracy in a cost-effective manner. We developed this assay for the *FKBP5* locus, an important gene in the regulation of the stress system and previously linked to stress-related disorders, but the method is applicable to any locus of interest.

Results: HAM-TBS enables multiplexed analyses of up to 96 samples and regions spanning 10 kb using the Illumina MiSeq. It incorporates a triplicate bisulfite conversion step, pooled target enrichment via PCR, PCR-free library preparation and a minimum coverage of 1000x. TBS was able to resolve DNA methylation levels with a mean accuracy of 0.72%. Using this method, we designed and validated a targeted panel to specifically assess regulatory regions within the *FKBP5* locus that are not covered in commercially available DNA methylation arrays.

Conclusions: HAM-TBS represents a highly accurate, medium-throughput sequencing approach for robust detection of DNA methylation changes in specific target regions.

Keywords: Targeted bisulfite sequencing, DNA methylation, Next-generation sequencing, 5-methylcytosine, *FKBP5*

Background

DNA methylation is the covalent addition of a methyl group at the 5-carbon ring of cytosine, resulting in 5-methylcytosine (5mC). In the mammalian genome, this occurs predominantly in the context of CpG dinucleotides. It is one of several epigenetic marks influencing gene expression and serving multiple other purposes such as genomic imprinting, X chromosome inactivation and maintenance of genomic stability [1, 2]. Aberrant regulation of the establishment, maintenance, erasure or recognition of DNA methylation has been associated with a range of disease phenotypes [3, 4]. In addition, lasting effects of environmental risk factors may be reflected by changes in DNA methylation [5]. The need

to measure DNA methylation in large human cohorts in a cost-effective manner is therefore of increasing interest for research in epidemiology and medicine [6].

Assessing DNA modifications with high accuracy and sensitivity in candidate loci would increase the power to detect and replicate such effects as well as to perform time course experiments in large numbers of samples to understand the stability of the environmentally induced changes during development. In addition, changes related to specific environmental exposure may only be present in specific cell types, although most studies rely on more complex tissues such as postmortem brain or blood samples. Assessing these effects in mixed tissues requires high accuracy in order to detect small changes emerging from a small number of cells. DNA bisulfite treatment followed by next-generation sequencing enabled the quantification of DNA methylation marks at single-base resolution. However, genome-wide bisulfite sequencing, although the best approach to identify DNA

*Correspondence: Nadine_Provençal@sfu.ca

[†]Simone Roeh and Tobias Wiechmann shared first authorship

³ Faculty of Health Sciences, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada

Full list of author information is available at the end of the article

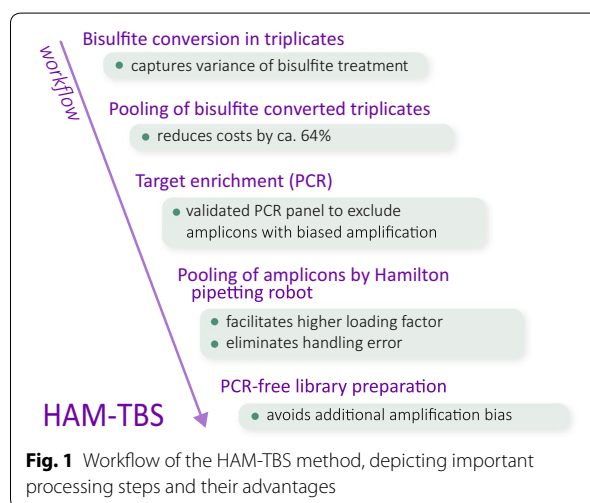


modifications, is still too cost intensive to be applied to large human cohorts at the coverage needed ($>60\times$) to detect differentially methylated sites [6]. Another set of accurate and cost-efficient measurement methods for DNA methylation at single CpG level are Illumina DNA methylation arrays. However, the ones currently available lack coverage in key enhancer regions that are important for environmentally driven changes and have a relatively small number of probes ($\sim 10\text{--}13$) covering each site. Targeted bisulfite sequencing (TBS) offers a candidate approach to perform such studies with high resolution by increasing depth of read coverage per CpG to detect small changes in DNA methylation in a cost-efficient manner. Recently, few applications of TBS have been developed with differences in accuracy, throughput and library preparation [7–10]. Our TBS approach focuses on the *FKBP5* gene, which encodes the FK506-binding protein (FKBP51), a co-chaperone tightly involved in stress regulation. Genetic and epigenetic factors have repeatedly been shown to increase the activity of this gene and associated with increased stress-reactivity and psychiatric disorders [11]. We have previously reported allele-specific demethylation of CpG sites located in intronic enhancer regions of *FKBP5* specific to posttraumatic stress disorder (PTSD) in patients who had experienced child abuse [12]. These gene \times environment interactions (G \times E) may be mediated by differential susceptibility to adversity-induced changes in DNA methylation in specific enhancers. Current methods do not cover the relevant enhancer regions of *FKBP5* affected by stress exposure. A highly accurate, cost- and time-efficient method to investigate *FKBP5* DNA methylation in a large number of samples is thus critical to gain more insight into how DNA methylation changes may mediate these G \times E. In this manuscript, we present a cost-effective, high-accuracy methylation measurement TBS (HAM-TBS) method to assess the regulatory regions of the *FKBP5* locus. Incorporating a triplicate bisulfite conversion step, PCR-free library preparation and rigorous quality control (validation of PCR target sites, $>95\%$ bisulfite conversion efficiency and $1000\times$ coverage minimum) ensures that our method is extremely robust (Fig. 1). Medium throughput and handling accuracy of up to 96 samples spanning approximately 10 kb is facilitated by embedding the Hamilton pipetting robot and TapeStation with the Illumina MiSeq sequencer.

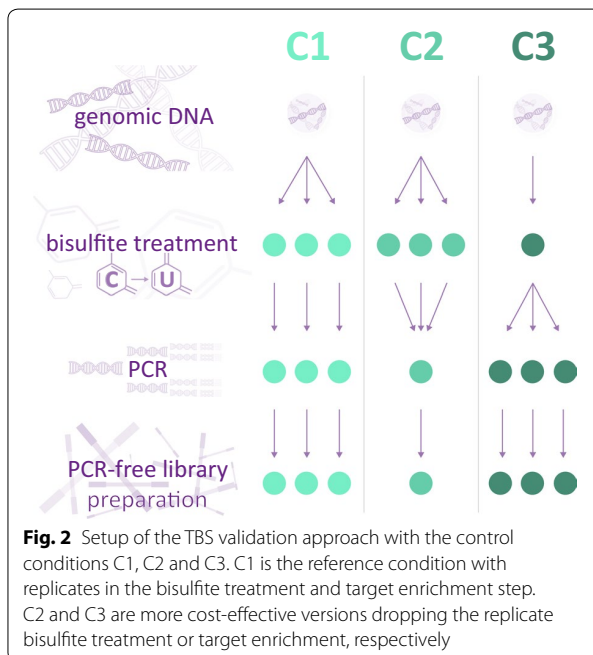
Results

QC, validation and optimization of the HAM-TBS method

TBS is based on bisulfite conversion coupled with targeted enrichment via PCR, library preparation for



sequencing and subsequent quantification of methylation levels. All steps are necessary and may influence the outcome by introducing bias to the assessment of methylation levels or by insufficient quality control of the data. The standard approach to minimize potential biases before sequencing is to produce replicates and assess the mean methylation levels during the analysis. In order to design a highly accurate yet cost-effective approach that is amenable to multiplexing, we assess at which step (bisulfite conversion or amplification) and to what extent technical variability would be introduced, as well as which quality control steps need to be performed on the sequencing data to ensure a robust analysis. To this end, we assessed the methylation level of 0, 25, 50, 75, 100% in vitro methylated bacterial artificial chromosome (BAC) control DNA for 3 different combinations of pooling strategies during the bisulfite treatment and PCR amplification (Fig. 2). Condition 1 (C1) assessed the methylation levels of control DNA using triplicate bisulfite treatments and PCR amplification for each replicate. C1 was considered the standard reference condition since each step was performed in triplicates. In condition 2 (C2), triplicate bisulfite treatments were pooled to perform one PCR amplification reducing the costs by approximately 64%. Finally, in condition 3 (C3) one bisulfite treatment of the control DNAs was performed followed by 3 separate PCR amplifications to assess the extent of the target enrichment bias. A smaller panel of 11 different PCRs (Fig. 3) within the *FKBP5* locus (see table in Additional file 1) served as basis for this analysis. Before comparing the three conditions, the collected sequencing data were subjected to three quality control steps in order to ensure accurate assessment of minimal methylation levels as well as small changes between samples.



1. *Bisulfite conversion rate* > 95%. We assessed the bisulfite conversion rate per sample and per amplicon and excluded rates lower than 95% from the analysis.
2. *Removal of PCR artefacts* During the target amplification, the PCR occasionally introduces artefacts presenting non-existent CpG sequences in the target region. They present at very low coverage and extreme levels of methylation (~0 or ~100%). In order to not exclude potential SNPs giving rise to CpGs, we removed artefacts on this basis rather than limiting the analysis to known CpGs according to the reference genome.
3. *Minimum coverage of 1000×* Higher sequencing depth and coverage of the CpGs yields higher accuracy of the methylation quantification. In order to determine the right balance between sequencing depth and thereby cost and sufficient accuracy, we took random subsamples of varying sequencing depth of an in silico created library representing methylation levels from 0 to 100% and assessed the standard deviation for each level of methylation with respect to coverage (Fig. 4a). To find a meaningful cutoff for coverage, we considered the trade-off between sum of the average standard deviation per amplicon (cost) present in various levels of coverage (Fig. 4b). In accordance with previous findings [7], we identified 1000× coverage as a useful cutoff for our analysis, as the gain in accuracy with increasing coverage above this threshold is low and 1000× is rea-

sonable to achieve for a larger locus, e.g., 9 kb in the *FKBP5* panel.

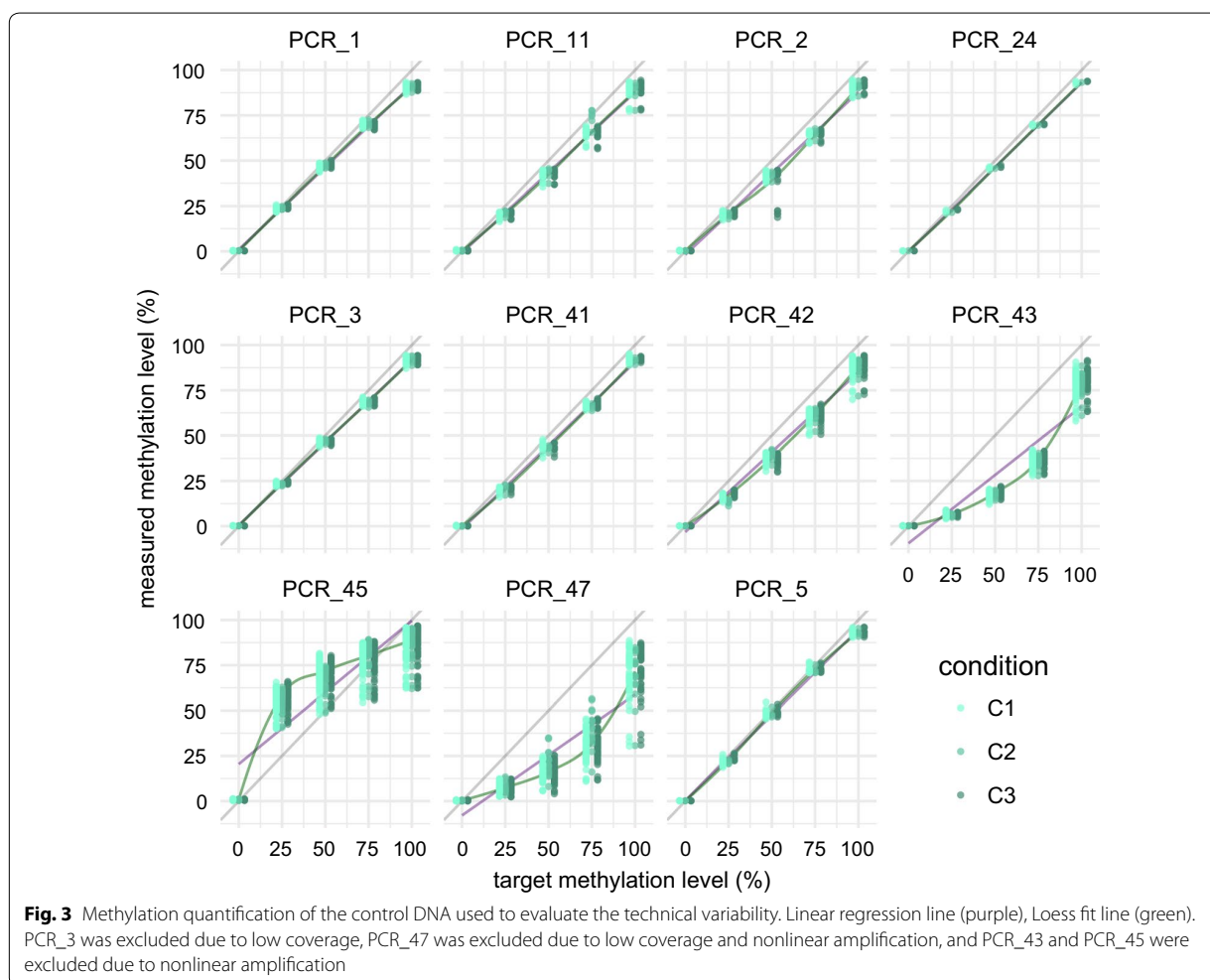
All PCRs for our validation experiment showed bisulfite conversion levels >99%. After QC, a total of 40 CpG spread across 7 amplicons remained in our analysis (1 PCR failed due to coverage < 1000×, 1 showed non-linear amplification and coverage < 1000×, 2 showed non-linear amplification). Methylation levels were very similar between all 3 conditions with an average error of <1% when comparing absolute methylation levels of C2 and C3 versus C1 (Fig. 5b). We calculated the R^2 values for each assessed CpG across the titration levels and used the mean per amplicon to compare the 3 conditions. R^2 is a measure for assessing linearity of amplification of the methylation signal, which is crucial when quantifying methylation changes in, e.g., cohort studies. Again, all conditions showed very high mean R^2 values above 0.99 (Fig. 5a). This confirms that all conditions are suitable for high-accuracy methylation detection. The introduced biases in our workflow, based on the control DNA, are minimal and enable very accurate methylation quantification even without including triplicates for the bisulfite conversion or target amplification. However, opposed to the target amplification, we cannot exclude slightly elevated variance of the bisulfite conversion on non-in vitro methylated DNA from, e.g., patients. Therefore, we chose to use C2 for our HAM-TBS method. While it still maintains a triplicate bisulfite conversion step, it is the most cost-effective of the tested conditions, an important factor when processing many samples from cohort studies.

Comparison of the technical accuracy of pyrosequencing to TBS

Next, we aimed to compare TBS to pyrosequencing, the reference method used for targeted DNA methylation analysis. We assessed the methylation levels of 5 CpGs within PCR_5 and PCR_11 measured by pyrosequencing as well as using HAM-TBS with the C1 protocol. The methylation analysis using pyrosequencing showed a high mean standard deviation of 4.68% with a maximum SD of 14.56%. The analysis using next-generation sequencing with C1 showed a much lower mean standard deviation of 0.72% with a maximum SD of 1.83%. This demonstrates a significantly lower technical variation and therefore higher accuracy when assessing methylation levels using a TBS approach.

Development of an extensive HAM-TBS *FKBP5* panel covering relevant regulatory sites

FKBP5 is an important gene in the field of psychiatry. The gene is larger than 100 kb rendering the assessment of the full locus including the adjacent up- and downstream

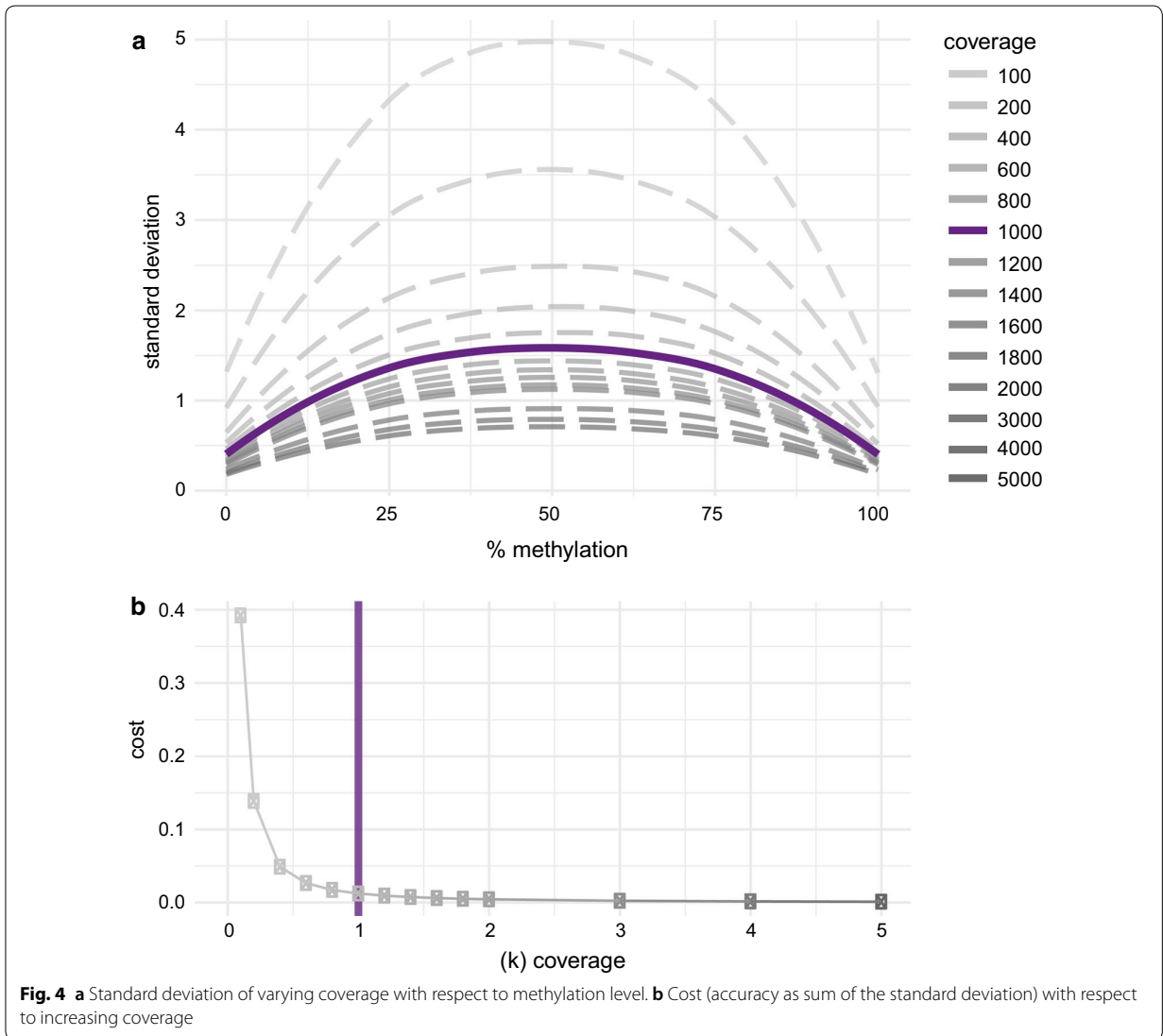


regions unfeasible and too cost intensive for TBS methods. We thus restricted our analysis to functionally relevant sites of interest to ensure compatibility with targeted measurement methods and enable the assessment in large cohorts. To this end, we designed and validated a comprehensive amplicon panel (Fig. 6) including the TSS, TAD boundaries, intergenic and proximal enhancers as well as *GR* and *CTCF* binding sites (see methods for further details). The resulting HAM-TBS *FKBP5* panel is composed of 29 amplicons passing our QC's threshold (described above) and covering 315 CpGs across the locus. The sequencing data showed sufficient bisulfite conversion for all amplicons when performed on control DNA using C2. In total, 27 of the amplicons included in the panel presented good linearity (see figure in Additional file 2) across the assessed methylation levels. Two amplicons located near the TSS showed a mild PCR bias, where methylation levels were lower than expected for the 50% and 75% controls (PCR_7, PCR_9).

These amplicons have a very high CpG content of >25%; hence, CpGs in the primer could not be avoided. It has been previously shown that methylation levels in this region are very low (<5%) across tissues [12], so that any bias at higher methylation levels would not impair accurate quantification of this region. We thus incorporated sites located in this region in the panel, but they should be used with caution if higher methylation levels are observed. PCR_26 of the HAM-TBS *FKBP5* panel is located in the *H19* locus [13] which is an imprinted gene and serves as an internal positive control with an expected methylation level ~50%.

Application and costs

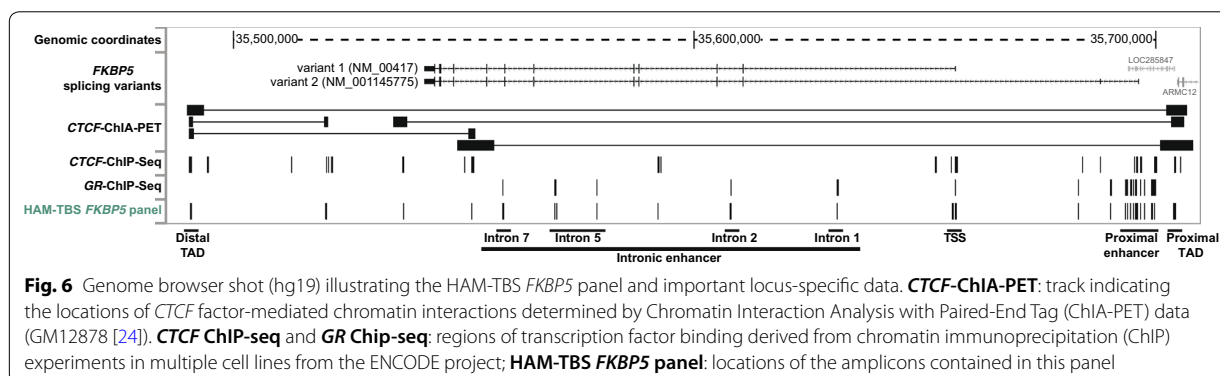
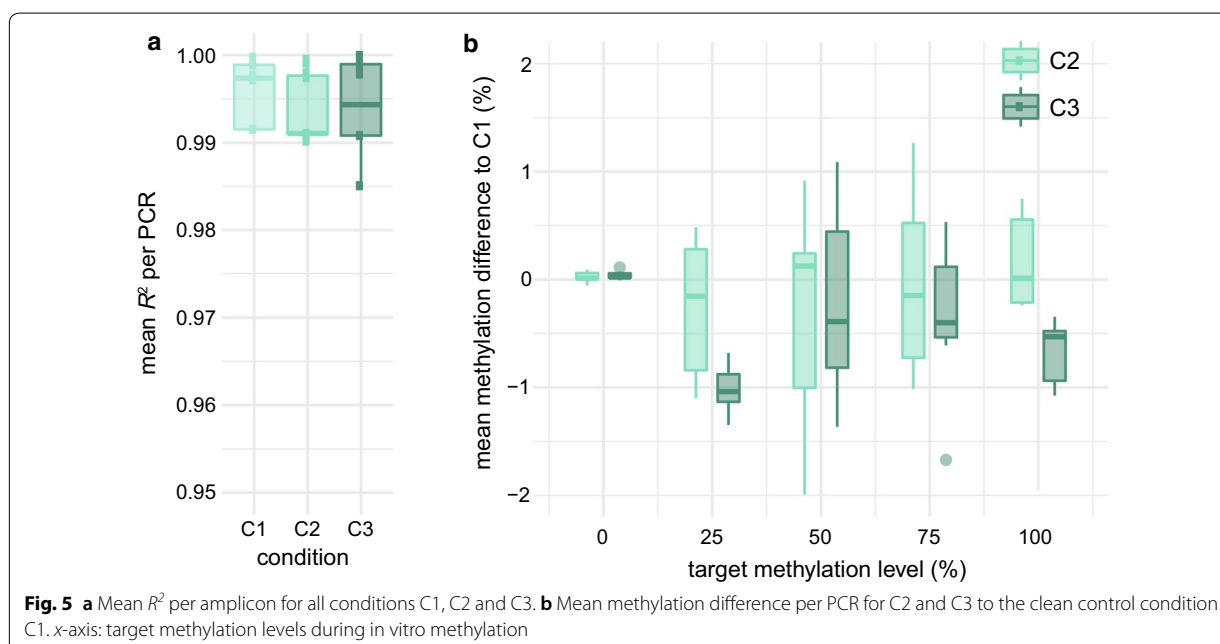
The HAM-TBS method can be multiplexed up to 96 samples in a medium-throughput manner. To demonstrate the applicability of our approach, quality control statistics of data derived from an experiment containing 95 blood samples from patients and the full *FKBP5*



panel of 29 amplicons are described here. After reads mapping and methylation calling, we identified PCR artefacts comprising ~1% of the methylation sites and removed them from the data, and 9 samples in 1 PCR showed insufficient bisulfite conversion rates (<95%) and were also removed. Two loci were identified as SNPs giving rise to a CpG sites in patients. In total, 91% of sample x amplicon data passed our filtering criteria. 27 amplicons passed QC with sufficient coverage and quality in >75% of samples, while two amplicons were dropped due to <1000× coverage (Additional file 3A, B). The control amplicon spanning the H19 imprinted locus for which methylation level is known to be ~50% [14] shows the expected methylation profile

in all samples (Additional file 3C). HAM-TBS approach allowed the quantification of 276 methylation sites for 95 samples in one single MiSeq run.

An assessment of the relative costs for each of the main reagents for this experiment containing 96 samples (95 patients and unmethylated control) with increasing number of amplicons assessed is depicted in Additional file 4. The quantifications using TapeStation and the PCR-free library preparation are the two most cost-intensive steps. The proportion of costs for the amplicon quantification using the TapeStation increases with the higher amount of amplicons investigated, while relative costs for the library preparation and sequencing chemistry decrease with the inclusion of more amplicons.



Discussion

We developed a targeted medium-throughput approach for measuring DNA methylation levels in multiple samples in parallel. This method enables cost-efficient high-resolution methylation measurements of target loci in cohorts of patients and probands at the *FKBP5* gene, a locus with large interest in the psychiatric and psychological community [11]. This cost-efficient, accurate method to determine *FKBP5* methylation levels would thus serve a large number of researchers. Our method is positioned between whole genome bisulfite sequencing and targeted approaches as pyrosequencing. The first is expensive and yields lower coverage and accuracy of single CpGs; the latter only allows to assess very small regions at a time and can generate significant variance between

replicates. HAM-TBS enables the analysis of a targeted but larger region (~10 kb) at high resolution and low costs. DNA methylation studies in large cohorts, investigating the impact of environment or association with disease status in mixed tissues, necessitate high accuracy at single-site resolution. In fact, TBS was able to resolve methylation levels with a mean accuracy of 0.72%. A high level of accuracy was maintained in more cost-efficient approaches using only one PCR amplification round. By pooling triplicate bisulfite treatments prior to PCR amplification, we can account for variance introduced by the bisulfite treatment but also reduce costs and hands-on time during the target amplification.

The accuracy of the method benefits from a PCR-free library preparation and rigorous quality control (prior

evaluation of linear PCR amplification of the target site, bisulfite conversion efficiency >95% and read coverage minimum of 1000×). Nonetheless, a proper assessment of possible amplification biases due to the choice of amplicon location in the design step is critical. Some loci can show nonlinear amplification curves, which renders them inappropriate for methylation quantification. Adjustment of primer design and PCR conditions may help solve this issue, but for some loci optimization may not be possible. For instance, in CpG islands with high CpG density, we observed that amplification curves were not linear, revealing a bias which became more pronounced as the level of methylation increased. Differential methylation results from these sites should be interpreted with caution and perhaps require additional replication. Besides validating each amplicon prior to usage, including controls such as in vitro unmethylated DNA, water and endogenous hemimethylated region, the H19 locus, during each HAM-TBS experiment is important and enables quality checks for each step of the protocol.

Additionally, reaching 1000× coverage is an important step to provide high resolution on methylation changes [8]. However, accurate quantification and pooling of many amplicons across multiple samples while reaching sufficient coverage of all regions has limitations. In theory, even though the MiSeq can handle a much higher loading factor (amplicons × samples) of almost 20,000 (disregarding uneven pooling of libraries, filtering of reads due to low quality or high amounts of PhiX), a maximum of 2500–3000 has proven to be feasible with minimal dropout rates. Assuming multiplexing of 96 samples and 25 amplicons at an average length of 400 bp, a region of approximately 10 kb can be comfortably covered with this approach. Notably, we streamlined the method to handle loading factors >2000 by implementation of Agilent's TapeStation and a pipetting robot for quantification and pooling of amplicons. Besides the throughput, this improves the robustness of the workflow. Our approach is designed to match the specifications of the Illumina MiSeq with its ability to run for 600 cycles resulting in 300 bp-long paired-end reads. This enables full-length coverage of amplicons up to a length of 600 bp. While our approach can be applied to different sequencers, such as the Illumina HiSeq for example, it would be necessary to design shorter amplicons due to the current limits of the sequencing chemistry. Using another sequencer, it is important to mention the index hopping phenomenon on the Illumina platforms [15]. It is less present on the MiSeq compared to other machines with pattern flow cells as our data show consistent levels of methylation close to 0% across all in vitro unmethylated control samples indicating no issue with

this specific bias. Nonetheless, it should be kept in mind that approaches like unique dual indexes when available or Illumina's Free Adapter Blocking Reagent are recommendable and gain importance, especially when using a different Illumina sequencer.

In the past years, only few TBS methods have been developed [8–10] with different methodological foci. Thus far, Bernstein et al. [10] allows a panel of 48 indices, while the approach by Chen et al. [9] could allow for a multiplexing rate of 1536 samples due to custom-made barcodes, but in practice only 478 have been used to date. In the latter method, the high multiplexing capacity comes at the cost of an additional PCR step potentially introducing additional bias. Moreover, increasing the number of samples needs to be weighed against the size of the target region in order to ensure sufficient coverage. We identified 1000× coverage as an optimal cutoff in terms of accuracy and cost in agreement with a publication by Masser et al. [8]. In the above-described study by Chen et al. [9], 100× was used as minimum cutoff. Based on our in silico analysis (Fig. 4a), this would lead to less accurate quantification of methylation levels. Besides the number of samples that can be processed, the size of the region of interest is also an important factor to be considered. The method by Masser et al. [8] has been applied to 2 amplicons (233 and 320 bp), while Chen et al. enable the assessment of larger loci around 10 kb—comparable to our HAM-TBS approach. Lastly, amplification-based library preparation methods have been adapted by most TBS approaches. At this point, HAM-TBS utilizes a PCR-free library preparation to avoid adding amplification biases.

Finally, using the optimized HAM-TBS workflow, we designed a panel comprising 29 amplicons to accurately assess methylation within the *FKBP5* locus using HAM-TBS. This panel covers ~9 kb and targets important regulatory regions of the *FKBP5* gene including the TSS, intergenic and proximal enhancers and TAD boundaries including *CTCF* binding sites. The HAM-TBS method and the *FKBP5* panel present valuable tools for epigenetic studies in which a highly accurate assessment of methylation levels is critical such as GxE studies in psychiatric research. It allows cost-efficient quantification of methylation in larger cohorts with optimized hands-on time due to automatization.

Conclusion

The presented method HAM-TBS offers a robust and low-cost method for researchers interested in DNA methylation measurements of specific target regions. In addition, we supply a validated panel of 29 amplicons to assess methylation levels of important regulatory regions

in the *FKBP5* locus, a gene of great interest in the field of psychiatry.

Methods

Generation of in vitro methylated control DNA

All primers designed for bisulfite PCR were first tested on in vitro methylated DNA to assess amplification efficiency and bias. For PCRs within the *FKBP5* gene, an in vitro methylated BAC (RP11-282I23, BACPAC) was used to generate control DNA. For PCRs outside the *FKBP5* locus (PCR_26, PCR_34, PCR_35), genomic DNA extracted from whole blood was amplified using the REPLI-g Mini Kit (QIAGEN GmbH, Hilden, Germany) to generate unmethylated DNA. 100% methylated DNA was achieved using in vitro methylation with M.SssI methyltransferase. After a first incubation (3 µg DNA, 0.5 µl SAM (32 mM), 1 µl M.SssI (20 U/µl, 40 µl NEB buffer 2 [10×], diluted with ddH₂O up to 400 µl) of 4 h at 37°C, 1 µl of M.SssI (20 U/µl) and 1 µl of SAM (32 mM) were added, and a second 4-h incubation was performed. Subsequently, the reaction was purified using the nucleotide removal kit (QIAGEN GmbH, Hilden, Germany). In vitro methylation was repeated with the eluted DNA for a second time. 25, 50 and 75% methylated control DNA was obtained by mixing 0 and 100% DNAs. In vitro methylation of control DNA was checked via pyrosequencing.

Bisulfite treatment of DNA

We used the EZ DNA Methylation Kit (Zymo Research, Irvine, CA) in column and plate format depending on the amount of DNA and throughput needed. Between 200 and 500 ng was used as input DNA and processed according to the manufacturer's instructions. DNA was eluted twice in 10 µl elution buffer which recovered over 90% of the input DNA after bisulfite conversion when using the column format. In order to quantify bisulfite treated DNA, we use a spectrophotometer with RNA quantification settings.

Target enrichment and amplicon pooling

The amplification of target locations from converted DNA (20 ng per amplicon) was achieved using the TaKaRa EpiTaq HS Polymerase (Clontech, Mountain View, CA; final concentration: 0.025 U/l), bisulfite-specific primers (final concentration of each primer: 0.4 M) and a touchdown cycling protocol with 49 cycles [for more details (see table in Additional file 5 and section HAM-TBS *FKBP5* panel)]. The amplicons of all PCR reactions were quantified using the Agilent 2200 TapeStation (Agilent Technologies, Waldbronn, Germany) and equimolar pooled with the Hamilton pipetting robot. After speed-vacuum and resuspension in 50 µl, a double-size selection was applied using Agencourt AMPure XP beads

(Beckman Coulter GmbH, Krefeld, Germany) to remove excess of primers and genomic DNA.

Control samples

For every TBS run, we included three different controls. First, up to three water controls in order to monitor cross-contamination with DNA and detect if the plate was accidentally rotated. Second, an unmethylated control DNA as a positive control and to detect failed steps throughout the workflow. And third, the H19 locus which is an imprinted region and presents with methylation levels ~50% as a positive control for bisulfite conversion in genomic DNA and detect outliers in patient samples. An amplicon located at this locus is incorporated in the *FKBP5* panel.

Library preparation and sequencing

For library generation, Illumina TruSeq DNA PCR-Free HT Library Prep Kit (Illumina, San Diego, CA) was used according to the manufacturer's standard protocol and obtained high-quality libraries using 500 ng of starting material (during optimization, input amounts as low as 100 ng were tested and showed no loss of quality on the QC level). Qubit 1.0 (Thermo Fisher Scientific Inc., Schwerte, Germany) was used for quantification, Agilent's 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany) for quality assessment and Kapa HIFI Library quantification kit (Kapa Biosystems Inc., Wilmington, MA) for final quantification before pooling. Libraries were pooled equimolarly. Sequencing of the libraries was performed on an Illumina MiSeq using Reagent Kit v3 (Illumina, San Diego, CA; 600 cycles) in paired-end mode, with 30% PhiX added.

Sequencing data processing

First, read quality was verified using FastQC [16]. Adapter sequences were trimmed using cutadapt v.1.9.1 [17]. For alignment to a restricted reference of hg19 based on the PCR locations, Bismark v.0.15.0 [18] was used. Due to the 600-cycle sequencing chemistry, PCRs shorter than 600 bp produce overlapping paired-end reads. Using an in-house developed Perl script, we trimmed low-quality overlapping ends. Quantification of methylation levels in CpG and CHH context was performed using the R package methylKit [19] with a minimum quality score of 20. The methylation calls were subjected to 3 quality control steps. First, we considered CHH levels for each sample and excluded samples if the conversion was less than 95% efficient. Second, we filtered PCR artefacts introduced by PCR amplification errors giving rise to CpG sites in some reads. As we do not restrict the analysis to known CpG sites, every read indicating the presence of a CpG will be considered and the information extracted. These

artefacts mostly present at very low levels of coverage and 0 or 100% methylation. Lastly, according to our coverage cutoff, we excluded CpG sites supported by less than 1000 reads. Subsequent analysis comparing methylation levels from the conditions C1, C2 and C3 as well as data from pyrosequencing was performed in R.

Coverage considerations

When performing a sequencing experiment, one will usually sequence part of the generated library and quantify the methylation levels on this basis rather than sequence the whole library to see the true level within. Therefore, each sequencing experiment corresponds to drawing a random subset of a certain size (sequencing depth) of the whole library and can be viewed as a subsampling problem. Depending on the sequencing depth, this will yield a different level of accuracy of the methylation levels. We created a dataset simulating CpGs methylated at levels from 0 to 100% supported by 100,000 “fragments” each. Therefore, e.g., for 10% methylation level, a set 10,000 Cs and 90,000 Ts was created. Accordingly, sets for 0–100% methylation were created. Using a bootstrapping approach, we drew 1000 random subsets of varying sequencing coverage (100, 200, 400, ..., 2000, 3000, 4000, 5000) from each set representing a certain level of methylation and the standard deviation (SD) was calculated. As a proxy for the increase in accuracy versus increase in sequencing depth (costs), the combined SD was divided by the sequencing depth. Of note, this is in concordance with results from the same analysis on highly covered amplicon data from our laboratory (data not shown).

Pyrosequencing

Methylation analysis by pyrosequencing of 5 CpGs covered within PCR_5 (CpG 35607969, CpG 35608022) and PCR_11 (CpG 35690280, CpG 35690318, CpG 35690365) was performed in triplicates on BAC control DNA. Bisulfite conversion of in vitro methylated control DNA was applied as described above. Target enrichment by PCR was achieved with a biotinylated reverse primer but otherwise performed as described above. Pre-treatment of PCR amplicons was facilitated with the PyroMark Q96 Vacuum Workstation (QIAGEN GmbH, Hilden, Germany). Sequencing of *FKBP5* CpGs was performed on a PyroMark Q96 ID system using PyroMark Gold Q96 reagents (QIAGEN GmbH, Hilden, Germany) and sequencing primers according to Klengel et al. [12]: P4 S1 (TTTGGAGTAGTAGGTAAA) GRE3 S1 MPI (GGGAATATGAGGTTG). The PyroMark Q96 ID Software 2.5

(QIAGEN GmbH, Hilden, Germany) was used for data analyses.

HAM-TBS *FKBP5* panel

We designed 29 primer pairs (see table in Additional file 5) using BiSearch [20, 21] targeting the *FKBP5* locus. Initially, 32 PCRs were included, but 3 PCRs were not selected for the panel due to QC failure. The excluded amplicons showed nonlinear amplification due to an elevated GC content in the region. Positions of amplicons covering glucocorticoid response elements (GREs) were selected from Klengel et al. [12] and the *GR* ChIP-Seq from the ENCODE project [22]. Amplicons covering *CTCF* binding sites were selected using HI-C peaks [23], *CTCF*-ChIA-Pet interactions from a lymphoblastoid cell line (GM12878, Tang et al. [24]) and *CTCF* ChIP-Seq information from the ENCODE project [22]. Lastly, amplicons located near the TSS were included in the panel. Only primers without CpGs in their sequence were chosen, with the exception of 2 amplicons close to the TSS where this could not be avoided due to the high CpG content of the region. The selected amplicons ranged from 200 to 450 bp in length.

Additional files

Additional file 1. A table containing the genomic coordinates of a smaller panel of 11 amplicons located within the *FKBP5* locus. These amplicons were used to assess the source of potential biases as well as variability between replicates.

Additional file 2. A figure displaying the bias assessment for all amplicons comprising the *FKBP5* HAM-TBS panel.

Additional file 3. A figure displaying the QC statistics of a HAM-TBS experiment with 95 samples using the *FKBP5* panel and methylation levels of the H19 locus.

Additional file 4. A figure displaying the relative and absolute costs for a HAM-TBS experiment with 96 samples.

Additional file 5. A table containing the genomic coordinates of the *FKBP5* HAM-TBS panel including primer sequences and cycling conditions.

Authors' contributions

TW, SR, NP, EBB contributed to experimental design. TW, SS, MK performed wet lab work. SR performed the data analyses. SR, TW, NP, EBB prepared the manuscript. All authors read and approved the final manuscript.

Author details

¹ Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Kraepelinstr. 2-10, 80804 Munich, Germany. ² Department of Psychiatry and Behavioral Sciences, Emory University Medical School, 12 Executive Park Dr NE #200, Atlanta, GA 30329, USA. ³ Faculty of Health Sciences, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada. ⁴ Healthy Starts Theme, BC Children's Hospital Research Institute, 938 West 28th Avenue, Vancouver, BC V5Z 4H4, Canada.

Acknowledgements

The authors would like to thank Jessica Keverne for professional English editing and formatting and Monika Rex-Haffner for her help regarding the sequencing of the libraries and Stoyo Karamihalev for graphics support.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This study was funded by the BMBF Grant Berlin-LCS (FKZ 01KR1301B) to EB and an ERC starting Grant (GxE molmech, Grant 281338) within the FP7 funding scheme of the EU to EB and fellowship from Canadian Institute of Health Research (CIHR) to NP.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 4 April 2018 Accepted: 28 June 2018

Published online: 04 July 2018

References

- Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 2002;16:6–21.
- Ehrlich M, Wang RY. 5-methylcytosine in eukaryotic DNA. *Science.* 1981;212:1350–7.
- Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol.* 2010;28:1057–68.
- Weng YL, An R, Shin J, Song H, Ming GL. DNA modifications and neurological disorders. *Neurotherapeutics.* 2013;10:556–67.
- Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, et al. Factors underlying variable DNA methylation in a human community cohort. *Proc Natl Acad Sci USA.* 2012;109(2):17253–60.
- Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA methylation in Newborns and Maternal Smoking in Pregnancy: genome-wide Consortium Meta-analysis. *Am J Hum Genet.* 2016;98:680–96.
- Ziller MJ, Hansen KD, Meissner A, Aryee MJ. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nat Methods.* 2015;12:230–2.
- Masser DR, Berg AS, Freeman WM. Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. *Epigenet Chromatin.* 2013;6:33.
- Masser DR, Stanford DR, Freeman WM. Targeted DNA methylation analysis by next-generation sequencing. *J Vis Exp.* 2015. <https://doi.org/10.3791/52488>.
- Bernstein DL, Kameswaran V, Le Lay JE, Sheaffer KL, Kaestner KH. The BisPCR(2) method for targeted bisulfite sequencing. *Epigenet Chromatin.* 2015;8:27.
- Chen GG, Gross JA, Lutz PE, Vaillancourt K, MauSSION G, Bramouille A, et al. Medium throughput bisulfite sequencing for accurate detection of 5-methylcytosine and 5-hydroxymethylcytosine. *BMC Genom.* 2017;18:96.
- Zannas AS, Wiechmann T, Gassen NC, Binder EB. Gene-stress-epigenetic regulation of FKBP5: clinical and translational implications. *Neuropsychopharmacology.* 2016;41:261–74.
- Klengel T, Mehta D, Anacker C, Rex-Haffner M, Pruessner JC, Pariante CM, et al. Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Nat Neurosci.* 2013;16:33–41.
- Huang RC, Galati JC, Burrows S, Beilin LJ, Li X, Pennell CE, et al. DNA methylation of the IGF2/H19 imprinting control region and adiposity distribution in young adults. *Clin Epigenet.* 2012;4(1):21.
- Illumina, Inc. Effects of index misassignment on multiplexing and downstream analysis. <https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf>.
- Andrews SR. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>. 2010.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:10–2.
- Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics.* 2011;27:1571–2.
- Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 2012;13:R87.
- Aranyi T, Váradi A, Simon I, Tusnády GE. The BiSearch web server. *BMC Bioinf.* 2006;7:431.
- Tusnády GE, Simon I, Váradi A, Aranyi T. BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic Acids Res.* 2005;33:e9.
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489:57–74.
- Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* 2015;159:1665–80.
- Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell.* 2015;163:1611–27.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Publication III

Identification of dynamic glucocorticoid induced methylation changes at the FKBP5 locus

Authors:

Tobias Wiechmann¹, **Simone Röh**¹, Susann Sauer¹, Darina Czamara¹, Janine Arloth^{1,7}, Maik Ködel¹, Madita Beintner¹, Lisanne Knop¹, Andreas Menke^{2,3}, Elisabeth B. Binder^{1,4} and Nadine Provençal^{1,5,6}

¹ Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany.

² Department of Psychiatry, Psychosomatics and Psychotherapy, University Hospital of Wuerzburg, Wuerzburg, Germany.

³ Comprehensive Heart Failure Center, University Hospital of Wuerzburg, Wuerzburg, Germany.

⁴ Department of Psychiatry and Behavioral Sciences, Emory University Medical School, Atlanta, GA, USA.

⁵ Faculty of Health Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

⁶ BC Children's Hospital Research Institute, Vancouver, BC, Canada.

⁷ Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany.

Publication details:

Published in *Clinical Epigenetics* 11:83 on May 23rd, 2019

<https://doi.org/10.1186/s13148-019-0682-5>


© The Author(s) 2018. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

RESEARCH

Open Access

Identification of dynamic glucocorticoid-induced methylation changes at the *FKBP5* locus



Tobias Wiechmann¹ , Simone Röh¹, Susann Sauer¹, Darina Czamara¹, Janine Arloth^{1,7}, Maik Ködel¹, Madita Beintner¹, Lisanne Knop¹, Andreas Menke^{2,3}, Elisabeth B. Binder^{1,4*} and Nadine Provençal^{1,5,6*}

Abstract

Background: Epigenetic mechanisms may play a major role in the biological embedding of early-life stress (ELS). One proposed mechanism is that glucocorticoid (GC) release following ELS exposure induces long-lasting alterations in DNA methylation (DNAm) of important regulatory genes of the stress response. Here, we investigate the dynamics of GC-dependent methylation changes in key regulatory regions of the *FKBP5* locus in which ELS-associated DNAm changes have been reported.

Results: We repeatedly measured DNAm in human peripheral blood samples from 2 independent cohorts exposed to the GC agonist dexamethasone (DEX) using a targeted bisulfite sequencing approach, complemented by data from Illumina 450K arrays. We detected differentially methylated CpGs in enhancers co-localizing with GC receptor binding sites after acute DEX treatment (1 h, 3 h, 6 h), which returned to baseline levels within 23 h. These changes withstood correction for immune cell count differences. While we observed main effects of sex, age, body mass index, smoking, and depression symptoms on *FKBP5* methylation levels, only the functional *FKBP5* SNP (rs1360780) moderated the dynamic changes following DEX. This genotype effect was observed in both cohorts and included sites previously shown to be associated with ELS.

Conclusion: Our study highlights that DNAm levels within regulatory regions of the *FKBP5* locus show dynamic changes following a GC challenge and suggest that factors influencing the dynamics of this regulation may contribute to the previously reported alterations in DNAm associated with current and past ELS exposure.

Keywords: DNA methylation, *FKBP5*, Glucocorticoid receptor, Early-life stress, Targeted bisulfite sequencing, Dexamethasone

Background

Epidemiological studies indicate a combined contribution of genetic and environmental factors in the risk for psychiatric diseases, which converge to alter gene regulation and consequently cell function [1]. Evidence suggests that epigenetic mechanisms play a major role in embedding environmental risk, including early-life adversity, but our understanding of the underlying mechanisms is limited. Epigenetic mechanisms encompass post-translational modifications of histone proteins and

chemical modifications of single nucleotides (most commonly in the form of methylation at cytosine guanine dinucleotides (CpGs)), which alter chromatin structure, and thus accessibility of the DNA to transcriptional regulators.

Even DNA methylation (DNAm), a stable chemical modification, undergoes highly dynamic regulation in post-mitotic cells. This property makes DNAm a suitable molecular mechanism to encode the impact of environmental cues in post-mitotic tissue [2, 3]. A mechanism that likely contributes to such dynamic, environmentally triggered DNAm changes is transcription factor-mediated DNA demethylation [4]. One example is local demethylation of glucocorticoid response elements (GREs) following activation of the glucocorticoid receptor (GR), a nuclear

* Correspondence: binder@psych.mpg.de; nadine_provençal@sfu.ca

¹Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Kraepelinstr. 2-10, 80804 Munich, Germany

Full list of author information is available at the end of the article



transcription factor [5]. The GR is activated by the glucocorticoid (GC) cortisol, a major mediator of the stress response.

Stress, especially in the form of early adverse life trauma, is a major environmental risk factor for psychiatric disorders [1, 6, 7]. Excessive GC release after stress exposure may induce long-lasting DNAm changes, thereby contributing to the biological embedding of risk trajectories. The mechanism of GR-induced local demethylation is not fully understood, but activation of DNA repair machinery is implicated. Demethylation of GREs facilitates the transcriptional effects of the GR on the target gene [8, 9].

FKBP5 is a stress-responsive gene and co-chaperone protein of GR. Increased activation of this gene by genetic or epigenetic factors has been repeatedly associated with increased stress-sensitivity and risk for psychiatric disorders in both animal and human studies (see [10, 11] for review). We have previously reported on GR-sensitive CpGs in GREs of the *FKBP5* locus. These are located in a functional GRE in intron 7 of the gene. Chromatin conformation capture experiments confirmed an interaction of this intronic enhancer with the transcription start site (TSS) of *FKBP5*. Reporter gene assays also demonstrated that higher DNAm of this enhancer region was associated with lower transcriptional activation of *FKBP5* by GCs [12]. Relative reduction of DNAm in this region has been reported both in peripheral blood and buccal cells of adults as well as children exposed to childhood trauma and in a hippocampal neuronal progenitor cell (HPC) line following exposure to GCs [12–16]. Changes in DNAm following exposure to child abuse seemed to be accentuated in individuals carrying the minor allele of a functional genetic variant in this locus (rs1360780). This variant, located in close proximity to a GRE in intron 2, alters the 3D conformation of the locus. The minor allele generates a TATA-box binding site which allows binding of this intron enhancer to the transcription start site. This is also associated with higher *FKBP5* induction following GR activation. We and others have shown that increased *FKBP5* leads to reduced GR sensitivity and impaired negative feedback regulation of the stress hormone axis [17, 18]. In fact, minor allele carriers have repeatedly been shown to have prolonged cortisol release following stress exposure [11]. Finally, this functional allele has consistently been shown to increase risk for a range of psychiatric disorders with exposure to early adversity [11, 19–21], suggesting that gene x environment interactions at the level of epigenetic regulation may contribute to disease risk.

These studies suggest that the CpGs associated with early trauma exposure may also be responsive to GCs and that increased GR activation with trauma may lead to DNAm changes of the sites. However, direct evidence

for this is so far missing. Furthermore, due to the limitation of the previously used pyrosequencing-based DNAm assessment of these enhancers, only a small number of CpGs had been investigated. Transcriptional regulatory sites of *FKBP5* are distributed throughout the locus and include several upstream, downstream, and intronic enhancer regions with GREs [22] as well as CCCTC-binding factor (CTCF) sites in addition to the TSS. CTCF creates boundaries between topologically associating domains (TADs) in chromosomes, and within these domains, CTCF facilitates interactions between transcription regulatory sequences [23, 24]. The extent of GR-associated DNAm changes in different categories of regulatory elements within the *FKBP5* locus has not yet been explored.

Here, we investigate the changes of DNAm following exposure to the selective GR agonist dexamethasone (DEX) in peripheral blood cells over 24 h and in relation to rs1360780 genotype, in two independent cohorts. DNAm levels were assessed using a high-accuracy methylation measurements via targeted bisulfite sequencing (HAM-TBS) approach [25], which extensively covers CpG sites located in the different categories of regulatory elements in the *FKBP5* locus. The changes are also compared to data generated by the widely used Illumina methylation arrays.

Results

DEX-induced dynamic changes at the *FKBP5* locus in human peripheral blood (study 1)

In order to test if GR activation is associated with changes in DNAm in vivo, we first analyzed serial blood samples from 19 subjects exposed to a single oral dose (1.5 mg) of DEX (see Table 1 for demographic details).

DEX-induced changes in *ACTH*, cortisol, and *FKBP5* mRNA levels

Analysis of serum adrenocorticotropin (ACTH) and cortisol levels showed the expected suppression following

Table 1 Description of study 1 and 2 subjects

	Study 1	Study 2
Samples	19	89
Male	19	67
Female	0	22
rs1360780 genotype	CC = 6; CT = 6; TT = 7	CC = 50; CT = 30; TT = 9
Time points of blood draw after DEX	0 h, 1 h, 3 h, 6 h, 23 h	0 h, 3 h, 18–24 h
Age (mean ± SD)	25.4 ± 2.9	41.6 ± 14.0
BMI (mean ± SD)	N/A	25.1 ± 3.8
Smoking score (mean ± SD)	N/A	−0.6 ± 4.9
Major depressive disorder	0	59 (M = 38; F = 21)

DEX administration with maximal effects observed at 3 and 6 h post-treatment. In addition, DEX induced a 4.2- and 4.0-fold increase in *FKBP5* mRNA levels after 3 and 6 h of treatment, respectively, and returned to baseline level after 23 h (Fig. 1a).

DEX-induced transient DNAm changes—dynamics

DNAm was analyzed using our HAM-TBS technique [25], where a total of 25 amplicons covering 228 CpGs at 5 time points from baseline (0 h) to 1, 3, 6, and 23 h following DEX administration passed QC. Three amplicons located in the proximal enhancer did not pass QC due to increased CHH methylation levels (PCR18) or low coverage (PCR 28 and PCR29; < 1000 reads, Additional file 6: Table S1). DNAm analysis across the loci at baseline revealed low methylation at the TSS and higher methylation within the gene body and 3' and 5' flanking regions (Fig. 1b “% Methylation baseline” track).

Following an acute dose of DEX, 44 CpG sites showed significant changes in DNAm over all time points ($FDR \leq 0.05$ and absolute delta methylation (T_i -baseline) $\geq |1\%$); Fig. 1b “Max. Δ % methylation” track, shaded regions; Additional file 7: Table S2). Significant DEX-induced differential DNAm was seen as early as 1 h after treatment ($n = 17$ sites, mean absolute Δ methylation = $|2.4\%$), and the largest effects were observed after 3 and 6 h ($n = 40$ sites, mean absolute Δ methylation = $|4.4\%$ and $|5.4\%$, respectively) with a range from -17 to $+0\%$. Seventy-four percent of the sites, however, showed decreases in DNAm levels following DEX treatment ranging from -17 to -1% compared to baseline. For the majority of the sites, DNAm levels returned to baseline after 23 h of treatment while only 8 sites remained differentially methylated at $FDR < 0.05$ with a small change compared to baseline (mean absolute Δ methylation = $|1.8\%$).

DEX-induced transient DNAm changes—localization

DEX-induced differentially methylated CpG sites (DMCs) were found in the proximal and intronic enhancers and co-localized with ENCODE GR binding sites and those located at the chromatin interaction blocks overlapped with ENCODE CTCF binding sites (examples are shown in Fig. 1c). Within the 82 CpGs analyzed surrounding the TSS, no DMCs were observed. Out of the 129 CpGs located in GR binding sites, 36 (28%) showed DEX-induced changes whereas only 8 (10%) out of 83 sites located in CTCF binding sites showed changes after DEX. Most of the DMCs located in GR binding sites ($n = 30$) showed reduction of DNAm following DEX (mean Δ methylation $-3.8 \pm 3.3\%$) with the exception of 6 sites located in the proximal enhancer ($n = 4$), intron 5 ($n = 1$) and intron 2 ($n = 1$) showing increased DNAm (mean Δ methylation $+2.9 \pm 1.3\%$).

DMCs within CTCF binding sites located in TAD boundaries and intron 3 showed increase in DNAm ($n = 5$ sites, mean Δ methylation $+3.6 \pm 3.3\%$) whereas those located in the proximal enhancer showed decrease in DNAm ($n = 3$ sites; mean Δ methylation $-3.0 \pm 1.1\%$). Demethylated CpGs in the proximal enhancer overlap with both GR and CTCF binding sites (see Fig. 1c).

To assess whether changes in DNAm might directly affect binding of GR and CTCF to DNA, we mapped the changes to their relative distance to GR and CTCF consensus binding motifs. DNAm of CpG sites within these motifs has previously shown to impair or decrease transcription factor binding [27, 28]. We used predicted DNA binding motif locations for GR and CTCF from [29] (<http://compbio.mit.edu/encode-motifs/>). Selecting CpGs within ± 50 bp of the consensus motif sequences ($n = 16$ for GR and $n = 9$ for CTCF), we observed that CpG sites directly in CTCF motifs consistently displayed very low DNAm levels ($0.57 \pm 0.10\%$) whereas those in NR3C1 motifs showed intermediate levels at baseline with a high variation ($39.98 \pm 18.43\%$; Fig. 2a). For CTCF motif regions, higher DNAm was observed at more distal sites at the edges of the motif. DEX-induced DMCs were found directly in GR motifs ($n = 4$) whereas none were observed within CTCF motif but at the edges of this motif ($n = 2$, Fig. 2b).

Increases in *FKBP5* mRNA levels occurred in parallel with the decrease of DNAm for 10 DMCs located within 50 bp of GR binding motifs (Fig. 2c and Additional file 6: Table S1).

Validation of DEX-induced DNAm changes in blood in an independent sample ($n = 89$, study 2)

Using study 2, see Table 1 for demographic details, we replicated the findings in an independent sample. Similar to study 1 and has been reported previously [30], DEX treatment induced a significant decrease in CORT and ACTH levels as well as an increase in *FKBP5* mRNA levels (Fig. 3a).

Ten amplicons covering 50 CpG sites (including 25 with significant DEX effect in study 1) were selected. Of these 50 sites, 21 showed significant changes ($FDR \leq 0.05$ and absolute Δ methylation $\geq |1\%$) in DNAm after DEX treatment validating 19 sites from the first study (Fig. 3b). Similar to the effects observed in the first study, DNAm changes were seen after 3 h of treatment (mean absolute Δ methylation = $|4.3\%$), and for most sites ($n = 13$), DNAm returned to baseline after 24 h. Eight DMCs remained significant after 24 h of treatment but showed a much smaller effect (mean absolute Δ methylation = $|1.8\%$). As observed before, the majority of sites (76%) show decrease in DNAm levels following DEX treatment with a range from -10 to -1% compared to baseline.

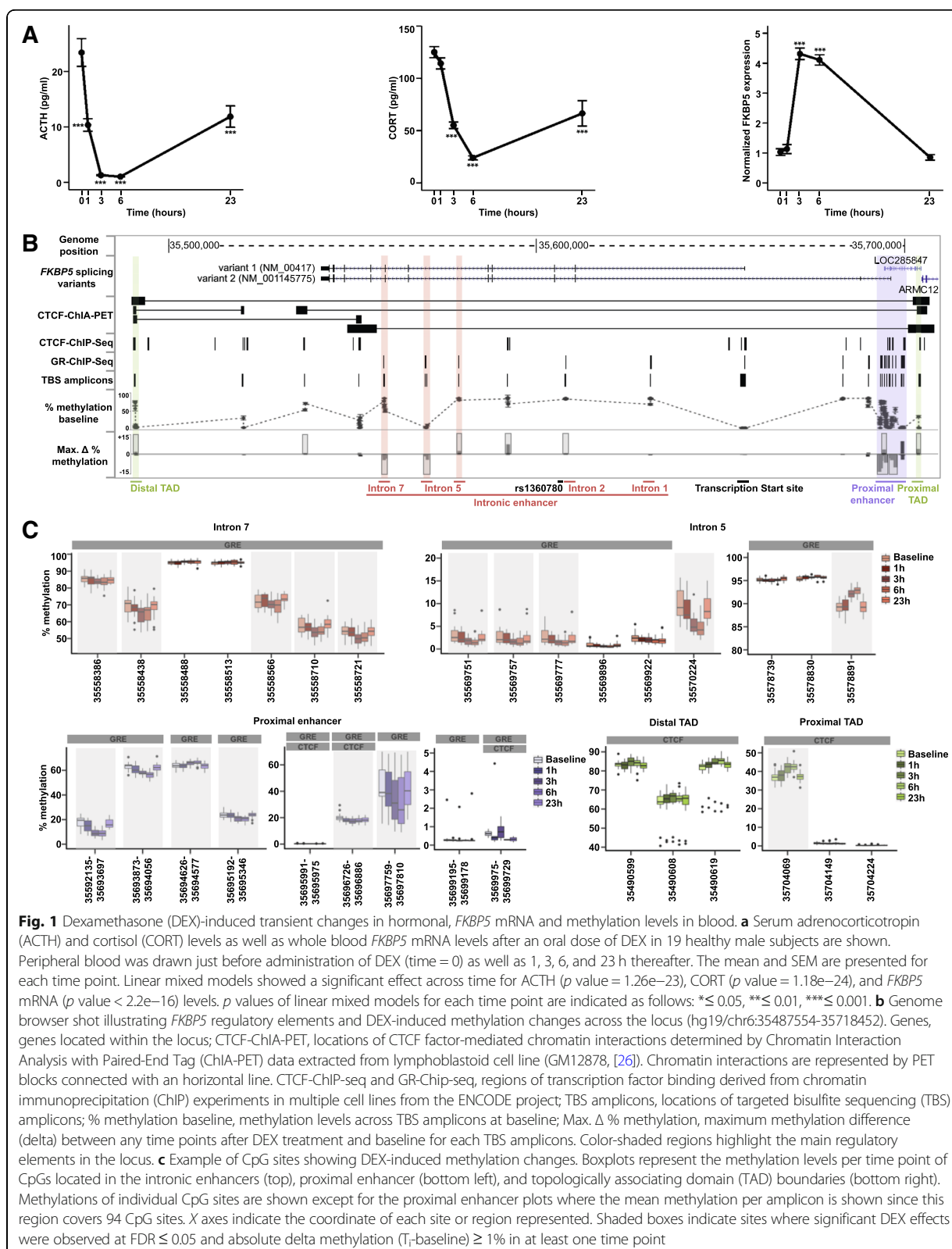
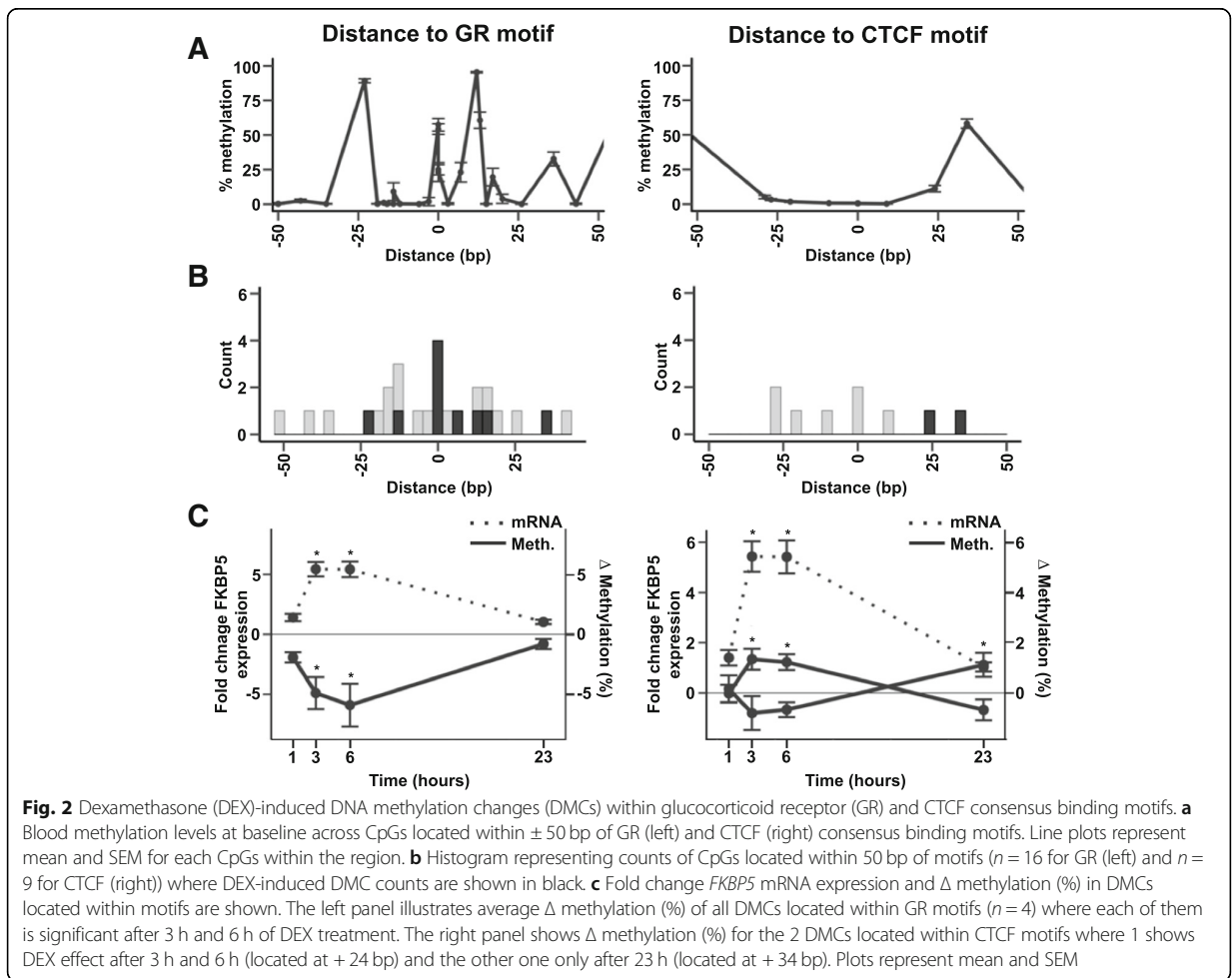


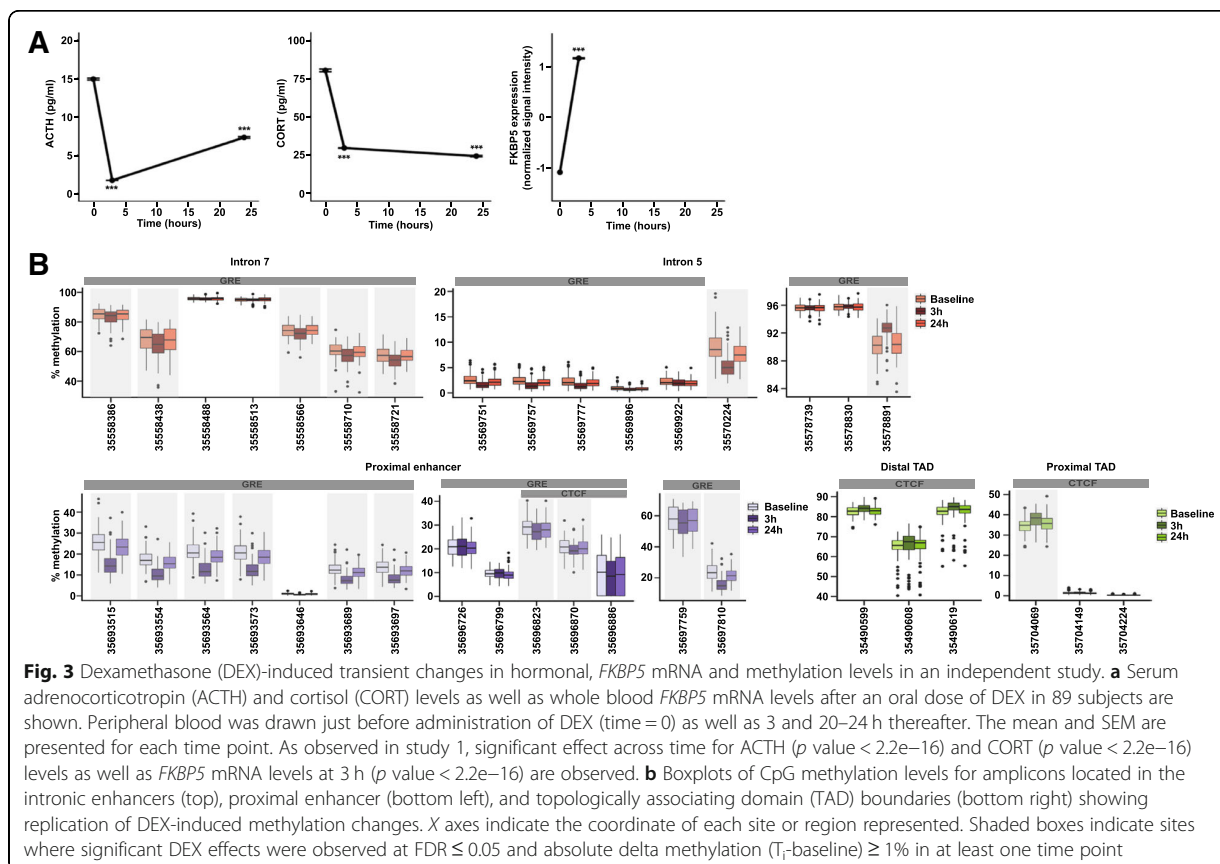
Fig. 1 Dexamethasone (DEX)-induced transient changes in hormonal, *FKBP5* mRNA and methylation levels in blood. **a** Serum adrenocorticotropin (ACTH) and cortisol (CORT) levels as well as whole blood *FKBP5* mRNA levels after an oral dose of DEX in 19 healthy male subjects are shown. Peripheral blood was drawn just before administration of DEX (time = 0) as well as 1, 3, 6, and 23 h thereafter. The mean and SEM are presented for each time point. Linear mixed models showed a significant effect across time for ACTH (p value = 1.26×10^{-23}), CORT (p value = 1.18×10^{-24}), and *FKBP5* mRNA (p value < 2.2×10^{-16}) levels. p values of linear mixed models for each time point are indicated as follows: * ≤ 0.05 , ** ≤ 0.01 , *** ≤ 0.001 . **b** Genome browser shot illustrating *FKBP5* regulatory elements and DEX-induced methylation changes across the locus (hg19/chr6:35487554-35718452). Genes, genes located within the locus; CTCF-ChIA-PET, locations of CTCF factor-mediated chromatin interactions determined by Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) data extracted from lymphoblastoid cell line (GM12878, [26]). Chromatin interactions are represented by PET blocks connected with a horizontal line. CTCF-ChIP-seq and GR-ChIP-seq, regions of transcription factor binding derived from chromatin immunoprecipitation (ChIP) experiments in multiple cell lines from the ENCODE project; TBS amplicons, locations of targeted bisulfite sequencing (TBS) amplicons; % methylation baseline, methylation levels across TBS amplicons at baseline; Max. Δ % methylation, maximum methylation difference (delta) between any time points after DEX treatment and baseline for each TBS amplicons. Color-shaded regions highlight the main regulatory elements in the locus. **c** Example of CpG sites showing DEX-induced methylation changes. Boxplots represent the methylation levels per time point of CpGs located in the intronic enhancers (top), proximal enhancer (bottom left), and topologically associating domain (TAD) boundaries (bottom right). Methylations of individual CpG sites are shown except for the proximal enhancer plots where the mean methylation per amplicon is shown since this region covers 94 CpG sites. X axes indicate the coordinate of each site or region represented. Shaded boxes indicate sites where significant DEX effects were observed at FDR ≤ 0.05 and absolute delta methylation (T_1 -baseline) $\geq 1\%$ in at least one time point



Inter-individual variability influencing DEX-induced methylation changes

Changes in DNA methylation in peripheral blood may reflect changes in immune cell composition. In study 2, we had data on blood cell counts (BCCs) as well as estimated immune cell types from the Illumina 450K array data. BCCs changed over time following DEX (see Additional file 1: Figure S1A) and lymphocyte counts significantly correlated with 9 DMCs (Additional file 8: Table S3). However, when we corrected for lymphocyte counts in the models testing DEX effects on DNAm across time, all 9 sites remained significant ($FDR \leq 0.05$, Additional file 8: Table S3). Since both DNAm and lymphocytes change over time with DEX, we next assessed how much of the variance in DNAm may still be explained by differences in BCCs. Comparing the standardized coefficients of lymphocyte counts change to the time changes in DNAm in a linear mixed model (LMM), we observed a significantly larger absolute coefficient at the 3-h time point (> 2.8 times larger) for all the associated sites for the changes in DNAm vs. the changes in lymphocytes (Additional file 8: Table S3 and Additional file 1: Figure S1B

showing the residuals of the null model correcting for lymphocyte count). Using estimated cell types from the Illumina 450K array data, we did not observe any effects of DEX on differential cell proportions (see Additional file 1: Figure S1C). Together, these analyses suggest that change in immune cell counts with DEX are likely not a major confounder of our results. In addition to immune cell counts, we also assessed the effects of other possible confounders including sex, age, ethnicity, body mass index (BMI), smoking, and depression symptoms. We observed significant main effects of age (7 sites), sex (5 sites), smoking (6 sites), BMI (8 sites), and major depressive disorder (MDD) (4 sites) on methylation but no significant interactions with DEX (Additional file 6: Table S1 and see Additional file 2: Figure S2A–C for examples). For MDD associations with DNAm, only men were included in the analysis ($n = 67$). We have previously reported no change at baseline for *FKBP5* mRNA levels between MDD patients and controls in a sub-sample of men from study 2 (29 cases and 31 controls, [30]). For the larger sample used here to test associations with MDD status in men (38 cases and 29 controls), there is also no



difference in *FKBP5* mRNA at baseline (p value > 0.05 and fold change ≥ 1.15). In contrast to the mRNA levels, out of the four *FKBP5* CpG sites associated with MDD over the three time points, three sites show a difference at baseline (p value ≤ 0.05 and absolute delta methylation (cases-controls) $\geq |1\%$) with less DNAm observed in cases (range from -1.4 to -2.2%, Additional file 6: Table S1).

Furthermore, we performed stepwise regression analyses using the Akaike Information Criterion (AIC) to select the main covariates influencing DNA methylation changes after DEX. This analysis was performed for the 17 CpG sites for which significant associations with DNA methylation were observed. Based on the AICs as well as the estimates of DNA methylation change after 3 h, correcting for smoking score gave the best models for all CpG sites (AIC smaller and largest variance explained). Sixteen out of seventeen sites remained significant ($FDR < 0.05$) after correcting for smoking (Additional file 10: Table S5). Adding the other covariates did not affect the significance of these 16 sites but increased AIC for most of them.

Effects of *FKBP5* genotype

We and others have previously described allele-specific DNAm changes (lower methylation) in intron 7 of the

FKBP5 gene in peripheral blood cells associated with exposure to child abuse only in carriers of the minor/risk T allele of rs1360780 [11]. Therefore, we investigated whether rs1360780 genotype (CC compared to CT/TT = high mRNA induction and disease risk) had an effect on the observed significant DEX-associated methylation changes in both studies. Genotype effect on DEX-induced methylation changes over time was tested using a model testing both additive as well as interactive effects. In the first study, we observed 17 CpGs showing significant interaction ($n = 13$) or additive effects ($n = 6$) on DNAm changes over time (Additional file 6: Table S1). Despite the different timelines, 2 CpGs showed significant genotype-dependent dynamic differences in both studies, cg35558710 located in intron 7 GRE and cg35570224 located in intron 5 next to the GRE (see Additional file 3: Figure S3). We next tested whether the direction of effects in the two different genotype conformations (CC vs CT/TT) was the same in study 1 and 2. When investigating all 50 CpGs common to both studies, 27 showed the same direction of effect for genotype x DEX at time point 3 h in study 2 as compared to the effect observed overall time points in study 1, which is more than expected by chance. This was not the case for effects of genotype on DNAm

differences at the time point 24 h in study 2. We then subdivided this analysis by regulatory regions. For CpGs within intronic and proximal enhancer GREs ($n = 30$), we observed a concordance of the direction of effects for genotype \times time interaction between the two studies significantly more than expected chance ($p = 0.049$), but this was not the case for the 20 CpGs annotated to TADs. Overall, T carriers displayed more methylation changes over time, with differences to CC genotype carriers ranging from 5.01 to 0.01%. This analysis supports that *FKBP5* rs1360780 genotypes associate with a differential DNAm sensitivity to GR activation within GREs but not in TADs.

Usefulness of Illumina methylation array for assessing DEX-induced DNA methylation changes in *FKBP5*

Most studies investigating DNA methylation in peripheral blood in large cohorts use Illumina methylation arrays. Overlapping Illumina 450K methylation data were available at baseline and 3 h post-DEX administration in study 2 ($n = 106$ subjects, [30]) allowing to assess the extent of coverage of DEX-reactive CpGs on these arrays. The 450K array covers 56 CpGs in the *FKBP5* locus (hg19/chr6:35487554-35718452, see Additional file 4: Figure S4 and Additional file 6: Table S1 for further details) which are located mainly around the TSS ($n = 12$) and the proximal enhancer ($n = 15$) with 12 CpGs in TAD boundaries and only sparse coverage within the gene body (11 CpGs) and 3' end (6 CpGs). None of the intronic GREs showing the biggest change in response to DEX are covered by the array. Analysis of 450K CpG sites that passed QC (52 sites) identified 13 DMCs following DEX ($FDR \leq 0.05$ and absolute Δ methylation $\geq |1\%$) identifying 9 additional DMCs not covered by HAM-TBS. Overlapping sites between TBS and 450K displayed high correlation in both datasets (17 CpGs with 0.96 mean correlation) where similar DNAm changes following DEX were observed (Additional file 4: Figure S4).

DEX-reactive sites reside in enhancer regions with cross-tissue relevance

To understand whether the observed changes could have cross-tissue relevance, as initially shown for the GRE in intron 7 [12], we compared the chromatin state segmentations from the Roadmap Consortium (<http://www.roadmapepigenomics.org/>) for blood/immune cells ($n = 29$ tissues), brain cells ($n = 10$), and fibroblasts ($n = 5$ tissues) (Additional file 5: Figure S5A) for the investigated regions within the *FKBP5* locus. Common active TSS marks at the TSS and marks indicating active transcription over the gene body support the well-documented active transcription of *FKBP5* across these tissues. Focusing on the regions in *FKBP5* that showed the most prominent effects of DEX on DNAm (GRE in intron 7, intron 5, and the

proximal enhancer), we observe that most of these regulatory elements show similar chromatin states, suggesting the comparable regulatory impact of these regions across these tissues (see Additional file 5: Figure S5B).

Discussion

Here, we investigated DNAm changes in response to GR activation in the *FKBP5* locus, a gene in which DNAm changes have been shown in association with exposure to childhood trauma in both children and adults [12, 14–16]. We here show that *FKBP5* DNAm in specific enhancers is highly responsive to GR activation by DEX. We observe dynamic methylation changes over time in longitudinal samples from two independent human studies. Significant effects of DEX exposure were detected as early as 1 h following oral ingestion of DEX, with maximal effects 3–6 h later. Most changes returned to baseline within 23 h. These effects remained significant when correcting for immune cell types as well as additional covariates such as age, sex, BMI, and depression status.

The sites dynamically responsive to an acute DEX challenge in blood overlap with sites correlating with 30-day cortisol load in healthy subjects [31] as well as CpGs differentially methylated in Cushing's syndrome patients, a disorder characterized by excess secretion of cortisol. This has been shown for CpGs within intronic GREs, especially the intron 2 and 7 GREs, for which differences were observed not only between patients with active Cushing's syndrome and controls but also in patients with cured Cushing's syndrome [32]. A second study using 450K data reported cg25114611, located in a GRE of the proximal enhancer, to show significantly lower methylation in patients with long-term remission of Cushing's syndrome [33]. In our study, this site showed lower methylation following DEX in blood that remained significant after 23 h. Such overlap suggests that GR activation may—under specific circumstances—result in lasting DNAm marks. In the above studies, patients had been in remission for 7–13 years [32, 33]. Such factors contributing to more lasting changes could be the level and duration of GR activation as well as developmental timing. We have previously reported that lasting changes in DNAm following DEX treatment in hippocampal progenitor cells were observed when cells were treated during proliferation and differentiation but not when they were treated post-differentiation [12].

The DEX-responsive sites also overlap with those reported to be demethylated in both children and adults exposed to early trauma [11]. Although these sites do not show durable demethylation after 23 h, the overlap suggests that these adversity-related epigenetic changes in *FKBP5* GREs could be mediated via GR-dependent effects through prolonged exposure to GCs, as observed in children exposed to trauma [13].

Our results give new insight into the previously described allele-specific methylation changes of intron 7 GRE CpGs that may contribute to the *FKBP5* x early trauma associations with risk for a number of psychiatric disorders ([10], for review). Several studies report that lower methylation of *FKBP5* GRE intron 7 is more pronounced in child abuse-exposed individuals carrying haplotypes tagged by the T allele of rs1360780 [12, 15, 34]. This functional allele has been associated with increased transcriptional activation of *FKBP5* by GR [12, 35]. Here, we show that healthy individuals carrying this allele have a different dynamic of DNAm changes following GR activation in GREs of *FKBP5*. Such a difference in epigenetic dynamics may also contribute to the fact that environmental risk factors linked to stress hormone activation, such as early adversity, could have more lasting effects in individuals carrying this specific allele. Such allele-specific differences in the dynamics of DNAm change may relate to the reported allele-specific differences in gene transcription. As illustrated in Fig. 2c, dynamic DNAm changes in *FKBP5* GREs inversely correlate with changes in RNA expression.

As expected from the literature, methylation levels of CTCF binding sites at the *FKBP5* locus were low ($0.57 \pm 0.10\%$). In general, it has been shown that CTCF occupancy is inversely correlated with DNA methylation and that DNA methylation at CpGs located directly in the core binding motif can inhibit CTCF binding [27, 36–40]. A gain of methylation directly at CTCF binding sites can lead to loss of CTCF binding and therefore disruption of chromatin interactions which can lead to a dysregulated gene expression [37]. DEX-induced DMCs were not found directly at the core motif of the CTCF binding sites but at close distance and show small changes ($< 2\%$, Fig. 2B). The upstream and downstream flanking regions around the CTCF core motif have been implicated in influencing CTCF binding stability [41]. However, how such a small change in DNA methylation at these flanking regions influences CTCF occupancy has not yet been experimentally addressed so that their exact role remains unknown. Overall, our data suggests that the DEX-induced methylation effects concentrate on GR and not CTCF binding sites. These changes in DNA methylation may thus alter enhancer function (as shown in reporter gene assays [12]) but will not likely result in more profound 3D chromatin changes, such as loop disruptions.

In addition to genetically induced altered dynamics of DEX responsiveness, other factors may also contribute to long-term effects on *FKBP5* methylation in the context of early adversity. Factors associated with child abuse such as smoking, BMI, and depression all had main effects on *FKBP5* methylation, including in dynamically responsive sites (see Additional file 6: Table S1 and Additional file 2: Figure S2). In addition, age also had a main effect on some

CpGs. The limited age range of our cohorts, however, prohibits to analyze the influence of age, including childhood and adolescence, in more depth. Whether such differences in baseline methylation contribute to the long-term effect of early adversity on *FKBP5* methylation needs to be investigated in more detail in longitudinal cohorts. A limitation of this study is the lack of information on early or more recent adversity. Future studies will need to address the influence of these environmental factors on GR-induced DNA methylation dynamics.

Overall, the effects of DEX on *FKBP5* methylation were mostly in the direction of lower methylation following GR activation. In fact, the observation that DEX can induce DNA methylation changes is not unique to the *FKBP5* locus and has been observed at different genomic loci [42]. Several mechanisms could contribute to differences in DNAm. One is a change in cell type composition, favoring cell types with no methylation at these sites within mixed tissues. While this is a possible explanation for the changes observed, the fact that these effects withstand correction for changes in cell composition over time (see Additional file 1: Figure S1B and Additional file 8: Table S3) and that DEX has no significant effect on specific immune cell types estimated from the genome-wide DNAm data suggests that at least some of these effects likely happen within specific cells. Re-assessing our results in sorted cells would give more information into which cell types or cell characteristics are associated with the highest epigenetic reactivity in humans. A study in mice, using genome-wide bisulfite sequencing after cell sorting, observed GC-induced methylation changes primarily in blood T cells [43]. On the other hand, mapping of enhancers across many tissues, including different immune cells (see Additional file 5: Figure S5), suggests that most of the GR-responsive enhancers exert a shared function and may thus show similar epigenetic responses to GR activation across different cell types and tissues.

A reduction in DNA methylation following GR activation could also be mediated via a transcription factor binding-mediated DNA demethylation which has been reported for GR binding to GREs [5]. The mechanisms for this targeted DNA demethylation are not fully understood, but mechanisms involving DNA repair have been proposed [9]. Similar to the GR-induced transient changes observed here in blood, rapid cyclical pattern of DNAm in response to estrogen stimulation in breast cancer cells has been reported. The ER α -responsive gene *pS2* undergoes rapid demethylation and remethylation cycles following activation of transcription with estrogen [44]. The authors [44] implicated a coordinated binding of DNA methyltransferases, glycosylase, and base excision repair proteins in these processes. The process of demethylation of the *pS2* promoter investigated in the

above paper is thought to involve Dnmt3a/b that is able to deaminate 5mC. The resulting abasic site (AP site) of this deamination is subsequently repaired by recruiting p68, TDG, and BER proteins (AP endonuclease, DNA polymerase β , and DNA ligase I). The rapid GR-induced demethylation followed by remethylation within 23 h observed here in blood cells may occur via similar mechanisms given the reported kinetics of this enzymatic process. In addition, when aggregating data from our study for all GREs and mapping DNAm changes to the distance from the consensus GR binding site, we observe high levels of methylation within the consensus binding site and these central CpGs are also the ones with dynamic reduction following GR activation (see Fig. 2a). These observations would support GRE-centric active DNA demethylation. The mechanisms that would then associate with the more lasting changes in remitted Cushing's patients and in individuals exposed to childhood trauma could relate to different actions of DNA-methyl binding proteins such as MeCP2 and polycomb complexes that would interfere with DNA-driven demethylation/remethylation [45].

Conclusions

Taken together, these data provide novel insight into possible mechanisms of stress and trauma-related changes in DNAm and gene \times stress interactions, suggesting a role of GR-dependent methylation changes at least for a subset of the effects. These effects are best investigated using targeted approaches, such as HAM-TBS [25], as most of the reactive enhancer CpGs are not covered on the current Illumina arrays. The observed dynamics of these changes in peripheral blood have consequences on epigenetic association studies in humans, where controlling for cortisol plasma levels appears to be an important factor. Given that dynamic changes in DNAm that can be induced by a single dose of DEX and given their overlapping sites correlating with 30-day cortisol load as well as with lasting changes observed in patients with Cushing's syndrome, critical questions arise for the long-term epigenetic consequences of the therapeutic use of GCs. Additional research in larger samples, with different exposure lengths and intensity, different tissues, and different developmental stages, will be necessary to better understand this phenomenon on a genome-wide and organism-wide level. Cataloging the moderation of these GR-induced epigenetic effects by common gene variants may further help in identifying genes contributing to risk and resilience to stress-related psychiatric disorders.

Methods

Study samples

Study 1

Healthy male participants ($n = 26$, age 25.4 ± 2.9) were given 1.5 mg of DEX orally at 12:00, see [46] for more

details on the study samples and [30, 47–50] for the choice of dose. Peripheral blood was drawn just before administration of DEX as well as 1, 3, 6, and 23 h thereafter. Nineteen of the 26 samples were selected for TBS for a balanced rs1360780 genotype distribution (7 subjects with TT genotype, 6 with TC, and 6 with the CC genotype).

Study 2

The second sample consisted of 89 Caucasian subjects and were also exposed to 1.5 mg of DEX orally as previously described in [42]. Here, DEX was administered at 6 pm and blood draws occurred immediately before the dose of DEX and then 3 hours as well as ~ 18 hours (from 17.5 to 21 h) later. The subset comprised of 30 healthy probands (female = 1; male = 29) and 59 inpatients with depressive disorders (female = 21; male = 38) with an age of 41.64 ± 13.96 (mean age \pm SD).

The demographics of both studies are reported in Table 1.

DNA and RNA extraction of study samples

For both studies, DNA was extracted from frozen EDTA blood using the Elmer Chemagic 360 Instrument (PerkinElmer chemagen Technologie GmbH, Baesweiler, Germany) in combination with the chemagic DNA Blood Kit special 400 (PerkinElmer chemagen Technologie GmbH, Baesweiler, Germany). Thirty-three blood samples of study 2 were only collected in PAXgene tubes (PreAnalytiX GmbH, Hombrechtikon, Switzerland) for the ~ 18 h time point. For these samples, DNA was extracted from PAXgene tubes using PAXgene Blood DNA Kit (QIAGEN GmbH, Hilden, Germany). Blood for RNA was stored in PAXgene tubes, and RNA was extracted using the PAXgene Blood RNA Kit (QIAGEN GmbH, Hilden, Germany). All samples had an RNA integrity number of at least 7.0.

Genotyping

Study 1

All participants were genotyped for the rs1360780 allele using hybridization probes (forward primer: CCTTATTC-TATAGCTGCAAGTCCC, reverse primer: TCTGAATAT-TACCAGGATGCTGAG, rs1360780_LC: Red640-AAAT TCTTACTTGCTACTGCTGGCACAAGAGA-Phosphate, rs1360780_FL: CAGAAGGCTTTCACATAAGCAAAGT-TACACAAAAC-Fluorescein). Genomic DNA was amplified using the LightCycler 480 Genotyping Master mix (Roche, Mannheim, Germany) with the following cycling conditions: 95 °C, 10 s; 45 \times (95 °C, 1 s; 56 °C, 10 s; 72 °C, 10 s); 95 °C, 1 min; and 40 °C, 1 min, ramped up to 85 °C using a ramp rate of 0.57 °C/s and one acquisition per °C on a LightCycler480 II Instrument (Roche, Mannheim, Germany).

Study 2

Genotyping for this cohort is described in [47] and was based on Illumina 660K genotyping arrays.

Assessment of endocrine and immune measures

Cortisol and ACTH levels were assessed as described in [48, 51]. For the measurement of plasma cortisol concentrations, a radioimmunoassay kit was used (INC Bio-medicals, Carlson, CA). Plasma ACTH concentrations were assessed by automated electrochemiluminescence immunoassay using Roche Cobas immunoassay analyzer (Roche, Basel, Switzerland). In the second study, additionally, plasma DEX levels were measured at the 3-h and 20–24-h time point using mass spectrometry as described in [48] and differential blood cell counts evaluated at all three time points as reported in [51].

Gene expression analysis via quantitative real-time PCR

Study 1

FKBP5 mRNA expression levels in blood samples of the first study were assessed as follows. The generation of cDNA was achieved using SuperScript™ II reverse transcriptase (Thermo Scientific Inc., Schwerte, Germany). Subsequently, the cDNA was amplified in duplicates in a LightCycler 480 Instrument II (Roche, Mannheim, Germany) using the LightCycler 480 SYBR Green I Master kit (Roche, Mannheim, Germany) and primer spanning Exon 10-11 (forward primer: AAAAGGCCAAGGAGCA-C AAC, reverse primer: TTGAGGAGGGGCCGAGTTC; cycling conditions: 95 °C, 10 s; 45× (95°,10 s; 58 °C, 10 s; 72 °C, 10 s)). C_t values were used to calculate relative expression levels according to [52] normalized on *YWHAZ* expression (Universal ProbeLibrary probe #77; cycling conditions according to the manufacturer's recommendations; Roche, Mannheim, Germany) for normalization and mean assay efficiencies.

Study 2

RNA expression from study 2 was done using Illumina HT12v4 arrays previously described in [48]. In this study, blood RNA samples were only available for two time points (baseline and 3 h after DEX exposure).

Assessment of DNA methylation in study samples

DNA methylation levels in both studies were analyzed using the HAM-TBS approach, comprising an optimized PCR panel of 28 amplicons in the *FKBP5* locus [25]. These data were complemented by Illumina 450K methylation arrays. An overview of the methylation data obtained can be found in Additional file 6: Table S1.

Study 1

HAM-TBS [25] on the *FKBP5* locus was run for all 28 amplicons on DNA from each blood sample (baseline, 1

h, 3 h, 6 h, and 23 h post-DEX administration) in a single sequencing run with 302 CpG sites analyzed.

Study 2

FKBP5 locus DNAm levels were assessed using both HAM-TBS [25] (10 amplicons covering 50 CpGs from each blood sample (baseline, 3 h and 18 h post-DEX administration)) and 450K methylation arrays (56 CpGs at baseline and 3 h post-DEX administration).

Targeted bisulfite sequencing of the *FKBP5* locus

This method has been described in detail in [25] and offers good performance of amplicon bisulfite sequencing assays in a technology comparison by the Blueprint consortium [53].

Amplicon selection and amplification by PCR

We optimized the amplifications of 28 regions covering 302 CpGs within GR and/or CTCF binding sites as well as the transcription start site of the *FKBP5* locus (see Additional file 9: Table S4 for primers and mapping of the amplicons). In order to reduce cost and maximize the number of samples per sequencing run, triplicate bisulfite treatments were performed for each sample and then pooled to run one PCR amplification per amplicon [25]. Overall, 200 ng to 500 ng of DNA was used per sample and bisulfite treated using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA). Twenty nanograms of bisulfite-converted DNA was then used for each PCR amplification employing Takara EpiTaq HS Polymerase (Clontech, Saint-Germain-en-Laye, France) and 49 amplification cycles. PCR amplicons were then quantified with the Agilent 2200 TapeStation (Agilent Technologies, Waldbronn, Germany) and pooled in equimolar quantities for each sample. AMPure XP beads (Beckman Coulter, Krefeld, Germany) were used for a double size selection (200–500 bp) to remove primer dimers and high molecular DNA fragments.

Sequencing

Libraries were generated using the TruSeq DNA PCR-Free HT Library Prep Kit (Illumina, San Diego, CA) according to the manufacturer's instructions. Each library was quantified with the Qubit® 1.0 (Thermo Fisher Scientific Inc., Schwerte, Germany), normalized to 4 nM and pooled. Library concentration and fragment sizes were checked via Agilent's 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany) and quantitative PCR using the Kapa HIFI Library quantification kit (Kapa Biosystems, Wilmington, MA). Paired-end sequencing was performed on an Illumina MiSeq Instrument (Illumina, San Diego, CA) with their MiSeq Reagent Kit v3 (2× 300 cycles) with the addition of 30% of PhiX Library.

Sequencing data processing

The quality of the sequencing reads was checked with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>), and Illumina adapter sequences were removed using Cutadapt v.1.9.1. Bismark v.0.15.0 was used for the alignment to a restricted reference limited to our PCR targets. In order to stitch paired-end reads, an in-house Perl script has been developed to remove the low-quality ends of the paired-end reads if they overlapped. The methylation levels for all CpGs, CHGs, and CHHs were quantified using the R package methylKit. The resulting DNAm calls were submitted to a 3-step quality control. First, PCR artifacts introducing false CpGs of low coverage at 0 or 100% methylation level were removed. Second, CHH methylation levels were analyzed, and samples with insufficient bisulfite conversion rate (< 95%) were removed. Finally, CpG sites with a coverage lower than 1000 reads were excluded.

Illumina 450K methylation arrays

Study 2

Illumina 450K arrays were processed as described in [42]. Smoking scores were predicted from DNAm data as described in [54] as this information was not available for all subjects. Blood cell ratios were estimated from the DNAm data using the Houseman algorithm [55]. Normalized beta values of 52 CpG sites located within the *FKBP5* locus (hg19/chr6:35,487,554-35,718,452) were extracted from the 425,883 probes that passed quality control (QC).

Statistical analysis

DEX effects in study 1 and study 2

Linear mixed models (LMMs) were used to assess the effects of DEX treatment over time on either ACTH, cortisol, *FKBP5* expression, or DNAm levels for each CpG sites in both studies. All models were run adjusting for intra-individual variability as random effect using the “lmer” function of the Lme4 package in R [56]. *p* values were calculated using the Wald chi-square test from the Car package [57]. False discovery rate (FDR) was applied to correct for multiple testing on methylation *p* values. Post hoc analysis comparing each time point to baseline was ran using LMMs for each site showing significant results from above (FDR ≤ 0.05) to determine at which time point the effect was observed. Differentially methylated CpG sites (DMCs) were called when FDR from the mixed model was ≤ 0.05 and absolute mean methylation differences between significant time points and baseline (*p* value ≤ 0.05) were ≥ 1%.

The power to detect significant DNAm changes after DEX administration was performed using the function “powerSim” from the R package SIMR [56] with 100 simulations for each CpG sites profiled in both studies.

These analyses used a significance (alpha) level of 0.05 and minimum effect of absolute methylation difference ≥ 1% between 3 or 6 h and baseline for study 1 (delta T_3-T_0 and T_6-T_0) and 3 hours and baseline for study 2 (delta T_3-T_0). In our discovery sample (study 1), an average power of 96.6% (bootstrap 95% CI = (94.7, 97.9)) overall the 228 CpG sites was predicted to detect a minimum difference of 1% in methylation after 3 and/or 6 h of DEX administration. Over all the 50 CpG sites profiled in the replication cohort (study 2), an average power of 88.8% (bootstrap 95% CI = (81.2, 93.4)) was predicted to detect a minimum difference of 1% in methylation after 3 h of DEX administration. These results indicate that both cohorts have sufficient power (> 80%) to detect a minimum difference of 1% in methylation after DEX administration with our repeated measures design (5 sampling times in study 1 and 3 in study 2). Although the power in both studies is sufficient to detect a 1% change in methylation, much of the effect observed was larger than 3% (for 66% and 62% of the total significant sites in study 1 and 2, respectively).

In addition, parametric bootstrap using the “bootMer” function of the “lme4 package” in R using 100 simulations, for the mixed models of the 50 CpG sites profiled in study 2, was performed. The bootstrap results including the measures of bias and standard error as well as confidence intervals are given in Additional file 11: Table S6. This analysis revealed that the results are stable as the 95% confidence intervals from the 3-h and 24-h time points indicate a change in DNA methylation for the all sites identified with LMM (FDR ≤ 0.05 and absolute delta methylation ≥ 1%).

Inter-individual factors influencing DEX-induced DNAm changes

To assess inter-individual factors influencing the observed changes in DNAm following DEX administration, each CpG site showing significant DEX effects in study 2 was tested (*n* = 21 CpGs). LMMs were used to assess the association between DNAm change over time and blood cell counts, age, sex, smoking, BMI, and MDD status for each CpG sites. All models were run adjusting for intra-individual variability as random effect using the “lmer” function of the Lme4 package in R [56]. *p* values were calculated using the Wald chi-square test from the Car package [57]. Stepwise regression analysis was also performed on 17 sites showing association with either age, sex, smoking, BMI, or MDD status to select the main confounding variables influencing DNAm change over time. AICs and the DNAm estimates at 3 h of these models were used to select the best model (Additional file 10: Table S5).

FKBP5 genotype effect on DEX-induced methylation changes over time was first assessed on 44 CpG sites in study 1 and 21 sites in study 2 showing DNAm changes at any time point post-administration of DEX. LMMs

were used as described above. Methylation for each CpG was regressed against the main effect of time (DEX) and rs1360780 risk allele (CC or CT/TT) with and without the interaction term of genotype * time point while adjusting for intra-individual variability. *p* values of the additive and interaction effects for each time point were calculated using the Wald chi-square test.

We assessed if the direction of effects was concordant across studies based on the binomial distribution. Assuming that a CpG site shows the same direction of effect in both studies by chance with a probability of 0.5, we determined the probability to observe the present or even a higher number of CpG sites with concordant directions.

Additional files

Additional file 1: Figure S1. Change in blood cell counts after DEX administration. A) Actual blood cell counts at baseline and after DEX administration for granulocytes, monocytes, and lymphocytes in 54 subjects from study 2. B) Boxplot of DNAm residuals from a null model correcting for associated variance in lymphocyte counts across time in 54 subjects from study 2. Post hoc analysis correcting for lymphocyte counts revealed significant change in DNAm after 3 h of DEX for all sites (p value $< 0.1e-18$). C) Predicted blood cell proportions from 450K methylation data in study 2 using the Houseman algorithm [55]. (PDF 260 kb)

Additional file 2: Figure S2. DEX-induced changes in DNAm are also influenced by factors associated with early life adversity. Examples of three CpG sites were significant associations with fixed factors including age, sex, BMI, smoking score, and major depression were observed. (PDF 150 kb)

Additional file 3: Figure S3. CpG sites with significant genotype-dependent dynamic methylation differences in both studies. Effects of rs1360780 genotype on DEX-induced DNA methylation changes in 2 sites located in intron 7 and 5 enhancers. The % methylation levels for rs1360780 risk allele carriers CT/TT and CC carriers following DEX exposure are shown for each study. Methylation of CpG 35558710 shows significant interaction effect at 23 h in study 1 ($X^2 = 5.69$, p value = 0.02) and additive effect at 3 h in study 2 ($X^2 = 4.15$, p value = 0.04) with risk allele genotype. Significant interactions between risk allele genotype and DEX on methylation were observed for CpG 35570224 at 6 h and 23 h post-treatment in study 1 ($X^2 = 7.59$, p value = 0.006 and $X^2 = 6.0$, p value = 0.01) and at 24 h in study 2 ($X^2 = 4.36$, p value = 0.04). Points and error bars represent mean and SEM for each genotype. (PDF 101 kb)

Additional file 4: Figure S4. Replication of dexamethasone (DEX)-induced methylation changes ($n = 106$ subjects) analyzed by Illumina 450K arrays. A) Genome browser shot illustrating the location of TBS amplicons assessed as well as the location of the 450K Illumina probes within the *FKBP5* locus (hg19/chr6:35487554-35718452). *CTCF-ChIA-PET*—track indicating the locations of CTCF factor mediated chromatin interactions determined by Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) data extracted from lymphoblastoid cell line (GM12878, [26]). Chromatin interactions are represented by PET blocks connected with an horizontal line; *CTCF-ChIP-seq* and *GR-ChIP-seq*—regions of transcription factor binding derived from chromatin immunoprecipitation (ChIP) experiments in multiple cell lines from the ENCODE project; *blood TBS amplicons*—locations of targeted bisulfite sequencing (TBS) amplicons assessed in blood of study 1; *450K probe locations*—locations of Illumina probes from the 450K array. B) Boxplot of DNAm levels using TBS or Illumina 450K approach for the overlapping CpG sites showing methylation changes after DEX using TBS. *p* values of linear mixed models for each time point compared to baseline or vehicle are indicated as follows: * ≤ 0.05 , ** ≤ 0.01 , *** ≤ 0.001 . Note that although cg125114611 show significant DEX effect using 450K array, this site has a methylation change after DEX of -0.4% which did not reach our threshold of $|1\%$. (PDF 480 kb)

Additional file 5: Figure S5. Comparison of chromatin states in *FKBP5* across brain, immune/blood, and fibroblasts. A) Genome browser shot illustrating the chromatin states of the *FKBP5* locus (hg19 / chr6:35487554-35718452) across brain, immune/blood, and fibroblasts. *FKBP5 splicing variants*—visualization of two splicing variants of *FKBP5*; *TBS amplicons*—locations of targeted bisulfite sequencing (TBS) amplicons; *450K probe locations*—locations of Illumina probes from the 450K array; *CTCF-ChIP-seq* and *GR-ChIP-seq*—regions of transcription factor binding derived from chromatin immunoprecipitation (ChIP) experiments in multiple cell lines from the ENCODE project; *CTCF-ChIA-PET* and *PoIII-ChIA-PET*—track indicating the locations of CTCF factor or PolII mediated chromatin interactions determined by Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) data extracted from lymphoblastoid cell line (GM12878, [26]). Chromatin interactions are represented by PET blocks connected with an horizontal line; *Ensembl Reg Build*—overview of the Ensembl regulatory build which represents an annotation of regions likely to be involved in gene regulation; *ChroHMM*—this track displays the chromatin state segmentation of the *FKBP5* locus for selected brain, immune/blood, and fibroblast cells from the Roadmap Consortium. The primary core marks segmentation has been used which visualize predicted functional elements as 15 states, which are displayed at the bottom of the figure. B) Quantification of the 15 chromatin states at key regulatory regions (transcription start site (TSS), topologically associating domains (TAD), proximal Enhancer (proxE), and intronic GREs (iGRE)) of the *FKBP5* locus. Chromatin states were averaged over brain ($n = 10$), immune/blood ($n = 29$), and fibroblasts ($n = 5$) cells. (PDF 390 kb)

Additional file 6: Table S1. Details on the CpG sites assessed in *FKBP5* locus and summary of the results obtained using HAM-TBS and Illumina 450K array in both studies. (XLSX 78 kb)

Additional file 7: Table S2. Summary statistic from linear mixed models testing the change in methylation after DEX for each CpG sites assessed in study 1 ($n = 228$) including post-hoc analysis for each time point. In bold are sites with significant DEX-induced methylation change. (XLSX 234 kb)

Additional file 8: Table S3. Summary statistic from linear mixed models testing the change in methylation after DEX including lymphocyte cells counts as covariate for each CpG sites associated with change in lymphocyte counts ($n = 9$) in study 2. (XLSX 50 kb)

Additional file 9: Table S4. Location of HAM-TBS amplicons and primer sequences used to analyze *FKBP5* CpGs. (XLSX 15 kb)

Additional file 10: Table S5. Results from stepwise regression analyses comparing the LMM models without covariate, with smoking score alone, or with all the associated covariates performed on 17 CpG sites showing association with either age, sex, smoking, BMI, or MDD status in study 2. (XLSX 35 kb)

Additional file 11: Table S6. Summary statistic from the linear mixed models testing the change in methylation after DEX for 50 CpG sites profiled in study 2 as well as the measures of bias, standard error, and confidence intervals using bootstraps with 100 simulations. (XLSX 46 kb)

Abbreviations

4-OHT: 4-Hydroxytamoxifen; BCCs: Blood cell counts; BER: Base excision repair; ChIA-PET: Chromatin Interaction Analysis with Paired-End Tag; ChIP: Chromatin immunoprecipitation; CpGs: Cytosine guanine dinucleotides; CTCF: CCCTC-binding factor; DEX: Dexamethasone; DMCM: Differentially methylated CpG site; DNAm: DNA methylation; DNMTs: DNA methyltransferases; ELS: Early-life stress; GC: Glucocorticoid; GR: Glucocorticoid receptor; GRE: Glucocorticoid response element; HAM-TBS: High-accuracy methylation measurements via targeted bisulfite sequencing; HPC: Hippocampal progenitor cell; HWE: Hardy-Weinberg equilibrium; LMM: Linear mixed model; MAF: Minor allele frequency; MDD: Major depressive disorder; PBMCs: Peripheral blood mononuclear cells; PcG: Polycomb complexes; SVA: Surrogate variable analysis; SVs: Surrogate variables; TAD: Topologically associating domains; TETS: Ten-eleven translocation methylcytosine dioxygenase proteins; TSS: The transcription start site

Acknowledgements

The authors would like to thank Monika Rex-Haffner for her help with the sequencing the libraries as well as to all individuals who agreed and provided

blood samples for this study. Moreover, the authors would like to thank Manfred Uhr and his team for their support in the assessment of ACTH and cortisol levels in the human blood samples.

Funding

This study was funded by the BMBF grant Berlin-LCS (FKZ 01KR1301B) to EBB and an ERC starting grant (GxE molmech, grant # 281338) within the FP7 funding scheme of the EU to EBB. NP was funded by a research fellowship from the Canadian Institute of Health Research (CIHR).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

TW, SS, MK, and MB performed the wet lab work. SR performed the sequencing data processing. JA performed the processing of 450K data. TW, DC, and NP analyzed and illustrated the data. MB, LK, and AM collected and processed the study samples. TW, NP, and EBB prepared the manuscript. NP and EBB conceptualized and supervised the study. All authors read and approved the final manuscript.

Authors' information

Not applicable

Ethics approval and consent to participate

All studies were approved by the local ethics committee of the Medical School of the Ludwig Maximilians University, and all participants gave informed consent.

Consent for publication

Not applicable

Competing interests

EBB is co-inventor of the European patent application *FKBP5*: a novel target for antidepressant therapy *European Patent# EP 1687443 B1* and receives a research grant from Böhlinger Ingelheim to investigate *FKBP5* as a candidate target in psychiatric disorders. The other authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Kraepelinstr. 2-10, 80804 Munich, Germany. ²Department of Psychiatry, Psychosomatics and Psychotherapy, University Hospital of Wuerzburg, Wuerzburg, Germany. ³Comprehensive Heart Failure Center, University Hospital of Wuerzburg, Wuerzburg, Germany. ⁴Department of Psychiatry and Behavioral Sciences, Emory University Medical School, Atlanta, GA, USA. ⁵Faculty of Health Sciences, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada. ⁶BC Children's Hospital Research Institute, Vancouver, BC, Canada. ⁷Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany.

Received: 27 November 2018 Accepted: 9 May 2019

Published online: 23 May 2019

References

- Kessler RC, Davis CG, Kendler KS. Childhood adversity and adult psychiatric disorder in the US National Comorbidity Survey. *Psychol Med*. 1997;27:1101–19.
- Sweatt JD. The emerging field of neuroepigenetics. *Neuron*. 2013;80:624–32.
- Weng YL, An R, Shin J, Song H, Ming GL. DNA modifications and neurological disorders. *Neurotherapeutics*. 2013;10:556–67.
- Feldmann A, Ivanek R, Murr R, Gaidatzis D, Burger L, Schubeler D. Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet*. 2013;9:e1003994.
- Thomassin H, Flavin M, Espinas ML, Grange T. Glucocorticoid-induced DNA demethylation and gene memory during development. *EMBO J*. 2001;20:1974–83.
- Kendler KS, Karkowski LM, Prescott CA. Causal relationship between stressful life events and the onset of major depression. *Am J Psychiatry*. 1999;156:837–41.
- Molnar BE, Buka SL, Kessler RC. Child sexual abuse and subsequent psychopathology: results from the National Comorbidity Survey. *Am J Public Health*. 2001;91:753–60.
- Kress C, Thomassin H, Grange T. Local DNA demethylation in vertebrates: how could it be performed and targeted? *FEBS Lett*. 2001;494:135–40.
- Kress C, Thomassin H, Grange T. Active cytosine demethylation triggered by a nuclear receptor involves DNA strand breaks. *Proc Natl Acad Sci U S A*. 2006;103:11112–7.
- Zannas AS, Wiechmann T, Gassen NC, Binder EB. Gene-stress-epigenetic regulation of FKBP5: clinical and translational implications. *Neuropsychopharmacology*. 2016;41:261–74.
- Matosin N, Halldorsdottir T, Binder EB. Understanding the molecular mechanisms underpinning gene by environment interactions in psychiatric disorders: the FKBP5 model. *Biol Psychiatry*. 2018;83:821–30.
- Klengel T, Mehta D, Anacker C, Rex-Haffner M, Pruessner JC, Pariante CM, Pace TW, Mercer KB, Mayberg HS, Bradley B, et al. Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Nat Neurosci*. 2013;16:33–41.
- Klengel T, Binder EB. FKBP5 allele-specific epigenetic modification in gene by environment interaction. *Neuropsychopharmacology*. 2015;40:244–6.
- Non AL, Hollister BM, Humphreys KL, Childebayeva A, Esteves K, Zeanah CH, Fox NA, Nelson CA, Drury SS. DNA methylation at stress-related genes is associated with exposure to early life institutionalization. *Am J Phys Anthropol*. 2016;161:84–93.
- Tyrka AR, Ridout KK, Parade SH, Paquette A, Marsit CJ, Seifer R. Childhood maltreatment and methylation of FKBP5 binding protein 5 gene (FKBP5). *Dev Psychopathol*. 2015;27:1637–45.
- Yehuda R, Daskalakis NP, Bierer LM, Bader HN, Klengel T, Holsboer F, Binder EB. Holocaust exposure induced intergenerational effects on FKBP5 methylation. *Biol Psychiatry*. 2016;80:372–80.
- Buchmann AF, Holz N, Boecker R, Blomeyer D, Rietschel M, Witt SH, Schmidt MH, Esser G, Banaschewski T, Brandeis D, et al. Moderating role of FKBP5 genotype in the impact of childhood adversity on cortisol stress response during adulthood. *Eur Neuropsychopharmacol*. 2014;24:837–45.
- Ising M, Depping AM, Siebertz A, Lucae S, Unschuld PG, Kloiber S, Horstmann S, Uhr M, Muller-Myhsok B, Holsboer F. Polymorphisms in the FKBP5 gene region modulate recovery from psychosocial stress in healthy controls. *Eur J Neurosci*. 2008;28:389–98.
- Halldorsdottir T, Binder EB. Gene x environment interactions: from molecular mechanisms to behavior. *Annu Rev Psychol*. 2017;68:215–41.
- Hawn SE, Sheerin CM, Lind MJ, Hicks TA, Marracini ME, Bountress K, Bacanu SA, Nugent NR, Amstadter AB. GxE effects of FKBP5 and traumatic life events on PTSD: a meta-analysis. *J Affect Disord*. 2019;243:455–62.
- Wang Q, Shelton RC, Dwivedi Y. Interaction between early-life stress and FKBP5 gene variants in major depressive disorder and post-traumatic stress disorder: a systematic review and meta-analysis. *J Affect Disord*. 2018;225:422–8.
- Paakinaho V, Makkonen H, Jaaskelainen T, Palvimo JJ. Glucocorticoid receptor activates poised FKBP51 locus through long-distance interactions. *Mol Endocrinol*. 2010;24:511–25.
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485:376–80.
- Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*. 2014;15:234–46.
- Roeh S, Wiechmann T, Sauer S, Kodel M, Binder EB, Provençal N. HAM-TBS: high-accuracy methylation measurements via targeted bisulfite sequencing. *Epigenetics Chromatin*. 2018;11:39.
- Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Włodarczyk J, Ruszczycki B, et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*. 2015;163:1611–27.
- Renda M, Baglivo I, Burgess-Beusse B, Esposito S, Fattorusso R, Felsenfeld G, Pedone PV. Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J Biol Chem*. 2007;282:33336–45.

28. Wiench M, John S, Baek S, Johnson TA, Sung MH, Escobar T, Simmons CA, Pearce KH, Biddie SC, Sabo PJ, et al. DNA methylation status predicts cell type-specific enhancer activity. *EMBO J*. 2011;30:3028–39.
29. Kheradpour P, Kellis M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res*. 2014;42:2976–87.
30. Menke A, Arloth J, Putz B, Weber P, Klengel T, Mehta D, Gonik M, Rex-Haffner M, Rubel J, Uhr M, et al. Dexamethasone stimulated gene expression in peripheral blood is a sensitive marker for glucocorticoid receptor resistance in depressed patients. *Neuropsychopharmacology*. 2012; 37:1455–64.
31. Lee RS, Mahon PB, Zandi PP, McCaul ME, Yang X, Bali U, Wand GS. DNA methylation and sex-specific expression of FKBP5 as correlates of one-month bedtime cortisol levels in healthy individuals. *Psychoneuroendocrinology*. 2018;97:164–73.
32. Resmini E, Santos A, Aulinas A, Webb SM, Vives-Gilabert Y, Cox O, Wand G, Lee RS. Reduced DNA methylation of FKBP5 in Cushing's syndrome. *Endocrine*. 2016;54:768–77.
33. Glad CA, Andersson-Assarsson JC, Berglund P, Bergthorsdottir R, Ragnarsson O, Johannsson G. Reduced DNA methylation and psychopathology following endogenous hypercortisolism - a genome-wide study. *Sci Rep*. 2017;7:44445.
34. Parade SH, Parent J, Rabemananjara K, Seifer R, Marsit CJ, Yang BZ, Zhang H, Tyrka AR. Change in FK506 binding protein 5 (FKBP5) methylation over time among preschoolers with adversity. *Dev Psychopathol*. 2017;29:1627–34.
35. Yeo S, Enoch MA, Gorodetsky E, Akhtar L, Schuebel K, Roy A, Goldman D. The influence of FKBP5 genotype on expression of FKBP5 and other glucocorticoid-regulated genes, dependent on trauma exposure. *Genes Brain Behav*. 2017;16:223–32.
36. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature*. 2000;405:482–5.
37. Flavahan WA, Drier Y, Liao BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suva ML, Bernstein BE. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. 2016;529:110–4.
38. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/*Igf2* locus. *Nature*. 2000;405:486–9.
39. Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R, Lobanenkov VV. Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol*. 2000;10:853–6.
40. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res*. 2012;22:1680–8.
41. Nakahashi H, Kieffer Kwon KR, Resch W, Vian L, Dose M, Stavreva D, Hakim O, Pruett N, Nelson S, Yamane A, et al. A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep*. 2013;3:1678–89.
42. Zannas AS, Arloth J, Carrillo-Roa T, Iurato S, Roh S, Ressler KJ, Nemeroff CB, Smith AK, Bradley B, Heim C, et al. Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling. *Genome Biol*. 2015;16:266.
43. Seifuddin F, Wand G, Cox O, Pirooznia M, Moody L, Yang X, Tai J, Boersma G, Tamashiro K, Zandi P, Lee R. Genome-wide Methyl-Seq analysis of blood-brain targets of glucocorticoid exposure. *Epigenetics*. 2017;0.
44. Metivier R, Gallais R, Tiffocche C, Le Peron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, et al. Cyclical DNA methylation of a transcriptionally active promoter. *Nature*. 2008;452:45–50.
45. Murgatroyd C, Spengler D. Polycomb binding precedes early-life stress responsive DNA methylation at the *Avp* enhancer. *PLoS One*. 2014;9:e90277.
46. Volk N, Pape JC, Engel M, Zannas AS, Cattaneo N, Cattaneo A, Binder EB, Chen A. Amygdalar microRNA-15a is essential for coping with chronic stress. *Cell Rep*. 2016;17:1882–91.
47. Arloth J, Bogdan R, Weber P, Frishman G, Menke A, Wagner KV, Balsevich G, Schmidt MV, Karbalai N, Czamara D, et al. Genetic differences in the immediate transcriptome response to stress predict risk-related brain function and psychiatric disorders. *Neuron*. 2015;86:1189–202.
48. Menke A, Arloth J, Best J, Namendorf C, Gerlach T, Czamara D, Lucae S, Dunlop BW, Crowe TM, Garlow SJ, et al. Time-dependent effects of dexamethasone plasma concentrations on glucocorticoid receptor challenge tests. *Psychoneuroendocrinology*. 2016;69:161–71.
49. Heuser I, Yassouridis A, Holsboer F. The combined dexamethasone/CRH test: a refined laboratory test for psychiatric disorders. *J Psychiatr Res*. 1994;28:341–56.
50. Hundt W, Zimmermann U, Pottig M, Spring K, Holsboer F. The combined dexamethasone-suppression/CRH-stimulation test in alcoholics during and after acute withdrawal. *Alcohol Clin Exp Res*. 2001;25:687–91.
51. Ramm C, Eichelkraut A, Best J, Czamara D, Rex-Haffner M, Uhr M, Binder EB, Menke A. Sex-related differential response to dexamethasone in endocrine and immune measures in depressed in-patients and healthy controls. *J Psychiatr Res*. 2018;98:107–15.
52. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001;29:e45.
53. consortium B. Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. *Nat Biotechnol*. 2016;34: 726–37.
54. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, Guan W, Xu T, Elks CE, Aslibekyan S, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet*. 2016;9:436–47.
55. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86.
56. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Software*. 2015;67.
57. Fox J, Weisberg S. An R companion to applied regression; 2011.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Discussion

Sequencing on the SOLiD 5500xl System – in-depth characterization of the GC bias

With the introduction of every kind of technology, it is important to assess its performance, to optimize it and put in the context of existing commonly used methods. With the advent of next generation sequencing, some competing technologies were introduced, most notably the SOLiD and Illumina platforms.

To date, existing attempts on evaluation, bias characterization and reduction largely focus on the upscale PCR during library construction, e.g. ramp rate of the thermocycler or the polymerase, specifically for the broadly Illumina sequencing technology (Aird et al., 2011; Oyola et al., 2012). In contrast, we revealed a strong GC bias specifically on the SOLiD system compared to the Illumina machine solely introduced post library preparation. An amplification-free library preparation of a microbial mixture of broad GC content and subsequent sequencing enabled the clear separation of library preparation and sequencing, hence allowing a much clearer identification of the source as well as the extent. The deviance of the sequenced GC distribution from the theoretical distribution based on the known genome sequence was measured and referred to as “shift”. Interestingly, the shift in GC content was much more pronounced in the high range on the SOLiD with 96% compared to 46% in the low range and reached best performance with a shift of 2% for the *P. torridus* at 36% mean GC content. The Illumina machine presented a more balanced and less extensive high/low bias at 23% and 21%, respectively, also reaching a minimum of 15% shift for the *P.torridus*.

Due to the absence of amplification in the library preparation, the emulsion PCR used by Life Technologies was identified as the main source of this bias – it prepares the readily made library for compatibility and subsequent loading on the flow cell and is the equivalent of Illumina’s bridge PCR. The uniformity of the heat distribution within an emulsion reaction cocktail has been shown to be crucial for proper enrichment in the past (Castellanos-Rizaldos, Milbury, & Makrigiorgos, 2012), therefore, optimization of this aspect was chosen. First, reducing the volume of the default E80 pouch by the smaller E20 pouch yielded a significant decrease of the shift. We achieved a reduction of 12% in the high range and 7% in the low, with the maximum improvement of 34% for *P.putida* (62% mean GC

content). In conclusion, the reduction of the volume allowed for a better more even heat distribution that clearly significantly improved the outcome. To reduce the volume even further, the emulsion PCR was performed in a 96-well plate. Again, an improvement of the shift could clearly be observed, however, to a lesser extent. In the high range the overall maximum of 12% reduction of the shift was measurable compared to the E20 pouch, however, on the low end the gain was minimal at 1%.

Overall, the optimization of the emulsion PCR heat distribution showed major improvement of the GC bias, both in the context of high and low GC content. While replacing the E80 pouch by the E20 can be recommended as a default for preparation, the additional expense of handling time to use a 96-well plate needs to be carefully considered in terms of gained signal for the specific experiment. Of note, while the extent of the bias to the margins of the distribution was more pronounced on the SOLiD system, in the intermediate range of *P.torridus* (36% mean GC content) the SOLiD achieved a minimal shift of only 1% (E20 pouch) compared to 15% shift as the best result of the Illumina system. Sequencing approaches residing in this range, potentially spanning up to the *E. coli* range (around 51%) with comparable shifts between both systems of 18% and 16% respectively, may be an almost ideal fit for the SOLiD platform. A close assessment of the technology of choice may harbor real benefits in this context.

The human genome has an average GC content of 40.9 % (Piovesan et al., 2019), however, genomic elements like promoters or CpG islands easily have elevated CG content. To shed light on the impact of this bias on a broadly used experimental design, a H3K9 histone ChIP sequencing was performed using both E80 and E20 emulsion PCR settings as well as the Illumina sequencer. This mark was chosen due to its mixed source distribution spanning multiple genomic features as e.g. CpG islands, exons, introns as well as promotor regions with variable CpG contents. Well in line with the characterization of the balanced microbial mixture, the Illumina machine was either performing similarly (promotor regions / introns) or outperforming the SOLiD depending on the GC content to variable extent. The difference was pronounced in exonic regions and UTRs but most striking in CpG islands. Especially when looking to CpG islands with a CpG content above 62%, data generated by the SOLiD sequencer was almost entirely void of coverage. Due to the functional relevance of these regions the SOLiD sequencer should be used with caution, the Illumina machine will be the better choice. While the switch to the E20 pouch did provide improvement, the performance of the Illumina machine could not be matched and for the H3K9

acetylation ChIP sequencing the Illumina sequencer is identified as the most fitting sequencing machine.

Additionally, checking publicly available SOLiD data on the same histone mark the same bias can be observed. The extent is variable depending on sequencing depth and tissue. While the machine itself is no longer in production, some laboratory still utilize it and awareness of its strength and weaknesses is imperative. Moreover, it is important to note that over the past years many data sets generated on the SOLiD sequencing system have been contributed to the scientific community. This data can still be used in a valuable way for various meta-analysis or validation approaches, hence, potential limitations need to be recognized.

Lastly, while the SOLiD sequencing approach uses an emulsion PCR step it is not the only application of it. Sequencing systems as Roche 545 and Ion Torrent follow the same approach and can benefit from this evaluation as well. In fact, optimization of this PCR step is important but there are further approaches one could consider. For instance, the polymerase or ramp rates could be adjusted when using an emulsion PCR. This may be especially beneficial in the context of extreme GC contents as e.g. when working with CpG islands or methylation measurements from bisulfite converted samples specifically. However, if available, an Illumina machine will mostly be the preferred choice.

HAM-TBS: high-accuracy methylation measurements via targeted bisulfite sequencing

To date, working with methylation measurements by NGS still faces numerous limitations, foremost the high requirements regarding coverage to enable sufficient sensitivity. Despite rapid increase of throughput and the accompanied drop of price, NGS is still a cost intensive method to produce genome wide methylation measurements of low resolution compared to the EPIC array. Moreover, since single sample prices are high, larger sample sizes quickly become unfeasible. Alternative enrichment-based approaches such as MeDIP are cheaper but sacrifice single base pair resolution which can be crucial. Therefore, methods providing affordable methylation measurements of a target set of CpGs at high resolution and in moderate to high sample numbers fill an important experimental need.

HAM-TBS enables the assessment of CpGs with a mean accuracy of 0.72% and within a genomic region of up to ~10 kb while remaining both affordable and

feasible to handle in the lab. This is achieved by introducing crucial steps both in the lab as well as during the extraction of the methylation signal from the data and the application exclusively to validated amplicons. Briefly, reduction of bias during the sample preparation was achieved by (1) introducing a triplicate bisulfite conversion step with subsequent pooling, (2) introduction of a pipetting robot and (3) and a PCR-free library preparation.

By means of in vitro methylated control DNA with 0, 25, 50, 75 and 100 % methylation level, a thorough evaluation of the variance introduced by the bisulfite conversion and library preparation method was performed. Interestingly, we reached very high accuracy for all tested conditions with R^2 values presenting amplification linearity across artificial methylation levels of >0.99 suggesting that our designed workflow is highly robust. Neither the bisulfite conversion nor the PCR-free library preparation introduced significant biases that would prevent robust and sensitive assessment of the methylation levels. However, when working with non-in vitro methylated samples from various tissues we cannot overall exclude higher variability during the bisulfite conversion step. Therefore, a triplicate bisulfite conversion step was installed and triplicates were subsequently pooled before proceeding to target enrichment. This maintained the cost efficiency by not inflating the sample numbers while increasing robustness against potential conversion issues due to different conditions in primary tissues. Next, the Agilent's TapeStation and a pipetting robot were introduced to the workflow. This is enabling a higher throughput which can be quantified as the "loading factor" of the experiment. Briefly, each experimental instance is defined by the number of samples, 96 being the limit due to the Illumina PCR-free library preparation, and the amplicons, usually averaging around 400 bp each, which are assessed per sample. Assuming a symmetric design with *loading factor = samples \times amplicons* a loading factor of >2000 for the HAM-TBS by increasing the atomization can comfortably be realized. Of note, the approach was designed to match the Illumina MiSeq capacity. In theory, the MiSeq can handle loading factors ranging up to 20.000, however, in practice pooling of amplicons and samples has limits in accuracy. A maximum loading factor as high as 2500-3000 has turned out to be practicable. In addition to the increase of throughput, the introduction of the TapeStation and pipetting robot reduce potential variance and error introduced by manual handling of the high number of samples. It does, however, presume access to this technology which may be restricted depending on the setup of the individual lab. Lastly, a PCR-free library preparation method was integrated into the workflow. The amplification bias introduces a source of variance and bypassing it during library preparation contributes to the high

accuracy achieved by HAM-TBS further distinguishing it from the few other TBS approaches presented to date.

To accurately extract the methylation measurements from the data, thorough quality control and rigorous filtering needs to be applied. Three steps were identified as crucial during data processing. First, bisulfite conversion rate is assessed for each sample and amplicon individually by extracting the conversion rate of cytosines in CHH context – meaning the cytosine is not followed by a guanine in the next two bases. Since cytosines in this context are largely unmethylated, a conversion rate close to 100 % is ideal. We installed a cutoff of 95 % to ensure proper bisulfite conversion has taken place, however, insufficient bisulfite conversion usually presents at much lower conversion rates. Second, we remove artefacts which present artificially introduced CpG sites. During the PCR to enrich the target region, errors can lead to artificial creation of CpG sites that are not present in the reference genome. However, limiting the analysis to the set of known CpG sites would remove the option to include SNPs in the analysis. Therefore, all CpG sites present in the data are quantified and filtering is applied afterwards. Artefacts commonly present with low coverage and extreme values of methylation of close to 0 or 100%. A heterogenous SNP will present with 50 % coverage compared to the average coverage across the full amplicon, hence, putting a threshold of e.g. 30 % coverage can be an appropriate choice. Lastly, the absolute coverage itself is to be considered. The aim of HAM-TBS is to provide accurate yet cost-effective measurements. A minimum of 1000 reads was identified as the threshold in agreement with others (Masser et al., 2013) to be both sensitive while still remaining cost efficiency.

The method was tailored to the Illumina MiSeq specifically. At the time, the longest sequencing read length was 300 bp paired-end available on this machine with 150 bp paired-end on other machines being the limit. Hence, the maximum size of amplicon to be covered in is 600 bp, however, sequencing quality drops with increasing length and shorter lengths can be of advantage. In addition, the library preparation limits the number of samples per batch to 96 since the kit is not able to multiplex higher. However, the HAM-TBS approach is designed to match the output capacity of the Illumina MiSeq and a maximum of 96 samples accommodates most medium-throughput experimental designs. Higher plexity would be of interest if CpGs on very few amplicons are being investigated, leaving capacity to increase sample number while not exceeding the maximum loading factor. By today, Illumina has introduced a sequencing kit for the NovaSeq6000 which is able to sequence 250 bp paired-end and therefore of interest to TBS. The yield is >25 times higher than on the MiSeq with less than 2-fold increase of price of the sequencing itself. For experiments exceeding 96 samples or a MiSeq

compatible loading factor the switch to the sequencer with higher capacity would be of interest. However, there are some obstacles. Currently, no commercially available kits exist that enable a PCR-free library preparation of higher plexity, a custom solution would need to be developed. Alternatively, one could consider replacing the library preparation by a PCR based one as Illumina's Nextera prep – however, this would be accompanied by a loss of accuracy. Additionally, Illumina sequencing technology suffers from an index hopping bias (Illumina, 2017) which is lower on the MiSeq. During the development of the HAM-TBS method occurrence of this bias has not been observed. Nonetheless, caution should be maintained and it is recommendable to e.g. use Illumina's Free Adapter Blocking Reagent or dual indexes whenever available.

In a second arm, the HAM-TBS publication releases a comprehensive panel targeting key regulatory regions of the *FKBP5* gene and highlights the importance of validating each amplicon individually. This gene is of high relevance to researchers in the field of psychiatry and is targeted by a panel consisting of 29 amplicons, spanning ~9 kb and thereby covering 315 CpGs located in significant regulatory regions. More specifically, the transcription start site, TAD boundaries, GR and CTCF binding sites as well as enhancers regions are included. In order to accurately assess methylation changes between samples it is imperative that the amplification of the target region is linear and independent of the absolute methylation level. Proper optimization as e.g. cycling conditions and primer design are crucial and can influence the linearity. However, it is conceivable that sites exhibiting extreme CpG density as CpG islands can present with skewed amplification across different levels of methylation. Indeed, when assessing methylation levels within a CpG island present at the transcription start site of *FKBP5* it was not possible to achieve linearity. The bias became more pronounced with increase in methylation level. However, two amplicons were included in the panel, it is important to interpret the results from these non-optimal amplicons with caution if measured methylation levels are high. Besides the evaluation of each amplicon, additional quality checks can be included. For instance, it may be useful to include samples containing unmethylated control DNA, a water control or an endogenous hemimethylated region as the H19 locus to control for index hopping, confirm orientation of the 96-well plate or as positive control.

Taken together, the HAM-TBS method presents a framework enabling highly accurate methylation measurements with medium throughput tailored to the MiSeq. The approach is delivered already equipped with a fully validated panel

targeting the *FKBP5* locus, hence, directly applicable to researchers in the field of psychiatry working with this gene.

Identification of dynamic glucocorticoid-induced methylation changes at the *FKBP5* locus

Psychiatric disorders are complex diseases often resulting from an interplay of genetic and environmental factors. Epigenetic mechanisms can serve to embed environmental cues during an individual's life time. The *FKBP5* locus is an interesting target as studies in both children and adults measuring methylation changes in this locus have repeatedly identified it as reactive in response to childhood trauma (Klengel et al., 2013; Non et al., 2016; Tyrka et al., 2015; Yehuda et al., 2016). Therefore, understanding the dynamics of *FKBP5* methylation in response to glucocorticoid exposure may provide valuable insight into underlying mechanisms embedding stress and gene x stress interactions.

In this scope, methylation was measured in peripheral blood samples of 2 independent cohorts across several timepoints following exposure to dexamethasone (DEX). This is a synthetic glucocorticoid with immunosuppressant and anti-inflammatory properties. In research, DEX is a powerful tool as it provides a way to stimulate the GR, setting the negative feedback loop in motion. Here, DEX was given to individuals with balanced alleles regarding the SNP rs1360780. This SNP has been associated with adult PTSD in individuals having experienced child abuse (Klengel et al., 2013). In this scope, the *FKBP5* panel and HAM-TBS method (Roeh et al., 2018) were key to interrogate samples from 19 (study 1) and 89 (study 2) individuals at baseline and 4 (study 1) or 2 (study 2) further timepoints between 1h and 24h following DEX exposure. Additionally, data generated from the Illumina EPIC array was utilized and complemented this study. As expected, *FKBP5* mRNA expression levels were significantly elevated at the 3h and 6h timepoints and returned to baseline after 23h. The *FKBP5* panel covers CpGs across multiple regulatory elements, however, DEX induced methylation changes (DMCs) were detected predominantly in proximal and intronic enhancers harboring GR and CTCF binding sites. However, methylation is low in CTCF binding sites in proximity but not directly at the binding site have been shown to obstruct CTCF binding (Bell & Felsenfeld, 2000; Wang et al., 2012). The detected changes were rather small and not located directly at the binding site but rather in proximity. This has been implicated in reducing the stability of CTCF binding (Nakahashi et al., 2013), however, it is not yet sufficiently researched to which extent such a small change

may have an impact. Moreover, most of the detected DMCs were located within GREs suggesting an impact on enhancer function rather than on chromatin confirmation. CpGs located surrounding the TSS did not present with DMCs and displayed rather low methylation – as opposed to the higher levels measured within the gene body. Amongst the significant DMCs some occurred as early as one hour after DEX, they reached the maximum at 3h and 6h post admission and more than 80% returned to baseline by 23h. In line, the sites identified as DEX responsive overlapped with sites identified by a study in healthy subjects correlating methylation with 30-day cortisol load (Lee et al., 2018). The same applies to sites with differential methylation in Cushing’s syndrome (Resmini et al., 2016), a disease exposing individuals to high levels of cortisol. Moreover, a study measuring methylation of patients in long-term remission identified a CpG located within a GRE the proximal enhancer of FKBP5 to be significantly demethylated (Glad et al., 2017). Our data supports this finding as this site remained significant after 23h and together indicates the presences of long lasting DNAm changes that may occur in response to GR activation.

Another aspect of the study is the evaluation of the risk allele regarding rs1360780 associated with child abuse (Matosin et al., 2018). Overall, CpG sites shown to be demethylated in children and adults in response to childhood trauma (Matosin et al., 2018) do overlap with those identified in this study to be responsive to DEX. This indicates a mechanism potentially mediated by glucocorticoid exposure, and while the methylation levels returned to baseline after 23h, prolonged exposure may lead to more lasting effects. Allele-specific dynamics of methylation changes over time following DEX were assessed, both additive as well as interactive effects were considered. Both were detected in study 1, the majority being interactive (n=13) with some additive as well (n=6). The timelines of the studies differ, however, of the 17 CpGs 2 – located in intron 7 GRE and next to intron 5 GRE - were also identified in study 2. Overall, the concordance of directionality of the genotype x DEX effect was assessed between both studies. While both studies showed higher than chance concordance of directionality for study 1 (all time points) and study 2 3h timepoint, this was not the case for the study 2 24h time point. Subsetting by regulatory region showed higher than chance concordance for GREs but not topologically associating domains (TAD) . Altogether, methylation changes displayed higher magnitude in GREs for carriers of the risk T allele indication an increased sensitivity to GR activation. This increased sensitivity to GR may also play a role in the potentially longer lasting

manifestation of environmental risk factors on methylation observed in carriers of the risk allele.

Of note, the data was generated from a mixed tissue, therefore, a shift in blood cell counts and immune cell type proportions needed to be considered as a potential source of the signal. As expected, we did observe a shift in lymphocyte counts following DEX, DMCs withstood adjustment for the shift. While this suggests a true effect of DEX on methylation levels, a more detailed analysis using sorted cells could expand these findings. Besides BCC and immune cell type proportions, additional potential confounders were evaluated containing sex, age, ethnicity, BMI, smoking and MDD. While all but ethnicity showed some expected main effects on methylation levels, no significant interaction with DEX could be detected. Of note, some CpGs presented with a main effect of age, however, it was not possible to further investigate its influence. While more detailed analysis of early baseline differences and their contribution to the long-term effects of early adversity on the methylation levels of *FKBP5* would be of interest, the narrow age range present in the cohorts prohibits this in the current study.

The study was carried out in whole blood. However, it is important to note that *FKBP5* chromatin states across various tissues (blood/immune cells, brain cells and fibroblasts) taken from the Roadmap Consortium ("Roadmap Consortium," n.d.) suggest active transcription across tissues in conjunction with current literature. In addition, GRE sites within intron 5 and 7 present with consistent chromatin state across tissues. Taken together, this indicates observed effects of methylation changes on transcription are likely not unique to blood but relevant in multiple other tissues including the brain.

Currently, the most widely used method to measure methylation in large sample sizes is the Illumina EPIC array. While it offers high resolution, it assesses a predefined set of CpGs only. Regarding the *FKBP5* locus the array covers predominantly the TSS and proximal enhancer (>50 % of covered CpGs). Here, EPIC methylation measurements assessed in study 2 yielded 52 CpGs located in the *FKBP5* locus that could be analyzed and compared to 17 CpGs overlapping the HAM-TBS results. As expected, the high mean correlation of 0.96 between the two approaches confirmed consistency between the methods. Overall, the array data allowed for identification of 13 DMCs, 9 of which were not covered by the *FKBP5* panel. However, it has to be noted that the highest DMCs located in the intronic GREs are entirely missing from the array data. This core finding was enabled by the HAM-TBS method and *FKBP5* panel allowing for customization of

the region of interest while remaining feasible at medium to high sample numbers.

Conclusion

Data generation is accelerating, constantly enabling the interrogation of countless biological questions. While the tools are powerful it is imperative to understand their individual strengths and weaknesses to draw the correct conclusions and, moreover, to be able to create customized tools to address specific queries. This thesis presents a full circle approach, opening with the thorough assessment of the SOLiD sequencing system itself and in regard to the broadly used Illumina technology. The relevance of this work in an experimental context is demonstrated using the example of a commonly used CHIP sequencing approach. Additionally, an optimized method termed HAM-TBS is introduced, enabling sensitive measurement of methylation of freely selected CpGs in a targeted fashion in medium to high sample sizes. We provide a validated panel of relevant target CpGs for the FKBP5 locus. This work lays the foundation for addressing the issue of dynamic methylation changes following DEX exposure in the FKBP5 locus with respect to a functional haplotype and sheds a little more light on this open field of research.

References

- Aird, D., Ross, M. G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., ... Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, 12(2), R18. doi: 10.1186/gb-2011-12-2-r18
- Appel, K., Schwahn, C., Mahler, J., Schulz, A., Spitzer, C., Fenske, K., ... Grabe, H. J. (2011). Moderation of Adult Depression by a Polymorphism in the FKBP5 Gene and Childhood Physical Abuse in the General Population. *Neuropsychopharmacology*, 36(10), 1982–1991. doi: 10.1038/npp.2011.81
- Babenko, V. N., Chadaeva, I. V., & Orlov, Y. L. (2017). Genomic landscape of CpG rich elements in human. *BMC Evolutionary Biology*, 17(Suppl 1), 19. doi: 10.1186/s12862-016-0864-0
- Ball, M. P., Li, J. B., Gao, Y., Lee, J.-H., LeProust, E. M., Park, I.-H., ... Church, G. M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature Biotechnology*, 27(4), 361–368. doi: 10.1038/nbt.1533
- Bell, A. C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature*, 405(6785), 482–485. doi: 10.1038/35013100
- Bernstein, D. L., Kameswaran, V., Lay, J. E. L., Sheaffer, K. L., & Kaestner, K. H. (2015). The BisPCR2 method for targeted bisulfite sequencing. *Epigenetics & Chromatin*, 8(1), 27. doi: 10.1186/s13072-015-0020-x
- Binder, E. B., Bradley, R. G., Liu, W., Epstein, M. P., Deveau, T. C., Mercer, K. B., ... Ressler, K. J. (2008). Association of FKBP5 Polymorphisms and Childhood Abuse With Risk of Posttraumatic Stress Disorder Symptoms in Adults. *JAMA*, 299(11), 1291–1305. doi: 10.1001/jama.299.11.1291
- Binder, E. B., Salyakina, D., Lichtner, P., Wochnik, G. M., Ising, M., Pütz, B., ... Muller-Myhsok, B. (2004). Polymorphisms in FKBP5 are associated with increased recurrence of depressive episodes and rapid response to antidepressant treatment. *Nature Genetics*, 36(12), 1319–1325. doi: 10.1038/ng1479
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & Development*, 16(1), 6–21. doi: 10.1101/gad.947102
- Blair, L. J., Nordhues, B. A., Hill, S. E., Scaglione, K. M., O'Leary, J. C., Fontaine, S. N., ... Dickey, C. A. (2013). Accelerated neurodegeneration through

chaperone-mediated oligomerization of tau. *Journal of Clinical Investigation*, 123(10), 4158–4169. doi: 10.1172/jci69003

- Buttenschön, H. N., Krogh, J., Nielsen, M. N., Kaerlev, L., Nordentoft, M., & Mors, O. (2017). Association analyses of depression and genes in the hypothalamus–pituitary–adrenal axis. *Acta Neuropsychiatrica*, 29(1), 59–64. doi: 10.1017/neu.2016.26
- Caspi, A., Hariri, A. R., Holmes, A., Uher, R., & Moffitt, T. E. (2010). Genetic Sensitivity to the Environment: The Case of the Serotonin Transporter Gene and Its Implications for Studying Complex Diseases and Traits. *American Journal of Psychiatry*, 167(5), 509–527. doi: 10.1176/appi.ajp.2010.09101452
- Caspi, A., & Moffitt, T. E. (2006). Gene–environment interactions in psychiatry: joining forces with neuroscience. *Nature Reviews Neuroscience*, 7(7), 583–590. doi: 10.1038/nrn1925
- Castellanos-Rizaldos, E., Milbury, C. A., & Makrigiorgos, G. M. (2012). Enrichment of Mutations in Multiple DNA Sequences Using COLD-PCR in Emulsion. *PLoS ONE*, 7(12), e51362. doi: 10.1371/journal.pone.0051362
- Cole, A. B., Montgomery, K., Bale, T. L., & Thompson, S. M. (2022). What the hippocampus tells the HPA axis: Hippocampal output attenuates acute stress responses via disynaptic inhibition of CRF+ PVN neurons. *Neurobiology of Stress*, 20, 100473. doi: 10.1016/j.ynstr.2022.100473
- Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, 25(10), 1010–1022. doi: 10.1101/gad.2037511
- Dick, A., & Provencal, N. (2018). Central Neuroepigenetic Regulation of the Hypothalamic–Pituitary–Adrenal Axis. *Progress in Molecular Biology and Translational Science*, 158, 105–127. doi: 10.1016/bs.pmbts.2018.04.006
- Edwards, J. R., Yarychivska, O., Boulard, M., & Bestor, T. H. (2017). DNA methylation and DNA methyltransferases. *Epigenetics & Chromatin*, 10(1), 23. doi: 10.1186/s13072-017-0130-8
- Gerritsen, L., Milaneschi, Y., Vinkers, C. H., Hemert, A. M. van, Velzen, L. van, Schmaal, L., & Penninx, B. W. (2017). HPA Axis Genes, and Their Interaction with Childhood Maltreatment, are Related to Cortisol Levels and Stress-Related Phenotypes. *Neuropsychopharmacology*, 42(12), 2446–2455. doi: 10.1038/npp.2017.118
- Glad, C. A. M., Andersson-Assarsson, J. C., Berglund, P., Bergthorsdottir, R., Ragnarsson, O., & Johannsson, G. (2017). Reduced DNA methylation and psychopathology following endogenous hypercortisolism – a genome-wide study. *Scientific Reports*, 7(1), 44445. doi: 10.1038/srep44445
- Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: A Landscape Takes Shape. *Cell*, 128(4), 635–638. doi: 10.1016/j.cell.2007.02.006

- Harris, C. J., Scheibe, M., Wongpalee, S. P., Liu, W., Cornett, E. M., Vaughan, R. M., ... Jacobsen, S. E. (2018). A DNA methylation reader complex that enhances gene transcription. *Science*, 362(6419), 1182–1186. doi: 10.1126/science.aar7854
- Häusl, A. S., Brix, L. M., Hartmann, J., Pöhlmann, M. L., Lopez, J.-P., Menegaz, D., ... Schmidt, M. V. (2021). The co-chaperone Fkbp5 shapes the acute stress response in the paraventricular nucleus of the hypothalamus of male mice. *Molecular Psychiatry*, 26(7), 3060–3076. doi: 10.1038/s41380-021-01044-x
- Heim, C., Bradley, B., Mletzko, T. C., Deveau, T. C., Musselman, D. L., Nemeroff, C. B., ... Binder, E. B. (2009). Effect of Childhood Trauma on Adult Depression and Neuroendocrine Function: Sex-Specific Moderation by CRH Receptor 1 Gene. *Frontiers in Behavioral Neuroscience*, 3, 41. doi: 10.3389/neuro.08.041.2009
- Illumina, Inc. (2017). Effects of index misassignment on multiplexing and downstream analysis. Retrieved from Effects of index misassignment on multiplexing and downstream analysis website: <https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf>
- Klengel, T., Mehta, D., Anacker, C., Rex-Haffner, M., Pruessner, J. C., Pariante, C. M., ... Binder, E. B. (2013). Allele-specific FKBP5 DNA demethylation mediates gene–childhood trauma interactions. *Nature Neuroscience*, 16(1), 33–41. doi: 10.1038/nn.3275
- Klose, R. J., & Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences*, 31(2), 89–97. doi: 10.1016/j.tibs.2005.12.008
- Koenen, K. C., Saxe, G., Purcell, S., Smoller, J. W., Bartholomew, D., Miller, A., ... Baldwin, C. (2005). Polymorphisms in FKBP5 are associated with peritraumatic dissociation in medically injured children. *Molecular Psychiatry*, 10(12), 1058–1059. doi: 10.1038/sj.mp.4001727
- Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell*, 128(4), 693–705. doi: 10.1016/j.cell.2007.02.005
- Lee, R. S., Mahon, P. B., Zandi, P. P., McCaul, M. E., Yang, X., Bali, U., & Wand, G. S. (2018). DNA methylation and sex-specific expression of FKBP5 as correlates of one-month bedtime cortisol levels in healthy individuals. *Psychoneuroendocrinology*, 97, 164–173. doi: 10.1016/j.psyneuen.2018.07.003
- Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., ... Ecker, J. R. (2009). Human DNA methylomes at base resolution show

widespread epigenomic differences. *Nature*, 462(7271), 315–322. doi: 10.1038/nature08514

- Liu, H., Zhou, J., Tian, W., Luo, C., Bartlett, A., Aldridge, A., ... Ecker, J. R. (2021). DNA methylation atlas of the mouse brain at single-cell resolution. *Nature*, 598(7879), 120–128. doi: 10.1038/s41586-020-03182-8
- Liu, Z., Liu, W., Yao, L., Yang, C., Xiao, L., Wan, Q., ... Xiao, Z. (2013). Negative life events and corticotropin-releasing-hormone receptor1 gene in recurrent major depressive disorder. *Scientific Reports*, 3(1), 1548. doi: 10.1038/srep01548
- Lövkvist, C., Dodd, I. B., Sneppen, K., & Haerter, J. O. (2016). DNA methylation in human epigenomes depends on local topology of CpG sites. *Nucleic Acids Research*, 44(11), 5123–5132. doi: 10.1093/nar/gkw124
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), 251–260. doi: 10.1038/38444
- Masser, D. R., Berg, A. S., & Freeman, W. M. (2013). Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. *Epigenetics & Chromatin*, 6(1), 33. doi: 10.1186/1756-8935-6-33
- Masser, D. R., Stanford, D. R., & Freeman, W. M. (2015). Targeted DNA Methylation Analysis by Next-generation Sequencing. *Journal of Visualized Experiments: JoVE*, (96), 52488. doi: 10.3791/52488
- Matosin, N., Arloth, J., Martinelli, S., Czamara, D., Maitra, M., Halldorsdottir, T., ... Binder, E. B. (2021). Brain expressed FKBP5 delineates a therapeutic subtype of severe mental illness. *BioRxiv*, 2021.01.27.428487. doi: 10.1101/2021.01.27.428487
- Matosin, N., Halldorsdottir, T., & Binder, E. B. (2018). Understanding the Molecular Mechanisms Underpinning Gene by Environment Interactions in Psychiatric Disorders: The FKBP5 Model. *Biological Psychiatry*, 83(10), 821–830. doi: 10.1016/j.biopsych.2018.01.021
- Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA*. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC392330/pdf/pnas00024-0174.pdf>
- McGowan, P. O., Sasaki, A., D'Alessio, A. C., Dymov, S., Labonté, B., Szyf, M., ... Meaney, M. J. (2009). Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nature Neuroscience*, 12(3), 342–348. doi: 10.1038/nn.2270
- Mifsud, K. R., & Reul, J. M. H. M. (2016). Acute stress enhances heterodimerization and binding of corticosteroid receptors at glucocorticoid

- target genes in the hippocampus. *Proceedings of the National Academy of Sciences*, 113(40), 11336–11341. doi: 10.1073/pnas.1605246113
- Misiak, B., Łoniewski, I., Marlicz, W., Frydecka, D., Szulc, A., Rudzki, L., & Samochowiec, J. (2020). The HPA axis dysregulation in severe mental illness: Can we shift the blame to gut microbiota? *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 102, 109951. doi: 10.1016/j.pnpbp.2020.109951
- Moore, L. D., Le, T., & Fan, G. (2013). DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, 38(1), 23–38. doi: 10.1038/npp.2012.112
- Murgatroyd, C., Patchev, A. V., Wu, Y., Micale, V., Bockmühl, Y., Fischer, D., ... Spengler, D. (2009). Dynamic DNA methylation programs persistent adverse effects of early-life stress. *Nature Neuroscience*, 12(12), 1559–1566. doi: 10.1038/nn.2436
- Nakahashi, H., Kwon, K.-R. K., Resch, W., Vian, L., Dose, M., Stavreva, D., ... Casellas, R. (2013). A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. *Cell Reports*, 3(5), 1678–1689. doi: 10.1016/j.celrep.2013.04.024
- Nichols, R. V., O'Connell, B. L., Mulqueen, R. M., Thomas, J., Woodfin, A. R., Acharya, S., ... Adey, A. C. (2022). High-throughput robust single-cell DNA methylation profiling with sciMETv2. *Nature Communications*, 13(1), 7627. doi: 10.1038/s41467-022-35374-3
- NIH National Human Genome Research Institute. (n.d.). Retrieved from <https://www.genome.gov/>
- Non, A. L., Hollister, B. M., Humphreys, K. L., Childebayeva, A., Esteves, K., Zeanah, C. H., ... Drury, S. S. (2016). DNA methylation at stress-related genes is associated with exposure to early life institutionalization. *American Journal of Physical Anthropology*, 161(1), 84–93. doi: 10.1002/ajpa.23010
- O'Donnell, K. J., & Meaney, M. J. (2020). Epigenetics, Development, and Psychopathology. *Annual Review of Clinical Psychology*, 16(1), 1–24. doi: 10.1146/annurev-clinpsy-050718-095530
- Oyola, S. O., Otto, T. D., Gu, Y., Maslen, G., Manske, M., Campino, S., ... Quail, M. A. (2012). Optimizing illumina next-generation sequencing library preparation for extremely at-biased genomes. *BMC Genomics*, 13(1), 1. doi: 10.1186/1471-2164-13-1
- Paakinaho, V., Makkonen, H., Jääskeläinen, T., & Palvimo, J. J. (2010). Glucocorticoid Receptor Activates Poised FKBP51 Locus through Long-Distance Interactions. *Molecular Endocrinology*, 24(3), 511–525. doi: 10.1210/me.2009-0443

- Piovesan, A., Pelleri, M. C., Antonaros, F., Strippoli, P., Caracausi, M., & Vitale, L. (2019). On the length, weight and GC content of the human genome. *BMC Research Notes*, 12(1), 106. doi: 10.1186/s13104-019-4137-z
- Renthal, W., & Nestler, E. J. (2008). Epigenetic mechanisms in drug addiction. *Trends in Molecular Medicine*, 14(8), 341–350. doi: 10.1016/j.molmed.2008.06.004
- Resmini, E., Santos, A., Aulinas, A., Webb, S. M., Vives-Gilabert, Y., Cox, O., ... Lee, R. S. (2016). Reduced DNA methylation of FKBP5 in Cushing's syndrome. *Endocrine*, 54(3), 768–777. doi: 10.1007/s12020-016-1083-6
- Roadmap Consortium. (n.d.). Retrieved from <http://www.roadmappigenomics.org/>
- Roeh, S., Wiechmann, T., Sauer, S., Kodel, M., Binder, E. B., & Provencal, N. (2018). HAM-TBS: high-accuracy methylation measurements via targeted bisulfite sequencing. *Epigenetics & Chromatin*, 11(1), 1–10.
- Romano, S., Staibano, S., Greco, A., Brunetti, A., Nappo, G., Ilardi, G., ... Romano, M. F. (2013). FK506 binding protein 51 positively regulates melanoma stemness and metastatic potential. *Cell Death & Disease*, 4(4), e578–e578. doi: 10.1038/cddis.2013.109
- Roy, A., Gorodetsky, E., Yuan, Q., Goldman, D., & Enoch, M.-A. (2010). Interaction of FKBP5, a Stress-Related Gene, with Childhood Trauma Increases the Risk for Attempting Suicide. *Neuropsychopharmacology*, 35(8), 1674–1683. doi: 10.1038/npp.2009.236
- Rulten, S. L., Kinloch, R. A., Tateossian, H., Robinson, C., Gettins, L., & Kay, J. E. (2006). The human FK506-binding proteins: characterization of human FKBP19. *Mammalian Genome*, 17(4), 322–331. doi: 10.1007/s00335-005-0127-7
- Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., ... Smith, M. (1977). Nucleotide sequence of bacteriophage ϕ X174 DNA. *Nature*, 265(5596), 687–695. doi: 10.1038/265687a0
- Siva, N. (2008). 1000 Genomes project. *Nature Biotechnology*, 26(3), 256–256. doi: 10.1038/nbt0308-256b
- Smith, Z. D., & Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nature Reviews Genetics*, 14(3), 204–220. doi: 10.1038/nrg3354
- Somarelli, J. A., Lee, S. Y., Skolnick, J., & Herrera, R. J. (2008). Structure-based classification of 45 FK506-binding proteins. *Proteins: Structure, Function, and Bioinformatics*, 72(1), 197–208. doi: 10.1002/prot.21908

- Tyrka, A. R., Ridout, K. K., Parade, S. H., Paquette, A., Marsit, C. J., & Seifer, R. (2015). Childhood maltreatment and methylation of FK506 binding protein 5 gene (FKBP5). *Development and Psychopathology*, 27(4pt2), 1637–1645. doi: 10.1017/s0954579415000991
- Waddington, C. H. (2012). The Epigenotype. *International Journal of Epidemiology*, 41(1), 10–13. doi: 10.1093/ije/dyr184
- Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., ... Stamatoyannopoulos, J. A. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Research*, 22(9), 1680–1688. doi: 10.1101/gr.136101.111
- Watson, J. D., & Crick, F. H. C. (1953). THE STRUCTURE OF DNA. *Cold Spring Harbor Symposia on Quantitative Biology*, 18(0), 123–131. doi: 10.1101/sqb.1953.018.01.020
- Weickert, C. S., Webster, M. J., Boerrigter, D., & Sinclair, D. (2016). FKBP5 Messenger RNA Increases After Adolescence in Human Dorsolateral Prefrontal Cortex - *Biological Psychiatry*. *Biological Psychiatry*. Retrieved from [https://www.biologicalpsychiatryjournal.com/article/S0006-3223\(15\)00957-9/fulltext](https://www.biologicalpsychiatryjournal.com/article/S0006-3223(15)00957-9/fulltext)
- Xie, P., Kranzler, H. R., Poling, J., Stein, M. B., Anton, R. F., Farrer, L. A., & Gelernter, J. (2010). Interaction of FKBP5 with Childhood Adversity on Risk for Post-Traumatic Stress Disorder. *Neuropsychopharmacology*, 35(8), 1684–1692. doi: 10.1038/npp.2010.37
- Yehuda, R., Daskalakis, N. P., Bierer, L. M., Bader, H. N., Klengel, T., Holsboer, F., & Binder, E. B. (2016). Holocaust Exposure Induced Intergenerational Effects on FKBP5 Methylation. *Biological Psychiatry*, 80(5), 372–380. doi: 10.1016/j.biopsych.2015.08.005
- Zannas, A. S., Wiechmann, T., Gassen, N. C., & Binder, E. B. (2016). Gene–Stress–Epigenetic Regulation of FKBP5: Clinical and Translational Implications. *Neuropsychopharmacology*, 41(1), 261–274. doi: 10.1038/npp.2015.235
- Zhou, W., Laird, P. W., & Shen, H. (2017). Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Research*, 45(4), e22–e22. doi: 10.1093/nar/gkw967
- Ziller, M. J., Hansen, K. D., Meissner, A., & Aryee, M. J. (2015). Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nature Methods*, 12(3), 230–232. doi: 10.1038/nmeth.3152
- Zimmermann, P., Brückl, T., Nocon, A., Pfister, H., Binder, E. B., Uhr, M., ... Ising, M. (2011). Interaction of FKBP5 gene variants and adverse life events in predicting depression onset: results from a 10-year prospective community

study. *The American Journal of Psychiatry*, 168(10), 1107–1116. doi:
10.1176/appi.ajp.2011.10111577

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den15.05.2023.....

.....Simone Röh-Karamihalev.....

(Unterschrift)

Erklärung

Hiermit erkläre ich, *

- dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.
- dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.
- dass ich mich mit Erfolg der Doktorprüfung im Hauptfach
und in den Nebenfächern
bei der Fakultät für der
(Hochschule/Universität)
unterzogen habe.
- dass ich ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

.....München, 15.05.2023.....

.....Simone Röh-Karamihalev.....