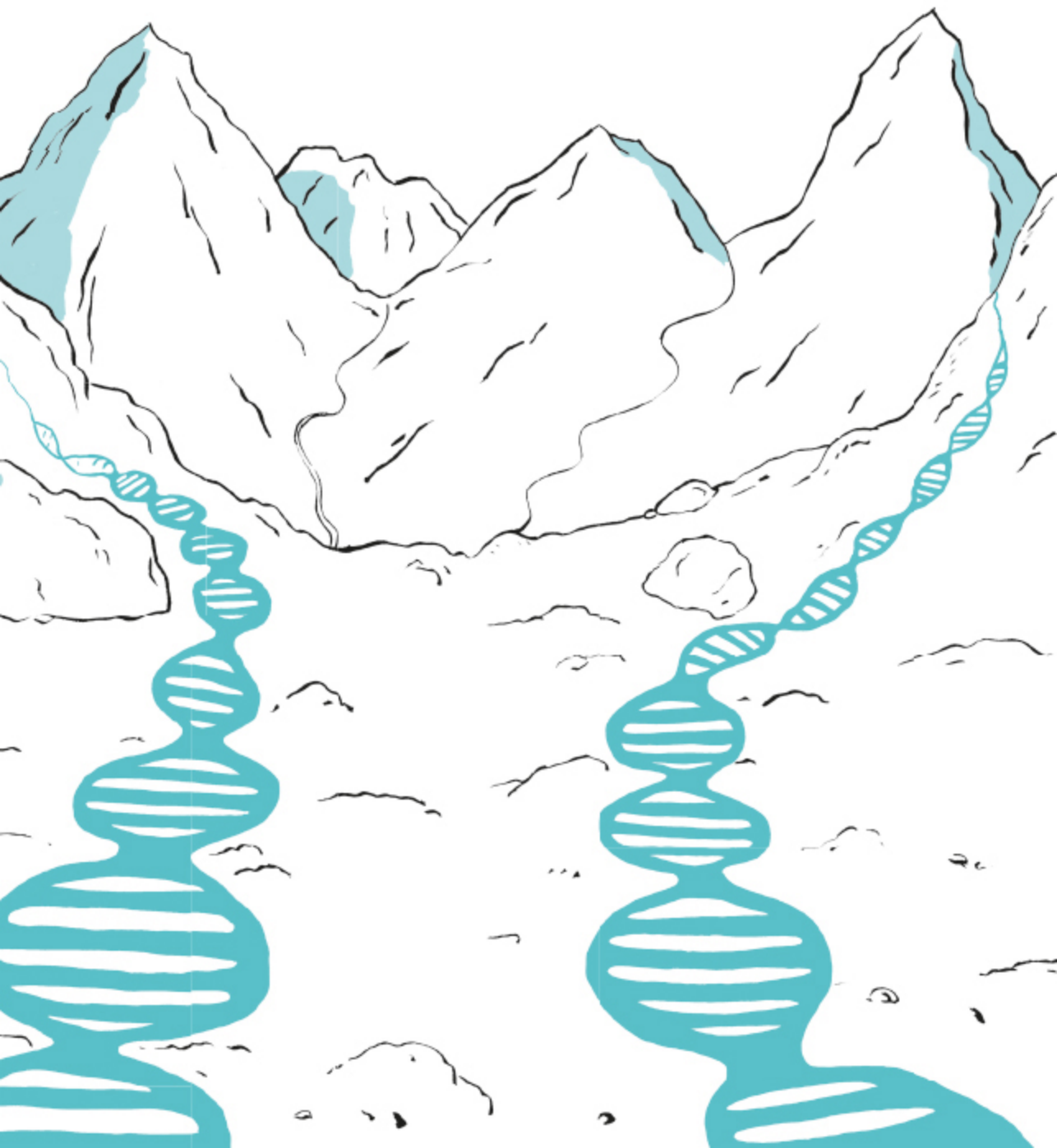
CONVERGING MOLECULAR NETWORKS AFFECTED IN KLEEFSTRA SYNDROME AND RELATED NEURODEVELOPMENTAL DISORDERS

Tom S. Koemans



Converging molecular networks affected in Kleefstra syndrome and related neurodevelopmental disorders

Tom Stefan Koemans

Cover concept: Tom Koemans
Lay-out and cover design: Remco Wetzels, Ridderprint BV
Printed by: Ridderprint BV, Ridderkerk

ISBN 978-94-6299-771-4

Converging molecular networks affected in Kleefstra syndrome and related neurodevelopmental disorders

Proefschrift

ter verkrijging van de graad van doctor

aan de Radboud Universiteit Nijmegen

op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,

volgens besluit van het college van decanen

in het openbaar te verdedigen op dinsdag 17 april 2018

om 10.30 uur precies

door

Tom Stefan Koemans

geboren op 15 september 1982

te Rheden

Promotor:
Prof. dr. ir. J.H.L.M. van Bokhoven

Copromotoren:
Dr. A. Schenck
Dr. J.M. Kramer (The University of Western Ontario, Canada)

Manuscriptcommissie:
Prof. dr. J.H. Jansen
Prof. dr. G.J.C. Veenstra
Dr. E. Storkebaum (Max-Planck-Institute für Molekulare Biomedizin Münster, Duitsland)

Paranimf :
Human Riahi Asl

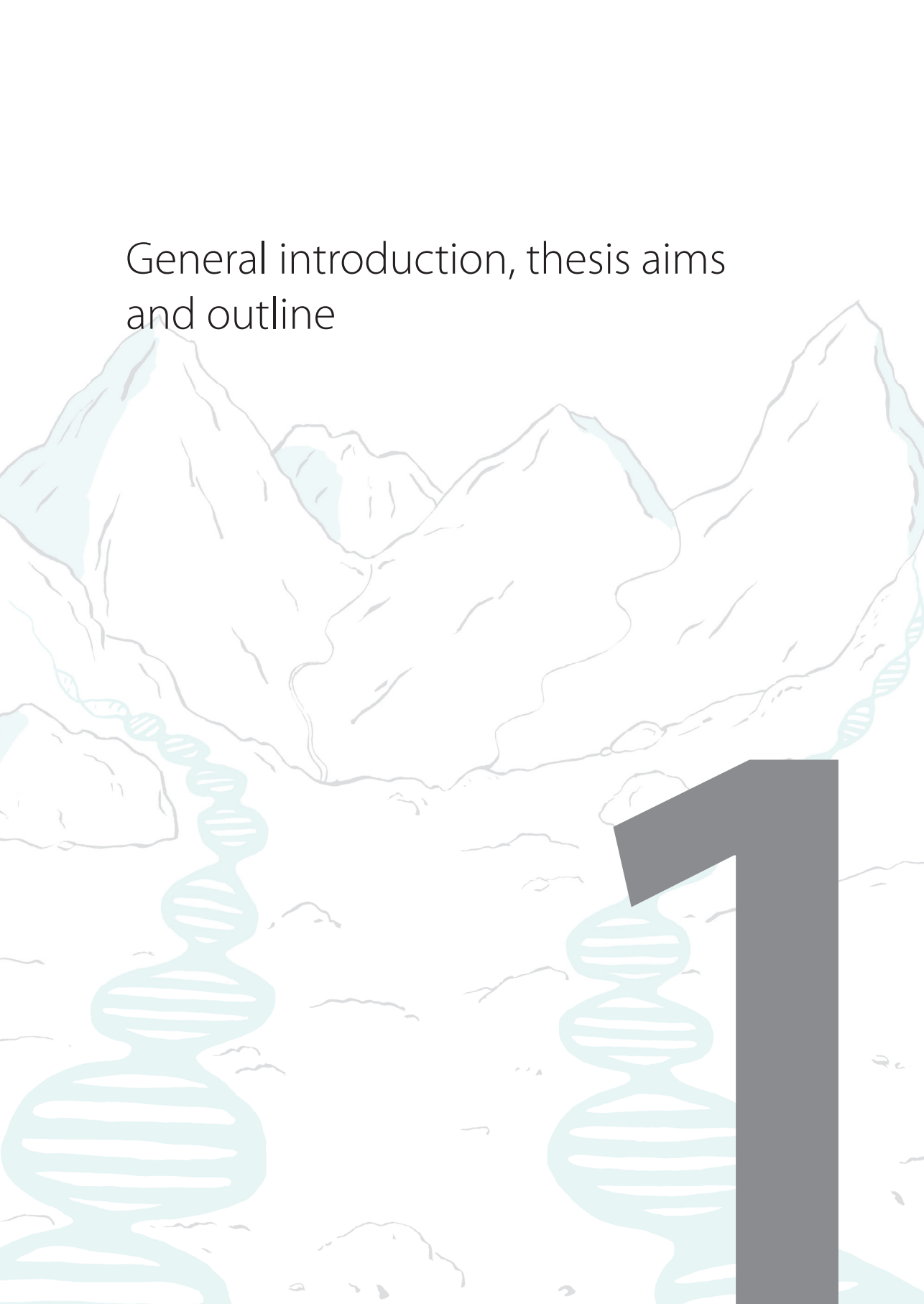
"Der Weg ist das Ziel"

Confucius (551-79 BC), Chinese teacher and philosopher

But also from "Die Eiger-Nordwand-Tragödie" (1936)

Table of contents

General introduction, thesis aims and outline



1

Intellectual disability

Intellectual disability (ID) is defined in the Diagnostic and Statistical Manual of Mental Disorder edition 5 (DSM-5) as a neurodevelopmental disorder with an onset during childhood and includes both difficulties in adaptive and intellectual functioning of the conceptual, social, and practical domains [1]. ID is diagnosed by the following three criteria: (1) an intelligence quotient (IQ) below 70, (2) significant limitations in adaptive behaviour such as communication, self-care, social skills, community use, self-direction, health, and safety, and (3) an onset before the age of eighteen [1]. A further classification is made based on the level of mental impairment: profound (IQ<20), severe (IQ 20-34), moderate (IQ 35-49), and mild (IQ 50-69) [2]. ID can manifest as "non-syndromic" where it is the sole recognizable symptom, or as "syndromic" where it is accompanied by additional mental and physical hallmarks such as dysmorphic features, metabolic or congenital defects [2]. Estimates of the prevalence are between 1 to 3 percent of the Western population [3] and recent estimates point towards 400.000 children born per year worldwide due to *de novo* mutations [4]. These numbers exceed the prevalence of many other mental illnesses and with the need for lifelong care, it is a challenge for our health care system and for caretakers [5]. Thus, ID represents one of the largest unsolved socio-economic challenges for society.

Approximately half of all the ID cases are of genetic origin and genetic testing has been a large contributor to the diagnosis of ID. Historically, examination of a common karyogram could confirm trisomy 21 under a microscope and thereby identify Down syndrome [6]. Later, technologies improved and cytogenetic banding and chromosomal abnormalities could be used to identify novel common causes of ID [7]. For example, mapping of breakpoints of translocations and deletions by fluorescent in situ hybridization (FISH) and linkage analysis combined with Sanger sequencing could target specific chromosomal regions [8]. Around 2000, genome wide array techniques were introduced which identified numerous novel micro deletion and duplication syndromes [9]. Nowadays, in the era of next generation sequencing, mutations are identified at a high frequency in trio-based approaches of exome sequencing [10-13].

Currently, nearly 1000 genes have been implicated in ID ([14] and http://sysid.cmbi.umcn.nl/) and new associations are still occurring at a high frequency [4, 10, 13, 15]. These genes vary a lot in function, however, certain common functionality can be identified. Systematic analysis of common themes identified highly enriched functional categories of similar biological function and phenotypes in human [14] and fly [16]. Well known examples are genes involved in synapse formation, intracellular signaling cascades, or mitochondria.

Another group that has been identified, and examples of this group are discussed in this thesis, is related to chromatin regulators which all have a function in epigenetics and/ or gene expression [17, 18]. This group consists of over 55 genes and generally encode epigenetic regulators such as histone modifiers, RNA binding proteins, and transcription factors [14, 19].

Epigenetic regulation of gene expression

The term "epigenetics" was coined by Conrad H. Waddington (1905-1975) as "the branch of biology which studies the causal interactions between genes and their products, which brings the phenotype into being" [20]. Epigenetics refers to the Greek prefix "*epi*" which means "on top of" or "in addition to" the traditional genetic basis of inheritance. Epigenetic modifications allow for heritable changes in gene expression through cell division without changing the DNA sequence. These changes often include alterations in the chromatin template. Nucleosomes are the building blocks of chromatin and consists of 147 basepairs DNA, linker histone H1, and dimers of histone H2A, H2B, H3, and H4 that form an octamer. The N-terminal "tails" of the four core histones and the C-terminal tail of histone H2A protrude from the nucleosome. These tails can be subjected to post-translational covalent modification of chemical groups and, classically, produce densely packed heterochromatin and more loosely packed euchromatin, directly influencing the accessibility of RNA polymerase. However, with recent technical and computational advances, local chromatin states have been described for a number of genomes [21-25]. These chromatin states are determined by co-association between epigenetic factors such as local histone modifications, transcription factor binding, non-coding RNA molecules, and nucleosome positioning. For example, the transcription start sites of active genes are marked by histone H3 Lysine (K) 4 trimethylation. On the other hand, silent regions of transcription are enriched for H3K9 di- and trimethylation, or H3K27 trimethylation [26]. Besides these classical, well studied histone modifications, many more posttranslational histone modifications have been identified, including more than 60 different residues on histones that can be modified by methylation, acetylation, phosphorylation, ubiquitylation, sumoylation, or ADP ribosylation [27, 28]. These findings could uncover totally new fields of epigenetic regulation of gene expression.

Another epigenetic mark influencing transcription is DNA methylation. The most abundant modified base in the mammalian genome is 5-methylcytosine, often clustered at CpG islands. These islands are often enriched at the promoter regions of genes and are generally associated with transcriptional repression when methylated. Active oxidation of 5-methylcytosine is catalyzed by ten eleven translocation enzymes to form 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine [29]. 5-hydroxymethylcytosine is shown to be highly present in the human brain and may uncover a new mechanism for epigenetic regulation of gene expression [30]. Additionally, adenine residues (N(6)methyladenine) are associated with transcriptional repression of long interspersed nuclear elements (LINE) transposons [31].

The above mentioned epigenetic processes play a crucial role during development of organisms. The totipotent zygote forms the specific germ layers endoderm, mesoderm, and ectoderm and develops into an embryo via many cell divisions [32]. During neurulation,

the ectoderm forms the neural plate that invaginates to form the neural tube and will later differentiate to the brain and spinal cord [32]. The early human brain first consists of the three primary vesicles prosencephalon, mesencephalon, and rhombencephalon and are later organized into the many brain regions formed by the walls and cavities [32]. These developmental processes require tightly controlled spatial and temporal regulation of neuronal gene expression [33]. When mutations occur in genes encoding for epigenetic regulators that mediate these transcriptional processes, normal development of an organism can be disrupted. In addition, epigenetic processes are crucial for proper processing of extracellular signals that lead to transcriptional changes during adulthood (see underneath) [34]. Kleefstra syndrome is an example of an ID disorder in which an epigenetic factor is affected.

Kleefstra syndrome

Kleefstra syndrome is an ID syndrome characterized by moderate to severe ID, developmental delay, childhood hypotonia, and characteristic craniofacial features like microcephaly, brachcephaly, hypertelorism, flat facies, midface hypoplasia, synophrys, arched eyebrows, anteverted nares, short nose, open mouth with protruding tongue, everted thicker lower lip, thin upper lip [35-37] (Figure 1). Neurobehavioral analysis of Kleefstra syndrome patients revealed a very high prevalence of autism spectrum disorder, extra vulnerability to develop obsessive–compulsive disorder and increased levels of depression compared to a control group that consists of non-related ID [38].

Kleefstra syndrome was first reported in 1994 by identification of a patient with ID and a cytogenetically visible terminal deletion of the chromosome band 9q34.3 [39], hence its former name 9q subtelomeric deletion syndrome. Three years later there was a second report of a patient with a similar genotype and phenotype [40]. In 2004, these reports were brought together with seventeen newly identified patients and this narrowed the deletion down to 1.2Mb spanning at that time 14 known genes [41]. This region was further limited to one affected gene, *euchromatin histone methyltransferase1 (EHMT1)* by the description of two patients with the same clinical features [42]. It is now widely accepted that heterozygous dominant deletions and mutations affecting the *EHMT1* gene cause Kleefstra syndrome (OMIM #610253) and more than 100 patients have been described [35, 42-47]. The syndrome has been identified in all ethnic groups, worldwide, and in both males and females [37, 48].

Figure 1: Kleefstra syndrome patients at different ages showing clinical hallmarks. Figure from Kleefstra *et al.* (2009) [43] and reprinted with permission.

The EHMT protein family

In mammals there are two EHMT paralogs, EHMT1 and EHMT2. Both proteins are widely expressed and contain ankyrin repeat sequences followed by a C-terminal pre-SET and SET domain [49-51]. The ankyrin repeats are known to be required for protein-protein interaction and are shown to bind H3K9me1 an H3K9me2 [52]. The SET/Pre-SET domains are capable of catalyzing mono- di- and trimethylation and are conserved during evolution as a family of proteins [53]. The acronym SET is named after the *Drosophila* genes *Su(var)3-9*, *Enhancer of zeste*, and *Trithorax*. *In vitro* experiments showed that EHMT1 and EHMT2 are required for methylation of histone H1 and H3 on lysine 9 (K9) and 27 (K27) [49, 54]. More specifically, it was shown that EHMT1 and EHMT2 are mainly responsible for H3K9 mono- and dimethylation *in vitro* at transcriptionally active euchromatic regions of the genome [55], in contrast to H3K9me3 which is enriched at pericentromric heterochromatin [56, 57]. Moreover, chemical inhibition of EHMT1 and EHMT2 in cultured cells lowers dimethylation of H3K9 [58]. Cultured mouse embryonic stem cells from EHMT1 or EHMT2 knockout mice also showed a reduction of mono- and dimethylation of H3K9 [59, 60]. Monomethylation of H3K9 is associated with actively transcribed genes, whereas H3K9me2 is associated

with transcriptional silencing [26]. Thus, EHMT1 and EHMT2 are considered to regulate transcription via H3K9me1 and H3K9me2 in euchromatin.

EHMT proteins are also associated with a range co-repression proteins and complexes. First, a direct protein-protein interaction was shown with repressor element-1 silencing transcription factor (REST) [61, 62] (**chapter 5**). In addition, the polycomb repressive complex [63, 64], a Suv39/EHMT1/EHMT2/SETBP1 module [65], a repressive complex that functions in the G0-phase of cell division [66], and CDP/cut associated repression [67] were shown to contain EHMT proteins. One other well studied repressive mechanism is the interaction with widely interspaced zinc finger (WIZ) [68] and ZNF644 [69]. It was shown that EHMT1 and EHMT2 can form a heterodimer via their c-terminal catalytic domains [60] in which EHMT1 binds WIZ and EHMT2 binds ZNF644 and contribute to DNA target specificity [70]. WIZ also contains pro c-terminal-binding protein1 (CtBP) binding sites which is associated with transcription repression [68, 71].

In contrast, EHMT proteins also mediate transcription activation which has mainly been studied in adipose cells. In brown adipose cells, deletion of EHMT1 causes obesity and insulin resistance [72]. In white adipose cells, EHMT2 has been shown to facilitate Wnt10a expression, which inhibits adipogenesis [73]. Subsequently, adipose-specific deletion of EHMT2 increases adipogenic gene expression and adipose tissue weight [73]. Taken together, EHMT proteins are involved in many gene regulatory protein complexes and are studied mainly via their role in H3K9me2.

EHMT proteins in synaptic plasticity

The brain is an organ made up of billions of neurons that form complex cellular networks and allows organisms to learn, memorize, forget, and adapt behavior. On the molecular level, memory acquisition can be initiated via activation of post-synaptic NMDA and AMPA receptors and leads to a depolarization of the postsynaptic neuron and an influx of Ca^{2+} [74]. Increased levels of Ca^{2+} activate multiple biochemical cascades and kinases. For instance, the kinase Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) can phosphorylate NMDA [75] and AMPA receptors [76] leading to a clustering of receptors and strengthening of the activated synapse [77]. Long lasting activated CaMKII induces intracellular signaling cascades such as cyclic AMP (cAMP) [78], MAPK/ERK [79], and CaMK [80] converging on the transcription factor cAMP response element binding protein (CREB) [81]. The transcription factor CREB can regulate activity of genes important for learning and memory, however, other transcription factors and chromatin regulators are also involved [82].

Evidence for a role of EHMT proteins in transcriptional plasticity of learning and memory comes from contextual fear memory in rats [83]. Contextual fear memory is a classical paradigm in which animals are placed in a new cage with or without a foot shock. It is shown that H3K9me2 levels are increased after one hour in rats with and without the foot

shock [83]. H3K9me2 decreases after 24 hours, with the exception of the entorhinal cortex [84]. In this brain region, H3K9me2 levels are increased upon contextual fear conditioning after 24 hours [84]. Pharmacological blocking of EHMT proteins in the hippocampus and lateral amygdala reduced long-term memory (LTM) upon contextual fear conditioning and auditory fear conditioning [84, 85]. In mouse knockout models, loss of *Ehmt1* causes phenotypes that are reminiscent of Kleefstra syndrome, including deficits in learning and memory, increased anxiety, hypotonia, cranial abnormalities, and developmental delay [86-89].

More recent work has investigated the role of EHMT1 in synaptic plasticity in cultured neurons. Cultured neurons in general tend to self-assemble into functional networks, from which the connectivity and activity can be measured using multi electrode arrays and whole-cell patch-clamp [90]. Upon knockdown of *EHMT1*, these neurons show impaired spontaneous network activity and lowered firing rates [91]. Additionally, after a period of delayed network bursting these networks show irregularity in the timing of network bursts [91]. Proper formation of neuronal networks is also essential for synaptic scaling, a process affected in EHMT deficient neurons *in vivo* and *in vitro* [92]. During synaptic scaling, EHMT1 and EHMT2 are required for brain-derived neurotrophic factor (BDNF) repression via H3K9me2 [92]. Dimethylation of H3K9 at the promoter of BDNF is shown to be dynamically regulated in an activity-dependent manner. This altered synaptic scaling could be related to neurogenesis in the dentate gyrus region of the hippocampus, shown to be increased in EHMT1 deficient neurons [93]. Thus, EHMT proteins are suggested to be essential for synaptic plasticity.

EHMT in *Drosophila melanogaster*

Drosophila has a single EHMT ortholog called G9a. Similar to the mammalian counterparts, it contains a series of ankyrin domains, followed by C-terminal pre-SET and SET domains **(chapter 5, figure 2A**). In *Drosophila* embryos, G9a translocates from the cytoplasm to the nucleus during nuclear division 8-11 [94]. In the nucleus, it is shown that G9a is responsible for downregulation of genes involved in many neurodevelopmental processes such as notch signaling and neuron differentiation [94]. *G9a* mutant flies are fully viable [95], but do show phenotypic differences compared to wildtype flies. Homozygous knockout *Drosophila* embryos are developmental delayed as observed by a longer time to hatch [94]. As flies progress into larval stages, G9a is required for proper dendrite branching of the sensory multidendrite neurons [96]. Additionally, larval foraging behaviour based on crawling is significantly reduced [96].

At adult stages, other cognitive functions can be tested such as habituation and courtship conditioning. Habituation is a simple form of non-associative learning and can be tested in flies using the light-off jump-reflex paradigm. An initial strong behavioural

response towards a repeated, non-threatening stimulus gradually wanes in control flies [97-99]. However, this decline was reduced in *G9a* knockout flies[96]. A more complex form of memory, *Drosophila* courtship conditioning, was tested in which male flies learn to discriminate between receptive and non-receptive females [100-102] (**chapter 2**). Trained control males reduce their courtship attempts after a training period with non-receptive females compared to socially naïve males. *G9a* mutant males are shown to have courtship conditioning short-term memory (STM) and LTM defects while the general learning capacity was unchanged [96] (**Figure 2A**). Moreover, it was shown that upon re-expression of G9a in the whole nervous system or neurons labeled by *7B-gal4* this courtship STM memory could be restored [96] **(Figure 2B**). *7B-gal4* is mainly expressed in the mushroom bodies, a brain region known to be involved in many forms learning and memory [103-105]. Lastly, it was shown that re-expression of G9a only in adult stages also rescues the STM phenotype **(Figure 2C**). Thus, G9a is involved in STM and LTM, but not in general learning. The memory phenotype can be rescued by re-expression of G9a in distinct brain regions such as the mushroom body or only in adult stages excluding developmental influences.

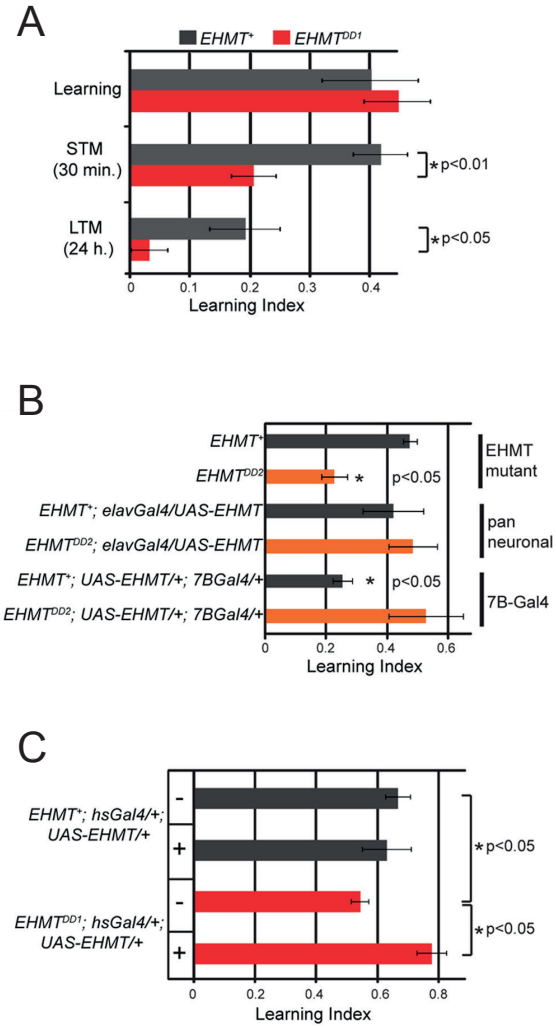


Figure 2: G9a (EHMT) in courtship conditioning learning and memory.

(A) The learning index (LI) of G9a mutant males ($EHMT^{DD1}$) was not in affected in general learning, but was significantly reduced at 30 minutes (short-term memory, STM) and 24 hours (long-term memory, LTM) after training. (B) The learning index of the $EHMT^{DD2}$ was affected in STM and was rescued by re-expression of G9a by the pan-neuronal driver elav-Gal4 and the mushroom body driver 7B-Gal4. Note the reduction of LI upon re-expression of G9a in the 7B-Gal4 expressing neurons, while the LI is not reduced in pan-neuronal re-expression (black bars). (C) Re-expression of G9a in adult stages by heat-shock induction did not alter the LI the control line (upper two bars) while the LI is significantly improved in the mutant background. + indicates G9a re-expression activated, - indicates not activated. * indicates a significant difference (Kruskall-Wallis and Mann-Whitney tests). This figure is adapted from Kramer *et al.* (2011) [96].

Kleefstra syndrome phenotypic spectrum[#]

Kleefstra syndrome is defined by mutations in *EHMT1* with the core features ID, autism spectrum disorder (ASD), childhood hypotonia, and distinctive facial features. However, within a clinically defined cohort of individuals with the core features of Kleefstra syndrome there is also heterogeneity of for instance renal anomalies and hearing loss. Around 25% of this cohort harbors *EHMT1* loss-of-function mutations. The other 75% was termed Kleefstra syndrome phenotypic spectrum (KSS) and it was hypothesized that these individuals have mutations in genes that share a biological function with EHMT1. Via a next generation sequencing approach, four genes with potentially causative *de novo* mutations were identified in four patients [106]. These genes are *methyl binding domain 5* (*MBD5*), *SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1* (*SMARCB1*), *nuclear receptor 1I3* (*NR1I3*), and *mixed-lineage leukemia 3* (*MLL3* (*KMT2C)*). The biochemical function of the proteins these genes encode have been well characterized. MBD5 contains a methyl binding domain which is required for localization to chromatin [107]. SMARCB1 is a member of the switch/sucrose nonfermentable (SWI/SNF) family of ATP-dependent chromatin-remodeling complexes, which affects transcription by destabilizing histone-DNA interactions and altering nucleosome positions [108]. NR1I3 is a nuclear hormone receptor that is associated with recruitment of chromatin-modifying complexes [109]. KMT2C trimethylates histone H3 at lysine 4 and is a central component of the activating signal cointegrator-2 (ASC-2) complex ASCOM, which acts as a transcriptional coactivator for nuclear hormone receptors [110, 111]. Taken together, mutations in multiple genes with comparable function give rise to a striking similar phenotype.

To validate a functional relationship between *EHMT1* and the four genes with *de novo* mutations, genetic interaction studies in the developing *Drosophila* wing were conducted [112]. Overexpression of *Drosophila G9a* in the wing consistently causes extra veins in defined regions of the wing [106]. Modulation of this *G9a*-induced wing phenotype was performed by manipulating expression of the orthologs of the four genes (*six banded* (*sba*) for *MBD5*, *trithorax related* (*trr*) for *KMT2C*, *Snf5-related-1* (*snr1*) for *SMARCB1* and the *ecdysone receptor* (*EcR*) for *NR1I3*) [106]. A synergistic interaction with G9a was established for MBD5/snr1 and NR1I3/EcR (**Figure 3**, green dotted lines) and an antagonistic interaction was established with KMT2C/trr (**Figure 3**, red dotted line). Together with published genetic interactions and protein-protein interactions, these studies identified an epigenetic module underlying clinically related ID syndromes (**Figure 3**) [106].

[#]: published in Kleefstra et al., 2012 (AJHG) [16]

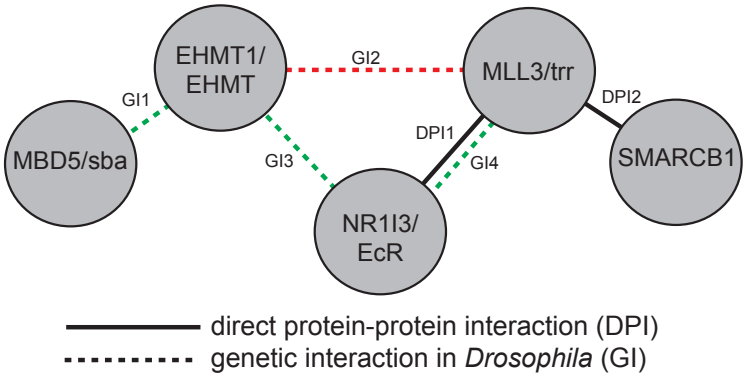


Figure 3[#]: An epigenetic network underlying KSS.

Functional studies indicate that genes implicated in KSS occur in a common chromatin-regulating module. This evidence comes from investigation of direct protein-protein interactions (solid lines) and from genetic interaction studies with *Drosophila melanogaster* (dashed lines). Green dashed lines indicate a synergistic interaction, and red dashed lines indicate an antagonistic interaction. It has been demonstrated in this study that *Drosophila EHMT* interacts genetically with *sba/MBD5, trr/MLL3*, and*EcR/NR1I3* (GI1, GI2, and GI3, respectively). Previously, genetic and physical interactions between trr and EcR (GI4 and DPI1, respectively), as well as physical association between SMARCB1 and MLL3 (DPI2), have been demonstrated. Figure from Kleefstra et al. (2012) and reprinted with permission.

Interestingly, the genetic interaction between *trr* and G9a was very strong. *Trr* knockdown alone in the fly wing caused a mild phenotype, consisting of mild loss of wing veins and mild upwards curvature of the wing. Combining *trr* knockdown with *G9a* overexpression resulted in fully penetrant pupal lethality resulting from necrosis of the entire developing wing tissue [106]. Thus, this dramatic compound phenotype indicates a very strong antagonistic relationship between *trr* and *G9a* and the human orthologs are involved in clinically related syndromes. This interaction is further investigated and described in chapter 3 of this thesis.

The KMT protein family

Another important player in the field of epigenetics crucial to cognitive functioning is the histone lysine methyltransferase (KMT) family, previously known as mixed lineage leukemia (MLL). The *KMT* genes were first identified by the *Drosophila* screens for enhancers and suppressors of position-effect variegation [113, 114]. Many of these hits turned out to contain a SET domain protein sequence, similar to the EHMT protein family domain, that catalyze mono-, di-, and trimethylation of lysine residues of histones [53]. The human *MLL1*

[#]: published in Kleefstra et al., 2012 (AJHG) [16]

gene was first cloned after its identification of translocations in patients with hematological malignancies [115]. The complete human KMT protein family was identified later and contains Set1A (KMT2F), Set1B (KMT2G), and KMT2A, KMT2B, KMT2C, KMT2D, and KMT2E [116]. However, KMT2E appears to lack histone methyltransferase activity [117], which in principle disqualifies the "KMT" name. KMT2A/KMT2B, KMT2C/KMT2D and SET1A/SET1B, are the three branches of the conserved human protein family and are direct orthologs of yeast Set1 [118] (**Figure 4**). The six human proteins contain Ash2L, RbBP5, WDR5, and Dpy30 as core components [118] (**Figure 4**).

Besides these common core components, SET1A and SET1B contain the specific structural and functional binding partners CXXC1, HCF1 and WDR82 [119, 120] (**Figure 4**). These complexes are key regulators of gene expression during development via trimethylation of H3K4 in mammalian cells [121]. KMT2A and KMT2B contain the specific structural and functional binding partners "menin tumor suppressor protein" and "HCF1" (**Figure 4**). KMT2B has been shown to trimethylate H3K4 at bivalent promoters of homeotic genes in mouse embryonic stem cells important for developmental processes [122]. KMT2C and KMT2D contain the specific structural and functional binding partners UTX (KDM6A), the PAX interacting protein 1 (PTIP), nuclear receptor coactivator 6 (NCOA6), and PTIP-associated 1 (PA1) [116] (**Figure 4**). KMT2C and KMT2D are involved in regulating hormone responsive genes [123]. In addition, KMT2C/KMT2D are associated with monomethylation of H3K4, a mark associated with active enhancers [124].

Next to the COMPASS complex, KMT2C/KMT2D have been associated with the activator signal cointegrator-2 (ASC-2) complex (ASCOM) complex. This large protein complex interacts with nuclear receptors and mediates H3K4 trimethylation, a hallmark of active transcription [110, 125]. In addition, SMARCB1 and KMT2C have been physically associated and were required for recruitment of the ATP dependant SWI/SNF complex and ASCOM to hormone response elements [126]. Although ASCOM and the COMPASS KMT2C/D branch have several subunits in common, there are clear distinctions. First, hDPY-30, PTIP, PA1, and UTX are absent from ASCOM [127, 128]. Second, α- and β-tubulins are absent from the COMPASS complex [128]. Third, KMT2C/D shows a more robust methyltransferase activity [127]. Taken together, the KMT protein family is a large protein family of conserved members involved in H3K4 methylation important for transcriptional regulation during development.

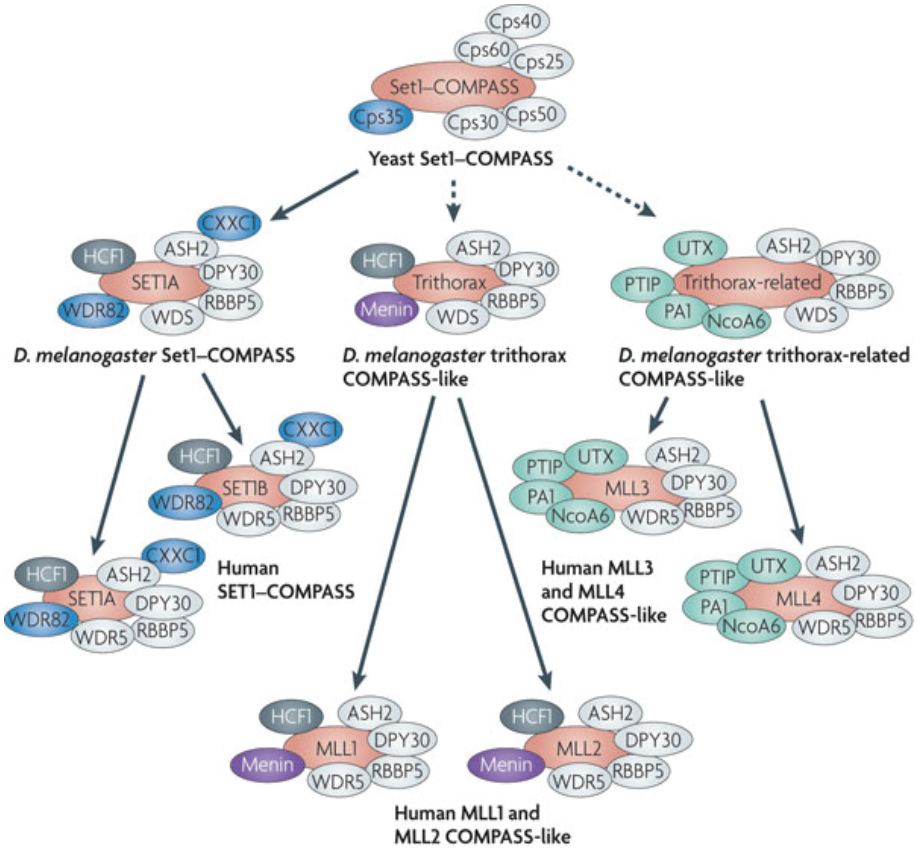


Figure 4: The yeast SET1 (COMPASS) is conserved in *Drosophila melanogaster* and human in three branches.

Yeast SET1-COMPASS is the sole H3K4 methyltransferase capable of monomethylation, dimethylation, and trimethylation. In *Drosophila*, there are three methyltransferases. SET1A is is the ortholog of yeast SET1-COMPASS shown by a solid arrow and two COMPASS-like complexes shown by a dashed arrows: trithorax and trithorax-related. In humans, SET1A and SET1B are homologues of *Drosophila* SET1A and yeast SET1-COMPASS. KMT2A (MLL1) and KMTB (MLL2) are homologs of *Drosophila* trithorax. KMT2C (MLL3) and KMT2D (MLL4) are homologs of *Drosophila* trithorax-related. The known common subunits shared between yeast, *Drosophila* and the human complexes are shown in grey. COMPASS component SWD2 (Cps35) in yeast and *Drosophila* and its homologue in human, WD repeat-containing protein (WDR82), are found only in COMPASS and the SET1A and SET1B complexes. The common subunits in COMPASS from yeast to human are shown in blue. Menin, a subunit of the trithorax and MLL and MLL2 complexes, is shown in purple. The shared subunits among the trithorax-related and MLL3 and MLL4 complexes are shown in green. Figure from Mohan *et al*. (2010) and reprinted with permission.

Drosophila trithorax related (trr)

The three branches of the human COMPASS complex are conserved in *Drosophila* by the single orthologs dSET1, trithorax (trx) and trithorax-related (trr) [123] (**Figure 4**). dSET1 is orthologous to human SET1A/B, trx is orthologous to human KMT2A/B, and trr is orthologous to human KMT2C/KMT2D. *Drosophila* dSET1 has been shown to be the main H3K4 di- and tri-methyltransferase during *Drosophila* development [129]. *Drosophila* trx mainly regulates homeotic gene expression [130]. Trr has a potential dual function in *Drosophila*. First, it interacts with the ecdysone receptor and is a co-activator of edcysone mediated gene expression during the larval to pupae transition [131]. In this developmental context, trr contains major H3K4 trimethylase activity [111]. Second, trr has been shown to contain H3K4 monomethylation activity [132]. Upon knockdown of trr in the wing imaginal disks, immunofluorescent imaging revealed a profound decrease of H3K4me1 which is the histone modification associated with enhancer functioning [26, 132, 133]. Moreover, chromatin immunoprecipitation experiments revealed trr binding sites mainly at promoters and enhancers [132]. Thus, the *Drosophila* orthologs of the COMPASS complexes are functionally conserved.

KMT-related proteins in ID

Autosomal dominant mutations in *KMT2A* have been associated with Wiedemann-Steiner syndrome (OMIM #605130) [134]. This syndrome is characterized by hypertrichosis cubiti (hairy elbows), short stature, facial features, mild to moderate ID, and behavioral difficulties [135]. A homozygous missense mutation in *KMT2B* have been identified in a Pakistani consanguineous family [136]. The three affected brothers had ID, short stature, childhood hypotonia and dysmorphic facial features. Remarkably, the phenotype of the boys had striking Kleefstra-like clinical features [136]. Mutations in *KMT2D* have been identified in Kabuki syndrome [137] (OMIM #147920) which is a rare, multiple malformation disorder characterized by postnatal dwarfism, a distinctive facial appearance, cardiac anomalies, skeletal abnormalities, immunological defects and mild to moderate ID [138] (**Chapter 4**). Kabuki syndrome 2 (OMIM #300867) is defined by mutations in *UTX* (*KDM6A*) and shows a phenotypic overlap [139, 140]. Moreover, a clinical overlap was found between the two types of Kabuki syndrome and Wiedemann-Steiner syndrome and comprises facial features, skeletal abnormalities of for instance the fingers, and neurological defects such as developmental delay, ID and hypotonia, growth problems, and problems of the internal organs [141]. Taken together, these ID syndromes with overlapping clinical features have shared underlying molecular mechanisms.

Modeling KSS in *Drosophila melanogaster*: Why the fly

Drosophila has been a model organism for fundamental biology and disease modeling for more than a century. Despite the evolutionary distance between human and fly, there are many similarities. First, comparative analysis showed a high conservation of important basic cellular and molecular processes between *Drosophila* and mammals [142]. Around 75% of human ID associated genes show sequence orthology [16]. Second, *Drosophila* is a very practical organism to work with. The generation time is very short and the housing and maintenance is inexpensive. Mutant fly lines can be generated by established protocols that for instance uses imprecise excision of transposable p-elements [143]. Briefly, these elements are incorporated into random positions of the genome. Imprecise DNA repair of the genome after p-element excision can also remove coding pieces of a gene of interest. This method has been used to generate the *G9a* null-mutant fly line [96], which is also used in the study described in chapter 3 of this thesis.

Besides the generation of a null-mutant, RNA interference (RNAi) can be used in *Drosophila* to study gene function. This can be done by expression of a double stranded RNA from a transgene. Using such transgenes in combination with the Gal4/UAS (upstream activating sequence) system, disruption of the expression levels of specific mRNA molecules can be achieved *in vivo* [144]. Worldwide stock centers have thousands of RNAi-lines and Gal4-lines that are available for research purposes. This enables gene specific knockdown in a time and tissue specific manner [145-147]. Thus, a gene of interest can be studied in a specific brain region or tissue as exemplified in **chapters 2, 3, and 5** of this thesis. Putative overexpression of a gene can be induced by cloning the coding sequence behind the UAS element. This has been used in **chapter 5** of this thesis. Thus, adding up all these benefits *Drosophila* has to offer, it provides a very good model to study Kleefstra syndrome and other clinical related ID syndromes.

Thesis objectives and outline

The functional analysis of genes involved in Kleefstra syndrome, KSS and other clinically related ID syndromes can help to provide insight in the molecular networks and mechanisms underlying human brain function in health and disease. I focus on the functional relationship between EHMT1/G9a and KMT2C/KMT2D/trr. In this way, I try to elucidate the following hypothesis:

Genes involved in clinically overlapping ID disorders such as Kleefstra syndrome, KSS, and other clinically related disorders encode proteins that function in common biological processes or pathways.

The aim of my research is to identify and understand these pathways.

The major objectives of my PhD thesis are:

1. To indentify the role of trr in *Drosophila* memory
2. To identify common gene expression regulation by factors associated with ID syndromes that have overlapping aspects with Kleefstra syndrome
3. To identify novel contributors of the EHMT1 molecular network

In order to achieve these objectives, I first optimized an established protocol for courtship conditioning to test learning, STM, and LTM in *Drosophila* [148] which is described in **chapter 2.** These optimizations include several novel practical aspects and an R-script to perform statistical validation. In **chapter 3** I show five additional patients with mutations in *KMT2C* which results in a clinically overlapping syndrome with Kleefstra syndrome and I performed genomic analysis to show that the *Drosophila* orthologs of EHMT1 and *KMT2C*, G9a and trr respectively, have molecular convergence. **Chapter 4** shows molecular convergence between EHMT1 and KMT2D. In this chapter I report differentially expressed genes in blood of Kleefstra syndrome (*EHMT1* mutations) and Kabuki syndrome (*KMT2D* mutations) and a significant subset of these genes that are commonly misexpressed. **Chapter 5** reports on the identification of novel contributors to the EHMT1 molecular network. I identified a novel zinc finger protein (CG9932) as interaction partner of *Drosophila* G9a. Finally, the results presented in this thesis are discussed in **chapter 6** in which I also evaluate more syndromes that have clinical overlap with Kleefstra syndrome and discuss the respective molecular functions that partly overlap with the function of EHMT1. Furthermore, prospects for future work are provided.

References

1. American Psychiatric Association. 2013.
2. van Bokhoven H. Genetic and epigenetic networks in intellectual disabilities. Annual review of genetics. 2011;45:81-104. doi: 10.1146/annurev-genet-110410-132512. PubMed PMID: 21910631.
3. Ropers HH. Genetics of intellectual disability. Current opinion in genetics & development. 2008;18(3):241-50. doi: 10.1016/j.gde.2008.07.008. PubMed PMID: 18694825.
4. Deciphering Developmental Disorders S. Prevalence and architecture of *de novo* mutations in developmental disorders. Nature. 2017. doi: 10.1038/nature21062. PubMed PMID: 28135719.
5. Ropers HH, Hamel BC. X-linked mental retardation. Nature reviews Genetics. 2005;6(1):46-57. doi: 10.1038/nrg1501. PubMed PMID: 15630421.
6. Lejeune J, Turpin R, Gautier M. [Chromosomic diagnosis of mongolism]. Archives francaises de pediatrie. 1959;16:962-3. PubMed PMID: 14415503.
7. van Karnebeek CD, Jansweijer MC, Leenders AG, Offringa M, Hennekam RC. Diagnostic investigations in individuals with mental retardation: a systematic literature review of their usefulness. European journal of human genetics : EJHG. 2005;13(1):6-25. doi: 10.1038/sj.ejhg.5201279. PubMed PMID: 15523501.
8. Ellison JW, Rosenfeld JA, Shaffer LG. Genetic basis of intellectual disability. Annual review of medicine. 2013;64:441-50. doi: 10.1146/annurev-med-042711-140053. PubMed PMID: 23020879.
9. Vissers LE, de Vries BB, Veltman JA. Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis. Journal of medical genetics. 2010;47(5):289-97. doi: 10.1136/jmg.2009.072942. PubMed PMID: 19951919.
10. Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, et al. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature. 2011;478(7367):57-63. doi: 10.1038/nature10423. PubMed PMID: 21937992.
11. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, et al. A *de novo* paradigm for mental retardation. Nature genetics. 2010;42(12):1109-12. doi: 10.1038/ng.712. PubMed PMID: 21076407.
12. Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet. 2012;380(9854):1674-82. doi: 10.1016/S0140-6736(12)61480-9. PubMed PMID: 23020937.
13. Deciphering Developmental Disorders S. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2015;519(7542):223-8. doi: 10.1038/nature14135. PubMed PMID: 25533962.
14. Kochinke K, Zweier C, Nijhof B, Fenckova M, Cizek P, Honti F, et al. Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. American journal of human genetics. 2016;98(1):149-64. doi: 10.1016/j.ajhg.2015.11.024. PubMed PMID: 26748517; PubMed Central PMCID: PMC4716705.
15. Stessman HA, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M, et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nature genetics. 2017. doi: 10.1038/ng.3792. PubMed PMID: 28191889.

16. Oortveld MA, Keerthikumar S, Oti M, Nijhof B, Fernandes AC, Kochinke K, et al. Human intellectual disability genes form conserved functional modules in *Drosophila*. PLoS genetics. 2013;9(10):e1003911. doi: 10.1371/journal.pgen.1003911. PubMed PMID: 24204314; PubMed Central PMCID: PMC3814316.
17. van Bokhoven H, Kramer JM. Disruption of the epigenetic code: an emerging mechanism in mental retardation. Neurobiology of disease. 2010;39(1):3-12. doi: 10.1016/j.nbd.2010.03.010. PubMed PMID: 20304068.
18. Kramer JM, van Bokhoven H. Genetic and epigenetic defects in mental retardation. The international journal of biochemistry & cell biology. 2009;41(1):96-107. doi: 10.1016/j.biocel.2008.08.009. PubMed PMID: 18765296.
19. Kleefstra T, Schenck A, Kramer JM, van Bokhoven H. The genetics of cognitive epigenetics. Neuropharmacology. 2014;80:83-94. doi: 10.1016/j.neuropharm.2013.12.025. PubMed PMID: 24434855.
20. Waddington C. The epigenotype. Endeavour. 1942:18-20.
21. Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, et al. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. Cell. 2010;143(2):212-24. doi: 10.1016/j.cell.2010.09.009. PubMed PMID: 20888037; PubMed Central PMCID: PMC3119929.
22. Kundaje A, Kyriazopoulou-Panagiotopoulou S, Libbrecht M, Smith CL, Raha D, Winters EE, et al. Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. Genome research. 2012;22(9):1735-47. doi: 10.1101/gr.136366.111. PubMed PMID: 22955985; PubMed Central PMCID: PMCPMC3431490.
23. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74. doi: 10.1038/nature11247. PubMed PMID: 22955616; PubMed Central PMCID: PMCPMC3439153.
24. Zentner GE, Henikoff S. Regulation of nucleosome dynamics by histone modifications. Nature structural & molecular biology. 2013;20(3):259-66. doi: 10.1038/nsmb.2470. PubMed PMID: 23463310.
25. de Graaf CA, van Steensel B. Chromatin organization: form to function. Current opinion in genetics & development. 2013;23(2):185-90. doi: 10.1016/j.gde.2012.11.011. PubMed PMID: 23274160.
26. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823-37. doi: 10.1016/j.cell.2007.05.009. PubMed PMID: 17512414.
27. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128(4):693-705. doi: 10.1016/j.cell.2007.02.005. PubMed PMID: 17320507.
28. Tan M, Luo H, Lee S, Jin F, Yang JS, Montellier E, et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell. 2011;146(6):1016-28. doi: 10.1016/j.cell.2011.08.008. PubMed PMID: 21925322; PubMed Central PMCID: PMCPMC3176443.
29. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. Genes & development. 2016;30(7):733-50. doi: 10.1101/gad.276568.115. PubMed PMID: 27036965; PubMed Central PMCID: PMCPMC4826392.
30. Jin SG, Wu X, Li AX, Pfeifer GP. Genomic mapping of 5-hydroxymethylcytosine in the human brain. Nucleic acids research. 2011;39(12):5015-24. doi: 10.1093/nar/gkr120. PubMed PMID: 21378125; PubMed Central PMCID: PMCPMC3130285.

31. Wu TP, Wang T, Seetin MG, Lai Y, Zhu S, Lin K, et al. DNA methylation on N(6)-adenine in mammalian embryonic stem cells. Nature. 2016;532(7599):329-33. doi: 10.1038/nature17640. PubMed PMID: 27027282; PubMed Central PMCID: PMCPMC4977844.
32. Gilbert SF. Developmental biology. 2006;8th edition.
33. Arney KL, Fisher AG. Epigenetic aspects of differentiation. Journal of cell science. 2004;117(Pt 19):4355-63. doi: 10.1242/jcs.01390. PubMed PMID: 15331660.
34. Borrelli E, Nestler EJ, Allis CD, Sassone-Corsi P. Decoding the epigenetic language of neuronal plasticity. Neuron. 2008;60(6):961-74. doi: 10.1016/j.neuron.2008.10.012. PubMed PMID: 19109904; PubMed Central PMCID: PMC2737473.
35. Willemsen MH, Vulto-van Silfhout AT, Nillesen WM, Wissink-Lindhout WM, van Bokhoven H, Philip N, et al. Update on Kleefstra Syndrome. Molecular syndromology. 2012;2(3-5):202-12. doi: 000335648. PubMed PMID: 22670141; PubMed Central PMCID: PMC3366700.
36. Kleefstra T, de Leeuw N, Wolf R, Nillesen WM, Schobers G, Mieloo H, et al. Phenotypic spectrum of 20 novel patients with molecularly defined supernumerary marker chromosomes 15 and a review of the literature. American journal of medical genetics Part A. 2010;152A(9):2221-9. doi: 10.1002/ajmg.a.33529. PubMed PMID: 20683990.
37. Stewart DR, Kleefstra T. The chromosome 9q subtelomere deletion syndrome. American journal of medical genetics Part C, Seminars in medical genetics. 2007;145C(4):383-92. doi: 10.1002/ajmg.c.30148. PubMed PMID: 17910072.
38. Vermeulen K, de Boer A, Janzing JGE, Koolen DA, Ockeloen CW, Willemsen MH, et al. Adaptive and maladaptive functioning in Kleefstra syndrome compared to other rare genetic disorders with intellectual disabilities. American journal of medical genetics Part A. 2017. doi: 10.1002/ajmg.a.38280. PubMed PMID: 28498556.
39. Schimmenti LA, Berry SA, Tuchman M, Hirsch B. Infant with multiple congenital anomalies and deletion (9)(q34.3). American journal of medical genetics. 1994;51(2):140-2. doi: 10.1002/ajmg.1320510211. PubMed PMID: 7522397.
40. Ayyash H, Mueller R, Maltby E, Horsfield P, Telford N, Tyler R. A report of a child with a deletion (9)(q34.3): a recognisable phenotype? Journal of medical genetics. 1997;34(7):610-2. PubMed PMID: 9222977; PubMed Central PMCID: PMC1051009.
41. Stewart DR, Huang A, Faravelli F, Anderlid BM, Medne L, Ciprero K, et al. Subtelomeric deletions of chromosome 9q: a novel microdeletion syndrome. American journal of medical genetics Part A. 2004;128A(4):340-51. doi: 10.1002/ajmg.a.30136. PubMed PMID: 15264279.
42. Kleefstra T, Brunner HG, Amiel J, Oudakker AR, Nillesen WM, Magee A, et al. Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. American journal of human genetics. 2006;79(2):370-7. doi: 10.1086/505693. PubMed PMID: 16826528; PubMed Central PMCID: PMC1559478.
43. Kleefstra T, van Zelst-Stams WA, Nillesen WM, Cormier-Daire V, Houge G, Foulds N, et al. Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype. Journal of medical genetics. 2009;46(9):598-606. doi: 10.1136/jmg.2008.062950. PubMed PMID: 19264732.
44. Harada N, Visser R, Dawson A, Fukamachi M, Iwakoshi M, Okamoto N, et al. A 1-Mb critical region in six patients with 9q34.3 terminal deletion syndrome. Journal of human genetics. 2004;49(8):440-4. doi: 10.1007/s10038-004-0166-z. PubMed PMID: 15258833.

45. Yatsenko SA, Cheung SW, Scott DA, Nowaczyk MJ, Tarnopolsky M, Naidu S, et al. Deletion 9q34.3 syndrome: genotype-phenotype correlations and an extended deletion in a patient with features of Opitz C trigonocephaly. Journal of medical genetics. 2005;42(4):328-35. doi: 10.1136/jmg.2004.028258. PubMed PMID: 15805160; PubMed Central PMCID: PMC1736036.
46. Kleefstra T, Koolen DA, Nillesen WM, de Leeuw N, Hamel BC, Veltman JA, et al. Interstitial 2.2 Mb deletion at 9q34 in a patient with mental retardation but without classical features of the 9q subtelomeric deletion syndrome. American journal of medical genetics Part A. 2006;140(6):618-23. doi: 10.1002/ajmg.a.31123. PubMed PMID: 16470689.
47. Kleefstra T, Smidt M, Banning MJ, Oudakker AR, Van Esch H, de Brouwer AP, et al. Disruption of the gene Euchromatin Histone Methyl Transferase1 (Eu-HMTase1) is associated with the 9q34 subtelomeric deletion syndrome. Journal of medical genetics. 2005;42(4):299-306. doi: 10.1136/jmg.2004.028464. PubMed PMID: 15805155; PubMed Central PMCID: PMC1736026.
48. Kleefstra T, Nillesen WM, Yntema HG. Kleefstra Syndrome. GeneReviews. Seattle (WA)2010
49. Tachibana M, Sugimoto K, Fukushima T, Shinkai Y. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. The Journal of biological chemistry. 2001;276(27):25309-17. doi: 10.1074/jbc.M101914200. PubMed PMID: 11316813.
50. Shinkai Y, Tachibana M. H3K9 methyltransferase G9a and the related molecule GLP. Genes & development. 2011;25(8):781-8. doi: 10.1101/gad.2027411. PubMed PMID: 21498567; PubMed Central PMCID: PMC3078703.
51. Kramer JM. Regulation of cell differentiation and function by the euchromatin histone methyltranserfases G9a and GLP. Biochemistry and cell biology = Biochimie et biologie cellulaire. 2016;94(1):26-32. doi: 10.1139/bcb-2015-0017. PubMed PMID: 26198080.
52. Collins RE, Northrop JP, Horton JR, Lee DY, Zhang X, Stallcup MR, et al. The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. Nature structural & molecular biology. 2008;15(3):245-50. doi: 10.1038/nsmb.1384. PubMed PMID: 18264113; PubMed Central PMCID: PMC2586904.
53. Herz HM, Garruss A, Shilatifard A. SET for life: biochemical activities and biological functions of SET domain-containing proteins. Trends in biochemical sciences. 2013;38(12):621-39. doi: 10.1016/j.tibs.2013.09.004. PubMed PMID: 24148750; PubMed Central PMCID: PMC3941473.
54. Weiss T, Hergeth S, Zeissler U, Izzo A, Tropberger P, Zee BM, et al. Histone H1 variant-specific lysine methylation by G9a/KMT1C and Glp1/KMT1D. Epigenetics & chromatin. 2010;3(1):7. doi: 10.1186/1756-8935-3-7. PubMed PMID: 20334638; PubMed Central PMCID: PMC2860349.
55. Rice JC, Briggs SD, Ueberheide B, Barber CM, Shabanowitz J, Hunt DF, et al. Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. Molecular cell. 2003;12(6):1591-8. PubMed PMID: 14690610.
56. Yokochi T, Poduch K, Ryba T, Lu J, Hiratani I, Tachibana M, et al. G9a selectively represses a class of late-replicating genes at the nuclear periphery. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(46):19363-8. doi: 10.1073/pnas.0906142106. PubMed PMID: 19889976; PubMed Central PMCID: PMC2780741.
57. Peters AH, Kubicek S, Mechtler K, O'Sullivan RJ, Derijck AA, Perez-Burgos L, et al. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Molecular cell. 2003;12(6):1577-89. PubMed PMID: 14690609.

58. Kubicek S, O'Sullivan RJ, August EM, Hickey ER, Zhang Q, Teodoro ML, et al. Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase. Molecular cell. 2007;25(3):473-81. doi: 10.1016/j.molcel.2007.01.017. PubMed PMID: 17289593.
59. Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, Ohki M, et al. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes & development. 2002;16(14):1779-91. doi: 10.1101/gad.989402. PubMed PMID: 12130538; PubMed Central PMCID: PMC186403.
60. Tachibana M, Ueda J, Fukuda M, Takeda N, Ohta T, Iwanari H, et al. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes & development. 2005;19(7):815-26. doi: 10.1101/gad.1284005. PubMed PMID: 15774718; PubMed Central PMCID: PMC1074319.
61. Kim HT, Jeong SG, Cho GW. G9a inhibition promotes neuronal differentiation of human bone marrow mesenchymal stem cells through the transcriptional induction of RE-1 containing neuronal specific genes. Neurochemistry international. 2016;96:77-83. doi: 10.1016/j.neuint.2016.03.002. PubMed PMID: 26952575.
62. Ooi L, Wood IC. Chromatin crosstalk in development and disease: lessons from REST. Nature reviews Genetics. 2007;8(7):544-54. doi: 10.1038/nrg2100. PubMed PMID: 17572692.
63. Mozzetta C, Pontis J, Ait-Si-Ali S. Functional Crosstalk Between Lysine Methyltransferases on Histone Substrates: The Case of G9A/GLP and Polycomb Repressive Complex 2. Antioxidants & redox signaling. 2015;22(16):1365-81. doi: 10.1089/ars.2014.6116. PubMed PMID: 25365549; PubMed Central PMCID: PMC4432786.
64. Mozzetta C, Pontis J, Fritsch L, Robin P, Portoso M, Proux C, et al. The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. Molecular cell. 2014;53(2):277-89. doi: 10.1016/j.molcel.2013.12.005. PubMed PMID: 24389103.
65. Fritsch L, Robin P, Mathieu JR, Souidi M, Hinaux H, Rougeulle C, et al. A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. Molecular cell. 2010;37(1):46-56. doi: 10.1016/j.molcel.2009.12.017. PubMed PMID: 20129054.
66. Ogawa H, Ishiguro K, Gaubatz S, Livingston DM, Nakatani Y. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. Science. 2002;296(5570):1132-6. doi: 10.1126/science.1069861. PubMed PMID: 12004135.
67. Nishio H, Walsh MJ. CCAAT displacement protein/cut homolog recruits G9a histone lysine methyltransferase to repress transcription. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(31):11257-62. doi: 10.1073/pnas.0401343101. PubMed PMID: 15269344; PubMed Central PMCID: PMC509191.
68. Ueda J, Tachibana M, Ikura T, Shinkai Y. Zinc finger protein Wiz links G9a/GLP histone methyltransferases to the co-repressor molecule CtBP. The Journal of biological chemistry. 2006;281(29):20120-8. doi: 10.1074/jbc.M603087200. PubMed PMID: 16702210.
69. Bian C, Chen Q, Yu X. The zinc finger proteins ZNF644 and WIZ regulate the G9a/GLP complex for gene repression. eLife. 2015;4. doi: 10.7554/eLife.05606. PubMed PMID: 25789554; PubMed Central PMCID: PMC4365668.

70. Simon JM, Parker JS, Liu F, Rothbart SB, Ait-Si-Ali S, Strahl BD, et al. A Role for Widely Interspaced Zinc Finger (WIZ) in Retention of the G9a Methyltransferase on Chromatin. The Journal of biological chemistry. 2015;290(43):26088-102. doi: 10.1074/jbc.M115.654459. PubMed PMID: 26338712; PubMed Central PMCID: PMC4646261.
71. Shi Y, Sawada J, Sui G, Affar el B, Whetstine JR, Lan F, et al. Coordinated histone modifications mediated by a CtBP co-repressor complex. Nature. 2003;422(6933):735-8. doi: 10.1038/nature01550. PubMed PMID: 12700765.
72. Ohno H, Shinoda K, Ohyama K, Sharp LZ, Kajimura S. EHMT1 controls brown adipose cell fate and thermogenesis through the PRDM16 complex. Nature. 2013;504(7478):163-7. doi: 10.1038/nature12652. PubMed PMID: 24196706; PubMed Central PMCID: PMC3855638.
73. Wang L, Xu S, Lee JE, Baldridge A, Grullon S, Peng W, et al. Histone H3K9 methyltransferase G9a represses PPARgamma expression and adipogenesis. The EMBO journal. 2013;32(1):45-59. doi: 10.1038/emboj.2012.306. PubMed PMID: 23178591; PubMed Central PMCID: PMC3545301.
74. Kovalchuk Y, Eilers J, Lisman J, Konnerth A. NMDA receptor-mediated subthreshold Ca(2+) signals in spines of hippocampal neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2000;20(5):1791-9. PubMed PMID: 10684880.
75. Sobczyk A, Scheuss V, Svoboda K. NMDA receptor subunit-dependent [Ca2+] signaling in individual hippocampal dendritic spines. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2005;25(26):6037-46. doi: 10.1523/JNEUROSCI.1221-05.2005. PubMed PMID: 15987933.
76. Diering GH, Gustina AS, Huganir RL. PKA-GluA1 coupling via AKAP5 controls AMPA receptor phosphorylation and cell-surface targeting during bidirectional homeostatic plasticity. Neuron. 2014;84(4):790-805. doi: 10.1016/j.neuron.2014.09.024. PubMed PMID: 25451194; PubMed Central PMCID: PMCPMC4254581.
77. Lisman J, Yasuda R, Raghavachari S. Mechanisms of CaMKII action in long-term potentiation. Nature reviews Neuroscience. 2012;13(3):169-82. doi: 10.1038/nrn3192. PubMed PMID: 22334212; PubMed Central PMCID: PMCPMC4050655.
78. Josselyn SA, Shi C, Carlezon WA, Jr., Neve RL, Nestler EJ, Davis M. Long-term memory is facilitated by cAMP response element-binding protein overexpression in the amygdala. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2001;21(7):2404-12. PubMed PMID: 11264314.
79. Stern CM, Luoma JI, Meitzen J, Mermelstein PG. Corticotropin releasing factor-induced CREB activation in striatal neurons occurs via a novel Gbetagamma signaling pathway. PloS one. 2011;6(3):e18114. doi: 10.1371/journal.pone.0018114. PubMed PMID: 21448293; PubMed Central PMCID: PMCPMC3063246.
80. Yan X, Liu J, Ye Z, Huang J, He F, Xiao W, et al. CaMKII-Mediated CREB Phosphorylation Is Involved in Ca2+-Induced BDNF mRNA Transcription and Neurite Outgrowth Promoted by Electrical Stimulation. PloS one. 2016;11(9):e0162784. doi: 10.1371/journal.pone.0162784. PubMed PMID: 27611779; PubMed Central PMCID: PMCPMC5017744.
81. Cortes-Mendoza J, Diaz de Leon-Guerrero S, Pedraza-Alva G, Perez-Martinez L. Shaping synaptic plasticity: the role of activity-mediated epigenetic regulation on gene transcription. Int J Dev Neurosci. 2013;31(6):359-69. doi: 10.1016/j.ijdevneu.2013.04.003. PubMed PMID: 23665156.

82. Nonaka M, Kim R, Sharry S, Matsushima A, Takemoto-Kimura S, Bito H. Towards a better understanding of cognitive behaviors regulated by gene expression downstream of activity-dependent transcription factors. Neurobiology of learning and memory. 2014;115:21-9. doi: 10.1016/j.nlm.2014.08.010. PubMed PMID: 25173698.
83. Gupta S, Kim SY, Artis S, Molfese DL, Schumacher A, Sweatt JD, et al. Histone methylation regulates memory formation. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2010;30(10):3589-99. doi: 10.1523/JNEUROSCI.3732-09.2010. PubMed PMID: 20219993; PubMed Central PMCID: PMC2859898.
84. Gupta-Agarwal S, Franklin AV, Deramus T, Wheelock M, Davis RL, McMahon LL, et al. G9a/GLP histone lysine dimethyltransferase complex activity in the hippocampus and the entorhinal cortex is required for gene activation and silencing during memory consolidation. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2012;32(16):5440-53. doi: 10.1523/JNEUROSCI.0147-12.2012. PubMed PMID: 22514307; PubMed Central PMCID: PMC3332335.
85. Gupta-Agarwal S, Jarome TJ, Fernandez J, Lubin FD. NMDA receptor- and ERK-dependent histone methylation changes in the lateral amygdala bidirectionally regulate fear memory formation. Learning & memory. 2014;21(7):351-62. doi: 10.1101/lm.035105.114. PubMed PMID: 24939839; PubMed Central PMCID: PMC4061426.
86. Balemans MC, Huibers MM, Eikelenboom NW, Kuipers AJ, van Summeren RC, Pijpers MM, et al. Reduced exploration, increased anxiety, and altered social behavior: Autistic-like features of euchromatin histone methyltransferase 1 heterozygous knockout mice. Behavioural brain research. 2010;208(1):47-55. doi: 10.1016/j.bbr.2009.11.008. PubMed PMID: 19896504.
87. Balemans MC, Kasri NN, Kopanitsa MV, Afinowi NO, Ramakers G, Peters TA, et al. Hippocampal dysfunction in the Euchromatin histone methyltransferase 1 heterozygous knockout mouse model for Kleefstra syndrome. Human molecular genetics. 2013;22(5):852-66. doi: 10.1093/hmg/dds490. PubMed PMID: 23175442.
88. Balemans MC, Ansar M, Oudakker AR, van Caam AP, Bakker B, Vitters EL, et al. Reduced Euchromatin histone methyltransferase 1 causes developmental delay, hypotonia, and cranial abnormalities associated with increased bone gene expression in Kleefstra syndrome mice. Developmental biology. 2014;386(2):395-407. doi: 10.1016/j.ydbio.2013.12.016. PubMed PMID: 24362066.
89. Schaefer A, Sampath SC, Intrator A, Min A, Gertler TS, Surmeier DJ, et al. Control of cognition and adaptive behavior by the GLP/G9a epigenetic suppressor complex. Neuron. 2009;64(5):678-91. doi: 10.1016/j.neuron.2009.11.019. PubMed PMID: 20005824; PubMed Central PMCID: PMC2814156.
90. Martens MB, Chiappalone M, Schubert D, Tiesinga PH. Separating burst from background spikes in multichannel neuronal recordings using return map analysis. International journal of neural systems. 2014;24(4):1450012. doi: 10.1142/S0129065714500129. PubMed PMID: 24812717.
91. Bart Martens M, Frega M, Classen J, Epping L, Bijvank E, Benevento M, et al. Euchromatin histone methyltransferase 1 regulates cortical neuronal network development. Scientific reports. 2016;6:35756. doi: 10.1038/srep35756. PubMed PMID: 27767173; PubMed Central PMCID: PMC5073331.
92. Benevento M, Iacono G, Selten M, Ba W, Oudakker A, Frega M, et al. Histone Methylation by the Kleefstra Syndrome Protein EHMT1 Mediates Homeostatic Synaptic Scaling. Neuron. 2016;91(2):341-55. doi: 10.1016/j.neuron.2016.06.003. PubMed PMID: 27373831.

93. Benevento M, Oomen CA, Horner AE, Amiri H, Jacobs T, Pauwels C, et al. Haploinsufficiency of EHMT1 improves pattern separation and increases hippocampal cell proliferation. Scientific reports. 2017;7:40284. doi: 10.1038/srep40284. PubMed PMID: 28071689; PubMed Central PMCID: PMC5223204.
94. Shimaji K, Konishi T, Tanaka S, Yoshida H, Kato Y, Ohkawa Y, et al. Genomewide identification of target genes of histone methyltransferase dG9a during *Drosophila* embryogenesis. Genes to cells : devoted to molecular & cellular mechanisms. 2015;20(11):902-14. doi: 10.1111/gtc.12281. PubMed PMID: 26334932.
95. Seum C, Bontron S, Reo E, Delattre M, Spierer P. *Drosophila* G9a is a nonessential gene. Genetics. 2007;177(3):1955-7. doi: 10.1534/genetics.107.078220. PubMed PMID: 18039887; PubMed Central PMCID: PMC2147950.
96. Kramer JM, Kochinke K, Oortveld MA, Marks H, Kramer D, de Jong EK, et al. Epigenetic regulation of learning and memory by *Drosophila* EHMT/G9a. PLoS biology. 2011;9(1):e1000569. doi: 10.1371/journal.pbio.1000569. PubMed PMID: 21245904; PubMed Central PMCID: PMC3014924.
97. Willemsen MH, Nijhof B, Fenckova M, Nillesen WM, Bongers EM, Castells-Nobau A, et al. GATAD2B loss-of-function mutations cause a recognisable syndrome with intellectual disability and are associated with learning deficits and synaptic undergrowth in *Drosophila*. Journal of medical genetics. 2013;50(8):507-14. doi: 10.1136/jmedgenet-2012-101490. PubMed PMID: 23644463.
98. van Bon BW, Oortveld MA, Nijtmans LG, Fenckova M, Nijhof B, Besseling J, et al. CEP89 is required for mitochondrial metabolism and neuronal function in man and fly. Human molecular genetics. 2013;22(15):3138-51. doi: 10.1093/hmg/ddt170. PubMed PMID: 23575228.
99. Lugtenberg D, Reijnders MR, Fenckova M, Bijlsma EK, Bernier R, van Bon BW, et al. *De novo* loss-of-function mutations in WAC cause a recognizable intellectual disability syndrome and learning deficits in *Drosophila*. European journal of human genetics : EJHG. 2016;24(8):1145-53. doi: 10.1038/ejhg.2015.282. PubMed PMID: 26757981; PubMed Central PMCID: PMC4970694.
100. McBride SM, Giuliani G, Choi C, Krause P, Correale D, Watson K, et al. Mushroom body ablation impairs short-term memory and long-term memory of courtship conditioning in *Drosophila melanogaster*. Neuron. 1999;24(4):967-77. PubMed PMID: 10624959.
101. Kamyshev NG, Iliadi KG, Bragina JV. *Drosophila* conditioned courtship: two ways of testing memory. Learning & memory. 1999;6(1):1-20. PubMed PMID: 10355520; PubMed Central PMCID: PMC311276.
102. Keleman K, Kruttner S, Alenius M, Dickson BJ. Function of the *Drosophila* CPEB protein Orb2 in long-term courtship memory. Nature neuroscience. 2007;10(12):1587-93. doi: 10.1038/nn1996. PubMed PMID: 17965711.
103. Heisenberg M. Mushroom body memoir: from maps to models. Nature reviews Neuroscience. 2003;4(4):266-75. doi: 10.1038/nrn1074. PubMed PMID: 12671643.
104. Davis RL. Traces of *Drosophila* memory. Neuron. 2011;70(1):8-19. doi: 10.1016/j.neuron.2011.03.012. PubMed PMID: 21482352; PubMed Central PMCID: PMC3374581.
105. Margulies C, Tully T, Dubnau J. Deconstructing memory in *Drosophila*. Current biology : CB. 2005;15(17):R700-13. doi: 10.1016/j.cub.2005.08.024. PubMed PMID: 16139203; PubMed Central PMCID: PMC3044934.

106. Kleefstra T, Kramer JM, Neveling K, Willemsen MH, Koemans TS, Vissers LE, et al. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. American journal of human genetics. 2012;91(1):73-82. doi: 10.1016/j.ajhg.2012.05.003. PubMed PMID: 22726846; PubMed Central PMCID: PMC3397275.
107. Laget S, Joulie M, Le Masson F, Sasai N, Christians E, Pradhan S, et al. The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. PloS one. 2010;5(8):e11982. doi: 10.1371/journal.pone.0011982. PubMed PMID: 20700456; PubMed Central PMCID: PMCPMC2917364.
108. Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. Nat Rev Cancer. 2011;11(7):481-92. doi: 10.1038/nrc3068. PubMed PMID: 21654818.
109. Aranda A, Pascual A. Nuclear hormone receptors and gene expression. Physiological reviews. 2001;81(3):1269-304. PubMed PMID: 11427696.
110. Kim DH, Lee J, Lee B, Lee JW. ASCOM controls farnesoid X receptor transactivation through its associated histone H3 lysine 4 methyltransferase activity. Molecular endocrinology. 2009;23(10):1556-62. doi: 10.1210/me.2009-0099. PubMed PMID: 19556342; PubMed Central PMCID: PMC2754897.
111. Sedkov Y, Cho E, Petruk S, Cherbas L, Smith ST, Jones RS, et al. Methylation at lysine 4 of histone H3 in ecdysone-dependent development of *Drosophila*. Nature. 2003;426(6962):78-83. doi: 10.1038/nature02080. PubMed PMID: 14603321; PubMed Central PMCID: PMC2743927.
112. Bier E. *Drosophila*, the golden bug, emerges as a tool for human genetics. Nature reviews Genetics. 2005;6(1):9-23. doi: 10.1038/nrg1503. PubMed PMID: 15630418.
113. Tschiersch B, Hofmann A, Krauss V, Dorn R, Korge G, Reuter G. The protein encoded by the *Drosophila* position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. The EMBO journal. 1994;13(16):3822-31. PubMed PMID: 7915232; PubMed Central PMCID: PMC395295.
114. Stassen MJ, Bailey D, Nelson S, Chinwalla V, Harte PJ. The *Drosophila* trithorax proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. Mechanisms of development. 1995;52(2-3):209-23. PubMed PMID: 8541210.
115. Gu Y, Nakamura T, Alder H, Prasad R, Canaani O, Cimino G, et al. The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to *Drosophila* trithorax, to the AF-4 gene. Cell. 1992;71(4):701-8. PubMed PMID: 1423625.
116. Shilatifard A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annual review of biochemistry. 2012;81:65-95. doi: 10.1146/annurev-biochem-051710-134100. PubMed PMID: 22663077; PubMed Central PMCID: PMC4010150.
117. Sebastian S, Sreenivas P, Sambasivan R, Cheedipudi S, Kandalla P, Pavlath GK, et al. MLL5, a trithorax homolog, indirectly regulates H3K4 methylation, represses cyclin A2 expression, and promotes myogenic differentiation. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(12):4719-24. doi: 10.1073/pnas.0807136106. PubMed PMID: 19264965; PubMed Central PMCID: PMC2651835.

118. Takahashi YH, Westfield GH, Oleskie AN, Trievel RC, Shilatifard A, Skiniotis G. Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(51):20526-31. doi: 10.1073/pnas.1109360108. PubMed PMID: 22158900; PubMed Central PMCID: PMC3251153.
119. Wysocka J, Myers MP, Laherty CD, Eisenman RN, Herr W. Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. Genes & development. 2003;17(7):896-911. doi: 10.1101/gad.252103. PubMed PMID: 12670868; PubMed Central PMCID: PMCPMC196026.
120. Lee JH, Tate CM, You JS, Skalnik DG. Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. The Journal of biological chemistry. 2007;282(18):13419-28. doi: 10.1074/jbc.M609809200. PubMed PMID: 17355966.
121. Wu M, Wang PF, Lee JS, Martin-Brown S, Florens L, Washburn M, et al. Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Molecular and cellular biology. 2008;28(24):7337-44. doi: 10.1128/MCB.00976-08. PubMed PMID: 18838538; PubMed Central PMCID: PMCPMC2593441.
122. Hu D, Garruss AS, Gao X, Morgan MA, Cook M, Smith ER, et al. The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nature structural & molecular biology. 2013;20(9):1093-7. doi: 10.1038/nsmb.2653. PubMed PMID: 23934151; PubMed Central PMCID: PMC3805109.
123. Morgan MA, Shilatifard A. *Drosophila* SETs its sights on cancer: Trr/MLL3/4 COMPASS-like complexes in development and disease. Molecular and cellular biology. 2013;33(9):1698-701. doi: 10.1128/MCB.00203-13. PubMed PMID: 23459940; PubMed Central PMCID: PMC3624181.
124. Hu D, Gao X, Morgan MA, Herz HM, Smith ER, Shilatifard A. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Molecular and cellular biology. 2013;33(23):4745-54. doi: 10.1128/MCB.01181-13. PubMed PMID: 24081332; PubMed Central PMCID: PMC3838007.
125. Ananthanarayanan M, Li Y, Surapureddi S, Balasubramaniyan N, Ahn J, Goldstein JA, et al. Histone H3K4 trimethylation by MLL3 as part of ASCOM complex is critical for NR activation of bile acid transporter genes and is downregulated in cholestasis. American journal of physiology Gastrointestinal and liver physiology. 2011;300(5):G771-81. doi: 10.1152/ajpgi.00499.2010. PubMed PMID: 21330447; PubMed Central PMCID: PMC3094144.
126. Lee S, Kim DH, Goo YH, Lee YC, Lee SK, Lee JW. Crucial roles for interactions between MLL3/4 and INI1 in nuclear receptor transactivation. Molecular endocrinology. 2009;23(5):610-9. doi: 10.1210/me.2008-0455. PubMed PMID: 19221051; PubMed Central PMCID: PMC2675954.
127. Cho YW, Hong T, Hong S, Guo H, Yu H, Kim D, et al. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. The Journal of biological chemistry. 2007;282(28):20395-406. doi: 10.1074/jbc.M701574200. PubMed PMID: 17500065; PubMed Central PMCID: PMCPMC2729684.
128. Goo YH, Sohn YC, Kim DH, Kim SW, Kang MJ, Jung DJ, et al. Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. Molecular and cellular biology. 2003;23(1):140-9. PubMed PMID: 12482968; PubMed Central PMCID: PMC140670.

129. Hallson G, Hollebakken RE, Li T, Syrzycka M, Kim I, Cotsworth S, et al. dSet1 is the main H3K4 di- and tri-methyltransferase throughout *Drosophila* development. Genetics. 2012;190(1):91-100. doi: 10.1534/genetics.111.135863. PubMed PMID: 22048023; PubMed Central PMCID: PMCPMC3249358.
130. Ingham PW. trithorax and the regulation of homeotic gene expression in *Drosophila*: a historical perspective. Int J Dev Biol. 1998;42(3):423-9. PubMed PMID: 9654027.
131. Johnston DM, Sedkov Y, Petruk S, Riley KM, Fujioka M, Jaynes JB, et al. Ecdysone- and NO-mediated gene regulation by competing EcR/Usp and E75A nuclear receptors during *Drosophila* development. Molecular cell. 2011;44(1):51-61. doi: 10.1016/j.molcel.2011.07.033. PubMed PMID: 21981918; PubMed Central PMCID: PMC3190167.
132. Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, et al. Enhancer-associated H3K4 monomethylation by Trithorax-related, the *Drosophila* homolog of mammalian Mll3/Mll4. Genes & development. 2012;26(23):2604-20. doi: 10.1101/gad.201327.112. PubMed PMID: 23166019; PubMed Central PMCID: PMC3521626.
133. Cheng J, Blum R, Bowman C, Hu D, Shilatifard A, Shen S, et al. A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. Molecular cell. 2014;53(6):979-92. doi: 10.1016/j.molcel.2014.02.032. PubMed PMID: 24656132; PubMed Central PMCID: PMC4031464.
134. Jones WD, Dafou D, McEntagart M, Woollard WJ, Elmslie FV, Holder-Espinasse M, et al. *De novo* mutations in MLL cause Wiedemann-Steiner syndrome. American journal of human genetics. 2012;91(2):358-64. doi: 10.1016/j.ajhg.2012.06.008. PubMed PMID: 22795537; PubMed Central PMCID: PMC3415539.
135. Polizzi A, Pavone P, Ciancio E, La Rosa C, Sorge G, Ruggieri M. Hypertrichosis cubiti (hairy elbow syndrome): a clue to a malformation syndrome. Journal of pediatric endocrinology & metabolism : JPEM. 2005;18(10):1019-25. PubMed PMID: 16355816.
136. Agha Z, Iqbal Z, Azam M, Ayub H, Vissers LE, Gilissen C, et al. Exome sequencing identifies three novel candidate genes implicated in intellectual disability. PloS one. 2014;9(11):e112687. doi: 10.1371/journal.pone.0112687. PubMed PMID: 25405613; PubMed Central PMCID: PMC4236113.
137. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature genetics. 2010;42(9):790-3. doi: 10.1038/ng.646. PubMed PMID: 20711175; PubMed Central PMCID: PMC2930028.
138. Bogershausen N, Wollnik B. Unmasking Kabuki syndrome. Clinical genetics. 2013;83(3):201-11. doi: 10.1111/cge.12051. PubMed PMID: 23131014.
139. Lederer D, Grisart B, Digilio MC, Benoit V, Crespin M, Ghariani SC, et al. Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. American journal of human genetics. 2012;90(1):119-24. doi: 10.1016/j.ajhg.2011.11.021. PubMed PMID: 22197486; PubMed Central PMCID: PMC3257878.
140. Miyake N, Mizuno S, Okamoto N, Ohashi H, Shiina M, Ogata K, et al. KDM6A point mutations cause Kabuki syndrome. Human mutation. 2013;34(1):108-10. doi: 10.1002/humu.22229. PubMed PMID: 23076834.
141. Miyake N, Tsurusaki Y, Koshimizu E, Okamoto N, Kosho T, Brown NJ, et al. Delineation of clinical features in Wiedemann-Steiner syndrome caused by KMT2A mutations. Clinical genetics. 2016;89(1):115-9. doi: 10.1111/cge.12586. PubMed PMID: 25810209.

142. Muqit MM, Feany MB. Modelling neurodegenerative diseases in Drosophila: a fruitful approach? Nature reviews Neuroscience. 2002;3(3):237-43. doi: 10.1038/nrn751. PubMed PMID: 11994755.
143. Lankenau DH, Gloor GB. In vivo gap repair in Drosophila: a one-way street with many destinations. BioEssays : news and reviews in molecular, cellular and developmental biology. 1998;20(4):317-27. doi: 10.1002/(SICI)1521-1878(199804)20:4<317::AID-BIES8>3.0.CO;2-M. PubMed PMID: 9619103.
144. Brand AH, Manoukian AS, Perrimon N. Ectopic expression in Drosophila. Methods in cell biology. 1994;44:635-54. PubMed PMID: 7707973.
145. Jenett A, Rubin GM, Ngo TT, Shepherd D, Murphy C, Dionne H, et al. A GAL4-driver line resource for Drosophila neurobiology. Cell reports. 2012;2(4):991-1001. doi: 10.1016/j.celrep.2012.09.011. PubMed PMID: 23063364; PubMed Central PMCID: PMC3515021.
146. Kvon EZ, Kazmar T, Stampfel G, Yanez-Cuna JO, Pagani M, Schernhuber K, et al. Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature. 2014;512(7512):91-5. doi: 10.1038/nature13395. PubMed PMID: 24896182.
147. Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, et al. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature. 2007;448(7150):151-6. doi: 10.1038/nature05954. PubMed PMID: 17625558.
148. Siegel RW, Hall JC. Conditioned responses in courtship behavior of normal and mutant Drosophila. Proceedings of the National Academy of Sciences of the United States of America. 1979;76(7):3430-4. PubMed PMID: 16592682; PubMed Central PMCID: PMC383839.

[1]Department of Human Genetics, Radboudumc, Nijmegen, The Netherlands
[2]Radboud Institute of Molecular Life Sciences, Nijmegen, The Netherlands
[3]Donders Institute for Brain, Cognition, and Behaviour, Centre for Neuroscience, Nijmegen, The Netherlands
[4]Research Institute of Molecular Pathology, Campus-Vienna-Biocenter 1, 1030 Vienna, Austria
[5]Department for Health Evidence, Radboud university medical center
[6]Janelia Research Campus, 19700 Helix Drive, Ashburn VA 20147, USA
[7]Department of Physiology & Pharmacology, Department of Biology, Western University, London, Ontario, Canada

Drosophila courtship conditioning as a measure of learning and memory

Tom S. Koemans[1,2,3]
Cornelia Oppitz[4]
Rogier A.T Donders[5]
Hans van Bokhoven[1,3]
Annette Schenck[1,3]
Krystyna Keleman[6]
Jamie M. Kramer[7]

Journal of Visualized Experiments, 2017 doi: 10.3791/55808

2

Short abstract

This protocol describes a *Drosophila* learning and memory assay called courtship conditioning. This classic assay is based on a reduction of male courtship behavior after sexual rejection by a non-receptive premated female. This natural form of behavioral plasticity can be used to test learning, short-term memory, and long-term memory.

Long abstract

Many insights into the molecular mechanisms underlying learning and memory have been elucidated through the use of simple behavioral assays in model organisms such as the fruit fly, *Drosophila melanogaster*. *Drosophila* is useful for understanding the basic neurobiology underlying cognitive deficits resulting from mutations in genes associated with human cognitive disorders, such as intellectual disability (ID) and autism. This work describes a methodology for testing learning and memory using a classic paradigm in *Drosophila* known as courtship conditioning. Male flies court females using a distinct pattern of easily recognizable behaviors. Premated females are not receptive to mating and will reject the male's copulation attempts. In response to this rejection, male flies reduce their courtship behavior. This learned reduction in courtship behavior is measured over time, serving as an indicator of learning and memory. The basic numerical output of this assay is the courtship index (CI), which is defined as the percentage of time that a male spends courting during a 10-min interval. The learning index (LI) is the relative reduction of CI in flies that have been exposed to a premated female compared to naïve flies with no previous social encounters. For the statistical comparison of LIs between genotypes, a randomization test with bootstrapping is used. To illustrate how the assay can be used to address the role of a gene relating to learning and memory, the pan-neuronal knockdown of *Dihydroxyacetone phosphate acyltransferase* (*Dhap-at*) was characterized here. The human ortholog of *Dhap-at*, *glyceronephosphate O-acyltransferase* (*GNPT*), is involved in rhizomelic chondrodysplasia punctata type 2, an autosomal-recessive syndrome characterized by severe ID. Using the courtship conditioning assay, it was determined that Dhap-at is required for long-term memory, but not for short-term memory. This result serves as a basis for further investigation of the underlying molecular mechanisms.

Introduction

The molecular mechanisms underlying learning and memory are conserved throughout many species. The pioneering work screening *Drosophila melanogaster* mutant lines for defects in olfactory learning and memory has provided key molecular insights into the processes underlying learning and memory [1]. These studies identified some of the first genes involved in learning and memory, such as *rutabaga* [2], *amnesiac* [3], and *dunce* [4], revealing a critical role for cyclic adenosine monophosphate (cAMP) signaling [5].

Early genetic screens for memory mutants were primarily conducted using olfactory conditioning. However, several methods to measure other forms of learning and memory have emerged over time. One of the most widely used learning and memory paradigms, and the assay described here, is known as courtship conditioning, which was first described by Siegel and Hall [6] and was later refined by several other research groups [7-9]. Courtship conditioning is dependent upon the presence of a specific pheromone, *cis*-vaccenyl acetate (cVA), on the female abdomen, which is deposited by the male during copulation. Sensing cVA on the female abdomen naturally reduces courtship behavior, and when coupled with the act of rejection by the female, the effect of cVA on male courtship reduction is dramatically enhanced. The response of male flies in this assay can be easily quantified by observing their distinct courtship behavior, which is characterized by orienting towards and following the female, tapping, extending and vibrating the wing, licking, and attempting copulation [10] (**Figure 1A**). Male flies learn to distinguish between receptive virgin and non-receptive mated females [11], and after sexual rejection, they display reduced courtship behavior towards non-receptive females for up to 9 days [8]. This natural behavior can be used to elucidate the mechanisms underlying learning, short term memory (STM), and long term memory (LTM) [8,9,12]. Learning is defined as the immediate reduction in courtship behavior that occurs during the training period and is often referred to as immediate recall memory, measured 0-30 min after exposure to a mated female [6,8]. STM is measured between 30 min and 1 h after training, while LTM is most often measured 24 h after training [8] (**Figure 1B**). STM can be induced using a 1-h training period, but it only lasts for 2-3 h [6,8]. In most learning paradigms, LTM can only be induced by spaced bouts of repeated training. Mcbride *et al*. (1999) [8] showed that three spaced, 1-h training sessions were sufficient to induce courtship memory lasting for up to 9 days, in contrast to the 2-3 h induced by a single 1-h training session. McBride *et al*. [8] also demonstrated that a single 5-h training session produced a similar LTM response for up to 9 days. Flies do not court constantly during this 5-h period, in effect producing their own spaced training to induce LTM in a single training session. This is very important from a practical perspective, vastly increasing the ease with which this assay can be used to investigate LTM. Current protocols predominantly use a single training session of 7-h for LTM [11,12]. Several studies have investigated different mutant conditions that have specific defects in different aspects

of courtship learning. For example, mushroom body ablation affects STM and LTM, but not learning [8]. Mutations in the *amnesiac* gene, which was first defined as a specific regulator of memory using olfactory conditioning [3], affect STM but not learning [6]. Disruption of the translation regulator orb2 (oo18 RNA-binding (orb) CPEB2 subfamily) and ecdysone signaling exclusively effect LTM [9,13]. Thus, courtship conditioning is a useful paradigm to dissect the mechanisms underlying the different stages of learning and memory.

This work demonstrates an optimized experimental setup that allows for the relatively high-throughput testing of courtship conditioning. Furthermore, it describes a statistical analysis script and discusses critical factors of the assay. It is shown here that the *Drosophila* gene *Dihydroxyacetone phosphate acyltransferase* (*Dhap-at*) is required in neurons for LTM, but not for STM. The human ortholog of this gene, *glyceronephosphate O-acyltransferase* (*GNPAT*), is mutated in rhizomelic chondrodysplasia punctata type 2 [14], an autosomal-recessive disorder characterized by severe intellectual disability, seizures, and several other clinical characteristics [15]. In this context, courtship conditioning can be used to functionally validate the role of human disease genes in learning and memory, providing a basis for mechanistic studies.

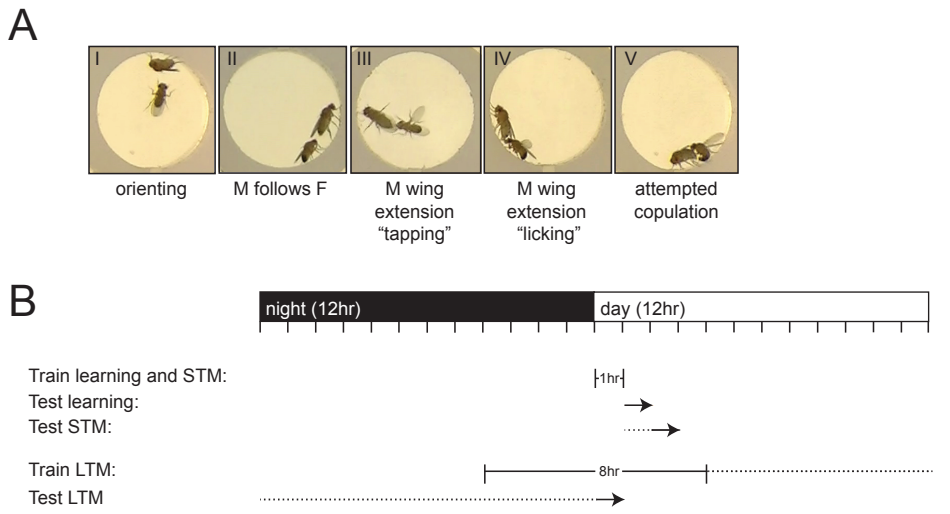


Figure 1: Determination of the courtship index and experimental overview.

(A) Images showing stereotypical male courtship behavior towards a female fly. Different stages of courtship behavior are shown: orientation (I), following (II), wing vibration (extension) and tapping (III), licking (IV), and attempted copulation (V). (B) Schematic overview of training and testing times relative to the incubator light cycle, marked in hours. Training times are indicated with bars, resting periods for STM and LTM are indicated with a dashed line, and the testing start point is indicated as an arrow. Note that the testing time for LTM is the day after after training.

Protocol

NOTE: In the protocol outlined below, one replicate of collection, training, and testing is described. In order to test the reproducibility of the results, these steps should be repeated in parallel, on multiple days, and with separate groups of flies (**Table 1**). The protocol is based on a 10-day life cycle from egg to adult, which is normal when rearing flies under constant conditions of 25 °C, 70% humidity, and a 12-h light/dark cycle. All aspects of this protocol assume that the conditions are kept constant throughout the entire assay. Times are indicated as hours before lights turn on (BLO) or after lights turn on (ALO) in the incubator, as this can be conveniently set depending on the researcher's preferred time of day. Use CO_2 gas only for the initial collection of naïve male flies and for the collection of premated females. This protocol for courtship conditioning is composed of the following steps:

1. Establishment of premated female collection cultures
2. Establishment of cultures for the collection of male test subjects
3. Preparation of housing blocks
4. Establishment of mating vials for the production of standardized premated females
5. Collection of male test subjects
6. Training
7. Testing
8. Video data analysis and statistics

Table 1: Example timeline for testing LTM over three replicates on individual days

	General	Collect	Train	Test
day -11	Start premated female collection cultures (step 1.3)			
day -10	Start cultures for the collection of male test subjects (step 2.2)			
day 1		rep. 1		
day 2		rep. 2		
day 3		rep. 3		
day 4		rep. 4	rep. 1	
day 5			rep. 2	rep. 1
day 6			rep. 3	rep. 2
day 7			rep. 4	rep. 3
day 8				rep. 4
day 9	Video data analysis and statistics (step 8)			

rep = repeat

1. Establishment of premated female collection cultures

1.1. Prepare powerfood [16]. Boil 0.8% (w/v) agar, 8% (w/v) yeast, 2% (w/v) yeast extract, 2% (w/v) peptone, 3% (w/v) sucrose, 6% (w/v) glucose, 0.05% (w/v) $MgSO_4$, and 0.05% (w/v) $CaCl_2$ in water for 15 min. Allow the solution to cool to 70 °C before adding 0.05% (w/v) methylparabene (CAUTION: toxic) and 0.5% (v/v) propionic acid (CAUTION: toxic). Mix well by stirring while cooling further to 50 °C to obtain a homogeneous solution.

1.2. Before the food solidifies at room temperature, add ~50 mL of powerfood to each 175-mL plastic vial. Allow the food to cool further. Close the vial with a plug.
NOTE: Powerfood is a specialized food mixture formulated specifically for the production of large numbers of flies, presumably by inducing egg laying. Powerfood is not used to produce male flies that will be used for behavior analysis (step 2) because the atypical diet and potential crowding might influence development.

1.3. On day -11 (**Table 1**), start 5-20 wildtype cultures with approximately 60-100 flies (a mix of males and females) in powerfood vials; these will be used in step 4 to produce standardized premated females [17]. Add a filter paper to each vial to increase the area upon which the larvae can pupate; this will increase the number of flies that can eclose.

1.4. Periodically repeat steps 1.1-1.3 throughout the experiment to obtain sufficient newly eclosing flies as input for the "establishment of mating vials for the production of standardized premated females" (step 4).

2. Establishment of cultures for the collection of male test subjects

2.1. Prepare normal food [16], made with 0.5% (w/v) agar, 2.75% (w/v) yeast, 5.2% (w/v) corn flour, 11% (w/v) sugar, 0.05% (w/v) methylparabene, and 0.5% (v/v) propionic acid in water, as described in steps 1.1 and 1.2. Close the 175-mL plastic vials with a fly vial plug.

2.2. On day -10 (**Table 1**), place about 10-20 males with approximately 30-75 virgin females (**Table of Specific Materials/Equipment**) in each 175-mL vial containing normal food. Add a filter paper to increase the surface area for pupation and to maximize productivity.

2.3. Establish three to six 175-mL vials per genotype to obtain the required number of test subject males.
NOTE: More vials may be needed, depending on the productivity of the desired genotype.

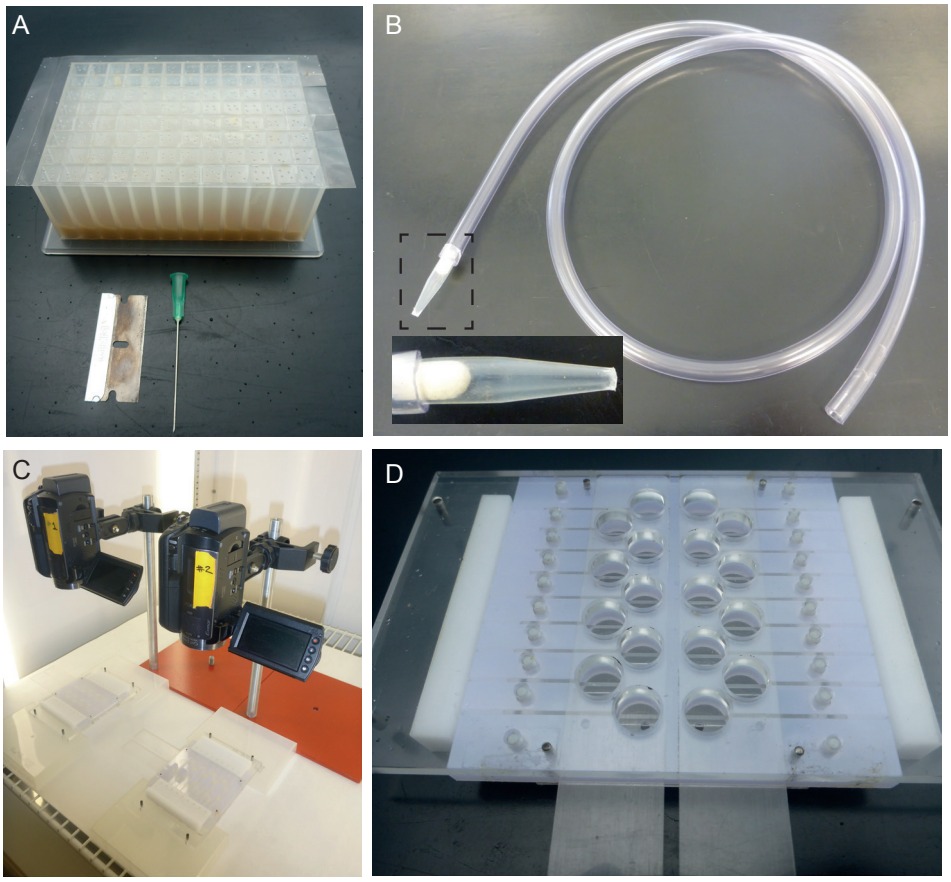


Figure 2: Equipment used for the *Drosophila* courtship conditioning assay.

(A) The housing block is a flat, bottom block with 500 µL of powerfood per well. It is sealed with a qPCR adhesive film with at least 4 holes per well that were created using a syringe needle with a 0.8-mm diameter. Individual rows are cut lengthwise into strips using a razor blade to allow opening and closing. (B) The aspirator is required for the gentle transfer of male and female flies without the use of anesthesia. The inset shows the tip of the aspirator, closed with a piece of cotton to keep the flies within the tip. (C) Setup of a two-camera system for the simultaneous recording of two courtship chambers. (D) A courtship chamber with 18 arenas. Sliding entry holes are used to place the flies in the arenas. The white dividers can be simultaneously opened to initiate interaction between the males and females.

3. Preparation of housing blocks (Figure 2A)

3.1. Melt approximately 50 mL of powerfood per housing block in a microwave, or prepare it fresh.

3.2. Add 500 µL of powerfood to each well of a 96-well, flat-bottom block using a multidispenser pipette.

3.3. Allow the food to solidify at room temperature.

3.4. Cover the blocks with PCR adhesive film and use a needle to make a least 4 holes per well to provide fresh air to the flies.

3.5. In order to be able to open each well, use a razor blade to cut the adhesive film lengthwise between each row. Leave the film intact on one end of the block.

3.6. The blocks can be stored at 4 °C for up to 2 d.
NOTE: Allow the blocks to re-equilibrate to room temperature prior to use.

4. Establishment of mating vials for the production of standardized premated females

4.1. On day -1, remove and discard all adult wildtype flies from premated female collection cultures at 2-5 h BLO.

4.2. Collect flies using the aspirator (**Figure 2B**) from these vials in 2- to 3-h intervals (*e.g.,* at 30 min, 2.5 h, and 5 h ALO) and place them in a new powerfood vial supplemented with a small amount of yeast paste and filter paper.

4.3. To avoid crowding and to promote an optimal mating atmosphere, do not transfer more than 150-200 flies to each new vial. Ensure the mating of all females by providing at least 25% males. Ensure that sufficient females are present in the mating vials to accommodate the size of the experiment.
NOTE: As this is a crucial step in the protocol, make sure that only freshly eclosed flies and no old flies, larvae, or pupae are transferred to the new mating vial.

4.4. Incubate these "mating vials" for four days to allow sufficient time for all females to have mated.

5. Collection of male test subjects

5.1. On day 1 (**Table 1**) at 2-3 h BLO, use CO_2 to remove all adult flies from the male collection vials (step 2), but let more flies eclose over the next few hours.

5.2. Over the next 5-6 h, remove newly eclosed flies every 20-30 min using CO_2 and put each male in an individual well of the housing block (step 3) using the aspirator (**Figure 2B**).

5.3. Re-seal the well with the adhesive PCR film.
NOTE: This is a crucial step in the protocol. The males should be collected

frequently. Collected males should be isolated in the housing block close to the time of eclosion, when they demonstrate pale pigmentation and the presence of the meconium in the translucent abdomen.

NOTE: Gentle use of the aspirator allows the transfer of flies; however, inappropriate use will stress the flies, causing variance in the assay (see the Discussion).

5.4. Aim to collect up to 48 males per genotype. This provides a small excess to the maximal number of males needed for the analysis of both naïve and trained conditions, allowing for some loss during later transfer steps.

6. Training

6.1. Remove the flies from the mating vial (step 4.2) using CO_2 and separate the premated females from the males.

6.2. Using the aspirator, add a single anesthetized, premated female to each well in one row of a new housing block.

6.3. Using the aspirator and without anesthesia, transfer an individual naïve male from the housing block set up in step 5.2 to the well containing a premated female. After the male is placed into the well, re-seal immediately with the adhesive film; do not allow the male to escape.

NOTE: Transfer male flies from the aspirator to the housing block by taking advantage of their natural "negative geotaxis" behavior (see the Discussion).

6.4. Repeat steps 6.2-6.3 until enough male-female pairs are established. Ideally, establish 24 pairs, two full rows of a housing block, per genotype. Leave the remaining naïve males in the original housing block set up in step 5.2.

6.5. Leave the male-female pairs undisturbed during the training period (**Table 2, Figure 1B**).

NOTE: During this time, the male will court and be rejected by the premated female. For learning and STM, the training period is 1 h and for LTM, the training period is 7-9 h.

6.6. End the training (**Table 2, Figure 1B**) by using an aspirator to gently separate the male from the premated female; do not use anesthesia. Place the separated male in a new housing block.

6.7. Use the aspirator to transfer all naïve males gently and without anesthesia from the housing block set up in step 5.2 to a new housing block.

NOTE: This step is optional for STM and learning, but it is very important for LTM because the flies are housed for an additional 24 h to test LTM.

6.8. For STM and LTM, allow the males to rest for 1 h and ~24 h, respectively (**Table 2, Figure 1B**) before testing (step 7).

6.9. For learning, immediately test the trained and naïve males (step 7).

Table 2: Training duration, training times, and testing times for learning, STM, and LTM

	Learning	STM	LTM
Training time	1 h.	1 h.	8 h.
Resting time	0 h.	1 h.	~ 24 h.
start training	0 h. ALO	0 h. ALO	4 h. BLO
stop training	1 h. ALO	1 h. ALO	4 h. ALO
start test	1 h. ALO	2 h. ALO	0 h. ALO (next day)

ALO = after lights turn on, BLO = before lights turn on, STM = short term memory, LTM = long term memory

7. Testing

7.1. Collect flies from mating vials (step 4.2) using CO_2 and separate the premated females from the males.

7.2. Let the females recover from the anesthesia for at least 1 h in a vial containing normal food.

7.3. Mount the video recorders in advance (**Figure 2C**), in order to have all equipment ready before the testing starts.

7.4. Start the testing according to the different timelines for learning, STM, and LTM (**Table 2, Figure 1B**). Perform testing immediately after training for learning, 1 h after training for STM, and 24 h after training for LTM.

7.5. Using the aspirator, gently transfer an individual male from the resting housing block or from the training housing block if learning is being tested (step 6.7, trained; step 6.8, naïve) to one half of a courtship arena with the divider closed (**Figure 2D; see File S1 for a building plan**).
NOTE: The use of the natural "negative geotaxis" behavior should be sufficient to transfer the male flies from the aspirator to the courtship arena (Discussion).

7.6. Quickly but gently move the entry hole to the next arena and repeat step 7.5 until all 18 arenas contain one male.

7.7. Using the aspirator and without CO_2, add one premated female (collected in step 7.2) to the other half of all 18 arenas.

7.8. Carefully place the courtship chamber under the camera, with the opening of the wells facing down (**Figure 2C**).

7.9. Remove the divider of the arenas to allow direct interaction between the males and premated females.

7.10. Immediately start recording the behavior for at least 10 min.
NOTE: When using a two-camera setup, the parallel recording of two courtship plates can be done in overlapping timeframes to maximize efficiency.

7.11. Empty the courtship arenas using a hand-held vacuum cleaner and allow the courtship chamber to ventilate before re-use.

7.12. Repeat step 7.4-7.11 until the testing of all genotypes and conditions (*i.e.,* naïve and trained) have been completed.

8. Video data analysis and statistics

8.1. Calculate the courtship index (CI), defined as the percentage of time that the male courts during the first 10 min of the testing period, for each individual male fly.

NOTE: This can be done manually by observing stereotypical courtship behavior (**Figure 1A**) or by using computer software for automated quantification of courtship behavior (Discussion).

NOTE: It is recommended to analyze 40-60 males per condition over the course of three days in order to achieve sufficient statistical power and to judge the consistency of the CI data.

8.2. Calculate the learning index (LI), defined as the percent reduction in the mean CI of trained males compared to naïve males ($LI = (CI_{naive} - CI_{trained}) / CI_{naive}$). Evaluate the LI for each day of testing and compare it to the cumulative LI calculated from all testing days combined.

8.3. Make a separate two-column tab data file with "Genotype" and "CI" as the headers.

NOTE: These headings are case sensitive. The name of the genotype for each CI must consist of a description of the genotype followed by an underscore and the training condition (*e.g.,* genotype_N and genotype_LTM, etc., where N = naïve and LTM = long term memory; see Supplemental File S2 for an example). This annotation is essential, as the function analearn will identify trained and naïve flies based on the characters present after the first underscore in the "Genotype" column.

8.4. Use the analearn R script (**Supplemental File S3**) to perform a randomization test to judge the statistical significance of differences between the LI values from different genotypes.

8.4.1 Source the script (**Supplemental File S3**) into R [18], which defines a function called "analearn."

NOTE: The function definition is: analearn <- function(nboot = 10000, naivelevel = 'N', refmutation = NA, datname = NA, header = TRUE, seed = NA, writeoutput = TRUE).

8.4.2 Start the function by entering "analearn()" in the R command line and selecting the data file to be analyzed (produced in step 8.3) via the pop-up window.

8.4.3 Choose the reference mutation, which is the control genotype, by entering the

corresponding number and pressing enter.
NOTE: After selecting the reference genotype, the script takes several seconds to perform 10,000 bootstrap replicates.

8.4.4 Observe the output table (**Table 3**), which contains the genotype, learning condition (*i.e.,* learning, STM, or LTM), mean CI naïve, mean CI trained, LI, the difference between the LI of the control compared to the experimental condition (LI dif), the lower limit (LL) and upper limit (UL) of the 95% confidence interval of the LI dif, and the *p*-value indicating the probability that there is no significant difference.
NOTE: analearn will store an output text file in the directory where the data file is located. However, the output table also appears in the R-Studio console. The default name is constructed based on the name of the data file provided.

8.4.5 There are several arguments in the analearn function that can be used to alter the default settings of the function to adjust the parameters of the bootstrapping.
NOTE: "nboot" defines the number of bootstrap replicates and is set to 10,000 by default. This value can be changed into any integer number larger than zero. **Table 4** enlists several arguments that can be used to alter the default settings of the function. However, it is not recommended to use data that is produced with a low number of bootstrap replicates.

Table 3: Statistical data produced from the analearn script

Genotype	Learning condition	CI naive	CI trained	LI	LI difference	Lower limit (95% conficence interval)	Upper limit (95% conficence interval)	*p*-value
Control	STM	0.467	0.116	0.752	NA	NA	NA	NA
Dhap-at-RNAi	STM	0.699	0.257	0.633	0.119	-0.030	0.265	0.116
Control	LTM	0.590	0.384	0.348	NA	NA	NA	NA
Dhap-at-RNAi	LTM	0.697	0.650	0.068	0.280	0.103	0.446	0.003

Statistical data produced from the analearn script. The output file of the bootstrapping R-script containing the genotype, learning condition *(i.e.*, learning, STM, or LTM), mean CI naïve, mean CI trained, LI, difference between LI of the control compared to experimental condition (LI dif), the lower limit (LL) and upper limit (UL) of the 95% confidence interval of LI dif, and the *p*-value indicating the probability that there is no significant difference.

Table 4: Arguments used in the analearn function that can alter the default settings of the function to adjust the parameters of the bootstrapping

"naivelevel" determines the text that will identify naïve values for each genotype. The default is "N," but this can be changed into any other alphanumerical text.
"refmutation" is set to "NA" (not applicable) by default, but can be changed to the name of the control or the genotype in order to perform statistical comparisons. This will cause the script to automatically select the control genotype.
"datname" refers to the name of the data file and can be specified in this argument instead of the default file selection.
"header" can be used to indicate whether or not the data file contains column headers. The default is "TRUE," but a file with no headers can be used when this argument is changed to "FALSE."
"seed" initializes the random number generator. This is set by default to "NA" and ensures a random number each time the script is used. By design, a bootstrap analysis will give slightly different results each time it is run, even when using the same data file. When the seed is specified by any integer number larger than zero, the same set of random bootstrap samples is obtained.
"writeoutput" can be set to "TRUE" or "FALSE" in order to determine whether an output file will be generated. The default is "TRUE."

Results

The courtship conditioning assay can be used to measure learning and memory in *Drosophila*. In order to demonstrate this, the results presented here analyze STM and LTM in control flies compared to flies with the neuron-specific knockdown of Dhap-at. Control males express an RNA interference hairpin sequence targeting a *Caenorhabditis elegans*-specific gene, *putative zinc finger protein C02F5.12* [19]. This control strain ensures an equal number of upstream activating sequence (UAS) elements between the knockdown and control in the same genetic background, and the control RNAi hairpin accounts for any non-specific effects of the RNAi system. Control males (genotype: *UAS-dcr2/+; P{KK108109}VIE-260B/+; elav-Gal4/+*) show a significant reduction in CI in trained versus naïve for both STM (**Figure 3A, $p = 1.2x10^{-4}$, Mann-Whitney U-test**) and LTM (**Figure 3B, $p = 1.2 \times 10^{-4}$, Mann-Whitney U-test**). This result reflects the normal capacity for learning and memory in these flies. *Dhap-at* knockdown flies (genotype: *UAS-dcr2/+; P{KK101437}VIE-260B/+; elav-Gal4/+*) also show a significant reduction of CI in trained versus naïve flies for STM (**Figure 3A, $p = 1.2 \times 10^{-4}$, Mann-Whitney U-test**). However, they do not show a significant reduction in CI after LTM training (**Figure 3B, $p = 0.33$, Mann-Whitney U-test**). The Mann-Whitney U-test was used to compare the mean CI of naïve and trained flies due to the non-parametric distribution of the CI data for some conditions (**Figure 3A and 3B**). The differences observed through the analysis of CIs are reflected by the LI for each genotype, which measures the percent reduction in CI in naïve versus trained flies (Figure 3C). There is

no significant difference in LI between controls and Dhap-at knockdown flies for STM (p = 0.115, 10,000 bootstrap replicates, **Figure 3C, Table 3**), whereas the LI for LTM is significantly lower in *Dhap-at* knockdown flies (p = 0.0034, 10,000 bootstrap replicates, **Figure 3C**, **Table 3**). For the comparison of LI between genotypes, a randomization test [20] adapted from a method first recommended by Kamyshev *et al.* [7] was used. Since the LI is a single value derived from population data (*i.e.,* mean CI-naïve and mean CI-trained), standard statistical methods that rely on experimentally derived distributions do not apply. The randomization test is distribution-free and uses bootstrapping (*i.e.,* random sampling with replacement) to generate hypothetical data sets that are derived from the actual data. The analearn script (**File S3**) generates a set of hypothetical LIs for each genotype and calculates the difference between the control and the test genotype (LIdiff). This process is repeated 10,000 times, and the resulting values are used to determine the 95% confidence interval of LIdiff (**Table 3**). This data is used to generate a p-value indicating the probability that the LI of the two genotypes is different. The results shown here demonstrate that Dhap-at is required in neurons for LTM but not for STM.

In order to control for day-to-day variability, CIs and LIs are compared between replicate days **(Table 5**). Although some fluctuation in LI is observed from day to day, the results are generally reproducible. It is important to note that CI data can vary greatly depending on the control strain used and the environmental conditions of testing. The CI data shown here is typical for this control genotype, but other genotypes may exhibit a higher or lower mean CI and distribution.

Table 5: CI and LI values obtained on separate testing days

		Control			***Dhap*-at-RNAi**		
		Average CI naïve	**Average CI trained**	**LI**	**Average CI naïve**	**Average CI trained**	**LI**
STM	Day 1	0.300	0.125	0.584	0.679	0.239	0.648
	Day 2	0.634	0.107	0.831	0.720	0.276	0.617
	All Days	0.467	0.116	0.752	0.699	0.257	0.633
LTM	Day 1	0.590	0.441	0.252	0.630	0.646	-0.027
	Day 2	0.640	0.363	0.432	0.709	0.710	-0.002
	Day 3	0.547	0.349	0.363	0.738	0.598	0.190
	All Days	0.590	0.384	0.348	0.697	0.650	0.068

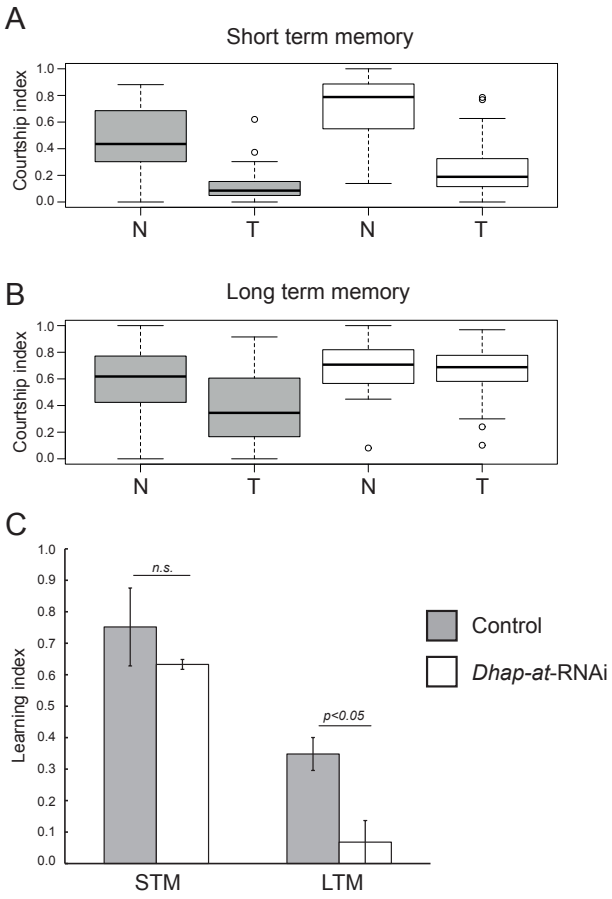


Figure 3: Analysis of STM and LTM in control and Dhap-at knockdown flies.

(A-B) Boxplots showing the distribution of CI values for naïve (N) and trained (T) flies of the control (gray) and *Dhap-at* knockdown (white) genotypes tested for STM (A) and LTM (B). (C) Corresponding LI values for control and *Dhap-at* knockdown flies tested for STM and LTM. Differences in LI between control and knockdown genotypes were compared using a randomization test (10,000 bootstrap replicates). Error bars indicate the standard error of the mean derived from the LI values calculated on different test days. The genotypes are: *w+*, *UAS-dcr2/+*; *P{KK108109}VIE-260B/+*; and *elav-Gal4/+* (Control) and *w+*, *UAS-dcr2/+*; *P{KK101437}VIE-260B/+*; and *elav-Gal4/+* (*Dhap-at*-RNAi).

Discussion

The courtship conditioning assay is a classic paradigm for the analysis of learning and memory in *Drosophila*. The protocol presented here follows the general methodology described previously [6-9] but includes unique aspects such as practical guidelines, specialized equipment, and a data analysis script [9,12] for randomization tests. Using this

protocol, it is possible to analyze large numbers of flies in parallel using 96-well flat-bottom blocks (**Figure 2A**) to collect and train males. The blocks are sealed with PCR adhesive film, which makes the flies easily accessible when required. Additionally, the unique courtship chambers described here allow for the simultaneous pairing of 18 male-female pairs in a nearly two-dimensional space that is optimal for video analysis. The custom-designed courtship chambers are easy to use, and a building plan is provided (**File S1, Figure 2D**). This protocol, from the establishment of the cultures used to collect test subjects to the acquisition of video data, takes approximately 20 days (**Table 1**). Additional time is required for the analysis of video data. In our experience, the STM assay is extremely robust. The LTM assay is also quite robust, but it is more sensitive to confounding environmental variables and therefore can be more difficult to master.

Animal behavior can be quite variable. Therefore, critical steps in the protocol must be performed with care to reduce this variance. First, gentle use of the aspirator (**Figure 2B**) will reduce the stress that can be imposed by rough handling or by blowing out too strongly. A suggested method of transferring individual flies out of the aspirator is by using negative geotaxis. As flies tend to walk up, one can simply point the tip of the aspirator up; just before the fly reaches the tip, a gentle blow is sufficient to let the fly out. Additionally, to let the males into the courtship chambers before testing, a blow is often not necessary.

Another important step is the collection and generation of male test subjects. All males must be collected when they are very young and socially naïve. This can be achieved by frequent collection during the peak periods of eclosion (step 5.2). If males are not collected in this tight timeframe, they can have early social interactions, which may result in poor learning or high variability in CI. Another factor of male test subjects that should be assessed is the genetic background. Different genetic backgrounds will exhibit different levels of naïve courtship and may also differ in general activity or locomotor ability. When comparing multiple genotypes, care should be taken with regard to genetic background in order to avoid these confounding factors that may influence LI scores. Additionally, the distribution of CI data should be carefully assessed. CI data can be both parametric and non-parametric, depending on the genotype or other environmental factors. In some cases, if the distribution of CI is dramatically skewed away from a normal distribution, it may be better to use the median CI rather than the mean for the calculation of LIs. However, in our experience, the use of median or mean CI does not make a difference in the statistical interpretation of the data, and the use of the mean CI is the common practice in the literature.

For successful courtship conditioning, the active rejection of male courtship attempts by premated females is crucial during the training period. It is important to ensure that the premated females used in this assay have been efficiently mated and are thus not allowing

copulation. This premating is established in the mating vials prepared in step 4, where male and females flies are housed together for 4 days (**Table 1**). Subsequently, mating can be monitored by regular examination of testing videos and by observing male-female pairs during training. If mating does occur, there are several measures that can be taken during the preparation of premated females. First, premated females should be reared under optimal breeding conditions. Vials can be supplemented with yeast paste and a folded filter paper to increase potential mating surfaces. The incubation of flies under the conditions described here has produced robust premated females in the past, but this may vary in different labs and with the use of different genetic strains. Therefore, it may be necessary to optimize the generation of premated females by varying the incubation time and conditions.

Quantification of courtship behavior is another critical step in this protocol. This can be done manually or automatically using specialized software programs [9]. Automated quantification is fast and, in principle, unbiased. Several programs have been published[21-23]; however, they are not straightforward to use, often requiring specialized video formats and advanced computational skills. Manual quantification is easy and accurate, but it is highly labor intensive and subject to individual variability and bias. It is important to emphasize that this protocol does not address the requirements for video formatting that are potentially required for the automated quantification of CIs. For manual quantification, use any simple video recording device that has the potential to produce a video of sufficient quality to accurately observe courtship behavior. For automated quantification, there will likely be different requirements depending on the software used, and users should investigate this thoroughly if automated quantification is desired.

In combination with the extensive tools that are available for the genetic manipulation of flies, the courtship conditioning assay provides a robust readout that can be used to dissect molecular mechanisms and neuronal networks involved in learning and memory.

Acknowledgements

We acknowledge the Vienna *Drosophila* Resource Center for providing the *Drosophila* strains. Additionally, stocks obtained from the Bloomington *Drosophila* Stock Center (NIH P40OD018537) were used in this study. This research was supported in part by the European Union's FP7 large-scale integrated network Gencodys to K.K., H.v.B., and A.S. and by NSERC Discovery and CIHR Project Grants to J.M.K.

Disclosure

The authors have nothing to disclose.

References:

1 Quinn, W. G., Harris, W. A. & Benzer, S. Conditioned behavior in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 71, 708-712 (1974).

2 Livingstone, M. S., Sziber, P. P. & Quinn, W. G. Loss of calcium/calmodulin responsiveness in adenylate cyclase of rutabaga, a *Drosophila* learning mutant. Cell 37, 205-215 (1984).

3 Quinn, W. G., Sziber, P. P. & Booker, R. The *Drosophila* memory mutant amnesiac. Nature 277, 212-214 (1979).

4 Dudai, Y., Jan, Y. N., Byers, D., Quinn, W. G. & Benzer, S. dunce, a mutant of *Drosophila* deficient in learning. Proc. Natl. Acad. Sci. USA 73, 1684-1688 (1976).

5 Dubnau, J. & Tully, T. Gene discovery in *Drosophila*: new insights for learning and memory. Annu. Rev. Neurosci. 21, 407-444, doi:10.1146/annurev.neuro.21.1.407 (1998).

6 Siegel, R. W. & Hall, J. C. Conditioned responses in courtship behavior of normal and mutant *Drosophila*. Proc. Natl. Acad. Sci. USA 76, 3430-3434 (1979).

7 Kamyshev, N. G., Iliadi, K. G. & Bragina, J. V. *Drosophila* conditioned courtship: two ways of testing memory. Learning & memory 6, 1-20 (1999).

8 McBride, S. M. et al. Mushroom body ablation impairs short-term memory and long-term memory of courtship conditioning in *Drosophila melanogaster*. Neuron 24, 967-977 (1999).

9 Keleman, K., Kruttner, S., Alenius, M. & Dickson, B. J. Function of the *Drosophila* CPEB protein Orb2 in long-term courtship memory. Nature Neurosci 10, 1587-1593, doi:10.1038/nn1996 (2007).

10 Hall, J. C. The mating of a fly. Science 264, 1702-1714 (1994).

11 Keleman, K. et al. Dopamine neurons modulate pheromone responses in *Drosophila* courtship learning. Nature 489, 145-149, doi:10.1038/nature11345 (2012).

12 Kramer, J. M. et al. Epigenetic regulation of learning and memory by *Drosophila* EHMT/G9a. PLoS Biol. 9, e1000569, doi:10.1371/journal.pbio.1000569 (2011).

13 Ishimoto, H., Sakai, T. & Kitamoto, T. Ecdysone signaling regulates the formation of long-term courtship memory in adult *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 106, 6381-6386, doi:10.1073/pnas.0810213106 (2009).

14 Kochinke, K. et al. Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. Am. J. Hum. Genet. 98, 149-164, doi:10.1016/j.ajhg.2015.11.024 (2016).

15 Buchert, R. et al. A peroxisomal disorder of severe intellectual disability, epilepsy, and cataracts due to fatty acyl-CoA reductase 1 deficiency. Am. J. Hum. Genet. 95, 602-610, doi:10.1016/j.ajhg.2014.10.003 (2014).

16 JoVE Science Education Database.*Essentials of Biology 1: yeast, Drosophila and C. elegans. Drosophila* Maintenance. JoVE, Cambridge, MA, doi: 10.3791/5084 (2017).

17 JoVE Science Education Database. *Essentials of Biology 1: yeast, Drosophila and C. elegans.* An Introduction to *Drosophila melanogaster*. JoVE, Cambridge, MA, doi: 10.3791/5082 (2017).

18 R Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna (Austria) (2016).

19 Stepien, B. K. et al. RNA-binding profiles of *Drosophila* CPEB proteins Orb and Orb2. Proc. Natl. Acad. Sci. USA, 113, E7030-E7038, doi:10.1073/pnas.1603715113 (2016).

20 Basu, D. Randomization Analysis of Experimental Data: The Fisher Randomization Test. J Am Stat Assoc 75, 575-582, doi:10.2307/2287648 (1980)

21 Dankert, H., Wang, L., Hoopfer, E. D., Anderson, D. J. & Perona, P. Automated monitoring and analysis of social behavior in *Drosophila*. Nature methods 6, 297-303, doi:10.1038/nmeth.1310 (2009).
22 Reza, M. A. et al. Automated analysis of courtship suppression learning and memory in *Drosophila melanogaster*. Fly 7, 105-111, doi:10.4161/fly.24110 (2013).
23 Schneider, J. & Levine, J. D. Automated identification of social interaction criteria in *Drosophila melanogaster*. Biol. Lett. 10, 20140749, doi:10.1098/rsbl.2014.0749 (2014).

Supplemental File S1: Building plan of a courtship chamber. The file can be opened with any application that allows .stp extensions (CAD-files).

Supplemental File S2: Example of an input file for the analearn script.

```
Genotype        CI
CG4625_N        0.819
CG4625_N        0.657
CG4625_N        0.805
CG4625_N        0.481
CG4625_N        0.696
CG4625_LTM      0.451
CG4625_LTM      0.548
CG4625_LTM      0.542
CG4625_LTM      0.947
CG4625_LTM      0.448
etc.
```

Supplemental File S3: The analearn.R script. The file can be opened with R-studio [18].

```
analearn <- function(nboot = 10000, naivelevel = 'N', refmutation = NA,
datname = NA, header = TRUE, seed = NA, writeoutput = TRUE) {
 if (!is.na(seed)) set.seed(seed)
 if (is.na(datname)) datname <- file.choose()
 datfile <- read.table(datname,header = header,stringsAsFactors = FALSE)
 idstr <- strsplit(datfile[,1],'_')
 datfile$mutation <- as.factor(unlist(lapply(idstr,'[[',1)))
 if (is.na(refmutation) || !(refmutation %in% levels(datfile$mutation))){
  cat('The following mutations are observed:\n')
  for(i in seq_along(levels(datfile$mutation))) {
   cat(i,': ',levels(datfile$mutation)[i],'\n', sep = "")
  }
```

```
  cat('Please indicate the reference mutation by giving a number from 1 to
',length(levels(datfile$mutation)),' followed by <Enter>.\n',sep = "")
  refnum <- readline()
  refmutation <- levels(datfile$mutation)[as.numeric(refnum)]}
 datfile$mutation <- relevel(datfile$mutation,refmutation)
 datfile$learncon <- as.factor(unlist(lapply(idstr,'[[',2)))
 datfile$learncon <- relevel(datfile$learncon,naivelevel)
 aggrdat <- aggregate(CI ~ mutation + learncon, data = datfile, FUN = mean)
  LIout <- data.frame(mutation = "", learningcondition = "", CInaive =
0, CItrained = 0, LI = 0, LIdif = 0,LL95CI = 0, UL95CI = 0 ,p = 0,
stringsAsFactors = FALSE)
 k = 0
 for (i in levels(aggrdat$learncon)[-1]) {
  k <- k + 1
  LIout[k,"mutation"] <- levels(aggrdat$mutation)[1]
  LIout[k,"learningcondition"] <- i
   LIout[k,"CInaive"] <- aggrdat[aggrdat$mutation == LIout[k,"mutation"] &
aggrdat$learncon == naivelevel,"CI"]
  LIout[k,"CItrained"] <- aggrdat[aggrdat$mutation == LIout[k,"mutation"] &
aggrdat$learncon == i,"CI"]
      LIout[k,"LI"]   <-   (LIout[k,"CInaive"]   -   LIout[k,"CItrained"])/
LIout[k,"CInaive"]
  LIout[k,"LIdif"] <- NA
  LIout[k,"LL95CI"] <- NA
  LIout[k,"UL95CI"] <- NA
  LIout[k,"p"]<- NA
  for(j in levels(aggrdat$mutation)[-1]) {
   k <- k + 1
   LIout[k,"mutation"] <- j
   LIout[k,"learningcondition"] <- i
   LIout[k,"CInaive"] <- aggrdat[aggrdat$mutation == j & aggrdat$learncon ==
naivelevel,"CI"]
   LIout[k,"CItrained"] <- aggrdat[aggrdat$mutation == j & aggrdat$learncon
== i,"CI"]
       LIout[k,"LI"]   <-   (LIout[k,"CInaive"]   -   LIout[k,"CItrained"])/
LIout[k,"CInaive"]
    LIout[k,"LIdif"] <- LIout[LIout$mutation == levels(aggrdat$mutation)[1]
& LIout$learningcondition == i,"LI"] - LIout[k,"LI"]
```

```
   LIout[k,"LL95CI"] <- NA
   LIout[k,"UL95CI"] <- NA
   LIout[k,"p"]<- NA

  }
 }
 datsplit <- with(datfile, split(datfile, list(mutation,learncon)))
 bootres = matrix(0, nrow = nboot, ncol = (length(levels(aggrdat$mutation))-
1)*(length(levels(aggrdat$learncon))-1))
 CIsplit = list()
 for (i in seq_len(nboot)) {
  for (j in names(datsplit)) {
   datstuk <- datsplit[[j]]
   CIsplit[j]<- mean(datstuk[sample(NROW(datstuk),replace = TRUE),"CI"])
  }
  CInames <- matrix(unlist(strsplit(names(CIsplit),'\\.')),ncol = 2, byrow
= TRUE)
   aggrtemp <- data.frame(mutation = as.factor(CInames[,1]), learncon =
as.factor(CInames[,2]),CI = unlist(CIsplit))
   aggrtemp$mutation <- relevel(aggrtemp$mutation,levels(datfile$mutation)
[1])
  aggrtemp$learncon <- relevel(aggrtemp$learncon,naivelevel)
  for (l in levels(aggrtemp$learncon)[-1]) {
   for (m in levels(aggrtemp$mutation)){
        aggrtemp[aggrtemp$mutation==m & aggrtemp$learncon == l,'LI'] <-
(aggrtemp[aggrtemp$mutation==m & aggrtemp$learncon == naivelevel,'CI']
- aggrtemp[aggrtemp$mutation==m & aggrtemp$learncon == l,'CI'])/
aggrtemp[aggrtemp$mutation==m & aggrtemp$learncon == naivelevel,'CI']
   }
  }
  k <- 0
  for (l in levels(aggrtemp$learncon)[-1]) {
   for (m in levels(aggrtemp$mutation)[-1]){
    k <- k + 1
     bootres[i,k] <- aggrtemp[aggrtemp$mutation==levels(aggrtemp$mutation)
[1] & aggrtemp$learncon == l,'LI'] - aggrtemp[aggrtemp$mutation==m &
aggrtemp$learncon == l,'LI']
   }
```

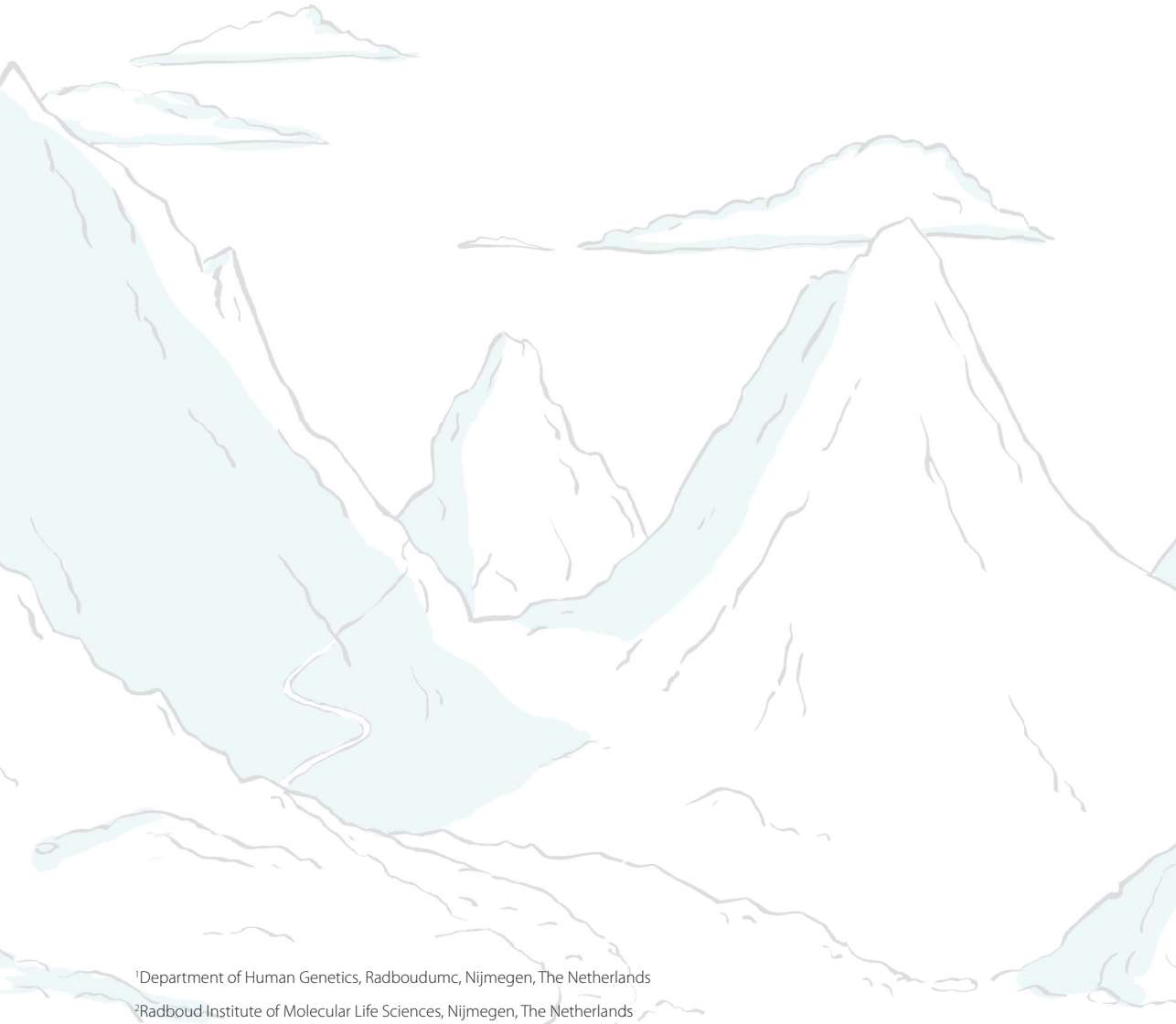
```
  }

 }
 k <- 0
 for (i in levels(aggrtemp$learncon)[-1]) {
  for(j in levels(aggrtemp$mutation)[-1]) {
   k <- k + 1
    LIout[LIout$mutation == j & LIout$learningcondition == i,"LL95CI"] <-
quantile(bootres[,k],probs = 0.025)
    LIout[LIout$mutation == j & LIout$learningcondition == i,"UL95CI"] <-
quantile(bootres[,k],probs = 0.975)
   p <- mean(bootres[,k]>0)
   if(p>0.5) p <- 1 - p
   if (p < 1) p <- p*2
   LIout[LIout$mutation == j & LIout$learningcondition == i,"p"] <- p

  }
 }
 if (writeoutput) {
  outname <- unlist(strsplit(datname,'\\.'))
   outname[length(outname)-1] <- paste(outname[length(outname)-1],'out',sep
= '')
  outname <- paste(outname, sep = '', collapse = '.')
  if (file.exists(outname)) {
   cat('Output file ', outname, ' already exists. Overwrite? <y/n>\n')
   if(!readline() == 'y') outname <- file.choose()
  }
  sink(outname)
  cat('Analyzing file ',datname,'\n')
  cat('Based on ',nboot,' bootstrap replications.\n\n')
  print(LIout)
  sink()
 }
 LIout
}
```

Table of specific materials

Name of Reagent/ Equipment	Company	Catalog Number	Comments/Description
General			
P{KK101437}VIE-260B	VDRC	101437	Gpat-RNAi in 60100 background
P{KK108109}VIE-260B	-	-	Control-RNAi in 60100 background (gift from K. Keleman)
w+, UAS-dcr2/yhh;;elav-Gal4 (III)	-	-	Panneuronal driver line
Containers for plant tissue culture	VWR	960177	175ml plastic vials
folded filters	Whatman	10311643	Filter paper to enlarge area flies can pupate on
Flat-bottom blocks (96-wells)	Qiagen	19579	Housing blocks
MicroAmp Clear Adhesive Film	Applied Biosystems	4306311	PCR adhesive film as lid on flat-bottom blocks
razor blade	-	-	Any sharp will do
needle (0.8 mm diameter)	-	-	Don't use too big diameter as flies can escape
aspirator	-	-	Cut a 1ml pipet tip with scissors in order to have two pieces. The narrow tip of the pipettip is placed as fly entrance in a ~80cm flexible hose. To prevent a fly from getting in the hose, a normal piece of cotton or small mesh gaze is placed in between the tip and the hose. The other half of the pipettip can be used as mouth piece at the end of the hose.
courtship chambers	-	-	File S1 can be opened with indicated CAD software
camcorder	Sony	-	>4M pixels, full HD

Table of specific materials *(continued)*

Name of Reagent/ Equipment	Company	Catalog Number	Comments/Description
Power food			
agar	Sigma	A7002	
yeast	Bruggeman	-	
yeast extract	MP biomedicals	0210330391	
peptone	Sigma	P6838	
sucrose	Sigma	S9378	
glucose	Sigma	G7021	
$MgSO_4$	Sigma	M2643	
$CaCl_2$	Merck	1023780500	
methylparabene (CAUTION)	Sigma	H5501	
propionic acid (CAUTION)	Sigma	P1386	
demineralized water	-	-	
yeast paste	-	-	Yeast grains and water mixture in a 1:1 ratio
Normal food			
agar	MP biomedicals	215017890	-
yeast	bruggeman	-	-
corn flour	de Molen	-	-
sugar	de Molen	-	-
methylparabene (CAUTION)	Sigma	H5501	-
propionic acid (CAUTION)	Sigma	P1386	-
demineralized water	-	-	-



[1]Department of Human Genetics, Radboudumc, Nijmegen, The Netherlands

[2]Radboud Institute of Molecular Life Sciences, Nijmegen, The Netherlands

[3]Donders Institute for Brain, Cognition, and Behaviour, Centre for Neuroscience, Nijmegen, The Netherlands

[4]Department of Biology, Western University, London, Ontario, Canada

[5]Children's Health Research Institute, London Ontario Canada

[6]Department of Clinical Genetics and School for Oncology & Developmental Biology (GROW), Maastricht University Medical Center, Maastricht, the Netherlands

[7]Department of Pediatrics, Máxima Medical Centre, Veldhoven, The Netherlands

[8]Clinical Genetics and Metabolism, Children's Hospital Colorado, Aurora

[9]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester

[10]Department of Molecular Developmental Biology, Radboud University, Nijmegen, The Netherlands

[11]Department of Physiology and Pharmacology, Western University, London Ontario, Canada

[12]These authors contributed equally to this work

Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder

Tom S. Koemans[1,2,3]
Tjitske Kleefstra[1,3]
Melissa C. Chubak[4]
Max H. Stone[4,5]
Margot R.F. Reijnders[1,3]
Sonja de Munnik[1,3]
Marjolein H. Willemsen[1,3]
Michaela Fenckova[1,3]
Connie T.R.M. Stumpel[6]
Levinus A. Bok[7]
Margarita Sifuentes Saenz[8]
Kyna A. Byerly[9]
Linda B. Baughn[9]
Alexander P.A. Stegmann[6]
Rolph Pfundt[1,3]
Huiqing Zhou[1,2,10]
Hans van Bokhoven[1,3,12]
Annette Schenck[1,3,12,]
Jamie M. Kramer[4,5,11,12]

PLoS Genetics, 2017 doi: 10.1371/journal.pgen.1006864

3

Abstract

Kleefstra syndrome (KS), caused by haploinsufficiency of euchromatin histone methyltransferase 1 (EHMT1), is characterized by intellectual disability (ID), autism spectrum disorder (ASD), characteristic facial dysmorphisms, and other variable clinical features. In addition to EHMT1 mutations, *de novo* variants were reported in four additional genes (MBD5, SMARCB1, NR1I3, and KMT2C), in single individuals with clinical characteristics overlapping KS. Here, we present a novel cohort of five patients with *de novo* loss of function mutations affecting the histone methyltransferase KMT2C. Our clinical data delineates the KMT2C phenotypic spectrum and reinforces the phenotypic overlap with KS and other related ID disorders. To elucidate the common molecular basis of the neuropathology associated with mutations in KMT2C and EHMT1, we characterized the role of the *Drosophila* KMT2C ortholog, trithorax related (trr), in the nervous system. Similar to the *Drosophila* EHMT1 ortholog, G9a, trr is required in the mushroom body for short term memory. Trr ChIP-seq identified 3371 binding sites, mainly in the promoter of genes involved in neuronal processes. Transcriptional profiling of pan-neuronal trr knockdown and G9a null mutant fly heads identified 613 and 1123 misregulated genes, respectively. These gene sets show a significant overlap and are associated with nearly identical gene ontology enrichments. The majority of the observed biological convergence is derived from predicted indirect target genes. However, trr and G9a also have common direct targets, including the *Drosophila* ortholog of Arc (Arc1), a key regulator of synaptic plasticity. Our data highlights the clinical and molecular convergence between the KMT2 and EHMT protein families, which may contribute to a molecular network underlying a larger group of ID/ASD-related disorders.

Author summary

Neurodevelopmental disorders (NDDs) like intellectual disability (ID) and autism spectrum disorder (ASD) present an enormous challenge to affected individuals, their families, and society. Understanding the mechanisms underlying NDDs may lead to the development of targeted therapeutics, but this is complicated by the great clinical and genetic heterogeneity seen in patients. Mutations in hundreds of genes have been implicated in NDDs, giving rise to diverse clinical presentations. However, evidence suggests that many of these genes lie in common biological pathways, and mutations in genes that are involved in similar biological functions give rise to more similar clinical phenotypes. Here, we define a novel ID disorder with comorbid ASD (ID/ASD) caused by mutations in KMT2C. This disorder is defined by clinical features that overlap with a group of other disorders, including Kleefstra syndrome, which is caused by EHMT1 mutations. In the fruit fly, we show that the KMT2 and EHMT protein families regulate a highly converging set of biological processes. Both EHMT1 and KMT2C, encode histone methyltransferases, which regulate gene transcription by modifying chromatin structure. Further understanding of the common gene regulatory networks associated with this group of ID- and ASD-related disorders may lead to the identification of novel therapeutic targets.

Introduction

Kleefstra syndrome (OMIM #610253) is a neurodevelopmental disorder that is caused by haploinsufficiency of EHMT1 [1, 2]. EHMT1 encodes a histone methyltransferase that regulates gene expression through modification of chromatin structure and through interactions with other transcription factors [2-4]. The core phenotype of Kleefstra syndrome is characterized by intellectual disability (ID), childhood hypotonia, and distinctive facial characteristics. Autism spectrum disorder(ASD) and other behavioural problems such as sleep disturbances and feeding difficulties are also frequently observed [5, 6]. Kleefstra syndrome shares considerable phenotypic overlap with several other disorders characterized by ID, ASD, other behavioural problems and hypotonia [7, 8]. These disorders include Pitt-Hopkins syndrome, Smith-Magenis syndrome, Rett syndrome, MBD5 deletion/duplication, and Angelman syndrome. We have collected a cohort of individuals with clinical characteristics that fit within this clinical spectrum. In the pre-exome sequencing era, around 25% of these cases were positive for EHMT1 haploinsufficiency characterizing Kleefstra syndrome [9]. Previously, we hypothesized that these unsolved, EHMT1 mutation-negative individuals with overlapping clinical features may carry mutations in other genes that code for proteins acting in EHMT1-related molecular pathways or biological processes. Genetic analyses of four individuals within our cohort using next generation sequencing methods revealed potentially causative *de novo* mutations in four genes (MBD5, SMARCB1, NR1I3, and KMT2C), all of which encode proteins involved in regulation of gene expression and/or chromatin structure [9].

Available physical interaction data between these proteins [10-14] and pairwise genetic interactions identified between the *Drosophila* orthologs of these genes [9], supported the hypothesis that EHMT1 and the genes identified in the individuals presenting with clinical overlap act in shared molecular pathways. In particular, a very strong antagonistic genetic interaction was observed between the *Drosophila* EMHT1 ortholog, G9a, and the KMT2C ortholog, trithorax related (trr), in the context of *Drosophila* wing development [9]. Genetic combination of G9a overexpression together with trr knockdown led to a complete developmental arrest and death of the targeted wing tissue, while the misregulation of either gene alone had only subtle effects on wing morphology [9]. Thus, although these data argued for a conserved functional relationship between the two genes, the underlying shared molecular basis remained unknown.

G9a and trr are both histone methyltransferases, but they have different substrates. G9a mediates mono- and dimethylation of histone H3 at lysine 9 (H3K9me2). This is generally regarded as a repressive histone mark, however, in mammals the EHMT1 protein has also been shown to activate gene expression independent of its methyltransferase activity [15, 16]. G9a mutant flies are fully viable [17, 18] but do show phenotypic differences compared to wildtype flies. Loss of *Drosophila* G9a delays embryonic development [19].

In larval stages, G9a mutants show defects in the morphology of multidendrite neurons and have altered crawling behaviour [18]. Adult G9a mutants have defects in habituation learning, short and long term courtship memory, and show decreased tolerance to virus infections [18, 20]. Genome-wide identification of genomic regions with reduced H3K9me2 in G9a mutant larvae revealed a large number of target sites in genes that are strongly enriched for neuronal functions [18]. In mouse knockout models, loss of Ehmt1 causes phenotypes that are reminiscent of Kleefstra Syndrome, including deficits in learning and memory, increased anxiety, hypotonia, cranial abnormalities, and developmental delay [21-24]. These studies show that the EHMT family of proteins are evolutionarily conserved regulators of neurodevelopmental processes and cognition. However, the underlying molecular mechanisms and functional partners of EHMT proteins in these processes are poorly understood.

Studies investigating the molecular biology of *Drosophila* trr and its binding partners have revealed a potential dual function for this protein in the regulation of gene expression. Trr mediates mono- and tri-methylation of histone H3 at Lysine 4 (H3K4me1 and H3K4me3) [14, 25], histone modifications that are found at enhancers and active gene promoters, respectively [26]. On one hand, trr and its mammalian orthologs KMT2C and KMT2D, are present in a conserved COMPASS-like (complex of proteins associated with Set1) protein complex that mediates H3K4me1 at enhancers [27, 28]. Additionally, trr interacts with the ecdysone receptor and is a co-activator of ecdysone-mediated transcription that deposits H3K4 trimethylation at promoters [14]. The role of trr in the nervous system and how it relates to G9a is unknown.

Here, we describe the first human cohort with *de novo* loss of function mutations in the trr ortholog, KMT2C, allowing us to delineate the phenotypic spectrum and define the core phenotype associated with mutations in this gene. In agreement with the single previously reported patient, these individuals show clinical overlap with Kleefstra syndrome and other related neurodevelopmental disorders; reemphasizing (1) an important role for KMT2C in neurodevelopment, and (2) a biological link with EHMT1. We further show that *Drosophila* trr shares the essential role in neurodevelopment with its human ortholog, and provide evidence for biological convergence between trr and G9a.

Results

De novo mutations in KMT2C cause ID and autism spectrum disorder

Through the large scale application of diagnostic whole exome sequencing for unexplained neurodevelopmental disorders in the clinical genetics centers at the Radboudumc and Maastricht UMC [29], we identified five *de novo* KMT2C mutations. All these mutations are predicted to cause loss of function, including c.5216del (p.Pro1739Leufs*2) in individual

1, c.7550C>G (p.Ser2517*) in mosaic (≈30% of blood cells) in individual 2, c.1690A>T (p.Lys564*) in individual 3, and c.10812_10815del (p.Lys3605fs) in individual 4 (**Figure 1A**). Using microarray-based comparative genomic hybridization for patient 5, an intragenic 203kb *de novo* deletion (Chr7: 151858920-152062163)x1 was identified (**Figure 1A**) and confirmed by locus-specific qPCR.

Identification of these five novel mutations in addition to the previously published case [9] has allowed us to establish the clinical phenotype associated with KMT2C mutations (**Table 1**). All individuals had ID, ranging from mild to severe, language and motor delay, and autism or Pervasive Developmental Disorder (PDD), a condition in the autistic spectrum. Other recurrent clinical features were short stature (2/6), microcephaly (3/6), childhood hypotonia (3/6), kyphosis/scoliosis (3/6) and recurrent respiratory infections (2/6). Kleefstra-like facial dysmorphisms, including flattened midface, prominent eyebrows, everted lower lip, thick ear helices were observed in several individuals (**Figure 1B**).

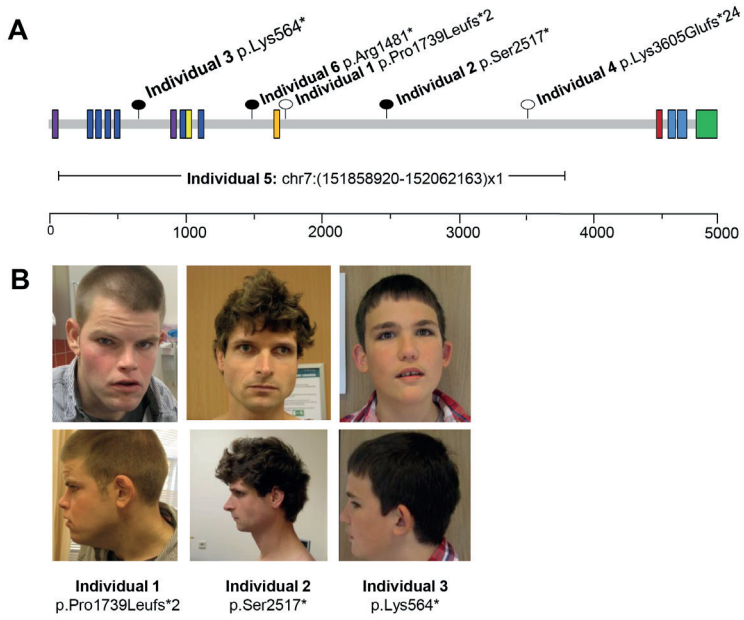


Figure 1: Patients and identified KMT2C mutations.

(A) Schematic view of the KMT2C protein with reported domains (purple: AT hook DNA binding domain; dark blue: zinc finger domain; yellow: cysteine rich; orange: High mobility group (HMG); dark red: Ring finger; light blue: "FY-rich" domain; dark green: SET and Post-SET domains) and identified frameshift mutations (open lollipops), nonsense mutations (closed lollipops) and deletion. Scale bar represents amino acid position. (B) Frontal and lateral photographs of individual 1 at age 29 years, individual 2 at age 31 years, individual 3 at age 15 years. Though there is variable facial features, KSS dysmorphisms are observed, including flattened midface (individual 1), prominent eyebrows (individuals 1 and 3), thick ear helices (individuals 1 and 3).

Table 1: Summary of molecular and clinical features of individuals with *KMT2C* mutations

	Individual 1	Individual 2	Individual 3	Individual 4	Individual 5	Individual 6 Kleefstra et al.(2012)
Gender	Male	Male	Male	Female	Female	Female
Age of examination	29 years	31 years	15 years	7 years	10 years	15 years
Mutation (NM_170606.2)						
Chromosome position (Hg19)	g.151880108del	g.151874988G>C	g.151947983T>A	g.151859847_151859850del	-	g.151891591G>A
cDNA change	c.5216del	c.7550C>G	c.1690A>T	c.10812_10815del	-	c.4441C>T
Amino acid change	p.(Pro1739Leufs*2)	p.(Ser2517*)	p.(Lys564*)	p.(Lys3605Glufs*24)	-	p.(Arg1481*)
Mosaic	-	+ (30% blood)	-	-	-	-
Deletion	-	-	-	-	7q36.1 (151858920-152062163)x1	-
Additional *de novo* mutations	*PHF21A*[1]	*UBR5*[2]	-	-	-	-
		C11orf35[3]				
Growth						
Height	171.5 cm (-1.7 SD)	179 cm (-0.5 SD)	160 cm (-2 SD)	109 cm (-3 SD)	N/A	148 cm (-2.5 SD)
Weight	63.5 kg (+0.6 SD)	53.8 kg (-1.5 SD)	55 kg (+1,7SD)	16 kg (-1.5 SD)	20 kg (-2.5 SD)	41 kg (0 SD)
Head circumference	56.6 cm (-0.5 SD)	57 cm (-0.5 SD)	55 cm (-0,6 SD)	47.5 cm (-2.25 SD)	49.5 cm (-2 SD)	52 cm (-2 SD)

Table 1: Summary of molecular and clinical features of individuals with KMT2C mutations *(continued)*

	Individual 1	Individual 2	Individual 3	Individual 4	Individual 5	Individual 6 Kleefstra et al.(2012)
Development						
Intellectual disability	+ (Moderate)	+ (Mild)	+ (Moderate – IQ 50)	+ (Mild - IQ63)	+ (Severe)	+ (Moderate - IQ 35)
Language and motor delay	+	+	+	+	+	+
Neurological						
Behavior problems	+ (Autistic-traits)	+ (Autism)	+ (PDD-NOS, ADHD)	+ (Autism, sleeping disorder)	+ (Automutilation)	+ (hyperactivity, aggressiveness)
Childhood hypotonia	-	-	+	-	+	+
Epilepsy	+	-	-	-	+	-
Skeletal						
Kyphosis/Scoliosis	+ (thoracal kyphosis)	+ (Scoliosis)	-	-	+ (Kyphosis)	-
Other						
	PKU, RRI	Strabismus, cryptorchidism	Bifid uvula, hypospadia, bilateral inguinal hernia	RRI, dry skin, hoarse voice	Plagiocephaly	-
MRI						
	Not performed	N/A	Normal	Normal	Non-progressive enlarged extracerebral space	N/A

Abbreviations: N/A = not available; PKU = phenylketonuria; RRI = recurrent respiratory infections; [1] *PHF21A*: c.1956del; p.(Ala653Profs*103), [2] *UBR5*: c.5720G>A; p.(Arg1907His), [3] C11orf35; p.(Pro602Leu)

The *Drosophila* KMT2C ortholog, trr, is required for short term courtship memory

To assess the functional role of KMT2C in neurons, we investigated the closest *Drosophila melanogaster* ortholog, trithorax related (trr), which shares a one-to-two evolutionary relationship with the human paralogs KMT2C and KMT2D [28]. Since homozygous mutations in trr are lethal [30], we used the UAS/Gal4 system [31] and inducible RNA interference (RNAi) [32] to assess the role of trr in the adult fly nervous system. Knockdown of trr was targeted specifically to the mushroom body (MB), the learning and memory center of the fly brain, using the R14H06-Gal4 driver line from the Janelia FlyLight collection [33]. To estimate the knockdown efficiency under these conditions we co-expressed UAS-trr-RNAi with UAS-mCD8::GFP and performed immunohistochemistry using a trr antibody [34] (**Figure 2A-2H**). Trr protein is localized in the nuclei of the mushroom body calyx, as seen by colocalization with DAPI. A clear reduction of trr staining in cells expressing trr-RNAi, which are marked by UAS-mCD8::GFP, is observed (**Figure 2E-2H**). Under these conditions we observed no gross morphological defects in the MB upon trr knockdown (**Figure 2I and 2J**). Next, to test these MB-specific knockdown flies for defects in learning and memory we used a classic behaviour paradigm known as courtship conditioning. In this assay male flies exhibit a learned reduction of courtship behaviour after rejection by a non-receptive premated female [18, 35]. We tested short term memory by measuring the courtship index (CI) in naïve males compared to males exposed to sexual rejection for one hour by a premated female. Control flies expressed a transgenic RNAi construct targeting mCherry, which is inserted into the same genetic background as the trr-RNAi. These controls showed significant reduction in CI in rejected flies compared to naïve (**Figure 2K**). Flies expressing the trr-RNAi construct did not exhibit a significant reduction in CI in response to rejection (**Figure 2K**), and as a result, had a significantly lower learning index (LI) than the controls (**Figure 2L**). These data show that trr is required in the mushroom body for short term memory.

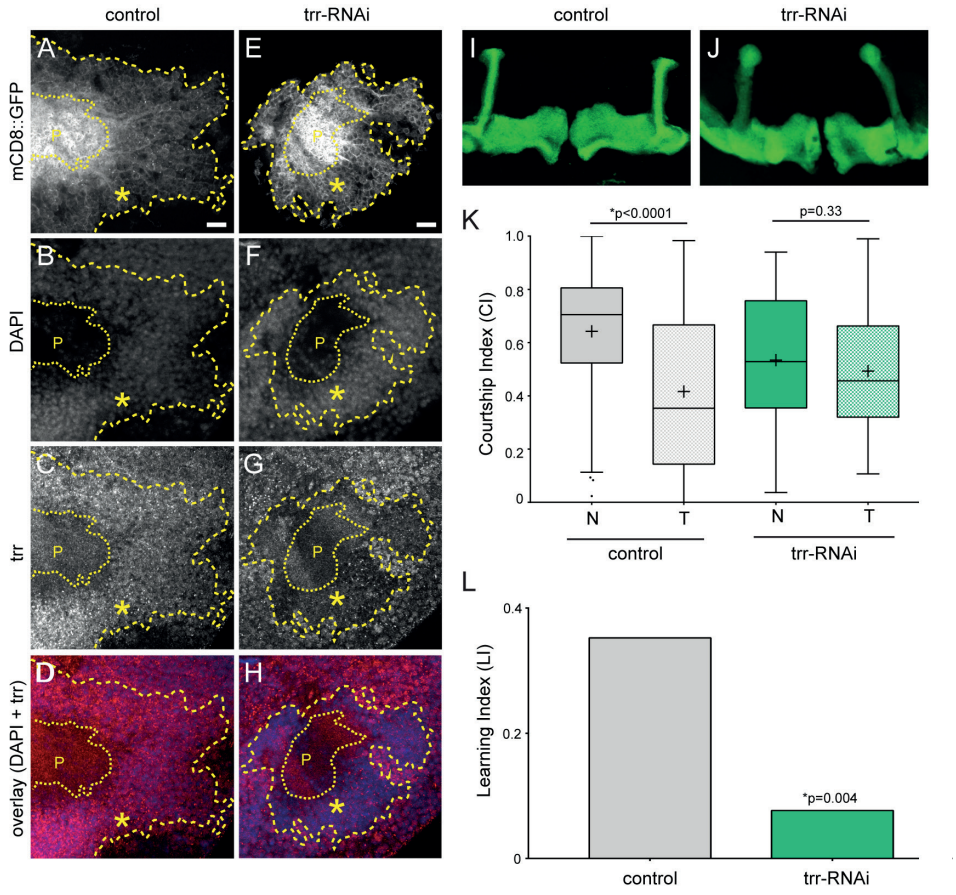


Figure 2: Trr is required in the mushroom body for short term memory.

Fluorescent confocal images of control (A-D) and trr knockdown (E-H) adult male brains. UAS-mCD8::GFP calyx (A, E) shows the expression domain of the R14H06-Gal4 driver in the mushroom body and is marked by yellow dashed lines and asterisk. The scale bar represents 10 µm. DAPI (B, F) is shown to identify nuclei and note the low nuclear density preduncle (p). Trr (C, G) is labeled by immunohistochemistry using an anti-trr antibody. The overlay of DAPI and trr signal (D, H) shows a reduction of trr in the target cells (blue and red channels). (I-J) Confocal projections showing the main axonal lobes of the mushroom body that are labeled by UAS-mDC8::GFP through expression with the R14H06-Gal4 driver in control flies (I) and trr knockdown flies (J). The scale bar represents 10 µm. (K) Standard boxplots representing the courtship indexes (CIs) resulting from courtship conditioning in control and trr knockdown flies. + indicates the mean. The mean CI for naïve and trained flies was compared using the Mann-Whitney test. (L) Learning Indexes (LI) for controls and trr knockdown flies derived from the CIs. Trr knockdown males have a significantly reduced LI (randomization test, 10,000 bootstrap replicates).

Trr binds to the promoter of genes important for the nervous system

Given the known role of trr as a transcriptional co-activator [14], we sought to identify genomic binding regions of trr to gain insight into its target genes and the processes that it regulates in the nervous system. To do this, we conducted chromatin immunoprecipitation combined with next generation sequencing (ChIP-seq) using wildtype adult *Drosophila* heads and a validated ChIP grade antibody directed against trr [14, 34]. After exclusion of poorly aligned and ambiguously mapped reads (**Table S1**), the MACS2 (model based analysis of chip-seq) algorithm [36] was used to identify genomic regions that were significantly enriched for trr binding in two biological replicates (**Table S1**). This analysis identified 3371 trr binding sites (**Table S2**) covering 1.27% of the genome. About 75% of trr binding sites (2564) were located within 1kb up- or downstream of the transcription start sites (tss) of 2362 unique genes. The remaining 25% of trr binding sites were associated to other genomic features (low complexity, 5'- or 3'-prime region, transcription termination site, intron/exon), each at a low frequency (**Figure 3A, Table S2**). Taking the genome-wide abundance of annotated features into account, trr binding at the tss is six fold enriched compared to random genomic positions (**Figure 3B**). The average trr occupancy profile over the tss of all genes shows a clear enrichment for trr when compared to the input control (**Figure 3C**). The observed association of trr with promoter regions in fly heads is consistent with previously published trr ChIP-seq data from cultured S2 cells [25]. Gene Ontology (GO) enrichment analysis of the 2362 unique genes with trr binding at the tss revealed a strong enrichment for neuronal terms, such as "axon extension" and "neuron recognition". Also enriched was the term "negative regulation of Ras protein signal transduction", a process that is corrupted in a series of ID disorders referred to as RASopathies [37] (**Figure 3D, Table S3**). Taken together, our data demonstrate that, in fly heads, trr binds to the promoter of many genes involved in neuronal processes.

G9a and trr show convergence in genomic targets

Given the clinical overlap between patients with KMT2C and EHMT1 mutations (**Figure 1, Table 1)**, and our previous studies showing a very strong genetic interaction between the two *Drosophila* orthologs [9], we reasoned that trr and G9a may regulate the expression of common genes and/or biological pathways. We investigated whether G9a and trr have common direct target genes (**Figure 3E-3F**). For this we compared the genes associated with trr binding sites identified here (**Figure 3A-3D**) with previously determined G9a target genes that were identified by investigating loss of H3K9me2 in G9a mutants [18]. This analysis revealed a significant overlap of 1047 genes based on hypergeometric probability ($p < 1.9*10\text{-}37$, fold change = 1.35, **Figure 3E**). These genes are primarily enriched for GO-terms associated with neuronal development and function, including "axon extension",

"neuron recognition", and again, "negative regulation of Ras protein signaling" (**Figure 3F, Table S4**). This suggests that trr and G9a have the potential to directly regulate a common set of genes with important functions in neurons.

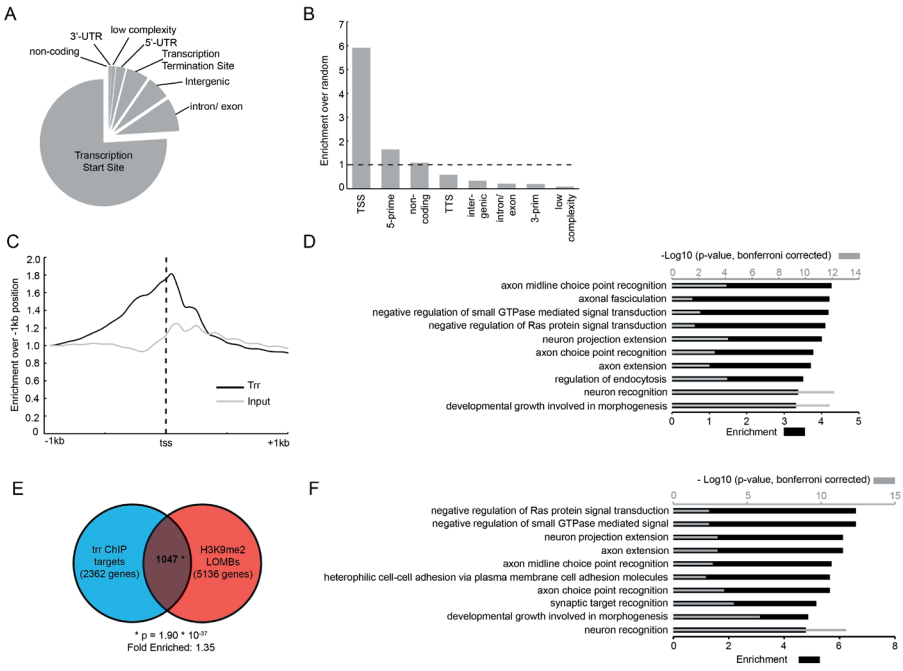


Figure 3: Trr localizes to promoters of neuronal genes in *Drosophila* heads and shows a significant overlap with G9a targets.

(A) Pie-chart representing the location of trr binding sites in annotated genomic features as specified by HOMER software. (B) Fold enrichment of trr binding sites in annotated genomic features compared to an equivalent group of random genomic positions. (C) Average trr occupancy (black) of all transcription start sites (tss) relative to the –1kb region compared to the average read depth in the input control (grey). (D) Gene ontology enrichment analysis of genes with a trr binding site near the tss. Shown here are the top 10 enriched terms. Enrichment is indicated by black bars (lower x-axis). The –log10 transformation of p-values is indicated by grey bars (upper x-axis). (E) Venn diagram showing the overlap between predicted trr target genes identified here, and predicted G9a targets that were previously published [18]. The overlap of 1047 genes is larger than expected by random chance, based on a hypergeometric test (p-value= 1.9*10-37 and 1.35 times enriched). (F) Top 10 enriched GO terms for biological processes identified for the 1047 overlapping predicted targets for G9a and trr. Enrichment is indicated by black bars (lower x-axis). The –log10 transformation of p-values is indicated by grey bars (upper x-axis).

G9a and trr regulate the expression of common genes and biological processes

To characterize the effect of trr and G9a mutations at the transcriptional level, we conducted whole-transcriptome mRNA sequencing (RNA-seq) on wildtype and mutant fly heads. Validation of pan-neuronal trr knockdown in fly heads was performed by RT-qPCR and revealed a 34% reduction in trr mRNA in whole fly heads, encompassing both the targeted neuronal knockdown cells and all other non-targeted cells (**Figure S1**). Only reads aligned unambiguously to exons were considered for RNA-seq analysis (**Table S5**). Upon trr knockdown, 613 genes were differentially expressed compared to controls (p-adj < 0.05, fold change > 1.5), with 341 genes downregulated, and 272 genes up regulated **(Figure 4A, Table S6**). GO enrichment analysis of the differentially expressed genes in trr knockdown heads revealed mainly terms related to metabolism (**Figure 4B**). In G9a null mutant heads, 1123 genes were differentially expressed compared to controls (p-adj < 0.05, fold change > 1.5), with 796 genes downregulated and 327 genes upregulated (**Figure 4C, Table S7**). These differentially expressed genes shared a striking overlap in GO enrichment when compared to differentially expressed genes in trr knockdown heads. In fact, 4 of the 6 GO terms that were enriched upon trr knockdown were also enriched in the analysis of G9a mutant heads (**overlapping GO terms in bold, Figure 4B and D**), including "gluconeogenesis", "steroid metabolic processes", "fatty acid beta oxidation", and "respiratory electron transport chain".

In addition to the overlapping enriched GO terms, we identified 119 differentially expressed genes represented in both mutant conditions (**Figure 4E-F**). This overlap is significantly larger than expected by random chance ($p < 6.4*10\text{-}23$, fold change = 2.7, hypergeometric test) (**Figure 4E**). GO enrichment analysis of these 119 genes again revealed an nearly identical set of enriched GO terms, with 9 of the 11 enriched terms represented in the analysis of all genes dysregulated in G9a and trr mutant conditions (**overlapping GO terms in bold, Figure 4F**). The striking overlap of specific GO terms associated with genes that are differentially expressed upon loss of trr and G9a suggest a high level of functional convergence between the two proteins.

Of the 119 commonly mis-regulated genes identified in G9a mutants and trr knockdown flies 5 genes were upregulated in both conditions, while 47 genes were downregulated in both conditions; 18 genes were upregualted in G9a and downregualted in trr, while 49 genes were downregulated in G9a and upregualted in trr (**Table S8**). Although this distribution is not fully random, there is no clear pattern. The significant enrichment observed in the majority of the categories suggests that many of the common differentially expressed genes may result from indirect regulation. Notably, only five of the overlapping genomic targets (**Figure 3E, 1047 genes**) are also found to be differentially expressed (**Figure 4E, 119 genes**). These five genes - PCB, mTTF, Acer, Reg-2, and Arc1 - represent potential common direct targets of G9a and trr that could act as hub genes to

account for the high degree of biological convergence among secondary targets (**Table S9, see discussion**).

To further explore the origin of the biological convergence of differentially expressed genes in trr RNAi and G9a mutant heads, we classified these genes into "potentially direct" and "potentially indirect" targets. Potentially direct targets for trr, a transcriptional activator, were defined as genes with an associated trr binding site that were down regulated in trr RNAi flies. Of 341 downregulated genes, 47 fulfill these criteria. For G9a, direct targets were defined as genes with an associated loss of H3K9me2 (a repressive modification) in G9a mutants (as defined in Kramer et al. [18]) that were up regulated in G9a mutant heads. Of the 327 genes upregulated in G9a mutant heads, 84 were identified as potential direct G9a targets using these criteria. GO enrichment analysis of predicted direct targets revealed metabolic terms and several terms relevant to learning and memory, such as "sensory perception" and "cellular calcium ion homeostasis" (**Figure S2**). However, no overlap in enriched GO terms was observed among the potentially direct target gene lists of trr and G9a. In contrast, GO enrichment analysis of predicted indirect targets (**Figure S2**) was nearly identical to that observed when all differentially expressed genes were analyzed (**Figure 4**) and the overlap of specific terms between G9a and trr datasets was even stronger (5 of the 6 GO terms that were enriched upon trr knockdown were also enriched in the analysis of G9a mutant heads) (**Figure S2**). Taken together, this analysis suggests that most of the biological convergence between dysregulated transcriptomes in trr and G9a mutant conditions happens via indirect regulation. This convergence could result from misregulation of a limited number of common direct targets, or by misregulation of different direct target genes that act in common biological pathways.

In summary, whole-transcriptome analysis revealed a large and significant overlap between genes and biological pathways that are differentially expressed upon loss of trr and G9a, suggesting functional convergence between these two proteins in the fly brain.

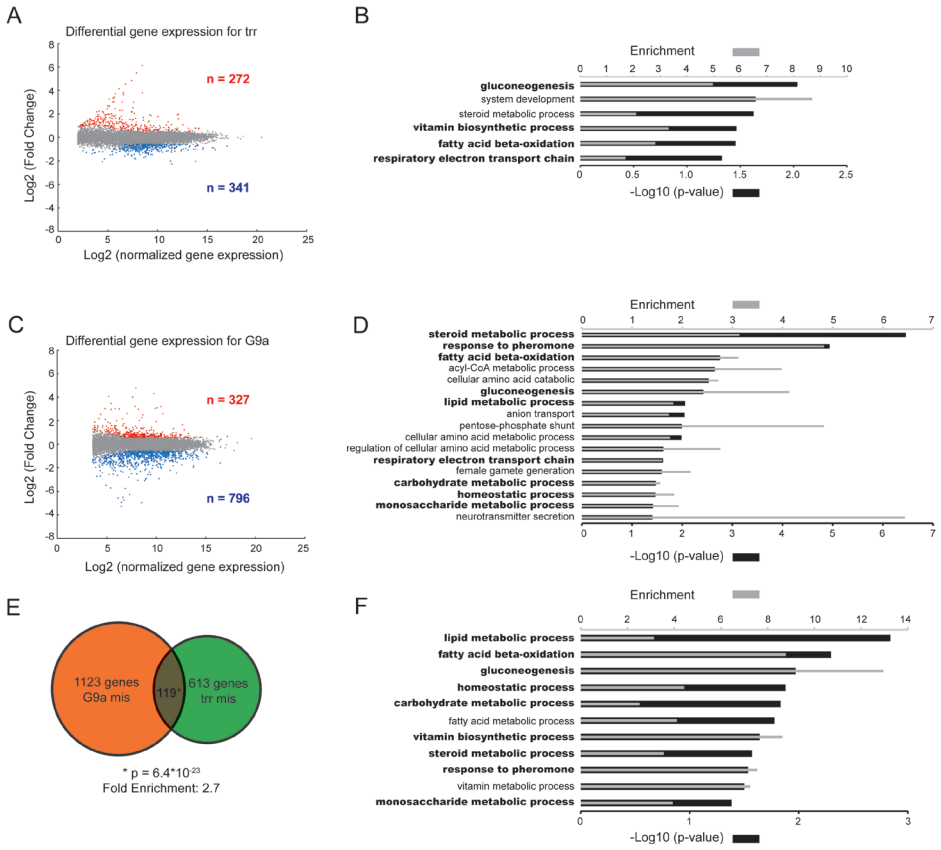


Figure 4: Differentially expressed genes in trr and G9a mutants show a strong biological overlap.

(A,C) Scatter plots showing Log2 fold changes plotted against the Log2 normalized expression. Significantly up- and downregulated genes (p-adj < 0.05, FC > 1.5) are represented by red and blue dots respectively in the G9a (A) and trr (C) mutant. (B,D,F) Enriched gene ontology terms identified using the Panther software (GO-slim setting) for differentially expressed genes. The –log10 transformation of p-values is indicated by black bars (lower x-axis). Grey bars (upper x-axis) indicate enrichment. Bold gene ontology terms indicate the overlap between the terms. GO terms are indicated for Trr (B) and G9a mutant heads (D). (E) Venn Diagram showing the overlap of differential expressed genes between the two mutant conditions. The overlap of 119 is significantly more than expected by random chance, based on a hypergeometric test (p-value = 6.4*10-23 and 2.7 times enriched). (F) GO term enrichment of the overlapping genes between trr- and G9a differentially expressed genes.

Discussion

Dynamic regulation of gene expression is essential for brain development and function. To date, at least 68 genes encoding chromatin regulating proteins have been implicated in the etiology of ID [38-40]. It is possible that disruption of chromatin related biological networks through mutations in different genes results in an overlapping phenotypic spectrum in the corresponding disorders. Indeed, several chromatin-related disorders, such as Pitt-Hopkins syndrome, Smith-Magenis syndrome, Rett syndrome, MBD5 deletion/duplication, and Angelman syndrome, have a clear clinical overlap characterized by ID/ASD and additional behavioral and neurological problems [7, 8]. Here, we describe five patients with *de novo* heterozygous loss of function mutations in KMT2C, and demonstrate a clinical overlap with this group of ID/ASD disorders (**Figure 1, Table 1**). In *Drosophila*, we demonstrate molecular convergence between trr, the KMT2C/D ortholog, and G9a, the ortholog of EHMT1, which is involved in Kleefstra syndrome. The observed clinical and molecular convergence between the KMT2 and EHMT protein families suggests that they are important players in an overlapping gene regulatory network that is critical for normal neuronal development and function.

KMT2 and EHMT histone methyltransferase families converge in the regulation of adult brain function

An increasing body of evidence suggests that chromatin modifiers contribute to gene regulatory networks that are important for complex brain processes like learning and memory. The molecular activity and protein interaction partners of trr have been well described, but only for early developmental stages and in cultured S2 cells [14, 25, 28]. Using conditional knockdown we showed that trr is required in post-mitotic MB neurons for normal short term courtship memory (**Figure 2L**). We did not observe any gross morphological defects in the MB upon trr knockdown (**Figure 2I and 2J**), and although we cannot rule out a subtle role for trr in MB morphogenesis, this suggests that trr may regulate behaviour and memory through a role in the post-developmental functioning of adult MB neurons. One key interaction partner for trr is the ecdysone receptor (EcR), a nuclear hormone receptor that binds to 20-hydroxyecdysone and requires trr-mediated H3K4me3 for optimal gene activation [14]. 20-hydroxyecdysone is actively synthesized in response to courtship conditioning and is required for long term courtship memory [41], supporting the idea that trr may be involved in the acute regulation of adult brain function through its role in EcR-mediated transcription. Indeed, thirteen known ecdysone responsive genes were identified in our trr genomic binding sites (**Table S2**), consistent with the known trr-EcR interaction. *Drosophila* G9a is also required in adult neurons for normal courtship memory [18], and the mouse G9a orthologs, Ehmt1 and Ehmt2, have been implicated in postnatal regulation of

fear memory [24]. Interestingly, the specific histone modifications that are deposited by the EHMT and KMT2 protein families (H3K9me2 and H3K4me3, respectively), are among the very few forms of histone methylation that are dynamically regulated in the rat brain in response to fear conditioning [42, 43]. Together, the evidence points towards a role for the EHMT and KMT2 families in adult brain plasticity, through active regulation of histone modifications and gene expression.

Molecular convergence of *Drosophila* G9a and trr

Here, we describe a high level of convergence in genes that are differentially expressed in trr and G9a mutant fly heads (**Figure 4E**). Although this overlap is significantly more than expected by random chance, the match at the level of biological processes is even more striking a nearly perfect overlap in the enriched GO terms between the two independent mutant conditions (**bold GO-terms in Figure 4B and 4D**). This suggests that trr and G9a may regulate; (1) distinct gene sets that operate in similar biological pathways, and / or (2) a limited number of joint key target genes that influence these pathways. In total, we observed five potential hub genes that are differentially expressed in both mutant conditions and identified in ChIP experiments as potential direct targets for both G9a and trr (**Table S9**). Two of these genes, pyruvate carboxylase (PCB), and mitochondrial transcription termination factor (mTTF), have a clear link to mitochondrial metabolism [44, 45], therefore, their misregulation could affect aspects of metabolism that are reflected in the GO terms associated with the differentially expressed genes in trr and G9a mutant heads (**Figure 4**). Loss of function mutations in PCB give raise to several disorders that include developmental delay, ID and seizures, and its misregulation may thus contribute to manifestation of these features in EHMT1 and KMT2C patients.

Interestingly, four of the five potential hub genes (Reg-2, ACER, PCB, and Arc-1) have a direct or indirect link to memory formation. Reg-2 and ACER are both linked to circadian rhythm [46, 47], which is known to be important for memory formation (reviewed in [48, 49]). The metabolic protein PCB has been linked to age-induced memory impairment due to the cumulative damaging effect of metabolic free radicals in the aging brain [44]. Arc-1 is the ortholog of mammalian Arc (activity regulated cytoskeletal protein), an immediate early gene that is activated in response to neuronal activity associated with learning and memory [50]. Arc is localized at the postsynaptic density and regulates many forms of synaptic plasticity including LTP, LTD and homeostatic synapse scaling [51]. A recent study has implicated mouse Ehmt1 in repression of Arc transcription in response to homeostatic synaptic scaling [52]. Arc was predicted to be indirectly regulated by Ehmt1 based on a lack of differential dimethylation of H3K9 at the Arc promoter [52]. Our data suggests that *Drosophila* G9a deposits H3K9me2 at the 3-prime end of Arc-1 [18], which has not been tested in mouse. Although, this might reflect a differential mechanism of gene regulation

between *Drosophila* and mammals, it appears that Arc and Arc-1 represent an evolutionarily conserved target for EHMT proteins. Having identified Arc-1 as a trr target gene in *Drosophila*, it will be interesting to see if this evolutionary relationship also holds true for KMT2 proteins.

It is possible that direct misregulation of specific genes, like Arc-1, may underlie cognitive defects in human and *Drosophila* with mutations in KMT2 and EHMT genes. However, our analysis of candidate genomic target genes in *Drosophila* revealed hundreds of other genes that may be involved. At the level of genomic targets, we observed more than 1000 genes that are potentially regulated by both G9a and trr. These genes show a strong enrichment for neuronal GO terms, such as "synapse assembly", "brain development", and "cognition" (**Table S4**). Notably, "negative regulation of Ras protein signaling", the most enriched GO-term (**Figure 3F**), is linked directly to a subset of ID disorders known as RASopathies [37]. However, the relevance of these targets remains in question as we do not observe differential expression for most of these genes in G9a and trr mutant fly heads, at least in the steady state conditions examined here. This, however, does not exclude the possibility that the histone modifications mediated by trr and G9a, may poise these genes for transcriptional activation. It is possible that trr- or EHMT-mediated regulation is only relevant in specific cells or in response to certain stimuli, such as the formation or retrieval of memory. Such activity-dependent or cell-specific gene expression mechanisms are likely to escape detection in our analyses in which steady-state expression is determined in whole heads.

Clinical convergence in chromatin-related ID disorders

The investigation into the molecular convergence between EMHT and KMT2 proteins in this study was motivated by the identification of an overlapping clinical phenotype resulting from mutations in these genes. This convergence likely goes beyond EHMT1/G9a and KMT2C/trr. Apart from KMT2C/trr, mutations in additional components of the well-defined COMPASS protein complex, including the KMT2C paralog KMT2D, and the histone demethylase KDM6A, are causative for Kabuki syndrome, another ID disorder which shares autistic behavior and developmental delay with Kleefstra syndrome and the patients described here [8]. Molecular and clinical convergence has also been demonstrated between Pitt-Hopkins syndrome (causative gene TCF4) and Kleefstra syndrome [8, 53]. Further fundamental research will likely reveal additional molecular connections between genes that are mutated in other ID/ASD disorders showing clinical overlap with KMT2C halpoinsufficiency. Promising candidates to be involved in a common molecular network include TCF4 [54], RAI1 [55], MECP2 [56], MBD5 [57, 58], KMT2D [59], UBE3A [60], which are all ID genes encoding proteins involved in gene regulation. Expanding our knowledge of these molecular networks may help to understand pathways underlying ID/ASD that could be exploited in the development of therapies for genetically distinct but clinically related disorders.

Materials and Methods

Identification and prioritization of variants

Individuals 1-4 had unexplained ID or developmental delay and were ascertained through family-based whole exome sequencing in a diagnostic setting in the Department of Human Genetics at Radboudumc and MaastrichtUMC. Exome sequencing and data analysis were performed as previously described in the probands and their unaffected parents [29]. Individual 5 was ascertained during a clinical genetic workup due to failure to thrive and lack of expected normal childhood development at Children's Hospital Colorado (Aurora, Colorado). Chromosomal microarray analysis for individual 5 was performed using both copy number and single-nucleotide polymorphism (SNP) probes on a whole genome array (CytoScan HD platform; Affymetrix, Santa Clara, CA). The *KMT2C* deletion was confirmed to be *de novo* by parental microarray testing. Additionally, this deletion was confirmed using locus-specific qPCR. The primer pairs used were directed to KMT2C and two adjacent negative control regions (**Table S10**). All data were analyzed and reported using the February 2009 (hg19) human genome build.

Although all individuals harbored *de novo* loss of function mutations in KMT2C (**Figure 1A**), exome results revealed that individual 1 and 2 carry one or two additional *de novo* mutations that do not appear to be related to the core phenotype of the individuals presented here. Individual 1 carried a second *de novo* truncating mutation in *PHF21A*, which is located in the 11p11.2 contiguous gene deletion syndrome named Potocki-Shaffer syndrome, PSS (OMIM #601224) [61, 62]. Three individuals with translocation breakpoints through *PHF21A* were reported previously and it was concluded that this gene was associated with the PSS. However, no intragenic loss of function mutations are reported in association with this syndrome so far. The mutation in individual 1 is present in the last exon of the gene and is therefore not subjected to nonsense mediated RNA decay, however it is possible that this variant may contribute to the phenotype observed in this patient.

In individual 2 we found the *KMT2C* change c.7550C>G (p.Ser2517*) in mosaic in blood. Individual 2 carries two additional *de novo* missense mutations in *C11orf35* (p.Pro602Leu) and *UBR5* (p.Arg1907His). Protein function or mutations associated with other diseases for C11orf35 are not known and thus we cannot rule out this variant to the patient phenotype. With respect to *UBR5*, the same mutation has been previously described in a family with adult myoclonic epilepsy [63], but that family was not associated with ID/Neuro Developmental Delay. The missense change in *UBR5* individual 2 is therefore unlikely to cause his ID and autism. Moreover, no epilepsy is present in this individual at age 10.

Fly stocks and genetics

Flies were reared on standard cornmeal-agar media at 25°C on a light/dark cycle of 12h/12h in 50% or 70% humidity. The following stocks were obtained from Bloomington *Drosophila* Stock Center: UAS-trr-RNAi generated by the Transgenic RNAi Project (TRiP) (stock #36916: y1 sc*v1;P{TRiP.HMS01019}attP2), UAS-mCD8::GFP (Stock #5137: y [1] w [*]; P{w [+mC]=UAS-mCD8::GFP.L}LL5, P{UAS-mCD8::GFP.L}2), UAS-mCherry-RNAi (stock #35785: y [1] sc [*] v [1]; P{y [+t7.7] v [+t1.8]=VALIUM20-mCherry}attP2), Elav-Gal4; UAS-dicer2 (stock #25750: P{w [+mW.hs] = GawB}elav [C155] w [1118]; P{w [+mC] = UAS-Dcr-2.D}2), and R14H06-Gal4 (stock #48667: P{GMR14H06-GAL4}fattP2) [33]. G9aDD1 mutants and the precise excision control were generated previously [18]. To study the effects of trr down regulation, the UAS/Gal4 system was used with UAS-trr-RNAi combined with pan-neuronal (elav-Gal4) or mushroom body (MB) specific (R14H06-Gal4) Gal4 driver lines. For all trr knockdown experiments, genetic control animals without the RNAi hairpin were generated using the (AttP2) genetic background control strain (Bloomington #36303) or the mCherry-RNAi strain (Bloomington #35785). For trr ChIP-seq experiments, a Nijmegen wild type strain was used.

Imaging of mushroom bodies and immunohistochemistry

R14H06-Gal4 combined with UAS-mCD8::GFP was used to visualize the calyx region of the *Drosophila* MB in controls and trr knockdown flies. The following genotypes were analyzed: UAS-mCD8::GFP/+;R14H06-Gal4/UAS-trr-RNAi (trr knockdown) and UAS-mCD8::GFP/+;R14H06-Gal4/+ (control). Adult brains were dissected in PBS (pH 7.2) and fixed with ice cold methanol for 2 minutes, washed three times with PBS for 5 minutes each and permeabilized in PBS with 0.3% Triton-X 100 (PBT) for 1 hour. Fixed and permeabilized brains were blocked (5% normal goat serum (NGS)) at room temperature for 2 hours and incubated in 300nM DAPI solution for three minutes, followed by three five minute washes in PBT. Next, brains were incubated with the primary antibody (rabbit anti-trr [24], 1:5000) in PBT with 5% NGS for 72 hour at 4°C followed by five 20 minute washes in PBT. Brains were incubated with the secondary antibody (goat anti-rabbit Alexa Fluor 568, 1:250; Invitrogen) for 48 hour in PBT with 5% NGS at 4°C and washed at room temperature in PBT five times for 20 minutes. Brains were mounted in Vectashield (Vector Laboratories) and imaged on a confocal microscope (Zeiss LSM 510 duo vario confocal microscope). Confocal stacks were processed using ImageJ software [64].

Courtship Conditioning Assay

Courtship conditioning assays were performed on 5-day-old males raised at 25 °C, 70% humidity, and a 12h day/night rhythm as previously described [18, 35]. The following genotypes were analyzed: R14H06-Gal4/UAS-mCherry-RNAi (control) and R14H06-Gal4/

UAS-trr-RNAi (trr-RNAi). These genotypes were generated by crossing females containing the trr-RNAi or mCherry-RNAi inserted into the TRiP attP2 genetic background (genotypes: trr RNAi - y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.HMS01019}, mCherry RNAi - y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8]=VALIUM20-mCherry}attP2), with identical males containing the Gal4 driver (genotype: P{GMR14H06-GAL4}fattP2. Males were randomly assigned to either trained ($n_{control}$ = 59, $n_{trr\text{-}RNAi}$= 62) or naïve ($n_{control}$ =56, $n_{trr\text{-}RNAi}$ = 60) groups. Training was performed by pairing individual males in with 5-day-old mated wild type females. All experiments were conducted with a one-hour training period, and tested after a one-hour isolation period. For each fly pair, a courtship index (CI) was calculated as the percentage of a 10 minute time period spent courting. Courtship behaviour was manually quantified by observing videos for all 237 fly pairs analyzed in this study. Quantification was performed by trained observers that were naïve to the nature of the experiment. Comparisons of average CI between naïve and trained groups of the same genotype were calculated using a Mann-Whitney test. A learning index (LI = (CInaive-CItrained)/CInaive) was calculated to compare courtship memory between genotypes. Statistical comparisons between genotypes were conducted using a randomization test [65] with a custom bootstrapping script created in R (10,000 replicates) (**chapter 2**).

ChIP-seq and identification of trr binding sites

Chip was performed using an antibody directed against trr [14, 34]. This antibody was originally validated by the lack of staining in trr homozygous null mutant embryos [14]. Additionally, this antibody shows complete lack of signal in mutant clones of the *Drosophila* salivary gland, and has been used for ChIP experiments showing very specific trr binding patterns at EcR response genes during ecdysone-mediated developmental transitions [34]. Here, we provide additional evidence for specificitiy of this antibody by showing reduced immunofluorescent staining in mushroom body target cells expressing a trr-RNAi construct (**Figure 2**).

Chromatin was extracted from 50 µL aliquots of frozen wildtype fly heads, aged between 0 and 5 days old, in biological duplicates. Fly heads were crushed in PBS (Sigma) and crosslinked with formaldehyde (Sigma) at a final concentration of 1% for 30 minutes. Crosslinking was terminated using glycine (Invitrogen) at a final concentration of 125mM and crosslinked material was immediately washed twice in PBS and centrifuged at 13000rpm, for 15 minutes at 4 degrees. The pellet was resuspended in buffer 1 (15 mM Tris-HCl (pH 7.5), 60 mM KCl, 15 mM NaCl, 1 mM EDTA, 0.1 mM EGTA, 0.15 mM spermine, 0.5 mM spermidine, 0.1 mM sucrose) and the solution was homogenized using a QiaShredder column (Qiagen). Cells were lysed using buffer 1 supplemented by 2% triton-X-100 (Sigma) and a crude nuclear extract was collected by centrifugation (6000 rpm, 10 minutes at 4 degrees). Nuclei were re-suspended in incubation buffer (0.15% SDS, 1% triton x-100, 150

mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris) supplemented with 0.1% BSA (Sigma) and 1x protease inhibitor (Roche) and subjected to sonication (Diagenode Bioruptor) for 30 minutes (30 seconds on/off cycle using the "high intensity" mode), yielding average DNA fragments of 150-300 base pairs. Immunoprecipitation reactions were performed overnight in incubation buffer with 3 ng of anti-trr antibody (gift from Dr. A. Mazo [14]), protease inhibitor cocktail (Roche), BSA, and pre-blocked protein A/G agarose beads (Santa Cruz). Chromatin-antibody-bead complexes were recovered by centrifugation (4000 rpm, 2 minutes at four degrees) and washed twice with low salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris pH 8, 150 mM NaCl), once with high salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris pH 8, 500 mM NaCl), once with LiCl wash buffer (10 mM Tris pH 8.0, 1% Na-deoxycholate, 1% NP-40, 250 mM LiCl, 1 mM EDTA) and twice with TE buffer. Chromatin was eluted in 1% SDS, 0.1M NaHCO3, 200 mM NaCl and de-crosslinked at 65° Celsius (C) for four hours. DNA was purified by phenol/chloroform extraction and ethanol precipitated using linear acrylamide (Ambion) and sodium acetate at -20 degrees. Library preparation for Illumina sequencing was performed using the Truseq DNA sample prep kit V2 (Illumina) with approximately 3 ng of starting DNA 15 PCR cycles for amplification. Library fragment size was assessed using the 2100 Agilent Bioanalyser and was shown to be between 200 and 400 bp. Cluster generation and sequencing-by-synthesis (50bp) was performed using the Illumina Hiseq 2000 according to standard protocols of the manufacturer. The image files generated by the HiSeq were processed to extract DNA sequence data. The raw sequence data from this study are available at the NCBI Gene Expression omnibus series (accession number GSE89459). We obtained between 30 and 54 million reads per sample. Reads were mapped to the *Drosophila* genome (BDGP R5/dm3) using the Burrows Wheeler Aligner (BWA, version 0.6.1) with standard settings allowing 1 mismatch [66]. Total alignment efficiency was more than 95%. Duplicate reads and reads with a mapping quality score (MAPQ) below 15 were excluded from downstream analysis (**Table S1**). For each biological replicate, trr binding sites were identified using MACS2 [36] with input DNA as control (**Table S1**). Bindings sites on chromosome U, Uextra and mitochondrial genome were not used for further analysis. Remaining putative trr binding regions were visualized in a heatmap that was sorted based on k-means clustering of read intensity around the centre of the peak using the python script fluff_heatmap.py (https://github.com/simonvh/fluff) (**Figure S3A**). Individual clusters were examined visually in the genome browser to assess the quality of the peaks within the clusters. Two clusters (8 and 15) of replicate 1 were identified that contained many false positive peaks with a relatively small binding region and cluster 15 had very few reads around the centre. Peaks within these clusters were removed from downstream analysis. For trr_rep2 all clusters appeared to represent high quality trr binding sites. Trr binding regions from both biological replicates were merged and concatenated and the number of reads present in these regions were

counted in each sample individually using HTSeq [67] and normalized to library size. The ratio and mean of reads in each binding region was calculated between the two biological duplicates to identify trr binding sites that were consistent between the two biological replicates. We included all binding sites with a mean number of reads > 100 and < 2 fold difference in normalized read count between the two biological replicates (**Figure S3B**). Using these criteria we identified 3371 predicted trr binding regions, which appeared to be very consistent upon visual inspection in the UCSC genome browser (**Figure S3D**). These predicted trr binding sites were allocated to the nearest genomic feature using the pearl script annotatePeaks.pl [68] with standard settings. In order to find out if detected binding sites of trr are overrepresented in any genomic regions, we compared trr binding sites to an equal set of randomly selected genomic positions using https://www.random.org/ **(Figure 3B)**. The average trr occupancy profile over all transcription start sites (Figure 3C) was calculated using ensemble annotations from release 84 [69] with read count data generated by HTseq-count [67] in 50-bp bins spanning all transcription start sites.

To validate the ChIP-seq data, independent ChIP reactions were performed as described above and tested by qPCR. Validation targets were selected based on ChIP-seq data with representative binding regions on the promoters of *mor*, *lis-1*, *Atg9*, *Hsc70-4*, *Socs36e*, *Tsp42ed*, *smid* and *med21*. Additionally, we selected two trr-negative regions in the promoters of *drm99B* and *CG1646* (**Table S10**). All primers were tested and approved for amplification efficiency according to standard methods. qPCR was performed on the ChIPed and input samples using SYBRgreen master mix (Promega) and the 7900HT Fast Real Time PCR system (Applied biosystems) according to the manufacturer's instructions. Fold enrichments per target were calculated using the mean of the percent input of the two negative regions relative to the positive region. Mean fold enrichments are plotted with standard error of the mean as error bars (**Figure S3C**). All regions tested confirmed the ChIP-seq results. In addition we compared our trr ChIP-seq targets to published trr ChIP-seq data from cultured *Drosophila* S2 cells that was obtained using a different trr antibody [25]. In S2 cells, 1482 genes were identified with a trr peak at the promoter [25]. 943 of these genes were also identified as trr targets in this study. This is a very high overlap considering the vastly different starting material (cultured S2 cells versus fly heads), suggesting a high degree of concordance between the two antibodies.

Transcriptional profiling

Mutant and control lines for trr and G9a were aged between 1 and 5 days and snap frozen in liquid nitrogen. Frozen fly heads were harvested by vortexing and separated from other body parts through a series of standard laboratory sieves. Total RNA was extracted from 50 µL aliquots of frozen fly heads using the RNAeasy lipid tissue mini kit (Qiagen). For trr knockdown, two biological replicates were used, for G9a mutants, three biological replicates

were used for each condition. mRNA was purified using the Oligotex kit (Qiagen) and cDNA was synthesized using the SuperScript III First Strand synthesis kit (Thermo Fisher) using random hexamers as primers. Second strand cDNA was synthesized using E. Coli polymerase and T4 ligase (New England Biolabs Inc. (NEB)). Remaining RNA was removed using 2 units RNaseH (NEB) before the cDNA was purified using the MinElute PCR purification kit (Qiagen). DNA end repair was performed followed by ligation of Illumina sequencing adaptors and size selection for 300 bp by 2% E-gel (Invitrogen). Fragments were linearly amplified (15 PCR cycles), as validated by quantitative real-time PCR (qPCR), and sample quality was assessed using the Agilent Bioanalyser 2100. Cluster generation and sequencing-by-synthesis (36bp) was performed using the Illumina Genome Analyser IIx (GAIIx) according to standard protocols of the manufacturer. The image files generated by the GAIIx were processed to extract DNA sequence data. The raw sequence data from this study are available at the NCBI Gene Expression omnibus series (accession number GSE89459). From the GAIIx, we obtained between 27 and 35 million reads. Reads were mapped to the *Drosophila* genome (BDGP R5/dm3) using the Burrows Wheeler Aligner (BWA, version 0.6.1) with standard settings allowing 1 mismatch [66]. Only the uniquely mapped reads were used for further analysis and total alignment efficiency was between 63% and 79% for G9a samples and between 79% and 81% for trr samples (**Table S5**). Total read count data was generated by the python script HTSeq-count (http://www-huber.embl.de/HTSeq/doc/overview.html) with gene annotations extracted from the file *Drosophila_melanogaster*.BDGP5.75.gtf, available at http://www ensembl.org. In all samples, over 96% of aligned reads, mapped unambiguously to exons. The unambiguously mapped reads, ranging from 18 to 25 million reads for G9a samples and from 26 to 28 million reads for trr samples, were used for further analysis of differential gene expression by DeSeq2 [70]. Hierarchical clustering based on Euclidean distances with Pearson correlation using the normalized expression values and variance stabilizing transformation, revealing a high degree of similarity between biological replicates (**Figure S4A and S4C**). In order to perform statistical comparisons, dispersion values were estimated using the DESEQ method. As expected, we observed a high degree of correlation between gene expression and dispersion values with decreasing dispersion upon increasing expression patterns (**Figure S4B and S4D**). We then used DESeq2 to identify genes that are differentially expressed in mutant fly heads based on negative binominal distribution (adjusted p-value < 0.05, fold change > 1.5, **Figure 4A and C, Table S6 and S7**).

Validation of trr knockdown was performed using using SYBRgreen master mix (Promega) and the 7900HT Fast Real Time PCR system (Applied Biosystems) according to the manufacturer's instructions on a new biological replicate cDNA (described above) performed in triplicate technical replicates. Beta-cop and RP49 were used as reference genes for normalization and calculation of fold change differences upon pan-neuronal

knockdown in fly heads using the ΔΔ Ct procedure (**Figure S1**). All RT-qPCR primers (**Table S10**) are validated for amplification efficiency according to standard procedures.

Gene Ontology enrichment and hypergeometric analysis

Gene Ontology (GO) enrichment analysis was performed using the Panther software version 11.1 [52] on http://geneontology.org/ (GO Ontology database released on 2015-08-06 and 2016-10-24) using the GO-SLIM function for GO enrichment datasets that showed a high degree of redundancy (**Figure 4 and Figure S2**). For GO enrichment analysis of differential gene expression, list was used with genes expressed in fly heads as background control. This list was generated by exclusion of 4579 genes that had less than 10 reads, leaving 11103 genes. Overlap between datasets was determined and visualized as a Venn diagram by BioVenn [71] (**Figure 3E and 4E**). Hypergeometric statistics on overlaps were calculated using https://www.geneprof.org/GeneProf/tools/hypergeometric.jsp.

Ethics Statement

All subjects or their legal guardians gave written informed consent. This study was approved by the institutional review boards of the Radboudumc or Maastricht University medical center.

Acknowledgements

We thank all families for their contribution. *Drosophila* stocks were obtained from the Bloomington *Drosophila* Stock Center (NIH P40OD018537). We thank Eva Janssen-Megens and Henk Stunnenberg for help with sequencing and Dr. Alex Mazo for providing the trr antibody.

Accession numbers

The raw data for RNA-seq and ChIP-seq is available at the NCBI Gene Expression omnibus (GEO), accession number GSE89459.

References

1. Kleefstra T, Smidt M, Banning MJ, Oudakker AR, Van Esch H, de Brouwer AP, et al. Disruption of the gene Euchromatin Histone Methyl Transferase1 (Eu-HMTase1) is associated with the 9q34 subtelomeric deletion syndrome. Journal of medical genetics. 2005;42(4):299-306. doi: 10.1136/jmg.2004.028464. PubMed PMID: 15805155; PubMed Central PMCID: PMC1736026.
2. Kleefstra T, Brunner HG, Amiel J, Oudakker AR, Nillesen WM, Magee A, et al. Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. American journal of human genetics. 2006;79(2):370-7. doi: 10.1086/505693. PubMed PMID: 16826528; PubMed Central PMCID: PMC1559478.
3. Benevento M, van de Molengraft M, van Westen R, van Bokhoven H, Kasri NN. The role of chromatin repressive marks in cognition and disease: A focus on the repressive complex GLP/G9a. Neurobiology of learning and memory. 2015;124:88-96. doi: 10.1016/j.nlm.2015.06.013. PubMed PMID: 26143996.
4. Kramer JM. Regulation of cell differentiation and function by the euchromatin histone methyltranserfases G9a and GLP. Biochemistry and cell biology = Biochimie et biologie cellulaire. 2016;94(1):26-32. doi: 10.1139/bcb-2015-0017. PubMed PMID: 26198080.
5. Stewart DR, Kleefstra T. The chromosome 9q subtelomere deletion syndrome. American journal of medical genetics Part C, Seminars in medical genetics. 2007;145C(4):383-92. doi: 10.1002/ajmg.c.30148. PubMed PMID: 17910072.
6. Willemsen MH, Vulto-van Silfhout AT, Nillesen WM, Wissink-Lindhout WM, van Bokhoven H, Philip N, et al. Update on Kleefstra Syndrome. Molecular syndromology. 2012;2(3-5):202-12. doi: 000335648. PubMed PMID: 22670141; PubMed Central PMCID: PMC3366700.
7. Tan WH, Bird LM, Thibert RL, Williams CA. If not Angelman, what is it? A review of Angelman-like syndromes. American journal of medical genetics Part A. 2014;164A(4):975-92. PubMed PMID: 24779060.
8. Mullegama SV, Alaimo JT, Chen L, Elsea SH. Phenotypic and molecular convergence of 2q23.1 deletion syndrome with other neurodevelopmental syndromes associated with autism spectrum disorder. International journal of molecular sciences. 2015;16(4):7627-43. doi: 10.3390/ijms16047627. PubMed PMID: 25853262; PubMed Central PMCID: PMC4425039.
9. Kleefstra T, Kramer JM, Neveling K, Willemsen MH, Koemans TS, Vissers LE, et al. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. American journal of human genetics. 2012;91(1):73-82. doi: 10.1016/j.ajhg.2012.05.003. PubMed PMID: 22726846; PubMed Central PMCID: PMC3397275.
10. Lee S, Kim DH, Goo YH, Lee YC, Lee SK, Lee JW. Crucial roles for interactions between MLL3/4 and INI1 in nuclear receptor transactivation. Molecular endocrinology. 2009;23(5):610-9. doi: 10.1210/me.2008-0455. PubMed PMID: 19221051; PubMed Central PMCID: PMC2675954.
11. Choi E, Lee S, Yeom SY, Kim GH, Lee JW, Kim SW. Characterization of activating signal cointegrator-2 as a novel transcriptional coactivator of the xenobiotic nuclear receptor constitutive androstane receptor. Molecular endocrinology. 2005;19(7):1711-9. doi: 10.1210/me.2005-0066. PubMed PMID: 15764585.

12. Underhill C, Qutob MS, Yee SP, Torchia J. A novel nuclear receptor corepressor complex, N-CoR, contains components of the mammalian SWI/SNF complex and the corepressor KAP-1. The Journal of biological chemistry. 2000;275(51):40463-70. doi: 10.1074/jbc.M007864200. PubMed PMID: 11013263.
13. Jyrkkarinne J, Makinen J, Gynther J, Savolainen H, Poso A, Honkakoski P. Molecular determinants of steroid inhibition for the mouse constitutive androstane receptor. Journal of medicinal chemistry. 2003;46(22):4687-95. doi: 10.1021/jm030861t. PubMed PMID: 14561088.
14. Sedkov Y, Cho E, Petruk S, Cherbas L, Smith ST, Jones RS, et al. Methylation at lysine 4 of histone H3 in ecdysone-dependent development of *Drosophila*. Nature. 2003;426(6962):78-83. doi: 10.1038/nature02080. PubMed PMID: 14603321; PubMed Central PMCID: PMC2743927.
15. Lehnertz B, Northrop JP, Antignano F, Burrows K, Hadidi S, Mullaly SC, et al. Activating and inhibitory functions for the histone lysine methyltransferase G9a in T helper cell differentiation and function. The Journal of experimental medicine. 2010;207(5):915-22. doi: 10.1084/jem.20100363. PubMed PMID: 20421388; PubMed Central PMCID: PMC2867284.
16. Ohno H, Shinoda K, Ohyama K, Sharp LZ, Kajimura S. EHMT1 controls brown adipose cell fate and thermogenesis through the PRDM16 complex. Nature. 2013;504(7478):163-7. doi: 10.1038/nature12652. PubMed PMID: 24196706; PubMed Central PMCID: PMC3855638.
17. Seum C, Bontron S, Reo E, Delattre M, Spierer P. *Drosophila* G9a is a nonessential gene. Genetics. 2007;177(3):1955-7. doi: 10.1534/genetics.107.078220. PubMed PMID: 18039887; PubMed Central PMCID: PMC2147950.
18. Kramer JM, Kochinke K, Oortveld MA, Marks H, Kramer D, de Jong EK, et al. Epigenetic regulation of learning and memory by *Drosophila* EHMT/G9a. PLoS biology. 2011;9(1):e1000569. doi: 10.1371/journal.pbio.1000569. PubMed PMID: 21245904; PubMed Central PMCID: PMC3014924.
19. Shimaji K, Konishi T, Tanaka S, Yoshida H, Kato Y, Ohkawa Y, et al. Genomewide identification of target genes of histone methyltransferase dG9a during *Drosophila* embryogenesis. Genes to cells : devoted to molecular & cellular mechanisms. 2015;20(11):902-14. doi: 10.1111/gtc.12281. PubMed PMID: 26334932.
20. Merkling SH, Bronkhorst AW, Kramer JM, Overheul GJ, Schenck A, Van Rij RP. The epigenetic regulator G9a mediates tolerance to RNA virus infection in *Drosophila*. PLoS pathogens. 2015;11(4):e1004692. doi: 10.1371/journal.ppat.1004692. PubMed PMID: 25880195; PubMed Central PMCID: PMC4399909.
21. Balemans MC, Ansar M, Oudakker AR, van Caam AP, Bakker B, Vitters EL, et al. Reduced Euchromatin histone methyltransferase 1 causes developmental delay, hypotonia, and cranial abnormalities associated with increased bone gene expression in Kleefstra syndrome mice. Developmental biology. 2014;386(2):395-407. doi: 10.1016/j.ydbio.2013.12.016. PubMed PMID: 24362066.
22. Balemans MC, Huibers MM, Eikelenboom NW, Kuipers AJ, van Summeren RC, Pijpers MM, et al. Reduced exploration, increased anxiety, and altered social behavior: Autistic-like features of euchromatin histone methyltransferase 1 heterozygous knockout mice. Behavioural brain research. 2010;208(1):47-55. doi: 10.1016/j.bbr.2009.11.008. PubMed PMID: 19896504.
23. Balemans MC, Kasri NN, Kopanitsa MV, Afinowi NO, Ramakers G, Peters TA, et al. Hippocampal dysfunction in the Euchromatin histone methyltransferase 1 heterozygous knockout mouse model for Kleefstra syndrome. Human molecular genetics. 2013;22(5):852-66. doi: 10.1093/hmg/dds490. PubMed PMID: 23175442.

24. Schaefer A, Sampath SC, Intrator A, Min A, Gertler TS, Surmeier DJ, et al. Control of cognition and adaptive behavior by the GLP/G9a epigenetic suppressor complex. Neuron. 2009;64(5):678-91. doi: 10.1016/j.neuron.2009.11.019. PubMed PMID: 20005824; PubMed Central PMCID: PMC2814156.
25. Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, et al. Enhancer-associated H3K4 monomethylation by Trithorax-related, the *Drosophila* homolog of mammalian Mll3/Mll4. Genes & development. 2012;26(23):2604-20. doi: 10.1101/gad.201327.112. PubMed PMID: 23166019; PubMed Central PMCID: PMC3521626.
26. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823-37. doi: 10.1016/j.cell.2007.05.009. PubMed PMID: 17512414.
27. Hu D, Gao X, Morgan MA, Herz HM, Smith ER, Shilatifard A. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Molecular and cellular biology. 2013;33(23):4745-54. doi: 10.1128/MCB.01181-13. PubMed PMID: 24081332; PubMed Central PMCID: PMC3838007.
28. Mohan M, Herz HM, Smith ER, Zhang Y, Jackson J, Washburn MP, et al. The COMPASS family of H3K4 methylases in *Drosophila*. Molecular and cellular biology. 2011;31(21):4310-8. doi: 10.1128/MCB.06092-11. PubMed PMID: 21875999; PubMed Central PMCID: PMC3209330.
29. de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. The New England journal of medicine. 2012;367(20):1921-9. doi: 10.1056/NEJMoa1206524. PubMed PMID: 23033978.
30. Sedkov Y, Benes JJ, Berger JR, Riker KM, Tillib S, Jones RS, et al. Molecular genetic analysis of the *Drosophila* trithorax-related gene which encodes a novel SET domain protein. Mechanisms of development. 1999;82(1-2):171-9. PubMed PMID: 10354481.
31. Brand AH, Manoukian AS, Perrimon N. Ectopic expression in *Drosophila*. Methods in cell biology. 1994;44:635-54. PubMed PMID: 7707973.
32. Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, et al. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. Nature. 2007;448(7150):151-6. doi: 10.1038/nature05954. PubMed PMID: 17625558.
33. Jenett A, Rubin GM, Ngo TT, Shepherd D, Murphy C, Dionne H, et al. A GAL4-driver line resource for *Drosophila* neurobiology. Cell reports. 2012;2(4):991-1001. doi: 10.1016/j.celrep.2012.09.011. PubMed PMID: 23063364; PubMed Central PMCID: PMC3515021.
34. Johnston DM, Sedkov Y, Petruk S, Riley KM, Fujioka M, Jaynes JB, et al. Ecdysone- and NO-mediated gene regulation by competing EcR/Usp and E75A nuclear receptors during *Drosophila* development. Molecular cell. 2011;44(1):51-61. doi: 10.1016/j.molcel.2011.07.033. PubMed PMID: 21981918; PubMed Central PMCID: PMC3190167.
35. Siegel RW, Hall JC. Conditioned responses in courtship behavior of normal and mutant *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America. 1979;76(7):3430-4. PubMed PMID: 16592682; PubMed Central PMCID: PMC383839.
36. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome biology. 2008;9(9):R137. doi: 10.1186/gb-2008-9-9-r137. PubMed PMID: 18798982; PubMed Central PMCID: PMC2592715.

37. Rauen KA. The RASopathies. Annual review of genomics and human genetics. 2013;14:355-69. doi: 10.1146/annurev-genom-091212-153523. PubMed PMID: 23875798; PubMed Central PMCID: PMC4115674.
38. Kochinke K, Zweier C, Nijhof B, Fenckova M, Cizek P, Honti F, et al. Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. American journal of human genetics. 2016;98(1):149-64. doi: 10.1016/j.ajhg.2015.11.024. PubMed PMID: 26748517; PubMed Central PMCID: PMC4716705.
39. Bjornsson HT. The Mendelian disorders of the epigenetic machinery. Genome research. 2015;25(10):1473-81. doi: 10.1101/gr.190629.115. PubMed PMID: 26430157; PubMed Central PMCID: PMC4579332.
40. Kleefstra T, Schenck A, Kramer JM, van Bokhoven H. The genetics of cognitive epigenetics. Neuropharmacology. 2014;80:83-94. doi: 10.1016/j.neuropharm.2013.12.025. PubMed PMID: 24434855.
41. Ishimoto H, Sakai T, Kitamoto T. Ecdysone signaling regulates the formation of long-term courtship memory in adult *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(15):6381-6. doi: 10.1073/pnas.0810213106. PubMed PMID: 19342482; PubMed Central PMCID: PMC2669368.
42. Gupta S, Kim SY, Artis S, Molfese DL, Schumacher A, Sweatt JD, et al. Histone methylation regulates memory formation. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2010;30(10):3589-99. doi: 10.1523/JNEUROSCI.3732-09.2010. PubMed PMID: 20219993; PubMed Central PMCID: PMC2859898.
43. Gupta-Agarwal S, Jarome TJ, Fernandez J, Lubin FD. NMDA receptor- and ERK-dependent histone methylation changes in the lateral amygdala bidirectionally regulate fear memory formation. Learning & memory. 2014;21(7):351-62. doi: 10.1101/lm.035105.114. PubMed PMID: 24939839; PubMed Central PMCID: PMC4061426.
44. Yamazaki D, Horiuchi J, Ueno K, Ueno T, Saeki S, Matsuno M, et al. Glial dysfunction causes age-related memory impairment in *Drosophila*. Neuron. 2014;84(4):753-63. doi: 10.1016/j.neuron.2014.09.039. PubMed PMID: 25447741.
45. Roberti M, Bruni F, Polosa PL, Gadaleta MN, Cantatore P. The *Drosophila* termination factor DmTTF regulates in vivo mitochondrial transcription. Nucleic acids research. 2006;34(7):2109-16. doi: 10.1093/nar/gkl181. PubMed PMID: 16648357; PubMed Central PMCID: PMC1450328.
46. Van Gelder RN, Bae H, Palazzolo MJ, Krasnow MA. Extent and character of circadian gene expression in *Drosophila melanogaster*: identification of twenty oscillating mRNAs in the fly head. Current biology : CB. 1995;5(12):1424-36. PubMed PMID: 8749395.
47. Carhan A, Tang K, Shirras CA, Shirras AD, Isaac RE. Loss of Angiotensin-converting enzyme-related (ACER) peptidase disrupts night-time sleep in adult *Drosophila melanogaster*. The Journal of experimental biology. 2011;214(Pt 4):680-6. doi: 10.1242/jeb.049353. PubMed PMID: 21270318.
48. Krishnan HC, Lyons LC. Synchrony and desynchrony in circadian clocks: impacts on learning and memory. Learning & memory. 2015;22(9):426-37. doi: 10.1101/lm.038877.115. PubMed PMID: 26286653; PubMed Central PMCID: PMC4561405.
49. Gerstner JR, Yin JC. Circadian rhythms and memory formation. Nature reviews Neuroscience. 2010;11(8):577-88. doi: 10.1038/nrn2881. PubMed PMID: 20648063.

50. Daberkow DP, Riedy MD, Kesner RP, Keefe KA. Arc mRNA induction in striatal efferent neurons associated with response learning. The European journal of neuroscience. 2007;26(1):228-41. doi: 10.1111/j.1460-9568.2007.05630.x. PubMed PMID: 17614950.
51. Korb E, Finkbeiner S. Arc in synaptic plasticity: from gene to behavior. Trends in neurosciences. 2011;34(11):591-8. doi: 10.1016/j.tins.2011.08.007. PubMed PMID: 21963089; PubMed Central PMCID: PMC3207967.
52. Benevento M, Iacono G, Selten M, Ba W, Oudakker A, Frega M, et al. Histone Methylation by the Kleefstra Syndrome Protein EHMT1 Mediates Homeostatic Synaptic Scaling. Neuron. 2016;91(2):341-55. doi: 10.1016/j.neuron.2016.06.003. PubMed PMID: 27373831.
53. Chen ES, Gigek CO, Rosenfeld JA, Diallo AB, Maussion G, Chen GG, et al. Molecular convergence of neurodevelopmental disorders. American journal of human genetics. 2014;95(5):490-508. doi: 10.1016/j.ajhg.2014.09.013. PubMed PMID: 25307298; PubMed Central PMCID: PMC4225591.
54. Zweier C, Peippo MM, Hoyer J, Sousa S, Bottani A, Clayton-Smith J, et al. Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). American journal of human genetics. 2007;80(5):994-1001. doi: 10.1086/515583. PubMed PMID: 17436255; PubMed Central PMCID: PMCPMC1852727.
55. Elsea SH, Girirajan S. Smith-Magenis syndrome. European journal of human genetics : EJHG. 2008;16(4):412-21. doi: 10.1038/sj.ejhg.5202009. PubMed PMID: 18231123.
56. Smeets EE, Pelc K, Dan B. Rett Syndrome. Molecular syndromology. 2012;2(3-5):113-27. doi: 000337637. PubMed PMID: 22670134; PubMed Central PMCID: PMCPMC3366703.
57. Talkowski ME, Mullegama SV, Rosenfeld JA, van Bon BW, Shen Y, Repnikova EA, et al. Assessment of 2q23.1 microdeletion syndrome implicates MBD5 as a single causal locus of intellectual disability, epilepsy, and autism spectrum disorder. American journal of human genetics. 2011;89(4):551-63. doi: 10.1016/j.ajhg.2011.09.011. PubMed PMID: 21981781; PubMed Central PMCID: PMC3188839.
58. Chung BH, Mullegama S, Marshall CR, Lionel AC, Weksberg R, Dupuis L, et al. Severe intellectual disability and autistic features associated with microduplication 2q23.1. European journal of human genetics : EJHG. 2012;20(4):398-403. doi: 10.1038/ejhg.2011.199. PubMed PMID: 22085900; PubMed Central PMCID: PMC3306850.
59. Bogershausen N, Wollnik B. Unmasking Kabuki syndrome. Clinical genetics. 2013;83(3):201-11. doi: 10.1111/cge.12051. PubMed PMID: 23131014.
60. Kyllerman M. Angelman syndrome. Handb Clin Neurol. 2013;111:287-90. doi: 10.1016/B978-0-444-52891-9.00032-4. PubMed PMID: 23622177.
61. Kim HG, Kim HT, Leach NT, Lan F, Ullmann R, Silahtaroglu A, et al. Translocations disrupting PHF21A in the Potocki-Shaffer-syndrome region are associated with intellectual disability and craniofacial anomalies. American journal of human genetics. 2012;91(1):56-72. doi: 10.1016/j.ajhg.2012.05.005. PubMed PMID: 22770980; PubMed Central PMCID: PMC3397276.
62. Montgomery ND, Turcott CM, Tepperberg JH, McDonald MT, Aylsworth AS. A 137-kb deletion within the Potocki-Shaffer syndrome interval on chromosome 11p11.2 associated with developmental delay and hypotonia. American journal of medical genetics Part A. 2013;161A(1):198-202. doi: 10.1002/ajmg.a.35671. PubMed PMID: 23239541.
63. Kato T, Tamiya G, Koyama S, Nakamura T, Makino S, Arawaka S, et al. UBR5 Gene Mutation Is Associated with Familial Adult Myoclonic Epilepsy in a Japanese Family. ISRN neurology. 2012;2012:508308. doi: 10.5402/2012/508308. PubMed PMID: 23029623; PubMed Central PMCID: PMC3458293.

64. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nature methods. 2012;9(7):671-5. PubMed PMID: 22930834.
65. Kamyshev NG, Iliadi KG, Bragina JV. *Drosophila* conditioned courtship: two ways of testing memory. Learning & memory. 1999;6(1):1-20. PubMed PMID: 10355520; PubMed Central PMCID: PMC311276.
66. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-95. doi: 10.1093/bioinformatics/btp698. PubMed PMID: 20080505; PubMed Central PMCID: PMC2828108.
67. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-9. doi: 10.1093/bioinformatics/btu638. PubMed PMID: 25260700; PubMed Central PMCID: PMC4287950.
68. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell. 2010;38(4):576-89. doi: 10.1016/j.molcel.2010.05.004. PubMed PMID: 20513432; PubMed Central PMCID: PMC2898526.
69. Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, Pignatelli M, et al. Ensembl comparative genomics resources. Database : the journal of biological databases and curation. 2016;2016. doi: 10.1093/database/bav096. PubMed PMID: 26896847; PubMed Central PMCID: PMC4761110.
70. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8. PubMed PMID: 25516281; PubMed Central PMCID: PMC4302049.
71. Hulsen T, de Vlieg J, Alkema W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC genomics. 2008;9:488. doi: 10.1186/1471-2164-9-488. PubMed PMID: 18925949; PubMed Central PMCID: PMC2584113.

Supporting information

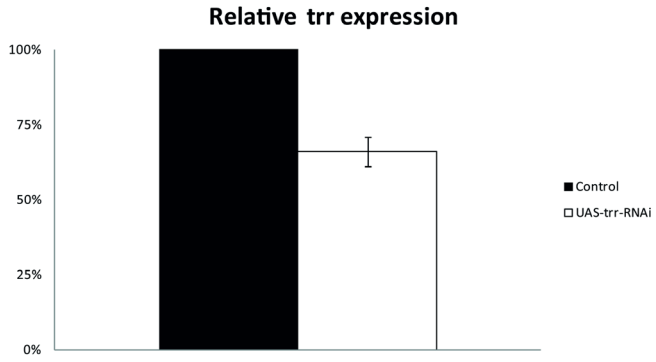


Figure S1: Pan-neuronal knockdown of trr in fly heads.

Knockdown as determined by RT-qPCR. Error bar represents standard deviation of mean knockdown relative to two reference genes.

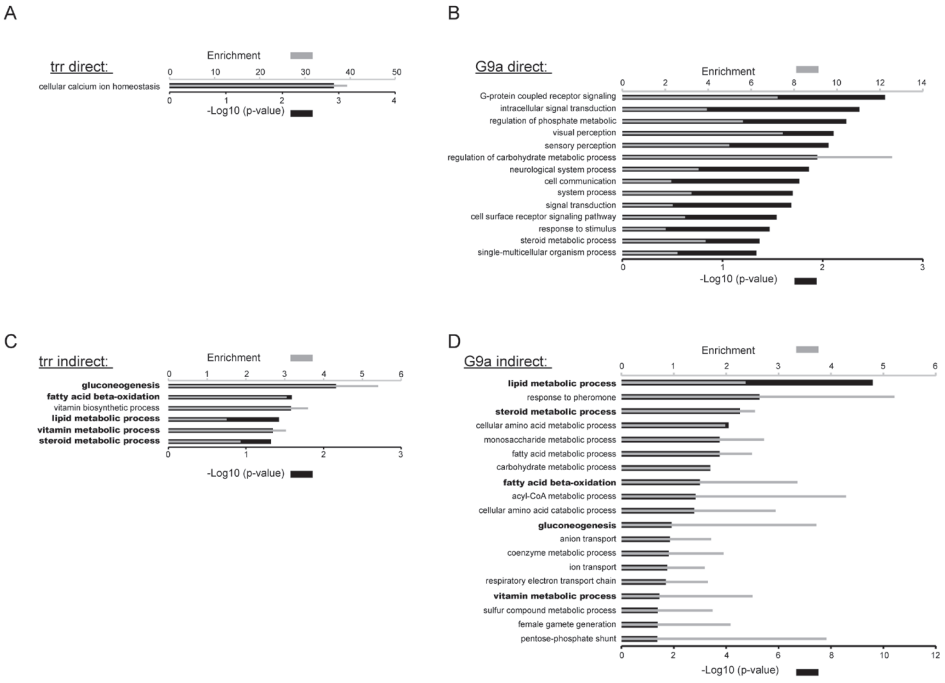


Figure S2: GO enrichment analysis of "potential direct" and "potential indirect" trr and G9a target genes.

(A-D) All enriched gene ontology (GO) terms identified using the Panther software (GO-slim setting) for potential direct trr (A) and G9a (B) target genes, and potential indirect and trr (C) and G9a (D) targets. Panther overrepresentation test from GO ontology database version 11.1 (released 2016-10-24). The reference list is a set of 11103 genes that show expression in our RNA-seq datasets (see materials and methods).

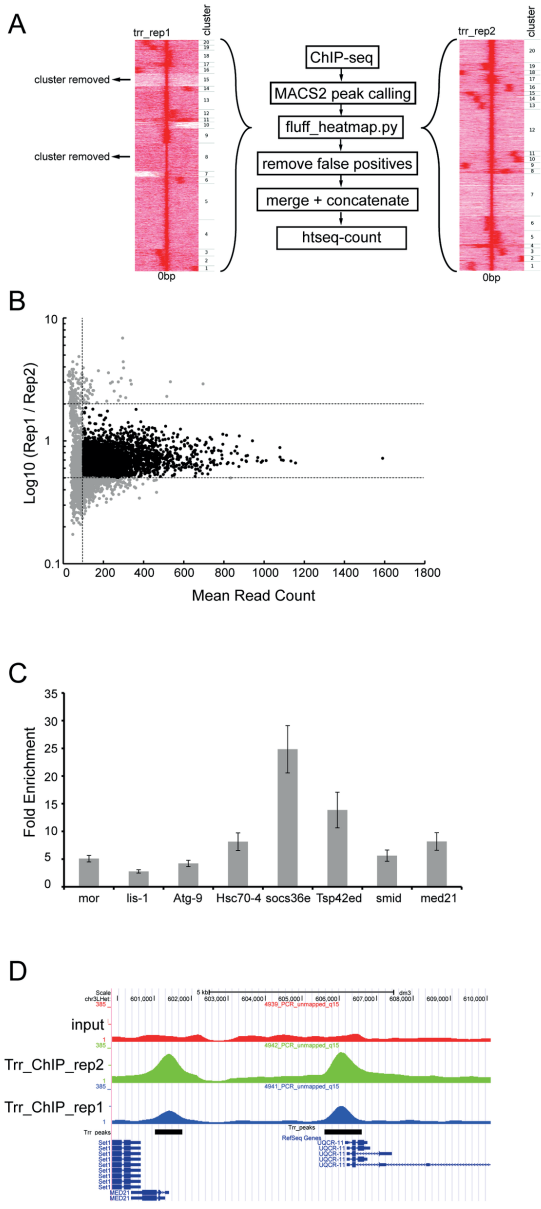


Figure S3: Filter criteria and validation of trr ChIP-seq.

(A) Flow chart describing the filtering steps for trr ChIP-seq starting with called peaks by MACS2 are quality assed by the python script fluff_heatmap.py (kmeans clustering). Regions of approved peaks are merged and concatenated and the number of reads are counted in each replicate individually. Average number of reads per peak in each replicate relative to the ratio of reads between replicates is plotted (B) and further quality assessed by visual inspection of the wiggle file in the genome browser. Black dots represents the 3371 high confident peaks with mean number of reads > 100 and ratio < 2 fold the number of reads in the replicate suggesting similar binding affinity (height). (C) Bar graph of ChIP-qPCR validation of targets identified by ChIP-seq are enriched over negative regions relative to the input. Error bars represent the standard error of the mean. (D) screenshot of the UCSC genome browser from two trr ChIP-seq replicates and input control at genomic loci identified as binding positions.

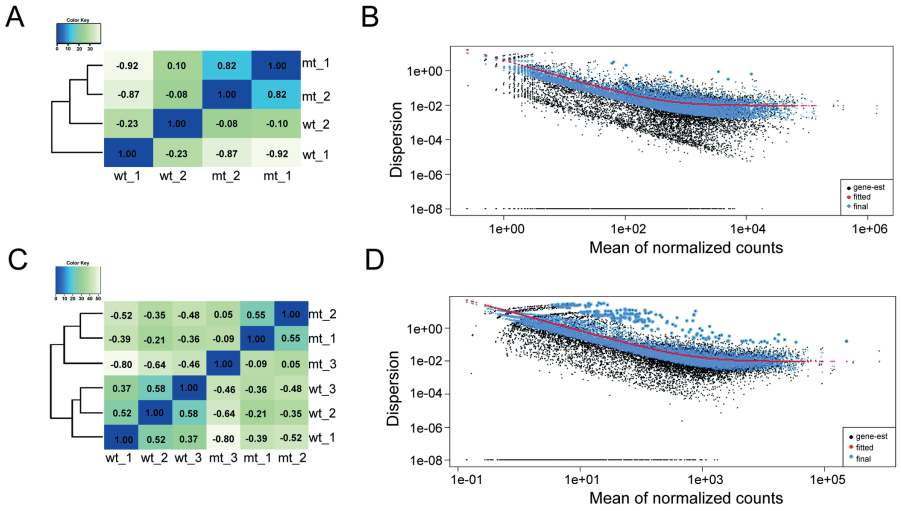


Figure S4: Validation of differential gene expression analysis.

(A,C) Dendrogram and heat map illustrating euclidean distances and Pearson correlation between samples of G9a- (A) and trr mutant (C). (B,D) Scatter plot showing dispersion estimates as determined using DESeq2, plotted against the mean of normalized reads for EMHT (B) and trr (D) RNA-seq datasets (black dots - gene-wise maximum-likelihood estimates, red dots - fitted values, blue dots - final dispersion. Genes with dispersion outliers were not used for further analysis. Note the decreasing dispersion values as the gene expression increases.

Supplemental tables

(available upon request)

Table S1: ChIP-seq depth, alignment, mapping efficiency and MACS2 settings. Total number of reads, and the percentage aligned reads are shown. Next, the percentage of aligned reads to unambiguous places relative to total aligned reads and the percentage of reads with MAPQ scores higher than 15 are shown. Lastly, the total number of high quality reads that was used for analysis is shown.

Table S2: Annotation of trr ChIP-seq peaks to nearest genomic feature using HOMER software (Raw data to Figure 3A).

Table S3: Gene ontology analysis of trr promoter associated genes (Raw data to Figure 3D). Gene Ontology analysis for trr associated genes. PANTHER overrepresentation test (release 20150430) from GO ontology database (released 2015-08-06). As reference list is all genes in database used.

Table S4: Gene ontology analysis of trr and H3K9me2 LOMB associated genes in the G9a mutant (Raw data to Figure 3F).
Gene Ontology analysis for genes associated with a trr binding site and a H3K9me2 LOMB. PANTHER overrepresentation test (release 20150430) from GO ontology database (released 2015-08-06). As reference list is all genes in database used.

Table S5: RNA-seq depth. Alignment and mapping efficiency of trr- and G9a mutant samples. Shown are the total number of reads, and the percentage aligned. Next, the percentage of aligned reads relative to the total number of aligned reads and unambiguous mapped reads are shown. Lastly, the total number of high quality reads that was used for analysis is shown.

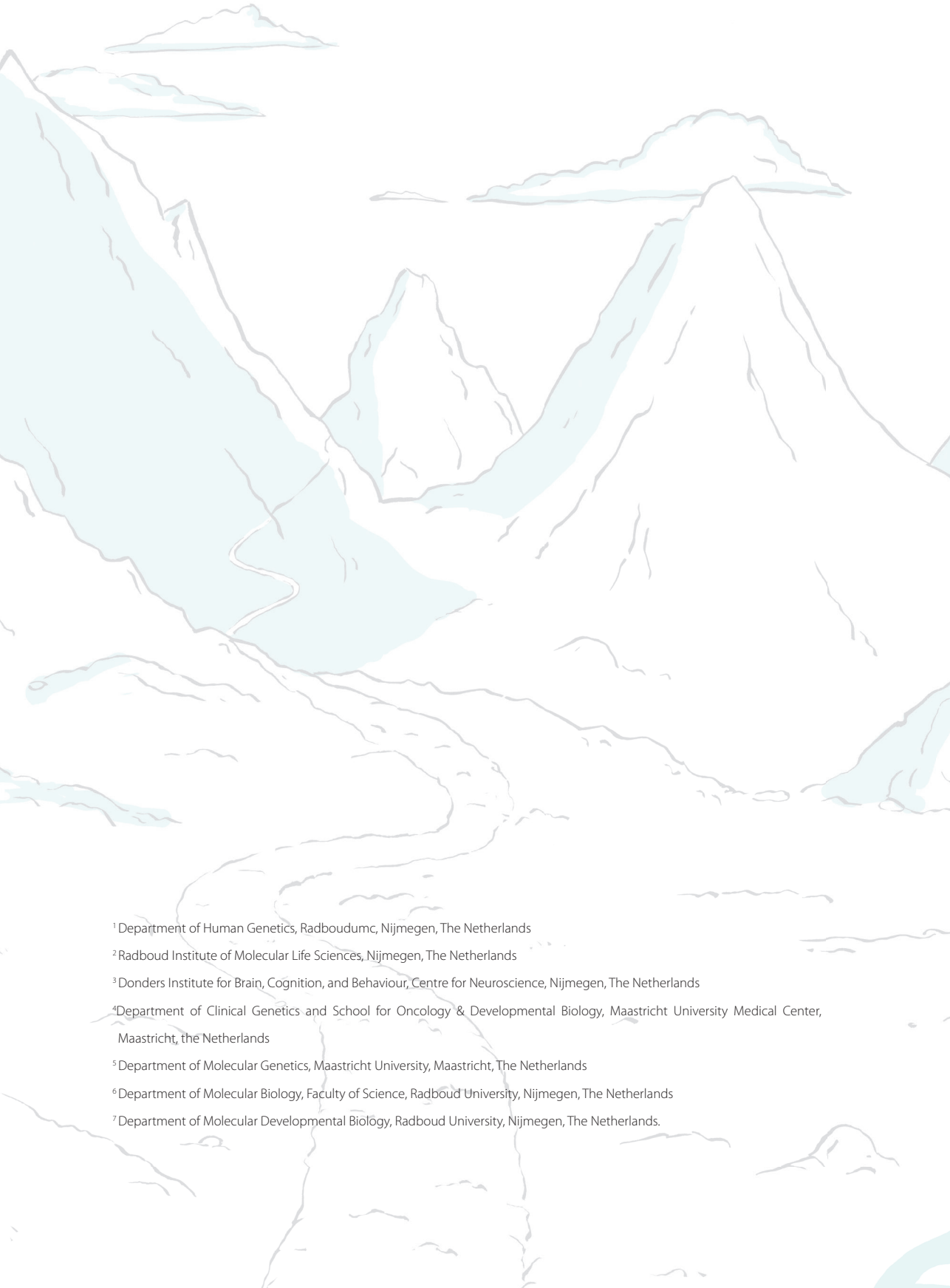
Table S6: Differential expressed genes in trr mutant (Raw data to Figure 4A).
All genes with altered expression in trr knockdown versus wildtype.

Table S7: Differential expressed genes in G9a mutant (Raw data to Figure 4C).
All genes with altered expression in G9a knockout versus wildtype.

Table S8: Statistical analysis of up and down regulated genes that are differentially expressed genes in both trr and G9a mutant fly heads.

Table S9: Gene ontology annotations for the five potential direct targets of both G9a and trr.

Table S10: List of primers used in this study.



[1] Department of Human Genetics, Radboudumc, Nijmegen, The Netherlands

[2] Radboud Institute of Molecular Life Sciences, Nijmegen, The Netherlands

[3] Donders Institute for Brain, Cognition, and Behaviour, Centre for Neuroscience, Nijmegen, The Netherlands

[4] Department of Clinical Genetics and School for Oncology & Developmental Biology, Maastricht University Medical Center, Maastricht, the Netherlands

[5] Department of Molecular Genetics, Maastricht University, Maastricht, The Netherlands

[6] Department of Molecular Biology, Faculty of Science, Radboud University, Nijmegen, The Netherlands

[7] Department of Molecular Developmental Biology, Radboud University, Nijmegen, The Netherlands.

Convergence of differentially expressed genes and their function in patients with Kleefstra- and Kabuki syndrome

Tom S. Koemans[1,2,3]
Ellen van Beusekom[1]
Constance Stumpel[4]
Willem Voncken[5]
Henk G. Stunnenberg[2,6]
Huiqing Zhou[1,2,7]
Tjitske Kleefstra[1]
Hans van Bokhoven[1,3]

4

Manuscript in preparation

Abstract

Kleefstra syndrome (KlS) and Kabuki syndrome (KaS) are rare intellectual disability (ID) disorders with unique characteristics. However, both syndromes also harbor common phenotypes such as ID, hypotonia, developmental delay, seizures and autism-like behaviour. Mutations in *EHMT1* and *KMT2D* are causative for KlS and KaS respectively, which encode broadly expressed histone H3 methyltransferases. Mutations in the genes encoding these enzymes are predicted to affect the epigenetic landscape in a variety of tissues. To investigate this, we analyzed the peripheral blood transcriptome of affected individuals. Comparison of genome wide mRNA expression profiles of five KlS, five KaS and five healthy control individuals revealed 595 differentially expressed genes in KlS patients and 1318 in KaS patients (1.5 fold change, p-value < 0.05). Strikingly, a significant overlap was observed between the identified misregulated genes of both syndromes. The direction (up- or downregulated) of these commonly misregulated genes was nearly identical. Gene ontology terms describing the misregulated genes were generated and revealed "cell-cell adhesion", "signaling" and "regulation of neuron differentiation" for KlS. For KaS, "nervous system development", "neurogenesis", and "signaling" were identified. When comparing all gene ontology terms, another striking overlap was observed between both disease entities on the functional level. Common terms include "central nervous system development", "brain development", and "cell surface signaling pathway". Taken together, these results support a shared etiology between KlS and KaS.

Introduction

Epigenetic control of developmental processes is a cellular mechanism by which spatial and temporal expression of distinct genes and pathways are regulated. Mutations in genes coding for epigenetic factors have been attributed to cause several syndromic disorders characterized by intellectual disability (ID) [1-3]. Kleefstra Syndrome (OMIM#610253, KlS) is one such disease caused by haploinsufficiency of *euchromatin histone methyl transferase 1* (*EHMT1*) [4, 5]. EHMT1 is involved in the mono- and dimethylation of histone H3 lysine (K) 9 [6-9]. KlS patients are characterized by ID, childhood hypotonia and facial characteristics [5, 10]. Additionally, other clinical features are described such as autism spectrum disorder, micro/brachycephaly, heart defects, seizures and difficulties in adaptive behaviour [11, 12]. Recently, additional mutations have been described in four genes (*MBD5*, *KMT2C*, *SMARCB1*, *NR1I3*) in individuals with Kleefstra syndrome phenotypic spectrum (KSS)[13]. Genetic interaction studies and protein-protein interactions between the KSS associated genes identified an epigenetic network underlying KlS and KSS [13]. Interestingly, a very strong antagonistic genetic interaction was identified between the *Drosophila* orthologs of *EHMT1* and *KMT2C/D*, called G9a and trithorax related (trr), respectively [13]. Additionally, a significant proportion of G9a and trr target genes and functions are overlapping (**chapter 3**) [14], suggesting that these proteins converge on similar biological processes.

Another example of an ID syndrome caused by mutations in an epigenetic factor is Kabuki syndrome (OMIM#147920, KaS) [15, 16]. Kabuki syndrome patients are characterized by postnatal dwarfism, distinct facial appearance, cardiac anomalies, skeletal abnormalities, immunological defects and mild to moderate ID [17]. Kabuki syndrome is associated with mutations in the histone methyltransferase *KMT2D* [15]. At the molecular level, KMT2D has been found as a component three unique protein complexes. First, KMT2D is a member of the complex of proteins associated with Set1 (COMPASS) which is a major contributor to mono- di- and trimethylation of H3K4 [18-20]. The COMPASS complex consists of several proteins that are conserved from yeast to human [21]. Human paralogs KMT2C and KMT2D are conserved as trr in *Drosophila* which is responsible for monomethylation of H3K4 at enhancer associated sites [22] and H3K4me3 in a developmental context [23, 24]. Second, KMT2C and KMT2D are part of the ASC-2 complex (ASCOM) [25, 26] which has been shown to regulate transcription by trimethylation of H3K4 [27]. Third, KMT2C and KMT2D have been associated with Pax transactivation domain-interacting protein (PTIP) [28]. This protein has been shown to be involved in the DNA damage response by interaction with 53BP1 [29]. In addition, PTIP is associated with histone H3K4 trimethylation as part of a complex consisting of the structural and functional proteins PA1, ASH2L, RBBP5, WDR5, hDPY-30, NCOA6, KMT2C, KMT2D, and UTX [28]. The main differences between ASCOM and PTIP containing complexes is that hDPY-30, PA1, and UTX are only present together with PTIP [28, 30]. Second, α- and β-tubulins are only present in ASCOM and, third, the

PTIP containing complex shows more robust methyltransferase activity [28, 31]. It should be noted that mutations in *UTX* (*KDM6A*) are reported to cause Kabuki syndrome type 2, which is characterized by highly similar clinical characteristics as KaS [32]. Thus, KMT2C and KMT2D are associated with a large range of interaction partners and are responsible for methylation of H3K4.

Phenotypically, KlS and KaS have many overlapping characteristics such as ID, hypotonia, developmental delay, seizures, feeding difficulties and autism-like behaviour [33]. Given this clinical overlap between both syndromes, the co-occurrence of KMT2C (associated with KSS) and KMT2D (associated with KaS) in molecular complexes, and our previous studies showing a very strong genetic interaction between the *Drosophila* orthologs of EHMT1 and KMT2D [13], we reasoned that both proteins may regulate the expression of common genes and/or biological pathways. Here, we describe the previously unknown transcriptome in whole blood of KlS (*EHMT1* mutations) and KaS (*KMT2D* mutations) patients. The gene ontology terms describing the function of differentially expressed genes are directly relevant for the human phenotype. Additionally, we show a significant overlap of differentially expressed genes and a striking common direction of these commonly misregulated genes. Lastly, we present putative protein-protein networks for the differentially expressed genes in KlS and KaS. Using this approach we identified shared differentially expressed genes that may lead to the observed common phenotype between Kleefstra and Kabuki syndrome.

Results

Transcriptional changes in Kleefstra syndrome

To explore the transcriptional changes brought about by *EHMT1* mutations in Kleefstra syndrome, we performed next generation mRNA sequencing (RNA-seq) on whole blood. We compared mRNA levels from five Kleefstra syndrome patients carrying an *EHMT1* loss-of-function mutation (**Table 1**) to five control individuals. Sequenced reads were aligned to the human genome and the number of unique reads mapped to each gene was quantified and normalized for library size. Differential gene expression was performed using DeSeq2 [34]. We identified a high number of differential expressed genes in Kleefstra syndrome with 595 genes misexpressed by at least 1.5-fold (p-value < 0.05) (**Figure 1A, Table S1**). Of these genes, 356 genes were upregulated and 239 genes downregulated. Of these, 11 genes were more than 2-fold upregulated and 22 genes were more than 2-fold downregulated. To provide insights into the functions of the up- and downregulated genes, we performed gene ontology (GO) enrichment analysis [35]. The genes that were upregulated showed enrichment for neuronal terms "dopamine catabolic process", "synapse assembly" and "neuron differentiation" (**Figure 1B, Table S2**). Additionally, terms related to "behaviour", "cell

adhesion", "cardiac-, and skeletal muscle formation" are also present along with many other terms. Considering only downregulated genes, neuronal terms involved in "signaling" and "axonogenesis" are found (**Figure 1C, Table S3**). Notably, 16 genes from the protocadherin gamma subfamily were downregulated in Kleefstra syndrome patients accounting for the high enrichment of cell adhesion GO terms. Additionally, "negative regulation of circadian sleep/wake cycle, wakefulness", "negative regulation of behavior" and "G-protein coupled receptor signaling pathway" are among the many associated GO terms. GO terms associated with all (595) misregulated genes revealed again "cell-cell adhesion", "signaling" and "regulation of neuron differentiation" among the many associated GO terms (**Figure 1D, Table S4**). Taken together, GO term analysis of differentially expressed genes from whole blood of Kleefstra syndrome patients indentified enriched terms that strikingly describe the clinical phenotype.

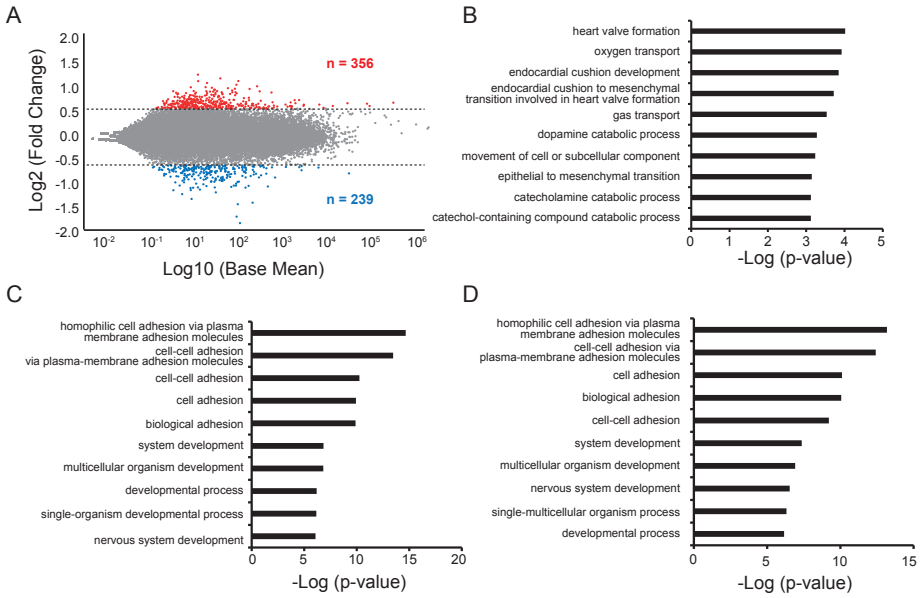


Figure 1: Transcriptional changes in Kleefstra syndrome.

(A) Scatter plots showing Log2 fold changes plotted against the Log10 normalized expression in whole blood of Kleefstra syndrome patients. Red and blue dots indicate significant up- and downregulated genes, respectively (fold change > 1.5; $p < 0.05$ (Wald test)). Top 10 enriched gene ontology terms identified using the Panther software for upregulated genes (B), downregulated genes (C), and misregulated genes (D).

Transcriptional changes in Kabuki syndrome

RNA-seq was also performed on mRNA from five Kabuki syndrome patients carrying a *KMT2D* loss-of-function mutation and compared to mRNA levels in the same five control individuals (**Table 1**). Using the same thresholds as the differential gene expression analysis in Kleefstra syndrome patients (1.5-fold, *p*-value < 0.05), we identified 1318 differential expressed genes in Kabuki syndrome patients (**Figure 2A, Table S5**). Of these genes, 779 genes are upregulated and 539 genes are downregulated. Of these, 73 and 81 genes are 2-fold up- and downregulted respectively. GO term analysis of the 1.5 fold upregulated genes revealed many genes important for "neuronal functioning" and "synapse assembly" (**Figure 2B, Table S6**). Go term analysis of the downregulated genes revealed terms describing "cell surface receptor signaling", "chemotaxis", and neuronal terms like "signaling", "neurogenesis" and "regulation of synaptic transmission" (**Figure 2C, Table S7**) along with many others. GO term analysis of the total (1318) misregulated genes revealed terms describing "nervous system development", "neurogenesis", and "signaling" (**Figure 2D, Table S8**). Interestingly, 11 misregulated genes were found to be involved in "learning". Thus, the differentially expressed genes may provide clues to the underlying mechanisms of the clinical phenotype of the patients.

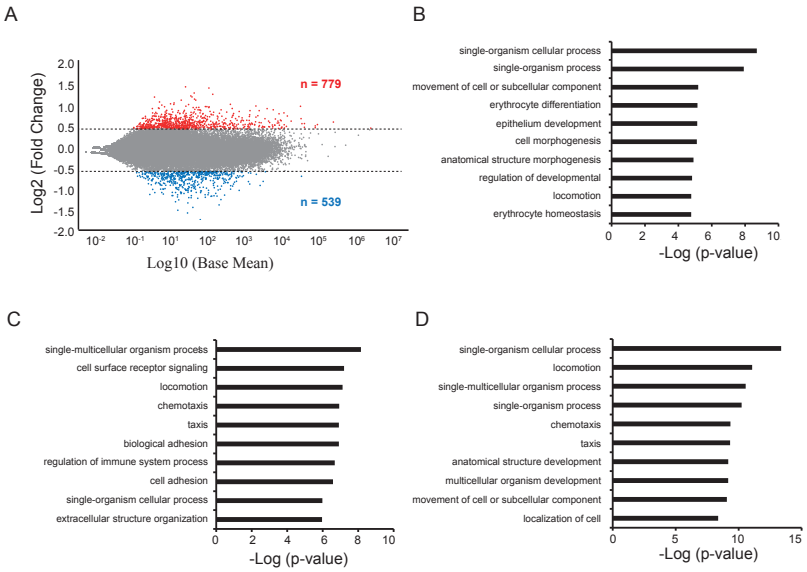


Figure 2: Transcriptional changes in Kabuki syndrome.

(A) Scatter plots showing Log2 fold changes plotted against the Log10 normalized expression in whole blood of Kabuki syndrome patients. Red and blue dots indicate significant up- and downregulated genes, respectively (fold change > 1.5; $p < 0.05$ (Wald test)). Top 10 enriched gene ontology terms identified using the Panther software for upregulated genes (B), downregulated genes (C), and misregulated genes (D).

A significant overlap between differential expressed genes in Kleefstra- and Kabuki syndrome

In order to identify commonly misregulated genes between KIS and KaS syndrome, an overlap analysis was performed using biovenn [36]. Of the 595 and 1318 genes that are 1.5-fold misregulated in KIS and KaS, respectively, 219 genes were identified to be present in both groups (10.43 fold enriched, *p*-value < 10^{-203} (hypergeometric distribution), **Figure 3A, Table S9**). Interestingly, 214 of these 219 genes, (98%) were regulated in the same direction (both up- or downregulated) (**Table S9**). Examples of GO terms describing this group of overlapping genes are "axon development", "MAPK signaling" and development of many brain structures such as "cerebellum", "forebrain", and "hypothalamus" (**Figure 3B, Table S10**). These GO terms can be directly relevant to the overlapping phenotypes between KIS and KaS. Additionally, comparison of associated GO terms from KIS (**Table S4,** 652 GO terms) to KaS (**Table S8,** 1129 GO terms) confirmed a strong functional overlap. 299 overlapping GO terms (of 20972 total GO terms) were found to be associated to both diseases (8.5x enriched, *p*-value 10^{-210} (hypergeometic distribution), **Figure 3C**). Thus, apart from the transcriptome that is unique for each disease, overlapping differential expressed genes and functionality could be established.

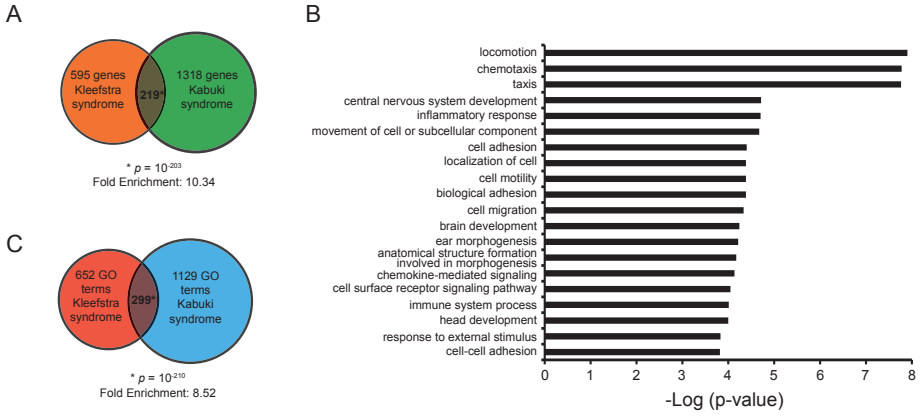


Figure 3: Genes misregulated in Kleefstra syndrome and Kabuki syndrome show a significant overlap.

(A) Venn diagram showing the overlap between misregulated genes in Kleefstra syndrome and Kabuki syndrome. The overlap of 219 genes is larger than expected by random chance, based on a hypergeometric test (***p***-value= 10^{-203} and 10.34 times). (B) Top 20 enriched gene ontology terms identified using the Panther software for 219 overlapping genes. (C) Venn diagram showing the overlap between gene ontology terms associated with Kleefstra - and Kabuki syndrome misregulated genes. The overlap of 299 gene ontology terms is larger than expected by random chance, based on a hypergeometric test (***p***-value= 10^{-210} and 8.52 times enriched).

Functional clusters of misregulated genes identified in Kleefstra and Kabuki syndrome

In order to identify more functional connections within the differential expressed genes, gene networks were assembled using String [37]. String combines several databases and computes scores for the level of protein-protein interaction. For all differentially regulated genes in KlS, a high confidence interaction network was generated (**Figure 4**) (PPI enrichment *p*-value: $2.33*10^{-15}$). One main network can be appreciated within several small connections and consists of chemokine receptors and G-protein coupled receptors (GPR17 as example). The other smaller networks are involved in kinase function (NTRK1), cell-adhesion (CADM1), dopamine hydroxylase (DBH), MAPK10, and neuropilin (NRP1).

A high confidence network for differentially expressed genes in KaS was also generated (**Figure 5**) (PPI enrichment *p*-value: $3.61*10^{-12}$). This network shows high interconnectivity and consists of three larger clusters. The first (middle, oval shape) cluster contains reelin/MAPK associated genes (RELN, SRC, DOK2, MAPK11 and CDH1 (Arc-1)). The second large cluster (right, round shape) contains chemokine receptors and G-protein coupled receptors (PGR55, PTGER3), whereas the third large cluster (left, round shape) contains RAS (RHOC), Rho (STARD8) and transcription factors (GATA1). Additionally, sub-clusters within the network are involved in potassium (KCNS1) and calcium (CACNA1I) voltage gated ion channels, choline (CHDH), wnt (SFRP5) and hedgehog (GLI1) signaling. Thus, these protein-protein networks give functional insights in the underlying mechanisms of KlS and KaS.

Next, a network was also generated from the 219 commonly differential expressd genes (PPI enrichment *p*-value: $7.44*10^{-9}$) in order to detect common connections (not shown in this thesis). 19 proteins show interconnection of which 10 are within the cytokine cluster (ie containing CCL5 wich is within the cluster of GPR17 in **Figure 4**). Thus, this network represents common pathways affected in KlS and KaS. However, the functional overlap between KlS and KaS is less pronounced than the common GO terms (**figure 3C**) which could be due to technical limitations of the String database.

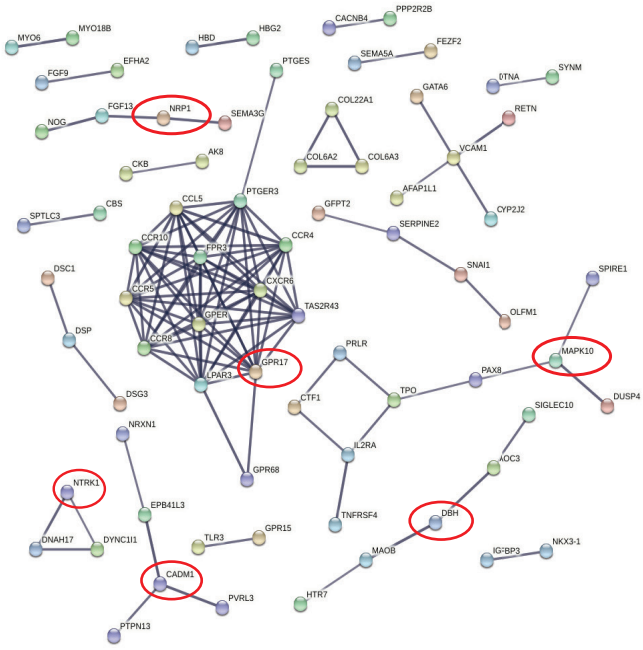


Figure 4: Protein network modules within the differentially expressed genes of Kleefstra syndrome.

Modules within the gene network of differrentially expressed genes reveal underlying pathways affected in Kleefstra syndrome. One mayor cluster of receptor function can be identified. Circled as example for this module is GRP17. Additional smaler clusters can also be appreciated. Examples are NTRK1 for kinase activity, CADM1 for cell adhesion, dopamine beta hydroxylase (DBH), MAPK, and neuropillin (NRP1).

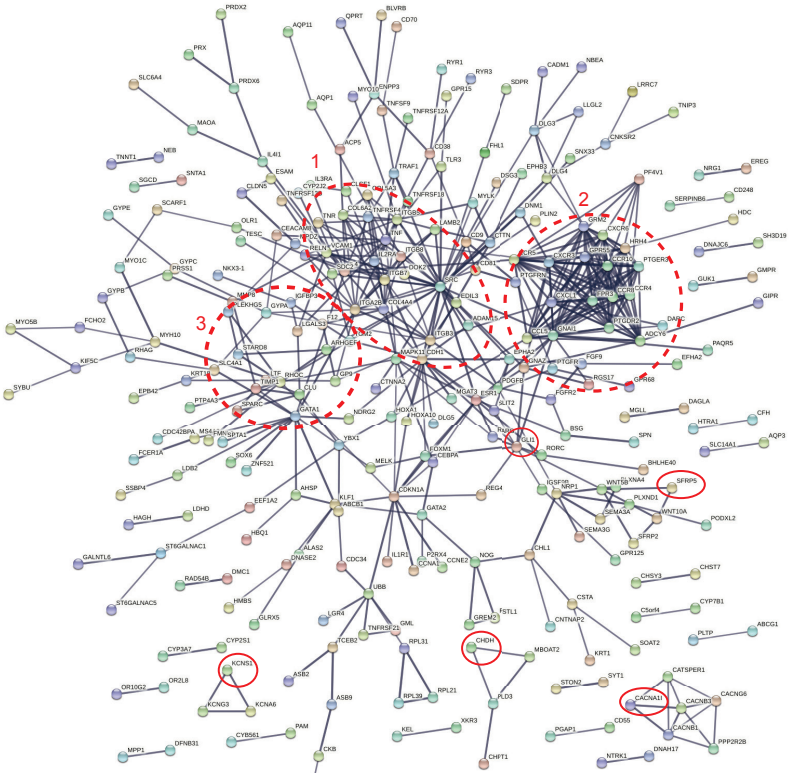


Figure 5: Protein network modules within the differentially expressed genes of Kabuki syndrome.

Modules within the gene network of differentially expressed genes reveal underlying pathways affected in Kabuki syndrome. Three mayor clusters can be indentified (red circles) within the dense network. The first one contains reeling/MAPK signaling factors (RELN, SRC, MAPK11, ARC-1(CDH1)). Additionally, a prominent cluster of receptors such as G-protein coupled receptors (GPR55) can be appreciated. The third large cluster contains members of the RAS family (RHOC and STARD8) and is closely linked to the first cluster. Smaller groups containing members of Potassium (KCNS1) and calcium (CACNA1I) gated ion channels and signaling pathways such as hedgehog (GLI1) and wnt signaling (SFRP5) are marked. Lastly, choline dehydrogenase (CHDH) forms a separate cluster.

Discussion

Kleefstra- and Kabuki syndrome have many common phenotypes such as ID and autism spectrum disorder, facial features, developmental delay, sleep- and eating difficulties, seizures, and hypotonia [33]. Mutations associated with these syndromes are in the genes *EHMT1* and *KMT2D* respectively. Additionally, mutations in *KMT2C*, the paralog of KMT2D, is associated with KSS, and is within an overlapping phenotypic spectrum. Here, we describe the identification of differentially expressed genes in blood of five KlS (*EHMT1* mutations) and five KaS (*KMT2D* mutations) compared to five healthy control individuals. We identified

many differential expressed genes in whole blood of both disease entities. Additionally, a high degree of overlap was identified between differentially expressed genes and associated GO terms describing the functions of these genes.

In this study, we used whole blood as surrogate tissue to the brain. Peripheral blood offers a major advantage in contrast to other tissues because it is easily accessible, minimally invasive to the individual to collect, and can support large population based collections. Blood cells are in contact with many other organs and carry bioactive molecules such as oxygen, metabolites, nutrients, cytokines and hormones and thereby potentially reflecting the physiological state of other organs. In addition, blood can be regarded as an organ by itself. However, dispite the large difference seen in expression profiles between blood and brain samples, comparative studies have revealed common patterns of deregulation, biological functions and targets of transcription factors [38]. Furthermore, over 80% of the human peripheral blood transcriptome is co-expressed in any given tissue [39]. Many studies report blood biomarkers in neurodevelopmental disorders such as ID syndromes [40], schizophrenia [41], and autism [42]. Thus, peripheral blood can be used to apply the functional genomics tools for discovering of biomarkers for other organs such as the brain. However, the observed differential expressed genes could still reflect a blood specific transcriptome.

GO enrichment analysis of misregulated genes in KIS revealed many terms relevant to the underlying pathogenesis of the human phenotype (**Table S2-S4**). For instance, upregulation of the catecholamine "dopamine", terms related to the "synapse" and "action potential propagation" could provide clues to the pathology. "Negative regulation of circadian sleep/wake cycle, wakefulness" and "positive regulation of heart growth" are terms describing the down regulated genes and are other examples of terms directly relevant to the phenotype. These underlying genes could provide clues to the etiology of the neurodevelopmental and behavioural phenotype. Next, a group of sixteen protocadherins was found to be downregulated in Kleefstra syndrome. Protocadherins are transmembrane proteins involved in cell-cell connectivity of synapses and neuronal growth cones. The protocadherins are known to be implicated in ID syndromes such as RETT syndrome [43], but may also contribute to the pathogenesis of the neurological features observed in Kleefstra syndrome.

Similar to the observations of differentially expressed genes in KIS, the KaS transcriptome shows GO terms directly relevant to the phenotype (**Table S6-S8**). First, the terms "cognition" (17 genes) and "learning" (11 genes) are directly relevant to the phenotype as these terms potentially give clues to the etiology of ID in Kabuki syndrome patients. Second, the term "odontogenesis" refers to a group of genes responsible for the proper development of teeth and their arrangement in the mouth. Dentition abnormalities, including wide spread teeth and/ or hypodontia have been reported frequently in Kabuki

syndrome [17]. Cleft lip/ palate is also observed however with lower frequencies [17]. Lastly, structural abnormalities of the brain such as ventricular dilatation, brain atrophy, agenesis/ hypoplasia of the corpus callosum and white matter abnormalities are observed frequently in Kabuki syndrome patients [17]. These observations could be related to the misregulated genes that clustered together in the term "brain development". Thus, these results provide a hint towards the underlying mechanisms of these structural brain abnormalities.

In the GO terms for the overlapping genes (**Figure 3**) and gene network analysis of both diseases entities (**Figure 4 and 5**) prominent themes are found. Many GO terms can be directly related to ID. For instance, "positive regulation of inhibitory G-protein coupled receptor phosphorylation", "calcium mediated signaling" and "signal transduction" are all related to mitogen-activated protein kinase (MAPK) signaling. This signaling pathway has been attributed to learning and memory but also to synaptic plasticity in the adult brain [44]. The MAPK signaling cascade is triggered by *N*-methyl-D-aspartate- and voltage gated calcium channels. Increased calcium inside the neuron will promote the RAS-GTP form and activate the RAF-MEK-ERK signaling cascade that is required for processes underlying synaptic plasticity by long term potentiation and/or long term depression. The MAPK signaling pathway is shown to regulate spacial memory in mice [45] and fear conditioning in rats [46]. Other examples of signaling pathways identified in this study are Wnt signaling which has been linked to learning and memory in the object recognition test [47] and sonic hedgehog signaling that has been implicated in neuronal plasticity and neuronal differentiation in the hippocampus [48]. These latter examples were seen as highly interacting genes in the network analysis of Kabuki syndrome (**Figure 5**). Thus, via clustering analysis of genes misregulated in KIS and KaS, common functional themes could be identified that hint towards the underlying molecular etiology of the ID aspect of both disease entities.

Another prominent theme in the results presented here is the large overlap of differentially expressed genes (**Figure 3A**) and the nearly identical direction (up- or downregulated) of commonly misregulated genes (**Table S9**). EHMT1 has been shown to regulate H3K9me2 involved in transcriptional repression [6] and on the other hand is KMT2D a methyltransferase involved in H3K4me3 and transcriptional activation [28, 49]. *Drosophila* genetic interaction studies has shown that there is a strong antagonistic interaction between the KMT2D ortholog trr and G9a, the ortholog of EHMT1 [13]. It has also been shown that the targets of the *Drosophila* G9a and trr significantly overlap on the genomic and transcriptome level (**Chapter 3**). The study described here presents a similar finding. Disruption of either EHMT1 or KMT2D results in a strikingly similar altered transcriptome. Thus, this study underscores that phenotypically overlapping syndromes that originate from mutations in distinct epigenetic factors show molecular convergence.

Materials and Methods

Patients selection

Participating patients were selected from the Kleefstra- and Kabuki syndrome cohort of the Radboudumc and the Maastricht university medical center. Control individual were healthy adult volunteers. All patients with KlS or KaS were clinically diagnosed as described before [50] and showed a heterozygous loss of function mutation in *EHMT1* (Kleefstra syndrome) or *KMT2D* (Kabuki syndrome) (**Table 1**). Full blood was drawn of individuals according to standard procedures. Whole blood was stored in PAXgene Blood RNA Tubes (Qiagen) at -80 °C until processed further.

Table 1: Control, Kleefstra- and Kabuki syndrome patients with corresponding variants in *EHMT1* or *KMT2D* used for this study.

Number	Patient number	Group	DNA variant	Protein variant	Gene	Inheritance type
KlS_1	RNA15-00280	Kleefstra syndrome	c.2704C>T	p.(Arg902*)	*EHMT1*	*de novo*
KlS _2	RNA15-00248	Kleefstra syndrome	c.3072_3073del	p.(Val1026fs)	*EHMT1*	*de novo*
KlS _3	RNA15-00423	Kleefstra syndrome	c.2380del	p.(Gln794fs)	*EHMT1*	*de novo*
KlS _4	RNA15-00150	Kleefstra syndrome	c.2587C>T/+	p.(Gln83*)	*EHMT1*	*de novo*
KlS _5	RNA15-00540	Kleefstra syndrome	c.2713-1G>T	p.(?)	*EHMT1*	*de novo*
KaS_1	RNA15-00351	Kabuki syndrome	c.9329delG	p.(Arg3110Profs*9)	*KMT2D*	*de novo*
KaS_2	RNA15-00298	Kabuki syndrome	c.11416C>T#	p.(Gln3806*)	*KMT2D*	*de novo*
KaS_3	RNA15-00147	Kabuki syndrome	c.11707C>T	p.(Gln3903*)	*KMT2D*	*de novo*
KaS _4	RNA15-00281	Kabuki syndrome	c.12164_12165-delCT	p.(Pro4055Argfs*6)	*KMT2D*	unknown
KaS _5	RNA15-00300	Kabuki syndrome	c.5320-2A>G	p.(?)	*KMT2D*	unknown
CON_1	RNA15-00279	Control	-	-	-	-
CON_2	RNA15-00278	Control	-	-	-	-
CON_3	RNA15-00526	Control	-	-	-	-
CON_4	RNA15-00527	Control	-	-	-	-
CON_5	RNA15-00530	Control	-	-	-	-

Control individuals were healthy adults. # =mosaic 5-20%. Genome reference version GRCh37 (hg19) was used for variant annotation.

RNA isolation, library preparation, and sequencing

Total RNA was extracted from whole blood in PAXgene Blood RNA Tubes (Qiagen) using standard procedures from the manufacturer. Ribosomal RNA and globin mRNA was depleted using the Globin-Zero Gold rRNA Removal Kit (Illumina). The isolated mRNA was subjected to 7.5 minutes fragmentation at 95 °C in fragmentation buffer (40 mM Tris-acetate, 10 mM Potassium Acetate, 30 mM Magnesium Acetate, pH 8.2) after which the samples were immediately placed on ice. First strand cDNA synthesis was performed using superscript (Invitrogen) with random hexamers as primers. Second strand synthesis was performed using *E. coli* polymerase (New England Biolabs) and random hexamers. DNA end repair was performed on 5 ng of double stranded cDNA followed by ligation of Illumina sequencing adaptors (Next-flex) and size selection for 300 bp. Fragments were amplified linearly (9 PCR cycles) as validated by quantitative real-time PCR (qPCR). Sample quality was assessed using the Agilent 2200 TapeStation system. Additionally, quality and quantity of the obtained libraries were assessed by library quantification by qPCR (Kapa technologies). Cluster generation and paired-end sequencing-by-synthesis (43 bp) was performed using the Illumina NextSeq500 according to standard protocols of the manufacturer. The image files generated by NextSeq 500 were processed to extract DNA sequence data. From the NextSeq 500, we obtained between 33 and 102 million reads (**Table S11**). Sequenced reads were aligned to the Human genome (GRCh37/hg19) using STAR [51] (version 2.5). Only reads that uniquely aligned to the genome were considered for further analysis and total alignment efficiency was between 21% and 36% (**Table S11**). The number of aligned reads mapping to gene coding sequences was counted in the same step as mapping using STAR (quantMode "GeneCounts" column four comparable with htseq-count [52] option -s reverse) with gene annotations extracted from the file Homo_sapiens.GRCh37.75.gtf, available at **http://www.ensembl.org**. The unambiguously mapped reads (ranging from 7.2–37.7 million reads) were used for further analysis. The number of reads per gene were normalized using DESeq2 software [34]. DESeq2 was then used to identify genes that were differentially expressed in the five individuals with Kleefstra syndrome relative to the five control individuals (p-value < 0.05 (Wald-test), fold change > 1.5). Additionally, we identified differential expressed genes in the five Kabuki syndrome patients relative to the (same) five control individuals (p-value < 0.05 (Wald-test), fold change > 1.5).

Gene Ontology enrichment, hypergeometric analysis, and String database

Gene Ontology (GO) enrichment analysis was performed using the Panther software [53] on **http://geneontology.org/** (GO Ontology database released on 2016-07-15). Overlap between datasets was determined and visualized as a Venn diagram by BioVenn [36]. Hypergeometric statistics on overlaps were calculated using https://www.geneprof.org/

GeneProf/tools/hypergeometric.jsp. The full data set of differentially expressed genes and GO-terms are in the supplemental tables (**Tables S1-S8**).

The String database (https://string-db.org/, version 10.0) was used to visualize protein networks of all differentially expressed genes per disease entity [37]. Only text mining, experiments and databases were visualized. Thickness of lines represents the interaction score with thin lines representing 0.700 ("high confidence") thicker lines represents 0.900 ("highest confidence"). Color of nodes represents individual proteins. Disconnected nodes are hidden.

References

1. Kochinke K, Zweier C, Nijhof B, Fenckova M, Cizek P, Honti F, et al. Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. American journal of human genetics. 2016;98(1):149-64. doi: 10.1016/j.ajhg.2015.11.024. PubMed PMID: 26748517; PubMed Central PMCID: PMC4716705.
2. Kleefstra T, Schenck A, Kramer JM, van Bokhoven H. The genetics of cognitive epigenetics. Neuropharmacology. 2014;80:83-94. doi: 10.1016/j.neuropharm.2013.12.025. PubMed PMID: 24434855.
3. van Bokhoven H. Genetic and epigenetic networks in intellectual disabilities. Annual review of genetics. 2011;45:81-104. doi: 10.1146/annurev-genet-110410-132512. PubMed PMID: 21910631.
4. Kleefstra T, Brunner HG, Amiel J, Oudakker AR, Nillesen WM, Magee A, et al. Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. American journal of human genetics. 2006;79(2):370-7. doi: 10.1086/505693. PubMed PMID: 16826528; PubMed Central PMCID: PMC1559478.
5. Willemsen MH, Vulto-van Silfhout AT, Nillesen WM, Wissink-Lindhout WM, van Bokhoven H, Philip N, et al. Update on Kleefstra Syndrome. Molecular syndromology. 2012;2(3-5):202-12. doi: 000335648. PubMed PMID: 22670141; PubMed Central PMCID: PMC3366700.
6. Tachibana M, Sugimoto K, Fukushima T, Shinkai Y. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. The Journal of biological chemistry. 2001;276(27):25309-17. doi: 10.1074/jbc.M101914200. PubMed PMID: 11316813.
7. Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, Ohki M, et al. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes & development. 2002;16(14):1779-91. doi: 10.1101/gad.989402. PubMed PMID: 12130538; PubMed Central PMCID: PMC186403.
8. Tachibana M, Ueda J, Fukuda M, Takeda N, Ohta T, Iwanari H, et al. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes & development. 2005;19(7):815-26. doi: 10.1101/gad.1284005. PubMed PMID: 15774718; PubMed Central PMCID: PMC1074319.
9. Shinkai Y, Tachibana M. H3K9 methyltransferase G9a and the related molecule GLP. Genes & development. 2011;25(8):781-8. doi: 10.1101/gad.2027411. PubMed PMID: 21498567; PubMed Central PMCID: PMC3078703.
10. Stewart DR, Kleefstra T. The chromosome 9q subtelomere deletion syndrome. American journal of medical genetics Part C, Seminars in medical genetics. 2007;145C(4):383-92. doi: 10.1002/ajmg.c.30148. PubMed PMID: 17910072.
11. Kleefstra T, van Zelst-Stams WA, Nillesen WM, Cormier-Daire V, Houge G, Foulds N, et al. Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype. Journal of medical genetics. 2009;46(9):598-606. doi: 10.1136/jmg.2008.062950. PubMed PMID: 19264732.
12. Vermeulen K, de Boer A, Janzing JGE, Koolen DA, Ockeloen CW, Willemsen MH, et al. Adaptive and maladaptive functioning in Kleefstra syndrome compared to other rare genetic disorders with intellectual disabilities. American journal of medical genetics Part A. 2017. doi: 10.1002/ajmg.a.38280. PubMed PMID: 28498556.

13. Kleefstra T, Kramer JM, Neveling K, Willemsen MH, Koemans TS, Vissers LE, et al. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. American journal of human genetics. 2012;91(1):73-82. doi: 10.1016/j.ajhg.2012.05.003. PubMed PMID: 22726846; PubMed Central PMCID: PMC3397275.
14. Koemans TS, Tjitske Kleefstra , Stone MH, Chubak MC, Reijnders MRF, Munnik Sd, et al. Functional Convergence of Histone Methyltransferases EHMT1 and KMT2C Involved in Intellectual Disability and Autism. PLoS genetics. accepted.
15. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature genetics. 2010;42(9):790-3. doi: 10.1038/ng.646. PubMed PMID: 20711175; PubMed Central PMCID: PMC2930028.
16. Bogershausen N, Gatinois V, Riehmer V, Kayserili H, Becker J, Thoenes M, et al. Mutation Update for Kabuki Syndrome Genes KMT2D and KDM6A and Further Delineation of X-Linked Kabuki Syndrome Subtype 2. Human mutation. 2016;37(9):847-64. doi: 10.1002/humu.23026. PubMed PMID: 27302555.
17. Bogershausen N, Wollnik B. Unmasking Kabuki syndrome. Clinical genetics. 2013;83(3):201-11. doi: 10.1111/cge.12051. PubMed PMID: 23131014.
18. Hu D, Gao X, Morgan MA, Herz HM, Smith ER, Shilatifard A. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Molecular and cellular biology. 2013;33(23):4745-54. doi: 10.1128/MCB.01181-13. PubMed PMID: 24081332; PubMed Central PMCID: PMC3838007.
19. Hu D, Garruss AS, Gao X, Morgan MA, Cook M, Smith ER, et al. The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nature structural & molecular biology. 2013;20(9):1093-7. doi: 10.1038/nsmb.2653. PubMed PMID: 23934151; PubMed Central PMCID: PMC3805109.
20. Mohan M, Herz HM, Smith ER, Zhang Y, Jackson J, Washburn MP, et al. The COMPASS family of H3K4 methylases in *Drosophila*. Molecular and cellular biology. 2011;31(21):4310-8. doi: 10.1128/MCB.06092-11. PubMed PMID: 21875999; PubMed Central PMCID: PMC3209330.
21. Shilatifard A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annual review of biochemistry. 2012;81:65-95. doi: 10.1146/annurev-biochem-051710-134100. PubMed PMID: 22663077; PubMed Central PMCID: PMC4010150.
22. Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, et al. Enhancer-associated H3K4 monomethylation by Trithorax-related, the *Drosophila* homolog of mammalian Mll3/Mll4. Genes & development. 2012;26(23):2604-20. doi: 10.1101/gad.201327.112. PubMed PMID: 23166019; PubMed Central PMCID: PMC3521626.
23. Johnston DM, Sedkov Y, Petruk S, Riley KM, Fujioka M, Jaynes JB, et al. Ecdysone- and NO-mediated gene regulation by competing EcR/Usp and E75A nuclear receptors during *Drosophila* development. Molecular cell. 2011;44(1):51-61. doi: 10.1016/j.molcel.2011.07.033. PubMed PMID: 21981918; PubMed Central PMCID: PMC3190167.
24. Sedkov Y, Cho E, Petruk S, Cherbas L, Smith ST, Jones RS, et al. Methylation at lysine 4 of histone H3 in ecdysone-dependent development of *Drosophila*. Nature. 2003;426(6962):78-83. doi: 10.1038/nature02080. PubMed PMID: 14603321; PubMed Central PMCID: PMC2743927.

25. Kim DH, Lee J, Lee B, Lee JW. ASCOM controls farnesoid X receptor transactivation through its associated histone H3 lysine 4 methyltransferase activity. Molecular endocrinology. 2009;23(10):1556-62. doi: 10.1210/me.2009-0099. PubMed PMID: 19556342; PubMed Central PMCID: PMC2754897.
26. Herz HM, Garruss A, Shilatifard A. SET for life: biochemical activities and biological functions of SET domain-containing proteins. Trends in biochemical sciences. 2013;38(12):621-39. doi: 10.1016/j.tibs.2013.09.004. PubMed PMID: 24148750; PubMed Central PMCID: PMC3941473.
27. Ananthanarayanan M, Li Y, Surapureddi S, Balasubramaniyan N, Ahn J, Goldstein JA, et al. Histone H3K4 trimethylation by MLL3 as part of ASCOM complex is critical for NR activation of bile acid transporter genes and is downregulated in cholestasis. American journal of physiology Gastrointestinal and liver physiology. 2011;300(5):G771-81. doi: 10.1152/ajpgi.00499.2010. PubMed PMID: 21330447; PubMed Central PMCID: PMC3094144.
28. Cho YW, Hong T, Hong S, Guo H, Yu H, Kim D, et al. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. The Journal of biological chemistry. 2007;282(28):20395-406. doi: 10.1074/jbc.M701574200. PubMed PMID: 17500065; PubMed Central PMCID: PMCPMC2729684.
29. Jowsey PA, Doherty AJ, Rouse J. Human PTIP facilitates ATM-mediated activation of p53 and promotes cellular resistance to ionizing radiation. The Journal of biological chemistry. 2004;279(53):55562-9. doi: 10.1074/jbc.M411021200. PubMed PMID: 15456759.
30. Goo YH, Sohn YC, Kim DH, Kim SW, Kang MJ, Jung DJ, et al. Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. Molecular and cellular biology. 2003;23(1):140-9. PubMed PMID: 12482968; PubMed Central PMCID: PMC140670.
31. Goo YH, Na SY, Zhang H, Xu J, Hong S, Cheong J, et al. Interactions between activating signal cointegrator-2 and the tumor suppressor retinoblastoma in androgen receptor transactivation. The Journal of biological chemistry. 2004;279(8):7131-5. doi: 10.1074/jbc.M312563200. PubMed PMID: 14645241.
32. Banka S, Lederer D, Benoit V, Jenkins E, Howard E, Bunstone S, et al. Novel KDM6A (UTX) mutations and a clinical and molecular review of the X-linked Kabuki syndrome (KS2). Clinical genetics. 2015;87(3):252-8. doi: 10.1111/cge.12363. PubMed PMID: 24527667.
33. Mullegama SV, Alaimo JT, Chen L, Elsea SH. Phenotypic and molecular convergence of 2q23.1 deletion syndrome with other neurodevelopmental syndromes associated with autism spectrum disorder. International journal of molecular sciences. 2015;16(4):7627-43. doi: 10.3390/ijms16047627. PubMed PMID: 25853262; PubMed Central PMCID: PMCPMC4425039.
34. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8. PubMed PMID: 25516281; PubMed Central PMCID: PMC4302049.
35. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics. 2000;25(1):25-9. doi: 10.1038/75556. PubMed PMID: 10802651; PubMed Central PMCID: PMC3037419.
36. Hulsen T, de Vlieg J, Alkema W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC genomics. 2008;9:488. doi: 10.1186/1471-2164-9-488. PubMed PMID: 18925949; PubMed Central PMCID: PMC2584113.

37. Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic acids research. 2000;28(18):3442-4. PubMed PMID: 10982861; PubMed Central PMCID: PMC110752.
38. Hess JL, Tylee DS, Barve R, de Jong S, Ophoff RA, KumarasingheN, et al. Transcriptome-wide mega-analyses reveal joint dysregulation of immunologic genes and transcription regulators in brain and blood in schizophrenia. Schizophr Res. 2016 Oct;176(2-3):114-124. doi: 10.1016/j.schres.2016.07.006.
39. Liew CC, Ma J, Tang HC, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. J Lab Clin Med. 2006;147(3):126-32. doi: 10.1016/j.lab.2005.10.005. PubMed PMID: 16503242.
40. Tang Y, Schapiro MB, Franz DN, Patterson BJ, Hickey FJ, Schorry EK, et al. Blood expression profiles for tuberous sclerosis complex 2, neurofibromatosis type 1, and Down's syndrome. Ann Neurol. 2004;56(6):808-14. doi: 10.1002/ana.20291. PubMed PMID: 15562430.
41. Tsuang MT, Nossova N, Yager T, Tsuang MM, Guo SC, Shyu KG, et al. Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2005;133B(1):1-5. doi: 10.1002/ajmg.b.30161. PubMed PMID: 15645418.
42. Glatt SJ, Tsuang MT, Winn M, Chandler SD, Collins M, Lopez L, et al. Blood-based gene expression signatures of infants and toddlers with autism. Journal of the American Academy of Child and Adolescent Psychiatry. 2012;51(9):934-44 e2. doi: 10.1016/j.jaac.2012.07.007. PubMed PMID: 22917206; PubMed Central PMCID: PMCPMC3756503.
43. Miyake K, Hirasawa T, Soutome M, Itoh M, Goto Y, Endoh K, et al. The protocadherins, PCDHB1 and PCDH7, are regulated by MeCP2 in neuronal cells and brain tissues: implication for pathogenesis of Rett syndrome. BMC neuroscience. 2011;12:81. doi: 10.1186/1471-2202-12-81. PubMed PMID: 21824415; PubMed Central PMCID: PMC3160964.
44. Thomas GM, Huganir RL. MAPK cascade signalling and synaptic plasticity. Nature reviews Neuroscience. 2004;5(3):173-83. doi: 10.1038/nrn1346. PubMed PMID: 14976517.
45. Atkins CM, Selcher JC, Petraitis JJ, Trzaskos JM, Sweatt JD. The MAPK cascade is required for mammalian associative learning. Nature neuroscience. 1998;1(7):602-9. doi: 10.1038/2836. PubMed PMID: 10196568.
46. Schafe GE, Atkins CM, Swank MW, Bauer EP, Sweatt JD, LeDoux JE. Activation of ERK/MAP kinase in the amygdala is required for memory consolidation of pavlovian fear conditioning. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2000;20(21):8177-87. PubMed PMID: 11050141.
47. Fortress AM, Frick KM. Hippocampal Wnt Signaling: Memory Regulation and Hormone Interactions. The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry. 2016;22(3):278-94. doi: 10.1177/1073858415574728. PubMed PMID: 25717070.
48. Yao PJ, Petralia RS, Mattson MP. Sonic Hedgehog Signaling and Hippocampal Neuroplasticity. Trends in neurosciences. 2016;39(12):840-50. doi: 10.1016/j.tins.2016.10.001. PubMed PMID: 27865563; PubMed Central PMCID: PMC5148655.
49. Lee S, Kim DH, Goo YH, Lee YC, Lee SK, Lee JW. Crucial roles for interactions between MLL3/4 and INI1 in nuclear receptor transactivation. Molecular endocrinology. 2009;23(5):610-9. doi: 10.1210/me.2008-0455. PubMed PMID: 19221051; PubMed Central PMCID: PMC2675954.

50. de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. The New England journal of medicine. 2012;367(20):1921-9. doi: 10.1056/NEJMoa1206524. PubMed PMID: 23033978.
51. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21. doi: 10.1093/bioinformatics/bts635. PubMed PMID: 23104886; PubMed Central PMCID: PMC3530905.
52. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-9. doi: 10.1093/bioinformatics/btu638. PubMed PMID: 25260700; PubMed Central PMCID: PMC4287950.
53. Gene Ontology C. Gene Ontology Consortium: going forward. Nucleic acids research. 2015;43(Database issue):D1049-56. doi: 10.1093/nar/gku1179. PubMed PMID: 25428369; PubMed Central PMCID: PMC4383973.

Supplemental Tables

(available upon request)

Table S1: Misexpressed genes in Kleefstra syndrome compared to control individuals based on Deseq2.

Table S2: Gene ontology terms identified using the Panther software for upregulated genes in Kleefstra syndrome.

Table S3: Gene ontology terms identified using the Panther software for downregulated genes in Kleefstra syndrome.

Table S4: Gene ontology terms identified using the Panther software for misregulated genes in Kleefstra syndrome.

Table S5: Misexpressed genes in Kabuki syndrome compared to control individuals based on Deseq2.

Table S6: Gene ontology terms identified using the Panther software for upregulated genes in Kabuki syndrome.

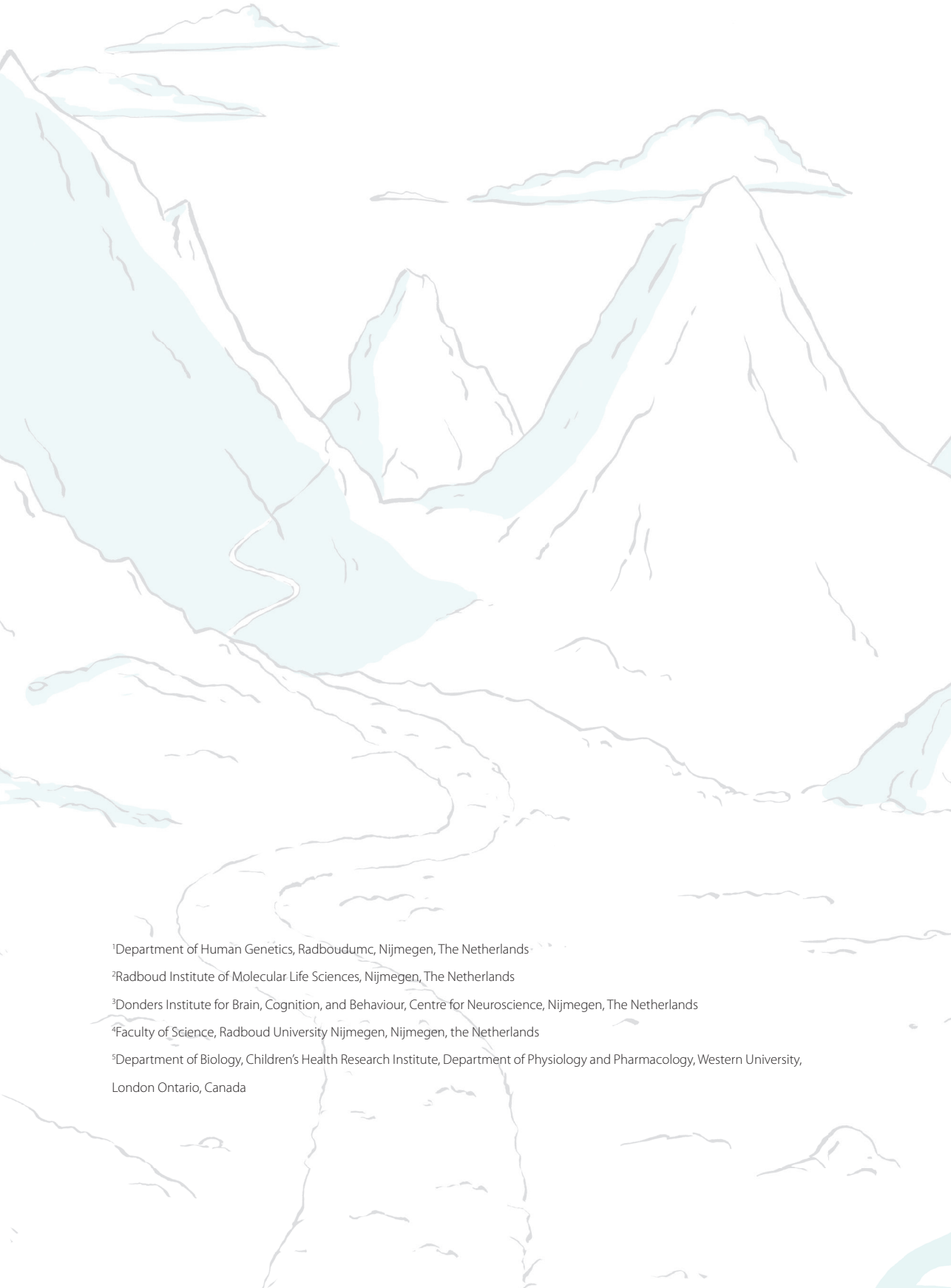
Table S7: Gene ontology terms identified using the Panther software for downregulated genes in Kabuki syndrome.

Table S8: Gene ontology terms identified using the Panther software for misregulated genes in Kabuki syndrome.

Table S9: List of overlapping differentially expressed genes and direction in the Kleefstra- and Kabuki syndrome transcriptome

Table S10: Gene ontology terms identified using the Panther software for the overlapping genes between Kleefstra- and Kabuki syndrome.

Table S11: Sequencing depths identified by STAR software.



[1]Department of Human Genetics, Radboudumc, Nijmegen, The Netherlands

[2]Radboud Institute of Molecular Life Sciences, Nijmegen, The Netherlands

[3]Donders Institute for Brain, Cognition, and Behaviour, Centre for Neuroscience, Nijmegen, The Netherlands

[4]Faculty of Science, Radboud University Nijmegen, Nijmegen, the Netherlands

[5]Department of Biology, Children's Health Research Institute, Department of Physiology and Pharmacology, Western University, London Ontario, Canada

Exome sequencing and protein-protein interaction studies enlarge the molecular network underlying Kleefstra syndrome

Tom S. Koemans[1,2,3]
Pascal W. Jansen[2,4]
Michiel Vermeulen[2,4]
Margot R.F. Reijnders[1]
Misa Fenckova[1,3]
Jamie M. Kramer[5]
Tjitske Kleefstra[1]
Hans van Bokhoven[1,3]
Annette Schenck[1,3]

5

Manuscript in preparation

Abstract

Kleefstra syndrome is defined by mutations in the *EHMT1* gene and is characterized by intellectual disability (ID), childhood hypotonia, autistic traits, characteristic facial dysmorphisms, and other variable clinical features. However, *EHMT1* mutations account for 25% of Kleefstra syndrome cases which leaves a large "EHMT1-negative" group with strikingly similar clinical characteristics, alluded to as the Kleefstra syndrome phenotypic spectrum (KSS). In a previous study we have shown that mutation in the KSS cohort include *MBD5*, *MLL3*, *SMARCB1*, and *NR1I3* and contribute to a epigenetic network involved in ID and autism. Here, we apply two approaches to further expand this EHMT1 network. First, clinical characterization followed by whole exome sequencing of two novel KSS patients identified *POGZ* and *DDX3X* as potentially causing factors. However, after initial findings, more patients were identified with mutations in these genes, but without the characteristic features of KSS. Additionally, genetic interaction studies showed no association between *EHMT1* and these two genes, which supports the paucity of clinical similarity between the associated phenotypes. Hence, *POGZ* and *DDX3X* are not added to the *EHMT1* network. The second approach is based on affinity purification followed by mass spectrometry in order to identify interaction partners of the *Drosophila* EHMT1 ortholog, G9a, by using a *Drosophila* fosmid line. This pull-down experiment identified a so far uncharacterized zinc finger protein CG9932 as a physical interaction partner of G9a which has orthology to human RE1-silencing transcription factor and ZNF462.

Introduction

Haploinsufficiency of the *EHMT1* gene causes Kleefstra Syndrome (KS, MIM #610253) which is characterized by intellectual disability (ID), autism spectrum disorder, hypotonia, facial dysmorphisms, and other features that occur in subsets of patients [1-3]. Around 25% of KS patients carry loss-of-function mutations in the *EHMT1* gene and we previously hypothesized that the "EHMT1 negative" cohort might carry mutations in genes with a shared biological function with EHMT1 [4]. Via a next generation sequencing approach in patients with Kleefstra phenotypic spectrum (KSS), *de novo* mutations have been found in *MBD5*, *MLL3*, *SMARCB1*, and *NR1I3,* all of which encode epigenetic regulators or transcription factors [4]. Our hypothesis that these genes are engaged in shared biological processes was further supported by established protein-protein interactions and genetic interaction studies of these genes in *Drosophila melanogaster*, which revealed a conserved epigenetic module underlying KSS [4]. Here we aimed to expand this module using two different approaches to identify additional members of the EHMT1 network related to KSS. The first approach is based on *de novo* variant identification in KSS patients using exome sequencing. The second approach uses affinity purification followed by mass spectrometry experiments to identify interactions partners of the *Drosophila* EHMT1 ortholog, G9a in a *Drosophila* fosmid line. This fly line expresses a G9a fusion protein with a molecular "tag" in an endogenous pattern [5, 6]. In this way, we identified a previously unknown protein, CG9932, as interaction partner. ZNF462 and RE-1 silencing transcription factor (REST) rank amongst the human genes with the highest sequence similarity to CG9932 and could be hits for further research to the human EHMT1 protein network.

Results

Identification of *de novo* mutations in individuals with clinical characteristics to Kleefstra syndrome

The two individuals described in this study with clinical features of KSS (**Figure 1**) were ascertained through family-based exome sequencing in a diagnostic setting in the Department of Human Genetics at the Radboudumc. Exome sequencing and data analysis were performed as previously described in the probands and their unaffected parents [7]. Individual 1 is diagnosed with severe ID, hypotonia, epilepsy, movement disorder, corpus callosum hypoplasia, hyperlaxity and visual problems. The proband carried a c.1126C>T (p.Arg376Cys) *de novo* variant in *DEAD-Box Helicase 3, X-Linked* (*DDX3X*). Individual 2 was diagnosed with severe ID, speech and language delay, motor delay, autism spectrum disorder, microcephaly, and behaviour difficulties such as characteristic mouth movements, sleeping problems, restless, and no eye contact. Patient 2 carried a c.2590C>T (p.Arg864*) *de*

novo variant in *pogo transposable element with ZNF domain* (*POGZ*). These results suggested that these genes may also be a part of the EHMT1-related network underlying KSS. In order to investigate this, we turned to *Drosophila* experiments to test genetic interaction between the respective orthologs.

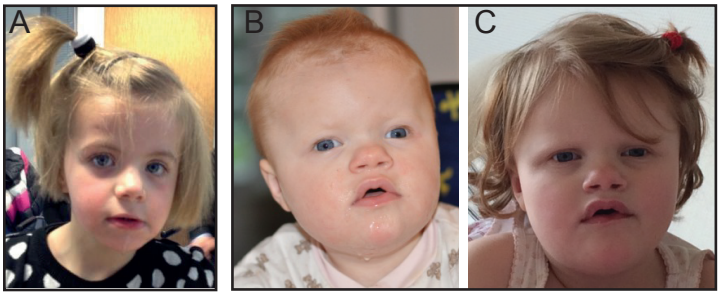


Figure 1: Clinical photograph of individual 1 (A) with a *de novo* mutation in DDX3X at the age of 4 and individual 2 with a *de novo* mutation in *POGZ* at the ages of 1 year (B) and 3 years (C). Reprints of figure panels are from Snijders-blok *et al.*, AJHG (2015) and Stessman *et al.*, AJHG (2016) with permission.

Absence of genetic interaction of *EHMT1* with *DDX3X* and *POGZ*

Previously, we used the *Drosophila* wing phenotype to test genetic interactions between KSS genes and to validate the EHMT1 network [4]. Here, we applied the same approach to test genetic interactions between the *Drosophila* orthologs of *EHMT1*, *DDX3X* and *POGZ*. Human paralogs EHMT1 and EHMT2 have one ortholog in *Drosophila*, G9a [8, 9]. Gene orthologs of *DDX3X* and *POGZ* were identified using the reverse BLAST method [10] and analysis of the Treefam database [11]. For DDX3X a clear ortholog was identified with 58% of the human amino acid sequence matched the sequence of belle including the ATP-dependent RNA helicase DEAD-box domain. 48% of the amino acid sequence of belle matched the sequence of human DDX3X. POGZ is less conserved in fly with 10% of the amino acid sequence of POGZ matching the amino acid sequence of relative of woc, row. 14% of the amino acid sequence of row matched the amino acid sequence of human POGZ including the C2H2-type zinc finger domain. Next, to test for genetic interaction we observed modulation of the established *G9a* induced overexpression phenotype in the fly wing by modulating gene expression of the orthologs of the putative KSS genes [4, 12]. Classification of wing vein phenotypes was performed as described previously with categories ranging from wild type (not affected) to severe (strongly affected) [4]. Overexpression of *G9a* in the *Drosophila* wing

consistently causes extra vein formation (**Figure 2, bar 1**). Rescue of the *G9a* overexpression phenotype was performed by using RNA interference against *G9a* (**Figure 2, bar 2**) and restored the wings to wild type levels. Next, we tested the modulation of the *G9a*-induced extra veins by genetic manipulation of *belle* and *row* expression (**Supplemental file S1**). Down regulation of *belle* expression by three different RNAi lines did not change ectopic wing formation caused by *G9a* overexpression (**Figure 2, bars 3-5**). Heterozygous deletion of *belle* by the p-element excision allele in combination to *G9a* overexpression gave a slight tendency towards more severe ectopic vein formation (**Figure 2, bar 6)**, but this trend is not sufficient to conclude positively about a genetic interaction. Overexpression of *belle* using the enhancer trap line (**Figure 2, bar 7**) showed no modulation of the *G9a* extra vein phenotype. Additionally, this phenotype is not opposite to the phenotype of lines that reduce *belle* expression. Modulation of *belle* expression alone (in absence of *G9a* overexpression) in the fly wing causes no phenotype in vein nor wing formation (not shown). Thus, these results indicate that *G9a* and *belle* do not genetically interact in the developing *Drosophila* wing.

Next, the genetic interaction between *G9a* and *row* was investigated by modulation of *row* expression in combination with *G9a* overexpression in the fly wing. Three different lines that reduce *row* expression were assessed. First, RNAi-mediated downregulation of *row* did not change the vein phenotype induced by *G9a* overexpression in two RNAi-lines (**Figure 2, bar 8-9**). Next, reduced expression of *row* by the p-element excision allele in combination with *G9a* overexpression caused a modest increase of ectopic vein formation (**Figure 2, bar 10**), but this trend is also not large enough to conclude a genetic interaction. Modulation of *row* expression alone (in absence of *G9a* overexpression) caused no phenotype in the fly wing (not shown). Taken together, these results suggest that *G9a* and *row* do not genetically interact.

In addition to these initial findings, more patients were identified with mutations in *POGZ* [13] and *DDX3X* [14], but without clinical features reminiscent of KSS. These latest clinical findings thus disqualified DDX3X and POGZ as members of the EHMT1 network, but confirmed the results of the genetic interaction experiments. In addition, these results show the power of genetic interaction studies to validate networks of genes that are associated with overlapping clinical characteristics.

After elimination of DDX3X and POGZ as potential components of an EHMT1-related module, we sought to identify additional components through identification of novel *Drosophila* G9a interaction partners using affinity purification followed by mass spectrometry. To do this, we used a *Drosophila* strain that expresses a "tagged" G9a protein [5].

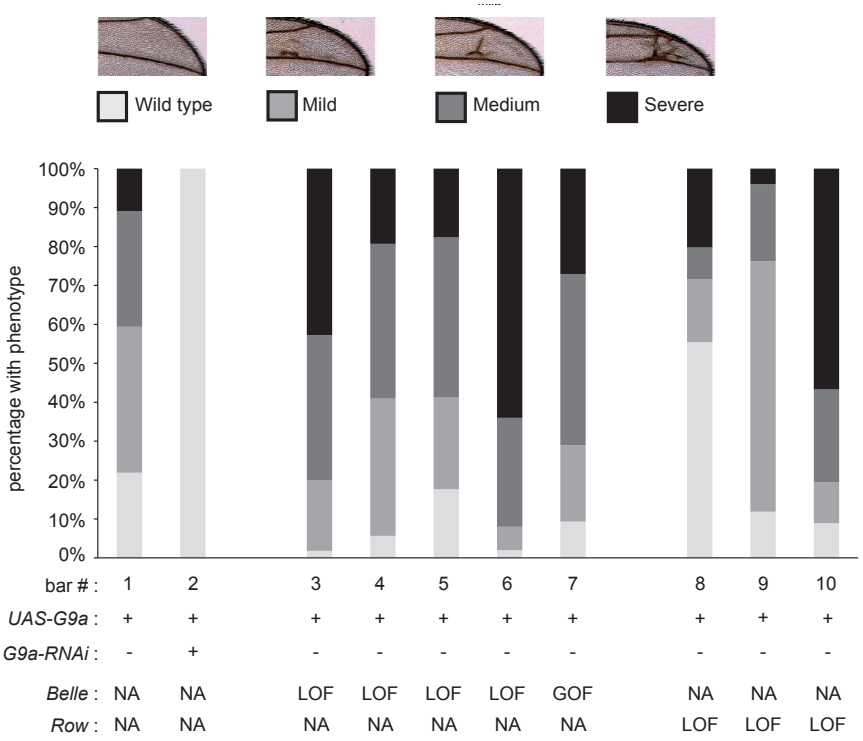


Figure 2: *Drosophila* orthologs of *DDX3X* and *POGZ* do not genetically interact with *G9a*.

Bar graph showing ectopic vein formation due to *G9a* overexpression in the wing in combination with various genetic elements. *G9a* overexpression causes a wing vein phenotype (bar 1). This phenotype is rescued by RNA interference (RNAi) against *G9a* (bar 2). Overexpression of *G9a* in combination with three individual Belle RNAi transgenes caused no increase in vein formation in the wing (bar 3-5). The heterozygous p-element allele *BelleL4740* in combination with *G9a* overexpression caused a modest increase of the *G9a* induced wing vein phenotype (bar 6), however, this is not convincing enough to conclude enhancement of the phenotype. The enhancer trap allele *BelleEY08943* in combination with *G9a* overexpression did not change the wing vein phenotype (bar 7). Reduced levels of row by two individual RNAi lines did not change the wing vein phenotype in combination with *G9a* overexpression (bar 8-9). The allele *rowBG02781* did not change the wing vein phenotype due to *G9a* overexpression (bar 10). All transgenes alone (without *UAS-G9a*) caused no aberrant vein formation (not shown). Genotypes are described in supplemental file S1. + indicates present, - indicates absent. LOF = loss of function, GOF = gain of function, NA indicates not applicable. Examples of severety of *G9a*-induced ectopic wing vein formation between veins L2 and L3 is variable in severity and can be quantified accordingly into wild-type, mild, medium, and strong. Upper panels of wings taken from Kramer *et al*. (2012) [4] with permission.

The G9a fusion protein is expressed in fly heads

In order to find physical interaction partners of G9a, affinity purification followed by mass spectrometry was performed. *Drosophila* fosmid lines are generated by using ϕC31-mediated

site-specific transgenesis [15] of a large genomic piece, including potential cis-regulatory DNA elements, using the Flyfos vector [5]. This approach resulted in a tagged version of the *G9a* gene in the *Attp2* site of the genome as additional copy to the natural occurring allele and is hereafter called "tagged G9a fosmid line" (for genoype see supplemental file S1). This approach ensures a stable and endogenous expression pattern and level of the G9a protein with a "multitag" [5], consisting of 2xTY1 amino acid sequence, super folder green fluorescent protein (sGFP), tobacco etch virus (TEV), and a biotin ligase recognition peptide (BLRP) (**Figure 3A**). The G9a protein is predicted to be 181 kDa and the tag 38 kDa by the Expasy program "compute pI/Mw" [16]. We first validated the expression of the tagged G9a protein in adult fly heads by Western blot, using an antibody directed against TY1 (**Figure 3B**) and two commercial generated antibodies against the sGFP epitope (**Figure 3C and 3D**). This analysis consistently revealed one band of the expected size (219 kDa) in two independent tagged G9a fosmid lines. As expected, this band is absent in the extract from wild type fly heads (**Figure 3B, C, D**). Taken together, these data show that the G9a fusion protein is expressed in fly heads and can be detected by Western blot analysis using the epitopes included in the tag.

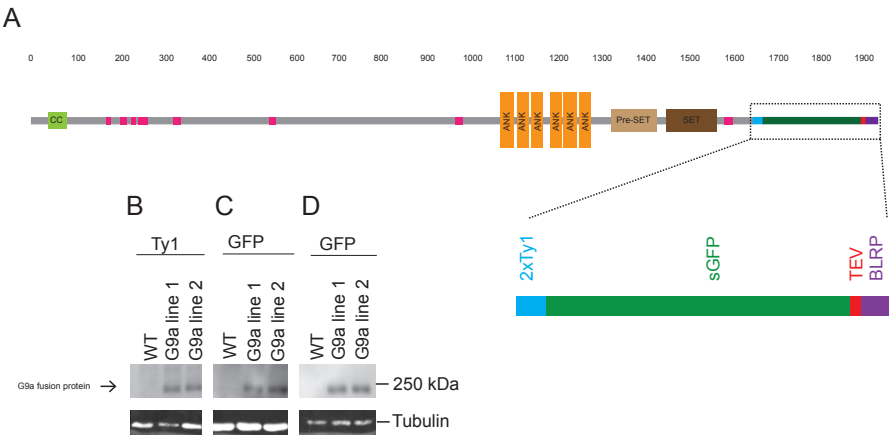


Figure 3: The G9a fusion protein is expressed in heads of the tagged G9a fosmid line

(A) Schematic protein motif overview and close up of the multi-tag of G9a fusion protein based on SMART analysis [18] of evolutionarily conserved protein domains. Domains are indicated with colours on the grey backbone: light green: coiled coil domain, pink: low complexity, orange: ankyrin domain, light brown: pre-SET domain, dark brown: SET domain, light bue: 2xTy1 peptide, dark green: sGFP, red: TEV, purple: BLRP, numbers indicate amino acid positions. (B-D) Western blot analysis visualizing the G9a fusion protein using antibodies from Diagenode directed against Ty1 (B), from Thermo Fischer (C), and from Abcam (D) directed against GFP. The following genotype was used: *FlyFos028501(pRedFlp-Hgr)(G9a[24641]::2xTY1-SGFP-3xFLAG)dFRT*

Abbreviations: cc: coiled coil, SET: Su(var)3-9, Enhancer-of-zeste, Trithorax, sGFP: super folder green fluorescent protein, TEV: tobacco etch virus, BLRP: biotin ligase recognition peptide

The G9a fusion protein can be purified from a protein extract

Next, we aimed to affinity purify the tagged G9a protein from a total fly head protein extract of tagged G9a fosmid line 1. Any interaction partners of G9a will be co-immunoprecipitated since this reaction is under physiological, non-stringent conditions and potential physical interactors can be subsequently identified using mass spectrometry. Briefly, the lysate is centrifuged to exclude intact cells and other debris (pellet) from the input. Next, this input is incubated with agarose beads coated with antibodies directed against super folder GFP (Chromotec). After the immunoprecipitation reaction, the beads bound by proteins are centrifugated to separate the bound from the unbound fraction. All fractions were loaded on a western blot and the blot was incubated with antibodies directed against GFP (Abcam #Ab290) and visualized using chemiluminescence (**Figure 4**). Indeed, we found G9a fusion protein in the input of the immunoprecipitation experiment (**Figure 4, lane 3**) and in the bound fraction (**Figure 4, lane 1**), confirming that the G9a fusion protein is bound to the antibody-coated beads. Next, this purified fraction was subjected to mass spectrometry to identify proteins that are physically associated to G9a.

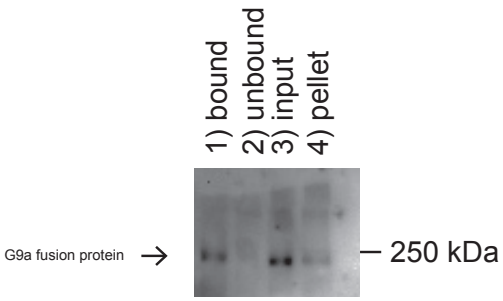


Figure 4: The G9a fusion protein can be purified in a protein extract from heads of the tagged G9a *Drosophila* fosmid line.

Western blot showing tagged G9a protein throughout different steps of the immunoprecipitation protocol visualized by an antibody directed against GFP (Abcam) in combination with chemiluminescence. G9a fusion protein can be detected in the input (lane 3), the bound fraction (lane 1), and the pellet (lane 4), but not in the unbound fraction (lane 2). The following genotype was analysed: FlyFos028501(pRedFlp-Hgr)(G9a[24641]::2xTY1-SGFP-3xFLAG)dFRT.

The zinc finger protein CG9932 is a novel binding partner of G9a

To address which proteins are physically associated to G9a, the immunoprecipitated fraction was subjected to mass spectrometry analysis. Briefly, lysates were subjected to single step GFP-affinity enrichments (ipGFP) in triplicate. In addition to these samples, the same lysates were incubated with beads that do not contain the GFP antibody (control). The precipitated proteins were then subjected to on-bead trypsin digestion after which peptide mixtures were analysed by nanoLC-MS/MS. Raw data processing was done using MAXQuant software [17] and the obtained normalized, log2 transformed, label free quantification (LFQ) intensities were used to determine outliers using a two-tailed t-test ($p<0.05$) (**Table 1**). Multiple testing correction was applied by using a permutation-based false discovery rate (FDR) method on a Perseus software platform. The ratio of mean LFQ values from GFP immunoprecipitated samples versus control were plotted against the FDR values (**Figure 5**) per protein. FDR and fold change (FC) cut-offs were determined based on a custom made R-script (Vermeulen lab). Since this method of protein identification also includes the bait itself, G9a, serves as a internal positive control and was significantly enriched among other proteins represented in the right upper corner of the graph (**Figure 5**). Other peptides that were identified were Q9VZQ7 and sallimus (sls). However, another co-immunoprecipitation experiment with an unrelated bait protein also identified these peptides as interaction partners (not shown), suggesting that these proteins were not specific G9a interactors but contaminants. The peptide Q9VK42 was identified as significant interaction partner, and this protein is encoded by the so far uncharacterized *CG9932* gene. Analysis of evolutionary conserved protein domains by SMART analysis [18] revealed that CG9932 contains sixteen C_2H_2 zinc finger domains and a coiled coil domain. Taken together, an uncharacterized zinc finger protein has been identified as novel interaction partner of G9a in *Drosophila* fly heads. Next, we reasoned that since G9a and CG9932 are physically associated, we also tested a potential genetic interaction.

Table 1: Significant hits of G9a interaction partners identified by mass spectrometry.

Shown are the normalized number of counts per peptide in the tagged G9a fosmid line in triplicates (GFP) and control line in triplicates (non). Additionally, -log 10 transformation of the *p*-value and log2 transformation of average GFP/control signal is shown.

GFP_1	GFP_2	GFP_3	NON_1	NON_2	NON_3	-Log10 (*P*-value)	2Log (GFP/ control)	peptide names
1	1	1	2	2	2	4.961	-6.405	Sgs7
47	48	41	0	0	0	2.962	8.033	G9a
4	4	4	1	2	1	3.697	7.555	sls
17	17	15	0	0	0	2.990	5.251	Q9VK42
3	4	6	1	1	2	3.699	5.866	Q9VZQ7

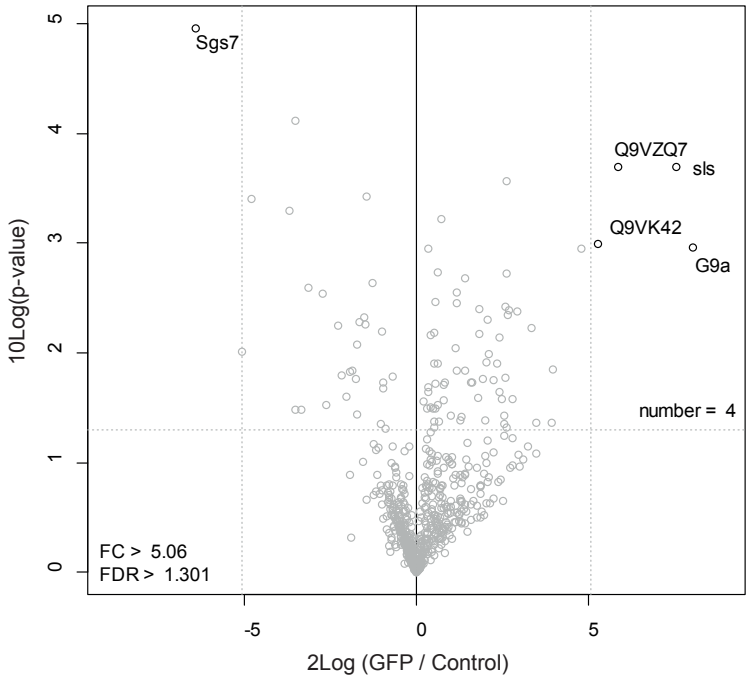


Figure 5: Identification of Q9VK42 as interaction partner of G9a.

Vulcanoplot showing significant binding partners of tagged G9a protein based on affinity purification followed by detection by mass spectrometry. The mean Log2 transformed LFQ intensity of the GFP pull-down over the control is plotted against the log10 FDR (p-value). Cut-offs (grey dashed lines) of Log2 transformed mean LFQ values and FDR are determined by a custom R-script based on the variability of the data points. The proteins in the right upper corner represent the bait and putative interactors.

Genetic interaction between G9a and CG9932

Besides the physical association between G9a and CG9932, a potential genetic interaction was investigated in the same way as described above (**Figure 2**). We first confirmed the extra wing vein formation due to overexpression of *G9a* in the fly wing in two different genetic backgrounds (**Figure 6A, bars 1-2**). Next, modulation of the *G9a* induced wing vein phenotype was investigated upon RNAi mediated down regulation of *CG9932* in the fly wing. We used three different lines in two different genetic backgrounds, corresponding to the background of the control lines, respectively. No major modification of the *G9a*-induced phenotype was observed in combination with *UAS-CG9932(IR1)* (**Figure 6A, bar 3**) while the RNAi alone caused no wing phenotype (**Figure 6B, bar 3**). However, the two RNAi lines alone in genetic background 2 (*UAS-CG9932(IR2)* and *UAS-CG9932(IR3)*) gave a very severe wing phenotype and this precludes interpretation of possible genetic interaction (**Figure**

6B, bar 4-5). Thus, since two out of three RNAi lines alone (without G9a overexpression) give such severe wing phenotypes, it was not possible to conclude on a genetic interaction between *G9a* and *CG9932*.

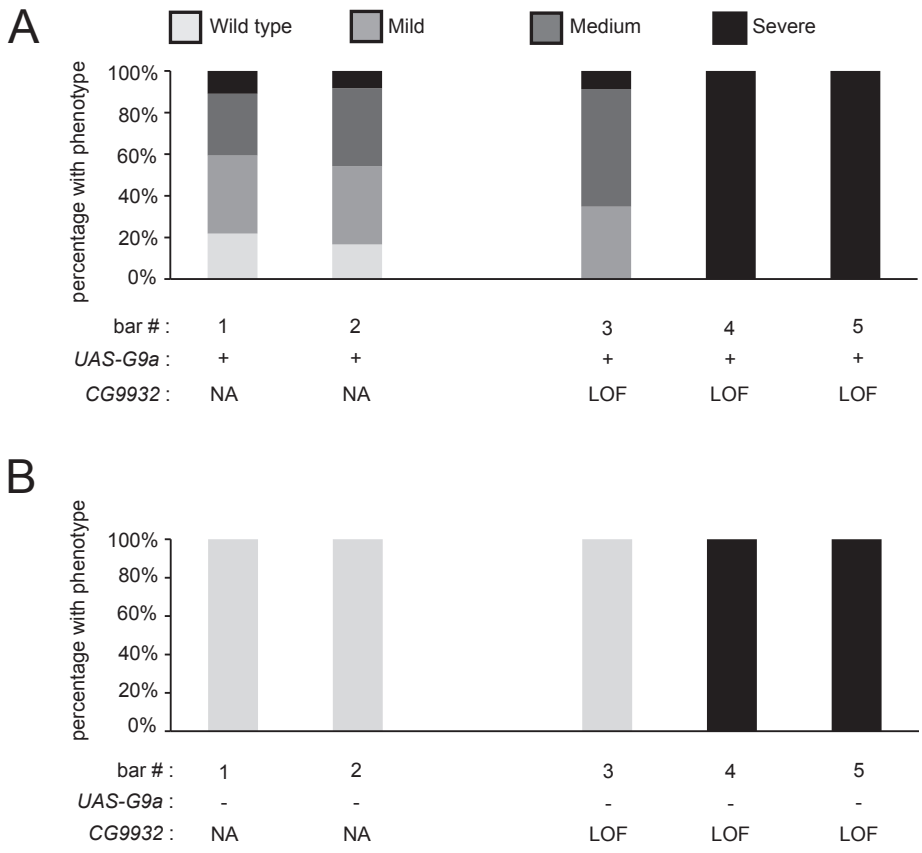


Figure 6: Lack of evidence to conclude on the genetic interaction between G9a and CG9932.

(A) Bar graph showing ectopic vein formation due to *G9a* overexpression in the wing in combination with various genetic elements. *G9a* overexpression causes a wing vein phenotype in two genetic backgrounds (bar 1 and 2). No difference in wing phenotype was observed by modulating *CG9932* expression using *UAS-CG9932(IR1) (bar 3)*. The phenotype of *UAS-CG9932(IR2)* and *UAS-CG9932(IR3)* in combination with *UAS-G9a* was very severe in all wings assessed (bar 4 and 5). (B) Wing phenotypes with the *row* transgenes alone (no G9a overexpression) revealed wild type wings in the genetic backgrounds (bar 1 and 2) and in the *UAS*-CG9932(IR1) alone (bar 3). *UAS-CG9932(IR2)* and *UAS-CG9932(IR3)* alone showed a very severe effect (bar 4 and 5). Genotypes are described in supplemental file S1. + indicates present, - indicates absent. LOF = loss of function, GOF = gain of function.

CG9932 is evolutionary conserved to human REST and zinc finger proteins

Next, we reasoned that CG9932 might have evolutionary conservation with human proteins. In order to identify homologous proteins to CG9932, the evolutionary conservation was investigated using the protein sequence of CG9932. This *Drosophila* sequence was analysed using a protein BLAST on the human proteome by the uniprot website (http://www.uniprot.org). 123 high quality and non-redundant protein sequences were identified with identity scores between 21% and 33%, mainly based on the zinc finger motifs of CG9932 (**Table S1**). The top hits on this list (**Table 2**) include RE1-silencing transcription factor (REST), and many zinc finger proteins with **RE-1-silencing transcription factor** showing the lowest E-value.

Table 2: Top-5 of human proteins with orthology to *Drosophila* CG9932 sorted on E-value.

Peptide	Gene name	Identity score	E-value
Q13127	RE1-silencing transcription factor	33%	2.70E-12
Q96JM2-2	Isoform 2 of Zinc finger protein 462	22%	1.30E-10
Q96JM2	Zinc finger protein 462	22%	1.40E-10
Q96JM2-3	Isoform 3 of Zinc finger protein 462	22%	1.40E-10
Q9P243-4	Isoform 4 of Zinc finger protein ZFAT	33%	4.70E-10

Discussion

Gene transcription is a tightly regulated process in which multi-protein complexes allow a cell to respond on environmental changes. Mutations in genes encoding epigenetic factors that regulate these transcriptional changes are implicated in ID [19]. Here we identified two genes (*POGZ* and *DDX3X*) involved in chromatin and gene transcription originally implicated in KSS by exome sequencing. However, after our initial findings, more patients were identified with mutations in *POGZ* [13] and *DDX3X* [14] but without phenotypic features suggestive of KSS. Additionally, no genetic interaction between the *Drosophila* orthologs of *EHMT1* with *POGZ* and *DDX3X* was identified, which is consistent with these clinical findings. For these reasons, the two proteins are not likely to be involved in an EHMT1 protein network. However, the absence of genetic interactions between *G9a* and *belle* and *row*, together with a clinical spectrum different from KSS strengthens the validity of the wing interaction assay. It confirms that the assay probably only identifies members of the EHMT1 network [4].

The second approach, protein-protein interaction studies to enlarge the network of EHMT1 underlying KS and KSS identified CG9932 as interaction partner of *Drosophila* G9a. While other zinc finger interactors have been identified in mice and human such as ZNF644 and WIZ [20-22], this is the first identified zinc finger interactor of G9a in *Drosophila*. However, no additional known protein interactors such as CoREST or HDAC proteins were

co-immunoprecipitated in this experiment. This lack of other interactors could be due to tissue specificity of protein-protein interactions or can be due to technical limitations. Homology searches revealed REST and ZNF462 as human orthologs with low E-values. It could be that mutations in REST and ZNF462 are among the elusive genes underlying the unexplained cases of KSS. Taken together, CG9932 is a novel binding partner of G9a in *Drosophila*. However, the functionality of CG9932 and the exact interaction with G9a remains to be elucidated.

Potential future experiments that could follow up on the findings of the approach using the *Drosophila* tagged fosmid line would first be a confirmation of the physical interaction between G9a and CG9932. This could be done by a reverse affinity purification reaction in which CG9932 is tagged. After immunoprecipitation, G9a should be confirmed as interactor in order strengthen the evidence of a physical interaction further. Next, besides a genetic interaction in the wing, other tissue could be more suitable for this interaction specifically. The *Drosophila* eye is another well known tissue to test such an interaction [12]. Furthermore, yeast-two-hybrid experiments could also be used to verify an interaction between CG9932 and G9a. Additionally, CG9932 together with G9a can be investigated for co-regulatory properties to identify the exact function of this protein-protein interaction for instance in the process of transcription and modification of the epigenetic landscape. Lastly, CG9932 can be studied in disease-relevant processes such as learning and/or memory for instance using courtship conditioning (**chapter 2**) [23]. Taken together, understanding the neuronal regulatory mechanisms controlled by the G9a protein network in relation to KS and KSS is essential. These two approaches could be used further to detect new genes and confirm members of this network.

REST-related ID disorders

Paralogs EHMT1 and EHMT2 are reported to be present as a heterodimeric protein complex [24] and are associated with several zinc finger proteins [20, 25]. For instance REST, a major factor in neuronal development [26-28], is a direct interaction partner of EHMT2 [29]. REST can bind the 21bp RE-1 silencing motif and recruit the co-repressor complexes CoREST and SIN3 [30]. Both complexes contain histone deacetylases (HDAC) 1 and 2 and catalyse de-acetylation [31] and demethylation [32] of histone H3 at genes important for neuronal functioning and thereby causing initial transcriptional silencing. Silencing of transcription at REST binding sites is prolonged through DNA methylation by methyl CpG binding protein 2 (MeCP2) [33]. Thus, these proteins have important functions in neuronal functioning. In *Drosophila*, panneuronal loss of *CoREST* results in a reduced olfactory memory [34]. Thus, EHMT proteins seem to be present in transcription repression complexes important for neuronal functioning.

Several of these proteins involved in REST-mediated neuronal gene repression are also implicated in ID. REST has been isolated in complex with SMCX (JARID1C/KDM5C), a histone demethylase implicated in X-linked ID and epilepsy [35]. Haploinsufficiency of *SIN3A* causes mild ID and autism spectrum disease [36]. MECP2, which has no counterpart in *Drosophila*, is a well-studied protein in brain function and is implicated in classic Rett syndrome [37]. These syndromes have several aspects in common, including ID, speech delay, seizures, and neurological regression [38]. Taken together, REST related transcriptional plasticity is crucial for normal neuronal functioning, however is not included in the EHMT1-network underlying Kleefstra syndrome and KSS. *Drosophila* experiments using modulation of expression of CG9932 and G9a could elucidate fundamental epigenetic processes underlying ID.

Materials and methods

Identification of variants in *POGZ* and *DDX3X*

Individuals were ascertained through family-based exome sequencing (ES) in individuals with unexplained ID or developmental delay in a diagnostic setting in the department of Human genetics at the Radboudumc. Exome sequencing in the probands and their unaffected parents and data analysis were performed as previously described [7]. These patients have been part of two previous studies in which large cohorts of individuals with *DDX3X* and *POGZ* mutations were presented [13, 14].

Fly stocks and maintenance

All flies were reared on standard medium (cornmeal/sugar/yeast) at 25 °C and 50% humidity with a 12-h light/dark cycle. Fly stocks were obtained from the Bloomington *Drosophila* Stock Center (Indiana University) and the Vienna *Drosophila* RNAi Center (Institute for Molecular Pathology) [39].

For belle we used stocks BL35302, BL28049, VDRC6299, BL10222, BL19945. For row we used BL25971, VDRC28196, BL13910. For CG9932 we used VDRC45686, VDRC45687, VDRC107846. For G9a we used VDRC110662, VDRC25473, VDRC25474. As controls we used VDRC60000, VDRC60100, BL36303. *UAS-G9a* flies were described previously [8]. *MS1096-Gal4* and *MS1096-Gal4/FM7*d;*UAS-G9a* were described previously [4]. G9a fosmid lines 1 and 2 were a gift from Dr. P. Tomancak and Dr. M. Zuberova (Max Planck Institute of Molecular Cell Biology and Genetics, Dresden Germany) and genotype is described in supplemental file S1. A Nijmegen specific control line was used for western blot experiments.

Genetic interaction assay

We performed genetic interaction studies in *Drosophila* by using the UAS/Gal4 ectopic expression system [40] to induce overexpression or RNAi-mediated knockdown of gene orthologs implicated in KSS. Gene orthologs were identified using the reverse BLAST method [10] and analysis of the Treefam database (Release 9, March 2013) [11]. Females of the genotypes *MS1096-Gal4* and *MS1096-Gal4/FM7d;UAS-G9a* were crossed to males from fly stocks of Belle, Row and CG9932 as indicated in Table 1. Wing phenotypes were examined in female progeny and scored for severity as described previously [4]. Genotypes of the progeny analysed are described in supplemental file S1.

Western blot analysis

G9a fusion protein line 1 and 2 and wild type flies (males and females) were aged between 1-5 days. The following genotype was analysed: *FlyFos028501(pRedFlp-Hgr)(G9a[24641]::2xTY1-SGFP-3xFLAG)dFRT.* Aged flies were snap frozen in liquid nitrogen. Frozen fly heads were harvested by vortexing and separated from other body parts through a series of standard laboratory sieves. Around 50 µl of frozen fly heads was crushed in lysisbuffer (10mM Tris/ HCl (pH=7.5), 150 mM NaCl, 0.5mM EDTA, 0.5% NP40) supplemented with 1x protease inhibitor cocktail (PIC) (Sigma) using Readyprep mini grinders (Bio-rad). Proteins in the extracts were separated using an 8% acrylamide gel with the Bio-rad electrophoresis system according to standard operating procedures. Proteins were transferred to nitrocellulose blotting paper (Sigma) using the Bio-rad Mini Trans-Blot cell. Blots were probed with mouse-anti-Ty1 at a 1/2000 dilution (Diagenode, catalog number C15200054), rabbit-anti GFP at a 1/1000 dilution (Abcam, catalog number Ab290), or rabbit-anti GFP at a 1/1000 dilution (Thermofisher, catalog number A-11122). Proteins were visualized using a goat-anti-mouse-HRP antibody or goat-anti-rabbit-HRP antibody at a 1/3000 dilution (Invitrogen) and the supersignal west pico chemiluminescent substrate (Pierce) on a ChemiDoc MP system (Bio-rad).

As a control for equal loading of samples the same blots were probed with a rat-anti-α-tubulin antibody (AbD Serotec) at a 1/1000 dilution and visualized using a goat-ant-rati-IRdye800 antibody (Invitrogen) on an Odyssey infrared imager (LI-COR Biosciences).

Immunoprecipitation of G9a fusion protein

Flies of the G9a fusion protein line 1 were aged between 1-5 days. The following genotype was analysed: *FlyFos028501(pRedFlp-Hgr)(G9a[24641]::2xTY1-SGFP-3xFLAG)dFRT*. Around 50 µl of frozen fly heads was crushed in lysis buffer (10mM Tris/HCl (pH=7.5), 150 mM NaCl, 0.5mM EDTA, 0.5% NP40) supplemented with 1x protease inhibitor cocktail (PIC) (Sigma) using Readyprep mini grinders (Bio-rad). Crude protein extracts were vigorously vortexed

for ten minutes while maintained on ice. Lysates were centrifuged at 14.000 rpm for 10 minutes at 4 °C and the supernatant was transferred to a new vial. The remaining pellet was resuspended in 300 µl dilution buffer (10 mM Tris/HCl (pH=7.5), 150 mM NaCl, 0.5 mM EDTA) and referred to as pellet. The supernatant (cell lysate) was adjusted to 500 µl with dilution buffer supplemented with 1x protease inhibitor cocktail (Sigma) and referred to as input. 200 µl of the cell lysate with 20 µl of equilibrated GFP-Trap agarose beads (Chromotec) and 1x PIC was used in a total volume of 500 µl and incubated over night at 4 °C on a gently rotating wheel. The next day, bead-immuno-protein complexes were isolated by centrifugation (2700x g for 2 minutes at 4 °C). Supernatant was referred to as 'unbound', and bound proteins were eluted from the beads in 25 µl washing buffer (10 mM Tris/Cl (pH=7.5), 150 mM NaCl, 0.5 mM EDTA) supplemented with 1x PIC and Western blot loading dye. All fractions of the immunoprecipiation were subjected to Western blot analysis (described above) using antibodies directed to GFP (Abcam, Ab290).

Mass spectrometry

Flies of the G9a fusion protein line 1 were aged between 1-5 days. The following genotype was analysed: *FlyFos028501(pRedFlp-Hgr)(G9a[24641]::2xTY1-SGFP-3xFLAG)dFRT*. Around 50 µl of frozen fly heads were lysed by adding RIPA lysis buffer (150mM NaCl; 50mM Tris pH 8.0; 1% NP-40; 1mM DTT supplemented with 1x protease inhibitor cocktail (PIC)). To help the lysis, heads were dounced with a type A pestle. The extract was cleared by centrifugation at 21,100 x g for 15 minutes at 4°C. The label free pull down was performed in triplicate by adding 7,5 µl GFP beads (chromotek, GFP-Trap) or 7,5 µl non-GFP (Chromotek, bab-20) per immunoprecipitation. The following mixture was used per immunoprecipitation: 100 µl extract, 300ul RIPA lysis buffer supplemented with Ethidiumbromide to a final concentration of 50 µg/µl. Samples were incubated for 90min at 4°C in order to let the protein complexes bind the antibody coated beads. The beads were then washed three times with 1ml RIPA lysis buffer and 2 times washed with 1ml PBS. Precipitated proteins were subjected to on-bead trypsin digestion and nanoLC-MS/MS mass spec analysis as described before [41]. The raw mass spectrometry data were analyzed using the MAXQuant software version 1.3.0.5 [17] to filter for contaminants, reverse hits, and number of peptides (>1) and to retrieve normalized log2 transformed LFQ values. These values were subjected to a two tailed t-test ($p<0.05$) to exclude outliers and the mean normalized LFQ values were used to calculate the ratio of LFQ (GFP)/ LFQ (control). False discovery rate (FDR) analysis was performed using a custom R-script based on a permutation test. The same R-script calculated the cut-off values for ratio (GFP/control) and FDR and was based on the variability of the dataset (**Table 2**). All amino acid sequences were searched against the *Drosophila* proteome (Uniprot, February 2012).

Identification the orthologs proteins of CG9932 in human

The protein sequence of CG9932 (Q9VK42) was downloaded from the Uniprot database (The Uniprot consortium, 2014). This sequence was used as input for protein BLAST version 2.2.29+ with BLOSUM62 as matrix. E-value threshold was set bigger than 10 and only curator-assessed orthologs were used as output to focus on the high confidence hits.

Acknowledgements

We would like to thank Dave Verkaik for technical assistance.

References

1. Kleefstra T, Brunner HG, Amiel J, Oudakker AR, Nillesen WM, Magee A, et al. Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. American journal of human genetics. 2006;79(2):370-7. doi: 10.1086/505693. PubMed PMID: 16826528; PubMed Central PMCID: PMC1559478.
2. Stewart DR, Kleefstra T. The chromosome 9q subtelomere deletion syndrome. American journal of medical genetics Part C, Seminars in medical genetics. 2007;145C(4):383-92. doi: 10.1002/ajmg.c.30148. PubMed PMID: 17910072.
3. Vermeulen K, de Boer A, Janzing JGE, Koolen DA, Ockeloen CW, Willemsen MH, et al. Adaptive and maladaptive functioning in Kleefstra syndrome compared to other rare genetic disorders with intellectual disabilities. American journal of medical genetics Part A. 2017. doi: 10.1002/ajmg.a.38280. PubMed PMID: 28498556.
4. Kleefstra T, Kramer JM, Neveling K, Willemsen MH, Koemans TS, Vissers LE, et al. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. American journal of human genetics. 2012;91(1):73-82. doi: 10.1016/j.ajhg.2012.05.003. PubMed PMID: 22726846; PubMed Central PMCID: PMC3397275.
5. Ejsmont RK, Sarov M, Winkler S, Lipinski KA, Tomancak P. A toolkit for high-throughput, cross-species gene engineering in *Drosophila*. Nature methods. 2009;6(6):435-7. doi: 10.1038/nmeth.1334. PubMed PMID: 19465918.
6. Langer CC, Ejsmont RK, Schonbauer C, Schnorrer F, Tomancak P. In vivo RNAi rescue in *Drosophila melanogaster* with genomic transgenes from *Drosophila* pseudoobscura. PloS one. 2010;5(1):e8928. doi: 10.1371/journal.pone.0008928. PubMed PMID: 20126626; PubMed Central PMCID: PMC2812509.
7. de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. The New England journal of medicine. 2012;367(20):1921-9. doi: 10.1056/NEJMoa1206524. PubMed PMID: 23033978.
8. Kramer JM, Kochinke K, Oortveld MA, Marks H, Kramer D, de Jong EK, et al. Epigenetic regulation of learning and memory by *Drosophila* EHMT/G9a. PLoS biology. 2011;9(1):e1000569. doi: 10.1371/journal.pbio.1000569. PubMed PMID: 21245904; PubMed Central PMCID: PMC3014924.
9. Seum C, Bontron S, Reo E, Delattre M, Spierer P. *Drosophila* G9a is a nonessential gene. Genetics. 2007;177(3):1955-7. doi: 10.1534/genetics.107.078220. PubMed PMID: 18039887; PubMed Central PMCID: PMC2147950.
10. Inlow JK, Restifo LL. Molecular and comparative genetics of mental retardation. Genetics. 2004;166(2):835-81. PubMed PMID: 15020472; PubMed Central PMCID: PMC1470723.
11. Li H, Coghlan A, Ruan J, Coin LJ, Heriche JK, Osmotherly L, et al. TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic acids research. 2006;34(Database issue):D572-80. doi: 10.1093/nar/gkj118. PubMed PMID: 16381935; PubMed Central PMCID: PMC1347480.
12. Bier E. *Drosophila*, the golden bug, emerges as a tool for human genetics. Nature reviews Genetics. 2005;6(1):9-23. doi: 10.1038/nrg1503. PubMed PMID: 15630418.
13. Stessman HA, Willemsen MH, Fenckova M, Penn O, Hoischen A, Xiong B, et al. Disruption of POGZ Is Associated with Intellectual Disability and Autism Spectrum Disorders. American journal of human genetics. 2016;98(3):541-52. doi: 10.1016/j.ajhg.2016.02.004. PubMed PMID: 26942287.

14. Snijders Blok L, Madsen E, Juusola J, Gilissen C, Baralle D, Reijnders MR, et al. Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling. American journal of human genetics. 2015;97(2):343-52. doi: 10.1016/j.ajhg.2015.07.004. PubMed PMID: 26235985; PubMed Central PMCID: PMC4573244.
15. Groth AC, Fish M, Nusse R, Calos MP. Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. Genetics. 2004;166(4):1775-82. PubMed PMID: 15126397; PubMed Central PMCID: PMC1470814.
16. Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, et al. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis. 1993;14(10):1023-31. PubMed PMID: 8125050.
17. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature biotechnology. 2008;26(12):1367-72. doi: 10.1038/nbt.1511. PubMed PMID: 19029910.
18. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. Nucleic acids research. 2015;43(Database issue):D257-60. doi: 10.1093/nar/gku949. PubMed PMID: 25300481; PubMed Central PMCID: PMC4384020.
19. Kleefstra T, Schenck A, Kramer JM, van Bokhoven H. The genetics of cognitive epigenetics. Neuropharmacology. 2014;80:83-94. doi: 10.1016/j.neuropharm.2013.12.025. PubMed PMID: 24434855.
20. Shinkai Y, Tachibana M. H3K9 methyltransferase G9a and the related molecule GLP. Genes & development. 2011;25(8):781-8. doi: 10.1101/gad.2027411. PubMed PMID: 21498567; PubMed Central PMCID: PMC3078703.
21. Bian C, Chen Q, Yu X. The zinc finger proteins ZNF644 and WIZ regulate the G9a/GLP complex for gene repression. eLife. 2015;4. doi: 10.7554/eLife.05606. PubMed PMID: 25789554; PubMed Central PMCID: PMC4365668.
22. Ueda J, Tachibana M, Ikura T, Shinkai Y. Zinc finger protein Wiz links G9a/GLP histone methyltransferases to the co-repressor molecule CtBP. The Journal of biological chemistry. 2006;281(29):20120-8. doi: 10.1074/jbc.M603087200. PubMed PMID: 16702210.
23. Siegel RW, Hall JC. Conditioned responses in courtship behavior of normal and mutant *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America. 1979;76(7):3430-4. PubMed PMID: 16592682; PubMed Central PMCID: PMC383839.
24. Tachibana M, Ueda J, Fukuda M, Takeda N, Ohta T, Iwanari H, et al. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes & development. 2005;19(7):815-26. doi: 10.1101/gad.1284005. PubMed PMID: 15774718; PubMed Central PMCID: PMC1074319.
25. Olsen JB, Wong L, Deimling S, Miles A, Guo H, Li Y, et al. G9a and ZNF644 Physically Associate to Suppress Progenitor Gene Expression during Neurogenesis. Stem Cell Reports. 2016;7(3):454-70. doi: 10.1016/j.stemcr.2016.06.012. PubMed PMID: 27546533; PubMed Central PMCID: PMCPMC5031922.
26. Chong JA, Tapia-Ramirez J, Kim S, Toledo-Aral JJ, Zheng Y, Boutros MC, et al. REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. Cell. 1995;80(6):949-57. PubMed PMID: 7697725.
27. Bithell A. REST: transcriptional and epigenetic regulator. Epigenomics. 2011;3(1):47-58. doi: 10.2217/epi.10.76. PubMed PMID: 22126152.

28. Lunyak VV, Burgess R, Prefontaine GG, Nelson C, Sze SH, Chenoweth J, et al. Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. Science. 2002;298(5599):1747-52. doi: 10.1126/science.1076469. PubMed PMID: 12399542.
29. Ooi L, Wood IC. Chromatin crosstalk in development and disease: lessons from REST. Nature reviews Genetics. 2007;8(7):544-54. doi: 10.1038/nrg2100. PubMed PMID: 17572692.
30. Huang Y, Myers SJ, Dingledine R. Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes. Nature neuroscience. 1999;2(10):867-72. doi: 10.1038/13165. PubMed PMID: 10491605.
31. Naruse Y, Aoki T, Kojima T, Mori N. Neural restrictive silencer factor recruits mSin3 and histone deacetylase complex to repress neuron-specific target genes. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(24):13691-6. PubMed PMID: 10570134; PubMed Central PMCID: PMC24126.
32. Lee MG, Wynder C, Cooch N, Shiekhattar R. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature. 2005;437(7057):432-5. doi: 10.1038/nature04021. PubMed PMID: 16079794.
33. Ballas N, Grunseich C, Lu DD, Speh JC, Mandel G. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. Cell. 2005;121(4):645-57. doi: 10.1016/j.cell.2005.03.013. PubMed PMID: 15907476.
34. Walkinshaw E, Gai Y, Farkas C, Richter D, Nicholas E, Keleman K, et al. Identification of genes that promote or inhibit olfactory memory formation in *Drosophila*. Genetics. 2015;199(4):1173-82. doi: 10.1534/genetics.114.173575. PubMed PMID: 25644700; PubMed Central PMCID: PMC4391555.
35. Tahiliani M, Mei P, Fang R, Leonor T, Rutenberg M, Shimizu F, et al. The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. Nature. 2007;447(7144):601-5. doi: 10.1038/nature05823. PubMed PMID: 17468742.
36. Witteveen JS, Willemsen MH, Dombroski TC, van Bakel NH, Nillesen WM, van Hulten JA, et al. Haploinsufficiency of MeCP2-interacting transcriptional co-repressor SIN3A causes mild intellectual disability by affecting the development of cortical integrity. Nature genetics. 2016;48(8):877-87. doi: 10.1038/ng.3619. PubMed PMID: 27399968.
37. Lombardi LM, Baker SA, Zoghbi HY. MECP2 disorders: from the clinic to mice and back. The Journal of clinical investigation. 2015;125(8):2914-23. doi: 10.1172/JCI78167. PubMed PMID: 26237041; PubMed Central PMCID: PMC4563741.
38. Mullegama SV, Alaimo JT, Chen L, Elsea SH. Phenotypic and molecular convergence of 2q23.1 deletion syndrome with other neurodevelopmental syndromes associated with autism spectrum disorder. International journal of molecular sciences. 2015;16(4):7627-43. doi: 10.3390/ijms16047627. PubMed PMID: 25853262; PubMed Central PMCID: PMC4425039.
39. Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, et al. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. Nature. 2007;448(7150):151-6. doi: 10.1038/nature05954. PubMed PMID: 17625558.
40. Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development. 1993;118(2):401-15. PubMed PMID: 8223268.
41. Smits AH, Jansen PW, Poser I, Hyman AA, Vermeulen M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic acids research. 2013;41(1):e28. doi: 10.1093/nar/gks941. PubMed PMID: 23066101; PubMed Central PMCID: PMC3592467.

Supplemental File S1: genotypes used in this study

Figure 2:

W118/MS1096-Gal4;+/UAS-G9a (not shown)
y,w[1118]/MS1096-Gal4;;P{attP,y[+],w[3`]}/UAS-G9a (column 1)
y[1] v[1] /MS1096-Gal4;;P{y[+t7.7]=CaryP}attP2/UAS-G9a (not shown)
y,w[1118]/MS1096-Gal4;;P{KK100579}VIE-260B/UAS-*G9a* (column 2)
+/MS1096-Gal4;;y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.GL00205}attP2/UAS-G9a (column 3)
+/MS1096;;y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.JF02884}attP2/UAS-G9a (column 4)
w[1118]/ MS1096-Gal4;;P{GD1324}v6299/UAS-G9a (column 5)
y[1] w[67c23]/MS1096-Gal4;;P{w[+mC] y[+mDint2]=EPgy2}bel[EY08943]/UAS-G9a (column 6)
y[1] w[1118]/MS1096-Gal4;;P{w[+mC]=lacW}bel[L4740]/UAS-G9a (column 7)
y1 v1/MS1096-Gal4;;P{TRiP.JF01993}attP2/UAS-G9a (column 8)
w1118/MS1096;;P{GD12377}v28196/UAS-G9a (column 9)
w1118/MS1096-Gal4; +/P{GT1}rowBG02781;+/UAS-G9a (column 10)

W118/*MS1096-Gal4;+* (not shown)
y,w[1118]/MS1096-Gal4;;P{attP,y[+],w[3`]} (not shown)
y[1] v[1] /MS1096-Gal4;;P{y[+t7.7]=CaryP}attP2 (not shown)
+/MS1096-Gal4;;y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.GL00205}attP2 (not shown)
+/MS1096;;y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.JF02884}attP2 (not shown)
w[1118]/ MS1096-Gal4;;P{GD1324}v6299 (not shown)
y[1] w[1118]/MS1096-Gal4;;P{w[+mC]=lacW}bel[L4740] (not shown)
y[1] w[67c23]/MS1096-Gal4;;P{w[+mC] y[+mDint2]=EPgy2}bel[EY08943] (not shown)
y1 v1/MS1096-Gal4;;P{TRiP.JF01993}attP2 (not shown)
w1118/MS1096;;P{GD12377}v28196 (not shown)
w1118/MS1096-Gal4; +/P{GT1}rowBG02781; (not shown)

Figure 3

FlyFos028501(pRedFlp-Hgr)(G9a[24641]::2xTY1-SGFP-3xFLAG)dFRT

Figure 6A:

y,w[1118]/MS1096-Gal4;;P{attP,y[+],w[3`]}/UAS-G9a (column 1)
W118/MS1096-Gal4;+/UAS-*G9a* (column 2)
y,w[1118]/MS1096-Gal4;; P{KK103450}VIE-260B/UAS-G9a (column 3)
w[1118] P{GD14726}v45686;; UAS-G9a (column 4)
w[1118] P{GD14726}v45687;;UAS-G9a (column 5)

Figure 6B:
y,w[1118]/MS1096-Gal4;;P{attP,y[+],w[3`]} (column 1)
W118/MS1096-Gal4; (column 2)
y,w[1118]/MS1096-Gal4;; P{KK103450}VIE-260B (column 3)
w[1118] P{GD14726}v45686;; (column 4)
w[1118] P{GD14726}v45687;; (column 5)

Supplemental Table S2: Complete list of human proteins with orthology to *Drosophila* CG9932

Peptide	Gene name	Identity score	E-value
Q13127	RE1-silencing transcription factor	33%	2.70E-12
Q96JM2-2	Isoform 2 of Zinc finger protein 462	22%	1.30E-10
Q96JM2	Zinc finger protein 462	22%	1.40E-10
Q96JM2-3	Isoform 3 of Zinc finger protein 462	22%	1.40E-10
Q9P243-4	Isoform 4 of Zinc finger protein ZFAT	33%	4.70E-10
Q9P243-2	Isoform 2 of Zinc finger protein ZFAT	33%	4.70E-10
Q9P243	Zinc finger protein ZFAT	33%	4.80E-10
Q8N8E2-3	Isoform 3 of Zinc finger protein 513	29%	6.40E-10
Q8N8E2-2	Isoform 2 of Zinc finger protein 513	29%	4.10E-09
Q8N8E2	Zinc finger protein 513	29%	4.60E-09
Q9NPA5-2	Isoform 2 of Zinc finger protein 64 homolog, isoforms 1 and 2	33%	2.00E-08
Q9NTW7-4	Isoform 6 of Zinc finger protein 64 homolog, isoforms 3 and 4	33%	2.10E-08
Q9NPA5	Zinc finger protein 64 homolog, isoforms 1 and 2	33%	2.20E-08
P52746	Zinc finger protein 142	33%	5.00E-08
Q9NTW7-3	Isoform 5 of Zinc finger protein 64 homolog, isoforms 3 and 4	28%	5.50E-08
Q6IV72	Zinc finger protein 425	30%	6.70E-08
Q9NTW7	Zinc finger protein 64 homolog, isoforms 3 and 4	28%	8.10E-08
P17035-2	Isoform 2 of Zinc finger protein 28	29%	8.30E-08
P17035	Zinc finger protein 28	29%	8.60E-08
Q8NI51-11	Isoform 11 of Transcriptional repressor CTCFL	24%	9.70E-08
Q13127-4	Isoform 4 of RE1-silencing transcription factor	28%	1.00E-07
Q13127-3	Isoform 3 of RE1-silencing transcription factor	35%	1.20E-07
Q13127-2	Isoform 2 of RE1-silencing transcription factor	21%	1.40E-07

Supplemental Table S2: Complete list of human proteins with orthology to *Drosophila* CG9932 *(continued)*

Peptide	Gene name	Identity score	E-value
Q8NI51-6	Isoform 6 of Transcriptional repressor CTCFL	24%	1.60E-07
Q8NI51-5	Isoform 5 of Transcriptional repressor CTCFL	24%	1.70E-07
Q8NI51-10	Isoform 10 of Transcriptional repressor CTCFL	24%	1.90E-07
Q8NI51-2	Isoform 2 of Transcriptional repressor CTCFL	24%	2.20E-07
Q8NI51	Transcriptional repressor CTCFL	24%	2.50E-07
Q8NI51-3	Isoform 3 of Transcriptional repressor CTCFL	24%	2.50E-07
Q8NI51-7	Isoform 7 of Transcriptional repressor CTCFL	24%	2.50E-07
Q13422	DNA-binding protein Ikaros	25%	2.70E-07
P10074	Telomere zinc finger-associated protein	30%	3.30E-07
Q6AHZ1	Zinc finger protein 518A	27%	4.20E-07
Q08AN1	Zinc finger protein 616	28%	4.60E-07
Q96IR2	Zinc finger protein 845	34%	5.00E-07
Q68EA5	Zinc finger protein 57	28%	5.00E-07
Q6ZNG1	Zinc finger protein 600	31%	5.80E-07
Q4V348	Zinc finger protein 658B	31%	6.10E-07
Q9NTW7-2	Isoform 4 of Zinc finger protein 64 homolog, isoforms 3 and 4	27%	6.90E-07
Q8NI51-8	Isoform 8 of Transcriptional repressor CTCFL	25%	9.20E-07
P49711	Transcriptional repressor CTCF	23%	1.00E-06
Q9P243-3	Isoform 3 of Zinc finger protein ZFAT	37%	1.10E-06
Q9H5V7	Zinc finger protein Pegasus	30%	1.20E-06
Q9BSK1-2	Isoform 2 of Zinc finger protein 577	30%	1.20E-06
Q86V71	Zinc finger protein 429	28%	1.30E-06
Q8N782	Zinc finger protein 525	27%	1.30E-06
Q9BSK1	Zinc finger protein 577	30%	1.30E-06
Q86XN6-3	Isoform 3 of Zinc finger protein 761	28%	1.70E-06
Q86XN6-2	Isoform 2 of Zinc finger protein 761	28%	1.70E-06
Q86XN6	Zinc finger protein 761	28%	1.70E-06
Q9UID6	Zinc finger protein 639	28%	1.80E-06
Q5TYW1	Zinc finger protein 658	30%	2.00E-06
Q96PE6	Zinc finger imprinted 3	30%	2.30E-06
P17010-2	Isoform 2 of Zinc finger X-chromosomal protein	28%	3.50E-06

Supplemental Table S2: Complete list of human proteins with orthology to *Drosophila* CG9932 *(continued)*

Peptide	Gene name	Identity score	E-value
P17010	Zinc finger X-chromosomal protein	28%	4.10E-06
P17010-3	Isoform 3 of Zinc finger X-chromosomal protein	28%	4.10E-06
Q9BY31	Zinc finger protein 717	28%	4.20E-06
Q6U7Q0	Zinc finger protein 322	32%	4.50E-06
P08048-2	Isoform 2 of Zinc finger Y-chromosomal protein	29%	4.70E-06
Q0VGE8	Zinc finger protein 816	29%	4.90E-06
P08048-3	Isoform 3 of Zinc finger Y-chromosomal protein	29%	5.10E-06
Q5VIY5-2	Isoform 2 of Zinc finger protein 468	28%	5.20E-06
P08048	Zinc finger Y-chromosomal protein	29%	5.30E-06
Q8N4W9-2	Isoform 2 of Zinc finger protein 808	30%	5.40E-06
Q8N4W9	Zinc finger protein 808	30%	5.50E-06
Q5VIY5	Zinc finger protein 468	28%	5.60E-06
O94892	Zinc finger protein 432	30%	6.40E-06
Q9C0G0-2	Isoform 2 of Zinc finger protein 407	26%	6.40E-06
Q9C0G0	Zinc finger protein 407	26%	6.60E-06
Q5HY98	Zinc finger protein 766	31%	6.80E-06
Q8WXB4	Zinc finger protein 606	28%	6.90E-06
Q96JB3-2	Isoform 2 of Hypermethylated in cancer 2 protein	28%	8.00E-06
O43830	Zinc finger protein 73	26%	8.10E-06
Q96JB3	Hypermethylated in cancer 2 protein	28%	8.20E-06
Q8N823-2	Isoform 2 of Zinc finger protein 611	28%	8.30E-06
Q8N823	Zinc finger protein 611	28%	8.70E-06
Q9NPC7-3	Isoform 3 of Myoneurin	27%	9.50E-06
Q9ULM2	Zinc finger protein 490	29%	9.80E-06
Q8IZC7-2	Isoform 2 of Zinc finger protein 101	29%	1.00E-05
Q16587-4	Isoform 4 of Zinc finger protein 74	29%	1.00E-05
Q16587-2	Isoform 1 of Zinc finger protein 74	29%	1.00E-05
Q9NPC7-2	Isoform 2 of Myoneurin	27%	1.00E-05
Q9NPC7	Myoneurin	27%	1.10E-05
Q16587-3	Isoform 3 of Zinc finger protein 74	29%	1.10E-05
Q16587	Zinc finger protein 74	29%	1.10E-05

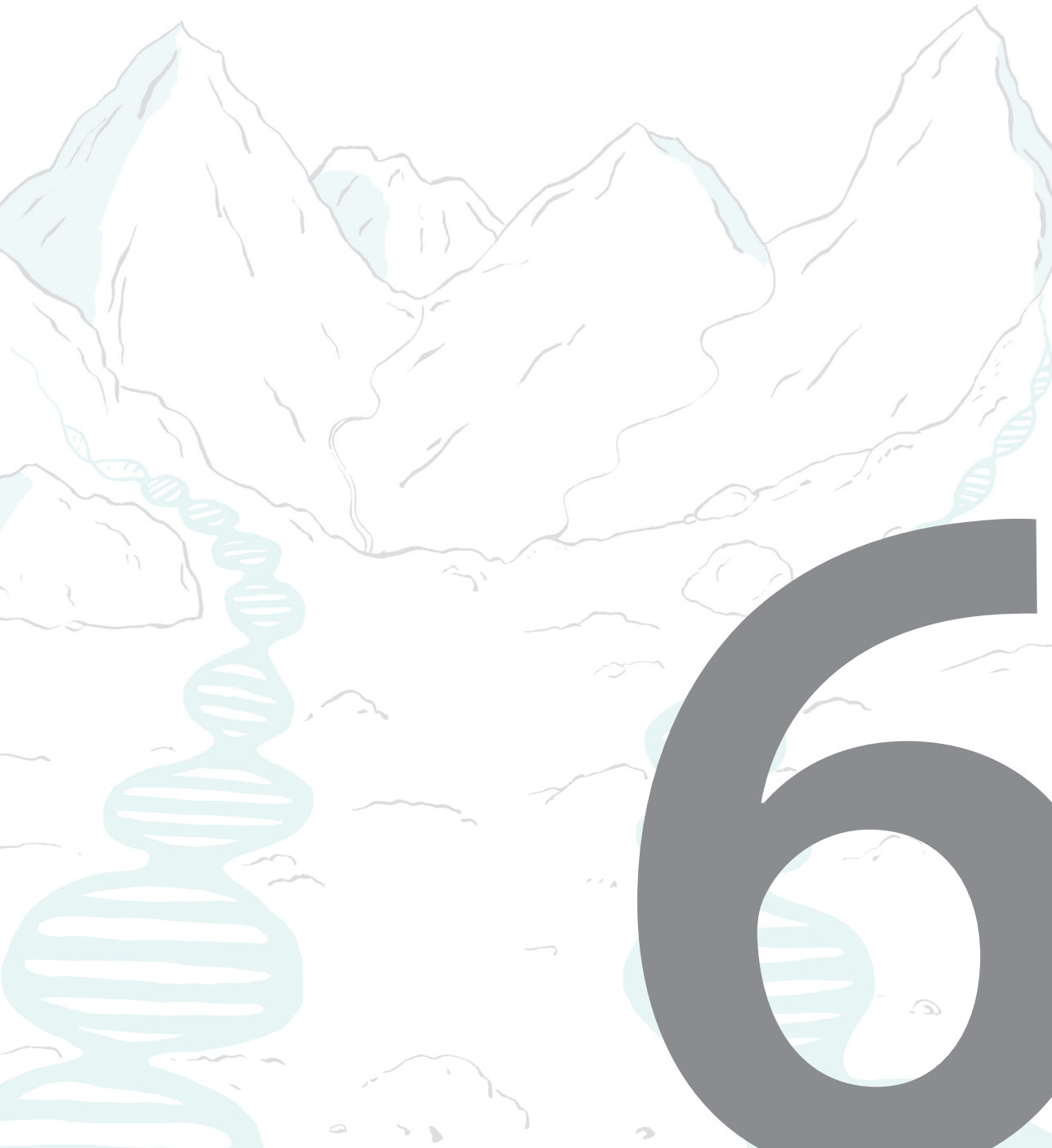
Supplemental Table S2: Complete list of human proteins with orthology to *Drosophila* CG9932 *(continued)*

Peptide	Gene name	Identity score	E-value
Q9NV72-2	Isoform 2 of Zinc finger protein 701	28%	1.20E-05
Q5JVG2-2	Isoform 2 of Zinc finger protein 484	30%	1.20E-05
Q5JVG2	Zinc finger protein 484	30%	1.20E-05
Q5JVG2-3	Isoform 3 of Zinc finger protein 484	30%	1.20E-05
Q96C28	Zinc finger protein 707	29%	1.30E-05
Q9NV72	Zinc finger protein 701	28%	1.30E-05
P15621-2	Isoform 2 of Zinc finger protein 44	29%	1.40E-05
P15621-3	Isoform 3 of Zinc finger protein 44	29%	1.40E-05
P15621	Zinc finger protein 44	29%	1.50E-05
Q8IZC7	Zinc finger protein 101	29%	1.50E-05
Q2VY69	Zinc finger protein 284	27%	1.80E-05
Q8N972	Zinc finger protein 709	29%	1.90E-05
Q6ZMW2	Zinc finger protein 782	29%	2.00E-05
Q9NZL3	Zinc finger protein 224	28%	2.00E-05
A6NDX5	Putative zinc finger protein 840	26%	2.00E-05
Q9P2J8-2	Isoform 2 of Zinc finger protein 624	28%	2.00E-05
Q8TA94	Zinc finger protein 563	27%	2.10E-05
Q9P2J8	Zinc finger protein 624	28%	2.10E-05
Q8NHY6	Zinc finger protein 28 homolog	28%	2.10E-05
Q6ZN19-2	Isoform 2 of Zinc finger protein 841	30%	2.20E-05
A2RRD8	Zinc finger protein 320	29%	2.20E-05
P35789-3	Isoform 3 of Zinc finger protein 93	27%	2.30E-05
Q96N38-3	Isoform 3 of Zinc finger protein 714	28%	2.40E-05
Q7Z7L9	Zinc finger and SCAN domain-containing protein 2	28%	2.40E-05
P35789	Zinc finger protein 93	27%	2.40E-05
Q17R98-3	Isoform 3 of Zinc finger protein 827	34%	2.60E-05
Q9UKN5	PR domain zinc finger protein 4	28%	2.70E-05
Q9NYT6	Zinc finger protein 226	29%	2.70E-05
Q6ZN19	Zinc finger protein 841	28%	2.70E-05
Q6ZN19-3	Isoform 3 of Zinc finger protein 841	28%	2.80E-05
P17019	Zinc finger protein 708	28%	2.80E-05
Q9NQX1-4	Isoform 4 of PR domain zinc finger protein 5	31%	2.80E-05

Supplemental Table S2: Complete list of human proteins with orthology to *Drosophila* CG9932 *(continued)*

Peptide	Gene name	Identity score	E-value
Q96N38-2	Isoform 2 of Zinc finger protein 714	28%	2.90E-05
P17031	Zinc finger protein 26	29%	3.00E-05
O43345-2	Isoform 2 of Zinc finger protein 208	28%	3.00E-05
Q8TBZ8	Zinc finger protein 564	27%	3.00E-05
Q96N38	Zinc finger protein 714	28%	3.00E-05
P49711-2	Isoform 2 of Transcriptional repressor CTCF	26%	3.10E-05

General discussion



6

The EHMT1 network underlying intellectual disabilities

The enzyme euchromatin histone methyltransferase1 (EHMT1) is involved in mono- and dimethylation of the lysine residue on the 9th position of the N-terminal tail of histone H3. Heterozygous loss-of-function mutations in *EHMT1* are known to be causative for Kleefstra syndrome (OMIM #610253). Kleefstra syndrome is a neurodevelopmental disorder characterized by deficits in cognitive functioning, facial characteristics, childhood hypotonia and autism spectrum disorder [1-3]. However, within the cohort of patients based on clinical features reminiscent of Kleefstra syndrome, around 25% have mutations in *EHMT1*. It was previously hypothesized that the other 75% of patients, the *EHMT1*-negative group, harbor mutations in genes with similar biological function. Via a next generation sequencing approach mutations in *MBD5, SMARCB1, NR1I3*, and *KMT2C* were identified [4]. The phenotype of these patients highly resembles that of Kleefstra syndrome, but are commonly referred to as Kleefstra syndrome phenotypic spectrum (KSS) [4]. Additionally, a conserved epigenetic network was suggested underlying intellectual disabilities (ID) [4] (**chapter 1, figure 3**).

In this doctoral thesis, I have aimed to further elucidate the molecular mechanisms underlying Kleefstra syndrome, KSS and other clinical overlapping syndromes. First, I focused on the strong genetic interaction between *EHMT1* and *KMT2C* using a next generation sequencing approach (**chapter 3**). At the same time, I made improvements to an existing protocol [5] to use *Drosophila* courtship conditioning to test general learning, short-term memory and long-term memory (**chapter 2**). Additionally, differential gene expression in whole blood of patients with Kleefstra syndrome and Kabuki syndrome (*KMT2D* mutations) are characterized (**chapter 4**). Lastly, I identified novel interaction partners of the *EHMT1* molecular network underlying KSS (**chapter 5**). Here, I discuss my main discoveries and current challenges in characterizing molecular networks underlying clinical overlapping ID syndromes such as Kleefstra syndrome, KSS and other neurodevelopmental disorders.

Shared pathway-shared phenotype paradigm

Within the "EHMT1-negative" KSS cohort, putative causative *de novo* mutations have been described and these factors are involved in a conserved epigenetic network (**Figure 1**) [4]. I have closely investigated the strong genetic interaction between the *Drosophila* orthologs of *EHMT1* and *KMT2C, G9a* and *trithorax related (trr)* respectively (**chapter 3**). By means of ChIP-sequencing, I found 3371 genomic targets of trr in *Drosophila* heads mainly at the transcription start site of genes (**chapter 3, figure 3**). These genes are generally involved in neuronal functioning. The predicted genomic targets of EHMT in *Drosophila* larvae were already reported [6] and I found a significant overlap of 1047 common targets between G9a and trr (**chapter 3, figure 3E**). Additionally, differentially expressed genes in *Drosophila*

heads lacking either G9a or trr were also significantly overlapping with 119 commonly misexpressed genes (**chapter 3, figure 4**). Furthermore, I identified 5 "hub genes" that are commonly misexpressed and show a binding site for G9a and trr suggesting direct transcriptional regulation. Screening of *trr* mutant *Drosophila* males revealed a courtship conditioning short-term memory phenotype that closely resembles that of *G9a* mutants (**chapter 3, figure 2**). Thus, EHMT1 and KMT2C are associated with the highly overlapping clinical phenotypes and the *Drosophila* orthologs G9a and trr show highly overlapping molecular targets and functionality.

As *Drosophila* trr is also orthologous to the KMT2C paralog KMT2D, I also investigated this protein (**chapter 4**). KMT2D is also a member of the COMPASS family of proteins and mutations in *KMT2D* are associated with Kabuki syndrome (OMIM #147920). Overlapping clinical phenotypes have been described between Kleefstra syndrome and Kabuki syndrome and include ID, hypotonia, developmental delay, seizures and autism-like behaviour [7]. Since both proteins possess histone methyltransferase activity it was postulated that both proteins could also affect transcriptional processes. I investigated this hypothesis, and I report many common differentially expressed genes in Kleefstra- and Kabuki syndrome patients (**chapter 4, figure 1, 2 and 3**). Although the investigated blood cells might not be strictly relevant to the mental impairments in Kleefstra- and Kabuki syndrome (in contrast to brain tissue), this tissue was chosen because it is relatively easy to obtain and analyze. Furthermore, it turned out that the identified differentially expressed genes might be disease relevant. For example, in Kleefstra syndrome cells I found a large group of protocadherin genes down regulated. Protocadherins are involved in cell-cell connectivity of synapses, and neuronal growth cones [8]. The protocadherins are known to be involved in ID syndromes such as RETT syndrome [9]. Other groups of genes that are differentially expressed were found via gene ontology term analysis [10]. The terms that described these groups of genes relevant for ID were "dopamine catabolic process", "synapse", "action potential propagation" and "negative regulation of circadian sleep/wake cycle, wakefulness". Further research of the corresponding genes could provide more insight into the nature of the cognitive and behavioural impairments in Kleefstra syndrome.

Similar to the observations of differentially expressed genes in Kleefstra syndrome patients, the transcriptome of Kabuki syndrome patients seems to be directly relevant to this phenotype (**chapter 4, figure 2**). Kabuki syndrome patients are characterized by postnatal dwarfism, distinct facial appearance (including wide spread teeth and/ or hypodontia), cardiac anomalies, skeletal abnormalities, immunological defects and mild to moderate ID [11]. Gene ontology terms like "cognition" and "learning" as well as "odontogenesis" appeared as examples that are directly relevant. Additionally, there are many specific terms related to brain regions in which those misregulated genes function. These genes could also be direct leads towards further investigation.

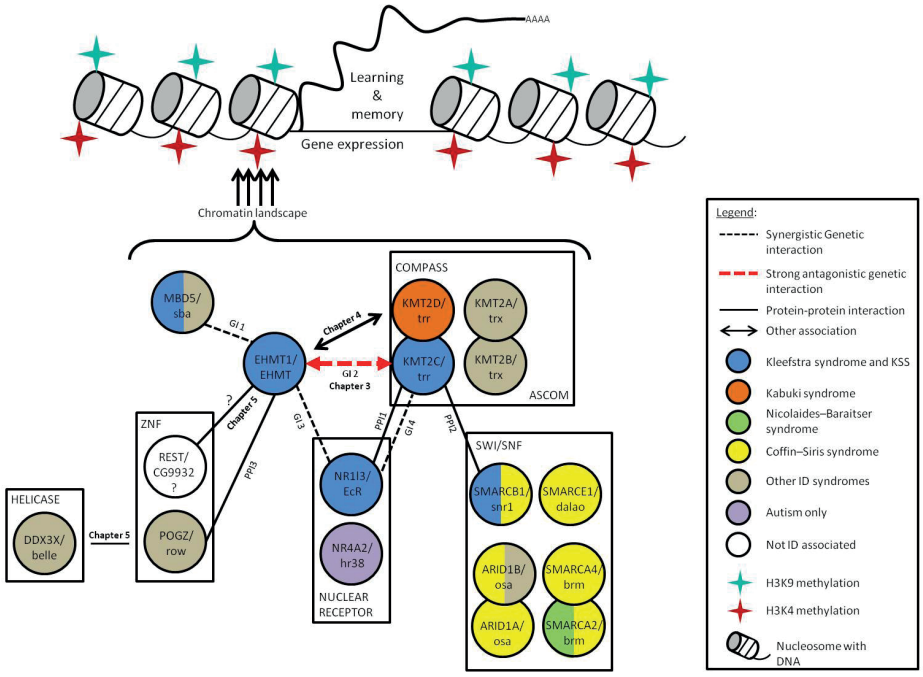


Figure 1: A conserved epigenetic network underlying neuronal functioning.

The Kleefstra associated protein EHMT1 is linked to various other Kleefstra syndrome phenotypic spectrum disorder (KSS) proteins (blue) via genetic interaction and protein-protein interaction. A very strong antagonistic interaction was observed between the *Drosophila* orthologs EHMT and trr. Other members of the COMPASS complex show phenotypic overlap with Kleefstra syndrome patients such as Kabuki syndrome (*KMT2D* mutations), and the candidate ID gene *KMT2B*. Members of the SWI/SNF have been implicated in ID of which *SMARCB1* is found mutated in one KSS patient. Other KSS genes are *NR1I3* and *MBD5*. In this thesis a protein-protein interaction was found between *Drosophila* EHMT and CG9932 (human ZNF462). No genetic interaction was found between *Drosophila EHMT* and *row* and *belle* respectively.

ZNF = zinc finger proteins, SWI/SNF = Switch/Sucrose nonfermentable, ASCOM = activating signal cointegrator-2 (asc-2) complex, COMPASS = complex of proteins associated with Set1. PPI = protein-protein interaction, GI = genetic interaction. PPI1 is based on Kim *et. al* and Sedkov *et al.* [17, 18]. PPI2 is based on Lee *et. al.,* Vicent *et al.,* and Choi *et al.* [19-21]. PPI3 is based on Simon *et. al* [22]. GI1,2,3 are based on Kleefstra *et al.*[4]. GI4 is based on Sedkov *et al.* [18] Figure adapted from Kramer (2013) [23].

When comparing the genes that are differentially expressed in both disease entities a large and significant overlap was observed (**chapter 4, figure 3**). Interestingly, some contributors of two major intracellular signaling pathways were detected (**chapter 4, figure 4 and 5**). First, RAS signaling that controls signal transduction was affected. Members of the RAS super family of proteins are corrupted in a series of ID disorders referred to as RASopathies [12]. Briefly, extracellular signaling molecules activate the RAF-

MEK-ERK kinases which can lead to synaptic plasticity involved in neuronal functioning. For instance, knock-out of *B-raf,* encoding a Ras-isoform, in the forebrain of mice disrupted hippocampus specific spatial learning and contextual discrimination [13]. Second, players of the wnt signaling pathway are identified which have important functions during synapse development and refinement [14]. Briefly, extracellular activation of the frizzled receptor causes an intracellular signaling cascade to ultimately result in gene expression induced by β-catenin and TCF4. RAS and wnt signaling pathways converge on transcriptional processes ultimately resulting in protein synthesis, which is required to consolidate "new" memories and thus suggesting a role in long-term memory [15, 16].

Other chromatin modification complexes in human cognitive disorders

Besides KMT2C and KMT2D, other members of the COMPASS complex of proteins are KMT2A and KMT2B (**Figure 1**). *KMT2A* is associated with Wiedemann-Steiner syndrome (OMIM #605130) and is characterized by a short stature, ID, and a distinctive facial appearance combined with excessive growth of terminal hair around the elbows. A phenotypic overlap is described between Wiedemann-Steiner syndrome and Kabuki syndrome type 1 (*KMT2D* mutations) and type 2 (*KDM6A* mutations) and includes similar facial features and hypertrichosis [24]. *KMT2B* is shown to be mutated in three brothers of a consanguineous Pakistani family [25]. These boys showed a remarkable resemblance to Kleefstra syndrome. Thus, mutations in genes coding for members of the COMPASS family of proteins result in a clinically distinguishable phenotypes with overlapping characteristics.

In addition to the COMPASS complex of proteins, the switch/sucrose non-fermentable (SWI/SNF) chromatin remodeling complex is also involved in ID (**Figure 1**). One patient with Kleefstra syndrome phenotypic spectrum is described with a mutation in *SMARCB1* [4]. Other members of this complex are associated with Nicolaides-Baraitser syndrome (SMARCA2) and Coffin-Siris syndrome (ARID1A, ARID1B, SMARCA2, SMARCA4, SMARCB1, and SMARCE1) [26-32]. A large phenotypic overlap is described for these syndromes such as ID and facial features [33, 34]. However, in contrast to the novel interactions between Kleefstra associated factors, some of these patients were identified by targeted sequencing based on the known protein-protein interactions of the SWI/SNF complex. Thus, the SWI/SNF complex also illustrates the occurrence of overlapping clinical phenotypes caused by mutations in different genes with similar biological function.

Another class of proteins involved in ID are the nuclear receptors (**Figure 1**). A *de novo* mutations in *NR1I3* has been described in one KSS patient (**Figure 1**) [4]. This gene is so far the only one in which mutations have been described causing ID. Thus, definite causality might need the identification of mutations in other patients. Nuclear receptors can bind a ligand, such as a steroid hormone and induce transcription. It has been shown that the COMPASS complex of proteins links the SWI/SNF complex to nuclear receptors

via direct protein-protein interactions [17, 19-21, 35]. The estrogen- and glucocorticoid receptor are other examples of factors known to be involved in learning and memory [36, 37]. However, it is largely unknown how exactly these factors play a role in learning and memory. It is plausible that stress can be a trigger to stimulate processes like learning and memory especially in light of evolution. It could be highly beneficial to organisms when stress-related cues are remembered well.

Another factor involved in neurodevelopmental delay is transcription factor 4 (TCF4) which is linked to Pitt-Hopkins syndrome (OMIM # 610954) [38, 39]. Pitt-Hopkins syndrome is a rare autism-spectrum ID syndrome and is associated with developmental delay, epilepsy, lack of motor coordination, language impairment, and breathing difficulties. Additionally, genome wide association studies correlate *TCF4* to schizophrenia [40, 41]. Modeling Pitt-Hopkins syndrome in *TCF4* heterozygous knockout mice revealed several deficits reminiscent of ID such as social interaction, prepulse inhibition, and deficits in associative and spacial learning and memory [42]. Many of the phenotypes of Pitt-Hopkins syndrome are also associated with Kleefstra syndrome such as ID, speech delay, seizures, sleep disturbances, and developmental delay [7]. Additionally, it has been shown that TCF4 and EHMT1 show molecular convergence [43].

Taken together, mutations in genes coding for members of the above mentioned molecular network (**Figure 1**) are examples in which the clinical spectrum of associated disorders partly overlap. However, clinical variability can be caused by several other factors such as the genetic context or background that enhance or suppress the core phenotype of a patient [44]. The consequences of pre- and postnatal environmental stressors are other examples that might influence core phenotypes of individuals. These aspects lead to a more diffuse spectrum of mental illnesses as seen in the clinic and this phenomenon is described as "dimensionality in mental illnesses" [45].

Experiments using *Drosophila* could uncover fundamental processes in cognitive disorders caused by mutations in epigenetic factors. Many experimental procedures have already been set up to elucidate underlying genetic defects that are reminiscent of ID such as habituation [46-48], aggression [49], circadian rhytm [50], spacial context [51] and the larval neuromuscular junction [52]. It would be interesting to investigate individual mutants of the complete library of *Drosophila* ID orthologs with epigenetic function in these paradigms. This could reveal common patterns and sub-clusters of shared phenotypes between mutants of one gene. On the molecular level, a library of tagged protein lines is available for the *Drosophila* community [53]. These lines could be used for genomic targets of all *Drosophila* ID orthologs with epigenetic function. Identification of overlapping genomic targets could reveal molecular convergence between certain factors.

Higher order chromatin in ID

In recent years it has become evident that the spacial organization of a genome is crucial for normal brain development. Correct folding of DNA is partly regulated by the nuclear lamina which is a filamentous protein mesh functioning as inner scaffold of the nuclear membrane [54]. Molecular mapping of DNA sequences has revealed that some regions are long-term repressed and others that are more dynamically regulated [55, 56]. Local chromatin composition is crucial for the interaction with lamins and it is shown that EHMT proteins are involved in this process [57]. Human G9a has been shown to co-immunoprecipitate with the Barrier-to-autointegration factor which is an essential component of the nuclear lamina [58]. It is also shown that human G9a is required for the silencing of late-replicating genes at the nuclear periphery [59]. H3K9me2 is enriched in lamina associated domains [60] which is potentially deposited by EHMT family members. Thus, a damaged association between G9a and the nuclear lamina could contribute to the transcriptional deregulation of gene expression observed in flies (**chapter 3**) and human (**chapter 4**).

The borders of lamina associated domains are often bound by CCCTC-binding factor (CTCF). CTCF is an insulator protein that is found between active and repressed forms of chromatin forming a molecular "barrier" [61]. Stabilization of these barrier complexes is done by structural proteins such as cohesin [62-64]. Mutations in genes coding for factors associated with the nuclear lamina are also associated with ID syndromes [65-67]. For instance, Cornelia de Lange syndrome (OMIM #122470) is a form of ID and is characterized by several facial features including arched eyebrows and synophrys and other multisystem malformations [68]. Mutations causing Cornelia de Lange syndrome so far include five members of the cohesin family with *NIPBL* being the most frequent one [65]. Mutations in *CTCF* are also causative for ID and microcephaly [66, 67]. Thus, the phenotypes observed in ID patients with mutations in genes coding for nuclear lamina associated proteins could be related to transcriptional de-repression of genomic regions important for proper brain functioning.

Besides association with cohesin, CTCF is also found in enhancer-promoter looping [69]. Enhancers are short DNA sequences that can be bound by transcription factors and influence transcription of distal genes [70-72]. Enhancers can be in an active or poised state with H3K27ac as a mayor discriminator for activation [73]. The histone acetyltransferase that is responsible for this acetylation is P300 [74, 75]. Activation of enhancer-promoter interactions are highly cell and tissue specific [76]. Mutations in the gene coding for P300 are found in patients with Rubinstein-Taybi syndrome subtype 2 (OMIM #613684) [77]. Rubinstein Taybi syndrome subtype 1 (OMIM #180849) is defined by mutations in the closely related histone acetyltransferase gene *CREBBP* [78]. Rubinstein-Taybi syndrome is characterized by ID, postnatal growth deficiency, microcephaly, broad thumps and halluces, and a distinct facial appearance [79, 80]. The protein products of these two genes

are involved in cAMP dependant formation of long-term memory [81, 82]. Thus, these examples show that wrongly formed enhancer-promoter interactions in the brain could contribute to the etiology of ID.

Enhancer function has also been associated with factors involved in Kleefstra syndrome phenotypic spectrum. KMT2C and its ortholog trithorax related have been shown to be involved in enhancer associated monomethylation of H3K4 [83, 84]. Additionally, the EHMT family member G9a has also been associated with enhancers [85, 86]. It was shown that in early mouse development G9a deposits H3K9me2 over active enhancers and thereby blocking their function [85]. Recently, it was shown that G9a co-regulates poised and active enhancers with the lysine-specific demethylase Jmjd2c (KDM4C) [86]. By using a ChIP-sequencing approach, a large overlap was found between genomic binding sites of G9a, Jmjd2c, P300, mediator and cohesin. These sites highly co-occur with the H3K27 acetylation which is the mark associated with active enhancers [86]. However, the exact mechanism through which enhancers regulate brain function is still poorly understood. Four models have so far been proposed which in principle could all co-occur [87]. First, distal sequences carrying a "cargo" of transcription factors into close proximity of the target gene promoter. With respect to learning and memory, this process has been shown for genes important for GABA synthesis [88]. The second mechanism involves locally transcribed enhancer RNA (eRNA). eRNA sequences can be bound by "cargo" transcription factors in order to facilitate DNA looping. This process has been shown at the promoter of the *Arc* gene with nuclear elongation factor (NELF) acting as a transcription repressor that binds these eRNA molecules [89]. The third mechanism described is based on competition between enhancer sequences over one promoter. This has been shown at the *GRIN2B* locus where silencer proteins compete over the binding of the promoter of *GRIN2B* [90]. *GRIN2B* has been shown to be involved in intellectual disabilities, and other neurodevelopmental and neurodegenerative disorders [91, 92]. The fourth mechanism of transcriptional regulation by enhancer sequences is associated with DNA double strand breaks in the context of immediate early genes like *Fos*, *Fosb* and *Npas4* [93]. However, it remains unclear if these double strand DNA breaks have negative side effects for important processes like learning and memory.

With the examples described above in mind, DNA enhancers could be involved in Kleefstra syndrome. It would be interesting to investigate how EHMT1 is associated with enhancers and how enhancers function in the context of ID. *Drosophila* seems to be a very good model to investigate this concept. Many factors involved in enhancer functioning are conserved during evolution in *Drosophila* such as CTCF, cohesins, EP300 [94-96]. For instance, one suggested experiment using *Drosophila* is to investigate transcriptional plasticity during the course of learning and memory. Male flies could be subjected to the courtship assay (**chapter 2**) and brain tissue could be isolated before and after the

training phase. However, since enhancer-promoter interactions are highly cell-type specific and since the brain consists of many different cell types, it would be beneficial to isolate neuronal subtypes using the cell enrichment technique based on UNC84 [97]. This technique genetically labels subsets of neurons using the Gal4/UAS system after which the labeled neurons can be isolated *ex vivo* by cell pull-down techniques with specific antibodies. Selected neural populations can be used to investigate transcriptional plasticity by RNA-sequencing [98], looping of DNA [99] by for instance HiC [100] or ChIA-PET [101], or other methods that elucidate open chromatin by accessibility such as ATAC-seq [102]. In this way courtship conditioning evoked learning and memory could be assessed in two ways. First, genome wide changes could be investigated in control males. It is currently poorly understood which molecular pathways are involved in normal courtship conditioning learning and memory. Second, by using G9a mutant males [6] it could be investigated which mechanisms underlie differentially expressed genes or enhancer-promoter loops in Kleefstra syndrome. Taken together, *Drosophila* can be used to uncover fundamental roles of DNA enhancers involved in Kleefstra syndrome.

Toward treatment of ID

ID disorders are generally considered to be untreatable [103] and only very few pharmacological intervention therapies are now established treatments of human ID disorders [104]. However, recently a number of mouse models has shown potential treatment in adulthood by pharmacological and/or genetic rescue for instance for neurofibromatosis I (*NF1*) [105], Rett syndrome (*MECP2*) [106], Rubinstein-Tabi syndrome (*CPB*) [107], Angelman syndrome (*UBE3A*) [108], and Kabuki syndrome [109]. For Kleefstra syndrome, it has been shown that re-expression of *G9a* in adult male flies restores short-term courtship conditioning memory [6]. Learning deficits of Fragile X flies were successfully rescued by mGluR agonists and lithium [110, 111]. Additionally, a large screen of 2000 compounds revealed nine hits that were able to rescue glutamate-induced lethality of Fragile X flies [112]. Taken together, despite some encouraging laboratory results, knowledge on potential treatment is still very limited [113]. *Drosophila* experiments have a high potential to elucidate fundamental aspects of potential treatments of ID syndromes.

An interesting novel therapeutic approach could come from the strong antagonistic genetic interaction between the *Drosophila* orthologs of EHMT1 and KMT2C [4]. These results imply that loss of one factor could potentially compensate for the loss of the other. It would therefore be interesting to measure courtship conditioning learning and memory (**chapter 2**) in double mutant flies. In short, it would require knockdown of the KMT2C ortholog *trr* in the mushroom bodies of *G9a* mutant animals and thereby rescuing the G9a [6] and trr courtship phenotype (**chapter 3**). Another angle of the same approach would be to use chemical inhibitors [114]. The inhibition of G9a with chemicals

UNC0638 [115] or A-366 [116] could be applied to *trr* mutant flies. The reverse approach of inhibiting KMT proteins in *G9a* mutant flies seems to be more difficult since specific KMT inhibitors are not available at present [117]. However, certain drugs are available to inhibit interactors of KMT such as the bromodomain and extra terminal (BET) family of proteins [118] or PI3K [119]. These drug have only been tested in the context of KMT fusion proteins in heamatological cancers. In my opinion, it would be unfeasible to subscribe these drugs for the treatment of Kleefstra syndrome based on a interaction with an MLL fusion protein that has genetic interaction with EHMT1.

Another experiment to get more insight into pharmacological rescue would be to screen a library of drugs that bind epigenetic regulators [120]. This library of drugs could be tested on mutant males in the courtship conditioning assay (**chapter 2**) in order to investigate improvements in learning and/or memory. Therapeutic intervention to compensate for disruption in the epigenetic machinery has been considered to have a great potential [121].

A potential treatment for Kleefstra syndrome and many other genetic disorders could come from genome-editing techniques. The use of CRISPR/Cas9 (clustered regulatory interspaced short palindromic repeats (CRISPR)/CRISPR associated protein 9) has the potential to relatively easy modify the genome at defined loci [122]. This bacteria-derived system combines a guide RNA with the RNA-guided endonuclease Cas9. Cas9 introduces double strand breaks at defined loci which can be repaired by homologous recombination. When using a specific application of CRISPR, the RNA-guide strand can also be designed to add the *EHMT1* gene to the genome. This approach can thus be used to reincorporate *EHMT1* in the genome of Kleefstra syndrome patients and potentially restore EHMT1 protein levels. However, there are many ethical and practical roadblocks ahead for this approach.

References

1. Kleefstra T, Smidt M, Banning MJ, Oudakker AR, Van Esch H, de Brouwer AP, et al. Disruption of the gene Euchromatin Histone Methyl Transferase1 (Eu-HMTase1) is associated with the 9q34 subtelomeric deletion syndrome. Journal of medical genetics. 2005;42(4):299-306. doi: 10.1136/jmg.2004.028464. PubMed PMID: 15805155; PubMed Central PMCID: PMC1736026.
2. Kleefstra T, Brunner HG, Amiel J, Oudakker AR, Nillesen WM, Magee A, et al. Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. American journal of human genetics. 2006;79(2):370-7. doi: 10.1086/505693. PubMed PMID: 16826528; PubMed Central PMCID: PMC1559478.
3. Vermeulen K, de Boer A, Janzing JGE, Koolen DA, Ockeloen CW, Willemsen MH, et al. Adaptive and maladaptive functioning in Kleefstra syndrome compared to other rare genetic disorders with intellectual disabilities. American journal of medical genetics Part A. 2017. doi: 10.1002/ajmg.a.38280. PubMed PMID: 28498556.
4. Kleefstra T, Kramer JM, Neveling K, Willemsen MH, Koemans TS, Vissers LE, et al. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. American journal of human genetics. 2012;91(1):73-82. doi: 10.1016/j.ajhg.2012.05.003. PubMed PMID: 22726846; PubMed Central PMCID: PMC3397275.
5. Siegel RW, Hall JC. Conditioned responses in courtship behavior of normal and mutant *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America. 1979;76(7):3430-4. PubMed PMID: 16592682; PubMed Central PMCID: PMC383839.
6. Kramer JM, Kochinke K, Oortveld MA, Marks H, Kramer D, de Jong EK, et al. Epigenetic regulation of learning and memory by *Drosophila* EHMT/G9a. PLoS biology. 2011;9(1):e1000569. doi: 10.1371/journal.pbio.1000569. PubMed PMID: 21245904; PubMed Central PMCID: PMC3014924.
7. Mullegama SV, Alaimo JT, Chen L, Elsea SH. Phenotypic and molecular convergence of 2q23.1 deletion syndrome with other neurodevelopmental syndromes associated with autism spectrum disorder. International journal of molecular sciences. 2015;16(4):7627-43. doi: 10.3390/ijms16047627. PubMed PMID: 25853262; PubMed Central PMCID: PMC4425039.
8. Gul IS, Hulpiau P, Saeys Y, van Roy F. Evolution and diversity of cadherins and catenins. Experimental cell research. 2017. doi: 10.1016/j.yexcr.2017.03.001. PubMed PMID: 28268172.
9. Miyake K, Hirasawa T, Soutome M, Itoh M, Goto Y, Endoh K, et al. The protocadherins, PCDHB1 and PCDH7, are regulated by MeCP2 in neuronal cells and brain tissues: implication for pathogenesis of Rett syndrome. BMC neuroscience. 2011;12:81. doi: 10.1186/1471-2202-12-81. PubMed PMID: 21824415; PubMed Central PMCID: PMC3160964.
10. Gene Ontology C. Gene Ontology Consortium: going forward. Nucleic acids research. 2015;43(Database issue):D1049-56. doi: 10.1093/nar/gku1179. PubMed PMID: 25428369; PubMed Central PMCID: PMC4383973.
11. Bogershausen N, Wollnik B. Unmasking Kabuki syndrome. Clinical genetics. 2013;83(3):201-11. doi: 10.1111/cge.12051. PubMed PMID: 23131014.
12. Rauen KA. The RASopathies. Annual review of genomics and human genetics. 2013;14:355-69. doi: 10.1146/annurev-genom-091212-153523. PubMed PMID: 23875798; PubMed Central PMCID: PMC4115674.

13. Chen AP, Ohno M, Giese KP, Kuhn R, Chen RL, Silva AJ. Forebrain-specific knockout of B-raf kinase leads to deficits in hippocampal long-term potentiation, learning, and memory. Journal of neuroscience research. 2006;83(1):28-38. doi: 10.1002/jnr.20703. PubMed PMID: 16342120.
14. Kwan V, Unda BK, Singh KK. Wnt signaling networks in autism spectrum disorder and intellectual disability. Journal of neurodevelopmental disorders. 2016;8:45. doi: 10.1186/s11689-016-9176-3. PubMed PMID: 27980692; PubMed Central PMCID: PMC5137220.
15. Goelet P, Castellucci VF, Schacher S, Kandel ER. The long and the short of long-term memory--a molecular framework. Nature. 1986;322(6078):419-22. doi: 10.1038/322419a0. PubMed PMID: 2874497.
16. Nader K, Schafe GE, Le Doux JE. Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. Nature. 2000;406(6797):722-6. doi: 10.1038/35021052. PubMed PMID: 10963596.
17. Kim DH, Lee J, Lee B, Lee JW. ASCOM controls farnesoid X receptor transactivation through its associated histone H3 lysine 4 methyltransferase activity. Molecular endocrinology. 2009;23(10):1556-62. doi: 10.1210/me.2009-0099. PubMed PMID: 19556342; PubMed Central PMCID: PMC2754897.
18. Sedkov Y, Cho E, Petruk S, Cherbas L, Smith ST, Jones RS, et al. Methylation at lysine 4 of histone H3 in ecdysone-dependent development of *Drosophila*. Nature. 2003;426(6962):78-83. doi: 10.1038/nature02080. PubMed PMID: 14603321; PubMed Central PMCID: PMC2743927.
19. Lee S, Kim DH, Goo YH, Lee YC, Lee SK, Lee JW. Crucial roles for interactions between MLL3/4 and INI1 in nuclear receptor transactivation. Molecular endocrinology. 2009;23(5):610-9. doi: 10.1210/me.2008-0455. PubMed PMID: 19221051; PubMed Central PMCID: PMC2675954.
20. Vicent GP, Nacht AS, Font-Mateu J, Castellano G, Gaveglia L, Ballare C, et al. Four enzymes cooperate to displace histone H1 during the first minute of hormonal gene activation. Genes & development. 2011;25(8):845-62. doi: 10.1101/gad.621811. PubMed PMID: 21447625; PubMed Central PMCID: PMC3078709.
21. Choi E, Lee S, Yeom SY, Kim GH, Lee JW, Kim SW. Characterization of activating signal cointegrator-2 as a novel transcriptional coactivator of the xenobiotic nuclear receptor constitutive androstane receptor. Molecular endocrinology. 2005;19(7):1711-9. doi: 10.1210/me.2005-0066. PubMed PMID: 15764585.
22. Simon JM, Parker JS, Liu F, Rothbart SB, Ait-Si-Ali S, Strahl BD, et al. A Role for Widely Interspaced Zinc Finger (WIZ) in Retention of the G9a Methyltransferase on Chromatin. The Journal of biological chemistry. 2015;290(43):26088-102. doi: 10.1074/jbc.M115.654459. PubMed PMID: 26338712; PubMed Central PMCID: PMC4646261.
23. Kramer JM. Epigenetic regulation of memory: implications in human cognitive disorders. Biomolecular concepts. 2013;4(1):1-12. doi: 10.1515/bmc-2012-0026. PubMed PMID: 25436561.
24. Miyake N, Tsurusaki Y, Koshimizu E, Okamoto N, Kosho T, Brown NJ, et al. Delineation of clinical features in Wiedemann-Steiner syndrome caused by KMT2A mutations. Clinical genetics. 2016;89(1):115-9. doi: 10.1111/cge.12586. PubMed PMID: 25810209.
25. Agha Z, Iqbal Z, Azam M, Ayub H, Vissers LE, Gilissen C, et al. Exome sequencing identifies three novel candidate genes implicated in intellectual disability. PloS one. 2014;9(11):e112687. doi: 10.1371/journal.pone.0112687. PubMed PMID: 25405613; PubMed Central PMCID: PMC4236113.

26. Helsmoortel C, Vulto-van Silfhout AT, Coe BP, Vandeweyer G, Rooms L, van den Ende J, et al. A SWI/SNF-related autism syndrome caused by *de novo* mutations in ADNP. Nature genetics. 2014;46(4):380-4. doi: 10.1038/ng.2899. PubMed PMID: 24531329; PubMed Central PMCID: PMC3990853.
27. Van Houdt JK, Nowakowska BA, Sousa SB, van Schaik BD, Seuntjens E, Avonce N, et al. Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. Nature genetics. 2012;44(4):445-9, S1. doi: 10.1038/ng.1105. PubMed PMID: 22366787.
28. Hoyer J, Ekici AB, Endele S, Popp B, Zweier C, Wiesener A, et al. Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. American journal of human genetics. 2012;90(3):565-72. doi: 10.1016/j.ajhg.2012.02.007. PubMed PMID: 22405089; PubMed Central PMCID: PMC3309205.
29. Santen GW, Aten E, Sun Y, Almomani R, Gilissen C, Nielsen M, et al. Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. Nature genetics. 2012;44(4):379-80. doi: 10.1038/ng.2217. PubMed PMID: 22426309.
30. Santen GW, Clayton-Smith J, consortium ABC. The ARID1B phenotype: what we have learned so far. American journal of medical genetics Part C, Seminars in medical genetics. 2014;166C(3):276-89. doi: 10.1002/ajmg.c.31414. PubMed PMID: 25169814.
31. Miyake N, Abdel-Salam G, Yamagata T, Eid MM, Osaka H, Okamoto N, et al. Clinical features of SMARCA2 duplication overlap with Coffin-Siris syndrome. American journal of medical genetics Part A. 2016. doi: 10.1002/ajmg.a.37778. PubMed PMID: 27264538.
32. Miyake N, Tsurusaki Y, Matsumoto N. Numerous BAF complex genes are mutated in Coffin-Siris syndrome. American journal of medical genetics Part C, Seminars in medical genetics. 2014;166C(3):257-61. doi: 10.1002/ajmg.c.31406. PubMed PMID: 25081545.
33. Samantha Schrier Vergano GS, Dagmar Wieczorek, Bernd Wollnik, Naomichi Matsumoto, and Matthew A Deardorff. Coffin-Siris Syndrome. internet. 2016:https://www.ncbi.nlm.nih.gov/books/NBK131811/.
34. Wieczorek D, Bogershausen N, Beleggia F, Steiner-Haldenstatt S, Pohl E, Li Y, et al. A comprehensive molecular study on Coffin-Siris and Nicolaides-Baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. Human molecular genetics. 2013;22(25):5121-35. doi: 10.1093/hmg/ddt366. PubMed PMID: 23906836.
35. Ananthanarayanan M, Li Y, Surapureddi S, Balasubramaniyan N, Ahn J, Goldstein JA, et al. Histone H3K4 trimethylation by MLL3 as part of ASCOM complex is critical for NR activation of bile acid transporter genes and is downregulated in cholestasis. American journal of physiology Gastrointestinal and liver physiology. 2011;300(5):G771-81. doi: 10.1152/ajpgi.00499.2010. PubMed PMID: 21330447; PubMed Central PMCID: PMC3094144.
36. Bean LA, Ianov L, Foster TC. Estrogen receptors, the hippocampus, and memory. The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry. 2014;20(5):534-45. doi: 10.1177/1073858413519865. PubMed PMID: 24510074; PubMed Central PMCID: PMC4317255.
37. Luine VN. Sex steroids and cognitive function. Journal of neuroendocrinology. 2008;20(6):866-72. doi: 10.1111/j.1365-2826.2008.01710.x. PubMed PMID: 18513207.
38. Van Balkom ID, Quartel S, Hennekam RC. Mental retardation, "coarse" face, and hyperbreathing: confirmation of the Pitt-Hopkins syndrome. American journal of medical genetics. 1998;75(3):273-6. PubMed PMID: 9475596.

39. Zweier C, Sticht H, Bijlsma EK, Clayton-Smith J, Boonen SE, Fryer A, et al. Further delineation of Pitt-Hopkins syndrome: phenotypic and genotypic description of 16 novel patients. Journal of medical genetics. 2008;45(11):738-44. doi: 10.1136/jmg.2008.060129. PubMed PMID: 18728071.
40. Lennertz L, Quednow BB, Benninghoff J, Wagner M, Maier W, Mossner R. Impact of TCF4 on the genetics of schizophrenia. European archives of psychiatry and clinical neuroscience. 2011;261 Suppl 2:S161-5. doi: 10.1007/s00406-011-0256-9. PubMed PMID: 21932083.
41. Quednow BB, Ettinger U, Mossner R, Rujescu D, Giegling I, Collier DA, et al. The schizophrenia risk allele C of the TCF4 rs9960767 polymorphism disrupts sensorimotor gating in schizophrenia spectrum and healthy volunteers. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2011;31(18):6684-91. doi: 10.1523/JNEUROSCI.0526-11.2011. PubMed PMID: 21543597.
42. Kennedy AJ, Rahn EJ, Paulukaitis BS, Savell KE, Kordasiewicz HB, Wang J, et al. Tcf4 Regulates Synaptic Plasticity, DNA Methylation, and Memory Function. Cell reports. 2016;16(10):2666-85. doi: 10.1016/j.celrep.2016.08.004. PubMed PMID: 27568567.
43. Chen ES, Gigek CO, Rosenfeld JA, Diallo AB, Maussion G, Chen GG, et al. Molecular convergence of neurodevelopmental disorders. American journal of human genetics. 2014;95(5):490-508. doi: 10.1016/j.ajhg.2014.09.013. PubMed PMID: 25307298; PubMed Central PMCID: PMC4225591.
44. Coe BP, Girirajan S, Eichler EE. A genetic model for neurodevelopmental disease. Current opinion in neurobiology. 2012;22(5):829-36. doi: 10.1016/j.conb.2012.04.007. PubMed PMID: 22560351; PubMed Central PMCID: PMC3437230.
45. Adam D. Mental health: On the spectrum. Nature. 2013;496(7446):416-8. doi: 10.1038/496416a. PubMed PMID: 23619674.
46. Willemsen MH, Nijhof B, Fenckova M, Nillesen WM, Bongers EM, Castells-Nobau A, et al. GATAD2B loss-of-function mutations cause a recognisable syndrome with intellectual disability and are associated with learning deficits and synaptic undergrowth in *Drosophila*. Journal of medical genetics. 2013;50(8):507-14. doi: 10.1136/jmedgenet-2012-101490. PubMed PMID: 23644463.
47. van Bon BW, Oortveld MA, Nijtmans LG, Fenckova M, Nijhof B, Besseling J, et al. CEP89 is required for mitochondrial metabolism and neuronal function in man and fly. Human molecular genetics. 2013;22(15):3138-51. doi: 10.1093/hmg/ddt170. PubMed PMID: 23575228.
48. Lugtenberg D, Reijnders MR, Fenckova M, Bijlsma EK, Bernier R, van Bon BW, et al. *De novo* loss-of-function mutations in WAC cause a recognizable intellectual disability syndrome and learning deficits in *Drosophila*. European journal of human genetics : EJHG. 2016;24(8):1145-53. doi: 10.1038/ejhg.2015.282. PubMed PMID: 26757981; PubMed Central PMCID: PMC4970694.
49. Anderson DJ. Circuit modules linking internal states and social behaviour in flies and mice. Nature reviews Neuroscience. 2016;17(11):692-704. doi: 10.1038/nrn.2016.125. PubMed PMID: 27752072.
50. van der Voet M, Harich B, Franke B, Schenck A. ADHD-associated dopamine transporter, latrophilin and neurofibromin share a dopamine-related locomotor signature in *Drosophila*. Molecular psychiatry. 2016;21(4):565-73. doi: 10.1038/mp.2015.55. PubMed PMID: 25962619; PubMed Central PMCID: PMC4804182.
51. Ofstad TA, Zuker CS, Reiser MB. Visual place learning in *Drosophila melanogaster*. Nature. 2011;474(7350):204-7. doi: 10.1038/nature10131. PubMed PMID: 21654803; PubMed Central PMCID: PMC3169673.

52. Nijhof B, Castells-Nobau A, Wolf L, Scheffer-de Gooyert JM, Monedero I, Torroja L, et al. A New Fiji-Based Algorithm That Systematically Quantifies Nine Synaptic Parameters Provides Insights into *Drosophila* NMJ Morphometry. PLoS computational biology. 2016;12(3):e1004823. doi: 10.1371/journal.pcbi.1004823. PubMed PMID: 26998933; PubMed Central PMCID: PMC4801422.
53. Sarov M, Barz C, Jambor H, Hein MY, Schmied C, Suchold D, et al. A genome-wide resource for the analysis of protein localisation in *Drosophila*. eLife. 2016;5:e12068. doi: 10.7554/eLife.12068. PubMed PMID: 26896675; PubMed Central PMCID: PMC4805545.
54. Kind J, van Steensel B. Genome-nuclear lamina interactions and gene regulation. Current opinion in cell biology. 2010;22(3):320-5. doi: 10.1016/j.ceb.2010.04.002. PubMed PMID: 20444586.
55. Amendola M, van Steensel B. Mechanisms and dynamics of nuclear lamina-genome interactions. Current opinion in cell biology. 2014;28:61-8. doi: 10.1016/j.ceb.2014.03.003. PubMed PMID: 24694724.
56. Kind J, Pagie L, de Vries SS, Nahidiazar L, Dey SS, Bienko M, et al. Genome-wide maps of nuclear lamina interactions in single human cells. Cell. 2015;163(1):134-47. doi: 10.1016/j.cell.2015.08.040. PubMed PMID: 26365489; PubMed Central PMCID: PMC4583798.
57. Kind J, Pagie L, Ortabozkoyun H, Boyle S, de Vries SS, Janssen H, et al. Single-cell dynamics of genome-nuclear lamina interactions. Cell. 2013;153(1):178-92. doi: 10.1016/j.cell.2013.02.028. PubMed PMID: 23523135.
58. Montes de Oca R, Shoemaker CJ, Gucek M, Cole RN, Wilson KL. Barrier-to-autointegration factor proteome reveals chromatin-regulatory partners. PloS one. 2009;4(9):e7050. doi: 10.1371/journal.pone.0007050. PubMed PMID: 19759913; PubMed Central PMCID: PMC2739719.
59. Yokochi T, Poduch K, Ryba T, Lu J, Hiratani I, Tachibana M, et al. G9a selectively represses a class of late-replicating genes at the nuclear periphery. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(46):19363-8. doi: 10.1073/pnas.0906142106. PubMed PMID: 19889976; PubMed Central PMCID: PMC2780741.
60. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature. 2008;453(7197):948-51. doi: 10.1038/nature06947. PubMed PMID: 18463634.
61. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature. 2008;451(7180):796-801. doi: 10.1038/nature06634. PubMed PMID: 18235444.
62. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature. 2010;467(7314):430-5. doi: 10.1038/nature09380. PubMed PMID: 20720539; PubMed Central PMCID: PMC2953795.
63. Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, et al. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(28):10684-9. doi: 10.1073/pnas.0600326103. PubMed PMID: 16815976; PubMed Central PMCID: PMC1484419.
64. Murrell A, Heeson S, Reik W. Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nature genetics. 2004;36(8):889-93. doi: 10.1038/ng1402. PubMed PMID: 15273689.

65. Boyle MI, Jespersgaard C, Nazaryan L, Ravn K, Brondum-Nielsen K, Bisgaard AM, et al. Deletion of 11q12.3-11q13.1 in a patient with intellectual disability and childhood facial features resembling Cornelia de Lange syndrome. Gene. 2015;572(1):130-4. doi: 10.1016/j.gene.2015.07.016. PubMed PMID: 26164757.
66. Gregor A, Oti M, Kouwenhoven EN, Hoyer J, Sticht H, Ekici AB, et al. *De novo* mutations in the genome organizer CTCF cause intellectual disability. American journal of human genetics. 2013;93(1):124-31. doi: 10.1016/j.ajhg.2013.05.007. PubMed PMID: 23746550; PubMed Central PMCID: PMC3710752.
67. Watson LA, Wang X, Elbert A, Kernohan KD, Galjart N, Berube NG. Dual effect of CTCF loss on neuroprogenitor differentiation and survival. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2014;34(8):2860-70. doi: 10.1523/JNEUROSCI.3769-13.2014. PubMed PMID: 24553927.
68. Allanson JE, Hennekam RC, Ireland M. De Lange syndrome: subjective and objective comparison of the classical and mild phenotypes. Journal of medical genetics. 1997;34(8):645-50. PubMed PMID: 9279756; PubMed Central PMCID: PMC1051026.
69. Holwerda SJ, de Laat W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2013;368(1620):20120369. doi: 10.1098/rstb.2012.0369. PubMed PMID: 23650640; PubMed Central PMCID: PMC3682731.
70. Cheng J, Blum R, Bowman C, Hu D, Shilatifard A, Shen S, et al. A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. Molecular cell. 2014;53(6):979-92. doi: 10.1016/j.molcel.2014.02.032. PubMed PMID: 24656132; PubMed Central PMCID: PMC4031464.
71. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823-37. doi: 10.1016/j.cell.2007.05.009. PubMed PMID: 17512414.
72. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature genetics. 2007;39(3):311-8. doi: 10.1038/ng1966. PubMed PMID: 17277777.
73. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(50):21931-6. doi: 10.1073/pnas.1016071107. PubMed PMID: 21106759; PubMed Central PMCID: PMC3003124.
74. Holmqvist PH, Mannervik M. Genomic occupancy of the transcriptional co-activators p300 and CBP. Transcription. 2013;4(1):18-23. doi: 10.4161/trns.22601. PubMed PMID: 23131664; PubMed Central PMCID: PMC3644037.
75. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011;470(7333):279-83. doi: 10.1038/nature09692. PubMed PMID: 21160473; PubMed Central PMCID: PMC4445674.
76. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature. 2009;457(7231):854-8. doi: 10.1038/nature07730. PubMed PMID: 19212405; PubMed Central PMCID: PMC2745234.

77. Bartsch O, Labonte J, Albrecht B, Wieczorek D, Lechno S, Zechner U, et al. Two patients with EP300 mutations and facial dysmorphism different from the classic Rubinstein-Taybi syndrome. American journal of medical genetics Part A. 2010;152A(1):181-4. doi: 10.1002/ajmg.a.33153. PubMed PMID: 20014264.
78. Hennekam RC. Rubinstein-Taybi syndrome. European journal of human genetics : EJHG. 2006;14(9):981-5. doi: 10.1038/sj.ejhg.5201594. PubMed PMID: 16868563.
79. Petrij F, Giles RH, Dauwerse HG, Saris JJ, Hennekam RC, Masuno M, et al. Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. Nature. 1995;376(6538):348-51. doi: 10.1038/376348a0. PubMed PMID: 7630403.
80. Roelfsema JH, White SJ, Ariyurek Y, Bartholdi D, Niedrist D, Papadia F, et al. Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease. American journal of human genetics. 2005;76(4):572-80. doi: 10.1086/429130. PubMed PMID: 15706485; PubMed Central PMCID: PMC1199295.
81. Dash PK, Hochner B, Kandel ER. Injection of the cAMP-responsive element into the nucleus of Aplysia sensory neurons blocks long-term facilitation. Nature. 1990;345(6277):718-21. doi: 10.1038/345718a0. PubMed PMID: 2141668.
82. Bourtchuladze R, Frenguelli B, Blendy J, Cioffi D, Schutz G, Silva AJ. Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. Cell. 1994;79(1):59-68. PubMed PMID: 7923378.
83. Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, et al. Enhancer-associated H3K4 monomethylation by Trithorax-related, the *Drosophila* homolog of mammalian Mll3/Mll4. Genes & development. 2012;26(23):2604-20. doi: 10.1101/gad.201327.112. PubMed PMID: 23166019; PubMed Central PMCID: PMC3521626.
84. Hu D, Gao X, Morgan MA, Herz HM, Smith ER, Shilatifard A. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Molecular and cellular biology. 2013;33(23):4745-54. doi: 10.1128/MCB.01181-13. PubMed PMID: 24081332; PubMed Central PMCID: PMC3838007.
85. Zylicz JJ, Dietmann S, Gunesdogan U, Hackett JA, Cougot D, Lee C, et al. Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. eLife. 2015;4. doi: 10.7554/eLife.09571. PubMed PMID: 26551560; PubMed Central PMCID: PMC4729692.
86. Tomaz RA, Harman JL, Karimlou D, Weavers L, Fritsch L, Bou-Kheir T, et al. Jmjd2c facilitates the assembly of essential enhancer-protein complexes at the onset of embryonic stem cell differentiation. Development. 2017;144(4):567-79. doi: 10.1242/dev.142489. PubMed PMID: 28087629; PubMed Central PMCID: PMC5312034.
87. Rajarajan P, Gil SE, Brennand KJ, Akbarian S. Spatial genome organization and cognition. Nature reviews Neuroscience. 2016;17(11):681-91. doi: 10.1038/nrn.2016.124. PubMed PMID: 27708356.
88. Bharadwaj R, Jiang Y, Mao W, Jakovcevski M, Dincer A, Krueger W, et al. Conserved chromosome 2q31 conformations are associated with transcriptional regulation of GAD1 GABA synthesis enzyme and altered in prefrontal cortex of subjects with schizophrenia. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2013;33(29):11839-51. doi: 10.1523/JNEUROSCI.1252-13.2013. PubMed PMID: 23864674; PubMed Central PMCID: PMC3713726.

89. Schaukowitch K, Joo JY, Liu X, Watts JK, Martinez C, Kim TK. Enhancer RNA facilitates NELF release from immediate early genes. Molecular cell. 2014;56(1):29-42. doi: 10.1016/j.molcel.2014.08.023. PubMed PMID: 25263592; PubMed Central PMCID: PMC4186258.
90. Bharadwaj R, Peter CJ, Jiang Y, Roussos P, Vogel-Ciernia A, Shen EY, et al. Conserved higher-order chromatin regulates NMDA receptor gene expression and cognition. Neuron. 2014;84(5):997-1008. doi: 10.1016/j.neuron.2014.10.032. PubMed PMID: 25467983; PubMed Central PMCID: PMC4258154.
91. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. Nature. 2012;485(7397):246-50. doi: 10.1038/nature10989. PubMed PMID: 22495309; PubMed Central PMCID: PMC3350576.
92. Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, et al. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell. 2012;149(3):525-37. doi: 10.1016/j.cell.2012.03.028. PubMed PMID: 22521361; PubMed Central PMCID: PMC3340505.
93. Madabhushi R, Gao F, Pfenning AR, Pan L, Yamakawa S, Seo J, et al. Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. Cell. 2015;161(7):1592-605. doi: 10.1016/j.cell.2015.05.032. PubMed PMID: 26052046; PubMed Central PMCID: PMC4886855.
94. Acemel RD, Maeso I, Gomez-Skarmeta JL. Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. Wiley interdisciplinary reviews Developmental biology. 2017. doi: 10.1002/wdev.265. PubMed PMID: 28251841.
95. Vietri Rudan M, Hadjur S. Genetic Tailors: CTCF and Cohesin Shape the Genome During Evolution. Trends in genetics : TIG. 2015;31(11):651-60. doi: 10.1016/j.tig.2015.09.004. PubMed PMID: 26439501.
96. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. Cell. 2012;148(3):458-72. doi: 10.1016/j.cell.2012.01.010. PubMed PMID: 22265598.
97. Henry GL, Davis FP, Picard S, Eddy SR. Cell type-specific genomics of *Drosophila* neurons. Nucleic acids research. 2012;40(19):9691-704. doi: 10.1093/nar/gks671. PubMed PMID: 22855560; PubMed Central PMCID: PMC3479168.
98. Hrdlickova R, Toloue M, Tian B. RNA-Seq methods for transcriptome analysis. Wiley interdisciplinary reviews RNA. 2017;8(1). doi: 10.1002/wrna.1364. PubMed PMID: 27198714.
99. de Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. Genes & development. 2012;26(1):11-24. doi: 10.1101/gad.179804.111. PubMed PMID: 22215806; PubMed Central PMCID: PMC3258961.
100. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nature reviews Genetics. 2013;14(6):390-403. doi: 10.1038/nrg3454. PubMed PMID: 23657480; PubMed Central PMCID: PMC3874835.
101. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. Nature genetics. 2011;43(7):630-8. doi: 10.1038/ng.857. PubMed PMID: 21685913; PubMed Central PMCID: PMC3436933.

102. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature methods. 2013;10(12):1213-8. doi: 10.1038/nmeth.2688. PubMed PMID: 24097267; PubMed Central PMCID: PMC3959825.
103. Ehninger D, Li W, Fox K, Stryker MP, Silva AJ. Reversing neurodevelopmental disorders in adults. Neuron. 2008;60(6):950-60. doi: 10.1016/j.neuron.2008.12.007. PubMed PMID: 19109903; PubMed Central PMCID: PMC2710296.
104. van der Vaart T, Overwater IE, Oostenbrink R, Moll HA, Elgersma Y. Treatment of Cognitive Deficits in Genetic Disorders: A Systematic Review of Clinical Trials of Diet and Drug Treatments. JAMA neurology. 2015;72(9):1052-60. doi: 10.1001/jamaneurol.2015.0443. PubMed PMID: 26168015.
105. Costa RM, Federov NB, Kogan JH, Murphy GG, Stern J, Ohno M, et al. Mechanism for the learning deficits in a mouse model of neurofibromatosis type 1. Nature. 2002;415(6871):526-30. doi: 10.1038/nature711. PubMed PMID: 11793011.
106. Guy J, Gan J, Selfridge J, Cobb S, Bird A. Reversal of neurological defects in a mouse model of Rett syndrome. Science. 2007;315(5815):1143-7. doi: 10.1126/science.1138389. PubMed PMID: 17289941.
107. Alarcon JM, Malleret G, Touzani K, Vronskaya S, Ishii S, Kandel ER, et al. Chromatin acetylation, memory, and LTP are impaired in CBP+/- mice: a model for the cognitive deficit in Rubinstein-Taybi syndrome and its amelioration. Neuron. 2004;42(6):947-59. doi: 10.1016/j.neuron.2004.05.021. PubMed PMID: 15207239.
108. van Woerden GM, Harris KD, Hojjati MR, Gustin RM, Qiu S, de Avila Freire R, et al. Rescue of neurological deficits in a mouse model for Angelman syndrome by reduction of alphaCaMKII inhibitory phosphorylation. Nature neuroscience. 2007;10(3):280-2. doi: 10.1038/nn1845. PubMed PMID: 17259980.
109. Bjornsson HT, Benjamin JS, Zhang L, Weissman J, Gerber EE, Chen YC, et al. Histone deacetylase inhibition rescues structural and functional brain deficits in a mouse model of Kabuki syndrome. Science translational medicine. 2014;6(256):256ra135. doi: 10.1126/scitranslmed.3009278. PubMed PMID: 25273096; PubMed Central PMCID: PMC4406328.
110. Bear MF, Huber KM, Warren ST. The mGluR theory of fragile X mental retardation. Trends in neurosciences. 2004;27(7):370-7. doi: 10.1016/j.tins.2004.04.009. PubMed PMID: 15219735.
111. McBride SM, Choi CH, Wang Y, Liebelt D, Braunstein E, Ferreiro D, et al. Pharmacological rescue of synaptic plasticity, courtship behavior, and mushroom body defects in a *Drosophila* model of fragile X syndrome. Neuron. 2005;45(5):753-64. doi: 10.1016/j.neuron.2005.01.038. PubMed PMID: 15748850.
112. Chang S, Bray SM, Li Z, Zarnescu DC, He C, Jin P, et al. Identification of small molecules rescuing fragile X syndrome phenotypes in *Drosophila*. Nature chemical biology. 2008;4(4):256-63. doi: 10.1038/nchembio.78. PubMed PMID: 18327252.
113. Castren E, Elgersma Y, Maffei L, Hagerman R. Treatment of neurodevelopmental disorders in adulthood. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2012;32(41):14074-9. doi: 10.1523/JNEUROSCI.3287-12.2012. PubMed PMID: 23055475; PubMed Central PMCID: PMC3500763.

114. Prachayasittikul V, Prathipati P, Pratiwi R, Phanus-Umporn C, Malik AA, Schaduangrat N, et al. Exploring the epigenetic drug discovery landscape. Expert opinion on drug discovery. 2017;12(4):345-62. doi: 10.1080/17460441.2017.1295954. PubMed PMID: 28276705.
115. Vedadi M, Barsyte-Lovejoy D, Liu F, Rival-Gervier S, Allali-Hassani A, Labrie V, et al. A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells. Nature chemical biology. 2011;7(8):566-74. doi: 10.1038/nchembio.599. PubMed PMID: 21743462; PubMed Central PMCID: PMC3184254.
116. Sweis RF, Pliushchev M, Brown PJ, Guo J, Li F, Maag D, et al. Discovery and development of potent and selective inhibitors of histone methyltransferase g9a. ACS medicinal chemistry letters. 2014;5(2):205-9. doi: 10.1021/ml400496h. PubMed PMID: 24900801; PubMed Central PMCID: PMC4027767.
117. Senter T, Gogliotti RD, Han C, Locuson CW, Morrison R, Daniels JS, et al. Progress towards small molecule menin-mixed lineage leukemia (MLL) interaction inhibitors with in vivo utility. Bioorganic & medicinal chemistry letters. 2015;25(13):2720-5. doi: 10.1016/j.bmcl.2015.04.026. PubMed PMID: 25987377.
118. Dawson MA, Prinjha RK, Dittmann A, Giotopoulos G, Bantscheff M, Chan WI, et al. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. Nature. 2011;478(7370):529-33. doi: 10.1038/nature10509. PubMed PMID: 21964340; PubMed Central PMCID: PMC3679520.
119. Spijkers-Hagelstein JA, Pinhancos SS, Schneider P, Pieters R, Stam RW. Chemical genomic screening identifies LY294002 as a modulator of glucocorticoid resistance in MLL-rearranged infant ALL. Leukemia. 2014;28(4):761-9. doi: 10.1038/leu.2013.245. PubMed PMID: 23958920.
120. Greschik H, Schule R, Gunther T. Selective targeting of epigenetic reader domains. Expert opinion on drug discovery. 2017:1-15. doi: 10.1080/17460441.2017.1303474. PubMed PMID: 28277835.
121. Millan MJ. An epigenetic framework for neurodevelopmental disorders: from pathogenesis to potential therapy. Neuropharmacology. 2013;68:2-82. doi: 10.1016/j.neuropharm.2012.11.015. PubMed PMID: 23246909.
122. Bassett AR, Tibbit C, Ponting CP, Liu JL. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. Cell reports. 2013;4(1):220-8. doi: 10.1016/j.celrep.2013.06.020. PubMed PMID: 23827738; PubMed Central PMCID: PMC3714591.

Appendix

Summary

Intellectual disability (ID) is a grouping of a large and heterogeneous collection of disorders characterized by an IQ below 70, significant limitations in adaptive behaviour, and an onset before the age of eighteen. It is estimated that the prevalence lies between 1 and 3 percent of the Western population and this equates to almost 400.000 children born per year worldwide. Because of this high frequency and the lifelong need for care, ID represents a major unsolved medical and socio-economic challenge for society. One of the genetic factors causing ID is the *euchromatin histone methyltransferase1 (EHMT1)* gene which encodes a histone methyltransferase. Haploinsufficiency of *EHMT1* cause Kleefstra syndrome that is characterized by ID, childhood hypotonia, general developmental delay, autism spectrum disorder, and facial characteristics. Around 25% of Kleefstra syndrome patients harbor mutations in *EHMT1*. It was thus hypothesized that the other 75% of patients, the "*EHMT1*-negative" group termed Kleefstra syndrome phenotypic spectrum (KSS), have mutations in other genes with a shared biological function. Via a next generation sequencing approach, four additional genes have been identified and their protein products form a conserved epigenetic module underlying ID. The main goal of this doctoral thesis is to uncover shared mechanisms through which affected proteins ultimately result in shared clinical phenotypes in human.

In the introductory chapter (**chapter 1**), I discuss the importance of epigenetic mechanisms underlying ID. It introduces Kleefstra syndrome and Kleefstra syndrome phenotypic spectrum and the known mechanisms through which the affected proteins function, focusing on the strong genetic interaction between *Drosophila G9a* and *trithorax related (trr)*. It concludes on the thesis aims and outline.

For functional analysis of proteins involved ID, reliable, reproducible, and well-controlled experimental procedures are needed. In **chapter 2**, I provide optimizations to an established protocol using *Drosophila* courtship conditioning. I describe guidelines, practical handling of the flies and an R-script to validate experimental data of general learning, short-term- and long-term memory.

In **chapter 3** I continue on the strong antagonistic genetic interaction between *EHMT1* and *KMT2C*. I show five additional ID patients disrupting *KMT2C* and thereby refining the clinical characteristics already known for one KSS patient. I then turned to the model organism *Drosophila* to further show that the ortholog of KMT2C, trithorax related (trr), is involved in courtship conditioning short-term memory, but not in the general morphology of the mushroom body which are neuropils known to be involved in this plastic behaviour. The chapter further continues on the genomic targets of trr in *Drosophila* heads and those are

mainly associated with the transcription start site of genes. The genes with a trr binding site at the transcription start site are generally involved in neuronal processes and are significantly overlapping with G9a associated genes. I further show that upon knockdown of *trr* or knockout of *G9a* in *Drosophila* heads the differentially expressed genes significantly overlap. I found five genes that are putative direct targets of both G9a and trr. Overall, in this chapter I show substantial proof of molecular convergence between EHMT1 and KMT2C.

Chapter 4 describes the study to further elucidate the common molecular mechanism underlying Kleefstra syndrome (*EHMT1* mutations) and Kabuki syndrome (*KMT2D* mutations). KMT2D is the paralog of KMT2C, which both share sequence orthology with *Drosophila* trr (chapter 3). The two disease entities are also associated with clinically overlapping characteristics such as ID, hypotonia, developmental delay, seizures and autism-like behaviour. I analyzed the transcriptome in whole blood of Kleefstra- and Kabuki syndrome patients and I show differentially expressed genes in both disease entities. Moreover, I identified a high overlap between differentially expressed genes and that the commonly misregulated genes have a highly similar function. In conclusion, I show in this chapter strong evidence for molecular convergence between EHMT1 and KMT2D.

Chapter 5 reports on the identification of novel interaction partners of EHMT1 using two approaches. The first approach uses clinical studies that describe two additional patients with characteristics of Kleefstra syndrome. The two patients have mutations in *DDX3X* and *POGZ* and I tested genetic interaction of *Drosophila G9a* with the orthologs "belle" and "row" respectively. However, I report lack of genetic interaction and this was later confirmed by further clinical characterization. Both genes are currently associated to other ID syndromes, not related to Kleefstra syndrome. The second approach uses a *Drosophila* line that expresses a molecularly tagged G9a protein. Affinity purification followed by mass-spectrometry identified a so far uncharacterized zinc finger protein, CG9932, interacting with G9a in *Drosophila* heads. CG9932 shares sequence orthology with human RE1-silencing transcription factor and zinc finger protein ZNF462. However, future studies are needed to understand the exact function of CG9932 and link with G9a. Taken together, this chapter expands the knowledge about interaction partners within the EHMT protein network.

In **chapter 6** I discuss the main findings in a broader context and in the light of recently published literature. I show other protein complexes involved in other ID syndromes and potential links with EHMT1. I also discuss the role of higher order chromatin conformation in ID. Finally, I discuss potential treatments of ID syndromes and how *Drosophila* can help to focus on specific novel therapeutic interventions based on epigenetic mechanisms underlying ID.

Samenvatting

Verstandelijke beperking is een overkoepelende term voor een groep aandoeningen die zowel fenotypisch als genotypisch erg heterogeen is. Cognitief disfunctioneren is één van de overeenkomstige kenmerken, waarbij een IQ lager dan 70, deficiënties in adaptief gedrag en een aanvangsleeftijd onder de achttien jaar gezamenlijke symptomen zijn. Verstandelijke beperking komt bij één tot drie procent van de Westerse bevolking voor. Extrapolerend naar de wereldbevolking zou dit betekenen dat er grofweg 400.000 aangedane kinderen per jaar geboren worden met een verstandelijke beperking. Door deze hoge prevalentie, in combinatie met de behoefte tot levenslange zorg, is verstandelijke beperking momenteel in zowel medisch- als socio-economisch opzicht een grote uitdaging. Eén van de genetische factoren die bij mutatie verstandelijke beperking kan veroorzaken is het gen *euchromatin histone methyltransferase1* (*EHMT1*), wat codeert voor een histon methyltransferase. Haploinsufficiëntie van het *EHMT1* gen veroorzaakt het Kleefstra syndroom wat gekenmerkt wordt door verstandelijke beperking, een lage spierspanning, een ontwikkelingsachterstand, autisme, en karakteristieke gezichtskenmerken. Ongeveer 25% van de patiënten met het Kleefstra syndroom heeft een mutatie in het *EHMT1* gen. Door dit gegeven werd de hypothese opgesteld dat de andere 75% van de patiënten, de *EHMT1*-negtieve groep en later genaamd "Kleefstra syndroom fenotypisch spectrum", wellicht mutaties hebben in andere genen, maar met een vergelijkbare biologische functie. Door middel van "next generation sequencing" zijn vier nieuwe genen geïdentificeerd in patiënten met het Kleefstra syndroom fenotypisch spectrum en de eiwit producten van deze genen vormen een geconserveerd epigenetisch netwerk wat onderliggend is aan verstandelijke beperkingen. Het hoofddoel van het hier beschreven onderzoek was daarom het ontdekken van gezamenlijke moleculaire mechanismen die uiteindelijk resulteren in gedeelde klinisch fenotypes in de mens.

In **hoofdstuk 1** bespreek ik de rol van epigenetische factoren die kunnen leiden tot verstandelijke beperking. Ik introduceer het Kleefstra syndroom en het Kleefstra syndroom fenotypisch spectrum en de bekende functies van de aangedane eiwitten, waarbij ik focus op de sterke genetische interactie tussen *Drosophila G9a* en *trithorax related* (*trr*). Ik eindig met een beschrijving van de doelstellingen van het onderzoek.

In **hoofdstuk 2** beschrijf ik optimalisaties van een bestaand protocol wat het baltsgedrag van *Drosophila* gebruikt om de algemene mogelijkheid tot leren (zeer kortetermijngeheugen), het korte- en langetermijngeheugen te testen. Er worden richtlijnen genoemd, praktische zaken omtrent de handelingen tijdens een experiment beschreven en een R-script om de onderzoeksresultaten te valideren.

In **hoofdstuk 3** ga ik dieper in op de sterke antagonistische genetische interactie tussen *EHMT1* en *KMT2C*. Ik beschrijf klinische karakteristieken in vijf patiënten met "Kleefstra syndroom fenotypisch spectrum disorder" met mutaties in *KMT2C*. Daarnaast beschrijf ik mijn resultaten met het model organisme *Drosophila* om verder aan te tonen dat de ortholoog van KMT2C, trithorax related (trr), betrokken is bij het baltsgedrag kortetermijngeheugen, maar niet in de algehele morfologie van het paddenstoelvormige lichaam welke de structuren zijn die betrokken zijn bij deze plastische vorm van gedrag. Het hoofdstuk beschrijft verder wat de genomische plekken zijn waar het trr eiwit bindt in cellen van de kop van de fruitvlieg en, specifiek, zijn dit met name de transcriptie start posities van genen in het genoom. De genen met een trr bindingsplek zijn over het algemeen betrokken bij neuronale processen en overlappen significant met genen die geassocieerd worden met binding van G9a. Tot slot laat ik in *Drosophila* zien dat door vermindering van trr mRNA moleculen en door disruptie van het *G9a* gen een significant aantal genen verkeerd tot expressie komt in beide condities. Bovendien vond ik vijf genen die mogelijk onder directe invloed van transcriptie staan door trr én G9a. Kortom, deze studie toont sterk bewijs voor moleculaire convergentie tussen EHMT1 en KMT2C.

Hoofdstuk 4 beschrijft de studie naar overeenkomstige moleculaire mechanismen onderliggend aan het Kleefstra syndroom (*EHMT1* mutaties) en Kabuki syndroom (*KMT2D* mutaties). KMT2D is de paraloog van KMT2C, welke beide sequentie homologie vertonen met het trr eiwit in *Drosophila* (hoofdstuk 3). Beide aandoeningen zijn ook geassocieerd met klinisch overeenkomstige kenmerken zoals verstandelijke beperking, een lage spierspanning, een ontwikkelingsachterstand, epileptische aanvallen en autisme. In mijn studie heb ik het transcriptoom in het bloed van Kleefstra- en Kabuki syndroom patiënten bestudeerd. Ik laat genen zien die verschillend tot expressie komen in beide syndromen vergeleken met een controle groep. Daarnaast laat ik een grote en significante overlap zien tussen de genen die verkeerd tot expressie komen, gecombineerd met een overlap in functie van deze groep genen. In dit hoofdstuk geef ik dus bewijs voor moleculaire convergentie tussen EHMT1 en KMT2D.

Hoofdstuk 5 beschrijft de identificatie van nieuwe interactiepartners van EHMT1 door gebruik te maken van twee verschillende methoden. De eerste aanpak maakt gebruik van klinische studies van twee nieuwe patiënten met een grote gelijkenis met het Kleefstra syndroom. Deze twee patiënten hebben mutaties in het *DDX3X* en *POGZ* gen en in deze studie ik heb een genetische interactie getoetst tussen *Drosophila G9a* en de orthologen "*belle*" en "*row*" respectievelijk. In mijn experiment vond ik geen genetische interactie en dit werd later bevestigd door verdere klinische bevindingen. Beide genen zijn op dit moment geassocieerd met andere type verstandelijke beperkingen die niet gerelateerd zijn aan het

Kleefstra syndroom. De tweede aanpak maakt gebruik van *Drosophila* lijnen waarin het G9a eiwit moleculair gemarkeerd is. Affiniteitzuiveringen gevolgd door massa spectometrie op basis van G9a heeft een vooralsnog niet getypeerd eiwit aangetoond, CG9932, met een zinc vinger motief. CG9932 heeft sequentie orthologie met de humane eiwitten "RE1-silencing transcription factor" en "zinc finger protein ZNF462". Er is echter meer onderzoek nodig om de exacte functie en link met G9a verder te ontrafelen. Dit hoofdstuk geeft een goede uitbereiding op de interactie partners binnen het EHMT1 netwerk.

In **hoofdstuk 6** bespreek ik mijn bevindingen in een breder kader en in het licht van recent gepubliceerde literatuur. Ik laat bijvoorbeeld andere eiwitcomplexen zien die betrokken zijn bij verstandelijke beperking met een mogelijke link met EHMT1. Daarna ben ik ingegaan op de rol van hogere orde chromatine vouwing; een fenomeen wat mogelijk betrokken is bij verstandelijke beperkingen. Tot slot bespreek ik potentiele therapeutische interventies waarbij *Drosophila* modellen, gebaseerd op epigenetische mechanismen, een rol kunnen spelen.

List of publications

Koemans TS, Kleefstra T, Chubak MC, Stone MH, Reijnders MRF, de Munnik S, Willemsen MH, Fenckova M, Stumpel CTRM, Bok LA, Sifuentes Saenz M, Byerly KA, Baughn LB, Stegmann APA, Pfundt R, Zhou H, van Bokhoven H, Schenck A, Kramer JM. Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder. PLoS Genet. 2017 Oct 25;13(10):e1006864. doi: 10.1371/journal.pgen.1006864.

Koemans TS, Oppitz C, Donders RAT, van Bokhoven H, Schenck A, Keleman K, Kramer JM. *Drosophila* Courtship Conditioning As a Measure of Learning and Memory. J Vis Exp. 2017 Jun 5;(124). doi: 10.3791/55808.

Snijders Blok L, Madsen E, Juusola J, Gilissen C, Baralle D, Reijnders MR, Venselaar H, Helsmoortel C, Cho MT, Hoischen A, Vissers LE, Koemans TS, Wissink-Lindhout W, Eichler EE, Romano C, Van Esch H, Stumpel C, Vreeburg M, Smeets E, Oberndorff K, van Bon BW, Shaw M, Gecz J, Haan E, Bienek M, Jensen C, Loeys BL, Van Dijck A, Innes AM, Racher H, Vermeer S, Di Donato N, Rump A, Tatton-Brown K, Parker MJ, Henderson A, Lynch SA, Fryer A, Ross A, Vasudevan P, Kini U, Newbury-Ecob R, Chandler K, Male A; DDD Study, Dijkstra S, Schieving J, Giltay J, van Gassen KL, Schuurs-Hoeijmakers J, Tan PL, Pediaditakis I, Haas SA, Retterer K, Reed P, Monaghan KG, Haverfield E, Natowicz M, Myers A, Kruer MC, Stein Q, Strauss KA, Brigatti KW, Keating K, Burton BK, Kim KH, Charrow J, Norman J, Foster-Barber A, Kline AD, Kimball A, Zackai E, Harr M, Fox J, McLaughlin J, Lindstrom K, Haude KM, van Roozendaal K, Brunner H, Chung WK, Kooy RF, Pfundt R, Kalscheuer V, Mehta SG, Katsanis N, Kleefstra T. Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling. Am J Hum Genet. 2015 Aug 6;97(2):343-52. doi: 10.1016/j.ajhg.2015.07.004.

Gupta VK, Scheunemann L, Eisenberg T, Mertel S, Bhukel A, Koemans TS, Kramer JM, Liu KS, Schroeder S, Stunnenberg HG, Sinner F, Magnes C, Pieber TR, Dipt S, Fiala A, Schenck A, Schwaerzel M, Madeo F, Sigrist SJ. Restoring polyamines protects from age-induced memory impairment in an autophagy-dependent manner. Nat Neurosci. 2013 Oct;16(10):1453-60. doi: 10.1038/nn.3512.

Kleefstra T, Kramer JM, Neveling K, Willemsen MH, **Koemans TS**, Vissers LE, Wissink-Lindhout W, Fenckova M, van den Akker WM, Kasri NN, Nillesen WM, Prescott T, Clark RD, Devriendt K, van Reeuwijk J, de Brouwer AP, Gilissen C, Zhou H, Brunner HG, Veltman JA, Schenck A, van Bokhoven H. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. Am J Hum Genet. 2012 Jul 13;91(1):73-82. doi: 10.1016/j.ajhg.2012.05.003.

van Dooren SH, Raijmakers R, Pluk H, Lokate AM, **Koemans TS,** Spanjers RE, Heck AJ, Boelens WC, van Venrooij WJ, Pruijn GJ. Oxidative stress-induced modifications of histidyl-tRNA synthetase affect its tRNA aminoacylation activity but not its immunoreactivity. Biochem Cell Biol. 2011 Dec;89(6):545-53. doi: 10.1139/o11-055.

Curriculum Vitae

Tom Koemans was born on September 15th 1982 in Rheden, and grew up in Arnhem, the Netherlands. A year after graduating from the Thomas à Kempis college in 2000 he started the "hoger laboratorium onderwijs" (HLO) in Nijmegen to study biochemistry. Tom graduated at the department of evolutionary biology in the lab of Dr. Johannes Hackstein under supervision of Dr. Rob de Graaf on the ribosomal gene cluster in hydrogen producing ciliates in the guts of cockroaches. The next steps in genetics followed soon after receiving the VSB scholarship that allowed him to move to Sydney, Australia, working in the lab of professor Susan Clark under supervision of Dr. Marcel Coolen. It is unknown whether the fascinating topic of epigenetics or the waves of Bondi beach motivated him to proceed in science. Tom thus started the master medical biology at the Radboud university of Nijmegen, the Netherlands, in 2007. During the master phase he studied autoimmunity processes in myositis at the Radboud Institute for Molecular Life Sciences in Nijmegen, the Netherlands and male fertility at the human genetics unit in Edinburgh, Scotland. It was in those Scottish highlands where Tom realized he should proceed his scientific career as a PhD student. Between 2011 and 2017 Tom did his PhD research on "epigenetic mechanisms underlying intellectual disabilities" at the department of human genetics of the Radboudumc under supervision of professor Hans van Bokhoven, Dr. Annette Schenck and Dr. Jamie Kramer. During this period he was given the opportunity to attend many conferences and courses and gain experience in the field of next generation sequencing and animal behavior experiments. At the moment, Tom works for Delft Imaging Systems in order to fight tuberculosis in developing countries.

Aside from his interest in molecular mechanisms of disease, Tom has a big interest in sports. Tom started at the age of eight as an all-round athlete with special interest in the middle-distances. Later, Tom started speed skating, wave surfing, mountain biking, and is rock- and alpine climber and guide.

Dankwoord / Acknowledgements

I have worked more than six years on the completion of this thesis. And, yes, it went by so fast. The sometimes-difficult moments were followed by successful periods and I am still very happy that I could stay focused on finishing. And with success. I met many great people along the way of which I would like to thank a few of them here.

Hans, bedankt dat je me de mogelijkheid hebt gegeven om deze ervaring aan te gaan. Bedankt ook dat je vertrouwen had in de goede afloop.

Annette, thank you for letting me perform my experiments within your group. Many thanks for all the discussions we had during the flycycle.

Jamie, your positive attitude has been a great inspiration for me. Even when the distance was very large you kept interest in my research and helped me wherever you could. Without you this thesis would not have been finished. Thanks for everything, I really appreciate this!

Ook wil ik graag mijn familie bedanken. Trudy en Eric, Sander en Marloes, bedankt dat jullie altijd interesse toonden in dit werk. Ik vond het in ieder geval erg leuk om over de fruitvliegen te vertellen.

Tot slot, lieve Goke, jij kent me misschien niet anders dan dat ik aan dit onderwerp werk. Ik ben erg blij dat jij toen besloot mee te gaan op klimweekend met de NijSAC. Op jou kan ik altijd terug vallen, ook al is het soms stressig en druk. Binnenkort beginnen we aan een heel nieuwe fase in ons leven. Ik hoop dat we altijd met veel plezier en enthousiasme onze nieuwe plannen kunnen maken en dit ook over kunnen brengen. Ik hou van je en kijk erg uit naar de komende tijd met z'n 3en!

Tom