Epigenetic alterations in autism spectrum disorders (ASD)

Aïda Homs Raubert

DOCTORAL THESIS UPF 2015

THESIS SUPERVISORS

Prof. Luis A. Pérez Jurado

Dra. Ivon Cuscó Martí

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

Universitat Pompeu Fabra Barcelona

Als meus pares,

a l'Alexandra

a l'Agustí

i als bessons que vindran

ACKNOWLEDGEMENTS

Aquesta no és nomes la meva tesi, en ella han contribuït moltes persones, tant de l'entorn del parc de recerca, de terres lleidatanes, Berguedanes i fins i tot de l'altra banda de l'Atlàntic.

Primer, volia agrair als directors de tesi, al Prof. Luis Pérez Jurado i a la Dra. Ivon Cuscó, tot el temps dedicat a revisar i corregir els raonaments i les paraules en aquesta tesi, ja que sempre han tingut la porta oberta per atendre qualsevol dubte. També per haver-me ensenyat una metodologia, un rigor i un llenguatge científic, on l'entrenament és necessari per assolir els conceptes per la recerca en concret, i pel món de la ciència i la genètica. Gràcies per la dedicació, la paciencia, la feina i energia dipositada.

No hagués arribat al mateix port si al laboratori no m'hagués trobat amb persones que m'inspiren. Primer de tot, a les nenes: a la Gabi, la companya de vaixell fins i tot el dia de dipositar la tesi, perquè sobretot ens hem sabut acompanyar i entendre malgrat tenir altres maneres de funcionar, gràcies. També a la Marta i la Cristina, que amb la seva honestedat i bona fe, omplen el laboratori de bones energies, gràcies per ser-hi en tot moment. I a la Marta, també per ajudar-me en moltes coses relacionades amb la recerca i la pantalla negra. També, a la Clara i la Maria, que encomanen el seu optimisme i l'amor a la genètica. I a totes elles per totes les vivències juntes, gràcies.

Al laboratori, també hi han hagut noves incorporacions, com la Roser, que ha portat un trosset de la manera de treballar americana, sempre disposada a ajudar i sempre amb un somriure. Gràcies per ajudar-me en moltes coses i per l'entesa a nivell personal. Per altra banda, també hi havia incorporacions des dels inicis del lab, com la Raquel, que m'ha ajudat amb moltes coses a nivell tècnic i que sempre s'ha mostrat disposada a ajudar en dissenys d'experiments i interpretacions. A l'Armand, al Marcos, i al Benja que sempre s'han mostrat disposats a ajudar amb la bioinformàtica. I també a tots els altres companys de laboratori, Judit, Mariví, Matilde i becaris que hem compartit estones i treball científic. També al Josep d'informàtica, que l'hem atabalat molt.

A l'altra banda de l'Atlàntic, agrair a la Dra. Janine LaSalle i la Dra. Diane Schroeder de la Universitat de California Davis per la seva acollida, la seva aportació al projecte científic i per tota l'experiència viscuda, que va canviar per complet la meva percepció de la ciència. De quina manera? Docns en la percepció de la ciència allà i aquí, per guanyar confiança en el projecte de recerca i en mi, per l'ambient independent on l'eficiència és palesa, per un altre tipus d'educació,etc. Gràcies a tots els companys del lab en especial a la Diane, per la seva amistat que encara segueix i pel seu suport, fins i tot, quatre dies abans de la tesi, quan on les seves paraules d'ànims reconforten.

Agrair als amics les bones estones passades aquests anys. Primer als companys de pis, el Jordi i la Txell, que hem compartit molts sopars, i moltes xerrades. També a la Marta, per ser-hi i anar coordinades en el camí, donant-nos suport i ajudant-nos quan és necessari. A l'Anna Graupera per ser tant bona persona i companya d'experiències per qualsevol bombardeig. A l'Anna Baulies, per la seva energia, per compartir l'amor al Pallars i per mil xerrades. A l'Emma pel cordó umbilical lligat a terres pallareses i per la seva senzillesa. A l'Estel per ser l'estel del nord que va pensar en mi en un moment en concret molt especial.

Agrair tot el suport personal als meus pares, la Lidia i el Joaquim, que m'han ajudat molt aquests anys i han simplificat la meva vida cada vegada que els visitava enmig de la naturalesa i els animals. També a la meva germana, l'Alexandra, que percep la vida d'una altra manera i es qüestiona els estàndards, per entendre'm i acompanyar-me. I sobretot, el meu gran amor, l'Agustí, sens dubte el millor regal d'aquests anys de tesis, i amb el qual he trobat una gran persona que m'acompanya i amb el qual m'encanta compartir la vida. També a la seva família.
Finalment, a la família que nosaltres formarem que es mereix un capítol especial. Als bessonets que m'han acompanyat en la redacció de la introducció de la tesi sense saber-ho i que han estat aguantant asseguts moltes hores durant la redacció dels altres capítols. Per tot el que imagino junts i per totes les vivències que tindrem plegats.

A totes les persones mencionades, ens ha costat deu i ajuda però hem arribat fins aquí!

ABSTRACT

The aetiology of autism spectrum disorders (ASD), a group of neurodevelopmental conditions with early onset, characterized by social and communication impairment and restricted interests, is unknown in about a third of the patients. The intense research done over the past decade has revealed a genetic contribution, while the epigenetic contribution barely begins to show. The epigenetic marks can exert an effect in gene expression without altering the underlying genetic sequence. In turn, these marks can be impaired by genetic mutations in their target sequence. Therefore, research in genomic, epigenomic and transcriptomic fields will provide convergent information to unravel the causes of ASD, necessary to establish improved diagnostic protocols and therapeutic strategies, allowing an earlier diagnosis and personalized treatment crucial for a better prognosis. Our data reveal variants associated to the phenotype which shows genetic-epigenetic interplay along with gene expression consequences. It also reveals region epigenetic variants, which follow a polygenic or complex model. Finally, we found ASD genotype-specific epigenetic marks. In the future, the progress in cost-efficiency technologies assessing epigenomics, and the availability of a reference epigenome in various tissues and cell types will provide the background to set a step-forward in establishing the developmental stage, cell types and tissues involved in the epigenetic mechanisms of the disorder.

RESUM

L'etiologia dels trastorns de l'espectre autista (TEA), un grup de malalties del neurodesenvolupament d'aparició primerenca caracteritzades per problemes de comunicació, relació social, i per la presencia d'interessos restringits, és desconeguda per un terç dels individus afectats. La intensa investigació feta durant l'última dècada ha revelat una gran contribució genètica en aquesta malaltia, mentre que de l'epigenètica tot just es comença a evidenciar. Les marques epigenètiques, sense alterar la seqüència genètica subjacent, tenen un efecte en l'expressió dels gens. A la vegada, aquestes marques epigenètiques es poden veure afectades per mutacions genètiques a la seqüència. Així doncs, la recerca en genòmica, epigenòmica i transcriptòmica proporcionarà informació convergent per determinar les causes

dels TEA, indispensable per establir millores en els protocols de diagnòstic i en estratègies terapèutiques, facilitant el diagnòstic precoç i el tractament personalitzat, crucial per a un millor pronòstic. Les nostres dades mostren que hi ha alteracions genètiques i epigenètiques associades al fenotip, que interactuen i tenen conseqüències sobre l'expressió gènica. També hem trobat regions amb alteracions epigenètiques, que sembla que contribueixen de manera additiva i seguint un model complex. Finalment, trobem marques epigenètiques específiques de grups de genotips TEA. En el futur, la millora de les tecnologies disponibles per avaluar l'epigenòmica, i la disponibilitat d'un epigenoma de referència en diversos teixits i tipus cel·lulars, serviran com a base per fer un pas cap endavant en l'establiment de l'etapa del desenvolupament, dels tipus cel·lulars i els teixits involucrats en els mecanismes epigenètics del trastorn.

PROLOGUE

The emergence of genome-wide techniques that permit the characterization to a single-base resolution is recent. For example, for the research work compiled in this thesis, we used several epigenetic techniques: methylation data from the 27k array (Illumina©) covering 27,000 positions, the 450k array (Illumina©) covering 450,000 positions, and more recently, whole methylome sequencing (approximately 10^{10} positions). This picture is an example of the unprecedented next generation sequencing era, which extrapolates also to the genomics and transcriptomics fields among others, and provides a large amount of data leading to a better knowledge of the genome/epigenome structure, and allowing studying new alterations and mechanisms leading to disease.

This thesis present the results of the application of *OMICS* strategy (whole genome-epigenome-transcriptome) in the study of the aetiology of autism spectrum disorders (ASD), a group of neurodevelopmental conditions, affecting about 1% of children, and with a higher prevalence in males, that involves substantial costs for health and educational system.

The thesis is divided in chapters following the classical structure:

The **introduction** is an overview of the ASD characteristics, classification and diagnosis. Also, the neurobiological characteristics and the genetic mechanisms involved in the disease in order to understand the architecture of ASD causes and the inheritance models for the disease. A multifactorial or complex model also involves the epigenetic contribution, which could explain the missing heritability seen in ASD. Therefore, an important section is the overview of the epigenetic general mechanisms, and the specific alterations detected in ASD. We also review the molecular techniques to study methylation alterations, from locus-specific to genome-wide. Finally, we revise the animal and in vitro models of the disease, the main features and gene functional studies. The combination of sections in the introduction gives a picture of the disease aetiology and focus of the thesis to interpret the results.

The **body** of the thesis contains the **scientific results**. The first chapter is the study of new epigenetic alterations in ASD idiopathic patients. The second chapter has an innovative design, and is the characterization of epigenetics in multiple ASD-genotypes subgroups. In the third chapter, methylation was studied and compared to another tissue data obtained by a methylome sequencing technique. All the chapters involved explanations of the design and the approaches used.

A general **discussion** analyzes the results and their possible interpretations and implications. A final chapter contains the summarized **conclusions** of the thesis.

CONTENTS

INTRODUCTION

INTRODUCTION

1. Autism Spectrum Disorders

Autism spectrum disorder (ASD) (OMIM 209850) is a neurodevelopmental condition diagnosed today on the basis of two symptom areas: social communication impairments, and restricted repetitive behaviours, interests or activities. The symptoms, however, show substantial variation in severity between individuals. ASD was first described by Leo Kanner in 1943, who observed children with severe social isolation and with language impairment, and described autism as an innate disorder[1]. ASD was not considered a separate psychiatric condition until the appearance in 1980 of the Diagnostic and Statistical Manual of Mental Disorders III (DSM-III). Nowadays, DSM-V is the standard since 2013. ASD onset is before three years of age and the prevalence is 0.4-1.25% of the children worldwide[2].

1.1. Disease characteristics and classification

ASD involve heterogeneous conditions with dysfunctional neurodevelopment, behaviourally defined by shared symptoms involving impairments in social communication and restrictive and repetitive behaviours. ASD is a lifelong condition diagnosed by the presence of observable behaviours. Early manifestations are normally perceived between 15 and 18 months of age by paediatricians and parents. Early signs include lack of gaze, no response to name, lack of interaction, delayed or absent joint attention behaviour, and social anomalies such as loss of eye contact, non-response to prize, or playing without imitation, among others[3,4]. For most children the onset of symptoms is gradual, although about 15-40% of the patients show a regressive onset meaning a normal development to the first year and a sudden -or gradual- loss of language and isolation[5,6]. Repetitive movements may develop later, around three or four years of age.

Other associated ASD developmental features include intellectual disability (IQ ≤70) seen in 31-55% of cases[2,7], attention-deficit hyperactivity disorder in 28-44% of cases, and motor abnormalities in 79% of cases[8]. Additionally, general medical problems as epilepsy are seen in

5-46% of cases[9], and also gastrointestinal problems and sleep disorders, in 9-70% and 50-80% respectively. Psychiatric and personality disorders can co-occur with ASD, being the more common depression in 12-70%, and anxiety in 42-26% of cases [8]. Besides, about 15-43% of ASD children present significant dysmorphology, such as head circumference (macrocephaly in 16.7% and macrocephaly in 7.3-15%[10]) and other minor and major dysmorphologies, reported in case-control studies[11,12].

Structural and functional brain anomalies have been observed in ASD by neuroimaging and electroencephalography techniques. Brain structural malformations are present in 40% of cases, such as variation in gray matter volumes, white-matter abnormalities and temporal lobe abnormalities[13,14]. However, the presence of brain anomalies is controversial as studies in high functioning ASD, term used for ASD individuals with better cognition features IQ>70, did not show neither brain morphology nor functioning alterations[15].

ASD affects predominantly males, with a ratio male-to-female estimated at approximately 4:1[16], which can vary and be higher in high-functioning cases and lower in cases with intellectual disabilities[17,18].

Classification of the symptoms and conditions of ASD change over the years. Multiple classification tools are available to describe the standard characteristics and categorization of mental disorders, including ASD. The worldwide most used standard manuals are The Diagnostic and Statistical Manual of Mental Disorders (DSM), published by the American Psychiatric Association, and The International Classification of Diseases (ICD) by the World Health Organization. The DSM latest release is the 5th edition (DSM-V) published on May 18, 2013[19]. The previous manual, the 4th edition (DSM-IV) was released in 1994 and "text reviewed" (DSM-IV-TR) in 2000[20] and many research studies were performed under its criteria. The ICD, currently in the 10th edition (ICD-10), was released in 1992 and revised in 1993[21].

Classification of symptoms by the manuals DSM-IV-TR and ICD-10 are defined completely on the basis of three core symptoms for autistic disorders: impairment in communication, deficits in social interaction and restricted repetitive behaviour, interests and activities. However, DSM-V collapsed the symptoms to only two: deficits in social communication and interaction and the restricted repetitive behaviours, interests or activities (RRBs). Therefore the spectrum

definition changes depending on the manual. DSM-IV and ICD-10 defined spectrum subcategories: autistic disorder, Rett's syndrome, childhood disintegrative disorder, Asperger's disorder, and pervasive developmental disorder-not otherwise specified (PDD-NOS), among others; whereas DSM-V does not maintain the ASD sub-classifications[19,22] (Table 1).

DSM-IV-TR	DSM-V	ICD-10
Autistic disorder Asperger Disorder Childhood disintegrative disorder Pervasive developmental disorder not otherwise specified (PPD-NOS)	ASD, Social communication disorder	Childhood autism Atypical autism Rett's syndrome Other childhood disintegrative disorder Asperger's syndrome Pervasive developmental disorder unspecified Overactive disorder associated with mental retardation and stereotyped movements

Table 1. Subcategories within ASD in each classification manual.

Asperger's disorder does not have cognitive impairment or communication impairment but present social alterations and restricted and repetitive interests and activities. PDD-NOS (atypical autism) present impairments in the three core symptoms but some do not reach the diagnostic threshold or appear in older ages. In DSM-V, Asperger's syndrome and PDD-NOS, which have typically higher communication and social functioning and present milder behaviour stereotypies, are only included if they meet ASD criteria, although many of them will no longer be. Additionally, DSM-V excludes Rett's syndrome, although they can be diagnosed of ASD if they meet the criteria[23]. These changes are supported by the difficulty of classification in the subtypes described in the DMS-IV following the standardized diagnostic instruments[24]. As a novelty, DMS-V categorizes ASD based on their degree of support needed in three levels and gives a single category of intellectual disability rather than severity level divisions. An example of the criteria to diagnose classic autism with each manual is listed in Table 2. Many studies compare DSM-IV and DSM-V and major changes are still being discussed by the scientific community, mainly the concern comes for services which will be provided to the patients and if the classification will affect comparison with previous research results[25,26].

	DSM-IV-TR (for autistic disorder)	DSM-V	ICD-10 (for childhood autism)
Communication alterations	**one of the following:** - delay or lack of spoken language - if speech is present, impairment to initiate and sustain a conversation - stereotyped and repetitive use of language - limited imitative play	**all three deficits of the following**: - Impairment social and emotional reciprocity - Inadequate nonverbal communicative behaviours - difficulties in developing and maintaining relationships	**two of the following:** - delay or lack of spoken language - failure to initiate maintain a conversational interchange - stereotyped and repetitive use of language - abnormalities in speech (pitch, stress, rate, rhythm and intonation)
Social alterations	**two of the following:** - multiple nonverbal inadequate behaviours - failure to develop peer relationships - lack of sharing with other people lack of social reciprocity		**one of the following:** - failure adequately to use non-verbal body language - failure to develop peer relationships - lack of socio-emotional reciprocity and modulation
Interests, activities and behaviours alterations	**one of the following:** - stereotyped and restricted patterns of interest - inflexible with non-functional routines or rituals - stereotyped and repetitive motor mannerisms - preoccupation with parts of objects	**two of the following**: - Stereotyped or repetitive speech, motor movements, or use of objects - Adherence to routines, ritualized patterns of verbal or nonverbal behaviour, or excessive resistance to change - restricted, fixated interests - Hyper/hypo-reactivity to sensory input or unusual interest in sensory aspects	**two of the following:** - stereotyped and restricted patterns of interest - compulsive adherence to specific, non-functional, routines or rituals - stereotyped and repetitive motor mannerisms - preoccupations with part-objects or non-functional elements - distress over changes in small, non-functional, details of the environment
Additional information	- In total, **six characteristics should be altered** from the items above. - Also, **impaired one of the following before the age of 3**: 1) social interaction, 2) social use of language, or 3) symbolic or imaginative play. - The items are not attributable to other classification subtypes of DSM-IV-TR (Asperger, Rett's syndrome, etc.).	- Symptoms must be present in early childhood - Symptoms together limit and impair everyday functioning.	- Also, impaired **one of the following before the age of 3**: 1) receptive or expressive language as used in social communication, 2) selective social attachments or social interaction, 3) functional/symbolic play - The items are not attributable to other classification subtypes (Asperger, Rett's syndrome, etc.) or if mental retardation, or schizophrenia, etc. are present.

Table 2. Criteria for ASD following DSM-IV-TR, DSM-V and ICD-10.

1.2. Epidemiology

The prevalence and incidence of ASD are controversial issues. ASD was considered a rare disease in the 1960s (1 every 10,000 individuals); today ASD is not considered a rare disease and the current estimate in the USA is 1 every 68[2]. These data shows that the prevalence has been reported to increase 20-30 folds since 1960s.

The increase in ASD incidence generates a strong concern for the "autism epidemic", which is attributed to various factors. Some factors include the earlier diagnosis due to the improvement of screening tools[27],and changes in the diagnostic chiteria[27–29] as in past surveys the milder cases as Asperger's syndrome and PDD-NOS, or comorbid diseases with ASD phenotype[29–31] were not included. Additionally, there is an increased awareness from parents and professionals[32,33] and the social services available make the society be more alert to the signs. Besides, milder cases having relatives affected may be diagnosed due to social influences[34]. Regarding all the influencing factors increasing ASD rate, it is still being discussed if these factors explain the fact.

Furthermore, the prevalence and its change over the years vary in different studies. Autism and Developmental Disabilities Monitoring Network (ADDM Network) created in year 2000 assess every two years systematically the ASD prevalence in the USA. ADDM Network current estimation is 1 every 68 children aged 8 years in 2010, which in boys is 1 every 42, showing a 78% increase between 2004 and 2010[2] (Figure 1). Whereas for the same period, in the UK, the prevalence was 1 in 256 boys aged 8 years[35] and incidence was maintained constant since 2004[35](Figure 1). Apart from these studies many ASD prevalence studies are constantly being published, which may follow certain protocols to assure reliable results[36].

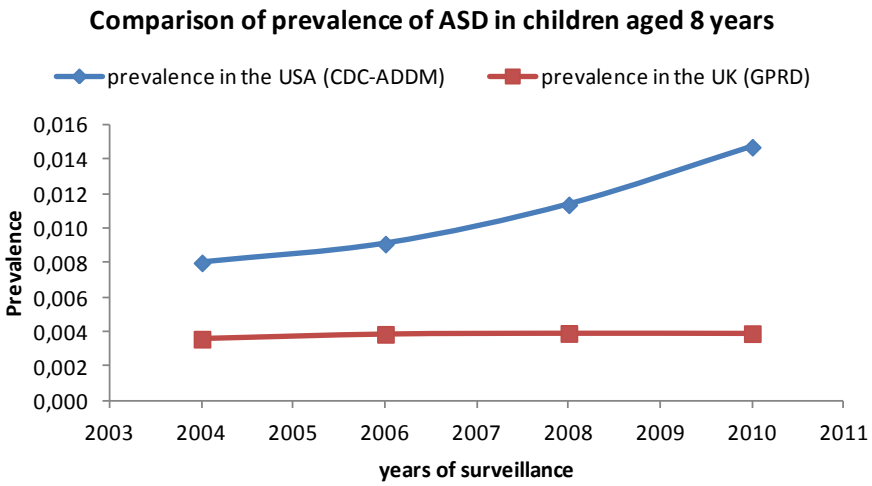


Figure 1. Comparison of two ASD prevalence studies of patients aged 8 years old from 2004 to 2010 from ADDM Network (CDC) in the USA, and General Practice Research Data-base (GPRD) in the UK. The data were extracted from global ASD rates (males and females) from both studies analyzing 363,749 children in the USA (9% of the USA), and 256,278 children in the UK. A higher prevalence for all the period was observed in the USA compared to UK Also an increase on incidence rate was seen in USA but not in UK. Both studies are comparable as the same psychological test was assessed under DSM-IV, and the children were studied that specific age because it shows the highest peak of prevalence[37].

1.3. Diagnosis

ASD has not a specific biochemical biomarker and the diagnosis is mainly a clinical diagnosis. The diagnostic assessment for ASD is described in many guidelines with recommendations for patient management. The guidelines describe the process from the first ASD suspicion of a primary-care physician to the diagnostic obtained by a multidisciplinary team[8]. A multidisciplinary team including a medical, biological and psychological assessment is needed.

If possible, following the recommended guidelines, when ASD is suspected, the diagnostic process starts gathering information from parents together with the pediatrician or primary-care physician. The physician refers to a professional responsible of an ASD multiprofessional team, which should be defined in the region. The multiprofessional team consist of a core team: a pediatrician and/or child and adolescent psychiatrist, a speech and language therapist, and a clinical and/or educational psychologist; and may include if the assessment is complex other support professionals as a neurologist or psychiatrist (patients with severe visual or

hearing impairments, motor disorders such as cerebral paralysis, severe intellectual disability and complex language disorders)[8,38] (Figure 2).

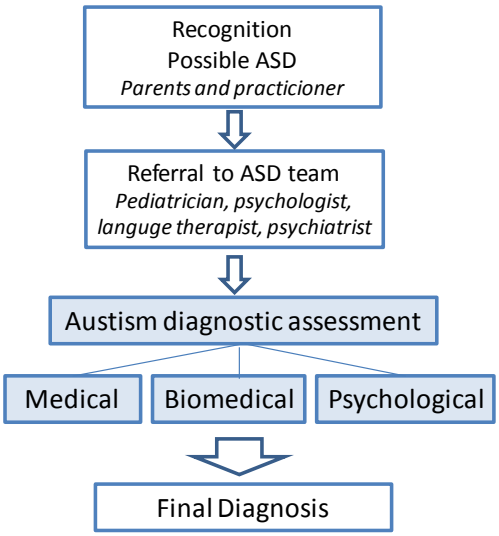


Figure 2. ASD diagnostic process following the recommended guidelines. Due to complex phenotype and the high comorbidity, differential diagnosis of ASD should be performed by a core ASD team with the help of additional professionals in order to assess medical, biomedical and pshychological assessment leading to a final diagnosis.

The diagnostic assessment by the multidisciplinary team is divided by three blocks: medical, biomedical and psychological assessment.

Medical assessment

Regarding the medical diagnosis, the perinatal and postnatal development, health and psychiatric history, and also familial information should be reported. Important information includes regression, neurological and seizure events, attention deficits, behaviour (self-harming, aggression, sleep and eating difficulties) and education problems[8,38].

Besides, many neurodevelopmental, mental or behavioural disorders can suggest ASD, due to the high comoridity, leading to an inaccurate diagnostic, so a differential diagnosis of ASD is crucial. A medical examination including parameters such as developmental growth, cranial perimeter, dimorphic features, dermal abnormalities (may suggest neurofibromatosis or tuberous sclerosis), audition and vision, among others[38]. Inappropriate diagnosis can lead to an inappropriate treatment and may cause anxiety and stress to the child and family[38].

Biochemical assessment

Although no specific biological biomarker has been detected, several analyses are recommended during the diagnostic process. Some of them are specially recommended if a certain comorbid syndrome is suspected. For example, neurological examinations with electroencephalography (EEG) to monitor the electric activity if epilepsy and seizures are present, regression occurred, or to discard Ladau-Kleffner syndrome (Orphanet Nº: ORPHA98818)[39]; neuroimaging studies if intracranial lesions, neurological familial problems or tuberose sclerosis are suspected. Besides, genetic studies are conducted when a genetic origin is suspected (familial history with a specific disorder or physical sign with a specific aetiology). Additionally, metabolic profiling can be selectively indicated when neurometabolic disorders are suspected[40]. Finally, since genetics is known to be involved in ASD aetiology, it is also recommended to assess genetic analyses. Genetic approaches include karyotype, Fragile-X and Rett Syndrome tests, and molecular arrays, among others.

Psychological assessment

Regarding the psychological assessment for speech, language and communication evaluation, various screening and diagnostic tools are used to compute ASD symptoms[41]. The instruments are designed to be adapted to certain age, sex and cognition abilities. There are different instrument formats: questionnaires completed by parents or health practitioners, and interviews or direct assessment performed by a trained certified individual in order to assure inter-rater reliability. One of the most widely used for research is Autism Diagnostic Interview Revised (ADI-R)[42]. It is a detailed parent interview which evaluates patients older than 12 months of any verbal level. Autism Diagnostic Observation Schedule (ADOS)[43] is a shorter version, which analyzes development- and language-dependent modules, and is considered a good combination with ADI-R[44]. Additionally there are checklists as M-CHAT[45], a parent-report instrument, which is the most commonly used with an assessment time of maximum 15 minutes, and easier to implement into the clinical practice. Some studies, recommend using a small numbers of complementary tests[46]. Despite the tests available, the scientific community is still not sure of the effectiveness of community screening programs for ASD because there is not still the gold standard instrument in order to identify and to ensure the appropriate patient management[47].

After following medical, biomedical and psychological assessment, the professionals diagnose the patient, leading to specific medical follow-ups and interventions depending on the diagnosis.

1.4. Neurobiolgy and cognitive consequences

ASD is a neurodevelopmental disorder, thus many studies searching brain structural and functional abnormalities have been performed. Neuropathological findings include abnormal development of cerebellum, limbic system including amygdala and hippocampus, and the frontal and temporal cortices, showing altered number of neurons and neuronal morphology[48] (Table 3, Figure 3). The most studied regions, cerebellum and amygdala, which were found enlarged in ASD patients, and with decreased number of Prukinje cells, and decreased size of neurons, respectively[49]. Besides, several brain areas have been implicated in each of the ASD impaired core behaviours[49].

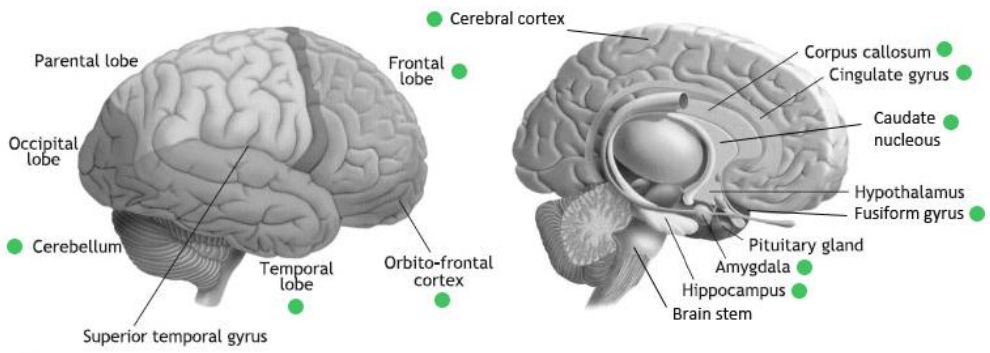


Figure 3. Brain main structures involved in ASD. The specific cortices are simplified as cerebral cortex.

Brain region	Sample	Results	Article
Amygdala-lateral, basal central nuclei	9 ASD; 10 CNT; M 10–44 years	12% decrease in total amygdala neurons. 14% decrease in neuron number in lateral nucleus.	*Schumann et al., 2006*
Fusiform gyrus (FG) and visual cortex	7 ASD; 4 M, 3 F 10 CNT; 8 M, 2 F 3–50 years	Neurons are fewer and smaller in size (~10–20%) in ASD.	*Van Kooten et al., 2008*

Brain region	Sample	Results	Article
Cerebellar Purkinje cells	6 ASD; 5 M, 1 F 4 CNT; 3 M, 1 F 17–54 years	Reduction in Purkinje cells number in 3 of the 6 ASD cases.	*Whitney et al., 2008*
Frontal, parietal and temporal cortices	10 ASD, 15 CNT 10–46 years	Increased spine density in layers II and V, especially in temporal cortex.	*Hutsler et al., 2010*
Dorsolateral and medial prefrontal cortices (PC)	7 ASD; 6 CNT; M 2–16 years	More neurons (67%) in PC in ASD with increased brain weight (17.6%)	*Courchesne et al., 2011*
Fronto-insular cortex	4 ASD; 2 M, 2 F 3 CNT; 2 M, 1 F 4–11 years	58% more neurons in ASD compared to controls.	*Santos et al., 2011*
Dorsolateral prefrontal cortex	8 ASD; 6 M, 2 F 8 CNT; 7 M, 1 F 4–66 years	Smaller pyramidal neuronal size (18%) in Brodmann areas(BA) 44 and 45 in ASD compared to controls. No change in cell number.	*Jacot-Descombes et al., 2012*

Table 3. Neuropathological findings in post mortem brains in ASD patients from 2003 to 2012. Adapted from Gadad et al, 2013[48].

Also abnormalities of brain white matter and in lesser extent in grey matter volume, which may lead to an increased brain size and macrochepaly, have been described[49,50]. At cellular level, alterations in neuron morphology, migration and maturation have been observed in these tissues[51].

Another alteration observed in ASD takes place in the frontal cortex neurons, which are distributed in cortical columns. This columns are found smaller but increased in density in patients causing higher local-connectivity but weak long-connectivity between the regions, known as the underconnectivity theory[52]. This theory could explain decreased joint attention, language ability and social behaviour with stereotyped patterns in ASD[50]. Added alterations that also would contribute are dysfunctional axon number and pathfinding, and synaptogenesis in the region[50].

Finally, altered synaptic mechanisms and circuitry (neurotransmission) have been reported in numerous ASD brain regions[51]. Many genes involved in synapses, reviewed in the following sections, are candidate genes implicated in ASD. Studies in animal models and recent exome studies (WES) reinforce the synaptic processes as a mechanism leading to ASD[53–55].

1.5. Heritability

ASD has a large genetic component. Heritability is calculated from the relative recurrence risk data, which is high (50-90%) among identical twins, but decreases to 0-30% for dizygotic twins, and 3-26% for siblings[56–60]. The enormous recurrence risk in MZ twins, which may share even germline mutations, compared to DZ twins, who share less genetic background, suggest a substantial genetic influence. Additionally, half siblings show two folds less concordance than full siblings[61], proving a genetic weight greater than 50%. Besides, for a second child, there is a two-fold increase in the probability of ASD diagnosis and the risk varies with sex, being three times higher for a male than a female[59]. The risk is even higher for a third born. Finally, the presence of milder autistic phenotype in patients' relatives, showing sub-clinical traits or language delay with autistic speech indicate indeed a genetic effect[59,62].

1.6. Genetics

The genetic architecture of ASD is complex and heterogeneous. Genetic encompasses from chromosomal alterations, insertions and deletions (indels), triplet base expansions, rare de novo or inherited single nucleotide variants (SNVs), copy number variants (CNVs), and single nucleotide polymorphisms (SNPs). Including the recent discoveries with next generation sequencing, the findings explain roughly the aetiology of 25-35% of the cases, remaining idiopathic the other 65-75%[63,64](Figure 4). Intriguingly, even the most common genetic event account for only 1-2% of cases[65].

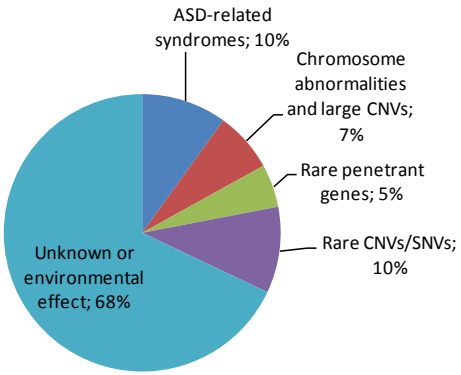


Figure 4. Genetic risk factors contributing to ASD. Genetic contributions can be modified directly or indirectly by environmental influences. The data differs from the genetic architecture proposed from Devlin et al., 2012[64] as recent discoveries and updated bibliography is reviewed (percentage found in this genetic section).

Comorbid conditions

Approximately 10% of ASD present a single gene disorder with ASD overlapping phenotype. The most commonly comorbid disorder is Fragile X syndrome (1-2% of syndrome incidence in ASD), caused by a CCG repeat expansion producing abnormal hypermethylation and transcript silencing in *FRM1*, which regulates synaptic plasticity by protein translation regulation[66,67]. Other monogenic disorders include Tuberous Sclerosis (1-4% of syndrome incidence in ASD) with deleterious mutations in *TSC1* and *TSC2* genes leading to tumour-like lesions in skin and multiple organs[68], and Neurofibromatosis type 1 (1.4% of ASD) encompassing mutations in *NFM1* and *NFM2* leading to various type of tumours and ophthalmological problems, as well as ASD[69] (Table 4). Besides, *PTEN* macrocephaly syndrome associated to a broad group of disorders known as PTEN hartoma tumour syndromes, is also associated with ASD having macrocephaly (1-17% of cases with PTEN mutations had ASD)[70]. These genes have been implicated in intellectual disability and in a less proportion to epilepsy, indicating that the diseases might share common basis[71]. Chromosomal abnormalities are also present in ASD comorbid diseases, as Down syndrome (5.6-8% of down syndrome patients have ASD)[30]. Finally, metabolic disorders showing ASD traits include untreated phenylcetonuria, Smith-Lemli-Opitz-Syndrome (SLOS) produced by a deficiency in a cholesterol enzyme (46-86% of patients present ASD)[72], creatinine deficit syndromes (CCDs) and mitochondrial disorders with

CNVs and mutations affecting crucial mitochondrial enzymes[73]. In fact, more than 100 syndromes are related to ASD[71] and in Table 4 we include the most prevalent ones.

Comorbid disorders	Gene/chr	Prevalence	Incidence of the syndrome in ASD	Incidence of ASD in the syndrome
Fragile X syndrome	*FMR1*	1/3500–1/9000	2.10%	18–33%
Tuberous sclerosis	*TSC1, TSC2*	1–1.7/10,000	1–4% (8–14% if seizures present)	25–60%
Neurofibromatosis type 1	*NF1*	1/3000–1/4000	<1.4%	4%
Untreated phenylketonuria	*PAH*	1/10,000–1/15,000	–	5.70%
Smith–Lemli–Opitz syndrome	*DHCR7*	1/10,000–1/60,000	<1%	46–86%
Down syndrome	chr21 trisomy	1/1000	1.7–3.7%	5.6–8%

Table 4. Some of the genetic syndromes associated with ASD. Adapted from Persico et al[74].

Recurrent Rearrangements

Traditional approaches as linkage analyses studying polymorphic loci and their segregation, reported few results in ASD despite not following parametric inheritance as Mendelian inheritance[75]. However, some successful results characterized several regions, as 7q21–q32 (known as ASD susceptibility locus AUTS1)[76] and 17q11–17q21 which have been replicated with enough support from independent studies[77]. Additionally, traditional approximations to genetics (e.g. Karyotype) used methods to characterize chromosomal abnormalities. The most observed cytogenetic abnormality is the maternal duplication of 15q11-q13, region for Prader-Willy/Angelman Syndrome when deleted, altered in 1-3% of ASD patients[78].

With the emergence in 2004 and development of microarray technologies allowed the detection of sub-microscopic copy number variants (CNVs) in the genome[79,80]. CNVs are rare genomic structural changes, duplicated or deleted regions greater than 1kb, *de novo* or inherited. The frequency in ASD of inherited CNVs proved to be equal than controls[81,82], but for rare *de novo* CNVs the frequency was higher in ASD cases than controls[81,83]. These initial studies plus following studies estimated that 8% of sporadic ASD cases carried a *de novo* CNV, compared to only 2% of their unaffected siblings and unrelated controls[84–86]. Therefore, *de novo* mutation plays an important role in ASD aetiology, representing highly deleterious and

penetrant mutations. The number of predicted genes is large (from 130 to 234 loci), although each gene contribute in a small number of cases[63,86]. Beside these rare CNVs, few recurrent microdeletions syndromes with ASD phenotype have been found, affecting regions 1q21, 2p15-p16.1, 15q13 and 16p11.2, and with a frequency of occurrence less than 1% each[64](Table 5). These CNVs are linked to a broad of clinical characteristics and symptoms and also to other psychopathies, presenting variable phenotype expressivity and making it difficult to be the unique cause of ASD. The most recurrent CNV is 16p11.2[87] (0.8% with deletions more penetrant than duplications), which is also observed in other diseases like developmental delay, obesity, and non-ASD psychiatric disorders.

To summarize, structural genomic rearrangements, detected by karyotype or array aCGH, are found in 5.8-8% of ASD cases and located in mostly all chromosomes[81,88]

Susceptibility variants

Several large whole-genome association studies have been assessed to detect ASD common genetic variants (CVs) (variants with a Minor Allele Frequency over 1%)[89–91]. Two large studies assayed 2 million SNPs and found association in 5p14.1 disrupting a non-coding RNA[91] and in 5p15.2, in a intergenic location[89]; a third one assayed 1 million SNPs and found association in 20p12.1 located in an intron of *MACROD2*[90]. Unfortunately none of these studies had overlapping risk locus, concluding that all the common variants obtained contribute in a small size effect to ASD risk[92,93]. Besides, numerous studies provided associated variants with solid evidences in ASD candidate genes, curiously being most of them intronic SNPs[77]. The weight that the recurrent SNPs carry in ASD heritability is discussed; a recent study showed that all common variants considered in combination (acting in widespread epistasis) contribute significantly to the aetiology of ASD[60], even reaching a 50% contribution; overestimated compared to other studies[60] and the large GWAS previously discussed[92]. It should be taken into account that the cohorts tested in GWAS studies need to be tested in larger cohorts.

Other susceptibility variants encompass rare inherited CNVs. It has been described that rare inherited CNVs were not enriched in ASD patients versus siblings, but showed a contribution to the disease due to the greater enrichment of biological relevant pathways. Additionally, some rare inherited CNVs, present only in cases, overlapped *de novo* CNVs[85]. Besides, significant effects for transmission of rare CNVs were also reported[94].

Monogenic events

Besides to the comorbid conditions with mutations in single genes (e.g. *PTEN, TSC1*...), many monogenic events have been described. High penetrant rare mutations have been identified causing ASD, or increasing considerably the risk, in many genes, mostly involved in synaptic processes. Gene families include neuroligins, SHANK and neurexin genes crucial in synapse formation and stabilization. Post-synaptic neuroligins (*NLGN3, NLGN 4, NLGN4Y*) interact with pre-synaptic neurexins (*NRXN1, NRXN2, NRXN3*) stimulating the formation of presynaptic bouton[95]. Besides, within the neuron, neurexins interact with the scaffolding proteins (*SHANK2* and *SHANK3*)[96]. *SHANK3* is involved in the synaptogenesis, development and maintenance of excitatory synapses in the neuronal postsynaptic density and the deletion encompassing *SHANK3* is found in 0.85% of ASD[97]. Deletion of *SHANK3* causes Phelan-McDermid syndrome, a developmental disorder characterized by language impairments, global developmental delay, and autistic behaviour (Table 5).

Locus	Band	CNV size	N of genes	N of events in cases; controls	*f* in ASD (n=2120 males)
CNV-16p11.2	16p11.2	700 kb	30 genes	18/2120; 3/2159	0.8%
CNV-7q11.23 (Williams–Beuren syndr)	7q11.23	~1.4 Mb	24 genes	4/2120; 0/2159	0.2%
CNV-22q11.2	22q11.2	~2.5 Mb	56 genes	4/2120; 1/2159	0.2%
CNV-1q21.1	1q21.1	~1.5 Mb	14 genes	4/2120; 3/2159	0.2%
CNV-15q13.3	15q13.3	~1.5 Mb	6 genes	5/2120; 0/2159	0.2%
CNV-15q11–q13	15q11–q13	~5 Mb	12 genes	2/2120; 0/2159	0.1%
PTCHD1/PTCHD1AS	Xp22.11	~1 Mb		10/1807; 0/786 M	0.5% (0.6%)
NRXN1	2p16.3	≥2 exons		9/2120; 1/2159	0.4%
SHANK2	11q13.3	≥2 exons		2/2120; 0/2159	0.1%
SHANK3	22q13.33	≥2 exons		1/2120; 0/2159	0.05%
NLGN3	Xq13.1	≥2 exons		1/1807; 0/786 M	0.05% (0.06%)
NLGN4X	Xp22.3	≥2 exons		1/1807; 0/786 M	0.05% (0.06%)

Table 5. Loci and genes often affected by CNVs in ASD. Based in two large CNVs studies of Pinto et al.2010[98] and Sanders et al.2011[85]. Table adapted from Devlin et al. [64]

The number of variants is increasing vastly since next-generation sequencing (NGS) technologies appeared in 2005. One issue arising from NGS is *de novo* variants interpretation.

Interpretation is difficult due to the lack of statistical support for rare variants. Only a fraction of these mutations are expected to be causative of ASD phenotype and the rest would be risk-conferring. Reports advise that given a rare variant, it cannot considered causative unless a) a large cohort is studied and achieve statistical significance (new statistical methods for smaller cohorts are needed), b) incorporate functional effect, c) gene expression data or d) protein interactomes, adding information to evaluate its consequences[93].

The first four studies of WES found the same frequency of *de novo* mutation in ASD than controls[99–102]. However, when studying *de novo* disrupting mutations (LoF mutations), the studies found that mutations affected 20% of ASD compared to 10% of the unaffected siblings[100,102] and therefore, implicated Loss of function mutations (LoF) (also called likely gene disrupting mutations (LDG)). LoF are defined as truncating mutations which create stop codons (nonsense), cause frameshift or alter splice-site preventing complete/functional protein. Another conclusion found was that LoF added to small insertions and deletions (indels) would contribute to the phenotype at least in 10% of ASD, suggesting an important role in ASD aetiology[100,102]. Recently, two WES studies analyzing more than 2500 and 3871 simplex families[55,103] found genes recurrently altered. Across these and other studies, thirteen genes with de novo LoF mutations with proven validated risk have been described associated to ASD (Table 6), estimated to contribute to 1% of ASD cases[74]. Aminoacid-altering missense mutations also contribute to the phenotype although not many association has been found, e.g. recent missense mutations in CDH8[55]. Among these genes there's a relatively small number of implicated pathways: synaptic function, transcription and splicing genes, and chromatin remodeling[55]. Given that only a fraction of mutations are expected to be causative, recent reports estimate that 13% of missense and 43% of LoF contribute truly to the diagnosis of ASD[103].

Appart from the LoF variats, mutations in non-coding sequencing explored by whole genome sequencing (WGS), may contribute to explain part of the ASD missing heritability[104].

adjusted p.val (FDR)	New ASD candidate genes with recurrent de novo disrupting (LoF) mutations
<0.0005	*CHD8*
<0.005	*ARID1B, DYRK1A, SYNGAP1*
<0.01	*ADNP, ANK2, DSCAM, SCN2A*

adjusted p.val (FDR)	New ASD candidate genes with recurrent de novo disrupting (LoF) mutations
<0.05	*CHD2, GRIN2B, KDM5B, POGZ, SUV420H1*
<0.2	*ANKRD11, ASXL3, ASH1L, BCL11A, CACNA2D3, CUL3, DIP2A, FOXP1, GIGYF1, ILF2, KATNAL2, KDM6B, MED13L, NCKAP1, PHF2, RANBP17, RIMS1, SPAST, TBR1, TCF7L2, TNRC6B, WAC, WDFY3, ZC3H4*

Table 6. ASD risk genes with disrupting LoF mutations discovered by Next Generation Sequencing. Reviewed by State et al, 2015[105].

The number of genes carrying de novo mutations (encompassing CNVs and SNVs) implicated from 200 to 1000 genes in ASD[106], and many of these genes have been associated not exclusively to ASD, but also to neuropsychiatric disorders indicating the possibility of having common pathways altered. The disorders include schizophrenia, attention deficit-hyperactivity disorder (ADHD), epilepsy and intellectual disability (ID) among others, pointing common pathways altered across these multiple neuropsychiatric disorders. One of the first mutation-centred approaches across multiple disorders was in 2009, leading to synaptic shared patways[107]. Many studies arose since then elucidating genes with alterations in several genes or pathways in multiple neurodevelopmental disorders[55,65,71,103,108].

Biological pathways altered in ASD

A way to determine the consequences of the ASD causative genes is to study their enrichment in molecular function, biological function, protein-protein interactions and perform pathways analysis to find common deregulated functions or pathways. For example, functional classification of SFARI and AutismKb databases candidate genes indicated that 61% are expressed in brain and 14% are involved in synaptic function[109]. In Table 7 the main genes associated to ASD have been grouped into different functional classes: neuronal activity, regulation and development; sodium and calcium channels; cell growth regulation; transcription, translation, and protein degradation (ubiquitination); chromatin remodelling genes and transcription factors.

Function	Gene Symbol	Protein Name	Band
Neuronal cell adhesion and/or synapse function	*NLGN3*	Neuroligin 3	Xq28
	NLGN4X (NLGN4)	Neuroligin 4	Xp22.33
	NRXN1	Neurexin 1	2p16.3
	SHANK3	SH3 & multiple ankyrin repeat domains 3	22q13
	CNTNAP2	Contactin-associated protein-like 2	7q36
	CNTN4 &CNTN3	Contactin 4 & Contactin 3	6p26-p25
	PCDH10	Protocadherin 10	4q28
	NRCAM	Neuronal cell adhesion molecule	7q31
	STXBP1	Syntaxin binding protein 1	9q34.11
	GRIN2B	Glutamate receptor	12p13.1
Neuronal activity regulation	*MECP2*	Methyl CpG binding protein 1	Xq28
	UBE3A	Ubiquitin protein ligase E3A	15q11-q13
	DIA1 (c3orf58)	Deleted in ASD	3q
	A2BP1	Ataxin 2-binding protein 1	16p13
Neuro-developmental genes	*EN2*	Engrailed 2	7q36
	HOXA1	Homeobox A1	17p15.3
	HOXB1	Homeobox B1	17q21-q22
	RELN	Reelin	7q22
	WNT2	Wingless-Type MMTV Integration Site Family Member 2	7q31
	FOXP2	Forkhead Box P2	7q31
	ARX	ARX homeobox gene 5	Xp22.13
	PTCHD1	Patched domain containing 1 gene	Xp22.11
	SYNGAP1	Synaptic Ras GTPase activating protein 1	6p21.32
Sodium channels	*SCN7A*	Sodium channel, voltage-gated, type VII	2q
	SLC9A9(NHE9)	Na+/H+ exchanger isoform 9	3q24
Calcium channels	*CACNA1C*	Calcium channel voltage-dependent L type alpha 1C subunit	12p13.3
	CACNA1H	Calcium channel voltage-dependent alpha 1H subunit 6	16p13.3
	CACNA1F	Calcium channel voltage-dependent L type, alpha 1F subunit 7	Xp11.23
Neurotransmitter genes	*GABRB3/A5/G3*	GABA receptor subunits	15q11.2-q12
	SLC6A4	Serotonin transporter	17q11.1-q12
Mitochondrial	*SLC25A12*	Mitochondrial aspartate/glutamate transporter	2q24
Signaling genes	*DYRK1A*	Dual-specificity tyrosine-(Y)-	21q22.13

Function	Gene Symbol	Protein Name	Band
		phosphorylation regulated kinase 1A	
	PTEN	Phosphatase and tensin homolog	10q23.31
	MAPK3	Mitogen-activated protein kinase 3	16p11.2
	MVP	Major vault protein	16p11.2
protein regulation	*CUL3*	Cullin 3	2q36.2
	KCTD13	Potassium channel tetramerization domain containing 13	16p11.2
	CYFIP1	Cytoplasmic FMR1 interacting protein 1	15q11.2
	RNF8	RING finger protein 8	6p21.3
transcription regulation (chromatin regulation, DNA methyltransferases and transcription factors)	*SATB2*	SATB homeobox 2	2q33.1
	ARID1B	AT rich interactive domain 1B (SWI1-like)	6q25.3
	CHD8	Chromodomain helicase DNA binding protein 8	14q11.2
	CHD2	Chromodomain helicase DNA binding protein 2	15q26.1
	SETBP1	SET binding protein 1	18q12.3
	ADNP	Activity-dependent neuroprotector homeobox	20q13.3
	DNMT3A	DNA (cytosine-5-)- methyltransferase 3 alpha	2p23.3
	KANSL1L	KAT8 regulatory NSL complex subunit 1-like	17q21.31
	SETD5	SET domain containing 5	3p25.3
	TBR1	T-box, brain, 1	2q24.2
	TBL1XR1	Transducin (beta)-like 1 X-linked receptor 1	3q26.32
	ZMYND11	Zinc finger, MYND-type containing 11	10p15.3
	FOXP1	Forkhead box P1	3p13
Other genes	*OXTR*	Oxytocin receptor	3p26.2
	LAMB1	Laminin beta 1	7q31.1

Table 7. Summary of genes involved in ASD[104,110].

One of the first pathways derived from gene onthology analyses for ASD were neuronal migration, synaptic development and axonal guidance[53], obtained studying candidate genes with CNVs using weighted functional background network (WGCNA). Additionally to these synaptic pathways, recently, signalling pathways (e.g. PI3K-Akt-mTOR and PI3K-RAS-MAPK), chromatin remodelling and transcription regulation are being found enriched in studies with

genes with rare *de novo* CNVs[86] (Figure 5) and in WES[55]. Another important breakthrough in ASD was the finding of *"Fragile X mental retardation Protein (FMRP) associated genes"* enrichment, which were first found in genes with rare variants from a WES[100]. The genes included *MET, PTEN, TSC1*, *TSC2* and *NF1* which are part of the post synaptic density and regulate protein translation via mTOR/PI3K pathway. For the proteins involved in the neuronal pathways implicated in ASD, see Figure 6.

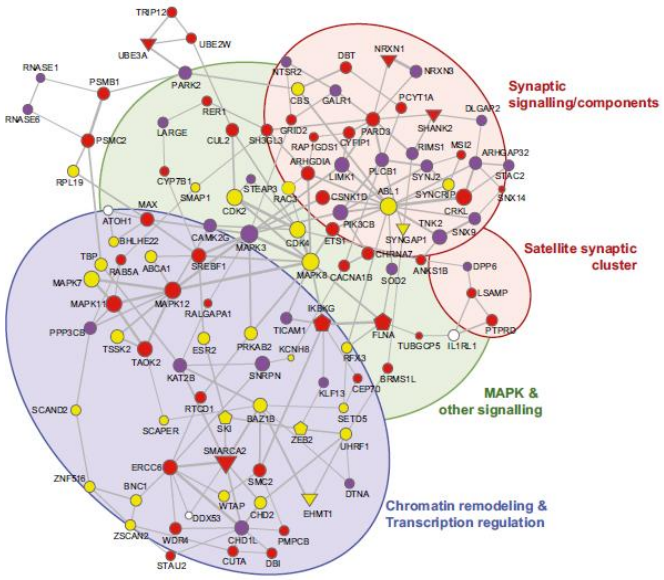


Figure 5. ASD altered genes by CNVs that converge on functional gene networks (adapted from Pinto et al. 2014[86] study).

Given the known pathways altered in ASD, research is moving a step forward to the "in vivo" neurobiology. This includes the study of genes in different brain regions, developmental stages and cell types in the pathology, which are called contextualized analyses. The first outcomes of these research arise in spatial and temporal studies identified chromatin transcription regulation during prenatal development as an important component for ASD[112]. Tools like Brain Span atlas for studying transcriptome in different areas and developmental stages is helping to this convergent neurobiology.

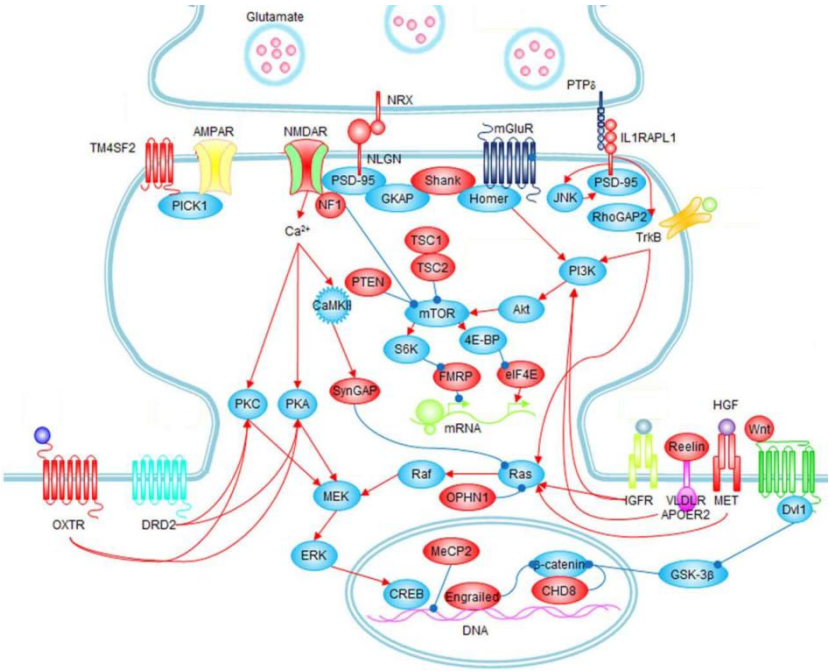


Figure 6. Main molecules involved in neuronal pathways altered in ASD (red) and related genes (in blue). Red arrows indicate inhibition and blue arrows activation. We see a representation of the pre and post synapses and gene transcription in the nucleus of the neuron. Genes include release of neurotransmitters, organization of the post synaptic density, scaffolding proteins, signalling cascades (FMR1 and mTOR), and transcription regulation (including epigenetic regulators). Modified from Won et al., 2012[111].

1.6.1. ASD genetic inheritance models

ASD genetic inheritance models have changed over time. Recently, the discovery of a great number of contributing *de novo* CNVs and SNVs made the scientific community reconsider the importance of each genetic variant (inherited or *de novo* and rare or common CNVs/SNVs) in the phenotype and rethink the previous genetic models proposed. At the same time, the complete rethinking brings to light the weight of environmental effect in the aetiology. The main proposed models are: Mendelian model, polygenic model and complex model with contribution of environmental factors[93].

Mendelian model

Genes with high penetrance causing syndromic ASD, and genes with SNVs found in cases but not in controls would follow this model under dominant, recessive or X-linked model[93].

Although, incomplete penetrance and variable expressivity would suggest additional factors: genetic, epigenetic, and environmental.

Polygenic model

Several variants/events are combined in multiple sub-models, all of them with evidences:

- Common variants: "A common disease equals common variants", true for many diseases, although arguments against are in ASD are the small size effect of common variants in large GWAS studies.
- Combination of common and rare variants: A major effect of rare variants added to a common variants background, in a model based on common variants explaining the main part of the genetic variability (>41%) with a lower weight for *de novo* variants (explaining <17% of the genetic variability, but with large deleterious effects)[60,113] (Figure 7). Examples of this model are two patients with 15q11-q13 CNV which also carried *de novo SHANK2* mutations[114], or ten patients with 16p11.2 microdeletion also carried an additional large CNV[115].
- Double-hit models of rare variants: Sebat et al.[83] proposed that the risk for ASD depended of spontaneous mutations in the affected child attributed to the intrinsic mutability of genome (many sites have more than tenfold the mutation average rate), added to the genetic background and a combination of the environment context, that would affect mutation rates in the male germline[104], as ASD risk increases in relation to the age of the father[116] and the gradual accumulation of new mutations in the father's germline[117]. Recent WES provide more evidence to the theory[103]. In this context, low-risk families would be affected by a combination of *de novo* rare variants and high-risk families transmit highly penetrant alleles, and would be affected by new *de novo* mutations[94].

An hypothesis has been formulated for the polygenic influence of a continuous and heterogeneous model covering all the above mentioned options, with combination of common and rare variants, that would exceed a threshold, would give an ASD phenotype[93].

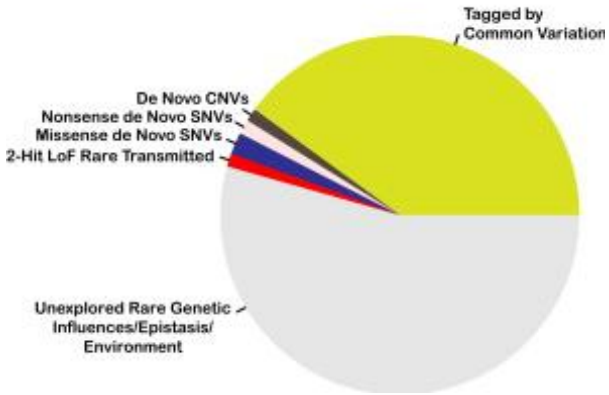


Figure 7. Genetic variability architecture and non-genetic risk factors contributing to ASD aetiology, including rare and common variants, and unknown genetic risk and environment weight. Figure from Geschwind et al., 2013[113].

Multifactorial or complex model

A complex disease involves multiple genes in combination with environmental factors. The knowledge gap in the genetic and the environmental weight in the ASD aetiology remains in concordant twin/sibling risk. ASD nature for monozygotic twins discordant for the disease could be explained by environment –epigenetics and in utero and postnatal events- or early somatic mutation[118]. Alike, dyzigotic twins, who can be considered equal as siblings in the genetic weight, but which have shared the same in utero environment, report a higher ASD risk than their sibling (although not in all prevalence studies[119]), which could be due to environmental factors or shared germline *de novo* events. Moreover, a twin study in California showed a larger proportion of variance by shared environmental factors (55%) as opposed to genetic heritability (37%)[120], concluding that shared environment plays a larger role than genetic factors. However, not a major weight to environment, although it would play a role in the disease, is given in other studies[60,83,113]. Additionally, it has been described a MZ twin pair with Rett Syndrome, a comorbid disease with ASD phenotype, that had the same parental-origin germline mutation in *MECP2* and after screening for other genetic variants or environmental factors, they only could find upstream differences in DNA methylation[121].

Finally, ASD having penetrant mutations show heterogeneous phenotypes, e.g. eight truncating de novo *CDH8* mutations were found in five individuals with intellectual disability whereas three had IQ>90[122], the variability in the IQ was also observed in *NLGN4X*

mutations[109]. These phenotypic variability could imply not only contribution of the genetic background, epistatic effects and incomplete penetrance but also the environmental context[74,104,109].

Arising questions to be solved in ASD research is the possibility of interactions of aetiology models[63]. Sebat[123] concluded that there is not still a unique model for the ASD aetiology and that the understanding of the neurobiology which points to convergent neurodevelopment and shared altered pathways (e.g. synapses) reconceptualise the approaches to psychiatric diseases.

1.7. Enviromental causes

Environmental influences through gene-environmental interactions or epigenetic modification of gene expression are the link between genetics and environment. Epigenetic mechanisms are crucial for reprogramming and cellular differentiation in fetal development, and the environment can have an important impact over epigenetics at that stage affecting the processes[124,125]. The *in utero* environment, perinatal and postnatal periods are crucial stages that have been involved in ASD[126–128]. However, no single major environmental agent has been proved to definitely contribute to ASD.

1.7.1. Environmental risk factors (prenatal, perinatal and postnatal)

Many external agents' exposures have been reported to increase susceptibility to ASD. Evidences for their effect come from epidemiological studies and animal models, which give convergent, or sometimes contradictory, results. One of the most relevant epidemiological study in ASD encompassing many environmental risk factors is the CHARGE study[129].

Some examples of the environmental agents' involved in the appearance of ASD symptoms would be the exposure to heavy metals (mercury and arsenic among others)[130], pesticides[131], and chemicals such as polybrominated diphenyl ethers (PBDEs) and polychlorinated biphenyls (PCBs). For the latter chemicals, they inhibit or disrupt the thyroid hormone action affecting critically brain development[132]. Additionally, air pollution including 23 risk parameters

identified by the US Environmental Agency is implicated in the pathology[133,134]. The external factors may act as an additive effect that triggers ASD in combination to the special susceptibility of certain individuals to environmental factors and to certain genetic backgrounds[135]. For example, an interaction of air pollution and a SNP in *MET* promoter increased significantly the risk of suffering ASD[136], reinforcing the gene-environment interaction.

Furthermore, prenatal and perinatal environment are of special concern. Pharmacological agents administered in pregnancy, such as thalidomide[137] and valproic acid (VPA) [138], known to be teratogens, and ethanol[139] are known to contribute to the disease. Mouse models in which the antiepileptic drug VPA has been administered in utero, promoted ASD-like traits and impaired synaptic connectivity in medial prefrontal cortex in the offspring [140,141]. Additionally, factors having an influence in perinatal period include caesarean section, birth complications, and even rare associations as births in summer period -due to deficits in melatonin-, among others[126].

Other factors as diet-influenced factors (low levels of vitamin D, folate[142] and zinc[143]), maternal diabetes have been described[130]. Also, excess of folate supplement could have adverse effects in ASD[144].

In addition to the early developmental stages (pre and peri natal) also there's an influence of the postnatal exogenous and endogenous factors. In ASD it has been observed that the methionine metabolism pathway with several impaired enzymes, lead to increased vulnerability to the extrinsic and intrinsic oxidative stress[145]. In fact, the pathway deregulation has consequences to the hypomethylation of DNA, being an example where genetics, epigenetics and environment interact[146]. Infections (e.g. Rubeolla) and vaccines effects have also been argued. Measles-mumps-rubella (MMR) vaccines is an example of traditionally associated factor which has been proved to not contribute to the disease[147].

Finally, psychological stress and depression in mothers[148], as well as psychological effects of population migration movements[149], also could have an impact on ASD aetiology.

2. Epigenetics

2.1. Basic concept

Conrad H. Waddington first described in 1940 the term 'epigenetics' as "…the interactions of genes with their environment which bring the phenotype into being" with the observation of vertebrate embryos development[150]. Additionally, he defined epigenetic landscapes as a result of natural gene selection and interactions leading to a canalized development. Holliday and Pug disassociated the term from development thirty years later with the discovery of the covalent chemical DNA modifications cytosine-guanine (CpG) methylation, defined as a stable layer of information with hereditary material. They also observed epigenetic control of gene activity and DNA-protein interactions [151,152].

In the 1990s the epigenetics suffered an expansion, the heritable nature was better defined unravelling genomic imprinting (parent-of-origin chromosome inactivation) and the X chromosome inactivation in mammals[153,154]. Also came one of the accepted definitions of epigenetics which describes epigenetics as the study of heritable changes in gene expression without a change in DNA sequence and included gene regulation in the definition[155]. Anomalies in the epigenetic regulatory machinery have been associated to several diseases such as cancer, autoimmune disorders, diabetes, and also neurodevelopmental disorders including Prader-Willi/Angelman syndrome, intellectual disability, schizophrenia, Alzheimer's disease, and ASD[156]. The term 'epigenomics' includes a global view of the following levels: the covalent modifications of the cytosines and histones, certain transcripts and nucleosomes positioning affecting chromatin positioning (Table 8). The epigenome at the same time, has an impact to the transcriptome (actively transcribed DNA to RNA), RNAome (including non coding RNA), proteome and metabolome[113].

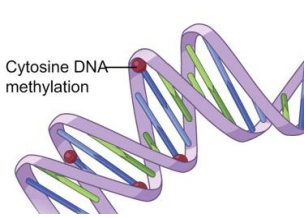
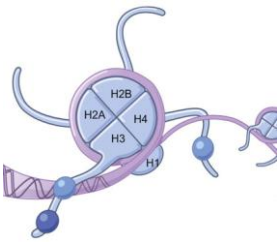
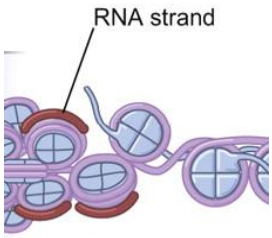
		DNA methylation	Histone marks	Non-coding RNA
				
Involved machinery or mechanisms	**Writers**	DNMT1 (meth. maintenance), DNMT3a/b (*de novo* meth.) Others: DNMT3L, DNMT1o	Histone acetylases (HATs), Methylases, Kinases	RNA condensation or clouds (e.g. Xist in ChrX), molecular decoy (trapping of molecules), molecular scaffold (binding of protein in complexes)
	Erasers	TET and TDG proteins	Histone deacetylases (HDACs), demethylates, phosphatases	
	Readers	MBD proteins, CFP1	Bromodomain proteins and many other proteins	
	Processes involved	Genomic stability Gene expression regulation Embryo development ChrX inactivation Imprinting Cell linage maintenance Nucleosome organization	Chromatin conformation and chromosome condensation Organization of DNA-based processes (replication, repair, transcription regulation) Nucleosome positioning	Chromatin state regulation Expression regulation Imprinting regulation Fast long-distant responses

Table 8. Epigenetic layers: methylation, histone marks and non coding ((ncRNAs) which include many types of RNA, being our focus the long non-coding RNAs (lncRNAs)). The table is a summary of the involved machinery (writers and erasers enzymes and reader domains) and cellular processes in which the marks have a role, which is reviewed more extensively in the following sections.

Among the processes regulated, epigenetics plays a role in RNA expression, DNA-protein interaction, and transposable mobility, in embryogenesis, cellular differentiation, X chromosome inactivation, genomic imprinting and tumorigenesis. Endogenous changes in epigenetics occur in early developmental stages, in tissue differentiation and with aging process. Exogenous factors as lifestyle (e.g. diet, smoking, drugs, and stress), as well as environment have an effect through epigenetics[157]. Epigenetic changes can arise at any time during life and can have a transgenerational effect; a known example is the maternal famine in Deutschland which predisposed to have smaller children, showing hypomethylation of the imprinted IGF2 region in blood, affecting up to six decades later[158]. An example of the effect of nutrition in gametes has been observed in mice, in which the oocyte maternal-age

deterioration was reduced by caloric restriction[159]. Currently, with the release of the epigenomes of 111 human beings, scientists are closer to define the epigenotypes, the epigenetic patterns of various cells, tissues, and organs in different time points, and disease states.

2.2.Mechanisms

2.2.1. DNA methylation

The addition of a methyl group to the fifth carbon of cytosine (5meC) is mostly observed in Cytosine-phosphodiester-Guanine (CpG) dinucleotides in mammals, although a methyl can be added also to non-CpG sites. Methylation in human varies between populations, individuals, developmental stage or age, sex, tissues, cell types and even spatio-temporally, so the methylation landscape is tissue-specific and cell-specific.

The distribution of CpGs throughout the genome is not uniform. In 98% of the genome the CpGs are located spread from each other and in 1-2% they cluster in CpG Islands[160]. The CpG islands (CGI) are regions of minimum 200bp to several kb in length with CpG content over 50%, and with a observed vs. expected CpG densities of ≥0.6. Although CGI only account for a small percentage of the genome, they are found in 60% of human gene promoters. The largest percentage of methylation is displayed in repetitive elements, they have 50% of 5mC in brain[161] (Figure 8 a)), although genome-wide studies might update the percentage. A study with higher resolution (genome-wide methylome) analyzing a wide range of human cell lines and tissues estimated that 64.2% would be located in gene regulatory elements as enhancers and transcription factor binding sites, highlighting their regulatory function[162].

Although the genome present mostly a methylated landscape (70-80% of CpGs), around 22% of the CpGs are dynamic[162] and the methylation status of CpGs correlates usually with the underlying genomic elements where they are located. A study in human brain observed that unmethylated domains were present in promoters, CGI and first exons and that hypermethylated domains were found in repetitive elements (SINEs, LINES and LTR) and non-

first exons[161]. CGI 2Kb flanking region, known as CGI-shores, have a lower CpG content and are associated with transcription repression and is associated mostly to tissue-specific methylation and methylation reprogramming[163]. The recent release of a reference epigenome of 111 human adds more data to the CpG status of different chromatin states with genome-wide methylome sequencing (Figure 8 b))[164].

Regarding the methylation status in cell types, as we mentioned, most cell types have relatively stable DNA methylation patterns and present mostly a methylated landscape (70-80% of CpGs), except germ cells and early embryos[162,165]. Methylation can vary in neurons and fibroblasts and tissues as placenta, where specific partial methylated domains can be found (PMDs, continuous domains of <70% of methylated CpGs) showing less expression than highly methylated domains. These PMDs domains can be dynamic and change to hypermethylated state associated to increased expression in immature neurons, evidencing developmental regulated genes[166].

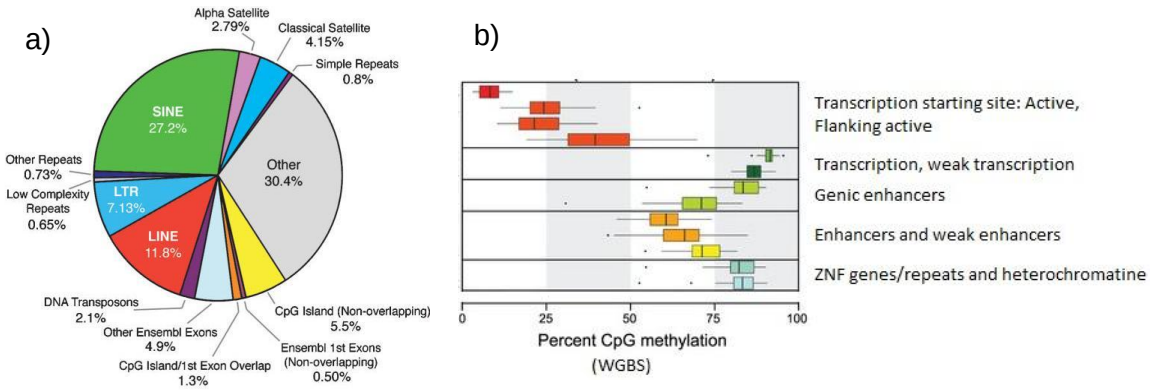


Figure 8. a) CpG distribution of the CpGs and b) methylation status depending on the different genomic elements or chromatin states. Edited from an article studying 5638 elements[161] and a recent article with 111 human epigenomes from the Roadmap epigenomic consortium[164].

Correlation between CpG methylation and gene expression depends on the genomic structure. In general methylation is associated with gene silencing. Two examples are genomic imprinting, in which the hypermethylation of one of the alleles repress its expression and leads to the monoalellic expression from the other allele, and chrX methylation in females which leads to a gene dosage reduction[167]. However, methylation can also lead to transcriptional

activation. This is the case of methylation in gene bodies, and may be necessary for transcription elongation or preventing the transcription in certain starting sites[168,169]. Correlation with expression shows a negative correlation (less methylation and higher expression) in transcription starting site (TSS), whereas in transcription end site (TES) shows a positive correlation (Figure 9). In fact correlation with expression models establish a) higher expression correlated with unmethylated promoters, methylated gene bodies and unmethylated intragenic CGI with enhancers; and b) silenced expression with the opposite methylation state in the same regions[162]. The changes in expression can occur immediately and last several hours or they can have a long-lasting effect[170].

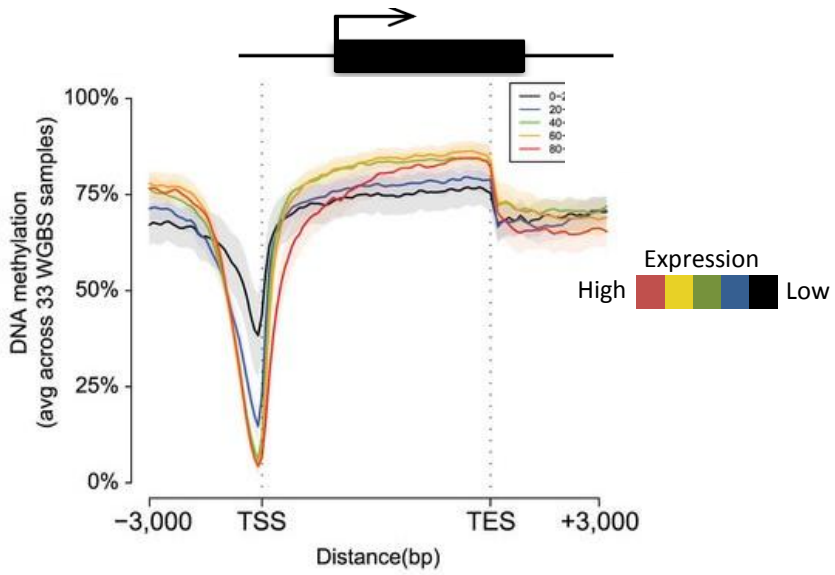


Figure 9. Relationship of DNA methylation, location from the transcription start site (TSS) and transcription end site (TES), and gene expression. Levels were assessed by genome-wide techniques and across 33 samples from different tissues[164]. For highly expressed genes a more pronounced hypomethylation is found in promoter regions (close to TSS), and higher hypermethylation in the 3' ends. Figure was modified from Roadmap Epigenomics Consortium with the study of 111 epigenomes[164].

Methylation sites other than CpG dinucleotides have been observed, called non-CpG methylation sites (CGH and CHH sites; H=A,C,T). This type of methylation account for 25% of total 5-methyl cytosine in stem cells and neurons and is thought to play a role in the maintenance of pluripotency [171], whereas in differentiated cells, non-CpG methylation sites levels are low or absent[172,173].

On the other hand, methylation forms other than 5mC have been detected, as the cytosine oxidized forms. In 2009 the discovery of 5-hydroxymethylcytosine (5hmC) found in 0.2% of the total nucleotides[174], and in 2011 the 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), signals for base excision and repair machinery, have been found in highly expressed promoters in stem cells[175]. The arousal of new high-throughput techniques to assess cytosine oxidized forms methylation status is bringing breakthroughs in the field.

2.2.1.1. DNA methylation machinery: methylases, binding proteins and demethylases.

Methylation machinery (writers)

Methylation is mediated by the DNMT protein family that transfer a methyl group from S-adenosyl methionine to DNA. In mammals, only DNMT1, DNMT3A and DNMT3B have a methyltransferase enzyme activity. DNMT1 is termed a maintenance methylase, and associated to UHRF1 protein, methylate sites generated in the replication fork, so is mostly expressed in dividing cells. It also can downregulate expression associated with HDAC2 and DMAP1[176]. DNMT3A and DNMT3B are known as *de novo* methylases, and with DNMT3L, establish the pattern in embryogenesis, so are highly expressed in embryonic stem cells and downregulated in differentiated cells[177]. Besides, these methylases help to anchor nucleosomes with hypermethylated DNA, and to methylate the sites missed by DNMT1 at the replication process[178]. Curiously, knock-out of any of the three enzymes is lethal in mice, but heterozygous mutants are viable[179,180] (Table 9). Several mechanisms are thought to help leading the methyltransferase machinery to the target DNA e.g. histone modifications and RNA or small inhibitory RNA (siRNA) which would hybridize with DNA[181].

Other ways to establish methylation states are through transcription factors. Twenty years ago it was observed that transcription factors also have a role establishing DNA methylation states, however the mechanisms are unknown[182,183].

Methylation recognition machinery (readers)

The CpGs are recognized by specific protein domains depending on their methylation status: a) methylated CpGs, especially in CGI, are recognized by methyl-CpG-binding domain proteins (MBD) which may contribute to expression silencing[184–186]. However, a more complex

regulation mechanism may be acting as MBD protein deletions do not re-establish expression. b) Unmethylated CpGs are recognized by CXXC protein domains (e.g. CXXC finger protein 1 (CFP1))[187] and unmethylated CGI for the histone demethylases KDM2A and KDM2B[188].

	Protein	Function	Mutant phenotype
Methyltransferases	Dnmt1	maintenance of methylation	embryonic lethal, LOI and X-linked gene expression, ES cells viable
	Dnmt1o	oocyte-specific isoform	loss of maternal imprints
	Dnmt2	non-CpG methylation in *Drosophila*	no phenotype
	Dnmt3a/b	*de novo* methyltransferases, establishment of methylation	embryonic lethal, ICF syndrome
	Dnmt3L	no catalytic activity, colocalizes with Dnmt3a/b	abnormal maternal imprinting
methyl binding prot.	MeCP2	methyl binding proteins, recruit histone deacetylases	Rett syndrome
	MBD1		NA
	MBD2		behaviour abnormalities
	MBD3		lethal
	MBD4	repair enzyme	increased mutation frequency

Table 9. Mouse mutants phenotypes of methylation regulatory factors. Edited from Jaenisch R. et al. 2013[189]. ES, Embrionic Stem cells; LOI, Loss of imprinting.

Demethylation machinery (erasers)

Demethylation can be lost passively by poor maintenance, or actively through the oxidation of 5mC to other cytosine forms. Through this active mechanism, ten-eleven translocation (TET) proteins convert 5mC to 5hmC, and subsequently to 5fC and 5caC, which can be effectively removed by thymine-DNA glycosylase (TDG) protein[190]. These mechanisms have been implicated in meiosis, development, stem-cell reprogramming and maintenance of imprinting. Loss of all three TET proteins leads to increased DNA methylation[191].

2.2.1.2. Role of DNA methylation

Chromosome stability

First of all, DNA methylation contributes to the genome integrity through the immobilization of transposable elements and silencing of repetitive elements. A reactivation of retrotransposons (IAP and LINE-1) has been observed in mouse embryos and germ cells lacking methyltransferases[192,193]. Secondly, methylation is crucial for chromosomal and centromere

stability, as methylation reduces recombination between homologous repeats and is required to maintain centromere functions[194]. Deletions in methyltransferases (DNMT1 and DNMT3B) induce global unmethylation and chromosomal abnormalities in humans[195,196] and telomere recombination in mouse[197]. ICF1 syndrome (OMIM: 242860), caused by a partial loss of DNMT3B, is associated with chromosomal and centromere instability, including chromosome breaks in peri-centromeric regions.

Gene expression

Methylation can regulate gene expression by different mechanisms: hypermethylation can directly prevent the binding of transcription factors and DNA binding proteins[198] and also can recruit methyl binding proteins (MBD) as MECP1 or MECP2, which can block the access of transcription factors to the DNA or facilitate repressive chromatin structures, by the recruitment of chromosome remodelling and histone modifying proteins[199] which deacetylate histones to compact the chromatin reducing gene expression. Removal of CpGs increases transcription and unmethylated regions, such as CGI, recruit methyltransferase complexes (Cfp1 and Setd1 proteins) creating domains with H3K4 histone methylation mark which generate an open chromatin domain promoting transcription[187]. However, correlation of methylation and reduced transcription, and unmethylation and overexpression not always occurs. For example, hypomethylation and expression downregulation has been reported involved in cell-type expression of developmental genes, e.g. maspin promoter[200].

Embryo development

Methylation performs a crucial role in mammalian embryo development, stage when organized DNA epigenetic reprogramming changes occur. In embryogenesis, within hours of oocyte fertilization a demethylation of the male genome occurs, whereas passive demethylation of the maternal genome in successive divisions takes longer. The process is followed by a *de novo* genome-wide methylation after implantation[201–203] (Figure 10). Hypomethylation after fertilization would lead to the embryo apoptosis[204].

Assisted reproductive techniques (ART) are controversial for the effects it can have in altering the methylation landscape in gametes[205–207]. For example, studies associate ART with imprinting disorders such as Beckwith-Wiedemann syndrome and Angelman syndrome, in

which a global hypomethylation at multiple loci was observed[208]. In ASD, a large scale study showed no association for in vitro fecundation (IVF)[209], except for IVF using intra-cytoplasmic spermatozoid injection (ICSI), which would manipulate gametes instead of embryos[210].

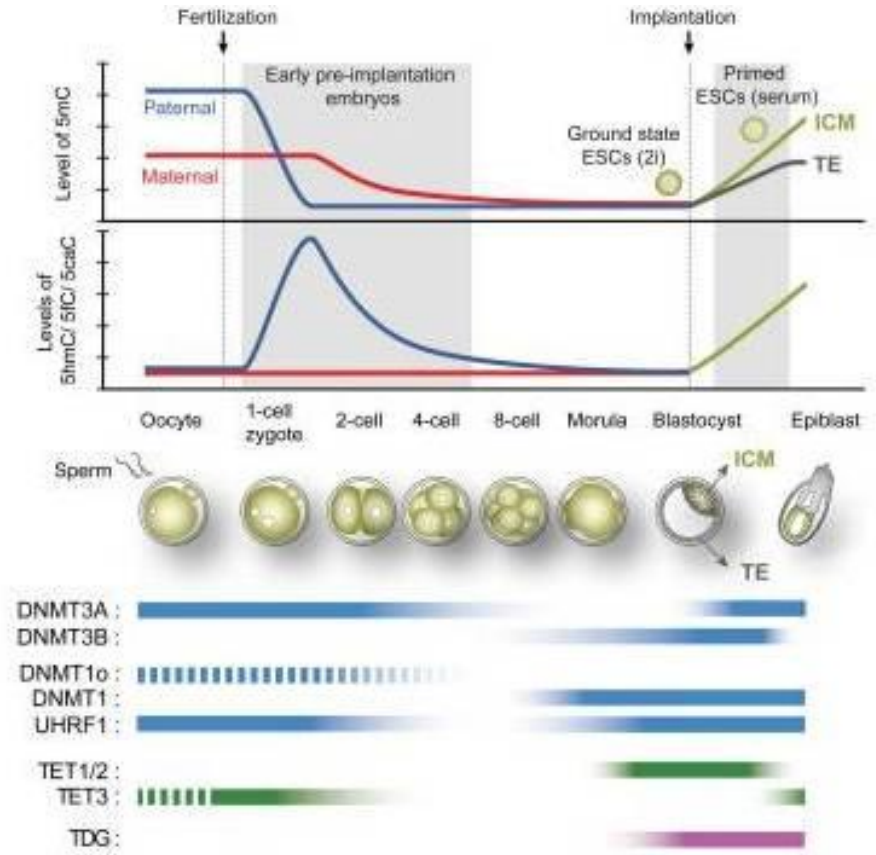


Figure 10. DNA methylation and enzymatic activity during fertilization, pre-implantation, and implantation of the mouse embryo. Level of 5mC and 5hmC methylation marks are erased after fertilization and re-established after implantation. Epigenetic machinery action is indicated with bars, showing an activity reduction after epigenetic marks removal and an increase before epigenetic reprogramming. Obtained from Wu H. et al. 2015[190].

Chromosome X inactivation

During development, in females, chromosome X inactivation is random in somatic cells, and the majority of genes in the inactive X chromosome are maintained methylated and transcriptionally silenced. Establishing X inactivation depends on Xist, a non coding transcript that recruits polycomb-repressing complexes (PRC) with methyltransferase and histone ubiquitinase activity[211]. Xist is unmethylated and expressed uniquely from the inactive X-chr,

but methylation in its promoter can vary, as well as its expression. Xist activation precedes methylation, so methylation is involved in the Chr-X inactivation maintenance, which in turn depends on DNMT1[189].

Imprinting

Methylation plays a role in genes that are expressed in a parent-of-origin specific manner, known as imprinted genes. Currently about 100 imprinted genes have been identified clustered in the regions and controlled by a coordinate mechanism from the imprint control regions (ICRs)[212]. Imprinting can take place in *cis* or *trans*. Genes imprinted regulate embryonic development, and placentation, among other processes. The imprinting takes place in the germ cells by *de novo* methylation through DNMT3A and DNMT3L, but the marks are erased during embryogenesis reprogramming, and is re-established afterwards. Other proteins added to DNMT3, as oocyte-specific isoform of Dnmt1 (DNMT1o), TRIM28 and ZFP57, would maintain the DNA methylation memory at imprints during cleavage/ reprogramming[189,212]. Some genes are imprinted in a tissue but are expressed biallelically in other tissues (ex.UBE3A, which is only expressed from the maternal allele in certain brain regions)[213]. The loss of imprinting (LOI) can lead to disease, e.g. LOI in 11p15.5 is related to cancer and Beckwith-Wiedeman syndrome, and in 15q11-q13 is associated to Prader Willy/Angelman syndrome, lacking of paternal and maternal gene transcription, respectively. To differentiate terms, methylated regions acquired post fertilization are known as differentially methylated regions (DMRs)[212].

Other processes

Methylation is involved in the maintenance of cell identity, through silencing expression of a subset of developmentally regulated genes to maintain cell lineage. Conditional inactivation of DNMT1 (the maintenance methyltransferase) in fibroblasts led to a general transcription activation of tissue-specific genes, evidencing the maintenance role of gene silencing in somatic cells[214]. Methylation also plays a role in nucleosomal and 3D organization. DNMT1 maintains the nucleolar compartment and elements are found closer, such as MeCP2 and a chromatin loop which participate for example, in expression at Dlx5-Dlx6 locus[215].

2.2.1.2.1. Genetics-epigenetic interaction

Methylation can be a source of variability among individuals dependent on the underlying DNA sequence or independently. Studying methylation in different human ethnicities evidenced that changes in methylation were mainly due to the genetic background (68% of the variants found) and a smaller percentage due to methylation variation itself (32%), all contributing to natural human variation[216]. Similarly, a large contribution of genetic sequence was also observed underlying changes in the chromatin modifications and DNA methylation[217,218].

Mutations affecting methylation targets can be found in described polymorphisms (meSNPs) (Minor allele frequency over 1%) or in rare variants (meSNVs). These mutations added to indels in regulatory regions could affect binding of transcription factor directly (allele-specific transcription factors or chromatin modifications), or alter regulatory regions as insulators/enhancers and exert a long-range effect. Besides, mutations in a CpG can influence the neighbouring CpGs to become cytosine-methylated and affect transcription factor and protein binding. These situations can lead to allele-specific methylation (ASM) and allele-specific expression (ASE) of mRNA and non-coding RNAs [219]. Some of these SNPs with a direct effect on methylation are called methylation quantitative trait loci (meQTLs), in which SNPs genotypes correlate with epigenomic profiles. Effects of meSNPs can be in *cis* or in *trans*, from5Kb up to 50Kb of distance[220], or even over 1 Mb apart or in different chromosomes [221].

2.2.2. Histone modification

DNA is compacted into chromosomes by nucleosomes -which contain the core histones- and linker histones. Depending on the chromatin condensation level, the DNA is more accessible to replication and transcription processes so histones are a major epigenetic layer. Two units of each core histone form the nucleosome (H2A, H2B, H3 and H4), and the DNA molecule is wrapped around the nucleosome helped by linker histones H1[222]. Histones have a globular structure except for tails which are post-translational modified (Figure 11). There are eight modifications described (methylation, phosphorilation, SUMOylation (adding a small ubiquitin-like modifier (SUMO)), ADP-ribosylation, ubiquitination, deimination and proline isomerisation) but is the combination of them, a total of 51 states, which establish the histone-code and the

final role[223,224]. The modifications can be located in the same aminoacid, in the same histone tail or in different tails. Histone modifications can be long-lived marks or dynamic.

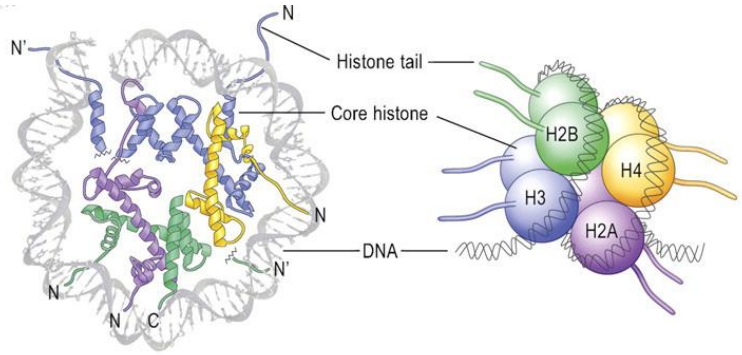


Figure 11. Nucleosome formed by the core histones subunits (H2A, H2B, H3, and H4), which show their tails which may be post-translationally modified, and a molecule of DNA wrapped around the nucleosome. Linker histones are not shown. Obtained from Alberini, 2009[225].

Many genomic elements, including promoters, imprinting regions or regions with/without transposable elements have been correlated with histone marks. For example, transposons-free sequences (e.g. Hox gene clusters), have the largest K27 domains and reflect evolutionary pressures against transposable elements. Also, most of TSSs in stem cells but not differentiated cells have K4 domains, which would define the initial epigenetic state[226].

2.2.2.1. Functional processes involving histones modifications

Histones are involved in several processes: chromosome condensation, transcriptional regulation, alternative splicing, and DNA replication and repair. To simplify, their function can be divided in the establishment of chromatin conformations and the organization of DNA-based processes[223].

Regarding chromatin conformation, DNA is accessible (euchromatin) or inaccessible (heterochromatin) for transcription. The euchromatin is actively transcribed allowing the access of RNA polymerase II (Pol II), and is triggered by histone acetylation and trimethylation of a specific set of histones, whereas the heterocromatin is transcriptionally inactive, except for certain non-coding RNA, and is signalled by low levels of acetylation and high levels of

methylation marks in certain histones[227](Figure 12). Then, models establish that H3K4me3, H3K79me1, H3K27ac and H4K20me1 can predict gene expression levels[228,229].

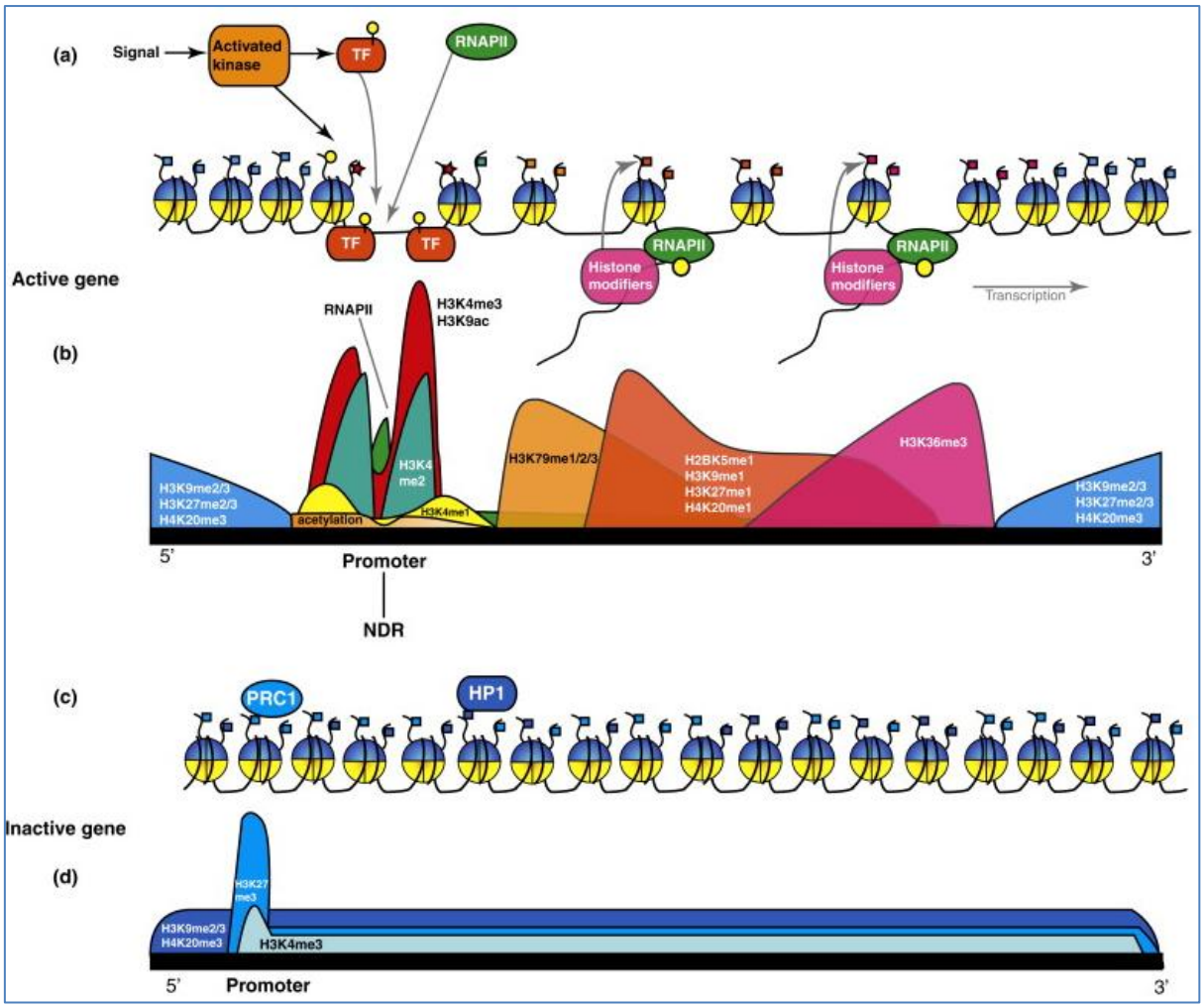


Figure 12. Nucleosome overview and modifications over the (a) active or (c) inactive gene. Histone modifications distribution in (b) active gene (acetylation and trimethylation of H3K4, H3K36 and H3K79) and (d) inactive genes (low levels of acetylation and high methylation levels of H3K9, H3K27 and H4K20), which can predict gene expression. Image from Barth T.K. et al, 2010[229].

Regarding the orchestration of processes that occur on the DNA molecule, histones modifications facilitate the display of local open chromatin domains for specific gene transcription or repair, or larger open domains for DNA replication or chromosome condensation.

Another key element, now regarded as a new epigenetic player, is nucleosome positioning. Their compaction level has a role in the initiation of transcription which also determines the accessibility to transcription factors. In fact, the loss of a nucleosome directly upstream of the TSS is correlated with gene activation[230]. Moreover, the nucleosome remodelling machinery incorporate other histones, called histone variants having different domain structure, tails and key aminoacids from core histones, which regulate the nucleosome positioning and gene expression[222,223]. Large chromatin remodelling complexes change the nucleosome, and as consequence, chromatin structures; the enzymatic families comprised in the complexes participate in various processes, from transcriptional activation, telomere regulation, chromosome recombination and segregation to DNA replication and repair[231].

Finally, the epigenetic layers are interconnected with each other, so methylation and histone modifications also influence nucleosome remodelling machinery. On the other way round, nucleosome position also contribute to the methylation landscape[232].

2.2.2.2. Mechanism of action

The mechanism of action of the processes regulated by histone modifications involve several steps such as the disruption of contacts between nucleosomes, and also the recruitment and binding of non-histones proteins with enzymatic activities, being different in each step of the process[223].

Adding and removing modifications to histones (Writers and Erasers)

Most histone modifications have been found to be dynamic and many enzymes adding or removing post-translational modifications on histones have been described. Writers' enzymes families are histone acetyltransferases (HATs), methylases, and kinases (for phosphorylation marks). For Erasers, histone deacetylases (HDACs), demethylases and phosphatases[223,233]. The activity of the enzymes can be influenced by the proteins associated (co-activators and co-repressors) which can mark a preference for nucleosomes or free histones, for the target residue or for the degree of modification it may have (e.g. mono-,di-tri- methylation)[223].

In transcribed regions, RNA pol recruits HATs which acetylate histones and facilitates transcription, whereas HDACs reset the chromatin state[234].

Reading histone modifications (Readers)

Many enzymes through different protein domains recognize a mark or a combination of marks. These proteins mediate the docking for protein effectors, or readers. As an example, acetylation is recognized by Bromodomain-containing proteins, and these proteins target histone acetyltransferases. Other histone marks are recognized by other proteins with specific domains[233,235].

Besides, DNA methylation in cytosines can interact with histones. Histone modifiers enzymes as histone methyltransferases can recruit methyltransferases DNMT enzymes and direct methylation promoting the transcription repression, and can modulate also the stability of the enzyme and regulate also the transcription. On the other way round, DNA methylation and methylation machinery can also direct histones, for example MeCP2 direct histone H3K9 and repress transcription[236], and DNMTL enzyme interacts with H3 and induce *de novo* methylation[212]..

2.2.3. Non coding RNA

Non coding RNAs can be classified in short and long non coding RNAs. Short non coding RNAs have been extensively subclassified, for example in miRNAs, which are 20-23nt in length and bind to mRNAs regulating their expression post-transcriptionally and many have been associated by mutations to ASD[237]. Epigenetic mechanisms have been defined for miRNAs, however mechanisms for lncRNAs are better characterized[238], so will focus on them. LncRNA are RNA of size over 200 bp, transcribed by the same transcriptional machinery and by RNA polymerase II, but lacking an open reading frame[239]. Around 15,000 different lncRNA were suggested to exist from the ENCODE project[240]. LncRNA is another epigenetic layer and player[211]. LncRNA not only regulate transcription, but also chromatin structure, cell differentiation, and other processes such as alternative splicing, miRNA abundance, protein-protein interactions and molecular signalling[241,242]. Many lncRNA show tissue-specific patterns of regulation[243]. Regarding transcription, other forms of RNA can regulate transcription as miRNA, circular RNAs, but only lncRNA or enhancer RNAs are considered epigenetic

layers[211,244]. Enhancer RNAs (eRNAs) are a new class of ncRNAs, transcribed bidirectionally from enhancers, which promote mRNA synthesis[245]. .

2.2.3.1. *General mechanisms of action*

General lncRNA mechanisms in *trans* include: 1) acting as a molecular decoy: binding transcription factors or miRNA and preventing their binding and action, 2) forming a RNA scaffold leading lncRNA-RNP (ribonucleoprotein) complex and 3) RNA can create a nuclear sub-domain (a RNA cloud) and target chromatin modification complexes[246] (Figure 13). For example, lncRNA can recruit repressive complexes to the chromosome X (Xist) or imprinted regions. Normally they have a repression role, but they can also upregulate transcription targeting activating complexes[247]. LncRNA can also act in *cis,* for example performing promoter occlusion or transcriptional interferences[248].

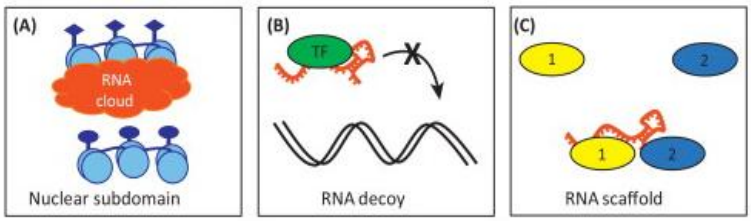


Figure 13. General lncRNA mechanisms. LncRNA can a) create RNA cloud domains, e.g.Xist subdomains in chrX inactivation; b) act as molecular decoys binding transcriptional factors or miRNAs, and c) act as a scaffold for large protein-RNA complexes. Extracted from LaSalle et al., 2013[242].

2.2.3.2. *Processes where lncRNA are involved*

Imprinted regions

Every cluster of imprinted genes contains at least one lncRNA regulated by DNA methylation. LncRNA genes promoters are located in their respective imprinting control region. For example, *Kcnq1ot1* accumulate at promoter of silenced alleles and mediate repressive histone modifications in an allele-specific manner[212].

Chromatine conformation regulation

As an epigenetic player, lncRNA regulate large domains of chromatin or chromosomes by binding to an epigenetic complex (such as Polycomb repressive complex 2 (PRC2)). For example, lncRNA together with DNA-bound factors (such as Xist), are loaded onto chromatin to silence the gene. Additionally PRC2 targets also participate in development and differentiation of stem cells[211].

Expression regulation through R loops and interference

LncRNA participate as a layer on chromosomes, modifying the structure of DNA (forming DNA-RNA hybrid or DNA-DNA-RNA tripkex loops) which protects from methylation, e.g. Prader Willi-Angelman Syndrome imprinting control[242,246]. Besides, transcriptional interferences of sense and antisense machineries by lncRNA polymerase binding might result in termination of one or both transcriptional events, as happens in Ube3a-AS which leads to paternal silencing in neurons[242].

Fast and far-reaching responses

Many examples of LncRNAs are now known, they all confer a temporal and spatial specificity, reaching far distances -not possible by proteins and small RNAs-. For example, nerurons stimulated (e.g. treatment with KCl) lead to a thousand of eRNA with unknown function transcribed bidirectionally from enhancers[244]. LncRNA adds a layer of complexity to the genome and the effect of mutations in the non-coding sequence.

2.3. Reference human epigenome

In contrast to the human genome, that has a reference sequence, variation annotation and displays association to diseases, the epigenome lacks of such a reference. With the appearance of new technologies to generate high throughput data, consortiums made considerable efforts to compile data from different epigenetic techniques characterizing various cell lines and tissues, to establish a reference epigenome.

ENCODE project

From 2003 to 2007, The National Human Genome Research Institute (NHGRI) formed a research consortium called ENCODE, the Encyclopaedia of DNA Elements on functional elements [encodeproject.org]. Functional elements included transcripts, novel non-coding RNA, new transcription starting sites, histone marks, DNA methylation, regulatory elements/modules (enhancers, insulators, silencers), chromatin accessibility, DNaseI hypersensitive sites which distinguishes histones domains from promoters, and transcription factors binding domains among others. Cell types cover 16 human cell lines and primary cell lines but it is being widened [genome.ucsc.edu/ENCODE/cellTypes.html]. The data is publicly available from UCSC browser [genome.ucsc.edu], and microarray studies are available e.g. in GEO databases [ncbi.nlm.nih.gov/geo/]. The first publication of the project was in 2007[249].

Roadmap epigenomics project

The NIH Roadmap Epigenomics Mapping Consortium, with the objective to develop reference epigenome maps and bringing new technologies to accomplish the objective, has released in 2015 with 111 human epigenomes. The resource maps similar elements than encode, but as a novelty, from multiple primary tissues and cell types, and in various developmental stages: adult, fetal, and stem cells. It includes also the ENCODE data. The data and protocols are publicly available [roadmapepigenomics.org], [compbio.mit.edu.sare.upf.edu/roadmap]. The first picture of the data was published recently by the consortium[164].

These data opened up the gateway to a new framework to better understand gene regulation, differentiation, reprogramming and human disease.

2.4. Epigenetics and disease

Anomalies in the epigenetic regulatory machinery have been associated with several diseases such as cancer, diabetes, hypertension, asthma and autoimmune disorders. Also epigenetics has been associated to addictive disorders such as cocaine and alcohol addiction. Finally, neurodevelopmental disorders including Fragile X, Rett and Prader-Willi/Angelman syndrome,

Alzheimer's and Hunghtington's diseases, intellectual disability, schizophrenia, attention deficit hyperactivity disorder, predisposition to stress, depression and ASD among others[157,250–253]. Epigenetic modifications, which as we have already mentioned can be dynamic or fixed, intergenerational or *de novo*, are crucial for understanding the molecular basis of complex phenotypes and diseases, and can explain difference in concordance of the diseases in individuals.

2.4.1. Epigenetics and ASD

The implication of epigenetic defects in the ASD pathology has been long hypothesized but currently is supported by evidences[254–259]. Twin studies showing discordant ASD phenotype and severity display arguments for epigenetic-environment interactions or early somatic mutations leading to ASD. Moreover, environmental factors in different developmental stages (pre-, peri- and postnatal) are proved to be have an influence to ASD aetiology. Therefore, environmental influences through gene-environmental interactions or epigenetic-environment can lead to disease. First evidences in ASD encompassed combined influence of genetic, epigenetic, and environmental factors that affected the methionine and glutathione metabolism. In the performed study, ASD children and their parents showed global DNA 5mC hypomethylation associated to deficits in methylation glutathione dependent antioxidant-detoxification capacity[260,261]. They concluded that the aberrant methylation levels in their mothers could predispose to ASD altering fetal methylation and gene expression[260]. Further methylation studies also found global DNA 5mC hypomethylation in ASD in blood[146], although similar studies in blood or in cerebellum could not replicate the result[262,263], remaining controversial. Besides, global methylation was seen to impair neuronal function and survival in mice[264]. Further genome-wide measurements with larger control cohorts are needed consolidate the evidences.

Imprinted regions associated with ASD

Other evidences for an epigenetic role in ASD are that numerous regions (1p,2q,3p,7q,15q and 17q) associated to ASD are located near regions that are subject to imprinting[265]. Loci on

chromosomes 15q11-13, 7q21-31.31 and 7q32.3-36.3 are the ones with most evidences for a combination of genetic and epigenetic factors. The maternally-derived duplication of the imprinted region 15q11-13 is recurrently associated with ASD, whereas paternal/maternal deletion of the region cause the neurodevelopmental disorders Prader Willi/Angelman Syndromes (PWS/AS) respectively. The regulation of the region is complex (Figure 14). The imprinting center is located in the *SNRPN* promoter, found hypermethylated in the maternal allele and which encodes SNURF and SmN polypeptides, as well as multiple non-coding RNAs. two genes which are preferentially expressed from the maternal chromosome in brain, which would be altered in ASD are: *UBE3A*, which encodes for a E6AP ubiquitin protein ligase, and *ATP10A*, which encodes for ATPase transporter[257]. Besides, gene dosage in gene UBE3A is increased along with its activity (leads to increased ubiquitinization)[266,267]. Specifically, in an ASD patient with the duplication, methylation alteration in 5'-CGI of *UBE3A* has been reported, along with a transcript down-regulation[268]. Other genes in the region encoding GABA receptor subunits (*GABRB3, GABRA5* and *GABRG3*) which are biallelically expressed in controls show monoallelic or highly skewed expression, suggesting epigenetic alterations in these genes in ASD and PWS/AS[269].

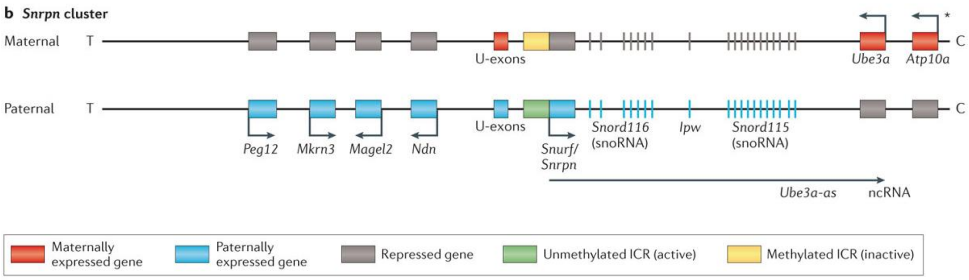


Figure 14. Chromosomal representation and regulation of the imprinted region on chr15q11-13 in the Prader-Willi syndrome (PWS)–Angelman syndrome (AS) locus, found duplicated in ASD. The expression and methylation status for the paternal and maternal chromosomes is indicated in different colours. The imprinting control region (ICR) is located in *SNRPN* promoter and controls the methylation of the genes in the region. In a maternal duplication, UBE3A is overexpressed. UBE3A expression has been shown to be regulated by *MeCP2* and *SNRPN*.

On chromosome 7q there are minimum two loci conferring susceptibility to ASD, however, epigenetic deregulation have not been found. The region 7q21-31.1 includes *SGCE*, *PEG10,* and an ASD candidate gene DLX5, which regulates expression a GABA neurotransmitter producing enzyme and plays a role in GABAergic neurons differentiation. Another region 7q32.2, contains an imprinted region with *MEST, COPG2, CPA4* genes, among others. It remains possible that epimutations are happening in a tissue-specific manner and further methylation studies should be done in these regions[257].

Epigenetic mechanisms involved in ASD comorbid diseases

Another evidence of the implication of epigenetics in ASD is the presence of alterations in genes found in single-gene disorders commonly associated with ASD, such as Fragile X, Rett and CHARGE syndromes, for the latter two, the genes are involved in epigenetic mechanisms. Fragile X, characterized by moderate to severe mental retardation, is caused by a CCG repeat expansion leading to hypermethylation and silencing of *FMR1* transcription[270]. Rett syndrome, a neurodevelopmental disorder characterized by developmental delay and regression, loss of speech, stereotypic behaviour and mental retardation, occurring almost exclusively in females, is caused by mutations in *MECP2*. Also, methylation studies proved aberrant methylation in several CpGs in the gene and certain mutations inhibited MECP2 of binding 5mC[271,272]. Other genetic disorders associated with ASD syndrome also involve epigenetic mechanisms, as CHARGE syndrome (involving mutations in chromatin remodelling factor *CDH7*), and Williams–Beuren syndrome among others conditions[128].

Enzymes regulating epigenetic mechanisms altered in ASD

Several enzymes involved in methylation and regulation of chromatin structure and function have been involved in ASD. A recent study found overexpression of DNMT3A, DNMT3B, TET1 and TET3 in ASD cerebellum compared to controls[273]. Sequencing specific regions in 226 ASD individuals, a study found 46 mutations in members of methyl-CpG-binding domain (MBD) family (MBD1, MBD2, MBD3, MBD4)[274]. Results included mutations in *MBD3* and *MBD4* which were prone to cause the disease. A similar study in 76 individuals involved a familiar mutation in *MBD1* and high-polymorphic variants in *MBD2*[275]. Besides, the histone desacetylase HDAC4, has been identified as the candidate gene to cause Brachydactyly mental retardation

syndrome (BDMR) characterized by craniofacial and skeletal abnormalities, intellectual disability, developmental delay, behavioural abnormalities, and ASD[276]. Another syndrome, the 9q34.3 subtelomeric deletion syndrome, which is an intellectual disability syndrome caused by haploinsufficiency of histone methyltransferase 1 (EHMT1) gene (the specific methyltransferase for H3K9), presents also autistic-like features[277]. Finally, a missense mutation in H3k4 demethylase *JARID1C* gene (also called *KDM5C* or *SMCX*) has been found in a patient with developmental delay and ASD[278]. Also, *ANKRD11*, which encodes a chromatin regulator and controls histone acetylation and gene expression in neurogenesis is involved in ASD[279,280]. In fact, an increasing number of genes involved in chromatin remodelling and transcription regulation are arising by WES[55,86,103], such as *CDH8,* which has been associated to ASD and their impairment by deleterious mutations[281].

Epimutations in ASD associated genes

Epimutations and epigenetic changes secondary to point mutations have been also implicated in ASD. Oxytosin receptor (OXTR) mutations confer risk to ASD and have been reported in numerous family-based and population-based association studies[282]. Oxytosin is a neuropeptide hormone with effects on social and repetitive behaviours and cognition. Hypermethylation of specific sites in OXTR gene in peripheral blood cells and temporal cortex, along with a reduced expression in temporal cortex has been reported in ASD patients[283]. In mice models, OXTR is regulated by methylation in the promoter[284].

Genes involved in synaptic processes as the synaptic scaffolding protein SHANK3, associated with Phelan-McDermid syndrome and ASD, has been proved to be regulated (expression) by DNA methylation of intragenic promoters[285]. Another ASD candidate gene, the EN-2 homeobox transcription factor involved in cerebellar development and serotonin/norepinephrine neurotransmitter systems, showed in cerebellum DNA hypermethylation in the promoter, reduced histone marks (H3K27me3 and H3K27me3) consistent with the increased expression and protein levels found[263]. Besides, decreased *MECP2* binding to *EN-2* promoter was observed, contributing to the aberrant overexpression[273].

Epimutations in novel genes involved in ASD

With the appearance in 2010 of methylomic arrays[286], becoming of wide use, novel methylation aberrations have been discovered. We summarize the relevant results obtained from the methylation studies in ASD up to date (Table 10). There are two studies encompassing discordant MZ twins studying methylation in lymphocyte cell lines (LCLs)[287] and whole blood[262]. In the study using LCLs, several alterations were found but focused on hypermethylation in the promoter of *RORA* a gene that codifies a hormone involved in neuronal development and differentiation, among other functions, and *BCL-2,* an anti-apoptotic protein. The alterations correlated with lower expression and lower protein levels in post-mortem cerebellum and prefrontal cortex samples of another set of patients and controls[287]. The twin study analyzing whole blood sample, found methylation aberrations in the multiple comparisons assessed (twins, family-specific, ASD-controls, sporadic-multiplex). They also found significant epigenetic heterogeneity studying differences in twins discordant for the core diagnostic traits (DSM-IV), and more importantly, found a relationship between the quantitative severity of ASD (CAST test) and methylation in some CpGs[262].

Two other methylation studies performed methylation assays in ASD post-mortem brain samples and in multiple regions (Table 10). A study performed in ASD prefrontal cortex in the area BA10 and anterior cingulate gyrus BA24 area found thousands of methylation differences compared to controls[288]. They specifically found hypomethylation and gene expression deregulations in 12 genes in BA10 area involved in immune responses, neurogenesis and synapses regulation, and the authors implicated specifically two cell types: microglia and neurons. The second study in brain samples analyzed different brain regions (dorsolateral prefrontal cortex, temporal cortex and cerebellum) searching for altered regions, instead of CpGs, and found four genes altered. Intriguingly, both brain studies summarized so far found the same gene altered (hypomethylated): *C11orf21/TSPAN32,* which is located in an imprinted region. *C11orf21* function is unknown, but TSPAN32 protein family (tetraspanins) play a role in immunity[289]. A third study in brain (not shown in the Table 9), did not found any differences between cases and controls, but revised the methylation array values published until then (*UBE3A, MECP2, RORA, OXTR* and *BCL2*) in cerebellar hemisphere cortex and BA19 cortex of ASD and controls[290]. However, they did not found any second case, and despite the small

number of samples studied, epigenetic deregulations might be present with low frequency among ASD patients.

Two curious studies in methylation found intriguing results, the first one found a methylation alteration as a possible ASD biomarker in blood, and another assessed methylation in ASD children with mothers of advanced age at birth in bucal epithelium (Table 10). In the first study, among the differentially methylated regions they found ENO2 gene[291], which encodes the neuron-specific enzyme related to ASD due to expression deregulation[292]. An hypermethylation of *ENO2* along with downregulation and lower protein level was found. As a biomarker, 19/131(14%) patients presented methylation alteration in blood. The study of ASD and typically developing children of advanced age mothers (AMA), a feature which was previously related to ASD[293], was not found to have an impact to the epigenetic deregulation[294]. However, they found thousands of CpGs differentially methylated, enriched for the post-synaptic transmission functions which interacted with ASD-associated genes. Besides, methylation alteration of the promoter in the olfactory receptor *OR2L13* was found. The gene was found previously altered in blood methylation[262] and in brain expression[292].

Summarizing the relevant over-represented pathways and functions found in the mentioned methylation studies, mainly neuronal related functions, such as central nervous system and neuron development, differentiation, and synaptic transmission processes are recurrently found enriched[287,288,294]. Additionally, enrichment in immune response genes, which in turn participated in neurogenesi and synapses [microgliogenesis (*SPI1* and *IRF8*), neuronal synaptogenesi in brain development (*C1q,C3, ITGB, CX3CR1*) and regulation in synaptic plasticity and signal transmission in mature brain (*TNF-α*)], were also found in one study. The authors speculated that, in ASD, changes in the microglia contribute to the immune response deregulation and changes in the neurons contribute to the synaptic genes deregulation seen in brain[288]. Another relevant function found is embryo development, which was also present in two of the studies[287,295]. Curiously, *ZFP57,* found altered in temporal cortex, is involved in imprinting maintenance in development, a process which has previously been implicated in ASD[295]. Finally, many genes or direct targets of the genes found altered have been previously described as ASD candidates[262,296]. It should be taken into account that these studies are

performed with a relatively small number of patients, and also in different tissues, although functions may converge to cause ASD phenotype.

Publication	Coverage	Tissue	Samples	DM parameters	Nº DMR	Findings
Nguyen et al., 2010	8,1k CpG Islands	LCLs	3 MZ twin pair (+ 2 TD siblings); 1MZ TD; 2 sibling pairs discordant for ASD; M; AGRE repository 2-19 years	NA	73CpGs discordant MZ twins; 201CpGs co-twins vs. Siblings	DMCpGs pathways: neurological disease, nervous system dev., cell organization, embryonic deve. , gene expression, cell death. 50 genes correlated with expression. Genes selected (*RORA, BCL2)* with LCLs hypermethylation plus transcription silencing and lower CBL/PFC protein level in post-mortem brain.
Wong et al., 2013	27k CpGs	whole blood	50 MZ twin pairs discordant for ASD; and CNT (individuals scoring low ASD traits); n.d. ethnicity	Difference ≥4% for MZ twins DM, diff ≥15% for familial and ASD vs. CNT	NA	Comparisons: all MZ twins; familial; ASD vs. CNT; sporadic-multiplex: *NFYC, PTPRCAP, MBD4, TINF2, RNF185;* 21 genes in discordance in core behaviours (DSM-IV), and between severities groups; 17 genes genes reported in ASD (e.g. *NRXN1, NLGN3, MBD4, AUTS2, UBE3A, GABRB3*)
Nardone S et al., 2014	485k CpGs	brain (PFC BA10, ACC BA24)	23 ASD; 23 CNT; M/F, 16-51 years; n.d. ethnicity	Difference > 5% q.val<0.05	Found 5329CpG (BA10), 10745 CpG (BA24)	PFC BA10: Hypomethylation correlated with increase in expression (n=12 genes), enriched for immune response and neuron differentiation and synapses: *C11orf21/TSPAN32, SPI1, IRF8, C1qA, C3, ITGB2, TNF-α,CTSZ, PTPN6, HLA-DMB,HDAC4, CX3CR1*
Ladd-Acosta C. et al., 2013	485k CpGs	brain (TC, DLPFC, CBL)	19 ASD; 21 CNT; M/F, 14-39 years;	Difference >6.6% q.val≤0.1	3DMRs (TC), 1DMR (CBL)	TC: *PRRT1, C11orf21/TSPAN32, ZFP57* (for *ZFP57*, sex differences); DLPFC: *SDHAP3* (possibly a CNV)
Wang et al., 2014	33,3k regions	whole blood	5 ASD; 5 CNT; M/F, 3-12 years; chinese ethnicity	1,5-fold change q.value<0.05	475 regions	13 genes altered in all ASD vs. CNT. Validated ENO2 Hyper-methylation correlated with downregulation and decreased protein levels. Replication in a larger cohort: a total of 19/131 (14%) with hypermeth.
Berko et al., 2014	485k CpGs	buccal epithelium	47 ASD; 48 CNT of mothers >35years at birth; M/F; 1-18 years; ethnicities	p<0.001 for the CpGs; Diff ≥7.5% for regions	3560 CpGs, 13 selected regions	Genes with post-natal synaptic transmission functions which interacted with ASD-associated genes. *OR2L13* region: promoter hypometh. Advanced maternal age may not involved in epigenetic deregulation.

Table 10. Methylation arrays studies in ASD. ACC, anterior cingulate gyrus; CBL, cerebellum; CNT, controls; F, female; TC,Temporal cortex; TD, typically developing; DLPFC, dorsolateral prefrontal cortex; M, male; LCLs, lymphoblast cell lines; PFC, Prefrontal cortex.

Histone marks altered in ASD

There are also evidences of alterations in histone tails' modifications in ASD. Excessive spreading of H3H4me3 mark into nucleosomes was observed in neurons in prefrontal cortex of 4 out of 16 ASD individuals compared to controls[297]. However, the studied patients did not show a global alteration of the histone mark compared to controls. Specifically, the mark was increased further away from the TSS (analyzing -2Kb to 3Kb) in 503 loci and decreased in 208 loci. Decreased H3H4me3 over the TSS associated with decreased transcript levels was observed in five genes involved in synapses processes (*ARC* and *USPx*), neuronal plasticity (*VGF*), involved in synaptic vesicle functions and being essential for social behaviour (*CPLX1*). They also found genes previously implicated in ASD with altered H3H4me3 marks in the promoter or gene bodies (e.g. *SEMA5A, CACNA1C, CACNA1H, AUTS2, PARK2* and *JMJD1C*). The authors explain the variation in the histone marks by 1) variation in the genetic sequence, although, they discarded CNVs or structural variants of having an effect of the observed changes in the studied cohort; 2) heritable transmission of histone modifications or 3) pathological consequences leading to cortical dysfunction.

LncRNA altered in ASD

Regulatory lncRNAs have also been involved in ASD. A study with microarrays covering 33,045 lncRNAs in ASD postmortem prefrontal cortex and cerebellum tissue compared to controls identified 200 differentially expressed lncRNAs located in or around protein-coding loci[298]. The genes were enriched for brain development genes and the GO term over-represented was cerebral cortex cell migration. Eleven of the genes near the lncRNAs have previously been described in ASD, and 18 exhibited differential expression in ASD brain previous studies.

Other studies in candidate genes involved lncRNA in their regulation. The gene *ST7,* which maps in an ASD susceptibility locus in chromosome 7, in which several rare variants were detected in ASD patients, contains at least four non-coding genes (*ST7OT1-4*) that potentially regulate *ST7*. Although it is unknown their role in ASD, its regulation involves several lncRNA[299]. Additionally, mutations in *PTCHD1* gene involved in X-linked Intellectual disability and ASD, with overlapping several anti-sense lncRNAs (*PTCHD1AS1-3*), has been involved in the disease, although their role remains unclear[300]. An example of *trans* lncRNA happens in moesin gene, MSN, located in chromosome X with a role in axons and dendrite development. An antisense

gene located in chromosome 5, MSNP1AS, when transcribed interacts with moesin transcripts and inhibit transcript expression. The authors found increased level of the antisense transcript in post-mortem ASD brain samples, meaning that aberrant levels of the protein in certain developmental stages may be involved in increasing ASD risk[301].

2.5. DNA methylation techniques for high-throughput assessment

The projects compiled in this thesis study specifically the methylation layer, its characterization and validation. For this reason, we review in this section the wide range of techniques for assessing methylation, giving further details for the ones specifically used in the projects (Table 11). Different tools are available to assess global methylation, genome-wide methylation (NGS or array based) with high resolution, and to target locus-specific or specific genes. Here we revise some basic techniques to assess 5mC, for assessing methylation of 5mC oxidative derivatives refer to a recent review[302].

Protocol Steps	Global methylation	Genome-wide NGS-based	Genome-wide Array-based	Locus-specific analysis	Gel-based analysis
sodium bisulfite conversion	Pyrosequencing (LINE-1, Alu, LUMA)	RRBS, WGBS, T-WGBS, WGSBS, PBAT, BC–seq, BSPP, Microdroplet PCR-BS, mTACL, LHC-BS (pre/post conv.), mRBBS,LCM-RBBS or sc RRBS	Infinium, GoldenGate, BiMP	EpiTYPER, Pyrosequencing, MethyLight, BS Patch PCR	Sanger BS, MSP MS-SNuPE, COBRA
enzyme digestion/ activity	SsI assay	Methyl–seq, MCA–seq, HELP–seq, MSCC, RSMA, MRE-seq	DMH, MCAM, HELP, MethylScope, CHARM, MMASS	HpaII-PCR	Southern blot, RLGS, MS-AP-PCR, AIMS
Affinity enrichment		Methylcap-seq, MeDIP–seq, MIRA–seq	MeDIP, mDIP mCIP, MIRA, MeKL-ChiP	MeDIP-PCR	

Protocol Steps	Global methylation	Genome-wide NGS-based	Genome-wide Array-based	Locus-specific analysis	Gel-based analysis
Chemical Separation	HPLC-MS, TLC-CE				

Table 11. Techniques available for DNA methylation quantification[302,303]. In bold are found the techniques reviewed below to obtain a general picture of the mostly used, and which some of them have been used in this thesis. Alu, alu element pyrosequencing; AIMS, amplification of inter-methylated sites; BC–seq, bisulphite conversion followed by capture and sequencing; BiMP, bisulphite methylation profiling; BSPP, bisulphite padlock probes; CHARM, comprehensive high-throughput arrays for relative methylation; COBRA, combined bisulphite restriction analysis; DMH, differential methylation hybridization; HELP, HpaII tiny fragment enrichment by ligation-mediated PCR; HPLC-MS high performance liquid chromatography -mass spectrum; LCM-RRBS, laser-capture microdissection-reduced representation bisulphite sequencing; LHC-BS (pre- and post-conversion) , liquid hybridization capture based bisulphite sequencing; LINE-1, long interspersed elements type-1 pyrosequencing; LUMA, luminometric methylation assay; MCA, methylated CpG island amplification; MCAM, MCA with microarray hybridization; MeDIP, mDIP and mCIP, methylated DNA immunoprecipitation; MeKL-ChIP, methylated-DNA kinase pre-treated ligation-mediated PCR amplification chromatin-immunoprecipitation; MethylCap, methylation DNA capture sequencing; MIRA, methylated CpG island recovery assay; MMASS, microarray-based methylation assessment of single samples; MRE-seq, methylation restriction enzyme sequencing; mRRBS, multiplexed reduced representation bisulphite sequencing; MS-AP-PCR, methylation-sensitive arbitrarily primed PCR; MSCC, methylation-sensitive cut counting; MSP, methylation-specific PCR; MS-SNuPE, methylation-sensitive single nucleotide primer extension; mTACL, methylation target capture and ligation; PBAT, post-bisulphite adaptor tagging; RLGS, restriction landmark genome scanning; RRBS, reduced representation bisulphite sequencing; RSMA, methylation-sensitive restriction enzyme-based assay; scRRBS, single-cell reduced representation bisulphite sequencing; SsI assay, CpG Methyltransferase transfering tritium-labeled methyl groups; TLC-CE, thin layer chromatography - capillary electrophoresis; T-WGBS, transposase-based library construction; WGBS, whole-genome bisulphite sequencing; WGSBS, whole-genome shotgun bisulphite sequencing.

Regarding the experiment design to perform methylomic assays or epigenome wide association analysis (EWAS), several confounding effects should be taken into account[304,305]. First the homogeneity within the studied groups, regarding age, ethnicity, sex, disease, treatments followed; Second, the homogeneity of the tissue and cell-types from which the DNA is extracted; Third, the replication of samples inter-array/inter-assay and the distribution in batches/in assays.

The steps used in the methylation techniques protocols encompass sodium bisulfite conversion, enzyme digestion, affinity or chemical separation, PCR, sequencing and have technical biases which should be also considered (Figure 15). For some of the techniques, the

first step encompassing sodium bisulfite conversion, a widely used treatment to differentiate methylated from unmethylated DNA, based on the conversion of non-methylated cytosines to uracil (converted to thymidine in PCR) has many sources of bias. Biases include incapacity to distinguish between 5mC and 5hmC and the incomplete conversion of the DNA which could lead to misleading quantification. Moreover, the resulting BS-DNA is a labile molecule with a highly repetitive sequence in which PCR can be performed using special taq polymerases and with increased number of cycles which result in PCR bias[306,307]. For other approaches, as chemical or enzymatic digestion, incomplete treatment also can lead to erroneous results or in sequencing, a minor coverage for CG-rich regions.

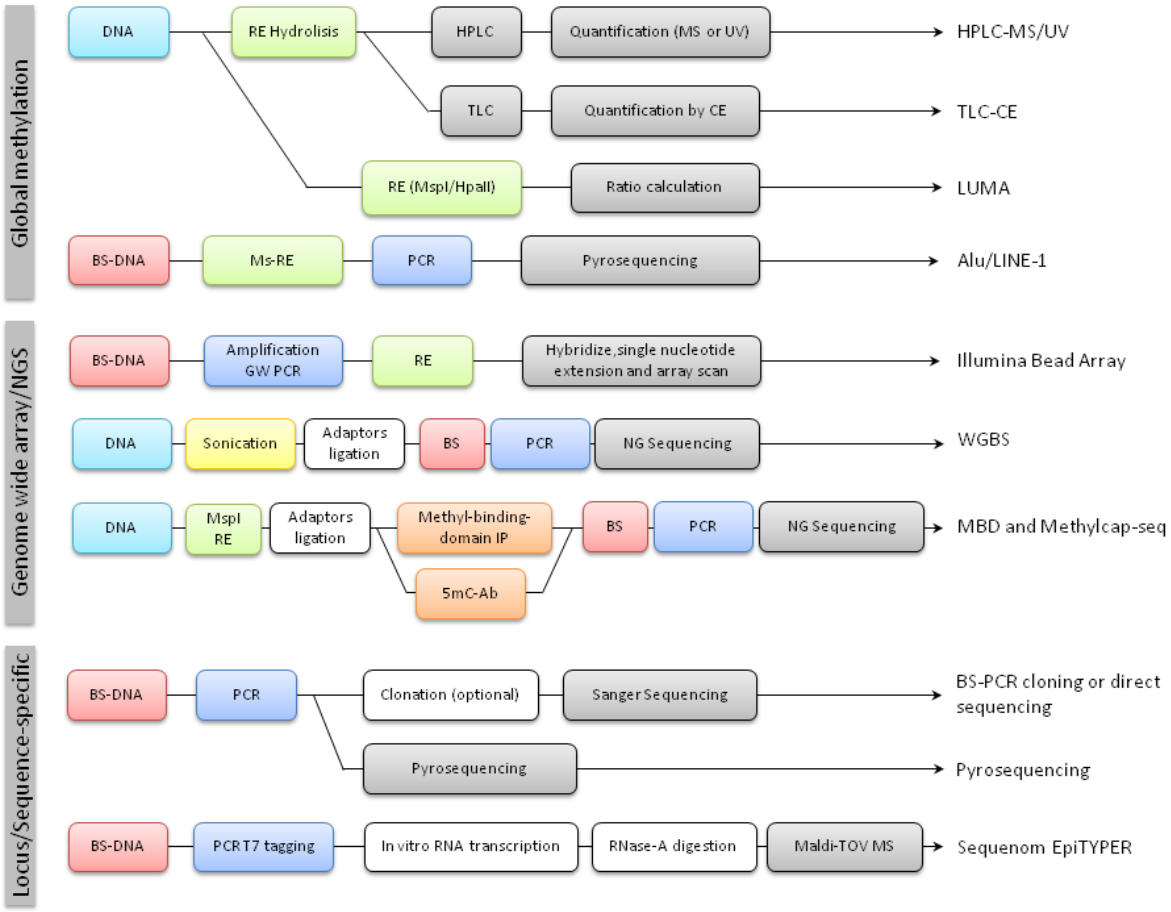


Figure 15. Protocol main steps of the global methylation, genome wide and locus-specific techniques reviewed in this section. Per each goal three to four techniques are detailed. The starting sample is genomic DNA or bisulfite (BS-DNA) and several steps of enzymatic restriction or digestion (green), PCRs (blue), bisulfite treatment (red), other steps (white/yellow) ended by sequencing, array scan or quantification by different methods (grey) are detailed in each technique. Choosing between the techniques depends within each goal depend on plenty of factors: the time-cost relation, the number of samples and quantity available, and the service availability.

Global methylation assessment

To perform global methylation measurements there are two traditional techniques. The high performance liquid chromatography HPLC or variants (coupled with mass spectrum), in which 5C and 5mC are separated by size and quantified, and thin layer chromatography (TLC) coupled with high performance capillary electrophoresis (CE), which also separate and quantify the nucleotides. However, these techniques require large amounts of DNA and the optimization is demanding[308]. Alternative techniques with improved methods are PCR based methods which estimate global methylation of repeat elements, Alu and LINE1, representing 11% and 17% of the human genome respectively. The PCR is performed over bisulfite converted DNA. The PCR is quantified after by pyrosequencing (reviewed below). Finally, the technique called luminometric methylation assay (LUMA) uses methylation sensitive enzymes (HpaII) and methylation unsensitive enzymes (MspI) which recognize a sequence present in 8% of the human genome, and the ratio is calculated. The isoschizomers MspI (methylation insensitive) and HpaII (methylation sensitive) cut differently and the resulting products are quantified by pyrosequencing[308].

Genome-wide techniques

To perform genome wide studies with locus resolution, there are array based methods which cover up to 2% of the genome, or genome wide techniques which covers >90% of the genome.

- Array-based techniques: Regarding currently used arrays, Infinium arrays provided by Illumina need a first step of sodium bisulfite followed by hybridization onto bead arrays. For the HumanInfinium27k array, covering 27000 CpGs, different beads (type I) hybridize to non-methylated CpGs and methylated CpGs; for the HumanInfinium450k array, covering 450 000CpGs, beads type I are used and also an additional bead type (type II) hybridizing to both methylated/unmethylated state use a different chemistry[309,310](Table

12). Although these arrays are widely used for their cost-efficiency, they need extensive normalization and polymorphisms in the CpGs or probes and cross-reactive probes should be removed. Several packages available in bioconductor perform different normalization types and follow different pipelines to obtain differentially methylated CpGs or regions[311,312].

Structural elements/regions	Nº CpGs
Promoter, 5'UTR, 1st exon	200,339
Body	150,212
3'UTR	12,383
intergenic	119,830
Coding	361,766 (98% Refseq genes)
Non-coding	4,168

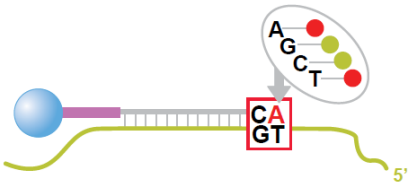


Table 12. Regions covered by the Illumina HumanInfinium450k Bead Chip array on the genome. The diagram shows the bead (probe) which hybridize with the genome sequence. Infinium uses two types of chemistries: Infinium II (used in 450k array): one bead hybridizes with both sequences (methylated and unmethylated) whereas in the Infinium I (used in the 450k and 27k arrays): one bead hybridizes with each methylation state. The nucleotides are labelled and the fluorescence for each incorporated (one base single-extension) dNTP is quantified. Several beads assess the same region in order that methylation can be quantitative.

- Regarding genome wide techniques, they cover from 15% to >90% of the genome[303]. With less resolution there is the reduced representation bisulfite sequencing (RRBs) technique (and variants), which select CpG-containing regions by BglII or MspI digestion. After, the library is prepared and gel-based selected, the fragments are bisulfite treated and the library is amplified and sequenced. With higher resolution, there are other methods as Methylcap-seq (covering about 40% of the regions) and MEDIP-seq (about 85% of the genome), both using non-bisulfite treated DNA. After sonication, Methylcap-seq captures the fragments by methyl-binding-domain protein affinity, followed by library preparation and sequencing. MEDIP-seq is based in library preparation after sonication, enrichment using 5mC antibody, library amplification, and sequencing. Finally, whole-genome bisulfite sequencing (EGBS or MethylC-seq) is used to sequence at single-base-pair

resolution, and uses an approach similar to RRBS, without selecting for regions. Extensive treatment of the reads, alignment and annotation is needed (Figure 16).

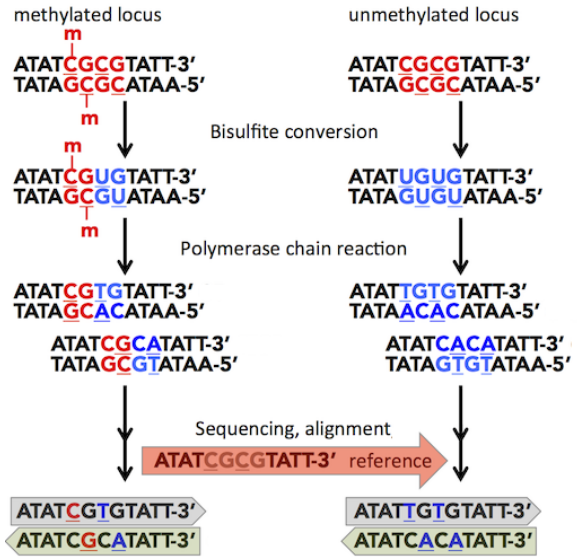


Figure 16. Alignment of the WGBS sequences. The BS-DNA is no longer complementary giving after PCR four different strands per methylation state. The alignment is performed for each state and a reference sequence is created allowing the methylation variations. A minimum of reads per strand are needed to overcome the PCR bias and sequencing errors and identify true individual alterations. Intensive sequencing is needed as the unmapped reads load is high, raising the cost for the technique.

Locus-specific methylation assessment

In order to assess methylation in a limited number of loci, several techniques are available. The gold standard is bisulfite sequencing followed by PCR. Once the BS-PCR is done, multiple approaches can be pursued: a) the fragments can be cloned and Sanger sequenced (from 10-20 clones to establish quantitatively the methylation); b) the PCR can be directly sequenced by Sanger sequencing without cloning, which has been demonstrated that also is quantitative[313,314]; c) pyrosequencing of the BS-PCR with an internal primer nearby the targeted CpG. By the first two approaches up to 400bp can be sequenced, and by pyrosequencing, only up to 50bp[315]. The advantage of the pyrosequencing is that can be coupled to ultradeep sequencing allowing deep coverage with a highly quantitative result and is less time demanding than cloning. To study a higher number of regions, from 10 to 150 regions, Sequenom EpiTYPER can be used, in which a primer flanking the CpG is designed and several primers can be multiplexed in the same analysis[316]. After DNA bisulfite treatment, a

PCR is performed (in which T7 promoter tag is introduced), followed by RNA transcription and RNase-digestion. MALDI-TOF Mass Spectrum analysis of the fragments gives a different pattern of the methylated versus non-methylated DNA. The technique quantifies the CpGs methylation level within 200-600bp.

A true quantification to understand the epigenome organization and function, and their implications in disease will only be possible when techniques become more affordable to study hundreds and thousands of individuals. Possibly, when methods using direct sequencing (not requiring bisulfite conversion, restriction enzymes or affinity enrichment) as single-molecule real-time, will lead to the direct quantification of methylation and hydroxymethylation in a single reaction in a good cost-efficiency relation, it will be a big step forward for the epigenetic field.

3. ASD in vivo and in vitro models

Finally, this section will review the animal models and the therapeutically strategies. To further understand the ASD physiopathology, models carrying mutations in candidate genes contribute to unravel the molecular mechanisms and the neuropathology involved in the disease. Moreover, the models recreate the physical and the behavioural consequences of mutations. The models also provide a framework to study the effect of different environmental factors contributing to ASD and to assay therapies to ameliorate the phenotype. A large number of animal models carrying mutations in ASD candidate genes have been reported. Currently, in the Mouse Genome Database (MGD-MGI), a total of 32 genes in ASD mice models are recopilated[317,318]. In the animal model module of SFARI [gene.sfari.org/autdb/AMHome.do], 179 genes involved in ASD having one or multiple animal models (up to 30 animal models for FMR1 (Fragile X)are compiled[319].

In vivo models: mice models

Animal models include chromosome-engineered mouse models (15q11-13 and 16p11.2), mutated genes involved in co-morbid conditions (e.g. *FMR1, TSC1/TSC2, NF1, PTEN*), in synaptic processes (*NLGN3/4, NRXN1, SHANK2/3*, *GABRB3)*, and other genes as *EN2, RELN* or

OXTR[48,320–322] (Table 13). For the models the phenotypic affectation, as well as the neuropathology involved, are analyzed. Regarding the phenotypic behaviour, most of the models show abnormalities in ASD core symptoms: communication and social interaction, and repetitive behaviours. The communication impairment in mice has been tested recording ultrasonic vocalizations; the social interaction with socialization tests encompassing other mice; and the repetitive behaviours with self grooming, jumping and marble burying behavioural patterns[323]. Some of this animal models show also other phenotypic traits associated e.g. intellectual disability, hyperactivity, sensory perception which can also be tested[323]. Several reviews characterize the mice models regarding the core symptoms[320,324].

	Mouse model	Social and communication	Repetitive behaviours	Other phenotypes	Neurobiological mechanism	References
Chr models	15q11-13 duplication	Impaired	Inflexibility	NA	Altered serotonergic signaling	Nakatani et al., 2009
	16p11.2	NA	Climbing deficits	Altered cyrcadian rhythm	Hypothalamic Deficits	Horev et al.,2011
	22q11.2 microdel.	NA	NA	Hyperactivity	Altered microRNA biogenesis	Stark et al.,2008
Syndromic ASD models	FMR1 KO	Impaired	ASD flapping	Learning def., Anxiety	mGluR hyperfunction	Bernardet et al., 2006
	TSC1 HT, TSC1CbKO	Impaired	Grooming, inflexibility	Ataxia	Cerebellar deficits	Auerbach et al., 2011; Tsai et al.,2012
	TSC2 HT	Impaired	Increased marble burying	Learning def.	mGluR hypofunction	Ehninger et al., 2008
	NF1 HT	Impaired	NA	Learning def.	Ras signaling hyperactivation	Costa et al.,2001, 2002
	PTEN KO	Impaired	NA	Learning def.	PI3K pathway hyperactivation, Macrocephaly	Kwon et al.,2006
	MeCP2 KO	Impaired (enhanced interaction)	Hindlimb clasping	Respiratory problem, Lethality	Decreased GABAergic transmission	Shahbazian et al., 2002; Moretti et al., 2005
Synaptopathy ASD models	NLGN1 KO	Minimal impairment	Grooming	Learning def.	NMDAR hypofunction	Blundell et al., 2010
	NLGN2 Tg	Impaired	Jumping	Seizure (EEG)	Increased GABAergic transmission	Hines et al.,2008

Mouse model	Social and communication	Repetitive behaviours	Other phenotypes	Neurobiological mechanism	References
NLGN3 R451C KI	Impaired	NA	Enhanced learning	Increased GABAergic transmission	Tabuchi et al., 2007
NLGN3 KO	Impaired	Normal	Hyperactivity	Decreased brain volume, Cerebellar deficit	Radyushkin et al., 2009; Baudouin et al., 2012;
NLGN4 KO	Impaired, Less aggression	Normal	NA	Decreased brain volume	Jamain et al., 2008
NRXN1 KO	Differences in novelty responsiveness	Increased		Impaired spatial memory	Blundell et al., 2010; Chubykin, 2007; Laarakker et al. 2012
Shank1 KO	Normal	Normal	Anxiety, Motor def.	Impaired glutamatergic transmission	Hung et al.,2008; Silverman et al., 2011
Shank2exon7KO	Impaired	Grooming	Hyperactivity, Anxiety	NMDAR hyperfunction	Schmeisser et al., 2012
Shank2exon6-7 KO	Impaired	Jumping	Hyperactivity, Anxiety	NMDAR hypofunction	Won et al.,2012
Shank3 HT	Impaired	NA	NA	Impaired glutamatergic transmission	Bozdagi et al., 2010
Shank3B KO	Impaired	Grooming	Anxiety	Striatal dysfunction	Peca et al.,2011
Shank3exon4-9 KO	Impaired	Grooming	Learning def.	Impaired glutamatergic transmission	Wang et al.,2011
Cadps2 KO	Impaired, Maternal neglect	NA	Hyperactivity, Anxiety	Decreased density of PV interneuron, Reduced BDNF release	Sadakata et al., 2007
Syngap1 HT	Impaired social recognition, isolation	Repeated stereotypy	Learning def., Hyperactivity, motor def.	Premature spine maturation and hyperexcitability	Guo et al.,2009; Clement et al., 2012
Scn1a KO	Impaired	Grooming	Seizure	Decreased GABAergic transmission	Han et al.,2012
CNTNAP2 KO	Impaired	Grooming	Seizure	Reduced number of interneurons, Abnormal neuronal migration	Penagarikano et al., 2011
Gabrb3−/−	Impaired,	Poor motor	Attention def.,	seizure	Culiat et al.,

	Mouse model	Social and communication	Repetitive behaviours	Other phenotypes	Neurobiological mechanism	References
		reduced social novelty	skills, tremors	learning and memory def., hyperactivity	susceptibility, cerebellar hypoplasia	1994; De Lorey, et al., 2008; Homanics et al., 1997
Other genes	Dvl1 KO	Impaired	NA	Sensory gating def.	Impaired Wnt signaling	Lijam et al., 1997; Long et al.,2004
	Oxtr KO	Impaired, Less aggression	Not impaired	NA	Impaired oxytocin signaling	Takayanagi et al., 2005; Crawley et al., 2007
	EN-2 KO	Impaired	Not impaired	Learning def.	Cerebellar deficits	Brielmaier et al., 2012
	Reeler mice (RELN)	Impaired, Increased social dominance	NA	Learning def., Ataxia	Brain lissencephaly	Salinger et al., 2003; Lalonde et al., 2004
	4E-BP2 KO	Impaired	Grooming, Increased marble burying	NA	Increased NLGNs translation, Increased excitation	Gkogkas et al., 2013
	eIF4E O/E	Impaired	Grooming, Increased marble burying	Impaired reversal learning, Impaired fear extinction	E/I imbalance in PFC, Increased LTP in striatum and hippocampus	Santini et al., 2013

Table 13. Main animal models (mice models) showing ASD phenotype (core symptoms impairments) and neuropathological mechanisms altered.

Important findings using animal models include common core pathways altered within different mouse models. For example, animal models for Fragile X syndrome ($Fmr1^{(-/y)}$), and $Nlgn3^{ko}$ mice overexpress the glutamate receptor mGluR5, whereas Tuberous sclerosis ($Tsc2^{+/-}$) downregulate it in hippocampus[325]. Another example is the impairment of the equilibrium of excitatory/inhibitory synapses balance, involved in ASD aetiology[109], which has been observed in multiple animal models (*NLGN2/3, NRXN1a, SHANK2/3, CTNAP2* and *FMR1*[93,326,327]) encompassing an increase or decrease in a certain type of synapses.

One of the goals of the mouse models is to successfully map the molecular mechanism to the behaviour. A recent successful study linking genetics to behavior is the study of two animal models (Nlgn3 KO and Nlgn3 R451C) which showed a specific striatal synapse impaired which caused a repetitive behaviour[328]. The authors highlight the complexity observing that a single mutation could have different synaptic effects in different brain regions, or even in the same neuron affect different synapses. It should be taken into account that mouse models also show phenotypic variability. For instance, NLGN4 KO showed communication and social interaction deficits in some studies[329] but not in others[330].

Multiple mouse models have been assayed for environmental factors as mercury or other environmental exposures[48]. For example, administration of valporoic acid injected to gestational mothers caused a reduction of cerebellar Purkinje cells and inhibitory circuits impairment[331].

Other animal models

Other animals models than mice are also used to increase the knowledge in molecular bases and behaviour of ASD. Primates models are relevant to mimic the human behaviour impairment, for example, the lack of gaze. Also, to study mechanisms that may have a role in the mimicking of language and behaviours, as the mirror neurons. Another model, Zebrafish, provides insights to the study of the molecular aspects of ASD, and to functional studies involving for example microinjection of synthetic mRNA, as well as to genetic screens of new regulators and candidate genes (e.g. *RELN* and *MET*[332]). Songbirds provide important clues about neuronal circuitry and language acquisition, as they show similar patterns to human speech. For example, *FOXP2* and its transcriptional target, *CNTNAP2*, revealed their spatio-temporal distribution during the bird song[333]. Studies in invertebrates (e.g. in Drosophila and c-elegans) provide molecular insights as well as neuronal synaptic communications. However, animal models do not always display human-specific behaviours for a complex disease such as ASD.

In vitro models: induced pluripotent stem cells (iPSCs)

Regarding in vitro models, the use of induced pluripotent stem cells (iPSCs) created from patients' fibroblasts (skin biopsies) is emerging as a patient-specific cell model for ASD. IPSCs

can be derived to a neural progenitor cell and to different cell types e.g. glutamatergic or GABAergic neurons. IPSCs-derived neurons mimic the neuronal development and patterning of the specific ASD patient[334] so can be used as a disease model (Table 14) and as a drug-screening model. Numerous studies can be done *in vitro* with the derived cell model such as morphological analysis, neurogenesis, migration and synaptogenesis studies, neuronal electrophysiology, protein assays and expression analysis[335]. In the future, therapies using patient-specific iPSCs-derived neurons might be a fact.

Disease	Time of neuron differentiation	Phenotype in iPSC-derived neurons	References
Fragile X syndrome	19 days	Fewer and shorter neurites	Sheridan et al., 2015
	4–6 weeks	Decreased Synaptic proteins expression (PSD95), decreased synaptic density, decreased neurite length, functionally abnormal neurons (calcium imaging), increased amplitude/frequency and altered response to glutamate uptake	Liu et al., 2012
Rett syndrome	4–5 weeks	Reduced spine density and number of synapses, smaller soma size, altered calcium signaling and electrophysiological defects	Marchetto et al., 2010
	30days	Reduced nuclear size	Ananiev et al., 2011
	8–9 weeks	Reduced neuronal soma size	Cheung et al., 2011
	25 days	Reduced expression of mature neuron markers	Kim et al., 2011
Timothy SD	43days	Impaired calcium signaling and electrophysiology, Defect in activity-dependent gene expression; Abnormality in differentiation, increased production of norepinephrine and dopamine.	Pasca et al., 2011
CDKL5-related disease (atypical Rett SD)	55–60days	Aberrant dendritic spines, significantly reduced number of synaptic contacts	Ricciardi et al., 2012
BCKDK assoc. mutations	~6 weeks	No disease-related phenotype reported	Novarino et al., 2012

Table 14. Phenotypes observed altered in iPSC-derived neurons in ASD comorbid diseases, compiled from Prilutsky et al., 2013[335]. BCKDK, Branched Chain Ketoacid Dehydrogenase Kinase; SD, Syndrome.

As a conclusion, further efforts should be done to understand the spatio-temporal alteration, and the additive genetics (oligogenic model), epigenetic and environmental factors

contributing to the phenotype. A further step would be to induce models with several mutations closer to the phenotype reality. Research in neuropathology, animal models and the study of converging pathways will truly help to detect and identify high-value therapeutic targets.

4. Therapy

4.1. Biomedical therapy

Drug therapy (pharmacology)

The objective of pharmacotherapy for ASD patients is to ameliorate the symptoms and behaviours. Targeted symptoms include repetitive behaviours, anxiety, depression, hyperactivity, aggression, self-injurious behaviour and sleep problems among others. However, no drugs target cognitive impairment.

Habitual prescribed drugs for ASD patients include antipsychotic drugs (Risperidone and Aripiprazole) which reduce aggression, self-injurious behaviour and repetitive behaviours more efficiently in children than in adults, although having some side-effects. For anxiety and repetitive behaviours, serotonin reuptake inhibitors are prescribed (e.g. Citalopram, escitalopram, fluoxetine and others) although effectiveness is doubtful considering the adverse effects they might have and that abnormalities in serotonin system are present in ASD patients. Stimulants are prescribed when attention-deficit hyperactivity disorder symptoms, although having no serious side-effects. Antiepileptic drugs (levetiracetam and divalproex sodium) proved successful to reduce repetitive behaviour, impulsivity and mood changes. A complete list of the "possibly efficacious" drugs can be found in this article[336].

Other drugs are on experimental trials, such as acetylcholinesterase inhibitors, agonist at N-methyl-D-aspartate (NMDA) glutamate receptors, and antagonists at metabotropic glutamate receptors (MGluR), as well as oxytosine, although this is difficult to administer[336].

Diet therapy

The use of nutrition to improve health created the concept of nutraceuticals in 1989. Nutraceutical-based management practices have been used in ASD treatment as a complementary therapy. Some of the enriched supplements provided are melatonin as a regulator of circadian rhythms, multivitamins (e.g. vitamin C with a role related to oxidative stress), gluten-casein free diet for the neurobehavioral symptoms and the gastrointestinal problems associated with the disease, aminoacids (carnosine for the neuroprotective effect and carnitine for mitochondrial disease), essential fatty acids (e.g. omega 3 fatty acids), folic acid, magnesium, probiotics among others[337]. Efficacy and safety of nutraceutical agents are being studied although experts point out the need of larger studies[338].

4.2. Behavioural and developmental interventions

Behavioural and developmental interventions are directed towards the core ASD symptoms to ameliorate patient living standard. There is a wide spectrum of behavioural and developmental interventions, although a metanalisis of available methods stated that no definitive intervention improves all symptoms for all individuals with ASD[339]. The US Health Resources and Services Administration, and the UK National Institute for Health and Care Excellence provided clinical guidelines for behavioural therapy[340,341]. They highlight that it should be individualized, start as earlier as possible after diagnosis, and involve the family. They underscore the communication interventions for non-verbal individuals and social-communication skills for the rest of individuals.

Experimental studies revealed that early intervention programmes improve the outcomes in terms of independent living and education for ASD patients. Studies performed with patients which did not receive early intervention showed that around 60% of adults with ASD have poor or very poor outcomes, e.g. phrase speech before age 6 years predict a better outcome[342]. Related to the neural mechanisms, early interventions with multi-sensory experiences have an impact in brain plasticity and development of new neural networks, and addresses social core impairments and mood control[343]. Early detection is essential for early intervention and efforts are in identifying ASD before 18 months instead of 24-30 months. Efforts are being made to identify trustworthy biomarkers in genetics, neuroimaging, gene expression, metabolism and

immunology. In the future, biomarkers and genetic tests may predict high risk for ASD diagnosis, meanwhile many ASD biomarkers have been proposed although none implemented in the clinical routine[344].

OBJECTIVES

OBJECTIVES

The aim of the thesis is to identify and explore the implication of epigenetics in ASD, defining genes, pathways and molecular mechanisms that may be involved in the aetiology of the disease.

To conduct this work we had availability of a cohort of 331 ASD patients clinically and psychologically diagnosed. Before exploring the epigenetic characteristics, a genetic screening was performed by standard karyotype, followed by Multiplex ligation-dependent probe amplification (MLPA) and molecular karyotype to discard recurrent genetic causes.

The specific objectives proposed are:

AIM 1

To depict genome-wide methylation alterations in blood in idiopathic ASD patients compared to controls, and correlate the alterations found with expression data for the same subset of patients. The final goal is to identify deregulated mechanisms in the disease caused by the epigenetic alterations found.

AIM 2

To define the genome-wide epigenetic landscape for ASD patients with different subgroups and search for convergences and divergences of each group between other groups and controls. Also to elaborate a classification method between groups using a combination of a methylation alterations.

AIM3

To study the correlation of methylation alterations at different tissues, peripheral blood and placenta, in ASD children and controls, in order to observe if common alterations are found, as an indicative of abarrations in early developmental stages.

MATERIALS AND METHODS

MATERIALS AND METHODS

The methodology followed to fulfill the objectives is represented in the following figure (Figure 17) and are described below.

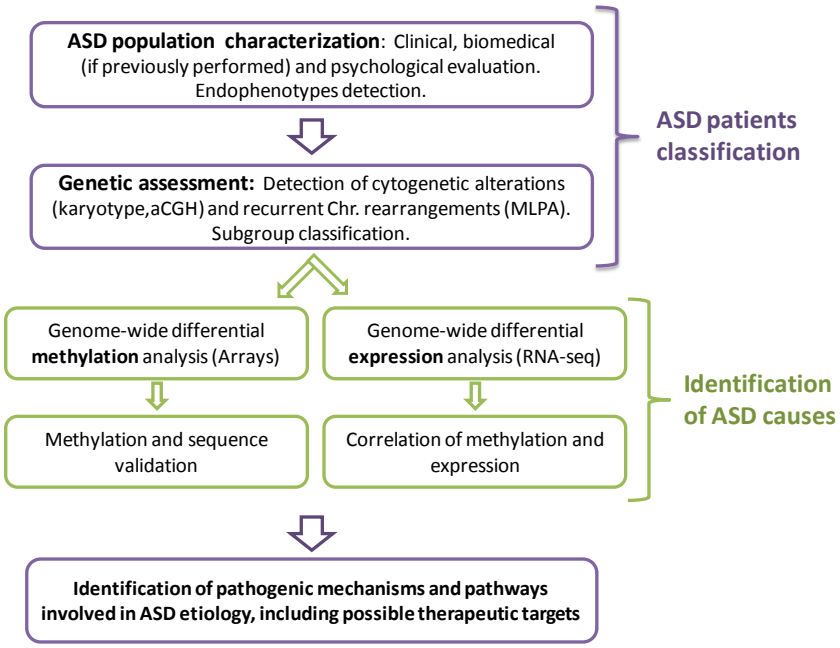


Figure 17. Approach followed to characterize the cohort and to identify the epigenetic or underlying genetic causes and their expression consequences, involved in the patients' aetiology. From the entire characterized cohort we selected an homogeneous cohort to perform methylation and expression analysis. aCGH, array Comparative Genomic Hybridization; MLPA, Multiplex ligation-dependent probe amplification.

Samples and diagnostic reports

Since 2006 the Genetic Unit has gathered from collaborating Spanish hospitals 331 ASD cases with ASD diagnosis. For all of them we obtained the **medical report** (with familial history, prenatal and developmental information), **biomedical tests reports** (mainly genetic tests but also neurobiological exploration) and also information for the evolution of the patients and adaptation to education. All the patients were diagnosed of ASD following DSM-IV-TR chriteria and most of them followed a **psychological assessment** by ADI-R (Autism Diagnostic Interview) and ADOS (Autism Diagnostic Observation Schedule). At the same time, we have 150 trios (patients plus parents) from the international repository Autism Genetic Resource Exchange (AGRE) and 200 samples from European Collection of cell Culture (HPA).

For the samples collected, we extracted blood and prepared DNA, RNA, and if possible, chromosomes and cell pellets. We also obtained samples of parents and relatives, if available.

Genetic assessment

All patients had a screening for cytogenetic abnormalities by standard karyotype, and if it did not reveal any alteration, a MLPA searching for recurrent genomic rearrangements was performed. In the MLPA results, if no chromosomal alterations were found, a molecular karyotype by array aCGH was performed.

MLPA consist of sets of two long primers which recognize a unique region, and which can be multiplexed in a panel (a detailed functioning can be found in Figure 18). The analysis consists with a comparison of quantified electrophoresis peaks between cases and controls, and allows identifying copy number variations and complete deletions. The panels we used to detect genetic abnormalities were panels covering subtelomeric regions (commercially available p36 and p70 kits by MRC-Holland, the Netherlands) and recurrent genomic disorders (in-house designed) covering regions such as 15q11, 15q13.3, 16p11.2, 7q11.2, 22q11, 1p36, 2q37, and which also allows sex determination (with Yp11.31 probe). For the commercially available MLPA, the probe sequences and capillary electrophoresis patterns can be found at [mlpa.com] For the in-house designed panel for genomic disorders, the genes assessed can be found in Table 15 and electrophoresis profile in Figure 19. With this assessment we can unravel some recurrent genetic abnormalities and conditions group patients by endogenotypes.

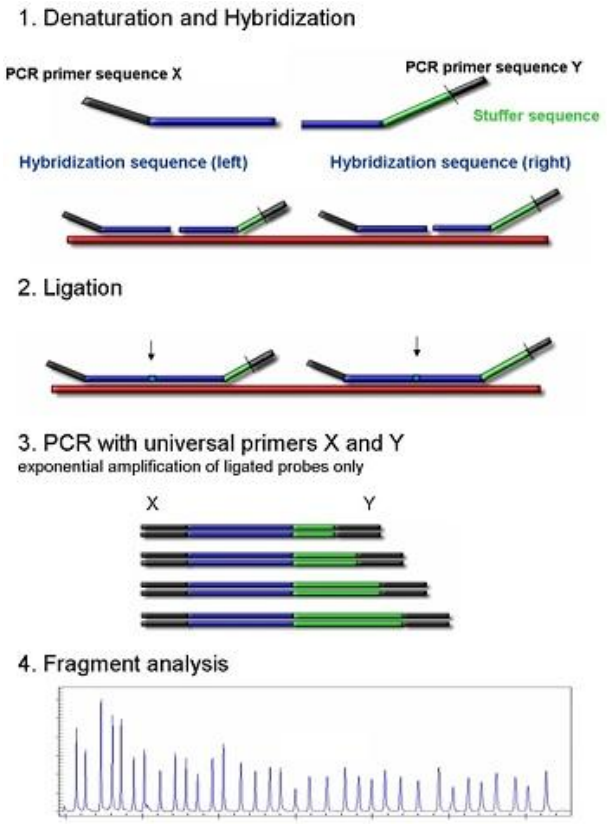


Figure 18. Scheme for the method followed by Multiplex ligation-dependent probe amplification (MLPA) technique. The primers contain a consensus sequence for a second set of universal primers, and a stuffer region of various lengths to create various sizes for the amplicons in order to be visualized in the capillary electrophoresis. Hybridization of primers (in a unique region and in the two alleles), followed by a ligation and amplification by the universal primers is analyzed by a fluorescent capillary sequencer and quantified.

Genomic Disorders in-house designed MLPA panel

Probe/Gene	Syndrome	Chr	Dye	Amplicon length (bp)
WBSCR1	Williams-Beuren Syn/ADS	7q11.23	FAM	89
HIRA	DiGeorge Syn/ASD	22q11.21	FAM	93
NSD1	Sotos Syn	5q35	FAM	96
SNRNP	Prader-Willi Syn/ASD	15q11.2	FAM	98
ARIH1	ID	15q24.1	FAM	102
TRIP3	ID	17q12	FAM	106
PML	ID	15q24.1	FAM	109
RAI1	Smith-Magenis Syn/ASD	17p11	FAM	112
COPS3	Smith-Magenis Syn/ASD	17p11	FAM	114
SHANK3	ASD	22q11	FAM	117
BAZ1B	Williams-Beuren Syn/ASD	7q11.23	FAM	119
EDC3	ID	15q24.1	FAM	122

PPCDC	ID	15q24.2	FAM	125
MAPT	ID	17q21.31	FAM	128
SPN	ID/ASD	16p	HEX	90
BCL9	ID	1q21.1	HEX	98
ALDOA	ID/ASD	16p	HEX	103
SCNNUG	ID	16p12.1	HEX	105
KLF13	ID/ASD	15q13.3	HEX	108
CDH9	Control	5p14.1	HEX	111
KIAA	Control	18q21.3	HEX	113
MECP2	ID	Xq28	HEX	120
GNB1L	DiGeorge Syn/ASD	22q11.21	HEX	123
FLJ20436	Control	12q13.1	HEX	129
SRY	Control	Yp11.31	HEX	131
SKI	1p deletion/ASD	1p36	HEX	148
TP73	1p deletion/ASD	1p36	HEX	151

Table 15. Genomic disorders panel. Genes covered by the panel probes, location (chromosomal band) and syndrome related with an anomaly in the region. ASD rearrangements include 7q11, 15q11-13, 16p11.2, 22q11.2 and deletion of 1p-terminal (1p36). Electrophoresis dyes are indicated as well as the approximate size of the peak generated, which can vary depending on the capillary electrophoresis machine. ID, Intellectual Disability; Syn, Syndrome. (Sequences for the right and left probes are available on demand).

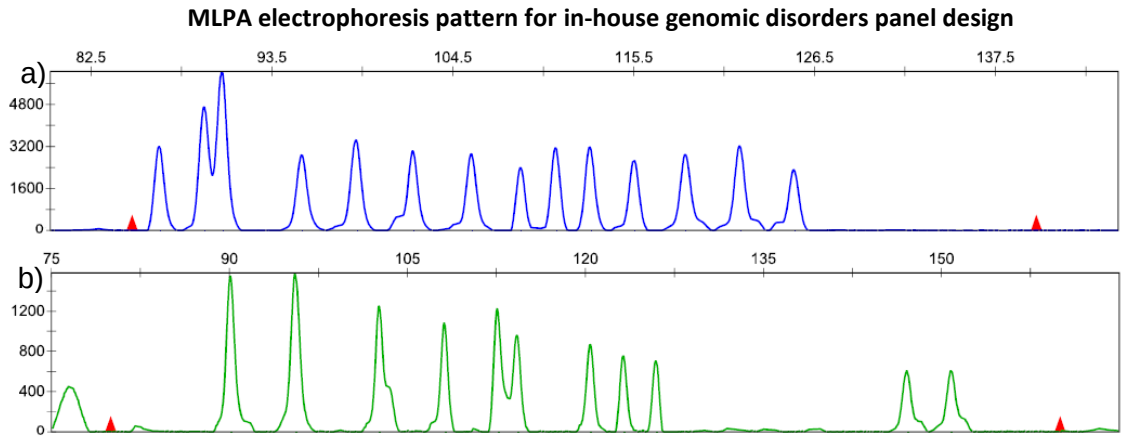


Figure 19. Electrophoresis pattern for in-house designed MLPA panels for the detection of genomic disorders. Two panels using different dye colors a) FAM™ (blue) and b) HEX™ (green) were used in the electrophoresis. Each peak accounts for a gene/region assessed following amplicon size (see Table 1).

If we did not observe any alteration by MLPA, a molecular karyotype by array comparative genome hybridization (aCGH) was performed. In the array, the questioned genome is hybridized to a reference genome to unravel the deletions and duplications with a higher resolution than the karyotype (reaching a resolution down to 100 kb with automated protocols).

The results for the genetic assessment for the 331 patients compiled in the UPF genetic Unit, classified the individuals as shown in Figure 20. Most of the patients (77%) remained with an idiopathic aetiology.

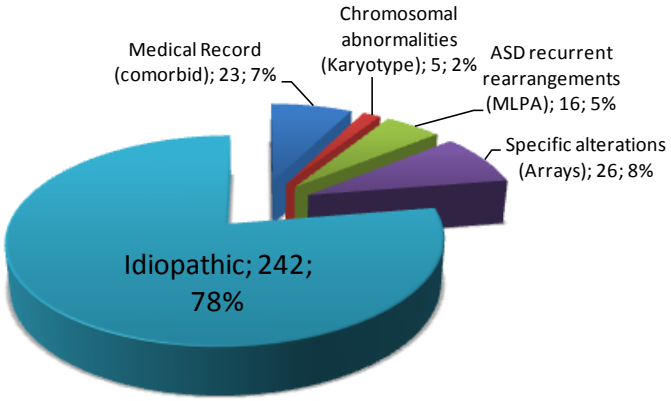


Figure 20. Classification of UPF ASD cohort after the genetic assessment (genetic causes from medical reports, previous biomedical tests and genetic assessment in the laboratory).

Additional research using next generation techniques could contribute to unravel the aetiology for some idiopathic patients. In a recent publication with this cohort, for 7 patients, which were first categorized as idiopathic, mutations were unraveled by exome sequencing and validated, with proven familial segregation[345].

Epigenetic assessment

We searched for genome-wide differential methylation patterns between patients (idopathic or with multiple subgroups) and controls with two genome-wide array technologies: Infinium HumanMethylation 27k and 450k by Illumina, San Diego, CA, USA.

Validation of the methylation alterations

Regarding the differential epigenetic regions obtained, we used additional tools to confirm methylation alterations:

1. Bisulfite-converted DNA PCR followed by direct Sanger sequencing (BS-PCR) or by cloning the PCR product. Sodium bisulfite deaminates unmethylated cytosine residues to uracil, but does not affect 5-mC or 5-hmC, which remains intact. Primers can be designed and PCR amplification can be performed. We assessed more than 40 BS-PCR

validations, most of them directly followed by direct Sanger Sequencing[314], as a qualitative approach. It should be taken into account that the methylation differences analyzed were high enough to be visualized (>18% to 50%).

2. Sequenom EPITYPER to quantify methylation of a CpG and ratio of a region. A total of 17 regions encompassing 39 CpGs for two samples and a control (showing a differential methylation >25%) were analyzed at Universitat de Valencia (Valencia, Spain) by Sequenom technology. Briefly, the DNA was bisulfite converted and a PCR reaction was performed with previously designed primers, followed by enzymatic specific cleavage (SAP enzyme) and an *in vitro* transcription and T-cleavage. The treated samples were run at matrix-assisted later desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) to quantify the methylation levels, showing the methylation pattern of the CpG and the neighbouring region up to 400pb. Each measurement was done in duplicate. Measures with a standard variation over a 0.1 were removed. T-test was done to calculate differential methylation ($p<0.01$).
3. Pyrosequencing was used to validate array probes, quantitatively sequencing 50bp around the probe. We validated 8 regions at Institut Català d'Oncologia, Barcelona, Spain, and at University California Davis, CA, USA. Using this technique we could also increase the cohort for 5 of these regions to 75 new individuals. DNA is bisulfite converted and followed by a PCR. A designed biotinilated primer binding nearby the CpG of interest, hybridize with the complementary strand and are incubated with an enzymatic mix (containing DNA polymerase, ATP sulfurylase, luciferase, and apyrase, adenosine 5' phosphosulfate (APS) and luciferin). A dNTP-type is added to the reaction and is incorporated if it is complementary to the template releasing a pyrophosphate (PPi), which is derived to ATP and oxyluciferin by the enzymes and is quantified. The height of the peak obtained is proportional to the nucleotides incorporated. The remaining nucleotides are degraded and a second nucleotide is added sequentially and generates a sequence.
4. Methylation Specific - Multiplex Ligation-dependent Probe Amplification (Ms-MLPA) was designed for 22 array regions to validate and apply to a larger cohort. Ms-MLPA principle is similar than a MLPA but includes a enzimatic target for a methylation-sensitive endonuclease (HhaI, AciI or others) in order to avoid amplification if the

target is unmethylated. In the protocol (Figure 21), a sample is treated in pararell with and without methylation-sensitive endonuclease, and the obtained quantification of the electrophoresis peaks of both reactions are compared. The patient pattern is compared also with the control pattern.

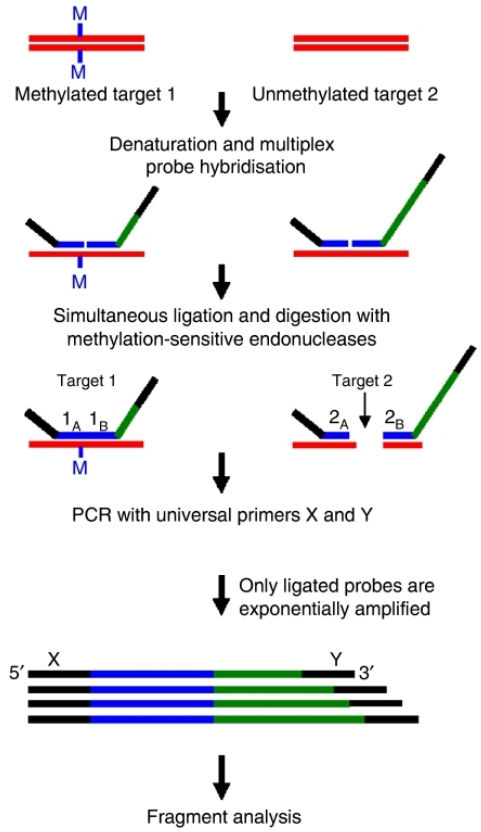


Figure 21. Scheme for the method followed by Methylation-specific Multiplex ligation-dependent probe amplification (Ms-MLPA) technique. The primers contain a consensus sequence for a second set of universal primers, and a stuffer region of various lengths to create various final sizes for the amplicons in order to be visualized in the capillary electrophoresis, and also a target sequence for methylation-sensitive endonuclease HhaI. Hybridization of primers (in a unique region and in the two alleles), followed by two parallel reactions: a ligation used as a control, and a ligation plus digestion by HhaI which gives a different pattern depending on the methylation status. Amplification by the universal primers is analyzed by a fluorescent capillary sequencer and quantified.

5. Whole-genome bisulfite sequencing (by MethylC-seq) was performed by collaborators in another tissue (placenta) studying ASD patients and controls. We correlated our data (DMCpGs, genes and regions) with their genome-wide methylation results. For

the preparation of the libraries, sonication of the genomic DNA and adapters Illumina adapters ligation, is followed by a bisulfite conversion and PCR amplification. The library was sequenced on Illumina HiSeq or GAII sequencers. Mapping was made with BS Seeker program. Genome-wide methylome was performed and processed at University California Davis, CA, USA.

Genomic DNA validation was implemented regarding that the BS-PCR sequence revealed sequence variants, located mainly in the CpG. For genetic validation we used:

1. Genomic PCR and Sanger Sequencing to validate more than 100 regions.
2. Sequenom for Genotyping was used to study 60 CpGs already confirmed to have a polymorphism, in 900 samples, and was performed at Universitat de València, València, Spain. The genotyping technique uses genomic DNA, and after performing a PCR, an hybridation with Sequenom primers is performed. A single extension with modified terminator nucleotides is performed and generates a specific peak. As primers can be multiplexed leading to different amplicons' sizes, a spectrum is obtained and the software extrapolates a genotype per polymorphism.
3. Genomic PCR followed by restriction enzymes was used for one high-throughput validation for one primer which could not be evaluated by Sequenom Genotyping. ApaI enzyme could differentiate both genotypes in the fragment analyzed and was used over 900 samples.

Expression studies

We correlated the epigenetic alterations with their consequences in expression. For these, we selected the patients with 1) available RNA sample and 2) which had already been studied by genome-wide epigenetic tecniques. We performed transcriptome studies (RNA-sequencing). Sequencing was performed paired-end 100bp reads using Illumina HiSeq™ 2000, and reads were processed by colleagues in our laboratory. The CpGs which correlated with expression were also genetically and epigenetically characterized (see *Validation of the methylation alterations* in this section).

CHAPTER 1

CHAPTER 1

GENETIC AND EPIGENETIC METHYLATION DEFECTS AND IMPLICATION OF THE *ERMN* GENE IN AUTISM SPECTRUM DISORDERS

Aïda Homs, Marta Codina, Benjamin Rodríguez-Santiago, Cristina M. Villanueva, David Monk, Ivon Cusco, and Luis A. Pérez-Jurado

In immediate submission to Molecular Psychiatry

Autism spectrum disorders (ASD) are a group of highly heritable neurodevelopmental conditions with a heterogeneous aetiology. Regarding that for only 25-30% of the patients the underlying causes have been found, many evidences support that a part of the missing heritability could be explained by an environmental contribution through the epigenetic mechanisms. For determining epigenetic aberrations in ASD patients, we performed methylation analysis of 53 male idiopathic ASD patients and 757 healthy controls in blood samples by Illumina HumanMethylation450k array. We also correlated the findings with expression alterations in blood by RNA-sequencing. We found differentially methylated CpGs, mostly hypomethylated, and having expression consequences in about 10% of the genes. We found regions with true epigenetic alterations, although inherited, and also a relevant amount of rare genetic variants altering methylation targets (meSNVs), also inherited from healthy parents. Six meSNVs were found significantly associated with ASD in a large-scale association study. One of them, *ERMN*, which showed expression consequences in one patient, also presented a higher mutation load in ASD patients in its coding sequence, being considered a novel ASD candidate gene. In summary, we found alterations which may contribute to the phenotype in a polygenic model and implicated *ERMN* as an ASD susceptibility-conferring gene.

Homs A, Codina-Solà M, Rodríguez-Santiago B, Villanueva CM, Monk D, Cuscó I, Pérez-Jurado LA. Genetic and epigenetic methylation defects and implication of the ERMN gene in autism spectrum disorders. Transl Psychiatry. 2016 Jul 12;6(7):e855. doi: 10.1038/tp.2016.120.

CHAPTER 2

CHAPTER 2

METHYLATION FINGERPRINTS AMONG SUBGROUPS OF AUTISM SPECTRUM DISORDERS WITH KNOWN AND UNKNOWN GENETIC CAUSE

Aïda Homs, Ivon Cusco and Luis A. Pérez-Jurado

In preparation

Autism spectrum disorders (ASD) comprise a group of neurodevelopmental disorders with clinical heterogeneity. In fact, ASD can be present associated to a hundred of comorbid conditions. ASD aetiology is contributed by a remarkable genetic background, evidenced by the high heritability of the conditions. However, for a third of cases their aetiology still remains unknown. Recently, an epigenetic involvement has been evidenced and detected in some ASD cases. Therefore, we have characterized blood methylation of ASD patients in three subgroups with known comorbid or recurrent chromosomal rearrangements: Down syndrome (DS) with trisomy 21, 15q11-q13 duplication (Dup15), and 1p36 terminal deletion (Del1p36). Also we studied patients with idiopathic ASD, some conceived by assisted reproductive techniques (ART), which are debated to have an influence to the methylation landscape. Although no global methylation pattern was identified in ASD versus controls, differentially methylated CpGs (DMCpGs) were identified in three groups of patients. For DS patients 354 DMCpGs and 7 regions altering genes were defined. The genes were involved in DS-related and ASD-related pathways. Dup15 patients presented hypermethylation in the region affected and at the imprinted domain at chr11p15. Apart from that, we obtained a methylation blood fingerprint which could classify some genetic causes of ASD. In summary, we detected specific methylation alterations according to the underlying genetic aetiology which also allowed the classification of the patients. The contribution of this work is defining new epigenetic mechanisms contributing to the disease for ASD patients, which could lead to improved therapies and interventions in the future.

METHYLATION FINGERPRINTS AMONG SUBGROUPS OF AUTISM SPECTRUM DISORDERS WITH KNOWN AND UNKNOWN GENETIC CAUSE

Aïda Homs, Ivon Cusco and Luis A. Pérez-Jurado

ABSTRACT

Background: Autism spectrum disorders comprise heterogeneous neurodevelopmental disorders diagnosed by the presence of two core symptoms: social communication deficits, and restricted repetitive behaviours, interests or activities (DSM-V). ASD can be comorbid with some conditions caused by chromosomal rearrangements, although the genetic aetiology is very complex and most cases are still idiopathic. Involvement of primary or secondary epigenetic alterations has been document in a proportion of ASD cases.

Methods: We have compared blood DNA methylation patterns of ASD patients, subclassified according to the presence or absence of three causal or comorbid chromosomal rearrangements: Down syndrome (DS) with trisomy 21, 15q11-q13 duplication (Dup15), and 1p36 terminal deletion (Del1p36). Among patients with idiopathic ASD, some had been conceived by assisted reproductive techniques (ART), which are debated to alter the methylation landscape.

Results: Although no global methylation pattern could differentiate the entire ASD group from controls, several differentially methylated CpGs (DMCpGs) were identified per group of patients. DS patients showed the most differentiated methylation pattern with 354 DMCpGs and 7 regions of clustered DMCpGs, 98.8% outside chromosome 21, altering genes involved in DS-related (embryo development and neuronal development) and ASD-related phenotypes (transcription regulation, axonogenesis). Dup15 patients presented hypermethylation in the region affected by the chromosomal imbalance as well as at the imprinted domain at chr11p15 and Del1p36 presented a few group-specific DMCpGs.

Conclusions: Specific methylation alterations can be detected in blood DNA of ASD patients according to the underlying genetic aetiology. These epigenetic marks provide a classification

tool for some genomic causes of ASD, and define pathogenic mechanisms of disease that could lead to more focused therapies and interventions.

INTRODUCTION

Autism Spectrum Disorder (ASD) comprises a group of severe neurodevelopmental diseases with complex multifactorial aetiology. Genetic and genomic approaches have unveiled the cause of roughly 20-35% of ASD cases, but it still remains elusive for a significant proportion[1,2]. ASD is found in association with other conditions, including known genetic disorders such as Fragile-X and Down syndromes[3], cytogenetic abnormalities, e.g. duplications at chr15q11.13 on the maternally inherited chromosome and terminal deletions of chromosomes 1p, 2q or 22q[4]. Additionally, submicroscopic rearrangements and point mutations are found in genes related to post synaptic density function, neuronal cell adhesion and genes related to transcription regulation among others[2,5,6].

Alterations of the epigenetic layer have also been demonstrated in ASD. Some diseases comorbid with ASD involve epigenetic mechanisms; for example, Fragile-X syndrome is caused by hypermethylation and transcriptional silencing of *FMR1*, Rett syndrome by mutations in the metyl-binding-domain chromatin remodeler *MECP2*, and CHARGE syndrome by mutations in the chromatin remodelling factor CDH7[7]. Epigenetic deregulation has also been implicated in some cases of ASD with chromosomal rearrangements altering imprinted regions in chr15q11-13, 7q21-31.31 and 7q32.3-36.3[8]. Finally, single genes presenting altered methylation patterns have been identified associated to ASD[9–15].

The genetic background is highly related and interconnected with the epigenetic genomic landscape. A recent report estimated that most of the differential methylation captured by an array was due to the genetic background (68%), and less than a third to the epigenetic changes (32%)[16]. This genetic-epigenetic interplay is well documented in several human disorders. For example, the triplet expansion causing Fragile-X syndrome at the *FMR1* promoter associates hypermethylation. Genomic rearrangements, genetic mutations and epimutations at several imprinting control regions of the human genome are the cause of developmental disorders, including Prader-Willi and Angelman syndromes, Beckwith-Wiedemann and Silver-Russell syndromes or pseudohypoparathyroidism, among others. Chromosomal aneuploidy disorders such as trisomy 21 are associated with remarkable alterations of genome-wide DNA methylation patterns[17].

Given the heterogeneity of the ASD causes, the determination of biological markers is challenging. Many studies have been addressed to identify biomarkers that would allow obtaining earlier diagnosis, prognosis or a better classification of ASD for therapeutic purposes. Examples include brain imaging by magnetic resonance (MRI), electroencephalography (EEG)[18–20], gene expression data[21–24], immune or hormonal parameters[25,26]. However, none of these biomarkers has been translated to the clinic.

The aim of this study was to explore genome-wide methylation patterns of blood DNA in several subgroups of ASD patients with comorbid chromosomal rearrangements (Down syndrome, 15q11-q13 duplication, and 1p36 terminal deletion) or idiopathic ASD including some individuals conceived with assisted reproductive technologies, along with controls, to define group-specific and/or divergent profiles that could be related to ASD.

MATERIALS AND METHODS

Subjects

We selected 14 individuals with either idiopathic ASD or representative comorbid chromosomal conditions out of a cohort of 331 well-studied patients with a diagnosis of ASD: four had Down syndrome (DS) due to trisomy 21 (T21), two had a maternal chromosome 15q11-q13 duplication, two had a chromosome 1p36 terminal deletion, and six had idiopathic ASD (ID), two of them conceived with assisted reproductive technology (ART) (Table 1). All patients were previously classified within one of the listed ASD categories in the Diagnosis and Statistical Manual of Mental Diseases (DSM-IV) using the ADI-R test. Patients' age ranged from 3 to 52 years old and both male and females, while controls were males under 11 years old. Genomic DNA was isolated from peripheral blood using the Blood DNA extraction kit (Qiagen, Hilden, Germany). Informed consent was obtained from the families or legal tutors. The study was approved by the medical ethical committee of the Parc de Salut Mar.

Additional data from 334 ASD patients and 327 healthy siblings (males, 6 to 25 years old, European descent) was downloaded from the database of Genotypes and Phenotypes (dbGAP) (accession number=phs000619) and used for comparison.

Patient	Genetic /Comorbid condition	Group	Sex	Age	ADI-R classif	Features
ASD_6	-	ID	M	7	AUTISM	IDis, stereotypia
ASD_22	-	ID	M	6	AUTISM (LF)	IDis
ASD_54	-	ID	F	11	ASPERGER	IDis, stereotypia
ASD_55	-	ID	F	11	AUTISM (HF)	IDis, stereotypia
ASD_56	-	ID (ART)	M	4	NQA	
ASD_57	-	ID (ART)	M	3	AUTISM (HF)	
ASD_58	Del 1p-ter	Del1pter	F	28	AUTISM	IDis, stereotypia, epilepsy, dymorphic
ASD_59	Del 1p-ter	Del1pter	F	4	AUTISM	
ASD_60	Dup 15q11-13 (maternal)	Dup15	M	6	AUTISM	IDis, stereotypia, epilepsy
ASD_61	Inv dup(15) (maternal)	Dup15	M	15	AUTISM	IDis, stereotypia, epilepsy
ASD_62	Down Syndrome (T21)	DS	M	36	AUTISM	stereotypia, epilepsy, dymorphic
ASD_63	Down Syndrome (T21)	DS	M	30	AUTISM	IDis, stereotypia, epilepsy, dymorphic
ASD_64	Down Syndrome (T21)	DS	M	47	AUTISM	IDis, epilepsy
ASD_65	Down Syndrome (T21)	DS	F	52	AUTISM	IDis, stereotypia, dymorphic

Table 1. Patients' characteristics, genetic and psychological diagnostics. ADI-R, Autism Diagnostic Interview-Revised; ART, conceived by Assisted Reproductive Technology; Del, Deletion; DS, Down Syndrome; Dup, duplication; HF, High Functioning; ID, Idiopathic aetiology; IDis, Intellectual Disability; LF, Low Functioning; NQA, Not quite Autism.

Illumina methylation array

500ng of genomic DNA per sample were modified by bisulfite using the EZ DNA Methylation Kit (Zymo Reasearch, Irvine, CA, USA) according to the manufacturer's instructions. Methylation analysis was performed using the Illumina Infinium Human Methylation27 BeadChip (2 chips per 12 samples each) (Illumina, San Diego, CA, USA). Bisulfite converted DNA was amplified, fragmented and hybridized to the BeadChip arrays at the National Center for Genotyping (CEGEN) (Madrid, Spain). Each CpG site is represented by 2 bead types: one for the methylated DNA sequence (M) and the other for unmethylated DNA (U). A single base extension was performed using labelled dNTPs. The arrays were imaged using a BeadArray™ Reader and the intensity data extraction was performed according to Illumina's instructions. The methylation status of a specific CpG site was calculated from the intensity of the M and U alleles, as the ratio of fluorescent signals β = intensity M Beads/[signal M+U beads]. DNA methylation β values are between 0 (absent methylation) and 1 (completely methylated). All samples passed the established quality control parameters, showed a complete conversion by bisulfite and had intensity values over 4000. The batch effect was assessed by a biological replicate in separate chips (Illumina treshold R^2>95%).

We first discarded probes that did not reach a significant threshold for detection signal (p.value<0.01), all chromosome X probes, those harbouring SNPs with Minor Allele Frequency

>0.002 (data from dbSNP v.138), and the reported cross-hybridization probes in this specific array[28,29]. The total number of probes analyzed was 23077 CpGs. The distribution across the chromosome and the CpGs obtained after filtering is represented in Figure . To compare ours with dbGAP data, we treated both data sets without normalization but with batch adjustment using Combat package[27], the method that provided the lowest variability between datasets.

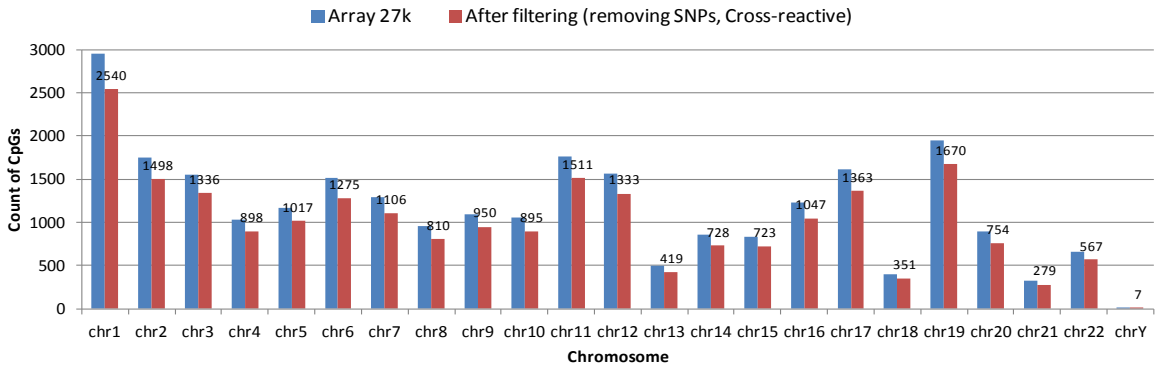


Figure 1. Number and chromosomal distribution of the probes before (blue) and after (red) filtering (number over the red bar).

Global methylation, clustering and statistical analyses

We studied the global methylation distribution of all the probes per individual in order to obtain a general view of the methylation profile. We assessed clustering of the samples using Euclidian clustering and also by Principal component Analysis (PCA).

To perform differential methylation analyses, we used Limma R package[30], which applies eBayes and Benjamini & Hochberg correction with False Discovery rate. For the comparison of all ASD or each group versus controls (our cohort controls and dbGAP controls), the significant adjusted p.value (adj.q.value) was set at 0.05. To perform region analysis we used COMB-P[31] package which combines the p.values retrieved from the Limma differential methylation and combines CpG positions with aberrant methylation. It also gives a p.value for the new region obtained, per sample. We analysed the presence of common altered regions within the groups. On the other hand, in the group-specific comparison versus all the other individuals, we set a p.value<0.05 and a delta-beta (Δβ) of 18%, which is the change of Beta compared to the reference group over 0.18 or under -0.18 (hypermethylation or hypomethylation respectively).

Linear Discriminant Analysis

Supervised analyses are classic classifiers which model data separating groups (given a conditional class/group). In LDA a linear distribution of data is obtained using eBayes predictions. We used the most differentially methylated CpGs (with a p.value<0.05 and a Delta Beta (Δβ) of /0.18/) comparing all the groups with the rest of individuals in order to assess a training group classification. The same LDA was applied to our cohort together with dbGAP data (patients and controls). Cross-validation with an independent cohort analyzed with the same genotypes was not possible, due to sample availability.

Gene Ontology

To search for enriched pathways, the list genes of genes associated with the DMCpGs were assessed with DAVIDGo database[32]. To search for over-represented molecular and biological functions we used the web resource CPDB[33].

RESULTS

Global methylation pattern

As a first approximation we performed a global methylation analysis. We observed a similar methylation distribution for all the samples and controls with a higher hypomethylation fraction, due to the biased coverage towards promoters in the array (Figure).

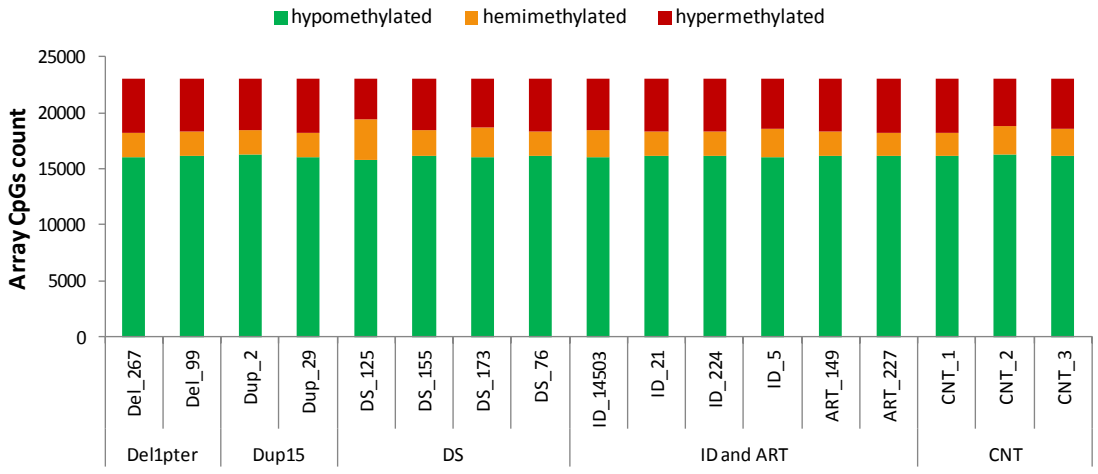


Figure 2. Distribution of the global methylation profile of the 27k probes in the 17 individuals studied. Most CpGs showed a hypomethylated state (Beta value <0.3) compared to hemimethylated state (0.3-0.7) and hypermethylated state (>0.7). No significant differences in global methylation were found comparing cases versus controls.

In order to define whether specific methylation alterations could correlate with the chromosomal regions altered in each ASD subgroup, we analyzed in detail the probes covering each chromosomal rearrangement. High correlation coefficients for β-values were obtained in all cases comparing subgroups versus controls, evidencing no significant methylation alterations.

To analyse the distribution of the samples and identify common methylation patterns or systemic differences between groups, we performed PCA and clustering analyses. We detected that three of the DS patients were located apart from the other samples, which were randomly distributed (Figure 3).

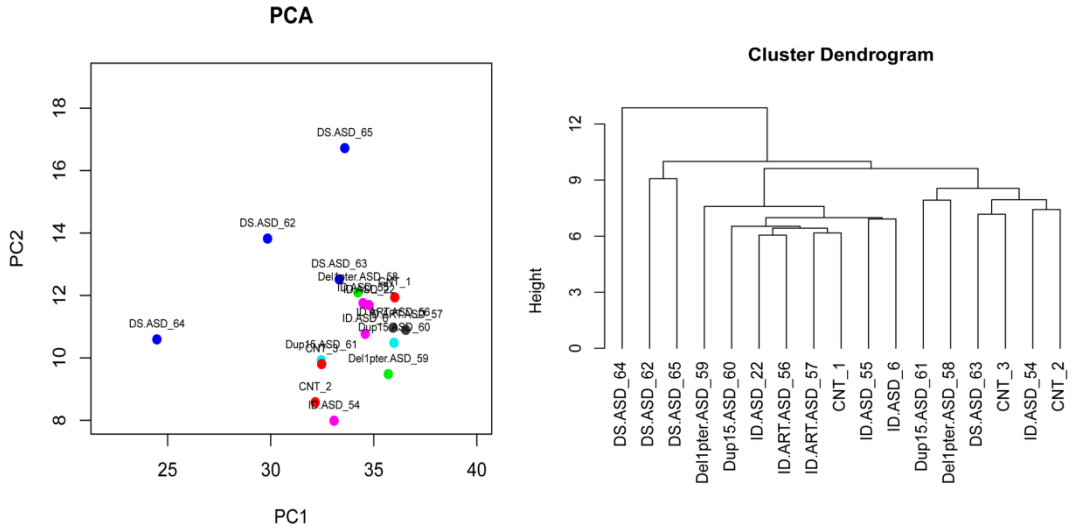


Figure 3. Sample distribution in a) PCA and b) dendrogram. A random distribution of controls (CNT) is observed in both figures (red dots in PCA), while 3 DS patients locate in PCA further than the rest of samples, and cluster also separately.

Differential methylation study

We performed differential methylation (DM) analysis to search for specific methylation following the approaches in Figure 4.

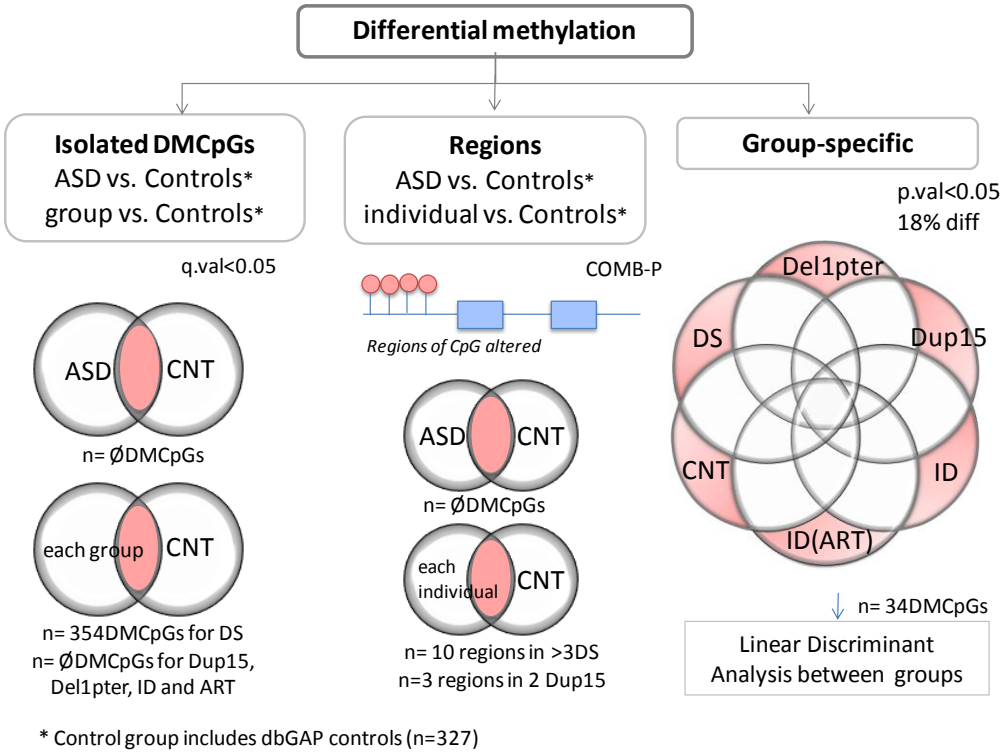


Figure 4. Strategies followed for differential methylation analyses.

As a first approach to identify common methylation aberrations, we compared all patients against controls (including dbGAP controls). This comparison did not show significant differences; therefore no common aberrations in ASD patients were detected. However, when comparing each group with controls, the DS group was the only one showing many statistically significant (q.value <0.05) differentially methylated CpGs (DMCpGs): a total of 354 DMCpGs encompassing 326 genes. Intriguingly, most of the DMCpGs (n=325, 91.8%) showed a relative hypermethylation, and were found widely spread among the autosomal chromosomes with only 4 DMCpGs (1.12%) located on chromosome 21. There is a high concordance and overlap (68 DMCpGs and 222 genes) with previously reported data on blood DNA of DS (Supplementary Table 1).

In addition to isolated CpGs, we also studied differentially methylated regions (DMRs) using the COMB-P algorithm[31]. Again, no common pattern among the different ASD groups could be detected. However, group-specific DMRs were found for DS and Dup15 (Table 2 and

Figure). We found 10 regions commonly altered in at least 3 of the DS patients and 41 regions commonly altered in 2 DS patients, none from chromosome 21. All these regions had been previously reported as altered in DS patients without ASD.

Group	27k probes Target_ID	Gene	Position (hg18)	Group mean	CNT mean
DS	cg11219178,cg12514506,cg07873128,cg23617121	*OSBPL5*	chr11:3077524-3078036	0.839	0.880
	cg09188980,cg14348532,cg06295856,cg09068492, cg22183706,cg01971122,cg26833169	*CALCA*	chr11:14949953-14950621	0.181	0.134
	cg21232015,cg02519218,cg20066677,cg24532476, cg17524886	*CHFR*	chr12:131934293-131934782	0.857	0.792
	cg17655614,cg11667754,cg11255163,cg23989635, cg00240312,cg22832044	*CDH1*	chr16:67328444-67329500	0.147	0.099
	cg22778981,cg26111030,cg11993754,cg16065186, cg27005179,cg13263114	*ERBB2*	chr17:35109658-35110596	0.163	0.128
	cg25602457,cg18841634,cg00399483,cg01839464, cg02624705,cg24084891	*DCC*	chr18:48122098-48122786	0.121	0.086
	cg07456645,cg21300318,cg20893022,cg01400401, cg03996793,cg04640886,cg02479575,cg26005082	*C19orf30*	chr19:4720145-4720660	0.200	0.177
	cg01381846,cg01354473,cg27009703,cg26521404, cg07778029,cg25047280	*HOXA9*	chr7:27171309-27172004	0.123	0.071
	cg09375488,cg25692621,cg26847490,cg11554605	*ASB4*	chr7:94952961-94953290	0.715	0.736
	cg19481686,cg04675937,cg08390209,cg18979223, cg10210238	*CDKN2B*	chr9:21995287-21995995	0.165	0.112
Dup15	cg06233503,cg19728223,cg17820828	*KCNQ1*	chr11:2769249-2769874	0.464	0.296
	cg12532169,cg13828758,cg18552939,cg01989224	*NDN*	chr15:21483462-21483851	0.498	0.410
	cg11265941,cg19803984,cg16321029	*SNRPN*	chr15:22619830-22619943	0.523	0.416

Table 2. Differentially methylated regions in DS and Dup15 groups. Clusters of DMCpGs encompassing several genes were found altered in at least 3 DS patients and 2 Dup15 patients.

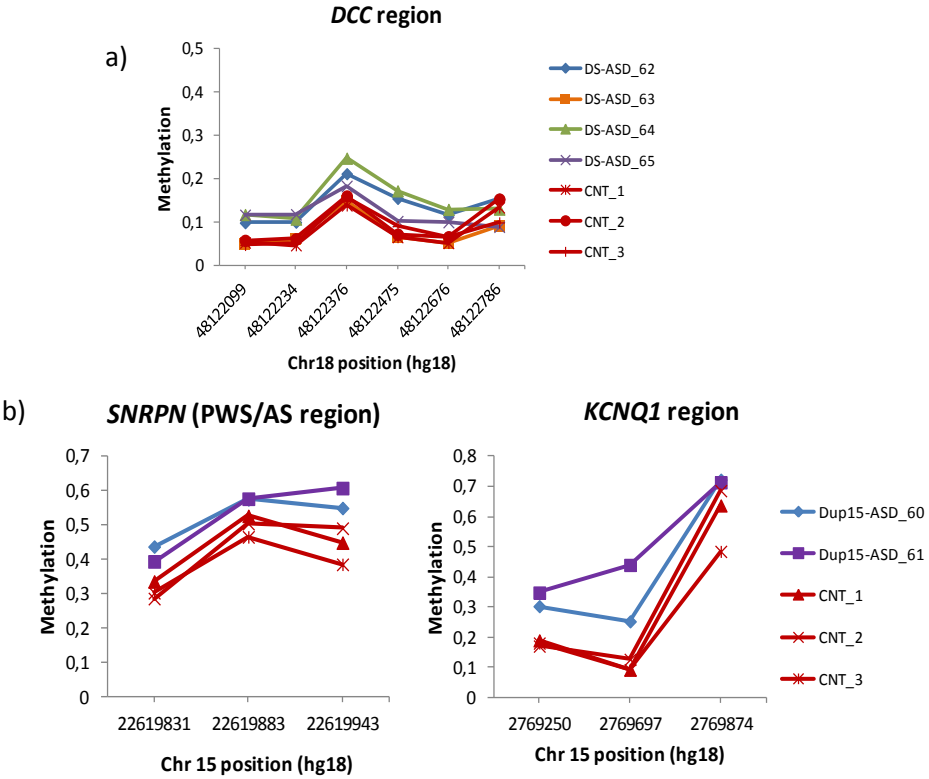


Figure 5. Example of the altered regions in DS and Dup15 groups. In a) DS patients, *DCC* showed several CpGs altered, and in b) Dup15 patients, two imprinted regions, chr15q11-13 (with *SNRPN* as example) and chr11 (*KCN*Q1), showed altered methylation. In all three regions, patients showed a relative hypermethylation.

We also found 3 DMRs specific of the Dup15 group, two of them located in the rearranged region itself and presenting relative hypermethylation of *NDN* and *SNRPN* as expected in association with the maternal duplication. The third DMR was located in chromosome 11p15, also an imprinted domain that contains a potassium channel involved in cardiac muscle cell repolarization (*KCNQ1*), showing relative hypermethylation. Reciprocal epigenetic alterations affecting this imprinted region are associated with either overgrowth or growth restriction of prenatal origin (OMIM *607542).

Classification of ASD subgroups

We performed a LDA to see if a methylation pattern could classify the multiple genotypes' subgroups. We identified group-specific DMCpGs by comparing each group versus all the other samples. For each analysis we selected the most extreme differences (Δβ≥/18%/ of methylation between groups and a p.value <0.05) and globally we obtained a total of 34 DMCpGs (from 5 to 14 CpGs group-specific) (Table 3). Performing LDA with these DMCpGs allowed classifying all three genotypes' subgroups as well as those ID individuals conceived by ART (ID-ART) from the controls group. The LDA analysis was able to discriminate 53.1 % (LD1) and 3.05% (LD2) of the individuals (Figure 6). The DMCpGs were located in various chromosomes and located in or nearby 27 genes.

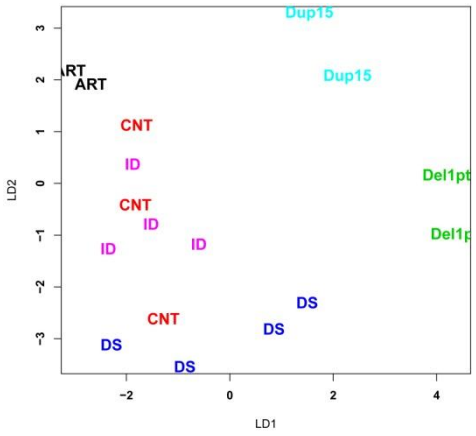


Figure 6. LDA classification of the genotypes' subgroups assessed using the methylation fingerprint of 34DMCpGs. The classification tool can differentiate Dup15, Del1pter, DS and ART from ID and CNT groups.

Group	Gene	CpGs	Position (hg18)	Mean β group	Mean β (others)	Δβ	P.value
ART	*FLJ90036*	cg12717203	chr4: 114341-114342	0.478	0.206	0.272	0.0305
ART	*HOP*	cg00019495	chr4: 57242281-57242282	0.509	0.288	0.221	0.0005
ART	*DUSP22*	cg11235426	chr6: 237521-237522	0.110	0.350	-0.240	0.0306
ART	*PSCA*	cg13446199	chr8: 143759867-143759868	0.190	0.409	-0.219	0.0052
ART	*UBL5*	cg17704839	chr19: 9800037-9800038	0.317	0.503	-0.186	0.0229
ART	*TNNI3*	cg18838701	chr19: 60360423-60360424	0.486	0.252	0.235	0.0118

Group	Gene	CpGs	Position (hg18)	Mean β group	Mean β (others)	Δβ	P.value
Del1p-ter	*ETNK2*	cg03718539	chr1: 202387211-202387212	0.006	0.196	-0.189	0.0141
Del1p-ter	*OR2L13*	cg20507276	chr1: 246167222-246167223	0.583	0.284	0.298	0.0498
Del1p-ter	*FLJ90036*	cg12717203	chr4: 114341-114342	0.002	0.270	-0.268	0.0334
Del1p-ter	*FLJ90036*	cg15792688	chr4: 114692-114693	0.029	0.225	-0.196	0.0318
Del1p-ter	*BXDC1*	cg05213296	chr6: 111409421-111409422	0.457	0.253	0.204	0.0019
Del1p-ter	*LOC349136*	cg19831077	chr7: 150738856-150738857	0.542	0.358	0.184	1.87E-05
Del1p-ter	*TP53INP1*	cg18059933	chr8: 96031638-96031639	0.391	0.206	0.185	0.0149
Del1p-ter	*FAM83A*	cg23067535	chr8: 124264313-124264314	0.439	0.757	-0.318	0.0022
Del1p-ter	*NRIP2*	cg05194726	chr12: 2814740-2814741	0.460	0.264	0.196	0.0006
Del1p-ter	*SLC38A4*	cg15584813	chr12: 45505892-45505893	0.551	0.265	0.286	3.79E-05
Del1p-ter	*SLC38A4*	cg07601320	chr12: 45506107-45506108	0.541	0.279	0.263	8.15E-06
Del1p-ter	*SMYD4*	cg04005701	chr17: 1680390-1680391	0.256	0.067	0.189	0.0019
Del1p-ter	*LOC126295*	cg08634464	chr19: 2852146-2852147	0.428	0.208	0.220	0.0087
Del1p-ter	*FLJ10781*	cg06851207	chr19: 51666977-51666978	0.210	0.438	-0.227	0.0007
DS	*ANKMY1*	cg08321346	chr2: 241145655-241145656	0.452	0.245	0.207	0.000
DS	*SH3BP2*	cg08822227	chr4: 2790265-2790266	0.348	0.136	0.212	0.000
DS	*SH3BP2*	cg07991621	chr4: 2790276-2790277	0.424	0.200	0.223	7.21E-05
DS	*HOXA4*	cg04317399	chr7: 27136837-27136838	0.637	0.435	0.201	0.000
DS	*HOXA4*	cg24169822	chr7: 27137518-27137519	0.634	0.418	0.215	0.000
DS	*RHOJ*	cg18771300	chr14: 62741489-62741490	0.720	0.466	0.254	0.026
DS	*MGC3207*	cg16474696	chr19: 13736013-13736014	0.485	0.255	0.229	0.037
DS	*C21orf56*	cg07747299	chr21: 46428479-46428480	0.404	0.205	0.198	0.027
DS	*C21orf56*	cg10296238	chr21: 46429601-46429602	0.542	0.340	0.202	0.032
Dup15	*ETNK2*	cg03718539	chr1: 202387211-202387212	0.403	0.143	0.261	0.0002
Dup15	*FLJ32569*	cg14893161	chr1: 204085873-204085874	0.124	0.514	-0.390	0.0167
Dup15	*KCNQ1*	cg06719391	chr11: 2510905-2510906	0.314	0.064	0.250	1.43E-05
Dup15	*KCNQ1*	cg19728223	chr11: 2769696-2769697	0.347	0.121	0.225	3.42E-05
Dup15	*MEOX1*	cg08471713	chr17: 39094418-39094419	0.804	0.547	0.257	0.0024

Table 3. Methylation marks (n=34DMCpGs) which in combination, produced a fingerprint which could differentiate DS, Del1pter, Dup15 and ART groups versus the rest of the groups (ID and CNT). Group-specific probes are shown. The tool was done using only our 27k data.

Using ontology analysis with genes altered by group-specific DMCpGs, we found over-represented pathways in the ID-ART group implicated in heart and muscle development (*HOPX* and *TNNI3*) (p.value<0.005), and the Dup15 group involved in embryonic development (*ETNK2, MEOX1*) (p.value<0.006). To search for convergent mechanisms among ASD subgroups, we performed ontology analysis together with all genes altered in the different groups. We found two over-represented pathways, chordate embryonic development represented by three genes (*ETNK2*, *HOXA4* and *MEOX1)* (p.value <0.03), and response to drugs represented by two genes (*TNNI3* and *KCNQ1)* (p.value <0.002). Other relevant functions found by CPDB resource were transcription regulation (*NRIP2, ZNF57/ LOC126295, TP53INP1* and *ZNF718/FLJ90036*), cell proliferation regulation and apoptosis (*PSCA, DUSP22, PM20D1/ FLJ32569*), cardiac growth and development (*HOP/HOPX, TNNI3*), signal transduction (*SH3BP2*), and pathways related to methylation (*SMYD4*) and reactive oxygen species (ROS) regulation (*TP53INP1, MGC3207*).

We performed LDA with the 34 DMCpGs as epi-markers in an additional cohort of ASD patients and siblings (dbGAP). All patients and healthy siblings for dbGAP clustered together, far from DS, Del1pter and Dup15 (Figure 7). The lack of other individuals with the same comorbid

chromosomal rearrangements in these cohorts precluded the validation of the used marker for classification.

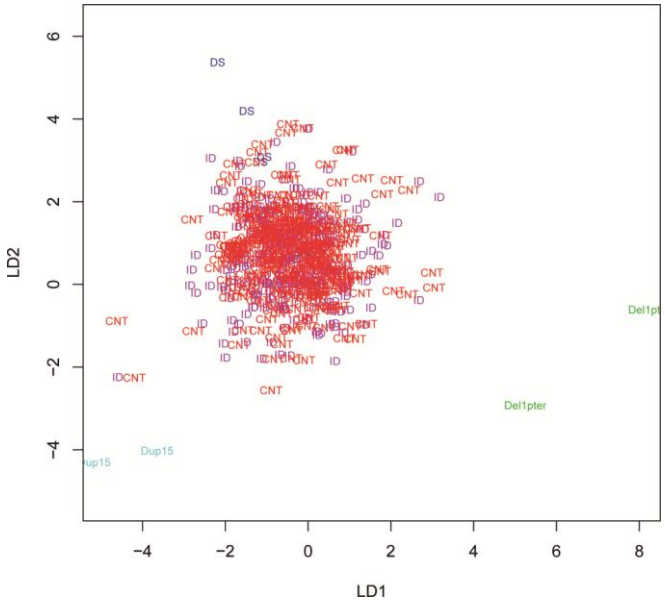


Figure 7. LDA classification of our cohort and dbGAP patients and controls assessed using the methylation fingerprint of 34DMCpGs. We observed that Dup15 and Del1pter are the most differentiated phenotypes. Two of the DS samples are also differentiated. ART, ID and CNT clustered together.

DISCUSSION

No global methylation alterations of blood DNA could be identified in the different subgroups of ASD cases with comorbid conditions (DS, Dup15, Del1p36) or idiopathic, as previously described in idiopathic ASD[11,12,34] and DS[35] methylation studies.

Although epidemiological studies show evidence for an increased risk of ASD related to ART procedures, the mechanisms for this risk are largely unknown and DNA methylation patterns have not been studied in detail[37–39]. ART has been associated with an increased risk of imprinting defects at the locus related to Beckwith-Wiedemann and Silver-Russell syndromes as well as with sporadic global demethylation at multiple loci[36]. However, we did not observe a genome-wide alteration or significant alterations at the known imprinting loci in the two ASD patients conceived by ART, although a few DMCpGs elsewhere in the genome were common to both patients.

Considering that the alterations may be local, we searched for subtle but specific methylation differences. We did not detect major methylation alterations on the aneusomic chromosomal regions in the comorbid conditions (chr21, chr15q11-13 and chr1p36) and we did not detect common alterations in cases with respect to controls. Therefore, convergent epigenetic marks detectable by the used technology do not appear to exist among these groups of patients with idiopathic or comorbid ASD, as seen in other studies using in brain[40] and blood[34] samples of ASD-ID patients. However, there are relevant methylations marks that are group-specific.

As previously shown in other studies and here replicated, trisomy21 is associated with widespread methylation alterations, mostly hypermethylation, not restricted to chr21[17,35,41–44]. The pattern in our patients, presenting with ASD associated to the trisomy 21, was not different from those with DS and no ASD. Genes affected by the aberrant methylation have been associated to DS, with functions in nervous system development, embryo development and transcription factor activity[17,43,44], and some have also been associated also to ASD, involved in calcium signalling, neuroactive ligand-receptor interaction, axon guidance, focal adhesion and Wnt signalling pathways[24,46–48]. It is interesting to highlight a DMRs altering the gene *DCC* which interacts with Down's syndrome Cell Adhesion Molecule (*DSCAM*) and plays a role in signalling and axon guidance in vertebrates[49]. *DSCAM* also has been associated to the congenital heart disease for DS patients[50]. Other interesting genes for the ASD-DS phenotype include *CDH1* coding for an adhesion molecule that plays a role in axonogenesis[51], the developmental genes (*ERBB2* and *HOXA9)* and two genes regulated by imprinting (*ASB4* and *OSBPL5*).

In the Dup15 group, not surprisingly, two altered loci were located in the duplicated region, showing hypermethylation of two maternally imprinted genes (*NDN* and *SNRPN*). These methylation levels correlate with previous studies of brain samples of ASD patients with similar genotypes[52]. Interestingly, these patients also revealed hypermethylation at the *KCNQ1* imprinted domain of chromosome 11p15. The imprinting control regions located in *SNRPN* and *KCNQ1* share a conserved structure and signals for the regulation and establishment of the imprinting signature during embryo development[53,54], reason why they might also share susceptibility to harbour concurrent alterations.

By LDA analysis we were able to classify the patients into subgroups with comorbid conditions, discriminating 56.15% of the individuals. Unfortunately, we have not been able to perform cross-validation in another set of samples to establish the classification power.

Among the group-specific alterations, we found several regions for Del1p36 patients and ART groups. For Del1pter patients, genes were involved in development (*ETNK2* and *SMYD4*), transcription, ribosome processes (*BXDC1*) and oxidative stress regulation (*TP53INP1*); the latter two associated together to ASD in a model in which ROS levels inversely depends on the copy number of ribosomal RNA genes[55]. For ART group, functions of deregulated genes include cardiac development (*HOP* and *TNNI3*) and cell proliferation and apoptosis (*DUSP22*, *PSCA and UBL5*), the latter gene associated to schizophrenia and depression[56,57] diseases which share genes with ASD. Regarding the convergence between the genotypes' subgroups deregulated functions; we find a significant enrichment in embryonic development and response to drugs pathways.

This study has some limitations such as the relative low coverage of the regions assessed for quantitative methylation, the study of blood DNA, and the different age range and gender which can affect methylation levels. However, the age differences assessed by array platforms described an small percentage of probes affected (<15%)[58,59], and both male and females have been assessed together in methylation studies, after removing the X-chromosome. Despite from that, our goal was to obtain convergences and divergences in methylation in patients with ASD with an altered genetic background, which would set the strongest pattern.

In conclusion, our results show that global methylation patterns, constrained by the platform resolution, are not significantly altered in any of the studied ASD samples or groups of patients. However, differential methylation was found in several sites and regions in groups of ASD patients with comorbid chromosomal aneusomies, allowing their classification. The genes and pathways altered by methylation aberrations in these groups of patients provide novel insight to unravel the complex and heterogeneous pathogenic mechanisms that can lead to the ASD phenotype.

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Differentially methylated CpGs (DMCpGs) in DS patients compared to controls.

REFERENCES

1 Krumm N, O'Roak BJ, Shendure J, Eichler EE. A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci* 2014; **37**: 95–105.

2 Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet* 2008; **9**: 341–355.

3 Betancur C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res* 2011; **1380**: 42–77.

4 Devlin B, Scherer SW, Emanuel B, Warren S. Genetic architecture in autism spectrum disorder. *Curr Opin Genet Dev* 2012; **22**: 229–237.

5 Brandler WM, Sebat J. From De Novo Mutations to Personalized Therapeutic Interventions in Autism. *Annu Rev Med* 2015; **66**: 487–507.

6 Jeremy Willsey a, State MW. Autism spectrum disorders: from genes to neurobiology. *Curr Opin Neurobiol* 2015; **30**: 92–99.

7 Tordjman S, Somogyi E, Coulon N, Kermarrec S, Cohen D, Bronsard G *et al.* Gene × Environment interactions in autism spectrum disorders: role of epigenetic mechanisms. *Front psychiatry* 2014; **5**: 53.

8 Schanen NC. Epigenetics of autism spectrum disorders. *Hum Mol Genet* 2006; **15 Spec No**: R138–50.

9 Nguyen A, Rauch T a, Pfeifer GP, Hu VW. Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB J* 2010; **24**: 3036–3051.

10 Nardone S, Sharan Sams D, Reuveni E, Getselter D, Oron O, Karpuj M *et al.* DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Transl Psychiatry* 2014; **4**: e433.

11 Ladd-Acosta C, Hansen KD, Briem E, Fallin MD, Kaufmann WE, Feinberg AP. Common DNA methylation alterations in multiple brain regions in autism. *Mol Psychiatry* 2013. doi:10.1038/mp.2013.114.

12 Wong CCY, Meaburn EL, Ronald A, Price TS, Jeffries AR, Schalkwyk LC *et al.* Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol Psychiatry* 2013. doi:10.1038/mp.2013.41.

13 Wang Y, Fang Y, Zhang F, Xu M, Zhang J, Yan J *et al.* Hypermethylation of the enolase gene (ENO2) in autism. *Eur J Pediatr* 2014; **173**: 1233–44.

14 Gregory SG, Connelly JJ, Towers AJ, Johnson J, Biscocho D, Markunas CA *et al.* Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. *BMC Med* 2009; **7**: 62.

15 James SJ, Shpyleva S, Melnyk S, Pavliv O, Pogribny IP. Complex epigenetic regulation of engrailed-2 (EN-2) homeobox gene in the autism cerebellum. *Transl Psychiatry* 2013; **3**: e232.

16 Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J *et al.* DNA methylation contributes to natural human variation. *Genome Res* 2013; **23**: 1363–72.

17 Bacalini MG, Gentilini D, Boattini A, Giampieri E, Pirazzini C, Giuliani C *et al.* Identification of a DNA methylation signature in blood cells from persons with Down Syndrome. *Aging (Albany NY)* 2015; **7**: 82–96.

18 Akshoomoff N, Lord C, Lincoln AJ, Courchesne RY, Carper RA, Townsend J *et al.* Outcome classification of preschool children with autism spectrum disorders using MRI brain measures. *J Am Acad Child Adolesc Psychiatry* 2004; **43**: 349–357.

19 Singh V, Mukherjee L, Chung MK. Cortical surface thickness as a classifier: Boosting for autism classification. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2008, pp 999–1007.

20 Ingalhalikar M, Parker D, Bloy L, Roberts TPL, Verma R. Diffusion based abnormality markers of pathology: Toward learned diagnostic prediction of ASD. *Neuroimage* 2011; **57**: 918–927.

21 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. *J Am Acad Child Adolesc Psychiatry* 2012; **51**. doi:10.1016/j.jaac.2012.07.007.

22 Kong SW, Collins CD, Shimizu-Motohashi Y, Holm IA, Campbell MG, Lee IH *et al.* Characteristics and Predictive Value of Blood Transcriptome Signature in Males with Autism Spectrum Disorders. *PLoS One* 2012; **7**. doi:10.1371/journal.pone.0049475.

23 Pramparo T, Pierce K, Lombardo M V., Carter Barnes C, Marinero S, Ahrens-Barbeau C *et al.* Prediction of Autism by Translation and Immune/Inflammation Coexpressed Genes in Toddlers From Pediatric Community Practices. *JAMA Psychiatry* 2015; **72**: 386–394.

24 Kong SW, Shimizu-Motohashi Y, Campbell MG, Lee IH, Collins CD, Brewster SJ *et al.* Peripheral blood gene expression signature differentiates children with autism from unaffected siblings. *Neurogenetics* 2013; **14**: 143–152.

25 Schwarz E, Guest PC, Rahmoune H, Wang L, Levin Y, Ingudomnukul E *et al.* Sex-specific serum biomarker patterns in adults with Asperger's syndrome. Mol. Psychiatry. 2011; **16**: 1213–1220.

26 Momeni N, Bergquist J, Brudin L, Behnia F, Sivberg B, Joghataei MT *et al.* A novel blood-based biomarker for detection of autism spectrum disorders. Transl. Psychiatry. 2012; **2**: e91.

27 Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012; **28**: 882–3.

28 Chen Y, Choufani S, Ferreira JC, Grafodatskaya D, Butcher DT, Weksberg R. Sequence overlap between autosomal and sex-linked probes on the Illumina HumanMethylation27 microarray. *Genomics* 2011; **97**: 214–22.

29 Chen Y, Choufani S, Grafodatskaya D, Butcher DT, Ferreira JC, Weksberg R. Cross-reactive DNA microarray probes lead to false discovery of autosomal sex-associated DNA methylation. *Am J Hum Genet* 2012; **91**: 762–4.

30 Du P, Zhang X, Huang C-C, Jafari N, Kibbe W a, Hou L *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010; **11**: 587.

31 Pedersen BS, Schwartz DA, Yang I V., Kechris KJ. Comb-p: Software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics* 2012; **28**: 2986–2988.

32 Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003; **4**: P3.

33 Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 2011; **474**: 380–4.

34 A. Homs, M. Codina, B. Rodríguez-Santiago, C. M. Villanueva, D. Monk IC and LAP-J. Genetic and epigenetic methylation defects and implication of the ERMN gene in autism spectrum disorders. *Prep* 2015.

35 Kerkel K, Schupf N, Hatta K, Pang D, Salas M, Kratz A *et al.* Altered DNA methylation in leukocytes with trisomy 21. *PLoS Genet* 2010; **6**. doi:10.1371/journal.pgen.1001212.

36 Owen CM, Segars JH. Imprinting disorders and assisted reproductive technology. Semin. Reprod. Med. 2009; **27**: 417–428.

37 Kissin DM, Zhang Y, Boulet SL, Fountain C, Bearman P, Schieve L *et al.* Association of assisted reproductive technology (ART) treatment and parental infertility diagnosis with autism in ART-conceived children. *Hum Reprod* 2014; **30**: 454–465.

38 Bay B, Mortensen EL, Hvidtjørn D, Kesmodel US. Fertility treatment and risk of childhood and adolescent mental disorders: register based cohort study. *BMJ* 2013; **347**: f3978.

39 Fountain C, Zhang Y, Kissin DM, Schieve LA, Jamieson DJ, Rice C *et al.* Association Between Assisted Reproductive Technology Conception and Autism in California, 1997–2007. *Am J Public Health* 2015; **105**: e1–e9.

40 Ginsberg MR, Rubin R a, Falcone T, Ting AH, Natowicz MR. Brain transcriptional and epigenetic associations with autism. *PLoS One* 2012; **7**: e44736.

41 Jin S, Lee YK, Lim YC, Zheng Z, Lin XM, Ng DPY *et al.* Global DNA Hypermethylation in Down Syndrome Placenta. *PLoS Genet* 2013; **9**. doi:10.1371/journal.pgen.1003515.

42 Yuen RK, Neumann SM, Fok AK, Peñaherrera MS, McFadden DE, Robinson WP *et al.* Extensive epigenetic reprogramming in human somatic tissues between fetus and adult. *Epigenetics Chromatin* 2011; **4**: 7.

43 Jones MJ, Farré P, McEwen LM, Macisaac JL, Watt K, Neumann SM *et al.* Distinct DNA methylation patterns of cognitive impairment and trisomy 21 in Down syndrome. *BMC Med Genomics* 2013; **6**: 58.

44 Eckmann-Scholz C, Bens S, Kolarova J, Schneppenheim S, Caliebe A, Heidemann S *et al.* DNA-methylation profiling of fetal tissues reveals marked epigenetic differences between Chorionic and Amniotic samples. *PLoS One* 2012; **7**. doi:10.1371/journal.pone.0039014.

45 Serra-Juhé C, Cuscó I, Homs A, Flores R, Torán N, Pérez-Jurado LA. DNA methylation abnormalities in congenital heart disease. *Epigenetics* 2015; **10**: 167–77.

46 Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* 2014; **94**: 677–694.

47 Zhang Y, Yuan X, Wang Z, Li R. The canonical Wnt signaling pathway in autism. *CNS Neurol Disord Drug Targets* 2014; **13**: 765–70.

48 Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare De Novo Variants Associated with Autism Implicate a Large Functional Network of Genes Involved in Formation and Function of Synapses. *Neuron* 2011; **70**: 898–907.

49 Ly A, Nikolaev A, Suresh G, Zheng Y, Tessier-Lavigne M, Stein E. DSCAM is a netrin receptor that collaborates with DCC in mediating turning responses to netrin-1. *Cell* 2008; **133**: 1241–54.

50 Barlow GM, Chen XN, Shi ZY, Lyons GE, Kurnit DM, Celle L *et al.* Down syndrome congenital heart disease: a narrowed region and a candidate gene. *Genet Med* 2001; **3**: 91–101.

51 Huynh MA, Stegmüller J, Litterman N, Bonni A. Regulation of Cdh1-APC function in axon growth by Cdh1 phosphorylation. *J Neurosci* 2009; **29**: 4322–4327.

52 Hogart A, Leung KN, Wang NJ, Wu DJ, Driscoll J, Vallero RO *et al.* Chromosome 15q11-13 duplication syndrome brain reveals epigenetic alterations in gene expression not predicted from copy number. *J Med Genet* 2009; **46**: 86–93.

53 Paoloni-Giacobino A, D'Aiuto L, Cirio MC, Reinhart B, Chaillet JR. Conserved features of imprinted differentially methylated domains. *Gene* 2007; **399**: 33–45.

54 Reinhart B, Paoloni-Giacobino A, Chaillet JR. Specific differentially methylated domain sequences direct the maintenance of methylation at imprinted genes. *Mol Cell Biol* 2006; **26**: 8347–8356.

55 Porokhovnik LN, Passekov VP, Gorbachevskaya NL, Sorokin AB, Veiko NN, Lyapunova NA. Active ribosomal genes, translational homeostasis and oxidative stress in the pathogenesis of schizophrenia and autism. *Psychiatr Genet* 2015; **25**: 79–87.

56 Bernstein H-G, Lendeckel U, Dobrowolny H, Stauch R, Steiner J, Grecksch G *et al.* Beacon-like/ubiquitin-5-like immunoreactivity is highly expressed in human hypothalamus and increased in haloperidol-treated schizophrenics and a rat model of schizophrenia. *Psychoneuroendocrinology* 2008; **33**: 340–51.

57 Kambe Y, Miyata A. Potential involvement of the mitochondrial unfolded protein response in depressive-like symptoms in mice. *Neurosci Lett* 2015; **588**: 166–71.

58 Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A *et al.* Distinct DNA methylomes of newborns and centenarians. 2012. doi:10.1073/pnas.1120658109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1120658109.

59 Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 2013; **49**: 359–67.

CHAPTER 3

CHAPTER 3

PLACENTA AND BLOOD METHYLATION PROFILING OF AUTISM SPECTRUM DISORDERS (ASD)

Aïda Homs, Diane Schroeder, Janine LaSalle, Luis A. Pérez-Jurado, Ivon Cuscó

Environmental factors can have an impact in embryo developmental periods affecting epigenetic mechanisms which regulate crucial differentiation and developmental processes. Placental tissue provides an epigenetic landscape which recapitulates the placental environment *in utero*, and is a valuable tissue to report aberrations leading to diseases such as ASD. Therefore, we compared methylation between ASD and healthy individuals, both in placental tissue (giving birth to an ASD or a typically developing child), and peripheral blood. Methylation data was assessed by genome-wide bisulfite sequencing (MethylC-seq) in placental tissue and by Infinium HumanMethylation450k array in blood. For this study, we used the results of CHAPTER 1 of this thesis. This study intended to find convergent methylation alterations. However, the placental data showed low sequencing depth, which compromised the outcome of the analyses. Despite this limitation, we found convergent methylation aberrations in a few isolated CpGs, but not in regions. Besides, we found regions in placenta encompassing ASD candidate genes. In the future, improved techniques will be needed to establish reliable methylation status, and the definition of common aberrations in ASD, and their onset in these tissues, awaits improved studies.

PLACENTAL AND BLOOD METHYLATION PROFILING OF AUTISM SPECTRUM DISORDERS (ASD)

Aïda Homs, Diane Schroeder, Janine LaSalle, Luis A. Pérez-Jurado, Ivon Cuscó

ABSTRACT

Background: Autism spectrum disorder (ASD) is currently considered a complex disorder with an epigenetic contribution. Environmental factors including xenobiotic exposures and life style factors can impact epigenetic mechanisms in all life stages and can exert intergeneracional effects. *In embryo* developmental period cell-specific re-programming and differentiation processes are driven by epigenetic mechanisms and patterns *and in utero* environment can affect the process. Therefore, studies *in utero* environment (e.g. placental tissue) may recapitulate biomarkers and provide developmental aberrations leading to diseases such as ASD.

Methods: We compared methylation data between ASD and healthy individuals in both placental tissue (giving birth to an ASD or typically developing child) and peripheral blood (MethylC-seq and Human Infinium 450k respectively) to identify common aberrations. We also compared previously identified genes located in partially or high methylated domains (PMDs/HMDs) in neurons and placenta, to genes with aberrant methylation in ASD blood.

Results/Discussion: This study is an attempt to identify convergent methylation alterations related to ASD in blood and placental tissues, as an indicative of an early developmental origin of the alterations found related to the disorder. However, we observed that technical features were an impediment to achieve our aim, as the sequencing performed by a next-generation sequencing technique (MethylC-seq) displayed low sequencing depth. This coverage issue limited the outcome of the analyses implying that the results are provisional, and that further studies are needed to confirm them. In spite of that, methylation comparisons performed between tissues evidenced the presence of similar methylation aberrations in 7CpGs in ASD patients, whereas no differentially methylated regions (DMRs) were found commonly altered. Besides, analyzing placental differential methylation in patients we found 18DMRs encompassing genes previously associated to the disorder. In the future, improved techniques to assess methylation will allow a better characterization of the methylation status, comparison between tissues, and in similar studies, the possibility to define the onset of convergent alterations found in a disease.

INTRODUCTION

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder characterized by social communication impairment, and restricted repetitive interests and behaviours[1]. ASD is highly heritable with a remarkable genetic component with an environmental contribution[2–5]. There are several evidences of environmental influences in ASD aetiology. In twin studies, the discordance for ASD diagnosis and severity within monozygotic twin pairs; and the higher ASD recurrence-risk for dyzigotic twins than their siblings, could be explained by environment–epi/genetics interactions, as well as early somatic mutations[3,6,7]. Moreover, heritability studies indicate a larger environmental contribution (55%) as opposed to genetic heritability (37%)[2], or a more moderate but still important environmental contribution[4,5,8]. Additionally, many studies involve contributing environmental factors in ASD aetiology, in prenatal, perinatal and post-natal stages, such as drugs (e.g. thalidomide and valporoic acid)[9,10], xenobiotics (heavy metals) and pollutants (pesticides, chemicals and air pollution) [11–13], alcohol and diet among many others[14,15]. Therefore, biological mechanisms modulated by environmental influences would be involved in the development of ASD, through interactions gene-environment or epigenetic-environment, leading to a regulation of gene expression[15].

DNA methylation is an epigenetic layer which can be altered by environment[16–20]. Evidences are demonstrated in an study with prenatal arsenic exposures, which showed hypermethylation in certain genes in cord blood[21,22] and generalized hypermethylation in leukocytes[23]. In ASD, pollutants with neurotoxic effects have been involved in hypomethylation of repetitive elements[24]. Additionally, methylation changes related to extrinsic or intrinsic oxidative stress have been implicated in the disease[25].

In prenatal development, post-zygotic reprogramming and differentiation processes are highly regulated by epigenetics processes, which can be negatively impacted by environmental factors[15]. Moreover, central nervous system development is also very sensitive to environment, which can alter fetal brain and have long-term consequences[26]. Therefore, methylation studies in early developmental stages using tissues as placental, which regulates pregnancy development, and is indicative of *in utero* environment and exposure in pregnancy, is a valuable tissue to study this pathology[19,27].

The placental epigenome displays lower level of methylation compared to most of the human tissues. It is composed by partially methylated domains (PMDs): large domains (>100Kb) which

show a reduced methylation (60-70%); spaced by highly methylated domains (HMDs) which show higher methylation levels (around 90%). These PMDs involve around 40% of the placental epigenome (about 3815 genes)[28], and are stable throughout pregnancy. PMDs associate to repressive chromatin state and reduced expression of irrelevant genes for placental development, whereas HMDs are enriched in important genes for placental development, pregnancy and immune response[29]. PMDs/HMDs are also present in other cell types as fibroblasts, in which were first discovered in 2009[30], neuronal cells, adipocytes, epithelial cells, and differentiated cells in breast and colorectal cancer[29], and may include other types in the future by further methylome sequencing studies. To define relevant genes in different tissues harbouring PMDs/HMDs, placenta domains were compared to neuronal cells (SH-SY5Y) and fetal lung fibroblasts (IMR90), and important specific-genes were established for each cell-type[28].

The central aim of this study is to compare the methylation alterations found (in CpGs and genes altered) in ASD peripheral blood (results from a previous study) to placental data (chorionic villus, CVS) obtained for a new subset of ASD patients and typically developing individuals. The goal is to find convergent alterations within the two tissues which reinforce the alteration and could indicate an early developmental origin of autism spectrum disorder.

METHODS

Infinium Illumina 450k array

Fifty-three ASD peripheral blood samples and 756 control samples, males from a wide range of age, were analyzed on the Illumina HumanMethylation 450k Bead Chip (Illumina, San Diego, CA, USA). The data is in publishing process[31] (please refer to the CHAPTER1 of the thesis to see the protocol followed and the analysis performed). Briefly, 53 ASD patients' data and 10 controls were from our cohort were added to 91 controls obtained in collaboration with the Center for Research in Environmental Epidemiology (CREAL) and Bellvitge Biomedical Research Institute (IDIBELL)[32,33]. The rest of the controls (n=656) were obtained from GEO dataset: GSE40279, males of a wide age range. Briefly, DNA was bisulfite-converted and hybridized to Bead Chips as the manufacturer's protocol. Quality control and data quantile normalization was applied (SSN). Additionally, cross-reactive probes and high frequency SNPs (MAF>0.005) within the CpG were removed. In the differential methylation analysis performed by Limma

package[34], a total of 700 differentially methylated CpGs (DMCpGs) were found comparing ASD individually versus controls.

Genome-wide methylation analysis

Fifteen ASD and twelve control placental samples from full-term pregnancies were obtained after cesarean sections, and chorionic villous tissue (CVS) was extracted from the placental fetal side. MethylC-seq or also called whole-genome bisulfite sequencing (WGBS) was assessed. The protocol was performed in duplicate per each sample in the Medical Microbiology and Immunology Department at University California Davis (Davis, CA, USA). The protocol is detailed in a previous publication[35]. Briefly, DNA was extracted, sonicated to ~300 bp and Illumina adapters were ligated. After, the library was bisulfite converted, amplified by PCR, and sequenced single-read on Illumina HiSeq obtaining 180M reads. Mapping of the reads was made over build hg18 using BS Seeker Package[36] and 55M reads 85-100pb were mapped (allowing a maximum of 3 mismatches). The minimum reads aligned were five reads (3X depth of coverage) although the final coverage obtained was 1.6X. To determine placental domains (PMDs and HMDs), CpG Island were removed to avoid bias due to their low methylation and two-state Hidden Markov Model (HMM) using StochHMM[37] program was applied with predictions on 2Mb windows of sequence (step size of 250Kb) as described also in another publication[28].

Global data exploration

For the subset of 700DMCpGs found in a previous study comparing ASD versus controls[31], we assessed correlation for 450k blood control samples (CNT-B) and MethylC-seq typically developing placenta samples (TD-P) in order to visualize the correlation. We assessed the Pearson correlation and calculated the coefficient between the combined values for TD-P and all CNT-B.

Differential methylation in isolated CpGs

We obtained for 688DMCpGs out of the 700DMCpGs, the MethylICseq values for the same positions for 15 ASD-P and 12 TD-P (two ASD were discarded after showing over 80% of empty values). For blood DMCpGs, we performed differential methylation analysis with Limma R package[34]. For placenta, we searched for alterations of a 20% methylation difference in ASD

versus controls, and with the same relative methylation change -hypomethylation or hypermethylation.

Differential methylation in regions

We searched for regions in the MethylC-seq data in the HMM defined HMDs/PMDs regions. We studied intervals from 125bp up to 6000Kb with a difference over 5% comparing between ASD-P and TD-P and selected the probes with a t-test p.value<0.01. We searched if the 700DMCpGs or regions found in the previous blood study were also found in these regions.

Validation of common alterations found in relevant genes

Genes located on/nearby the 700DMCpGs were searched in placental PMDs/HMDs and neuronal PMDs/HMDs defined in a previous study. Characterization of the variants was performed. The genes were assessed by Genomic PCR and Sanger Sequencing and by pyrosequencing in order to find sequence or epigenetic variants. Specific protocols are defined in a publication in submission[31] and in the thesis CHAPTER 1.

Sequenom Genotyping

We performed Sequenom Genotyping to determine if sequence variants found were statistically more prevalent in ASD compared to controls. We only selected those with a MAF<0.02 (1000Genomes, Exome variant server and within all the control cohorts available). For these we analyzed additional 394 ASD (Spanish and ECACC collection) and 500 Spanish control samples. We used 250ng of genomic DNA and Sequenom Genotyping was processed at Universitat de Valencia (Valencia, Spain). Briefly, locus-specific PCR reaction is performed, and after a single nucleotide primer extension is performed over the polymorphic nucleotide. Using Matrix Assisted Laser Desorption Ionization Time-of-Flight (MALDI-TOF mass spectrometry), a peak ratio analysis identifies the SNP allele. Statistical association by Fisher exact tests and False Discovery Rate (FDR) multiple-testing correction (qvalue R function[38]) were performed.

RESULTS

In Figure 1 we summarize the approaches followed to compare blood and placental methylation data for ASD patients and healthy individuals (Figure 1).

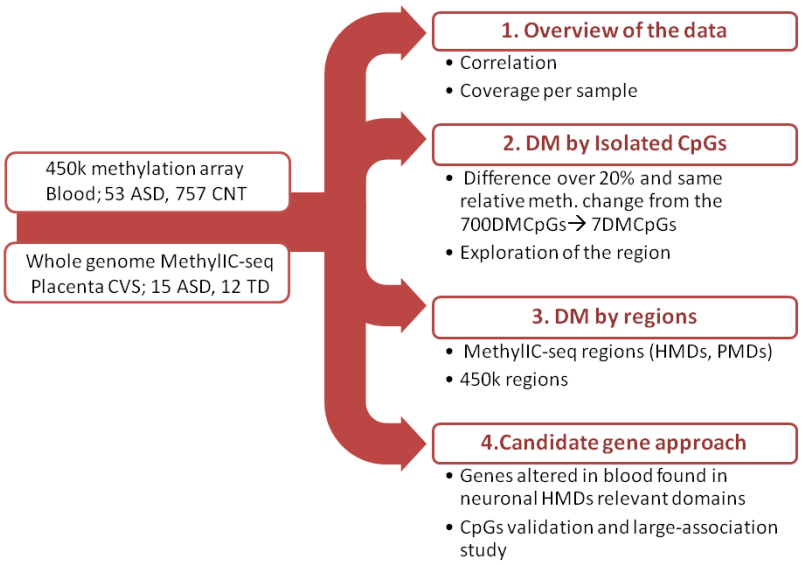


Figure 1. Approaches followed to compare the data obtained for ASD patients and healthy individuals of blood and placental tissue, by array 450k and whole genome bisulfite sequencing (MethylC-seq) respectively.

As a first exploration we compared the data for the same positions in placenta by 450k array and MethylC-seq. The selected positions studied (n=688 DMCpGs) were found altered in ASD in a case-control study in blood (Homs A. et al, submitted[31]). Correlation of control samples, analyzed as a pool, between the two techniques showed a low correlation coefficient (R^2 around 0.5) (Figure 2). This low correlation could be due either to the different methodologies used or due to the intrinsic differences of the tissues. A previous study assessed also the correlation for the same array and sequencing technologies, both for placental tissue samples, and obtained a slightly higher correlation coefficient (R^2 around 0.7)[28], pointing out that the differences we observed were probably produced by the technique. We detected that the MethylC-seq method had a main coverage depth of 4 reads for the CpGs analyzed, which could set a study limitation (Figure **3**).

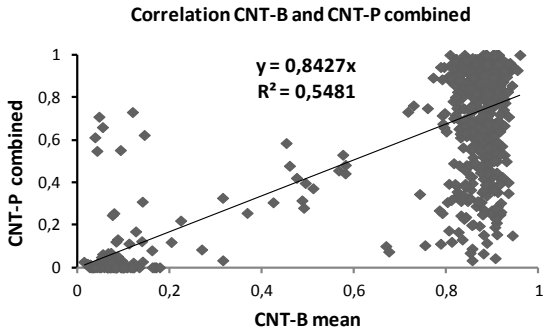


Figure 2. Correlation of methylation values for CNT-B mean and CNT-P with combined reads.

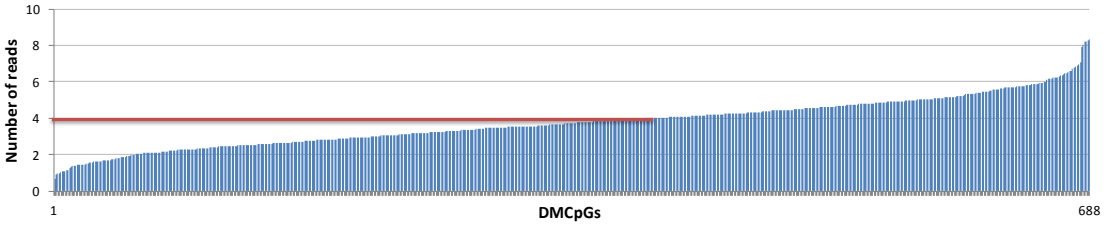


Figure 3. Approximate mean reads of selected CpG per sample. We calculated the mean reads per sample from the total number of reads by group (ASD-P or TD-P) divided by the total number of individuals.

Common isolated DMCpGs found differentially methylated in blood and placental tissue

For the same subset of CpGs altered in ASD in a previous study, we searched for common alterations in placental tissue. In order to solve the coverage limitation, we combined all ASD-P values and TD-P values and made a pool of reads. We obtained 7 CpGs altered in the same direction in both tissues (Table 1, Figure **4**).

Target_ID	Gene	Position	ΔM Placenta	ASD-P	TD-P	ASD-P	TD-P
cg22949832	*PRDM16*	chr1:3047342	0.222	2	1	26	17
cg24127050	*SEMA4C*	chr2:97530800	0.227	4	5	50	64
cg07367602	intergenic	chr7:115963592	0.29	2	2	29	26
cg17580798	MESTIT1;MEST	chr7:130132199	0.272	5	6	66	66
cg14078687	*CEL*	chr9:135943307	0.211	4	5	51	56
cg20169197	*B4GALNT4*	chr11:376100	0.219	4	3	47	39
cg22961727	*KRTAP21-1*	chr21:32128708	0.229	2	3	22	41

Table 1. Common CpGs obtained from the differential methylation in blood and placenta and mean sample coverage and combined sample coverage. The combined coverage (reads) is obtained combining all the reads of the individuals.

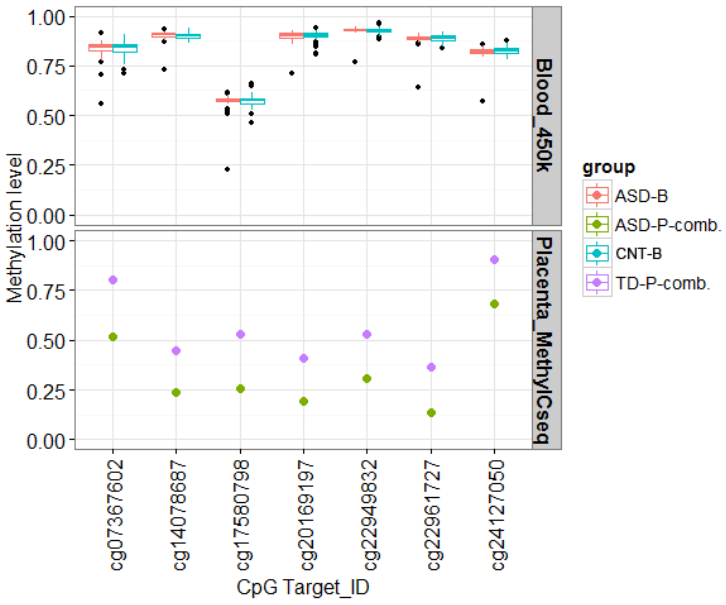


Figure 4. Common CpGs obtained from the differential methylation in blood and placenta (450k and MethylC-seq) showing a minimum of 20% change. For 450k technique all the individual values are plot, whereas for the MethylC-seq we used the value for the pool of individuals (combined).

However, when we studied the methylation values of the specific positions at individual placental samples we identified a high variability between them (Figure 5). We also observed in each sample a high proportion of extreme values of methylation (60%), which are rarely seen in placental tissue (Figure 6).

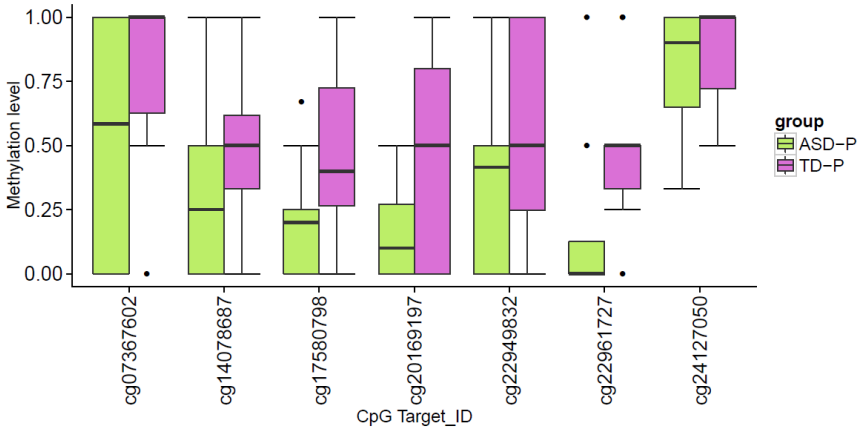


Figure 5. Individual values for the placenta samples (ASD-P and CNT-P) for the 7DMCpGs altered. We see a lot of variability in their methylation level.

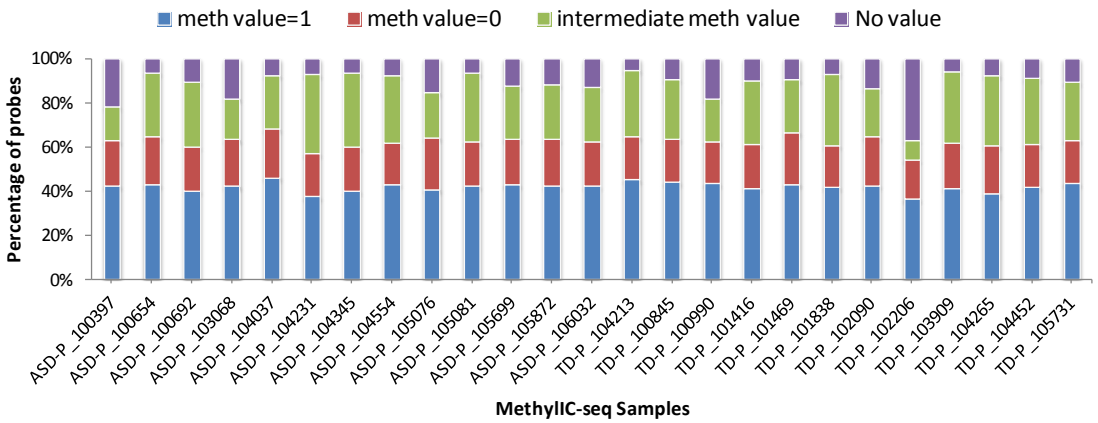


Figure 6. Distribution of the methylation value for 688CpGs for the samples (ASD-P and TD-P) sequenced by MethylC-seq. The samples show extreme values (equal to 0 or 1) in 60% of the CpGs (around 300DMCpGs) and empty values around 10% of the CpGs. Two ASD-P were excluded for this analysis because showed more than 80% of empty values.

Differentially methylated regions in placental tissue and blood

Given that isolated CpGs display a lot of variability, another approach followed was to search for regions in patients altered in both tissues. We searched for differentially methylated regions (DMRs) (over 5% of methylation difference) in placental samples between ASD-P and TD-P. We identified a total of 125 DMRs, having a maximum methylation difference of 10.1%. After selection by p.value<0.01, there remained 5DM-PMDs and 13DM-HMDs (Table **2**, Figure **7**). Regarding the function of the genes located in these domains we observed that genes are involved in neuroactive ligand receptors interactions (*GCGR* and *NPFFR2*), transcription factors (*GTF2H2B*) and regulation of transcription (*SLC27A1*). Moreover, some genes are associated with ASD by CNVs (SCL27A1, *GTF2H2B, NDRG4, ZNF653*), mutations (*NPFFR2),* expression studies (PRRT2), and by linkage *(OBSCN, LCN10).*

Class	Chr coordinates (hg19)	Kb	Genes	Mean TD-B	Mean ASDP	ΔM	P.val T-test
PMD1	chr17:80292395-80293464	1,1		0,397	0,463	-0,066	0,0070
PMD2	chr17:79769554-79770468	0,9	*GCGR*	0,394	0,471	-0,076	0,0094
PMD3	chr16:58535597-58537422	1,8	*NDRG4*	0,552	0,624	-0,073	0,0018
PMD4	chr14:95236570-95237621	1,1		0,263	0,315	-0,052	0,0071
PMD5	chr9:139640886-139642893	2,0	*LCN10, LCN6*	0,466	0,557	-0,091	0,0071
PMD6	chr8:986273-989166	2,9	*ERICH1-AS1*	0,672	0,621	0,050	0,0098
PMD7	chr5:158532503-158533556	1,1	*AK123543*	0,307	0,401	-0,094	0,0002
PMD8	chr5:69711309-69783245	72	*SMA4, GTF2H2B*	0,842	0,898	-0,056	0,0042
PMD9	chr19:35323953-35329722	5,8		0,287	0,340	-0,052	0,0041
PMD10	chr19:11593644-11594371	0,7	*ZNF653*	0,263	0,333	-0,070	0,0015

PMD11	chr19:3583313-3585356	2,0		0,488	0,565	-0,077	0,0055
PMD12	chr18:12092177-12093606	1,4		0,457	0,363	0,094	0,0026
PMD13	chr1:228559247-228561425	2,2	*OBSCN*	0,378	0,446	-0,068	0,0070
HMD1	chr19:17608407-17611344	2,9	*SLC27A1*	0,542	0,616	-0,074	0,0014
HMD2	chr16:29823907-29826661	2,8	*PRRT2, BOLA2*	0,631	0,693	-0,061	0,0063
HMD3	chr4:72923008-73015787	92	*NPFFR2*	0,738	0,668	0,070	0,0001
HMD4	chr2:175192124-175193397	1,3	*LOC285044*	0,365	0,458	-0,093	0,0073

Table 2. Regions found significantly altered between ASD-P and TD-P by MethylC-seq (p.value<0.01). Mean values for ASD and TD for the region and methylation difference (ΔM) between them.

a)

b)

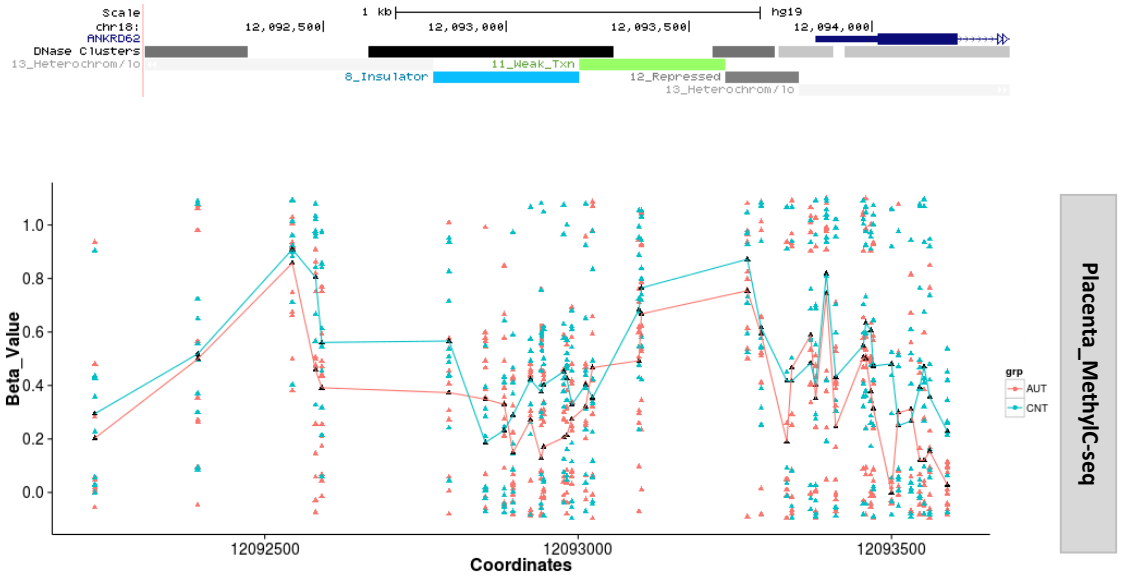


Figure 7. Region and methylation values the region PMD10, which was found significantly differentially methylated between ASD-P and TD-P (CNT in the figure). We show in a) a detailed region, with a gene downstream the region (*ANKRD62*), which show open chromatin domains (DNAse clusters) and also several insulator, weak transcription and heterocromatin domains. In b) the representation of the methylation values for the detected region for ASD-P and TD-P.

The differences observed in placenta were subtle, reaching a maximum 9.4% of difference compared to the 20% of methylation difference in the blood study. Therefore, we used COMB-P[39] algorithm to detect subtle changes in methylation in blood. We identified 6 new regions, although none of these regions were found altered in placenta (Table 3).

Band	Start	End	Genes	NºCpG	COMB-P pval
chr11p15.4	2890189	2891495	*KCNQ1DN*	27	0.0075
chr16p13.3	2652695	2653987	*PDKP1*	14	0.0043
chr17p13.1	7832478	7833663	*KCNAB3*	11	0.00018
chr3p21.2	51740740	51741473	*GRM2*	7	0.0016
chr5q14.1	78985424	78986160	*CMYA5*	12	5,63E-02
chr6p22.1	30038253	30040291	*RNF39*	44	0.0074

Table 3. Regions found altered between ASD-B and CNT-B by 450k array. The number of CpGs altered in the region, the positions and the resulting p.values are indicated. We selected the regions by nominal p.value, only one CpG altered had a q.value< 0.05.

Altered blood genes and developmental methylation domains

In order to see if genes within the DMCpGs found in ASD blood were involved in neurodevelopment in placental or brain tissue, we searched if those genes were found in the methylation domains (Neuronal-HMDs/PMDs (N-HMDs or N-PMDs) and placental-HMDs/PMDs (P-PMDs/P-HMDs)), described previously[28]. From the 467 genes altered we detected 68 genes (14.55%) in one of the methylation domains. Although the genes were located in N-HMDs domains in a higher proportion, we did not observe a significantly enrichment in any domain (Table **4**).

	P-HMDs	P-PMDs	N-HMDs	N-PMDs	Total genes
Schroeder et al.[28] (PMDs/HMDs genes)	507	735	1170	435	20500
450k (467 genes)	5	22	34	7	467
χ2	0.07	0.24	0.17	0.4	

Table 4. Statistical analyses for each domain (neuronal and placental PMDs and HMDs). We calculated the enrichment of the total number of genes in a domain versus all the genes in the genome. Similarly, we calculated the total number of differentially methylated genes in the blood 450k versus all genes differentially methylated in the same assay. We calculates chi square test, which was significant for the N-HMDs.

Regarding the 34 genes found in the N-HMDs, which are important for neurodevelopment, we studied 9 genes (10DMCpGs) selected by their gene function (Table **5**). Validation showed that 80% of the CpGs analyzed were rare inherited sequence variants, although they did not contain any common polymorphism described (dbSNP v138) and were not found having similar values in 656 controls (GEO dataset). Two genes showed no single nucleotide variant

(SNVs), although bisulfite pyrosequencing did not validate the epigenetic alteration, becoming false positive results.

Target_ID	Sample	Location (hg19)	Gene	Validation	B-val ASD	B-val CNT	Function
cg12315391	ASD_3	3:157815145	*SHOX2*	as reference	0,823	0,823	Specifying neural systems
cg10315366	ASD_9	10:134911047	*GPR123*	as reference	0,399	0,467	Neurological signaling
cg08921491	ASD_1	15:88577641	*NTRK3*	HTZ C>T; M	0,467	0,437	Neuron fait
cg18865080	ASD_9	11:21263031	*NELL1*	HTZC>T; M	0,417	0,405	Cranial bone and neural tissues dev.
cg01661235	ASD_4	16:23988974	*PRKCB*	HTZ C>T;M	0,437	0,491	PKC-signalling
cg18454510	ASD_40	6:40363184	*LRFN2*	HTZC>T; M	0,405	0,398	Neurite outgrowth
cg00155846	ASD_39	9:138011566	*OLFM1*	HTZ G>A; P	0,411	0,823	Neural crests dev.
cg07137277	ASD_4	10:1252782	*ADARB2*	HTZ G>A; M	0,491	0,399	RNA editing
cg02955989	ASD_37	19:30944506	*ZNF536*	HTZ C>T; M	0,398	0,467	TF. Assoc. to ASD
cg17609948	ASD_28	10:1668625	*ADARB2*	HTZ C>T; P	0,395	0,411	RNA editing

Table 5. Candidate genes within PMDs and Neuronal-HMDs regions. Genomic DNA PCR validation. Most of them were DMCpGs showing single nucleotide variant (SNVs). M, Maternal inherited; P, Paternal inherited; TF, transcription factor.

From these 9 variants, we studied the probes which presented a low MAF frequency. We analyzed 500 Spanish controls and 350 patients by Sequenom Genotyping, which added to the individuals analyzed by 450k arrays. We did not obtain a significant association for the meSNVs analyzed, which could be due to the fact that are rare variants (the maximum number for the controls altered is two) and the cohort could not be enough to prove their association (Table **6¡Error! No se encuentra el origen de la referencia.**)

Target_ID	Nº ASD altered/total	Nº CNT altered/total	MAF ASD-CNT	Fisher p.val; OR(CImin-CImax)
cg08921491	1/446	2/1349	0.0011-0.0007	0.576; 1.51(0.03-29.1)
cg01661235	1/446	1/1349	0.0011-0.0004	0.435; 3.03(0.04-237.24)
cg18454510	1/445	1/1329	0.0011-0.0004	0.439; 2.99(0.04-234.25)
cg07137277	1/445	0/1347	0.0011-0	0.248; Inf(0.08-Inf)
cg02955989	2/445	1/1349	0.0022-0.0004	0.154; 6.07(0.32-357.66)

Table 6. Large-scale association study results for selected 5DMCpG harbouring meSNVs. None of the DMCpGs obtained a significant p.value, although minor allele frequency indicates a low frequency in both groups. MAF, Minor Allele Frequency; OR, Odds Ratio; CI, Confidence Interval.

DISCUSSION

Most of the studies in methylation using peripheral blood and placental tissues search for differential methylation, most of them looking for differences useful for prenatal non-invasive tests[19,40–44]. We instead searched for convergent methylation in tissues given a disease condition, ASD. We performed differential methylation between patients and healthy individuals in both tissues, to determine if convergent alterations could be found altered in initial developmental stages, due to extrinsic environmental factors (*in utero* environment) which may promote aberrations in the early developmental stages[17,23,45]. The control individuals used in placenta were typically developing healthy siblings, as they might have shared more similar *in utero* environment.

The placenta data was obtained by next generation sequencing (Whole genome bisulfite sequencing (WGBS) or MethylIC-seq). We first found that the correlation among blood and placenta was very low (R^2=0.5) and that these differences can be tissue-specific or intrinsic to the technique comparisons. However, samples from the same tissue obtained using the same technologies, and treated with smoothing algorithms, displayed important but not high correlation. (R^2=0.7)[28], reinforcing that the technique may limit the data concordance.
We also observed that the data displayed had not enough coverage to establish individual site-specific methylation percentages, so we worked with data obtained from pools of individuals by groups. Therefore, this low coverage, and the possible technical bias of the technique[46] can influence to the interpretation of the data obtained. Results may be further compared when data with higher coverage is available. Despite the coverage limitation, we performed comparisons between placenta and blood datasets focusing on the previously identified 700 DMCpGs detected in ASD blood (previous ASD-control study[31]) and 7DMCpGs displayed the same methylation change in both tissues. However the high variability can lead to erroneous results and comparisons should be made with new data with better coverage.

Besides, we also identified multiple DMRs (4 PMDs and 13 HMDs) between patients and controls in placenta. Comparing the results obtained, we did not see correlation with blood regions (DMRs previously reported in the ASD blood study and regions showing lower differences obtained with other algorithms). This low amount of DMRs obtained could be due to the fact that the regions are normally patient-specific and that it is difficult to find

convergent DMRs among all the patients in the study, due to the heterogeneity for ASD aetiology, as seen in most of the ASD methylation studies, which scarcely find commonalities. Moreover, it is difficult to detect DMRs in large domains with subtle methylation alterations, such as 5-10%. In spite of that, the regions found altered in placenta displayed genes involved the disease.

On the other hand, as regions located in P-HMDs play a role in development, and more interestingly, regions in N-HMDs in neurodevelopment[28]; we searched if the previously identified altered genes in blood (within the 700DMCpGs) where located in these domains. Specifically, in neurons, ASD candidate genes were enriched within N-HMDs and N-PMDs in SH-SY5Y, for example, a cluster of cadherin genes in a PMD strongly associated to ASD[47]. From the studied genes, we found that 14.1% of our genes located in PMDs and HMDs in neurons and placenta without significant association. From them, we selected 9 genes (10DMCpGs) located in these regions based on their function. Validation showed that the vast majority were rare sequence variants affecting methylation targets (meSNVs) inherited from a healthy progenitor, and two were false positive methylation aberrations. It has been reported that genetics exert a large effect on methylation, for example a study showed that sequence variants produced methylation changes in about two thirds of the sites studied[48]. These variants could affect allelic methylation and expression with a larger range effect[49]. For 5CpGs out of this 10CpGs, we genotyped the sequence variants in a larger cohort of patients (n=394) and controls (n=500) to see if the variants associated to the phenotype. Unfortunately, we could not discard the association given that up to 2 controls had the variant. Therefore, we would need a larger cohort to increase the statistical power and determine their association to the disease.

Finally, the comparisons among tissues leading to convergent mechanisms are not easy, as there is interlocus, intertissue and interindividual epigenetic variation and each tissue has its epigenetic signature reflecting their functions. This complexity was reflected in a study assessing methylation comparisons between placenta, coord blood and saliva, which concluded that the tissues were not comparable nor interchangeable to assess methylation in infancy[50]. Current limitations are also in the methodologies, as WGBS needs extensive sequencing to obtain enough coverage to establish methylation. In fact, estimated coverage for assessing methylation is still not defined in guidelines[51]. And for example, recent

methylome data obtained from the Epigenomics Roadmap display a coverage of 3 reads minimum for WGBS[52]. It is a fact that BS-DNA needs more sequencing than the genomic DNA to obtain a good coverage. First, due to the inherent complexity of the BS-genome as the unmethylated cytosine is converted to thymine increasing the library complexity and mapping complexity. Additionally, BS-converted DNA is no longer complementary, so in the process to merging reads, lots of reads are unmapped, and as consequence more sequencing is needed to obtain a reliable percentage of methylation. This extensive sequencing raises the cost to prohibitive for large genomes[53]. However, efforts to this direction are being performed with techniques that do not need DNA treatment nor amplification as Single molecule real time sequencing (SMRT) and nanopore[54].

In summary, we performed a first exploration comparing case-control differential methylation data of blood versus placental tissues, using site-specific, regions (DMRs) and gene comparisons. Despite finding some aberrations common to both tissues, we could not assure their truly implication. For that, we would need further studies assessing larger coverage for the sequencing samples. Nevertheless, we have done a first exploration of the placental tissue which can give clues of the alterations that could contribute in early development to the ASD aetiology. Finally, we found genes located in N-HMDs, having a role in neurodevelopment. Analyzing a subset of these genes we found that they showed inherited sequence alterations in the CpGs (meSNVs), being some of them absent in a large cohort of controls. Increased sample-size genotyping would be needed to establish their final association with ASD phenotype. Development of techniques in cost efficiency and effective way, capable to obtain data with enough coverage to establish genome-wide methylation percentage, are strongly needed for the advancement of the epigenetic field.

REFERENCES

1 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5)*. 2013 doi:10.1176/appi.books.9780890425596.744053.

2 Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T *et al.* Genetic Heritability and Shared Environmental Factors Among Twin Pairs With Autism. Arch. Gen. Psychiatry. 2011; **68**: 1095–1102.

3 Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet* 2014; **15**: 133–41.

4 Gaugler T, Klei L, Sanders SJ, Bodea C a, Goldberg AP, Lee AB *et al.* Most genetic risk for autism resides with common variation. *Nat Genet* 2014; **46**: 881–885.

5 Stein JL, Parikshak NN, Geschwind DH. Rare Inherited Variation in Autism: Beginning to See the Forest and a Few Trees. Neuron. 2013; **77**: 209–211.

6 Ronald A, Hoekstra RA. Autism spectrum disorders and autistic traits: A decade of new twin studies. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. 2011; **156**: 255–274.

7 Miyake K, Yang C, Minakuchi Y, Ohori K, Soutome M, Hirasawa T *et al.* Comparison of Genomic and Epigenomic Expression in Monozygotic Twins Discordant for Rett Syndrome. *PLoS One* 2013; **8**: e66729.

8 Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T *et al.* Strong association of de novo copy number mutations with autism. *Science (80-)* 2007; **316**: 445–449.

9 Strömland K, Nordin V, Miller M, Akerström B, Gillberg C. Autism in thalidomide embryopathy: a population study. *Dev Med Child Neurol* 1994; **36**: 351–6.

10 Harden CL. In utero valproate exposure and autism: long suspected, finally proven. *Epilepsy Curr* 2013; **13**: 282–4.

11 Gardener H, Spiegelman D, Buka SL. Prenatal risk factors for autism: Comprehensive meta-analysis. Br. J. Psychiatry. 2009; **195**: 7–14.

12 Haglund NGS, Källén KBM. Risk factors for autism and Asperger syndrome. Perinatal factors and migration. *Autism* 2011; **15**: 163–183.

13 St-Hilaire S, Ezike VO, Stryhn H, Thomas MA. An ecological study on childhood autism. *Int J Health Geogr* 2012; **11**. doi:10.1186/1476-072X-11-44.

14 Guinchat V, Thorsen P, Laurent C, Cans C, Bodeau N, Cohen D. Pre-, peri- and neonatal risk factors for autism. *Acta Obstet Gynecol Scand* 2012; **91**: 287–300.

15 LaSalle JM. Epigenomic strategies at the interface of genetic and environmental risk factors for autism. *J Hum Genet* 2013; **58**: 396–401.

16 Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. *Nat Rev Genet* 2007; **8**: 253–62.

17 Tordjman S, Somogyi E, Coulon N, Kermarrec S, Cohen D, Bronsard G *et al.* Gene Environment Interactions in Autism Spectrum Disorders: Role of Epigenetic Mechanisms. *Front Psychiatry* 2014; **5**: 53.

18 Marsit CJ. Influence of environmental exposure on human epigenetic regulation. *J Exp Biol* 2015; **218**: 71–9.

19 Novakovic B, Yuen RK, Gordon L, Penaherrera MS, Sharkey A, Moffett A *et al.* Evidence for widespread changes in promoter methylation profile in human placenta in response to increasing gestational age and environmental/stochastic factors. *BMC Genomics* 2011; **12**: 529.

20 Dolinoy DC, Weidman JR, Jirtle RL. Epigenetic gene regulation: Linking early developmental environment to adult disease. Reprod. Toxicol. 2007; **23**: 297–307.

21 Kile ML, Baccarelli A, Hoffman E, Tarantini L, Quamruzzaman Q, Rahman M *et al.* Prenatal Arsenic Exposure and DNA Methylation in Maternal and Umbilical Cord Blood Leukocytes. *Env Heal Perspect* 2012.http://dx.doi.org/10.1289/ehp.1104173.

22 Intarasunanont P, Navasumrit P, Waraprasit S, Chaisatra K, Suk W a, Mahidol C *et al.* Effects of arsenic exposure on DNA methylation in cord blood samples from newborn babies and in a human lymphoblast cell line. *Environ Health* 2012; **11**: 31.

23 Koestler DC, Avissar-Whiting M, Andres Houseman E, Karagas MR, Marsit CJ. Differential DNA methylation in umbilical cord blood of infants exposed to low levels of arsenic in utero. *Environ Health Perspect* 2013; **121**: 971–977.

24 Mitchell MM, Woods R, Chi LH, Schmidt RJ, Pessah IN, Kostyniak PJ *et al.* Levels of select PCB and PBDE congeners in human postmortem brain reveal possible environmental involvement in 15q11-q13 duplication autism spectrum disorder. *Environ Mol Mutagen* 2012; **53**: 589–598.

25 James SJ, Cutler P, Melnyk S, Jernigan S, Janak L, Gaylor DW *et al.* Metabolic biomarkers of increased oxidative stress and impaired methylation capacity in children with autism. *Am J Clin Nutr* 2004; **80**: 1611–1617.

26 Riccio A. Dynamic epigenetic regulation in neurons: enzymes, stimuli and signaling pathways. *Nat Neurosci* 2010; **13**: 1330–1337.

27 Maccani MA, Marsit CJ. Epigenetics in the placenta. *Am J Reprod Immunol* 2009; **62**: 78–89.

28 Schroeder DI, Blair JD, Lott P, Yu HOK, Hong D, Crary F *et al.* The human placenta methylome. *Proc Natl Acad Sci U S A* 2013; **110**: 6037–42.

29 Schroeder DI, Lasalle JM. How has the study of the human placenta aided our understanding of partially methylated genes? *Epigenomics* 2013; **5**: 645–54.

30 Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009; **462**: 315–22.

31 A. Homs, M. Codina, B. Rodríguez-Santiago, C. M. Villanueva, D. Monk IC and LAP-J. Genetic and epigenetic methylation defects and implication of the ERMN gene in autism spectrum disorders. *Prep* 2015.

32 Salas LA, Bustamante M, Gonzalez JR, Gracia-Lavedan E, Moreno V, Kogevinas M *et al.* DNA methylation levels and long-term trihalomethane exposure in drinking water: An epigenome-wide association study. *Epigenetics* 2015; **10**: 650–61.

33 Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A *et al.* Distinct DNA methylomes of newborns and centenarians. 2012. doi:10.1073/pnas.1120658109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1120658109.

34 Smyth G. limma: Linear Models for Microarray Data. In: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. 2005, pp 397–420.

35 Schroeder DI, Lott P, Korf I, LaSalle JM. Large-scale methylation domains mark a functional subset of neuronally expressed genes. *Genome Res* 2011; **21**: 1583–1591.

36 Chen P-Y, Cokus SJ, Pellegrini M. BS Seeker: precise mapping for bisulfite sequencing. *BMC Bioinformatics* 2010; **11**: 203.

37 Korf Lab, Genome Center, University of California D. StochHMM - A Flexible hidden Markov model application and C++ library. 2007.https://github.com/KorfLab/StochHMM.

38 Storey J. qvalue: Q-value estimation for false discovery rate control. 2015; **R package** .http://qvalue.princeton.edu/, http://github.com/jdstorey/qvalue.

39 Pedersen BS, Schwartz DA, Yang I V., Kechris KJ. Comb-p: Software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics* 2012; **28**: 2986–2988.

40 Papageorgiou EA, Fiegler H, Rakyan V, Beck S, Hulten M, Lamnissou K *et al.* Sites of differential DNA methylation between placenta and peripheral blood: molecular markers for noninvasive prenatal diagnosis of aneuploidies. *Am J Pathol* 2009; **174**: 1609–1618.

41 Xiang Y, Zhang J, Li Q, Zhou X, Wang T, Xu M *et al.* DNA methylome profiling of maternal peripheral blood and placentas reveal potential fetal DNA markers for non-invasive prenatal testing. *Mol Hum Reprod* 2014; **20**: 875–884.

42 Yuen RK, Peñaherrera MS, von Dadelszen P, McFadden DE, Robinson WP. DNA methylation profiling of human placentas reveals promoter hypomethylation of multiple genes in early-onset preeclampsia. *Eur J Hum Genet* 2010; **18**: 1006–1012.

43 Chavan-Gautam P, Sundrani D, Pisal H, Nimbargi V, Mehendale S, Joshi S. Gestation-dependent changes in human placental global DNA methylation levels. *Mol Reprod Dev* 2011; **78**: 150.

44 Turan N, Katari S, Gerson LF, Chalian R, Foster MW, Gaughan JP *et al.* Inter- and intra-individual variation in allele-specific DNA methylation and gene expression in children conceived using assisted reproductive technology. *PLoS Genet* 2010; **6**: 1–14.

45 Bollati V, Baccarelli A. Environmental epigenetics. *Heredity (Edinb)* 2010. doi:hdy20102 [pii] 10.1038/hdy.2010.2.

46 Ji L, Sasaki T, Sun X, Ma P, Lewis ZA, Schmitz RJ. Methylated DNA is over-represented in whole-genome bisulfite sequencing data. *Front Genet* 2014; **5**: 341.

47 Schroeder DI, Lott P, Korf I, LaSalle JM. Large-scale methylation domains mark a functional subset of neuronally expressed genes. *Genome Res* 2011; **21**: 1583–1591.

48 Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J *et al.* DNA methylation contributes to natural human variation. *Genome Res* 2013; **23**: 1363–72.

49 Tycko B. Allele-specific DNA methylation: beyond imprinting. *Hum Mol Genet* 2010; **19**: R210–20.

50 Armstrong DA, Lesseur C, Conradt E, Lester BM, Marsit CJ. Global and gene-specific DNA methylation across multiple tissues in early infancy: Implications for children's health research. *FASEB J* 2014; **28**: 2088–2097.

51 Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 2014; **15**: 121–32.

52 Consortium RE, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 2015; **518**: 317–330.

53 Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat Protoc* 2015; **10**: 475–83.

54 Plongthongkum N, Diep DH, Zhang K. Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nat Rev Genet* 2014; **15**: 647–61.

DISCUSSION

DISCUSSION

Many efforts have been taken during the last decade to unravel the complex causes of ASD, which affects about 1% of children. Currently, the aetiology of ASD can be explained in about one third of cases[64], although each the causes found so far accounts for a small proportion of cases.

It is estimated that thousand of genes, with monogenic, polygenic or complex inheritance models, are involved in ASD pathogenesis[93,106]. Recent evidences also support that epigenetic mechanisms, intermediary between environmental effects, provide additional contribution to the disorder. The epigenetic contribution to ASD is supported by the known implication of genes regulating epigenetics[128,270–272] and by the documented environmental effects during early life stages[126,129,135]. Specifically, epigenetic alterations associated to ASD have been found in known genes and pathways involved in neurodevelopment[263,283,285,287], and novel susceptibility genes have been defined, pending confirmation in larger cohorts[262,288,289,291,294,346].

Our contribution to the field was centered in the characterization of the blood DNA methylation profile of ASD patients from different genotype subgroups (with known ASD-related chromosomal rearrangements) or with unknown causes (idiopathic), to find convergent and divergent genes, mechanisms and pathways involved. We depicted the effect of epigenetic aberrations in expression and studied commonalities with other tissues, providing an integrated representation of the mechanisms leading to the disease. Given that no biomarker has been translated into the clinical practice, and that the early diagnostic of the cases significantly improves their life quality, the better characterization of the layers and tissues affected for ASD patients brings valuable knowledge to the field. Moreover, the benefits of unraveling the aetiology will translate in the future to earlier diagnosis, better prognosis, personalized treatments and adequate health and support services for the patients.

Epigenetic landscape in ASD

We first characterized the genome-wide epigenetic landscape in ASD patients' DNA samples from blood. Although ASD is a neurodevelopmental disorder and some alterations might be

found specifically in brain, postmortem brain as well as other tissues as blood, are widely used to assess ASD epigenetic studies. Regarding peripheral blood, many studies proved its utility to unravel ASD epigenetic alterations common to many tissues[262,291,347].

Global methylation landscape

The DNA methylation landscape, constrained to the 2% of CpGs analyzed by the 450k array platform showed that the idiopathic ASD individuals did not display a systemic methylation alteration compared to controls, as previously shown in other studies in blood and brain[262,290]. It also showed that ASD patients did not present common aberrations (in CpGs or regions), and despite the small cohort studied, it pointed out the low-recurrent nature of the ASD epigenetic events. This low prevalence of the epimutations is also reinforced by the low convergence among the published results for methylation case-control studies in ASD.

Reinforcing the results in idiopathic ASD patients, no convergent alterations were found shared by all the subgroups with multiple ASD genotypes (comorbid conditions as Down Syndrome (DS), and recurrent rearrangements as chr15q11-13 duplication (Dup15) and 1-pter deletion (Del1pter)) compared to controls. Their global methylation distribution, understood by the proportion of a certain methylation state, did not display differences in any group and no global alterations were detected in the affected genetic regions of the subgroups (chr21, chr1p36, chr15q11-13). However, when assessing methylation by subgroups compared to controls, few alterations in the DS group aroused. The trisomy21 has been studied genome-wide and individuals did not display global[348] but local genome-wide distributed alterations[348–353]. The probes and genes altered in the patients were mainly represented in DS previous studies, but also in ASD studies. Among the other subgroups studied, idiopathic patients conceived by assisted reproductive technology showed no global differences, although some studies described genome-wide alterations in methylation as consequence of its use[208].

Specific alterations found in regions in ASD

The low prevalence of the epimutations made the changes undetectable analyzing groups, therefore individual-specific approaches have been proved useful in ASD to depict individual or shared (by a small number of patients) alterations. Through the study of individual-specific methylation of ASD idiopathic group, a small number of patient-specific regions (DMRs) aroused; considered relevant findings as a region is more prone to alter gene regulation and

function[354]. The regions displayed (encompassing *SMG7, PIK3CD, PHACTR1, ZCCHC9* and *RAB26*) showed mainly a relative hypermethylation and cis-acting expression effects in four of the regions, associated to both over and down-regulation. Regarding the inverse correlation found (higher methylation and overexpression), it is being accepted that expression correlation is evolving from a traditional perspective towards of a gene dosage matter, and also that methylation in different structural genomic elements, such as gene bodies, where hypermethylation is positively correlated with expression[242,262].

The methylation aberrations found were rare and inherited from healthy progenitors. Their intergenerational transmission might be due to evasion of post-zygotic reprogramming[355,356] through different mechanisms such as nucleosomes retention or epigenetic modifiers dosage/impairment[357], or also due to cis-acting mutations effects[219].

Several DMRs were studied in a larger ASD patient's cohort but we did not find a second individual harbouring the epigenetic alteration, again evidencing the low recurrence of the events.

Regarding DMRs altered in the subgroups, we found regions common to multiple patients in DS and Dup15 groups. All DS DMRs genes were previously found with methylation aberrations, for example *DCC,* which interacts with Down's syndrome Cell Adhesion Molecule (*DSCAM)* and plays a role in axonogenesis[358]. Regarding the Dup15 patients, 2 out of 3 regions were located in the rearranged region (*NDN* and *SNRPN*) and displayed an expected pattern (hypermethylation expected from a maternal duplication of the region). Also, Dup15 presented a DMR involving a voltage-gated potassium channel (*KCNQ1)* located in a imprinting region of chromosome 11p15, a structurally-similar region to the imprinting control region in 15q. *KCNQ1* plays a role in cardiac repolarization and is associated with multiple pathologies, including the imprinting disorder Beckwith-Wiedemann syndrome (BWS). Apart from these altered regions, we also found group-specific alterations, which had a 20% of methylation difference between the group and the rest of the groups, in Del1p-ter and ART groups. Del1p-ter group showed two genes located in the chr1 located far from the altered region, one of them involved in embryonic development (*ETNK2),* and also genes involved in transcriptional processes (NRIP2, LOC126295 and *BXDC1*); whereas for ART patients, genes were involved in cardiac development (*HOP* and *TNNI3*) and cell proliferation and apoptosis (*DUSP22*, *PSCA and UBL5*).

Specific alterations found in isolated CpGs in ASD

The study of individual-specific aberrations for idiopathic ASD patients gave an average of 13DMCpGs altered for each individual (700DMCpGs for 53 patients in total). In this case, most of them showed a relative hypomethylation, and also cis-acting gene expression effects, mostly with a negative correlation. The hypomethylation was reported in previous reports showing global hypomethylation in cerebellum and blood of ASD patients[146,263], although strangely all clustered regions showed hypermethylation in ASD cases.

Regarding the isolated CpGs studied (selected if present in more than one individual or had associated expression), they consisted of rare inherited genetic variants affecting methylation targets (meSNVs). This large amount of meSNVs detection, could be due to the filters applied in the differential methylation analysis, as we searched for large effects (>18%) and prioritized methylation-status change. These selection parameters were used in some studies[27], although ASD studies analyze smaller differences (4-7%[262,288,295]) or even rank the top differentiated genes without filtering[262].

From the obtained rare meSNVs, six showed association with ASD through a large case-control association study, including more than 1300 controls; and for additional seven meSNVs no control were found altered, so the association remain to be studied. The proved associated genes involved axonogenesis (*ERMN*), apoptosis (*USP24),* signal transduction *(PDE10A),* a methyltransferase which regulates neuroinflamation chaperones (*METTL21C),* a gene which takes part in BCKD complex, associated to ASD by mutations *(DBT)*, and with nominal association, a synaptic vesicle transport gene (*STX16*).

Rare DMCpGs/meSNVs confirmation as ASD susceptible variants needs additional studies in larger datasets.

Interaction of genetics with epigenetics

It is known that genetics has large effect on methylation, it has even been quantified to be around 70% in array platforms studying three different ethnicities. In our study, we found this percentage even higher (85%); however, as we said, might be due to the selection filters applied. In any case, the meSNVs or meSNPs (Minor allele frequency over 1%) have an effect

on methylation and also can exert an effect on the cis-CpGs around, affecting regulatory regions and impair the methyl-binding-proteins and transcription factors binding[360]. Therefore, both genetics and epigenetics are implicated in the allele specific methylation (ASM) and expression (ASE) in *cis* or even in *trans*[220,221]. So the interplay between genetics-epigenetics-transcription is a feature described and also found in this study.

Inheritance model followed

The rare epigenetic and genetic alterations found would have an additive contribution to other underlying aberrations causing ASD, following a polygenic model. This is explained by the fact that all the alterations were inherited from unaffected parents.

Besides, the DMCpGs/DMRs found in DS and Dup15 might be causative for their phenotype. However, for the DS group many genes are found also in individuals with exclusively DS phenotype, remaining to be further compared to larger DS with comorbid ASD.

Implicated a new susceptibility gene in ASD: *ERMN*

The meSNV most significantly altered compared to controls (cg05777410, chr2q24.1) was located over *GALNT5* and correlated with overexpression of its nearby gene, *ERMN,* in one of the two patients. We also conducted a study of sequence mutations (missense and deleterious mutations) in cases and controls (n>1,000 and n>60,000, respectively) and found that *ERMN* carried a higher mutation load in ASD individuals. The gene encodes for ermin protein, an oligodendroglia-specific cytoskeletal protein, involved in myelinisation and maintenance of neuronal plasticity, both important process in mood disorders. *ERMN* is located in the boundaries of *AUTS5* locus, known to be associated to ASD and ID. Deletions of the region were also associated to developmental delay and impaired communication[47], and downregulation was observed in patients with epileptic seizures[362]. Therefore, the study brings enough arguments to associate *ERMN* as an ASD susceptibility gene which can be altered by both rare point meSNVs, missense or LoF mutations.

Testable epigenetic biomarkers in ASD

Blood tissue has the advantage that marks found can be easily translated to testable epigenetic biomarkers. Until now only one epigenetic biomarker, *ENO2,* encoding a neuron-specific enzyme, identifying 14% of ASD patients through its hypermethylation has been described[291]. Other biomarkers assessing risk and prognosis by imaging techniques, blood expression profiles and hormone quantification have been found, although none of them has been translated into clinics. For this aim, we obtained a classification tool using supervised analysis (LDA) and identified group-specific methylation pattern that classified four groups (DS, Dup15, Del1pter and ART) (56.15% of the studied individuals can be classified with the tool) from idiopathic patients and controls. We had not availability of an equivalent sample to perform cross-validation, fact that remains to be proved in the future.

Genes and pathways involved ASD by methylation analysis

Around one thousand genes and many different pathways, going from neuronal development, transcription and chromatin regulation to metabolic processes, have been involved to ASD[86,93]. However, some databases as AutismKb and SFARI databases conserve candidate core genes with solid evidences. In the idiopathic patients, we found methylation alterations in a subset of genes, which 5.8% of them were present in these core genes. The genes included syndromic genes (*YWHAE*), genes with mutations (*NRXN2, APBA2, CACNA1H, SLC9A9, RBMS3,* and *AFF2*), recurrent CNVs (*TUBGCP5*, and *CYFIP1*), and others discovered with association or linkage studies. From the genes identified altered in the idiopathic patients methylation study some were also associated by some evidence to ASD (although not being in the core gene list). Important ASD associations were by expression deregulation (n=5, *PHACTR1, LCP1, SERPINF1, PRCD, SETD1A*), mutations (*C15orf62/ DNAJC17*) or ASD-associated microdeletion syndromes (*RNF166*).

Regarding pathways and functions, studying all altered CpGs in the idiopathic patients, we highlight the statistically significant pathways such as axonogenesis, cell adhesion and

immune response signalling known to be involved in ASD. Also, functions included development and function of central nervous system, apoptosis, oxidative stress regulation, vesicle trafficking and nonsense-mediated mRNA decay (NMD) with a role in transcription/translation regulation, associated to ASD[363–366].

Regarding the genotype subgroups studied, we found significant enrichment in convergent pathways as embryonic development. Besides, studying the group-specific pathways, DS displayed genes previously associated to the trisomy21 and to ASD. Nervous system development, transcription factor activity and embryo development were described in trisomy21, although the first two also in ASD. And added to the previous associated ASD pathways, DS displayed enriched-pathways as axon guidance, neuroactive ligand-receptor interaction, focal adhesion and Wnt signalling pathways and calcium signaling[367,368]. Del1pter patients displayed also altered development and transcription processes, as well as oxidative stress regulation, and ART patients altered cardiac development and apoptosis, being described associated with ASD.

Comparison with other tissues

Most differential methylation studies between blood and placenta tissues done so far have been searching for biomarkers for prenatal non-invasive tests applications. We however, have performed one of the first studies searching for convergences between both tissues in ASD to see if the alteration was found in the first developmental stages. Nevertheless, methylation comparison between tissues is challenging, as inter-tissue and inter-individual epigenetic variation could confound the results.

We studied genome-wide methylation by different techniques, array platforms was used in blood and whole genome bisulfite sequencing (WGBS or MethylC-seq) in placenta. The methylation landscape showed that placenta displayed a lot of variability in methylation values, with a low correlation within tissues. The correlation coefficient, in another study using the same techniques but in the same tissue (placenta), was higher than ours (R^2=0.7 and

R^2=0.5, respectively). Therefore, this low correlation could be to the tissue-specific pattern, to a bias of WGBS technique which is described to over-represent hypermethylated regions[369], or other techniques issues. However, we are conscious that the coverage displayed by the WGBS technique (which was x1.6, with a mean of 4 reads per sample) was not enough to establish a reliable methylation status.

Despite the coverage limitation, we found interesting findings. Searching for common aberrations we found a small number of DMCpGs (n=7) common and altered in ASD in both tissues. The 7DMCpGs showed in both tissues significant alterations (>20% methylation) and followed the same methylation-status changes. The DMCpGs showed a lot of variability, limiting the effect of the alterations found. Apart from the individual CpGs, we searched for DMRs in both tissues separately, but we did not find convergent alterations. The reason for not finding convergent DMRs is that before we showed that DMRs were prone to be patient-specific, being difficult to find convergent DMRs either among all the patients in the study or in other tissues from other individuals. Also, the detection of DMRs in large domains with such a subtle methylation change involves local variability making it difficult to establish their implication in the disease.

Finally, we found genes differentially methylated in blood which located in neuronal highly methylated domains (N-HMDs) and which involved genes with a role in neurodevelopment[166]. From the 63 genes, we analyzed 10DMCpGs, which 8DMCpGs rare inherited meSNVs and 2 were false positive methylation aberrations. Some were assayed in a large case-control cohort searching for variants associated to ASD, and for 5 CpGs we could not discard the association given that only up to 2 controls were found with the meSNVs. However, as the variants are very rare, new approaches and statistical tools are needed to address rare variant association studies as the ones performed.

EWAS design and techniques features in genome-wide methylation

When addressing epigenome wide association studies (EWAS), the study design is crucial, involving as main elements a homogeneous cohort with enough size-effect. Homogenity of the

individuals involve gender, age, disease phenotypic variations, population cohorts effects, and in some studies, exhaustive environment/exposure characterization. Regarding sex, many epigenetic studies mix male and female and remove chromosome X from the studies. In our case, for the subgroups analysis we did not analyze the X-chromosome as we had individuals from both genders, hierarchical clustering methods did not separate males versus females. Regarding age, many methylation studies are assessed in aging processes. Two studies from newborns to centenarians involved a small percentage of CpGs in the process and the CpGs did not show a large methylation effect[290,370]. In our studies, we used an age paired cohort for the first differential methylation analysis, but used a wide age range control cohort, to discard common variation, which may also put down several age-variation CpGs, but were a minority. Also, cell population should be controlled[371], and also tissue correction, for example in brain, to assure homogeneity between samples[372]. However, these methods are evolving and ameliorating constantly. In the future more standardized protocols should be established. On the other side, larger cohorts for ASD methylation studies are needed to disassociate the individual changes from the pathogenic changes. In our case, although the relatively small cohort used, we used a large cohort to replicate and associate the findings.

The epigenetic field faces with reference epigenomes and techniques limitations. First, consortiums made an effort to construct a reference epigenome assessing different cell/tissues, as ENCODE and the Roadmap epigenomics consortium with 111 human epigenomes[164,249], in contrast to the genomic (exome) database gathering 60,000 individuals (Exome aggregation Consortium[373]). The challenge that reference epigenome faces is that methylation should be assessed in different tissues/cell-types in the same individual to avoid inter-individual differences, but also they might include enough people to avoid individual-specific variation. Additionally, extractions should be assessed similarly between individuals (due to tissue/cell specific methylation variation), and the technique used to quantify methylation should be as similar as possible. Also, the reference epigenome should include individuals from different developmental stages. In the future, all this data should be gathered for various diseases conditions.

On the other side, genome wide methylation assessment faces limitations due to the techniques available. The genome-wide array platforms are useful platforms to assess

methylation in large cohorts, they however only cover up to 2% of the CpGs (450k Infinium Illumina array) although nearly all genes in the genome. Its array predecessor (27k Infinium Illumina array) has even less coverage and is focused to analyze the promoters. These arrays need extensive normalization and data treatment to obtain reliable results and validation is highly recommended. Regarding WGBS, the limitations are that extensive sequencing is needed to obtain enough coverage to establish a reliable methylation percentage. The complexity of the BS-treated genome, as the resulting DNA has one base less (part of the cytosines are thymines) and the sequence is no longer complementary, increases the mapping complexity. Therefore, it is an expensive technique for organisms with large genomes as human, and to be applied to a lot of samples. Future efforts should be centered in techniques without BS treatment as direct DNA sequencing.

Contribution and future perspective

We have contributed to the epigenetic knowledge of the ASD aetiology field with epigenome-wide association studies (EWAS) between ASD (of multiple genotypes) and controls, and with the characterization of additional tissues in *in utero* environment, crucial for first developmental stages. Despite the heterogeneous nature of ASD and aetiology, we found epigenetic alterations and sequence variants altering DNA methylation targets which were found in a small number of ASD patients and were rare in a large control cohort, even determining ASD-statistically significant associated variants. For example, *ERMN* gene, which also harbored a larger mutation load in ASD samples. We also established expression consequences for 3,2% of the CpGs altered. We also characterized different ASD genotype subgroups in ASD, important for the better knowledge of additional layers contributing to their aetiology and also for designing personalized therapies in the future. Moreover, we established a methylation fingerprint which is an approximation to the aim of finding reproducible biomarkers for earlier diagnosis and patient prognosis and also for obtaining individualized treatments. Finally, the study of additional tissues as placenta, brings the in utero environmental exposures, crucial in early developmental stages for processes such as neurodevelopment. We found several genes which were part of domains involved in neurodevelopment processes and which association should be further studied.

In the future, many efforts should be performed in order to give comprehensive and solid data for epigenetics in general, and specifically in ASD, where the twins/siblings recurrence points out that the heritability is high, but genetic causes do not explain the aetiology alone and epigenetic contribution has demonstrated a collaborative role to the causative alterations.

CONCLUSIONS

CONCLUSIONS

1. We performed an epigenome-wide association study (EWAS) of blood DNA using methylation microarrays and found no global differences in the methylation patterns between idiopathic ASD and controls. The low-recurrence nature of the aberrations suggested a high heterogeneity of the disorder at the epigenetic level.

2. We also studied the blood methylation profile of a few small groups of ASD caused by several genomic rearrangements (DS, Dup15 and Del1pter). A subset of DMCpGs not present in the rearranged regions could be used to differentiate each group. DMCpGs in DS altered genes relevant for the phenotype, such as *DCC* that participates in axonogenesis. Dup15 patients showed methylation anomalies at the 11p15 imprinted domain containing the potassium-voltage channel *KCNQ1*. Patients with Del1p-ter and ART involved genes participating in development, transcription and apoptosis.

3. A group-specific methylation fingerprint achieved four groups classification (DS, Dup15, Del1pter and ART) from controls. This tool was assessed with the aim of finding blood biomarkers to classify the subgroups, although equivalent samples to perform cross-validation would be needed to be translated into practice.

4. We have found 700 DMCpGs patient-specific or shared by a small number of patients with idiopathic ASD. Most of these DMCpGs showed relative hypomethylation and 3,2% of them correlated with cis-acting effects on expression of regional genes, as shown by blood RNA-sequencing. A significant proportion of genes altered by DMCpGs (5.8%) are core candidates for ASD. We found enriched pathways involved in ASD such as axonogenesis, cell adhesion and immune chemokine signaling.

5. We identified five differentially methylated regions (DMRs) (*SMG7, PIK3CD, PHACTR1, ZCCHC9* and *RAB26*), mostly hypermethylated and associated with expression consequences, in four unrelated ASD cases. The DMRs harboured rare epimutations inherited from unaffected parents. Their absence in controls and in additional cases after

screening an additional ASD cohort indicated that the epimutations are family-specific, also reinforcing a low recurrence nature of aberrations.

6. We found that most of the individual-specific DMCpGs corresponded to rare inherited genetic variants affecting methylation targets (meSNVs). Six of the meSNVs (*ERMN, USP24, PDE10A, METTL21C, DBT* and *STX16*) proved to be ASD-susceptibility variants following a large-scale case-control association study.

7. *ERMN*, involved in myelinization, was overexpressed in two ASD individuals, one associated to a rare meSNV. Targeted-resequencing of the coding region revealed a significantly increased load of missense deleterious variants in ASD individuals compared to controls. *ERMN* is located in the AUTS5 locus, previously related to developmental delay, impaired communication and epileptic processes. Our data strongly indicate that *ERMN*, altered by either meSNV or missense mutations, is a novel ASD susceptibility gene.

8. The placental tissue is indicative of the *in utero* environment, which is crucial in neurodevelopment, having direct implications in disorders such as ASD. We found 7 DMCpGs altered in ASD in both tissues, but we did not find convergent DMRs. These results should be considered a first approach as the data for placental whole methylome available displayed low coverage, which limited the study, so the comparison of tissues awaits comparable technologies and ideally assessment of the same individuals.

9. We found differentially methylated genes in ASD blood that encompassed genes located in neuronal highly methylated domains (N-HMDs), previously characterized comparing such domains between tissues. For 6 selected genes (7DMCpGs), the aberrations were rare inherited meSNVs, and 5 of them were studied in a large association study, but we could not discard their implication given that few controls were found with the alteration. Larger cohorts are needed to assert their association.

10. The underlying genetic sequence (meSNVs) had a remarkable effect on methylation in the study. And both, the rare epigenetic and genetic alterations, both inherited, contributed under a polygenic model to the disease.

ANNEX

Codina-Solà M, Rodríguez-Santiago B, Homs A, Santoyo J, Rigau M, Aznar-Laín G, Del Campo M, Gener B, Gabau E, Botella MP, Gutiérrez-Arumí A, Antiñolo G, Pérez-Jurado LA, Cuscó I. Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders. Mol Autism. 2015 Apr 15;6:21. doi: 10.1186/s13229-015-0017-0.

Serra-Juhé C, Cuscó I, Homs A, Flores R, Torán N, Pérez-Jurado LA. DNA methylation abnormalities in congenital heart disease. Epigenetics. 2015;10(2):167-77.
doi: 10.1080/15592294.2014.998536

REFERENCES

REFERENCES

1 Kanner L. Autistic disturbances of affective contact. *Acta Paedopsychiatr* 1968; **35**: 100–136.

2 Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR Surveill Summ* 2014; **63**: 1–21.

3 Johnson CP, Myers SM. Identification and evaluation of children with autism spectrum disorders. *Pediatrics* 2007; **120**: 1183–215.

4 Brewer N. Parental identification of early behavioural abnormalities in children with autistic disorder. 2015; : 125–143.

5 Hansen RL, Ozonoff S, Krakowiak P, Angkustsiri K, Jones C, Deprey LJ *et al.* Regression in autism: prevalence and associated factors in the CHARGE Study. *Ambul Pediatr* 2008; **8**: 25–31.

6 Baird G, Charman T, Pickles A, Chandler S, Loucas T, Meldrum D *et al.* Regression, developmental trajectory and associated problems in disorders in the autism spectrum: the SNAP study. *J Autism Dev Disord* 2008; **38**: 1827–36.

7 Baird G, Simonoff E, Pickles A, Chandler S, Loucas T, Meldrum D *et al.* Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *Lancet* 2006; **368**: 210–215.

8 Lai M-C, Lombardo M V, Baron-Cohen S. Autism. *Lancet* 2014; **383**: 896–910.

9 Spence SJ, Schneider MT. The role of epilepsy and epileptiform EEGs in autism spectrum disorders. *Pediatr Res* 2009; **65**: 599–606.

10 Fombonne E, Rogé B, Claverie J, Courty S, Frémolle J. Microcephaly and macrocephaly in autism. *J Autism Dev Disord* 1999; **29**: 113–119.

11 Ozgen HM, Hop JW, Hox JJ, Beemer FA, van Engeland H. Minor physical anomalies in autism: a meta-analysis. *Mol Psychiatry* 2010; **15**: 300–7.

12 Ozgen H, Hellemann GS, Stellato RK, Lahuis B, van Daalen E, Staal WG *et al.* Morphological features in children with autism spectrum disorders: a matched case-control study. *J Autism Dev Disord* 2011; **41**: 23–31.

13 Chen R, Jiao Y, Herskovits EH. Structural MRI in autism spectrum disorder. *Pediatr Res* 2011; **69**: 63R–8R.

14 Boddaert N, Zilbovicius M, Philipe A, Robel L, Bourgeois M, Barthélemy C *et al.* MRI findings in 77 children with non-syndromic autistic disorder. *PLoS One* 2009; **4**: e4415.

15 Vasa RA, Ranta M, Huisman TAGM, Pinto PS, Tillman RM, Mostofsky SH. Normal rates of neuroradiological findings in children with high functioning autism. *J Autism Dev Disord* 2012; **42**: 1662–70.

16 Fombonne E. Epidemiology of pervasive developmental disorders. *Pediatr Res* 2009; **65**: 591–8.

17 Baron-Cohen S, Lombardo M V., Auyeung B, Ashwin E, Chakrabarti B, Knickmeyer R. Why are Autism Spectrum conditions more prevalent in Males? *PLoS Biol* 2011; **9**. doi:10.1371/journal.pbio.1001081.

18 Lai M-C, Lombardo M V., Auyeung B, Chakrabarti B, Baron-Cohen S. Sex/Gender Differences and Autism: Setting the Scene for Future Research. *J Am Acad Child Adolesc Psychiatry* 2015; **54**: 11–24.

19 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5)*. 2013 doi:10.1176/appi.books.9780890425596.744053.

20 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders – 4th Ed. (DSM-IV-TR™, 2000)*. 2000 doi:10.1001/jama.285.6.811.

21 World Health Organization. *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. 1993 doi:10.1002/1520-6505(2000)9:5<201::AID-EVAN2>3.3.CO;2-P.

22 American Psychiatric Association. Highlights of Changes from DSM-IV-TR to DSM-5. *Am Psychiatr Assoc Washington, …* 2013.http://www.dsm5.org/Documents/Autism Spectrum Disorder Fact Sheet.pdf.

23 W. David L, Peter T. DSM-5 and Proposed Changes to the Diagnosis of Autism. *Pediatr Ann* 2013; **42**: 161–166.

24 Lord C, Petkova E, Hus V, Gan W, Lu F, Martin DM *et al.* A multisite study of the clinical diagnosis of different autism spectrum disorders. *Arch Gen Psychiatry* 2012; **69**: 306–13.

25 Maenner MJ, Rice CE, Arneson CL, Cunniff C, Schieve L a, Carpenter L a *et al.* Potential impact of DSM-5 criteria on autism spectrum disorder prevalence estimates. *JAMA psychiatry* 2014; **71**: 292–300.

26 Volkmar FR, Reichow B. Autism in DSM-5: progress and challenges. *Mol Autism* 2013; **4**: 13.

27 Wing L, Potter D. The epidemiology of autistic spectrum disorders: is the prevalence rising? *Ment Retard Dev Disabil Res Rev* 2002; **8**: 151–161.

28 Fisch GS. Nosology and epidemiology in autism: Classification counts. *Am J Med Genet Part C Semin Med Genet* 2012; **160C**: 91–103.

29 King M, Bearman P. Diagnostic change and the increased prevalence of autism. *Int J Epidemiol* 2009; **38**: 1224–34.

30 Kent L, Evans J, Paul M, Sharp M. Comorbidity of autistic spectrum disorders in children with Down syndrome. *Dev Med Child Neurol* 1999; **41**: 153–158.

31 Johansson M, Råstam M, Billstedt E, Danielsson S, Strömland K, Miller M *et al.* Autism spectrum disorders and underlying brain pathology in CHARGE association. *Dev Med Child Neurol* 2006; **48**: 40–50.

32 Blumberg SJ, Bramlett MD, Kogan MD, Schieve LA, Jones JR, Lu MC. Changes in Prevalence of Parent-reported Autism Spectrum Disorder in School-aged U.S. Children: 2007 to 2011– 2012. *Natl Health Stat Report* 2013; : 1–11, 1 p following 11.

33 Wiggins LD, Baio J, Rice C. Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. *J Dev Behav Pediatr* 2006; **27**: S79–S87.

34 Liu K-Y, King M, Bearman PS. Social influence and the autism epidemic. *AJS* 2010; **115**: 1387–1434.

35 Taylor B, Jick H, Maclaughlin D. Prevalence and incidence rates of autism in the UK: time trend from 2004-2010 in children aged 8 years. *BMJ Open* 2013; **3**: e003219.

36 Rutter M. Incidence of autism spectrum disorders: changes over time and their meaning. *Acta Paediatr* 2005; **94**: 2–15.

37 Turner LM, Stone WL, Pozdol SL, Coonrod EE. Follow-up of children with autism spectrum disorders from age 2 to age 9. *Autism* 2006; **10**: 243–265.

38 Children's NCC for W and. NICE Clinical Guidelines, No. 128. Recognition, Referral and Diagnosis of Children and Young People on the Autism Spectrum. 2011.http://www.ncbi.nlm.nih.gov/books/PMH0042124/ (accessed 20 May2015).

39 Ruiz-Lázaro PM, Posada de la Paz M, Hijano Bandera F. Trastornos del espectro autista: Detección precoz, herramientas de cribado. *Pediatría Atención Primaria*; **11**: 381–397.

40 Schaefer GB, Mendelsohn NJ. Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. *Genet Med* 2013; **15**: 399–407.

41 Charman T, Gotham K. Measurement Issues: Screening and diagnostic instruments for autism spectrum disorders - lessons from research and practise. *Child Adolesc Ment Health* 2013; **18**: 52–63.

42 Lord C, Rutter M, Couteur AL. Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 1994; **24**: 659–685.

43 Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC *et al. Autism Diagnostic Observation Schedule (ADOS)*. 2000 doi:10.1007/BF02211841.

44 De Bildt A, Sytema S, Ketelaars C, Kraijer D, Mulder E, Volkmar F *et al.* Interrelationship between Autism Diagnostic Observation Schedule-Generic (ADOS-G), Autism Diagnostic Interview-Revised (ADI-R), and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) classification in children and adolescents with menta. *J Autism Dev Disord* 2004; **34**: 129–137.

45 Robins DL, Fein D, Barton ML, Green JA. The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *J Autism Dev Disord* 2001; **31**: 131–44.

46 McCray AT, Trevvett P, Frost HR. Modeling the Autism Spectrum Disorder Phenotype. Neuroinformatics. 2013; : 1–15.

47 Al-Qabandi M, Gorter JW, Rosenbaum P. Early autism detection: are we ready for routine screening? *Pediatrics* 2011; **128**: e211–e217.

48 Gadad BS, Hewitson L, Young KA, German DC. Neuropathology and Animal Models of Autism: Genetic and Environmental Factors. *Autism Res Treat* 2013; **2013**: 731935.

49 Amaral DG, Schumann CM, Nordahl CW. Neuroanatomy of autism. *Trends Neurosci* 2008; **31**: 137–145.

50 Geschwind DH, Levitt P. Autism spectrum disorders: developmental disconnection syndromes. *Curr Opin Neurobiol* 2007; **17**: 103–111.

51 Wegiel JJ, Kuchna I, Nowicki K, Imaki H, Marchi E, Ma SY *et al.* The neuropathology of autism: defects of neurogenesis and neuronal migration, and dysplastic changes. *Acta Neuropathol* 2010; **119**: 755–770.

52 Casanova MF, van Kooten IAJ, Switala AE, van Engeland H, Heinsen H, Steinbusch HWM *et al.* Minicolumnar abnormalities in autism. *Acta Neuropathol* 2006; **112**: 287–303.

53 Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare De Novo Variants Associated with Autism Implicate a Large Functional Network of Genes Involved in Formation and Function of Synapses. *Neuron* 2011; **70**: 898–907.

54 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**: 246–250.

55 De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**: 209–215.

56 Colvert E, Tick B, McEwen F, Stewart C, Curran SR, Woodhouse E *et al.* Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample. *JAMA Psychiatry* 2015; : 1–9.

57 Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A. The familial risk of autism. *JAMA* 2014; **311**: 1770–7.

58 Ronald A, Hoekstra RA. Autism spectrum disorders and autistic traits: A decade of new twin studies. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. 2011; **156**: 255–274.

59 Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L *et al.* Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics* 2011; **128**: e488–95.

60 Gaugler T, Klei L, Sanders SJ, Bodea C a, Goldberg AP, Lee AB *et al.* Most genetic risk for autism resides with common variation. *Nat Genet* 2014; **46**: 881–885.

61 Constantino JN, Todorov A, Hilton C, Law P, Zhang Y, Molloy E *et al.* Autism recurrence in half siblings: strong support for genetic mechanisms of transmission in ASD. *Mol Psychiatry* 2013; **18**: 137–8.

62 Piven J, Palmer P, Jacobi D, Childress D, Arndt S. Broader autism phenotype: Evidence from a family history study of multiple-incidence autism families. *Am J Psychiatry* 1997; **154**: 185–190.

63 Krumm N, O'Roak BJ, Shendure J, Eichler EE. A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci* 2014; **37**: 95–105.

64 Devlin B, Scherer SW, Emanuel B, Warren S. Genetic architecture in autism spectrum disorder. *Curr Opin Genet Dev* 2012; **22**: 229–237.

65 Geschwind DH. Genetics of Autism Spectrum Disorders. *Trends Cogn Sci* 2013; **15**: 409–416.

66 Todd PK, Oh S, Krans A, He F, Sellier C, Frazer M *et al.* CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* 2013; **78**: 440–455.

67 Hatton DD, Sideris J, Skinner M, Mankowski J, Bailey DB, Roberts J *et al.* Autistic behavior in children with fragile X syndrome: Prevalence, stability, and the impact of FMRP. *Am J Med Genet Part A* 2006; **140**: 1804–1813.

68 Baker P, Piven J, Sato Y. Autism and tuberous sclerosis complex: prevalence and clinical features. *J Autism Dev Disord* 1998; **28**: 279–285.

69 Williams PG, Hersh JH. Brief Report: The Association of Neurofibromatosis Type 1 and Autism. *J Autism Dev Disord*; **28**: 567–571.

70 Butler MG, Dasouki MJ, Zhou X-P, Talebizadeh Z, Brown M, Takahashi TN *et al.* Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. 2005.

71 Betancur C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res* 2011; **1380**: 42–77.

72 Manzi B, Loizzo AL, Giana G, Curatolo P. Autism and metabolic diseases. *J Child Neurol* 2008; **23**: 307–314.

73 Smith M, Spence MA, Flodman P. Nuclear and mitochondrial genome defects in autisms. *Ann N Y Acad Sci* 2009; **1151**: 102–132.

74 Persico AM, Napolioni V. Autism genetics. *Behav Brain Res* 2013; **251**: 95–112.

75 O'Roak BJ, State MW. Autism genetics: strategies, challenges, and opportunities. *Autism Res* 2008; **1**: 4–17.

76 Bailey A, Hervas A, Matthews N, Palferman S, Wallace S, Aubin A *et al.* A full genome screen for autism with evidence for linkage to a region on chromosome 7q. *Hum Mol Genet* 1998; **7**: 571–578.

77 Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet* 2008; **9**: 341–355.

78 Baker P, Piven J, Schwartz S, Patil S. Brief report: duplication of chromosome 15q11-13 in two individuals with autistic disorder. *J Autism Dev Disord* 1994; **24**: 529–35.

79 Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P *et al.* Large-Scale Copy Number Polymorphism in the Human Genome. *Science (80-)* 2004; **305**: 525–528.

80 Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y *et al.* Detection of large-scale variation in the human genome. *Nat Genet* 2004; **36**: 949–951.

81 Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J *et al.* Structural Variation of Chromosomes in Autism Spectrum Disorder. *Am J Hum Genet* 2008; **82**: 477–488.

82 Autism Genome Project Consortium Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Skaug JL, Thompson AP, Senman L, Feuk L, Qian C, Bryson SE, Jones MB, Marshall CR, Scherer SW, Vieland VJ, Bartlett C, Mangin LV, Goedken R, Segre A, Per SP. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 2007; **39**: 319–328.

83 Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T *et al.* Strong association of de novo copy number mutations with autism. *Science (80-)* 2007; **316**: 445–449.

84 Levy D, Ronemus M, Yamrom B, Lee Y ha, Leotta A, Kendall J *et al.* Rare De Novo and Transmitted Copy-Number Variation in Autistic Spectrum Disorders. *Neuron* 2011; **70**: 886–897.

85 Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D *et al.* Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. *Neuron* 2011; **70**: 863–885.

86 Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* 2014; **94**: 677–694.

87 Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R *et al.* Association between Microdeletion and Microduplication at 16p11.2 and Autism. *N Engl J Med*

2008.http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18184952.

88 Xu J, Zwaigenbaum L, Szatmari P, Scherer SW. Molecular Cytogenetics of Autism. Curr. Genomics. 2004; **5**: 347–364.

89 Weiss LA, Arking DE, Daly MJ, Chakravarti A. A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 2009; **461**: 802–808.

90 Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR *et al.* A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* 2010; **19**: 4072–82.

91 Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 2009; **459**: 528–533.

92 Devlin B, Melhem N, Roeder K. Do common variants play a role in risk for autism? Evidence and theoretical musings. Brain Res. 2011; **1380**: 78–84.

93 Berg JM, Geschwind DH. Autism genetics: searching for specificity and convergence. *Genome Biol* 2012; **13**: 247.

94 Zhao X, Leotta A, Kustanovich V, Lajonchere C, Geschwind DH, Law K *et al.* A unified genetic theory for sporadic and inherited autism. *Proc Natl Acad Sci U S A* 2007; **104**: 12831–12836.

95 Craig AM, Kang Y. Neurexin-neuroligin signaling in synapse development. *Curr Opin Neurobiol* 2007; **17**: 43–52.

96 Gerrow K, Romorini S, Nabi SM, Colicos MA, Sala C, El-Husseini A. A preformed complex of postsynaptic proteins is involved in excitatory synapse development. *Neuron* 2006; **49**: 547–562.

97 Waga C, Okamoto N, Ondo Y, Fukumura-Kato R, Goto Y-I, Kohsaka S *et al.* Novel variants of the SHANK3 gene in Japanese autistic patients with severe delayed speech development. *Psychiatr Genet* 2011; **21**: 208–211.

98 Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010; **466**: 368–72.

99 Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012; **485**: 242–5.

100 Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J *et al.* De Novo Gene Disruptions in Children on the Autistic Spectrum. *Neuron* 2012; **74**: 285–299.

101 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**: 246–250.

102 Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey a J *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**: 237–41.

103 Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**: 216–221.

104 Brandler WM, Sebat J. From De Novo Mutations to Personalized Therapeutic Interventions in Autism. *Annu Rev Med* 2015; **66**: 487–507.

105 Jeremy Willsey a, State MW. Autism spectrum disorders: from genes to neurobiology. *Curr Opin Neurobiol* 2015; **30**: 92–99.

106 He X, Sanders SJ, Liu L, De Rubeis S, Lim ET, Sutcliffe JS *et al.* Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS Genet* 2013; **9**: e1003671.

107 Guilmatre A, Dubourg C, Mosca A-L, Legallic S, Goldenberg A, Drouin-Garraud V *et al.* Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Arch Gen Psychiatry* 2009; **66**: 947–956.

108 Pescosolido MF, Gamsiz ED, Nagpal S, Morrow EM. Distribution of disease-associated copy number variants across distinct disorders of cognitive development. *J Am Acad Child Adolesc Psychiatry* 2013; **52**: 414–430.e14.

109 Huguet G, Ey E, Bourgeron T. The genetic landscapes of autism spectrum disorders. *Annu Rev Genomics Hum Genet* 2013; **14**: 191–213.

110 Miles JH. Autism spectrum disorders--a genetics review. *Genet Med* 2011; **13**: 278–294.

111 Ben-David E, Shifman S. Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. *Mol Psychiatry* 2012; : 1–2.

112 Won H, Mah W, Kim E. Autism spectrum disorder causes, mechanisms, and treatments: focus on neuronal synapses. *Front Mol Neurosci* 2013; **6**: 19.

113 Stein JL, Parikshak NN, Geschwind DH. Rare Inherited Variation in Autism: Beginning to See the Forest and a Few Trees. Neuron. 2013; **77**: 209–211.

114 Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, Huguet G *et al.* Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. *PLoS Genet* 2012; **8**. doi:10.1371/journal.pgen.1002521.

115 Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A *et al.* A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* 2010; **42**: 203–209.

116 Reichenberg A, Gross R, Weiser M, Bresnahan M, Silverman J, Harlap S *et al.* Advancing paternal age and autism. *Arch Gen Psychiatry* 2006; **63**: 1026–1032.

117 Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G *et al.* Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 2012; **488**: 471–5.

118 Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet* 2014; **15**: 133–41.

119 Brian J, Bryson SE, Garon N, Roberts W, Smith IM, Szatmari P *et al.* Clinical assessment of autism in high-risk 18-month-olds. *Autism* 2008; **12**: 433–456.

120 Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T *et al.* Genetic Heritability and Shared Environmental Factors Among Twin Pairs With Autism. Arch. Gen. Psychiatry. 2011; **68**: 1095–1102.

121 Miyake K, Yang C, Minakuchi Y, Ohori K, Soutome M, Hirasawa T *et al.* Comparison of Genomic and Epigenomic Expression in Monozygotic Twins Discordant for Rett Syndrome. *PLoS One* 2013; **8**: e66729.

122 O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 2012; **338**: 1619–22.

123 State MW, Levitt P. The conundrums of understanding genetic risks for autism spectrum disorders. *Nat Neurosci* 2011; **14**: 1499–506.

124 Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. *Nat Rev Genet* 2007; **8**: 253–62.

125 Novakovic B, Yuen RK, Gordon L, Penaherrera MS, Sharkey A, Moffett A *et al.* Evidence for widespread changes in promoter methylation profile in human placenta in response to increasing gestational age and environmental/stochastic factors. *BMC Genomics* 2011; **12**: 529.

126 Guinchat V, Thorsen P, Laurent C, Cans C, Bodeau N, Cohen D. Pre-, peri- and neonatal risk factors for autism. *Acta Obstet Gynecol Scand* 2012; **91**: 287–300.

127 Yirmiya N, Charman T. The prodrome of autism: early behavioral and biological signs, regression, peri- and post-natal development and genetics. *J Child Psychol Psychiatry* 2010; **51**: 432–58.

128 Tordjman S, Somogyi E, Coulon N, Kermarrec S, Cohen D, Bronsard G *et al.* Gene Environment Interactions in Autism Spectrum Disorders: Role of Epigenetic Mechanisms. *Front Psychiatry* 2014; **5**: 53.

129 Hertz-Picciotto I, Croen L a., Hansen R, Jones CR, van de Water J, Pessah IN. The CHARGE study: an epidemiologic investigation of genetic and environmental factors contributing to autism. *Env Heal Perspect* 2006; **114**: 1119–1125.

130 Gardener H, Spiegelman D, Buka SL. Prenatal risk factors for autism: Comprehensive meta-analysis. Br. J. Psychiatry. 2009; **195**: 7–14.

131 Shelton JF, Geraghty EM, Tancredi DJ, Delwiche LD, Schmidt RJ, Ritz B *et al.* Neurodevelopmental disorders and prenatal residential proximity to agricultural pesticides: the CHARGE study. *Environ Health Perspect* 2014; **122**: 1103–9.

132 Haglund NGS, Källén KBM. Risk factors for autism and Asperger syndrome. Perinatal factors and migration. *Autism* 2011; **15**: 163–183.

133 Volk H, Lurmann F, Penfold B, Irva H-P, Rob M. Traffic-related air pollution, particulate matter, and autism. *{JAMA} psychiatry* 2013; **70**: 71–77.

134 St-Hilaire S, Ezike VO, Stryhn H, Thomas MA. An ecological study on childhood autism. *Int J Health Geogr* 2012; **11**. doi:10.1186/1476-072X-11-44.

135 Dietert RR, Dietert JM, DeWitt JC. Environmental risk factors for autism. *Emerg Health Threats J* 2011; **4**: 1–11.

136 Volk HE, Kerin T, Lurmann F, Hertz-Picciotto I, McConnell R, Campbell DB. Autism spectrum disorder: interaction of air pollution with the MET receptor tyrosine kinase gene. *Epidemiology* 2014; **25**: 44–7.

137 Strömland K, Nordin V, Miller M, Akerström B, Gillberg C. Autism in thalidomide embryopathy: a population study. *Dev Med Child Neurol* 1994; **36**: 351–6.

138 Harden CL. In utero valproate exposure and autism: long suspected, finally proven. *Epilepsy Curr* 2013; **13**: 282–4.

139 Nanson JL. Autism in fetal alcohol syndrome: a report of six cases. *Alcohol Clin Exp Res* 1992; **16**: 558–65.

140 Martin HGS, Manzoni OJ. Late onset deficits in synaptic plasticity in the valproic acid rat model of autism. *Front Cell Neurosci* 2014; **8**: 23.

141 Schneider T, Przewłocki R. Behavioral alterations in rats prenatally exposed to valproic acid: animal model of autism. *Neuropsychopharmacology* 2005; **30**: 80–9.

142 Surén P, Roth C, Bresnahan M, Haugen M, Hornig M, Hirtz D *et al.* Association between maternal use of folic acid supplements and risk of autism spectrum disorders in children. *JAMA* 2013; **309**: 570–7.

143 Lakshmi Priya MD, Geetha A. Level of trace elements (copper, zinc, magnesium and selenium) and toxic elements (lead and mercury) in the hair and nail of children with autism. *Biol Trace Elem Res* 2011; **142**: 148–58.

144 Beard CM, Panser LA, Katusic SK. Is excess folic acid supplementation a risk factor for autism? *Med Hypotheses* 2011; **77**: 15–7.

145 Damodaran LPM, Arumugam G. Urinary oxidative stress markers in children with autism. *Redox Rep* 2011; **16**: 216–22.

146 Melnyk S, Fuchs GJ, Schulz E, Lopez M, Kahler SG, Fussell JJ *et al.* Metabolic imbalance associated with methylation dysregulation and oxidative damage in children with autism. *J Autism Dev Disord* 2012; **42**: 367–377.

147 Demicheli V, Rivetti A, Debalini MG, Di Pietrantonj C. Vaccines for measles, mumps and rubella in children. *Cochrane database Syst Rev* 2012; **2**: CD004407.

148 Rai D, Lee BK, Dalman C, Golding J, Lewis G, Magnusson C. Parental depression, maternal antidepressant use during pregnancy, and risk of autism spectrum disorders: population based case-control study. *BMJ* 2013; **346**: f2059–f2059.

149 Magnusson C, Rai D, Goodman A, Lundberg M, Idring S, Svensson A *et al.* Migration and autism spectrum disorder: population-based study. *Br J Psychiatry* 2012; **201**: 109–15.

150 WADDINGTON CH. Organisers and genes. 1940.http://www.cabdirect.org/abstracts/19401601308.html (accessed 28 Mar2015).

151 Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. *Science* 1975; **187**: 226–232.

152 Jablonka E, Lamb MJ. The changing concept of epigenetics. *Ann N Y Acad Sci* 2002; **981**: 82–96.

153 Willard HF, Brown CJ, Carrel L, Hendrich B, Miller AP. Epigenetic and chromosomal control of gene expression: molecular and genetic analysis of X chromosome inactivation. *Cold Spring Harb Symp Quant Biol* 1993; **58**: 315–22.

154 Monk M. Genomic imprinting. *Genes Dev* 1988; **2**: 921–925.

155 Wolffe AP, Matzke MA. Epigenetics: regulation through repression. *Science (80-)* 1999; **286**: 481–486.

156 Hirst M, Marra MA, Kargul J, Laurent GJ. Epigenetics and human disease. *Int J Biochem Cell Biol* 2009; **41**: 1.

157 Latham KE, Sapienza C, Engel N. The epigenetic lorax: gene-environment interactions in human health. *Epigenomics* 2012; **4**: 383–402.

158 Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A* 2008; **105**: 17046–17049.

159 Selesniemi K, Lee H-J, Muhlhauser A, Tilly JL. Prevention of maternal aging-associated oocyte aneuploidy and meiotic spindle defects in mice by dietary and genetic strategies. *Proc Natl Acad Sci U S A* 2011; **108**: 12319–24.

160 Straussman R, Nejman D, Roberts D, Steinfeld I, Blum B, Benvenisty N *et al.* Developmental programming of CpG island methylation profiles in the human genome. *Nat Struct Mol Biol* 2009; **16**: 564–571.

161 Rollins RA, Haghighi F, Edwards JR, Das R, Zhang MQ, Ju J *et al.* Large-scale structure of genomic methylation patterns. *Genome Res* 2006; **16**: 157–163.

162 Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* 2013; **500**: 477–81.

163 Doi A, Park I-H, Wen B, Murakami P, Aryee MJ, Irizarry R *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* 2009; **41**: 1350–1353.

164 Consortium RE, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 2015; **518**: 317–330.

165 Hackett J a, Surani MA. DNA methylation dynamics during the mammalian life cycle. *Philos Trans R Soc Lond B Biol Sci* 2013; **368**: 20110328.

166 Schroeder DI, Blair JD, Lott P, Yu HOK, Hong D, Crary F *et al.* The human placenta methylome. *Proc Natl Acad Sci U S A* 2013; **110**: 6037–42.

167 Hellman A, Chess A. Gene body-specific methylation on the active X chromosome. *Science* 2007; **315**: 1141–1143.

168 Ball MP, Li JB, Gao Y, Lee J-H, LeProust EM, Park I-H *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* 2009; **27**: 361–8.

169 Jones PA. The DNA methylation paradox. *Trends Genet* 1999; **15**: 34–37.

170 Guo JU, Ma DK, Mo H, Ball MP, Jang M-H, Bonaguidi MA *et al.* Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat Neurosci* 2011; **14**: 1345–51.

171 Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009; **462**: 315–22.

172 Pinney SE. Mammalian Non-CpG Methylation: Stem Cells and Beyond. *Biology (Basel)* 2014; **3**: 739–51.

173 Ziller MJ, Müller F, Liao J, Zhang Y, Gu H, Bock C *et al.* Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet* 2011; **7**: e1002389.

174 Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 2009; **324**: 929–930.

175 Neri F, Incarnato D, Krepelova A, Rapelli S, Anselmi F, Parlato C *et al.* Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Rep* 2015; **10**: 674–683.

176 Rountree MR, Bachman KE, Baylin SB. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat Genet* 2000; **25**: 269–277.

177 Esteller M. Epigenetic gene silencing in cancer: The DNA hypermethylome. Hum. Mol. Genet. 2007; **16**. doi:10.1093/hmg/ddm018.

178 Jones PA, Liang G. Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet* 2009; **10**: 805–811.

179 Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 1999; **99**: 247–257.

180 Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 1992; **69**: 915–926.

181 Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet* 2013; **14**: 204–20.

182 Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Nemes A *et al.* Sp1 elements protect a CpG island from de novo methylation. *Nature* 1994; **371**: 435–438.

183 Schübeler D. Function and information content of DNA methylation. 2015. doi:10.1038/nature14192.

184 Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 1989; **58**: 499–507.

185 Hendrich B, Bird A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* 1998; **18**: 6538–6547.

186 Klose RJ, Bird AP. Genomic DNA methylation: The mark and its mediators. Trends Biochem. Sci. 2006; **31**: 89–97.

187 Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 2010; **464**: 1082–1086.

188 Long HK, Blackledge NP, Klose RJ. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem Soc Trans* 2013; **41**: 727–40.

189 Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 2003; **33 Suppl**: 245–254.

190 Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* 2014; **156**: 45–68.

191 Lu F, Liu Y, Jiang L, Yamaguchi S, Zhang Y. Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev* 2014; **28**: 2103–19.

192 Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 2004; **431**: 96–9.

193 Walsh CP, Chaillet JR, Bestor TH. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 1998; **20**: 116–7.

194 Ehrlich M. DNA methylation: Normal development, inherited diseases, and cancer. In: *Journal of Clinical Ligand Assay*. 2000, pp 144–146.

195 Chen T, Hevi S, Gay F, Tsujimoto N, He T, Zhang B *et al.* Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nat Genet* 2007; **39**: 391–396.

196 Karpf AR, Matsui SI. Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. *Cancer Res* 2005; **65**: 8635–8639.

197 Gonzalo S, Jaco I, Fraga MF, Chen T, Li E, Esteller M *et al.* DNA methyltransferases control telomere length and telomere recombination in mammalian cells. *Nat Cell Biol* 2006; **8**: 416–424.

198 Tate PH, Bird AP. Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr Opin Genet Dev* 1993; **3**: 226–231.

199 Nan X, Cross S, Bird A. Gene silencing by methyl-CpG-binding proteins. *Novartis Found Symp* 1998; **214**: 6–16; discussion 16–21, 46–50.

200 Futscher BW, Oshiro MM, Wozniak RJ, Holtan N, Hanigan CL, Duan H *et al.* Role for DNA methylation in the control of cell type specific maspin expression. *Nat Genet* 2002; **31**: 175–179.

201 Li E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* 2002; **3**: 662–673.

202 Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Demethylation of the zygotic paternal genome. *Nature* 2000; **403**: 501–502.

203 Jaenisch R. DNA methylation and imprinting: Why bother? Trends Genet. 1997; **13**: 323–329.

204 Stancheva I, Hensey C, Meehan RR. Loss of the maintenance methyltransferase, xDnmt1, induces apoptosis in Xenopus embryos. *EMBO J* 2001; **20**: 1963–1973.

205 Niemitz EL, Feinberg AP. Epigenetics and assisted reproductive technology: a call for investigation. *Am J Hum Genet* 2004; **74**: 599–609.

206 Le Bouc Y, Rossignol S, Azzi S, Steunou V, Netchine I, Gicquel C. Epigenetics, genomic imprinting and assisted reproductive technology. *Ann Endocrinol* 2010; **71**: 237–238.

207 Maher ER, Afnan M, Barratt CL. Epigenetic risks related to assisted reproductive technologies: epigenetics, imprinting, ART and icebergs? *Hum Reprod* 2003; **18**: 2508–2511.

208 Owen CM, Segars JH. Imprinting disorders and assisted reproductive technology. Semin. Reprod. Med. 2009; **27**: 417–428.

209 Lehti V, Brown a S, Gissler M, Rihko M, Suominen a, Sourander a. Autism spectrum disorders in IVF children: a national case-control study in Finland. *Hum Reprod* 2013; **28**: 812–8.

210 Sandin S, Nygren K-G, Iliadou A, Hultman CM, Reichenberg A. Autism and mental retardation among offspring born after in vitro fertilization. *JAMA* 2013; **310**: 75–84.

211 Lee JT. Epigenetic Regulation by Long Noncoding RNAs. *Sci (Washington, DC, U S)* 2012; **338**: 1435–1439.

212 Adalsteinsson BT, Ferguson-Smith AC. Epigenetic control of the genome-lessons from genomic imprinting. *Genes (Basel)* 2014; **5**: 635–55.

213 Prows CA, Hopkin RJ. Prader Willi and Angelman syndromes: exemplars of genomic imprinting. *J Perinat Neonatal Nurs* 1999; **13**: 76–89.

214 Jackson-Grusby L, Beard C, Possemato R, Tudor M, Fambrough D, Csankovszki G *et al.* Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat Genet* 2001; **27**: 31–39.

215 Horike S, Cai S, Miyano M, Cheng J-F, Kohwi-Shigematsu T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet* 2005; **37**: 31–40.

216 Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J *et al.* DNA methylation contributes to natural human variation. *Genome Res* 2013; **23**: 1363–72.

217 Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A *et al.* Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife* 2013; **2013**. doi:10.7554/eLife.00523.

218 McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A *et al.* Identification of genetic variants that affect histone modifications in human cells. *Sci (New York, NY)* 2013; **342**: 747–749.

219 Tycko B. Allele-specific DNA methylation: beyond imprinting. *Hum Mol Genet* 2010; **19**: R210–20.

220 Bell JT, Tsai P-C, Yang T-P, Pidsley R, Nisbet J, Glass D *et al.* Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* 2012; **8**: e1002629.

221 Lemire M, Zaidi SHE, Ban M, Ge B, Aïssi D, Germain M *et al.* Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. *Nat Commun* 2015; **6**: 6326.

222 Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol* 2015; **16**: 178–189.

223 Kouzarides T. Chromatin Modifications and Their Function. *Cell* 2007; **128**: 693–705.

224 Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* 2010; **28**: 817–825.

225 Alberini CM. Transcription factors in long-term memory and synaptic plasticity. *Physiol Rev* 2009; **89**: 121–45.

226 Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 2006; **125**: 315–26.

227 Li B, Carey M, Workman JL. The Role of Chromatin during Transcription. Cell. 2007; **128**: 707–719.

228 Karlić R, Chung H-R, Lasserre J, Vlahovicek K, Vingron M. Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A* 2010; **107**: 2926–31.

229 Barth TK, Imhof A. Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem Sci* 2010; **35**: 618–26.

230 Cairns BR. The logic of chromatin architecture and remodelling at promoters. *Nature* 2009; **461**: 193–8.

231 Ho L, Crabtree GR. Chromatin remodelling during development. *Nature* 2010; **463**: 474–84.

232 Chodavarapu RK, Feng S, Bernatavichute Y V, Chen P-Y, Stroud H, Yu Y *et al.* Relationship between nucleosome positioning and DNA methylation. *Nature* 2010; **466**: 388–92.

233 Arrowsmith CH, Bountra C, Fish P V, Lee K, Schapira M. Epigenetic protein families: a new frontier for drug discovery. *Nat Rev Drug Discov* 2012; **11**: 384–400.

234 Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W *et al.* Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* 2009; **138**: 1019–31.

235 Musselman CA, Lalonde M-E, Côté J, Kutateladze TG. Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* 2012; **19**: 1218–27.

236 Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J Biol Chem* 2003; **278**: 4035–40.

237 Toma C, Torrico B, Hervás A, Salgado M, Rueda I, Valdés-Mas R *et al.* Common and rare variants of microRNA genes in autism spectrum disorders. *World J Biol Psychiatry* 2015; : 1–11.

238 Peschansky VJ, Wahlestedt C. Non-coding RNAs as direct and indirect modulators of epigenetic regulation. Epigenetics. 2014; **9**: 3–12.

239 Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol* 2013; **20**: 300–7.

240 Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012; **22**: 1775–89.

241 Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 2009; **458**: 223–7.

242 LaSalle JM, Powell WT, Yasui DH. Epigenetic layers and players underlying neurodevelopment. *Trends Neurosci* 2013; **36**: 460–70.

243 Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ *et al.* Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res* 2012; **22**: 577–591.

244 Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs Kevin. *Mol Cell* 2012; **43**: 904–914.

245 Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* 2010; **465**: 182–187.

246 Rinn JL, Chang HY. Genome Regulation by Long Noncoding RNAs. *Annu Rev Biochem* 2012; **81**: 145–166.

247 Bernard D, Prasanth K V, Tripathi V, Colasse S, Nakamura T, Xuan Z *et al.* A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J* 2010; **29**: 3082–3093.

248 Osato N, Suzuki Y, Ikeo K, Gojobori T. Transcriptional interferences in cis natural antisense transcripts of humans and mice. *Genetics* 2007; **176**: 1299–1306.

249 Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007; **447**: 799–816.

250 Robertson KD. DNA methylation and human disease. *Nat Rev Genet* 2005; **6**: 597–610.

251 Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol* 2010; **28**: 1057–68.

252 Gräff J, Mansuy IM. Epigenetic codes in cognition and behaviour. *Behav Brain Res* 2008; **192**: 70–87.

253 Schumacher A, Petronis A. Epigenetics of complex diseases: from general theory to laboratory experiments. *Curr Top Microbiol Immunol* 2006; **310**: 81–115.

254 Lopez-Rangel E, Lewis M. Loud and clear evidence for gene silencing by epigenetic mechanisms in autism spectrum and related neurodevelopmental disorders. *Clin Genet* 2005; **69**: 21–22.

255 Flashner BM, Russo ME, Boileau JE, Leong DW, Gallicano GI. Epigenetic factors and autism spectrum disorders. *NeuroMolecular Med* 2013; **15**: 339–350.

256 Rangasamy S, D'Mello SR, Narayanan V. Epigenetics, autism spectrum, and neurodevelopmental disorders. *Neurotherapeutics* 2013; **10**: 742–56.

257 Schanen NC. Epigenetics of autism spectrum disorders. *Hum Mol Genet* 2006; **15 Spec No**: R138–50.

258 Mbadiwe T, Millis RM. Epigenetics and autism. *Autism Res Treat* 2013; **2013**: 826156.

259 Mostafavi Abdolmaleky H. Horizons of psychiatric genetics and epigenetics: where are we and where are we heading? *Iran J psychiatry Behav Sci* 2014; **8**: 1–10.

260 James SJ, Melnyk S, Jernigan S, Hubanks A, Rose S, Gaylor DW. Abnormal Transmethylation/transsulfuration Metabolism and DNA Hypomethylation Among Parents of Children with Autism. *J Autism Dev Disord* 2008; **38**: 1976.

261 James SJ, Cutler P, Melnyk S, Jernigan S, Janak L, Gaylor DW *et al.* Metabolic biomarkers of increased oxidative stress and impaired methylation capacity in children with autism. *Am J Clin Nutr* 2004; **80**: 1611–1617.

262 Wong CCY, Meaburn EL, Ronald A, Price TS, Jeffries AR, Schalkwyk LC *et al.* Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol Psychiatry* 2013. doi:10.1038/mp.2013.41.

263 James SJ, Shpyleva S, Melnyk S, Pavliv O, Pogribny IP. Complex epigenetic regulation of engrailed-2 (EN-2) homeobox gene in the autism cerebellum. *Transl Psychiatry* 2013; **3**: e232.

264 Fan G, Beard C, Chen RZ, Csankovszki G, Sun Y, Siniaia M *et al.* DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. *J Neurosci* 2001; **21**: 788–797.

265 Folstein SE, Rosen-Sheidley B. Genetics of autism: complex aetiology for a heterogeneous disorder. *Nat Rev Genet* 2001; **2**: 943–55.

266 Baron CA, Tepper CG, Liu SY, Davis RR, Wang NJ, Schanen NC *et al.* Genomic and functional profiling of duplicated chromosome 15 cell lines reveal regulatory alterations in UBE3A-associated ubiquitin-proteasome pathway processes. *Hum Mol Genet* 2006; **15**: 853–869.

267 Herzing LBK, Cook EH, Ledbetter DH. Allele-specific expression analysis by RNA-FISH demonstrates preferential maternal expression of UBE3A and imprint maintenance within 15q11- q13 duplications. *Hum Mol Genet* 2002; **11**: 1707–1718.

268 Jiang Y-HH, Sahoo T, Michaelis RC, Bercovich D, Bressler J, Kashork CD *et al.* A mixed epigenetic/genetic model for oligogenic inheritance of autism with a limited role for UBE3A. *Am J Med Genet A* 2004; **131**: 1–10.

269 Hogart A, Nagarajan RP, Patzel KA, Yasui DH, Lasalle JM. 15q11-13 GABAA receptor genes are normally biallelically expressed in brain yet are subject to epigenetic dysregulation in autism-spectrum disorders. *Hum Mol Genet* 2007; **16**: 691–703.

270 Warren ST. The Epigenetics of Fragile X Syndrome. *Cell Stem Cell* 2007; **1**: 488–489.

271 Nagarajan RP, Patzel KA, Martin M, Yasui DH, Swanberg SE, Hertz-Picciotto I *et al.* MECP2 promoter methylation and X chromosome inactivation in autism. *Autism Res* 2008; **1**: 169–178.

272 Liyanage VRB, Rastegar M. Rett syndrome and MeCP2. *Neuromolecular Med* 2014; **16**: 231–64.

273 James SJ, Shpyleva S, Melnyk S, Pavliv O, Pogribny IP. Elevated 5-hydroxymethylcytosine in the Engrailed-2 (EN-2) promoter is associated with increased gene expression and decreased MeCP2 binding in autism cerebellum. *Transl Psychiatry* 2014; **4**: e460.

274 Cukier HN, Rabionet R, Konidari I, Rayner-Evans MY, Baltos ML, Wright HH *et al.* Novel variants identified in methyl-CpG-binding domain genes in autistic individuals. *Neurogenetics* 2010; **11**: 291–303.

275 Li H, Yamagata T, Mori M, Yasuhara A, Momoi MY. Mutation analysis of methyl-CpG binding protein family genes in autistic patients. *Brain Dev* 2005; **27**: 321–325.

276 Williams SR, Aldred MA, Der Kaloustian VM, Halal F, Gowans G, McLeod DR *et al.* Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. *Am J Hum Genet* 2010; **87**: 219–228.

277 Balemans MCM, Huibers MMH, Eikelenboom NWD, Kuipers AJ, van Summeren RCJ, Pijpers MMCA *et al.* Reduced exploration, increased anxiety, and altered social behavior: Autistic-like features of euchromatin histone methyltransferase 1 heterozygous knockout mice. *Behav Brain Res* 2010; **208**: 47–55.

278 Adegbola A, Gao H, Sommer S, Browning M. A novel mutation in JARID1C/SMCX in a patient with autism spectrum disorder (ASD). *Am J Med Genet Part A* 2008; **146**: 505–511.

279 Willemsen MH, Fernandez BA, Bacino CA, Gerkes E, de Brouwer APM, Pfundt R *et al.* Identification of ANKRD11 and ZNF778 as candidate genes for autism and variable cognitive impairment in the novel 16q24.3 microdeletion syndrome. *Eur J Hum Genet* 2010; **18**: 429–35.

280 Gallagher D, Voronova A, Zander MA, Cancino GI, Bramall A, Krause MP *et al.* Ankrd11 Is a Chromatin Regulator Involved in Autism that Is Essential for Neural Development. *Dev Cell* 2015; **32**: 31–42.

281 Pagnamenta AT, Khan H, Walker S, Gerrelli D, Wing K, Bonaglia MC *et al.* Rare familial 16q21 microdeletions under a linkage peak implicate cadherin 8 (CDH8) in susceptibility to autism and learning disability. *J Med Genet* 2011; **48**: 48–54.

282 Li X, Zou H, Brown WT. Genes associated with autism spectrum disorder. *Brain Res Bull* 2012; **88**: 543–552.

283 Gregory SG, Connelly JJ, Towers AJ, Johnson J, Biscocho D, Markunas CA *et al.* Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. *BMC Med* 2009; **7**: 62.

284 Mamrut S, Harony H, Sood R, Shahar-Gold H, Gainer H, Shi YJ *et al.* DNA Methylation of Specific CpG Sites in the Promoter Region Regulates the Transcription of the Mouse Oxytocin Receptor. *PLoS One* 2013; **8**. doi:10.1371/journal.pone.0056869.

285 Uchino S, Waga C. SHANK3 as an autism spectrum disorder-associated gene. Brain Dev. 2013; **35**: 106–110.

286 Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL *et al.* Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol* 2010; **28**: 1097–105.

287 Nguyen A, Rauch T a, Pfeifer GP, Hu VW. Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB J* 2010; **24**: 3036–3051.

288 Nardone S, Sharan Sams D, Reuveni E, Getselter D, Oron O, Karpuj M *et al.* DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Transl Psychiatry* 2014; **4**: e433.

289 Gartlan KH, Belz GT, Tarrant JM, Minigo G, Katsara M, Sheng K-C *et al.* A complementary role for the tetraspanins CD37 and Tssc6 in cellular immunity. *J Immunol* 2010; **185**: 3158–66.

290 Ginsberg MR, Rubin R a, Falcone T, Ting AH, Natowicz MR. Brain transcriptional and epigenetic associations with autism. *PLoS One* 2012; **7**: e44736.

291 Wang Y, Fang Y, Zhang F, Xu M, Zhang J, Yan J *et al.* Hypermethylation of the enolase gene (ENO2) in autism. *Eur J Pediatr* 2014; **173**: 1233–44.

292 Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 2011; **474**: 380–4.

293 Sandin S, Hultman CM, Kolevzon A, Gross R, MacCabe JH, Reichenberg A. Advancing maternal age is associated with increasing risk for autism: A review and meta-analysis. *J Am Acad Child Adolesc Psychiatry* 2012; **51**. doi:10.1016/j.jaac.2012.02.018.

294 Berko ER, Suzuki M, Beren F, Lemetre C, Alaimo CM, Calder RB *et al.* Mosaic Epigenetic Dysregulation of Ectodermal Cells in Autism Spectrum Disorder. *PLoS Genet* 2014; **10**. doi:10.1371/journal.pgen.1004402.

295 Ladd-Acosta C, Hansen KD, Briem E, Fallin MD, Kaufmann WE, Feinberg AP. Common DNA methylation alterations in multiple brain regions in autism. *Mol Psychiatry* 2013. doi:10.1038/mp.2013.114.

296 Sarachana T, Hu VW. Genome-wide identification of transcriptional targets of RORA reveals direct regulation of multiple genes associated with autism spectrum disorder. *Mol Autism* 2013; **4**: 14.

297 Shulha HP, Cheung I, Whittle C, Wang J, Virgil D, Lin CL *et al.* Epigenetic signatures of autism: trimethylated H3K4 landscapes in prefrontal neurons. *Arch Gen Psychiatry* 2012; **69**: 314–24.

298 Ziats MN, Rennert OM. Aberrant Expression of Long Noncoding RNAs in Autistic Brain. *J Mol Neurosci* 2013; **49**: 589–593.

299 Vincent JB, Petek E, Thevarkunnel S, Kolozsvari D, Cheung J, Patel M *et al.* The RAY1/ST7 tumor-suppressor locus on chromosome 7q31 represents a complex multi-transcript system. *Genomics* 2002; **80**: 283–94.

300 Noor A, Whibley A, Marshall CR, Gianakopoulos PJ, Piton A, Carson AR *et al.* Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability. *Sci Transl Med* 2010; **2**: 49ra68.

301 Kerin T, Ramanathan A, Rivas K, Grepo N, Coetzee GA, Campbell DB. A Noncoding RNA Antisense to Moesin at 5p14.1 in Autism. Sci. Transl. Med. 2012; **4**: 128ra40–128ra40.

302 Plongthongkum N, Diep DH, Zhang K. Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nat Rev Genet* 2014; **15**: 647–61.

303 Laird PW. Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* 2010; **11**: 191–203.

304 Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Greally JM, Gut I *et al.* Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods* 2013; **10**: 949–55.

305 Rakyan VK, Down T a, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 2011; **12**: 529–41.

306 Grunau C, Clark SJ, Rosenthal a. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res* 2001; **29**: E65–5.

307 Warnecke PM, Stirzaker C, Melki JR, Millar DS, Paul CL, Clark SJ. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res* 1997; **25**: 4422–4426.

308 Lisanti S, Omar WAW, Tomaszewski B, De Prins S, Jacobs G, Koppen G *et al.* Comparison of methods for quantification of global DNA methylation in human cells and tissues. *PLoS One* 2013; **8**. doi:10.1371/journal.pone.0079044.

309 Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* 2011; **98**: 288–295.

310 Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R *et al.* Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics* 2009; **1**: 177–200.

311 Morris TJ, Beck S. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods* 2015; **72**: 3–8.

312 Marabita F, Almgren M, Lindholm ME, Ruhrmann S, Fagerström-Billai F, Jagodic M *et al.* An evaluation of analysis pipelines for DNA methylation profiling using the illumina humanmethylation450 BeadChip platform. *Epigenetics* 2013; **8**: 333–346.

313 Lewin J, Schmitt AO, Adorján P, Hildmann T, Piepenbrock C. Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplificates. *Bioinformatics* 2004; **20**: 3005–12.

314 Jiang M, Zhang Y, Fei J, Chang X, Fan W, Qian X *et al.* Rapid quantification of DNA methylation by measuring relative peak heights in direct bisulfite-PCR sequencing traces. *Lab Invest* 2010; **90**: 282–90.

315 Ronaghi M, Uhlén M, Nyrén P. A sequencing method based on real-time pyrophosphate. *Science* 1998; **281**: 363, 365.

316 Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A* 2005; **102**: 15785–90.

317 MGI-Autism. http://www.informatics.jax.org/disease/209850.

318 Blake JA, Bult CJ, Kadin JA, Richardson JE, Eppig JT. The mouse genome database (MGD): Premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res* 2011; **39**. doi:10.1093/nar/gkq1008.

319 Banerjee-Basu S, Packer A. SFARI Gene: an evolving database for the autism research community. *Dis Model Mech*; **3**: 133–5.

320 Moy SS, Nadler JJ. Advances in behavioral genetics: mouse models of autism. *Mol Psychiatry* 2008; **13**: 4–26.

321 Provenzano G, Zunino G, Genovesi S, Sgadó P, Bozzi Y. Mutant mouse models of autism spectrum disorders. *Dis Markers* 2012; **33**: 225–239.

322 Banerjee S, Riordan M, Bhat M a. Genetic aspects of autism spectrum disorders: insights from animal models. *Front Cell Neurosci* 2014; **8**: 58.

323 Silverman JL, Yang M, Lord C, Crawley JN. Behavioural phenotyping assays for mouse models of autism. *Nat Rev Neurosci* 2010; **11**: 490–502.

324 Ey E, Leblond CS, Bourgeron T. Behavioral profiles of mouse models for autism spectrum disorders. *Autism Res* 2011; **4**: 5–16.

325 Baudouin SJ, Gaudias J, Gerharz S, Hatstatt L, Zhou K, Punnakkal P *et al.* Shared Synaptic Pathophysiology in Syndromic and Nonsyndromic Rodent Models of Autism. Science (80-.). 2012; **338**: 128–132.

326 Schmeisser MJ, Ey E, Wegener S, Bockmann J, Stempel AV, Kuebler A *et al.* Autistic-like behaviours and hyperactivity in mice lacking ProSAP1/Shank2. *Nature* 2012; **486**: 256–60.

327 Tabuchi K, Blundell J, Etherton MR, Hammer RE, Liu X, Powell CM *et al.* A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science* 2007; **318**: 71–76.

328 Rothwell PE, Fuccillo M V, Maxeiner S, Hayton SJ, Gokce O, Lim BK *et al.* Autism-associated neuroligin-3 mutations commonly impair striatal circuits to boost repetitive behaviors. *Cell* 2014; **158**: 198–212.

329 Jamain S, Radyushkin K, Hammerschmidt K, Granon S, Boretius S, Varoqueaux F *et al.* Reduced social interaction and ultrasonic communication in a mouse model of monogenic heritable autism. *Proc Natl Acad Sci U S A* 2008; **105**: 1710–1715.

330 Ey E, Yang M, Katz AM, Woldeyohannes L, Silverman JL, Leblond CS *et al.* Absence of deficits in social behaviors and ultrasonic vocalizations in later generations of mice lacking neuroligin4. *Genes, Brain Behav* 2012; **11**: 928–941.

331 Gogolla N, Leblanc JJ, Quast KB, Südhof TC, Fagiolini M, Hensch TK. Common circuit defect of excitatory-inhibitory balance in mouse models of autism. *J Neurodev Disord* 2009; **1**: 172–81.

332 Rice DS, Curran T. Role of the reelin signaling pathway in central nervous system development. *Annu Rev Neurosci* 2001; **24**: 1005–1039.

333 Panaitof SC. A songbird animal model for dissecting the genetic bases of autism spectrum disorder. *Dis Markers* 2012; **33**: 241–249.

334 Ardhanareeswaran K, Coppola G, Vaccarino F. The Use of Stem Cells to Study Autism Spectrum Disorder. *Yale J Biol Med* 2015; **88**: 5–16.

335 Prilutsky D, Palmer NP, Smedemark-Margulies N, Schlaeger TM, Margulies DM, Kohane IS. IPSC-derived neurons as a higher-throughput readout for autism: Promises and pitfalls. Trends Mol. Med. 2014; **20**: 91–104.

336 Farmer C, Thurm A, Grant P. Pharmacotherapy for the core symptoms in autistic disorder: current status of the research. *Drugs* 2013; **73**: 303–14.

337 Akins RS, Angkustsiri K, Hansen RL. Complementary and alternative medicine in autism: An evidence-based approach to negotiating safe and efficacious interventions with families. *Neurotherapeutics* 2010; **7**: 307–319.

338 Alanazi AS. The role of nutraceuticals in the management of autism. *Saudi Pharm J SPJ Off Publ Saudi Pharm Soc* 2013; **21**: 233–43.

339 Ospina MB, Seida JK, Clark B, Karkhaneh M, Hartling L, Tjosvold L *et al.* Behavioural and developmental interventions for autism spectrum disorder: A clinical systematic review. PLoS One. 2008; **3**. doi:10.1371/journal.pone.0003755.

340 Maglione MA, Gans D, Das L, Timbie J, Kasari C. Nonmedical Interventions for Children With ASD: Recommended Guidelines and Further Research Needs. Pediatrics. 2012; **130**: S169–S178.

341 Pilling S, Baron-Cohen S, Megnin-Viggars O, Lee R, Taylor C. Recognition, referral, diagnosis, and management of adults with autism: summary of NICE guidance. BMJ. 2012; **344**: e4082–e4082.

342 Howlin P, Moss P, Savage S, Rutter M. Social outcomes in mid- to later adulthood among individuals diagnosed with autism and average nonverbal IQ as children. *J Am Acad Child Adolesc Psychiatry* 2013; **52**: 572–581.

343 Sullivan K, Stone WL, Dawson G. Potential neural mechanisms underlying the effectiveness of early intervention for children with autism spectrum disorder. *Res Dev Disabil* 2014; **35**: 2921–2932.

344 Ruggeri B, Sarkans U, Schumann G, Persico AM. Biomarkers in autism spectrum disorder: The old and the new. Psychopharmacology (Berl). 2014; **231**: 1201–1216.

345 Codina-Solà M, Rodríguez-Santiago B, Homs A, Santoyo J, Rigau M, Aznar-Laín G *et al.* Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders. *Mol Autism* 2015; **6**: 21.

346 A. Homs, M. Codina, B. Rodríguez-Santiago, C. M. Villanueva, D. Monk IC and LAP-J. Genetic and epigenetic methylation defects and implication of the ERMN gene in autism spectrum disorders. *Prep* 2015.

347 Loke YJ, Hannan AJ, Craig JM. The Role of Epigenetic Change in Autism Spectrum Disorders. *Front Neurol* 2015; **6**: 107.

348 Kerkel K, Schupf N, Hatta K, Pang D, Salas M, Kratz A *et al.* Altered DNA methylation in leukocytes with trisomy 21. *PLoS Genet* 2010; **6**. doi:10.1371/journal.pgen.1001212.

349 Jin S, Lee YK, Lim YC, Zheng Z, Lin XM, Ng DPY *et al.* Global DNA Hypermethylation in Down Syndrome Placenta. *PLoS Genet* 2013; **9**. doi:10.1371/journal.pgen.1003515.

350 Yuen RK, Neumann SM, Fok AK, Peñaherrera MS, McFadden DE, Robinson WP *et al.* Extensive epigenetic reprogramming in human somatic tissues between fetus and adult. *Epigenetics Chromatin* 2011; **4**: 7.

351 Bacalini MG, Gentilini D, Boattini A, Giampieri E, Pirazzini C, Giuliani C *et al.* Identification of a DNA methylation signature in blood cells from persons with Down Syndrome. *Aging (Albany NY)* 2015; **7**: 82–96.

352 Jones MJ, Farré P, McEwen LM, Macisaac JL, Watt K, Neumann SM *et al.* Distinct DNA methylation patterns of cognitive impairment and trisomy 21 in Down syndrome. *BMC Med Genomics* 2013; **6**: 58.

353 Eckmann-Scholz C, Bens S, Kolarova J, Schneppenheim S, Caliebe A, Heidemann S *et al.* DNA-methylation profiling of fetal tissues reveals marked epigenetic differences between Chorionic and Amniotic samples. *PLoS One* 2012; **7**. doi:10.1371/journal.pone.0039014.

354 Nativio R, Sparago A, Ito Y, Weksberg R, Riccio A, Murrell A. Disruption of genomic neighbourhood at the imprinted IGF2-H19 locus in Beckwith-Wiedemann syndrome and Silver-Russell syndrome. *Hum Mol Genet* 2011; **20**: 1363–74.

355 Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* 2010; **42**: 1093–100.

356 Carone BR, Fauquier L, Habib N, Shea JM, Hart CE, Li R *et al.* Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* 2010; **143**: 1084–1096.

357 Daxinger L, Whitelaw E. Transgenerational epigenetic inheritance: More questions than answers. Genome Res. 2010; **20**: 1623–1628.

358 Ly A, Nikolaev A, Suresh G, Zheng Y, Tessier-Lavigne M, Stein E. DSCAM is a netrin receptor that collaborates with DCC in mediating turning responses to netrin-1. *Cell* 2008; **133**: 1241–54.

359 Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Degner JF, Gilad Y *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* 2011; **12**: R10.

360 Shoemaker R, Deng J, Wang W, Zhang K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res* 2010; **20**: 883–9.

361 Newbury DF, Warburton PC, Wilson N, Bacchelli E, Carone S, Lamb JA *et al.* Mapping of partially overlapping de novo deletions across an autism susceptibility region (AUTS5) in two unrelated individuals affected by developmental delays with communication impairment. *Am J Med Genet Part A* 2009; **149A**: 588–597.

362 Wang T, Jia L, Lv B, Liu B, Wang W, Wang F *et al.* Human Ermin (hErmin), a new oligodendrocyte-specific cytoskeletal protein related to epileptic seizure. *Brain Res* 2011; **1367**: 77–84.

363 Fatemi SH, Halt AR, Stary JM, Realmuto GM, Jalali-Mousavi M. Reduction in anti-apoptotic protein Bcl-2 in autistic cerebellum. *Neuroreport* 2001; **12**: 929–33.

364 Chauhan A, Chauhan V. Oxidative stress in autism. *Pathophysiology* 2006; **13**: 171–81.

365 Onore C, Careaga M, Ashwood P. The role of immune dysfunction in the pathophysiology of autism. *Brain Behav Immun* 2012; **26**: 383–92.

366 Jolly LA, Homan CC, Jacob R, Barry S, Gecz J. The UPF3B gene, implicated in intellectual disability, autism, ADHD and childhood onset schizophrenia regulates neural progenitor cell behaviour and neuronal outgrowth. *Hum Mol Genet* 2013; **22**: 4673–87.

367 Kalkman HO. A review of the evidence for the canonical Wnt pathway in autism spectrum disorders. *Mol Autism* 2012; **3**: 10.

368 Piton A, Michaud JL, Peng H, Aradhya S, Gauthier J, Mottron L *et al.* Mutations in the calcium-related gene IL1RAPL1 are associated with autism. 2008; : ddn300.

369 Ji L, Sasaki T, Sun X, Ma P, Lewis ZA, Schmitz RJ. Methylated DNA is over-represented in whole-genome bisulfite sequencing data. *Front Genet* 2014; **5**: 341.

370 Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A *et al.* Distinct DNA methylomes of newborns and centenarians. 2012. doi:10.1073/pnas.1120658109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1120658109.

371 Jaffe AE, Irizarry R a. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 2014; **15**: R31.

372 Guintivano J, Aryee MJ, Kaminsky ZA. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics* 2014; **8**: 290–302.

373 Exome Aggregation Consortium (ExAC), Cambridge M. Exome Aggregation Consortium (ExAC) Browser. http://exac.broadinstitute.org (accessed 1 Jan2015).