11-2019

THE ASSOCIATION OF DNA AND HISTONE METHYLTRANSFERASE GENES WITH DIFFERENT METHYLATION LEVELS IN FRAGILE X SYNDROME INDIVIDUALS

Sara Humaid Sembaij

United Arab Emirates University

College of Science

Department of Biology

THE ASSOCIATION OF DNA AND HISTONE METHYLTRANSFERASE GENES WITH DIFFERENT METHYLATION LEVELS IN FRAGILE X SYNDROME INDIVIDUALS

Sara Humaid Sembaij

This thesis is submitted in partial fulfillment of the requirements for the degree of Master of Science in Molecular Biology and Biotechnology

Under the Supervision of Dr. Khaled Amiri

November 2019

Declaration of Original Work

I, Sara Humaid Sembaij, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this thesis entitled "*The Association of DNA and Histone Methyltransferase Genes with Different Methylation Levels in Fragile X Syndrome Individuals*", hereby, solemnly declare that this thesis is my own original research work that has been done and prepared by me under the supervision of Dr. Khaled Amiri, in the College of Science at UAEU. This work has not previously been presented or published, or formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my thesis have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this thesis.

Student's Signature: _____ Date: 30/1/2020

Approval of the Master Thesis

This Master Thesis is approved by the following Examining Committee Members:

1) Advisor (Committee Chair): Dr. Khaled Amiri
   Title: Associate Professor and Chairman
   Department of Biology
   College of Science
   Signature _____ Date 3/11/2019

2) Member: Prof. Rabah Iratni
   Title: Professor
   Department of Biology
   College of Science
   Signature _____ Date 3/11/2019

3) Member (External Examiner): Dr. Khalid Bajou
   Title: Assistant Professor
   Department of Biology
   Institution: Sharjah University
   Signature _____ Date 03/11/2019

This Master Thesis is accepted by:

Dean of the College of Science: Professor Ahmed Murad

Signature _____ Date 02/02/2020

Dean of the College of Graduate Studies: Professor Ali Al-Marzouqi

Signature _____ Date 2/2/2020

Copy 2 of 10

Abstract

Fragile X Mental Retardation 1 (FMR1) gene produces a FMR protein (FMRP) which is known to regulate translation process in various organs. It has a significant role in neurons function, maturation and synaptic plasticity. FMR1 gene encompasses 5 - 30 CGG repeats and the greater number of repeats has the potential to expand during gametogenesis. The expansion depending on the number of CGG repeat undergoes hyper-methylation and considered as a dynamic mutation in which the expansion increases through generation. Methylation of expanded CGG repeats result in inhibiting the transcription and silencing the gene. There are two types of affected individuals of Fragile X Syndrome, one has the methylated full mutation (no protein produced) and the other is mosaic (low amount of protein produced), which are a result of having different size expansion or both methylated and unmethylated alleles. It is not clear yet what causes the differential methylation in different individuals. Therefore, we hypothesize that background gene, such as methyltransferase genes varies between individuals causing the observed epigenetic differences. The phenotypic severity depends on the number of repeats and the degree of methylation, which corresponds to the concentration of FMRP. The study is focusing on DNA and Histone methyltransferase genes, which have an important role in genome imprinting, gene regulation, X chromosome inactivation, and embryonic development. In this study, we identified nine genetic variations in lysine methyltransferase genes only. No variation was identified in DNA and other histone methyltransferase genes. Whole exome analysis resulted in a total of 37 variations which were presented in more than 35 % of mosaic and full mutation samples, 17 were novel variants and >28 variants were presented in more than 50 % of mosaic and full mutation samples, and not found in the control. In this preliminary study of fragile X syndrome, we found more variations in introns and intergenic regions that might be associated with methylation level, and physical and mental phenotypes. This preliminary data requires further genetic and functional studies, which can ultimately use genetic counseling, precision medicines and early interventions.

Keywords: Fragile X syndrome, mosaic, full mutation, DNA methyltransferase, histone methyltransferase, fragile X mental retardation 1.

Title and Abstract (in Arabic)

ارتباط جينات الحمض النووي والهيستون الناقلات للميثيل مع مستويات مختلفة من المثيلة لدى افراد متلازمة اكس الهشة

الملخص

ينتج جين اكس الهشة للأعاقة الذهنية بروتين المعروف بتنظيم عملية الترجمة في الأعضاء المختلفة ولديه دور كبير في وظيفة الخلايا العصبية، والنضج ومرونة التشابك العصبي. يحتوي الجين على 50 – 3 من ال CGG المتكررة في الاكسون الاول وكلما زاد التكرار زاد احتمال توسعه خلال عملية تكوين الأمشاج. فزيادة تكرار عدد ال CGG يخضع لارتفاع نسبة الميثيل ويعتبر بمثابة طفرة ديناميكية حيث يزداد خلال الأجيال. يؤدي تكرار ال CGG إلى تثبيت النسخ والتعبير لجين اكس الهشة للأعاقة الذهنية. هناك نوعان من الأفراد المتأثرين بمتلازمة اكس الهشة، أحدهما لديه طفرة كاملة ميثليته (لا ينتج بروتين) والاخر مثيليته جزئية (ينتج كمية منخفضة من البروتين)، والتي هي نتيجة لوجود اختلافات في حجم التكرار او نتيجة وجود كلا الأليلات المثيلية والغير مثيليه. ليس من الواضح بعد ما الذي يسبب الميثلة التفاضلية في مختلف الأفراد. لذلك نفترض أن الجينات الخلفية، مثل جينات ناقلة المثيلة تختلف بين الأفراد مسببة اختلافات جينية ملحوظة. تعتمد شدة تغير المظهر الخارجي على عدد التكرارات ودرجة المثيلة التي تتوافق مع تركيز بروتين. ستركز الدراسة على جينات الحمض النووي والهيستون الناقلة للميثيل التي لها دور هام في طبع الجينوم، وتنظيم الجينات، وتعطيل الكروموسوم X، والتطور الجنيني. في هذه الدراسة، حددنا تسعة اختلافات وراثية في جينات ليسين الناقلة للميثيل فقط، ولم يتم تحديد أي تباين في الحمض النووي وغيرها من جينات الهيستون الناقلات للميثيل. أسفر تحليل إكسوم الكامل عن 37 تغير في جينات مختلفة في أكثر من 35٪ من عينات الطفرة الكاملة والجزيئية، 17 متغيرات جديدة وأكثر عن 28 متغير موجود في أكثر من 50٪ من عينات الطفرة الكاملة والجزيئية، وتلك التغيرات غير الموجودة في عينات التحكم. في هذه الدراسة الأولية لمتلازمة اكس الهشة، وجدنا في هذه الدراسة الابتدائية على ان المتغيرات في الإنترونات وبين الجينات قد تترافق في مستويات مثيلة مختلفة وكذلك الأنماط الجسدية والعقلية في أفراد متلازمة X الهشة. تتطلب هذه البيانات الأولية مزيدًا من الدراسات الجينية والوظيفية التي يمكنها في النهاية استخدامها للاستشارة الوراثية والعلاجات الدقيقة والتدخلات المبكرة.

مفاهيم البحث الرئيسية: متلازمة اكس الهشة، مثيل الجزيئي، كامل المثيل حمض النووي الناقل للمثيل، الهيستون الناقل للمثيل، اكس الهشة للأعاقة الذهنية.

Acknowledgements

First and foremost, Alhamdulilla for everything done throughout my research work. I would like to express my deep and sincere gratitude to my research supervisor, Dr. Khaled Amiri, the head of the biology department of UAE University for giving me the opportunity to do research on fragile X syndrome and providing invaluable guidance throughout this research. It was a great privilege and honor to work and study under his guidance. I want to express my deepest appreciation to Mr. Naganeeswaran Sudalaimuthuasari for guiding me with bioinformatics analysis. I would also like to thanks Mrs. Hidaya Mohammed Abdul Kader, and Mr. Biduth Kundu for helping me in the laboratory and improve my techniques. My sincere thanks also goes to Prof. Flora Tassone from University of California, Davis, MIND Institute (USA) for providing the samples used in the project and to Sandooq Alwatan for partially funding the project. I want to thank a precious person to my heart for his continuous support and motivation during his life which made me continue doing my best and develop myself further, my father Humaid Sembaij may allah rest his soul in peace. His guidance and encouragement during my thesis project and life were the ones helped me go through every obstacle on my way. My special thanks of gratitude to my mother for her love, support and raising me to a person I am today. Finally, I am grateful for all the people, professors, doctors, family, friends and colleagues who have encouraged or supported me directly or indirectly whether from the university or outside.

Dedication

To my father Humaid Sembaij and my Grandfather Marei

May Allah rest their souls in peace

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3PUTRV	3 Prime UTR Variant
5PUTRV	5 Prime UTR Variant
5UTRSGV	5 Prime UTR Premature Start Codon Gain Variant
AdMet	S-Adenosyl-L-Methionine
DGV	Downstream Gene Variant
DInfD	Disruptive Inframe Deletion
DInfI	Disruptive Inframe Insertion
DNMT	DNA Methyltransferase
FV	Frameshift Variant
FXS	Fragile X Syndrome
ICV	Initiator Codon Variant
InfD	Inframe Deletion
InfI	Inframe Insertion
IntgV	Intergenic Variant
IntV	Intron Variant
MV	Missense Variant
PRMT	Protein Arginine Methyltransferase
SG	Stop Gained
SL	Stop Lost
SO(C)	Sequence Ontology (Combined)
SPAV	Splice Acceptor Variant
SPDV	Splice Donor Variant
SPRV	Splice Region Variant

STRV	Stop Retained Variant
SV	Synonymous Variant
TVAF	Total Variants After Filter
TVWF	Total Variants Without Filter
UGV	Upstream Gene Variant

Chapter 1: Introduction

1.1 Overview

Genes are the basic unit of heredity; each human has two copies of each gene, inherited from their parents. In humans, all genes are organized in 46 chromosomes (44 autosomes + XX in female and 44 autosomes + XY in male). Each chromosome has different types of genes that produce different types of proteins that determine specific characteristics or functions. The sequences of a particular gene could vary between people (at genotype and phenotype levels). Genome variation is due to mutation occurs on the molecular level that might specify a specific variability in the trait. Furthermore, some of these variations are associated with genetic diseases. Fragile X syndrome (FXS) is a classic example of dynamic variation that occurs in Fragile X mental retardation 1 (FMR1) gene, which causes the neurodevelopment disorder. FXS results in the expansion of triplet repeat (CGG) in FMR1 gene. The expanded repeats (>200) results in methylation of C residues and consequently silencing FMR1 gene function.

1.2 Fragile X Mental Retardation 1 [FMR1] Gene and Protein

1.2.1 Characteristics and Location

FMR1 gene is located on Xq27.3, which is ~ 40 kb in length, containing 17 exons (Lozano et al., 2014). The FMR1 gene promoter region extends to CGG repeats (5 – 30 repeats), including CGG repeats and CpG island (52 nucleotide) (Kraan et al., 2019). The FMR1 promoter is bidirectional not only transcribes FMR1 gene but also several long non coding RNAs including ASFMR1/FMR4 and FMR6 (Loesch et al., 2011; Budworth, & McMurray, 2013; Pastori et al., 2014). There are two other regions,

fragile X-related element1 (FREE1) located 5’ the promoter region and fragile X-related element2 (FREE2) located at 3’ of the CGG expansion within the intron of FMR1 gene (Figure 3a). If hyper-methylated in FXS, it leads to a decline in the transmental retardation protein (FMRP) (Kraan et al., 2019).

As shown in Figure 1, FMRP contains (a) three K homology (KH) domains (KH0, KH1, and KH2) (b) Agenet domains (AG1 and AG2) and (c) dimerization (1&2) domains (d) Unstructural regions such as, a glycine-arginine (RGG) box, nuclear export sequence (NES) and C- terminus domain. It can exist as monomer or dimer Figure1 (Dockendorff & Labrador, 2019). The three KH domains and RGG box have RNA and protein binding capacity (Valverde et al., 2008; Blackwell et al., 2010). The Agenet domains act as an intermediate for the interaction with methylated lysine and arginine residues of other proteins (Myrick et al., 2014a). The dimerization ability of FMRP helps to increase its strength and stability within the protein complex and might have other regulatory function (Dockendorff & Labrador, 2019).

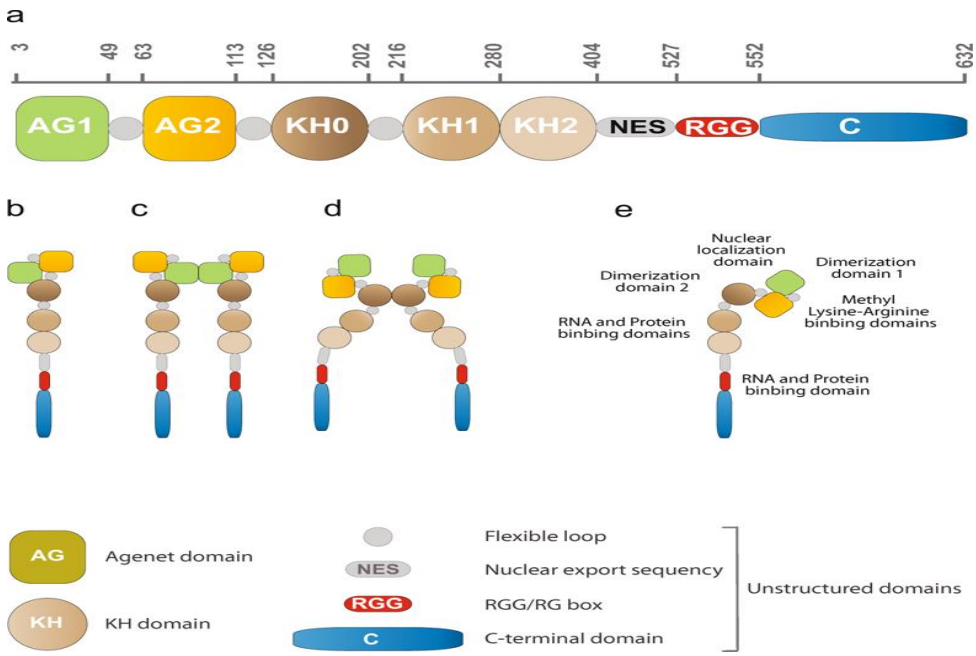

Figure 1: FMRP Structure. (a) A linear representative structure of FMRP (b) Monomer form of FMRP. (c) - (d) Different forms of FRMP dimers. (e) The FMRP components (Dockendorff et al., 2019).

1.2.2 Function

The FMR1 gene produces FMRP, a RNA binding protein of 632 amino acids in length known as a synaptic regulator, where it regulates a large number of mRNAs in postsynaptic neurons. It regulates a number of genes that are associated with autism spectrum disorders. It also regulates RNA stability, subcellular transport and translation process in various organs including ovaries, testis and more prominently the brain (Ascano et al., 2012). It is abundant in nerve cells, especially in dendrites where it has a significant role in neurons function and maturation (Halevy et al., 2015). It also has a critical part in synaptic plasticity, which in turn has a main role in memory and learning (Rosenberg et al., 2014). Recent discovery revealed the existence of FMRP within the nucleoplasm of neurons upon the analysis of its structure, which suggests that FMRP has a regulatory effect throughout the cells and acts as FRMP nuclear protein by modulating the RNA post-translational modification in the alternative splicing, as nucleocytoplasmic transport and involved in RNA editing. It has a pleiotropic function in neurons due to its ability to act as a scaffold platform with multi-interaction potential to proteins, RNAs and chromatins (Dockendorff et al., 2019; Davis and Broadie, 2017). These characteristics of FMRP demonstrate its essential roles in regulating neuronal development and function and its function within the nucleoplasm. The absence of the FMRP results in gene misregulation, an enhanced dysregulation of neural protein production, dendritic spine dysmorphogenesis and an excitation/inhibition imbalance of the special membrane receptor, metabotropic glutamate receptors signaling (Glutamate/GABA), which associated with Fragile X Syndrome (FXS) phenotypes (Hagerman et al., 2017). A study done by Fatemi et al (2011) found a significant decrease of FMRP levels in the brain in Adults with autism

and psychiatric disorders such as, bipolar disorder, major depression and schizophrenia.

1.3 Heredity & Expansion Dynamics

FMR1 gene mutations are X- linked, effecting male more than female due the X inactivation ratios (Hagerman et al., 2009). The longer the repeats, the higher propensity for expansion (anticipation). Intermediate carriers are found to transmit premutation alleles to the offspring. However, a premutation can transmit a full mutation to the next generation Figure1 (Fernandez-Carvajal et al., 2009; Nolin et al., 2003). It should be noted that the stretch of CGG repeats is found to be interrupted by AGG sequence and it was reported that AGG sequence can prevents expansion of CGG repeats (Figure 2).

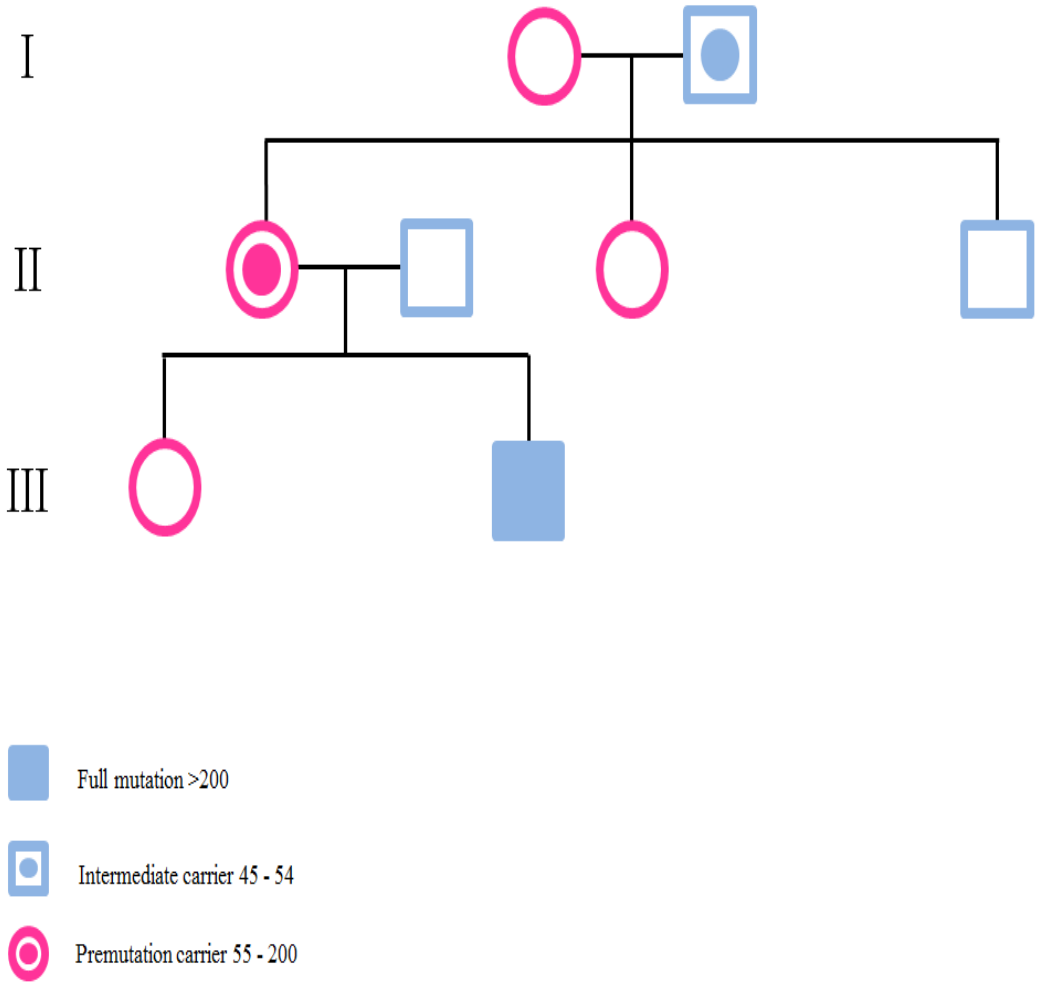

Figure 2: Pedigree of typical transmission of CGG repeats in Fragile X Syndrome Family.

Three different mutations are classified by expansion size. The categories are clustered in terms of pathological involvement and propensity for expansion Table 1.

Table 1: Different types of mutation that occurs in FMR1 gene

Type of mutation	Number of CGG Repeats	Disorder	Prevalence rates
Intermediate Mutation (Gray zone)	41 – 54 repeats	1. Could expand to Full or premutation. 2. Might be associated with disorders, neurological conditions or common features similar to premutation carriers. (Int.M1& Int.M2)	Varies
Premutation	55 – 200 repeats	Fragile X-associated tremor/ataxia syndrome (FXTAS)	1: 430 males 1: 209 females
		Premature ovarian insufficiency (POI)	
Full mutation	>200 repeats	Fragile X Syndrome (FXS)	1: 4000 males 1: 8000 Females

Premutation are shown to be associated with a score of physiological symptoms including premature ovarian insufficiency (POI) and high risk of fragile X syndrome associated tremor and ataxia (FXTAS). FXTAS is a neurodegenerative disorder due to low production of FMRP. Effected individuals develop several medical problems: (A) Psychiatric disorders (such as, anxiety and depression), (B) Chronic pain syndromes (such as, Sufibromyalgia and chronic migraine) and (c) Some can have neurodevelopmental disorder (such as, intellectual disability and autism spectrum disorder ASD) (Hagerman & Hagerman, 2015). Most of premutation carriers do not exhibit any defects or medical problems until the age of 60’s. Individuals with FXTAS of more than 50 years old could have a memory defects and symptoms that resemble

Parkinson and Alzheimer disorders (Hall et al., 2014). POI begins before 40, where women experience an irregular menstrual periods or amenorrhoea and have a high risk of infertility because of the loss function of the ovaries due to abnormal production of estrogen hormone (Barasoain et al., 2016).

1.4 Fragile X Syndrome

1.4.1 Definition

Fragile X syndrome (FXS) is a genetic neurodevelopmental disorder and the most common cause of intellectual challenges and autism caused by a single gene which results from the expansion of CGG triplet repeats.

1.4.2 The Cause of FXS

Normally, the FMR1 gene region is not methylated and the associated chromatin allows the transcription of the gene through active chromatin markers. There are two DNA region acts as methylation boundary: (1) At the 5' of the promoter discovered to be 650 to 800 nucleotides upstream of the CCG repeats, separates the FMR1 gene promoter from the methylated area. (2) At 3' of the promoter of the FMR1 gene within the intron. These boundaries interact with chromatins, conserving the FMR1 promoter region from being methylated but it is lost in FXS individuals, when the CGG repeats expands up to 200 nucleotides which results in methylation across the whole FMR1 gene sequence including FREE1 and FREE2 (Figure 3b) (Kraan et al., 2019). The chromatin undergoes a conformational change and the histones (H3 and H4) interact with lysine residues of FMR1 5'UTR region causing deacylation and methylation. Both of them results in chromatin condensation which in turn prevent transcription, therefore silencing the gene (Barasoain et al., 2016). However, a deletion

and point mutations were observed in FMR1 gene in individuals with number of expansions. This cohort represents <1% of FXS individuals. These mutation causes impairment or the absence of the FMRP. Therefore they might resemble or differ with the symptoms of FXS patients with full mutation (Myrick et al., 2014b; Handt et al., 2014; Quan et al., 1995). The gene silencing occurs at 11 weeks of gestation.

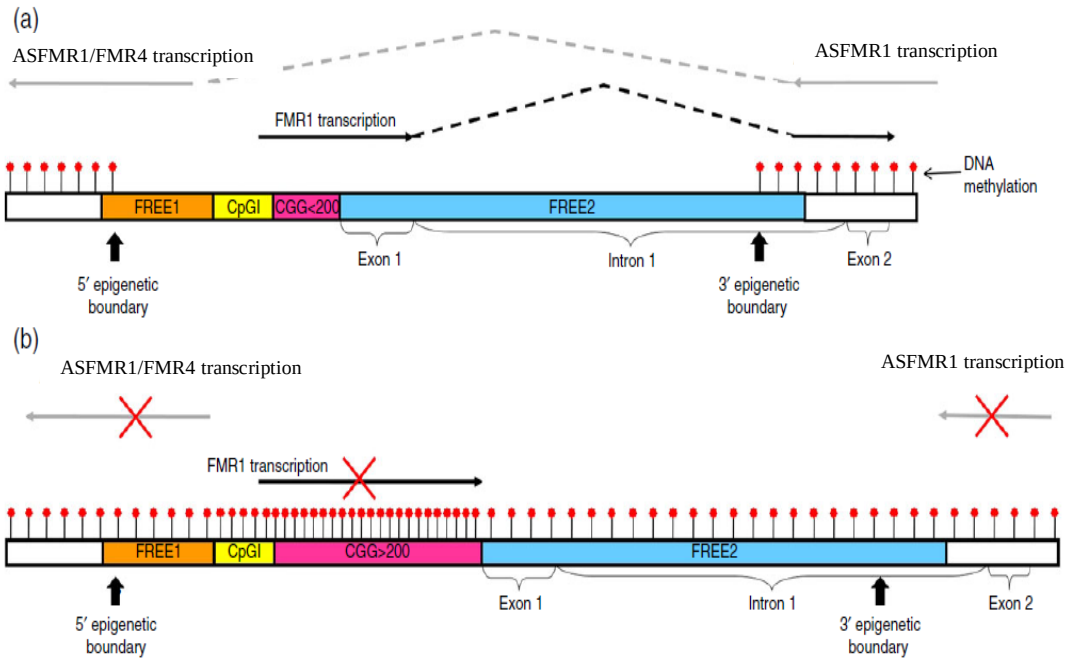

Figure 3: DNA methylation and the bidirectional transcription at FMR1 promoter in males (Kraan et al., 2019). (a) Normal unmethylated FMR1 gene with the epigenetic boundaries. (b) Full methylation of the DNA Strand and loss of the epigenetic boundaries.

1.4.3 Prevalence

About 1 in 4000 to 1 in 7000 of general population is affected by FXS (Lozano et al., 2016). However, this prevalence varies in different regions of the world. That might be attributed to environmental and genetics/epigenetics factors.

1.4.4 Symptoms and Methylation Levels

Patients with FXS suffer with clinical manifestation shown in Figure 4 and Figure 5 (Rajaratnam et al., 2017). The prevalence of this clinical manifest differs between gender and population not all may exhibit the same features. The most common behavior, cognitive and learning disability they exhibit, are low attention, hyperactivity, anxiety, tend to solitude, short term memory, hyperarousal to sensory stimuli, delayed speech and language, poor eye contact and difficulties in performing certain tasks such as, planning and organizing (Garber et al., 2008; Barasoain et al., 2016). The most prevalent clinical physical features of FXS individuals are flat feet, large ears, unusual flexible fingers, long and narrow face and a prominent jaw and forehead. Males also manifest macroorchidism after puberty (Barasoain et al., 2016; Rajaratnam et al., 2017).

These phenotypic and clinical severities depend on the number of CGG repeats and the degree of methylation which corresponds to the concentration of FMR protein (FMRP) (Saldarriaga et al., 2014). If the protein was absent or present in low amount (loss of function), the defects severity increases and vice versa. They are two types of mosaicism: (1) Methylation mosaicism occurs if some cell populations carried unmethylated alleles and others carried methylated alleles which are expressed within or across different tissue, and (2) Repeat size mosaicism is when different size of CGG expansion on FMR1 alleles are present within or across various cells including, mosaic

full mutation/premutation, mosaic full mutation/normal size & mosaic full mutation/deletion (Jiraanont et al., 2017). Mosaic males with full mutation have shown to produce low amount of FMRP and has less severe phenotype depending on FMRP levels (LaFauci et al., 2016).

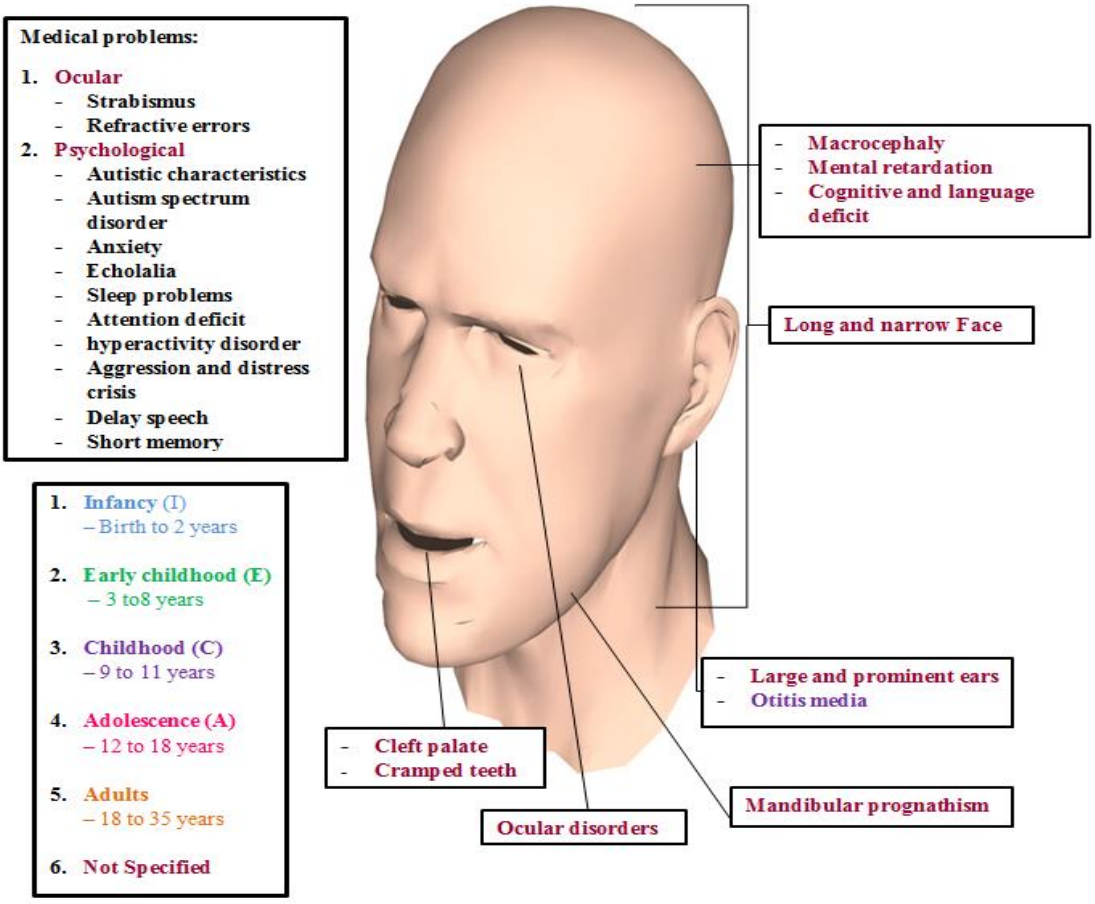

Figure 4: Facial features and clinical manifest of fragile X individuals.

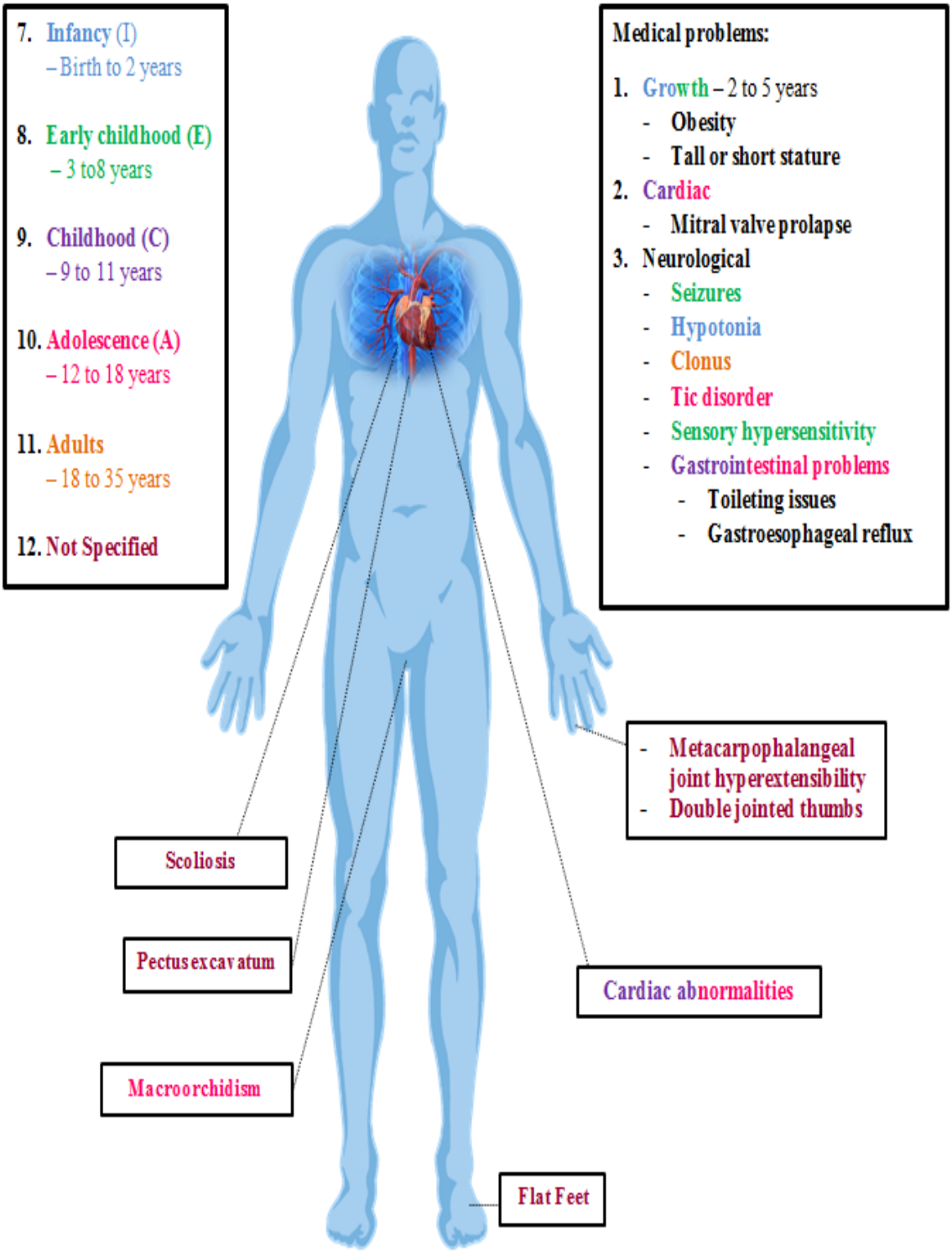

Figure 5: Body clinical manifest of Fragile X syndrome individuals.

1.5 Epigenetics

Epigenetic is a way of controlling or changing the gene activity or expression turning the genes ON/OFF without altering the DNA sequence by acting on the chromatin level and gene expression mechanisms according to the needs of the cellular system to maintain a normal healthy functions and structures of our being. It has important roles in all biological process involving cell differentiation, maintenance and cell cycle. However, abnormal epigenetic expression could result in cancer, syndromes, disorders or diseases. The two main factors in epigenetic for gene regulations are DNA methylation and histone modification.

1.5.1 DNA Methyltransferase Genes

DNA methylation is one of the epigenetic modifications that do not alter the DNA sequence, but it's involved in transferring a methyl group from S-adenosyl-L-methionine (AdoMet) at 5' position of the cytosine of the DNA segment and mostly within CpG island by DNA methyltransferases which has an important role in genome imprinting, gene regulation, X chromosome inactivation, cell fate determination, embryonic development and chromosome stability (Jin et al., 2011).

Types of DNA methyltransferase genes

There are four types of DNA methyltransferases: (a) DNMT1 (b) DNMT3A (c) DNMT3B and (d) DNMT3L**.** Variations in these genes contributes to several disorder is listed in Table 2 with their functions.

Table 2: The types of DNMT genes and their function and some associated disorders

Gene	Location	Function	Some Associated Disorders
DNMT1	19p13.2	1.Methylation maintenance during cell division 2. A preference of hemi-methylation	1. Cerebellar ataxia, deafness, and narcolepsy 2. Gastric cancer
DNMT3A	2p23.3	1.de novo methylation 2. Involved in gametogenesis and embryogenesis	1. Gastric cancer 2. Colorectal cancer
DNMT3B	20q11.21	1.de novo methylation 2. Involved in gametogenesis and embryogenesis	1. Schizophrenia in males 2. Parkinson disease
DNMT3L	21q22.3	1. Stimulates the methyltransferase activity by interacting with 3A & 3B 2. No catalytic site	1.DNA hypo-methylation 2. Schizophrenia in males

Structure of DNA Methyltransferase Proteins

DNMTs genes have both amino terminal containing regulatory domains and carboxyl terminal containing catalytic domain except DNMT3L doesn't contain a catalytic domain. DNMT1 gene has seven regulatory domains on its N terminal: (1) NLS (nuclear localization sequence) an ATRX zinc finger DNA-binding (cysteine-rich) (2) DMAP1 (DNA methyltransferase associated protein 1) (3) PBD (PCNA-proliferating cell nuclear antigen-binding) (4) RFTS (replication foci targeting sequence) (5) CXXC zinc domain an allosteric site containing eight conserved cytosine residues assembled into two CXXCXXC repeats binds to two zinc ions (6) PBHD (polybromo homology domain) which consists of two motifs: (a) BAH1 (Bromo-adjacent homology1) (b) BAH2 (Bromo-adjacent homology 2) (Bestor, 2000; Kar et al., 2012). Between N-terminal and C-terminal region is KG linker composed of multiple of lysine and glycine residues which has a role in localizing the DNMT1 near the replication fork. The carboxylic terminal region consist of ten conserved motifs (I–X) are divided into two Folds small and large domains separated by a big

cleft. The Large domain composed of motifs I-VIII and part of motif X forms the binding site for AdoMet and cytosine targeting. The small domain composed of a called TRD (target recognition domain) consists of catalytic site between VIII and IX motifs, the conserved motif IX and part of motif X that allows the binding of the target DNA into the active site and other regulatory substrates essential for gene regulation (Jeltsch and Jurkowska, 2016).

DNMT3A and DNMT3B consist of two domains in the N terminal: (a) PWWP (proline-tryptophan-tryptophan-proline domain) (b) ADD (an ATRX, DNMT3, and DNMT3L-type zinc finger). DNT3L contains ADD domain only in its N-terminal. The carboxyl-terminal for DNMT3 proteins are same having the methyltransferase domain that binds to AdoMet but DNMT3L has some substitution and deletions of amino acids within the conserved domain that makes it unable to harbor a catalytic activity, so its domain is so called Methylranferase like domain (Figure 6) (Cheng and Blumenthal, 2008; Tajima et al., 2016).

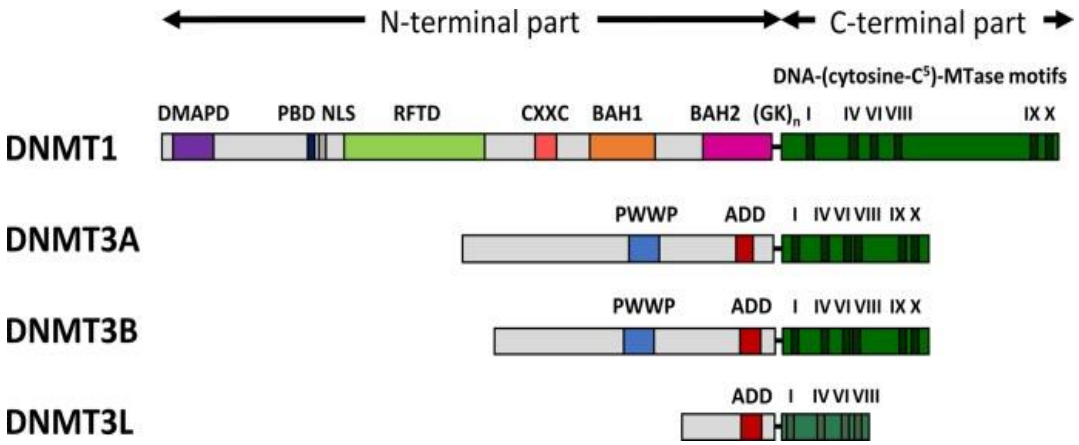

Figure 6: The Four types of DNMT proteins and different domains (Jeltsch et al., 2016).

Table 3: DNMT genes domain in the N-terminal and their function

DNMT genes	Domain / Motif	Function
DNMT1	1. NLS	Targets DNMT1 into the cell nucleus
	2. DMAP	A. Interacts with DMAP1 a transcriptional repressor. B. Facilities DNMT1's stability and its binding to DNA within the CpG dinucleotide region at the replication foci (S phase). C. Affects the methylation maintenance in early development.
	3. PBD	Locates the DNMT1 to the to the replication foci
	4. RFTS	A. Locates the DNMT1 to the to the replication foci. B. Targets it to centromeric of the chromatin. C. Involves in the dimerization of DNMT1.
	5. CXXC	A. Involves in the recognition of the unmethylated CpG island B. Induces the catalytic activity of DNMT1 by allowing the interaction of PBHD
	6. PBHD • BAH1 • BAH2	Acts as protein- protein interaction module causing the silencing of the gene.
DNMT3A	1. PWWP	A. Recognizes the H3K36 trimethylation B. Targets DNMT3A to the DNA and to pericentric heterochromatin C. Acts as protein-protein interaction module which effects the chromatin remodeling and the transcription
	2. ADD	Acts as intermediate for protein – protein interaction with regulatory factors and proteins
DNMT3B		
DNMT3L	ADD	

D. Function of DNA Methyltransferase Genes

The DNA methylation takes place during gametogenesis and embryogenesis (Table 3). It is initiated when a new methyl marker is added to the unmethylated cytosine within the CpG Island this is called de novo methylation, which is done by DNMT3A and DNMT3B and stimulated by a regulatory factor DNMT3L, the interaction happens through their C- terminal domain. A complex of tetramer is formed (3L-3a-3a-3L) which stabilize the conformation structure of the Catalytic site loop of

DNMT3Aby DNMT3L (Figure 7).This complex access the DNA by flipping the target cytosine and stimulate the methyltransferase activity. The (3a-3a) interfaces could methylate two-separated CpG in one binding event. The DNMT1 then maintains the methylation of the DNA strands and ensures that the hemimethylated daughter strands are harboring the accurate DNA patterns across the cell generation during chromosome replication and DNA repair (Chen et al., 2004; Tajima et al., 2016).

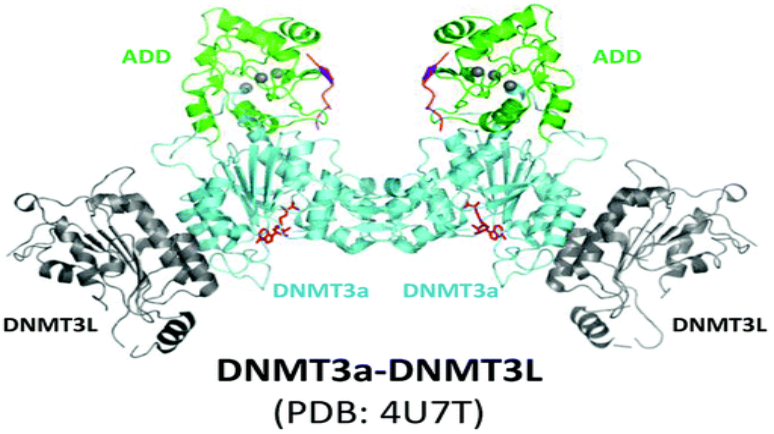

Figure 7: The Structure of DNMT3a-DNMT3L complex (Ravichandran et al., 2019).

1.5.2 Histone Methyltransferase Genes

Histone methyltransferase genes are enzymes involved in histone modification by inducing the transfer of methyl group(s) from S-adenosyl-L-methionine (AdoMet) to lysine or arginine residues of the histone proteins mainly in their N terminal tails which are positively charged. The methylation that occurs in the lysine residues could be mono, di or tri methylation whereas the Arginine residues could be only mono or di methylated. The DNA is wrapped around two of each four types of histone proteins H3, H4, H2A & H2B to form a chromatin with H1 as linker. The gene expression or the activation of the transcription depends on the chromatin structure according to the

compaction of the chromatin. Histone methyltransferase genes are classified into two genes: (A) lysine methyltransferase (B) Protein Arginine methyltransferase (Table 4).

Table 4: Different types of Histone methyltransferase genes

Arginine Methyltransferase	Lysine Methyltransferase (SET Domain)	Lysine Methyltransferase (Non-SET Domain)
PRMT1	EZH1 / EZH2	DOT1L
PRMT2	KMT2A / KMT2B / KMT2C / KMT2D / KMT2E	
PRMT3	SET1A / SET1B / SETB1 / SETB2 / SETD2 / SETD5 / SETD7 / SETD8 (KMT5A) / SETMAR	
CARM1(PRMT4)	SUV39H1 / SUV39H2	
PRMT5	EHMT1 / EHMT2	
PRMT6	ASH1L / ASH2L	
PRMT7	NSD1	
PRMT8	WHSC1 (NSD2) / WHSC1L1 (NSD3)	
PRMT9	SMYD1 / SMYD2 / SMYD3	
	SUV420H1 (KMT5B) / SUV420H2 (KMT5A)	
	PRDM2 / PRDM5 / PRDM6 / PRDM7 / PRDM8 / PRDM9 / PRDM16	
	MECOM (PRDM3)	

1.5.2.1 Lysine Methyltransferase Gene

A. Types of Lysine methyltransferase genes

The lysine methyltransferase are subbed group into two: (1) SET domain-containing lysine methyltransferase (2) Non-SET domain lysine methyltransferase (Table 5). The SET domain-containing lysine methyltransferase contains conserved SET domain with a methyltransferase activity, having 130 amino acids. The non-SET domain lysine methyltransferase has only one gene which is Dot1L doesn’t contain the SET domain as the name indicates (Zhang et al., 2003).

Table 5: Some of Lysine methyltransferase genes and their function (Dillon et al., 2005).

Types protein lysine methyltransferase	Histone lysine methylation site	HMT Genes	Function
lysine methyltransferases (SET Domain)	H1 k26	EZH2	Transcriptional silencing
	H3 K4	MLL1/MLL2/MLL3/ SET7/9 / SMYD3	Transcriptional activation
		SET 1	1. Transcriptional activation 2. Transcriptional elongation
	H3 K9	SUVAR39H1/ UVAR39H2/ G9a/ GLP1/ESET/ RIZ	1.DNA methylation 2. Heterochromatic silencing 3. Euchromatic silencing 4. Transcriptional activation or silencing
	H3 K27	EZH1 / EZH2/ G9a	1. Euchromatic silencing 2. X inactivation
	H3 k36	NSD1	1. Transcriptional elongation 2. Transcriptional silencing
	H4 K20	SET8	Cell cycle-dependent silencing, mitosis, and cytokinesis
		SUV4-20H1 / SUV4-20H2/ NSD1	Heterochromatic silencing
lysine methyltransferases (Non-SET Domain)	H3 K79	DOT1L	Demarcation of euchromatin and DNA repair

B. Structure of lysine methyltransferase proteins

The SET is categorized into pre SET which presents in the amino terminus and post SET presents in carboxyl terminus. The lysine methyltransferase proteins could have either SET domains, both of them or additional domain (i-SET) within the SET domains. Both pre SET and post SET contains a number amount of cysteine residues

that might be separated by various numbers of amino acids. The numbers of cysteine residues are different between the methyltransferase genes and could have similarity. There are four conserved motifs :(A) SET motif I (GxG) (B) SET motif II (YxG) (C) SET motif III (RFINHxCxPN) (D) SET motif IV (ELxFDY). These motifs are organized in such way to facilitate its methyltransferase activity (Table 6). These SETs form a multiple folded β stands that's creates curved small β sheets surrounds a structural pseudo-knot that brings two conserved motifs III and IV, next to AdoMet (methyl-donor-binding pocket) and the target lysine of the histone (peptide-binding cleft) (Qian et al., 2006) binding sites which are located on the opposite sides, near forming an active catalytic site. These binding sites are connected by a deep channel that allows multiple transformation of methyl group (multiple methylations) from AdoMet to the ε-amino group of the lysine without its dissociation from the SET domain. The lysine channel is formed via residues on the carboxyl terminus by having α-helix structure or metal center (zinc), onto the active site where they are required for the enzymatic activity. For non-SET domain (DOT1L) gene its catalytic activity located in the N-terminal where it methylates the Lysine residue in the globular core of the histone (H3 k79) (Figure 8) (Dillon et al., 2005).

Table 6: Lysine methyltransferase motifs and their functions

Motifs	Function
Motif I, first Half of motif (RFINH) and of motif IV (last Y)	Responsible for AdoMet binding
motif II (Y)	Involved in methylation
the second half of motif III (CxPN) and motif IV	formation of the hydrophobic target lysine-binding channel

Figure 8: Some of different lysine methyltransferase and their domains (Yang et al., 2018).

C. Function of Lysine methyltransferase genes

The methylation activity occurs in lysine residues of the histone where a methyl group is transferred from AdoMet by the SET domains to lysine residue, forming a cofactor byproduct **S**-adenosyl-L-homocysteine (AdoHcy) and a methylated lysine. The methylated lysine has specific function in gene expression; act as activation or in activation chromatin marker that helps in changing the chromatin configuration by recruiting other proteins and in elongation by the help of RNA polymerase II (Dillon et al., 2005). The mechanism depends on the methylated lysine location and its type of methylation (mono, di or tri) (Table 7).

1.5.2.2 Protein Arginine Methyltransferase Genes

A. Types of Protein Arginine Methyltransferase genes

Protein Arginine methyltransferases gene are classified into three groups, according to the transferred amount of methyl group and methylation status: (A) Type 1 (PRMT1, PRMT2, PRMT3, PRMT4 (CARM1), PRMT6, and PRMT8) catalyzes asymmetric dimethylation arginine (ADMA) by adding two methyl groups to the terminal nitrogen atoms, (B) Type2 (PRMT5, PRMT7 and PRMT9) induce the symmetric dimethylation arginine (sDMA) by adding only one methyl group, and (C) Type 3 forms monomethyl arginine (MMA) by (PRMT7). Both Type 1 and Type 2 genes catalyze the formation of MMA (Bedford et al., 2005).

B. Structure of Arginine methyltransferase proteins

The general structure of the Protein Arginine methyltransferases (PRMTs) are organized into four parts: (1) AdoMet -binding domain has Rossman fold (2) a β-barrel that involved in substrate binding (3) dimerization arm (4) N-terminus could consider

as protein-protein interaction core or could contain motifs depending on PRMT genes. The structure arrangement of these four parts differs between these genes (Figure 9) (Schapira and De Feritas, 2014).

Table 7: Different types of protein arginine methyltransferase and their function adopted from (Yang & Bedford, 2013).

PRMTs	Function
PRMT1	1. Transcription activation. 2. Signal transduction. 3. RNA splicing. 4. DNA repair.
PRMT2	Transcription regulation
PRMT3	Ribosomal homeostasis
CARM1	1. Transcription activation. 2. RNA splicing. 3. Cell cycle progression. 4. DNA repair.
PRMT5	1. Transcription repression, 2. Signal transduction and 3. piRNA pathway
PRMT6	Transcription regulation
PRMT7	Male germline gene imprinting
PRMT8	Brain-specific function
PRMT9	Unknown

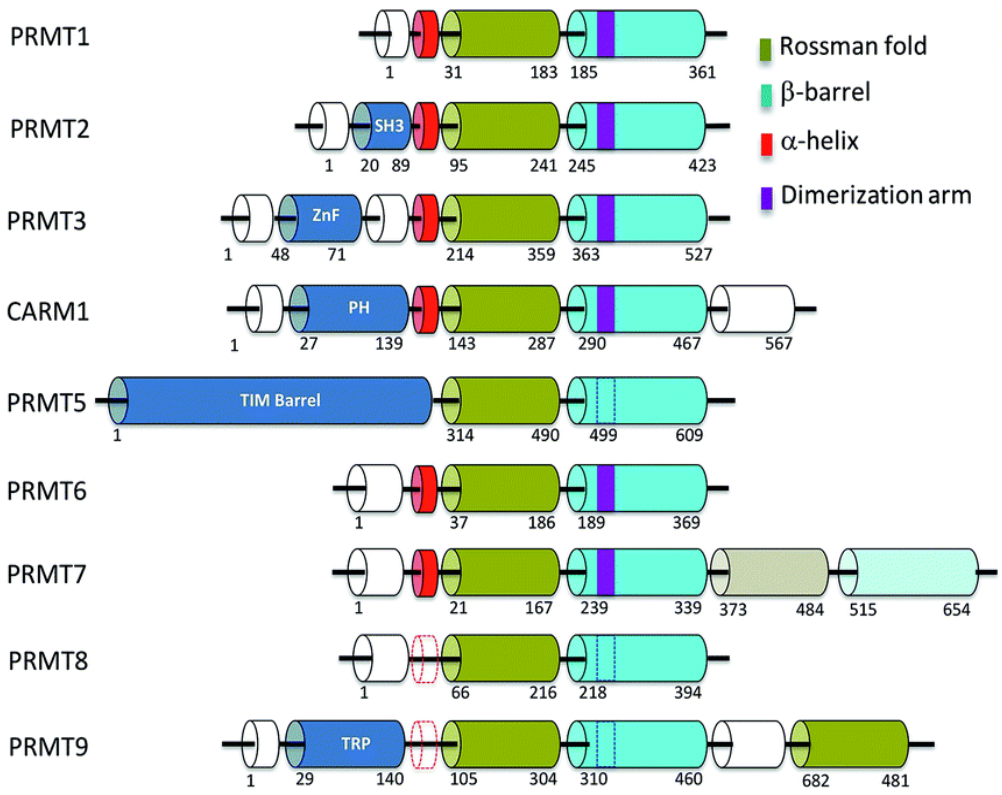

Figure 9: Different type of protein arginine methyltransferase genes and their domains (Schapira and De Feritas, 2014).

C. Function of Arginine methyltransferase proteins

PRMT genes transfer a methyl group from AdoMet to the guanidino group of arginines in protein substrates. Most of them methylate the glycine and arginine-rich (GAR) motifs in the protein substrates. In each methyl group transfer, a hydrogen bounds to a methyl in the PRMTs gene loses potential hydrogen. These genes has important role in chromatin remodeling, as transcriptional co activator and other cellular process including cell growth, proliferation and differentiation (Figure 10) (Bedford et al., 2009).

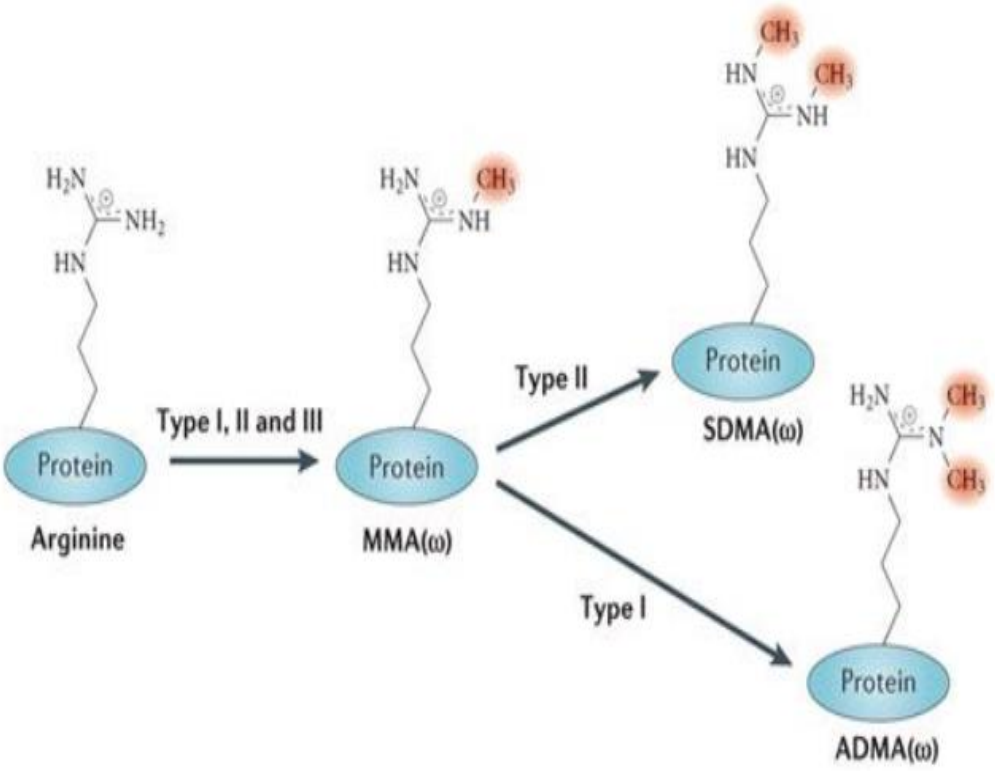

Figure 10: Different class of Arginine methyltransferase genes (Yang & Bedford, 2013).

1.6 Next Generation Sequencing

Ages ago human curiosity and circumstances were the ones kept them moving forward leading to loads of discoveries and inventions in science and other disciplines. Since DNA discovery as that code of life, scientist became more curious and motivated to gain more knowledge. Many attempts were done to sequence the nucleic acid and multiple methods were used until they found the original sequencing methodology. The original sequencing methodology which is called Sanger sequencing was invented by Fred Sanger and his colleagues was the first to sequence a whole DNA genome from bacteriophage ϕX174 and where human genome sequencing began. The principle of Sanger technique is relying on primers that identify specific location in the genes.

The sequencing reaction takes place in the presence of genomic DNA, deoxynucleoside triphosphates (dNTPs) and four different dideoxynucleotides (ddNTPs) A, T, C or G which are attached to aflorescent dye to allow DNA detection. These bases bind to the growing DNA that are intiated at 3' end by DNA polymerase and terminate the replication yielding a various length of DNA sequence (Sikkema-Raddatz et al., 2013). This technique was slow and expensive that led the researcher to improve the sequencing methods to become faster, high throughput (billons of reactions) and reduce the costs. The Next generation sequencing (NGS) was invinted using the sanger principle with massive parallel sequencing platform. This technique sequence DNA in three steps. DNA library is created from fragmented DNAs which are ligated to custom adapters. Then then these DNAs are amplified, followed by sequencing generation (Shendure and Ji, 2008). This led to efficient genome and whole exome sequencing. Whole exome sequencing targets exons and small stretch of flanking introns regions.

1.7 Hypothesis

In this experiment, we hypothesize that DNA and histone methyltransferase genes are associated in different methylation levels of fragile X syndrome individuals.

1.8 Objectives

1. To detect variations of methyltransferase genes among individuals.
2. To compare variation obtained between different groups (control, mosaic and full mutation patients of fragile x syndrome).
3. Identify specific DNA polymorphisms association with specific epigenetic status (methylation level).

Chapter 2: Methods

2.1 Samples Collection

Twenty-eight human male DNA samples were obtained from Professor Flora Tassone, University of California, Davis, MIND Insttitute (USA). Obtained samples were further classified into 3 groups; (A) five samples of controls, (B) ten samples of mosaic patients with fragile X syndrome and (C) Thirteen samples of full mutated patients with fragile X syndrome.

2.2 DNA Quality and Quantity Confirmation

DNA initial quality was checked using agarose gel electrophoresis (1%) method and DNA quantification was carried out using Nanodrop/Qubit method. Further DNA samples were diluted into 20 ng/ul concentration for the NGS library preparation

2.3 DNA Library Preparation and Whole Exome Sequencing

Exome sequencing was performed by Novogene and microgene company. Briefly, exomic regions found in the samples were captured and enriched using SureSelect V6-Post kit. Illumina compatible NGS short gun library was prepared using SureSelectXT Library Prep Kit according to manufacturer instructions. Prepared library quality and insert size was confirmed by Agilent Technologies 2100 Bioanalyzer using a DNA 1000 chip and exome sequencing was carried out using Illumina-NovoSeq platform.

2.4 Bioinformatics Data Analysis

All bioinformatics works were performed in the biology department laboratory at UAEU. The raw data (fastq files) obtained from illumina-NovoSeq platform, initial quality was checked using FastQC program. The low quality and adapter regions found in the raw data were trimmed using Trimmomatic program. The reference human genome (Build 37) was retrieved from NCBI database (Pruitt et al., 2005) and reference index was created using BWA (Houtgast et al, 2015) program.

Trimmed fastq reads were aligned against the human reference genome using BWA-MEM program. Aligned SAM files were sorted and converted into BAM files using Samtools was used to mask the duplicated reads from the alignment files and GATK pipeline was used to call the variants from the BAM file. Identified variants were annotated using dbSNP (Sherry et al., 2001) database, Clinvar (Landrum et al., 2015) database. The effect of the variant was predicted using SnpEff program. The circular chromosome map was created using circus program. An in house perl script was used for the variant filtration process.

Chapter 3: Results

3.1 DNA Quantification Using Nanodrop

The DNA initial quantification was carried out using Nanodrop. We obtained ~1.9 - ~279.5 µg of total DNA from the samples (Table 8). Samples were further diluted into ~20 ng/µl concentration for the downstream process.

Table 8: Nanodrop using UV spectrometer method for DNA quantification

Categories	Sample ID	Conc. (ng/µl)	260/280	260/230	Total Amount (µg)
Control	329-05-AE	100	1.91	1.73	5.0
	125-08-FM	70.8	1.89	1.65	3.5
	529-08-VG	58.0	1.92	1.78	2.9
	479-09-MT	53.4	1.91	2.03	2.6
	551-10-SH	42.9	1.93	1.37	2.1
Mosaic	209-12-NS	59.5	1.95	1.48	2.9
	225-12-RN	99.1	1.92	2.35	4.9
	473-12-CR	38.5	1.89	1.04	1.9
	141-13-TF	1361.1	1.88	2.04	68.0
	245-13-MB	72.4	1.94	2.89	3.6
	380-11-NS	1089.1	1.88	1.84	54.4
	120-13-SP	1569.9	1.89	2.01	78.4
	310-13-NO	770.5	1.85	2.07	38.0
	481-13-MK	1436.5	1.88	2.01	71.8
	005-14-BS	2003.3	1.88	2.20	100.1
Full mutation	17-12-ML	71.7	1.86	1.27	3.5
	009-12-GU	89.6	1.94	1.77	4.4
	699-11-EC	72.0	1.86	1.16	3.6
	273-12-TM	61.2	1.91	1.17	3.06
	197-12-JA	53.9	1.88	1.66	2.6
	311-12-TE	1139.4	1.87	1.95	56.9
	544-12-TM	382.7	1.89	1.77	19.1
	521-12-DW	861.3	1.81	1.85	43.06
	089-13-OA	1379.8	1.91	1.97	68.9
	113-13-JM	42.1	1.87	3.07	2.1
	148-13-LW	1311.6	1.88	1.88	65.5
	305-13-JG	5591.1	1.88	2.23	279.5
	299-14-EC	1495.5	1.88	2.00	74.7

3.2 DNA Quality Check Using Gel Electrophoresis

Diluted samples DNA quality was confirmed using agarose gel electrophoresis method. Figure 11 showing the DNA (single bands) quality, compare to the control and DNA ladder. We could not find any RNA contamination in the samples.

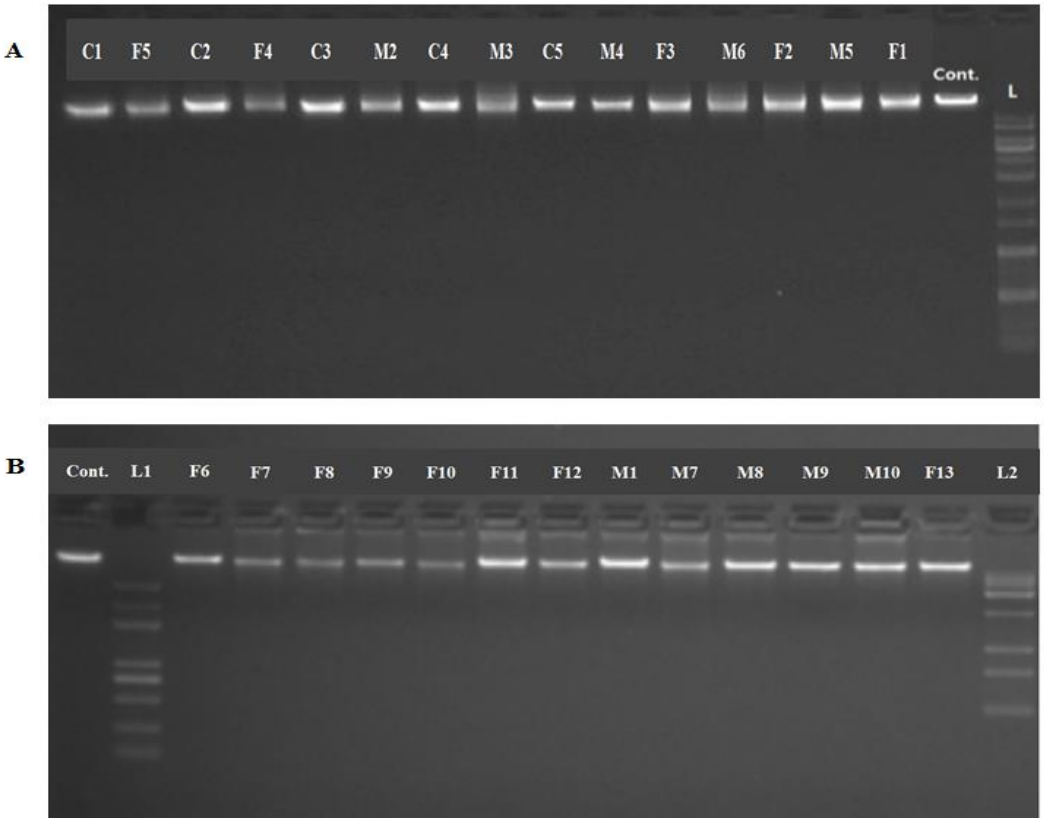

Figure 11: Gel electrophoresis for DNA quality check of the twenty-eight samples. L (1KB), L1 (2Kb), L2 (15Kb) (ladders) and Cont. (Control). A) C1-C5 control samples, M2-M6 mosaic samples, F1-F5 full mutation samples. B) M1, M7-M10 mosaic samples, F6- F13 full mutation samples.

3.3 Raw Data Quality Analysis

More than 20 million paired-end (PE) reads were generated using Illumina NovoSeq platform; overall, we obtained ~ 96 to 98% good quality reads. (>Q20) (Table 9). We found ~ 50 – 53% of GC content from the exome raw data.

Table 9: Raw Data Statistics for twenty-eight samples

Categories	Sample ID	Q20 (%)	Q30 (%)	GC (%)	AT (%)
Control	C1	98.02	94.56	51.61	48.39
	C2	97.65	93.47	51.27	48.73
	C3	98.01	94.53	51.43	48.57
	C4	97.68	93.93	51.68	48.32
	C5	97.79	94.04	51.45	48.55
Mosaic	M1	97.71	93.74	51.66	48.34
	M2	96.89	92.06	51.18	48.82
	M3	97.89	94.24	51.61	48.39
	M4	97.62	93.66	50.86	49.14
	M5	97.48	93.17	51.21	48.79
	M6	97.40	92.86	53.18	46.82
	M7	97.91	94.42	51.18	48.82
	M8	97.83	94.06	53.66	46.34
	M9	97.63	93.58	52.31	47.69
	M10	97.76	93.83	52.63	47.37
Full mutation	F1	97.81	94.18	51.16	48.84
	F2	97.81	94.1	52.0	48.0
	F3	97.9	94.33	51.53	48.47
	F4	97.76	94.06	51.64	48.36
	F5	97.93	94.33	51.44	48.56
	F6	97.76	93.86	51.63	48.37
	F7	97.53	93.35	52.89	97.53
	F8	97.48	93.18	52.70	47.30
	F9	97.69	93.67	52.69	47.31
	F10	97.79	94.06	51.38	48.62
	F11	97.66	93.66	53.47	46.53
	F12	97.67	93.68	53.48	46.52
	F13	97.81	94.00	52.54	47.46

3.4 Filtered Reads Quality Check and Reference Alignment Statistics

After quality trimming, more than 85% of reads were retained for the downstream analysis. Initial reference alignment resulted ~97 to 99% of reads aligned against the reference genome and ~93 to ~97% of whole exome regions were sequenced at 10X coverage. Detailed alignment and exome coverage statistics for all 28 samples are provided in Tables 10 -12 and Figures 12 and 13.

Table 10: Read quality check for five control samples

Read QC	C1	C2	C3	C4	C5
Raw data count	27,902,034	22,767,126	26,366,576	29,074,344	33,807,040
Filtered data count	23,895,063	19,741,845	23,221,446	25,028,725	29,097,666
Alignment %	99.04%	99.03%	98.94%	98.52%	99%
Average depth	65.05%	52.88%	62.22%	66.68%	77.38%
Coverage at (100x)	17.10%	10.25%	15.34%	17.82%	24.66%
Coverage at (50x)	55.87%	43.27%	53.16%	57.46%	67.57%
Coverage at (20x)	91.36%	86.05%	90.38%	92.22%	94.17%
Coverage at (10x)	96.29%	94.96%	96.02%	96.51%	96.84%
Coverage at (2x)	97.80%	97.64%	97.78%	97.86%	97.87%
Coverage at (1x)	98.01%	97.89%	98.00%	98.05%	98.05%

Table 11: Read quality check for ten mosaic samples

Read QC	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Raw data count	27,595,146	39,439,781	32,498,055	35,048,944	22,823,731	24,763,345	34,053,056	25,242,312	25,164,351	21,958,289
Filtered data count	26,509,988	33,778,485	29,563,881	30,367,765	21,810,343	23,403,205	29,199,090	24,152,558	24,040,327	20,954,999
Alignment %	99.26	98.68	98.11	98.75	99.37	99.33	98.08	99.39	99.39	99.36
Average depth	51.93%	87.84%	77.92%	76.56%	44.86%	52.35%	76.09%	54.66%	49.49%	47.55%
Coverage at (100x)	11.37%	32.1039	25.92%	23.94%	7.12%	12.31%	24.25%	13.93%	10.58%	9.21 %
Coverage at (50x)	37.17%	74.80%	64.02%	68.93%	31.00%	36.83%	65.92%	37.96%	34.55%	33.11%
Coverage at (20x)	82.49%	94.93%	92.32%	94.64%	80.08%	80.32%	93.38%	79.58%	79.14%	79.19%
Coverage at (10x)	94.86%	96.85%	96.41%	96.97%	94.08%	94.02%	96.61%	93.87%	93.97%	93.76%
Coverage at (2x)	97.79%	97.73%	97.85%	97.94%	97.68%	97.67%	97.80%	97.73%	97.76%	97.67%
Coverage at (1x)	97.97%	97.90%	98.05%	98.13%	97.90%	97.87%	97.98%	97.93%	97.96%	97.88%

Table 12: Read quality check for thirteen full mutation samples

Read QC	F1	F2	F3	F4	F5	F6
Raw data count	33,345,350	21,779,340	25,268,642	32,660,384	40,194,723	26,980,216
Filtered data count	27,425,752	19,129,499	21,966,359	28,563,104	34,965,395	25,992,280
Alignment %	97.91	98.47	98.83	97.86	99.07	99.32%
Average depth	72.27%	51.90%	59.30%	75.14%	91.70%	52.58%
Coverage at (100x)	21.57%	10.58%	13.78%	23.62%	34.40%	11.11%
Coverage at (50x)	63.58%	40.84%	50.00%	64.08%	74.92%	38.52%
Coverage at (20x)	93.09%	84.02%	89.33%	93.26%	94.93%	84.40%
Coverage at (10x)	96.55%	94.57%	95.84%	96.65%	97.00%	95.21%
Coverage at (2x)	97.78%	97.68%	97.76%	97.83%	97.92%	97.79%
Coverage at (1x)	97.96%	97.92%	97.98%	98.02%	98.08%	97.98%

Table 12: Read quality check for thirteen full mutation samples (continued)

Read QC	F7	F8	F9	F10	F11	F12	F13
Raw data count	21,517,487	20,362,275	22,099,117	33,534,744	25,638,704	20,690,825	27,322,097
Filtered data count	20,551,473	19,495,358	21,322,275	28,546,975	24,443,077	19,571,128	26,229,844
Alignment %	99.32	99.42	99.42	98.87%	99.37	99.44	99.21
Average depth	44.43%	41.71%	45.30%	76.14%	54.87%	45.24%	55.60%
Coverage at (100x)	8.69%	7.14%	8.65%	24.54%	13.90%	9.17%	13.68%
Coverage at (50x)	29.73%	27.22%	30.59%	65.04%	38.31%	30.44%	39.83%
Coverage at (20x)	73.18%	71.73%	75.86%	92.97%	80.48%	73.47%	83.07%
Coverage at (10x)	92.12%	91.67%	93.21%	96.50%	94.16%	91.95%	94.94%
Coverage at (2x)	97.68%	97.62%	97.75%	97.76%	97.74%	97.60%	97.77%
Coverage at (1x)	97.92%	97.87%	97.96%	97.94%	97.92%	97.84%	97.95%

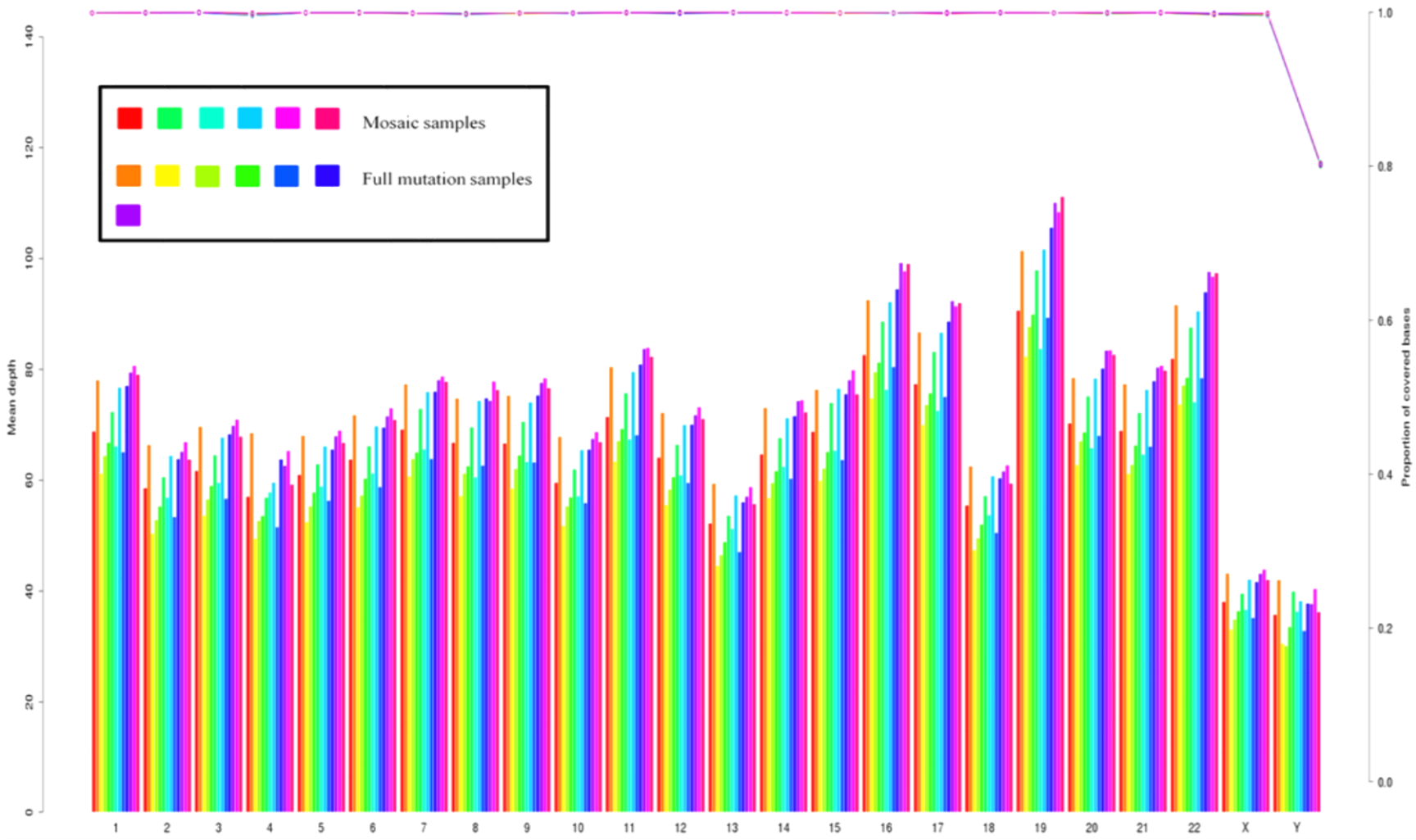

Figure 12: The average depth (bar plot) and coverage (dot plot) of Mosaic and full mutation samples in each chromosome.

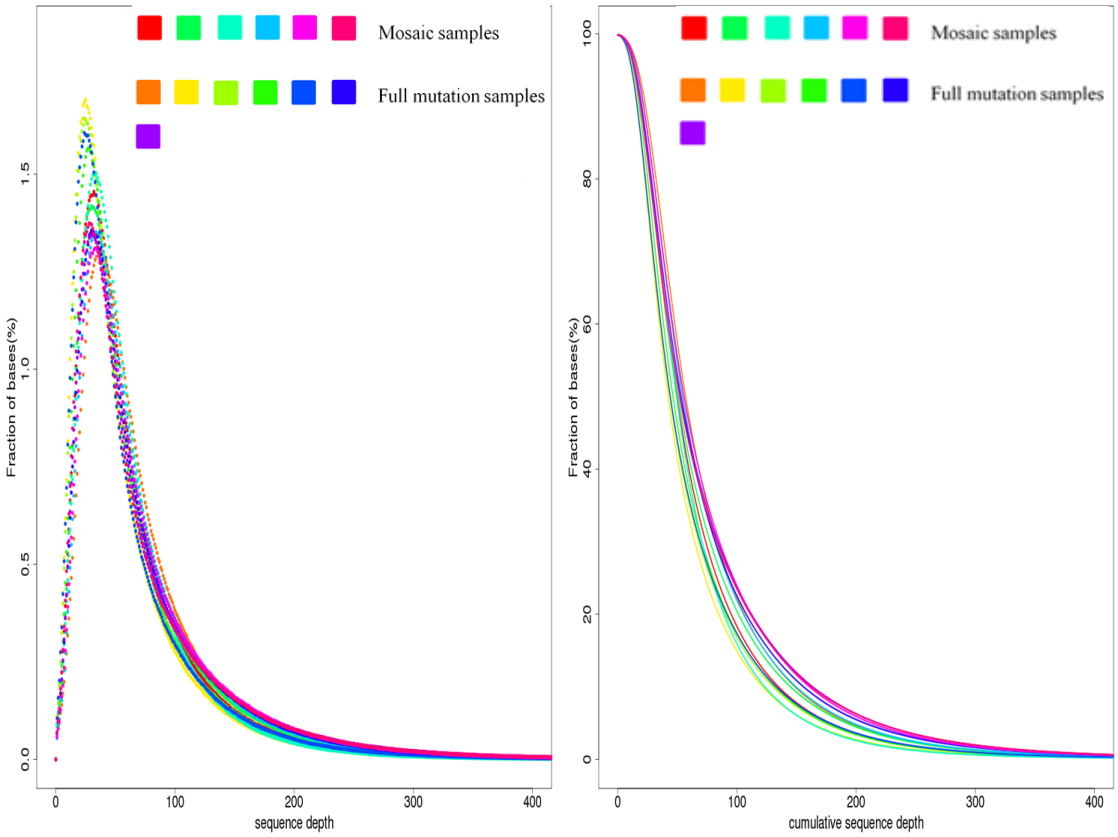

Figure 13: Sequencing depth and the cumulative depth of mosaic and full mutation samples

3.5 Variants Analysis

From the read alignment files (BAM), variations found in the samples were identified using GATK tool. The primary variant analysis resulted in ~0.4 to ~0.8 million variants from each studied sample (Tables 13-18). The distributions of identified variant types were shown in Figures 14 and 15. Figures 16 and 17 describe the location of the mutation in the human chromosomal level. Further identified variants were filtered based on the high read depth (depth > 10), mapping quality >20 and alignment quality >20 and obtained high confident exome SNPs for the downstream process.

Table 13: Variants found in five control samples without filter

Variants	C1	C2	C3	C4	C5
TVWF	576751	541898	541242	588155	635762
3PUTRV	9174	8493	9196	9677	10161
5PUTRV	4502	4047	4289	4672	4599
5UTRSGV	159	144	140	148	151
UGV	1	1	1	1	1
DGV	4	5	5	5	8
SO(C)	1	1	1	1	1
IntV	241336	225408	230933	247902	264640
IntgV	292944	275055	268411	297029	328000
SPRV	2485	2376	2446	2515	2470
SPDV	45	53	50	56	56
SPAV	100	87	99	99	93
ICV	38	35	33	26	34
STRV	10	12	12	12	14
SG	127	127	123	118	125
SL	15	17	18	13	16
DInfI	3	2	2	1	2
DInfD	6	5	8	4	4
InfI	203	189	193	198	203
InfD	221	207	198	217	225
FV	342	347	366	354	393
MV	12205	12380	12115	12453	12211
SV	12782	12835	12555	12624	12319

Table 14: Variants found in ten mosaic samples without filter

Variants	M1	M2	M3	M4	M5
TVWF	595135	772086	709581	743171	478441
3PUTRV	8047	11411	11096	11403	7213
5PUTRV	3862	4884	4671	4892	3496
5UTRSGV	129	144	153	157	112
UGV	1	1	1	2	NA
DGV	9	5	8	8	6
SO(C)	1	1	1	1	1
IntV	244306	313744	297657	308608	200884
IntgV	311130	412555	366450	388977	238511
SPRV	2403	2652	2524	2630	2394
SPDV	49	56	49	53	51
SPAV	94	103	105	108	94
ICV	29	40	26	37	26
STRV	8	12	15	10	12
SG	117	135	113	125	127
SL	19	17	17	16	19
DInfI	3	4	3	2	2
DInfD	5	6	8	6	6
InfI	186	229	189	191	207
InfD	220	247	224	239	224
FV	348	386	358	399	360
MV	11865	12479	12684	12506	12201
SV	12268	12927	13169	12771	12435

Table 14: Variants found in ten mosaic samples without filter (continued)

Variants	M6	M7	M8	M9	M10
TVWF	451071	731120	466433	518249	416895
3PUTRV	6897	10613	7271	7539	7054
5PUTRV	3619	4652	3827	3735	3747
5UTRSGV	130	149	121	116	136
UGV	2	1	1	1	1
DGV	4	5	5	4	5
SO(C)	1	1	1	1	1
IntV	192902	295703	198272	216259	184485
IntgV	219577	391804	229333	262705	190177
SPRV	2394	2521	2419	2338	2701
SPDV	40	47	45	44	51
SPAV	95	102	105	95	108
ICV	32	33	31	29	28
STRV	13	10	12	15	12
SG	129	137	114	124	141
SL	14	16	17	11	16
DInfI	3	3	3	2	3
DInfD	9	4	6	7	7
InfI	202	197	204	207	211
InfD	248	214	244	261	252
FV	365	375	362	350	384
MV	12038	12035	11889	12045	13417
SV	12303	12444	12091	12301	13928

Table 15: Variants found in thirteen full mutation samples without filter

Variants	F1	F2	F3	F4	F5	F6
TVWF	711166	488931	541718	663002	854509	517781
3PUTRV	10256	8299	8998	10084	12378	7703
5PUTRV	4617	3998	4295	4674	5503	3669
5UTRSGV	164	135	146	160	177	144
UGV	2	NA	NA	1	1	2
DGV	8	4	3	6	10	8
SO(C)	1	1	1	1	1	1
IntV	288701	205765	228612	273812	349025	219397
IntgV	379351	243218	271337	346586	455794	258753
SPRV	2513	2381	2411	2446	2777	2322
SPDV	49	44	55	50	51	49
SPAV	102	102	94	99	99	98
ICV	28	31	30	35	36	32
STRV	10	9	12	13	12	14
SG	123	119	124	115	124	128
SL	14	15	18	15	18	12
DInfI	2	2	2	3	5	2
DInfD	6	6	6	6	6	9
InfI	202	204	202	205	229	200
InfD	212	206	220	230	243	253
FV	338	335	333	341	416	380
MV	12063	11886	12229	11794	13500	12191
SV	12344	12153	12458	12272	14038	12372

Table 15: Variants found in thirteen full mutation samples without filter (continued)

Variants	F7	F8	F9	F10	F11	F12	F13
TVWF	447467	468208	414127	693562	469697	442696	548837
3PUTRV	7024	8032	7039	9830	7096	7527	7979
5PUTRV	3689	4228	3646	4501	3799	4125	3961
5UTRSGV	114	127	124	142	117	140	133
UGV	3	1	1	NA	2	2	2
DGV	6	5	5	7	5	4	4
SO(C)	1	1	1	1	1	1	1
IntV	191524	204359	182981	279404	200999	194846	228300
IntgV	217066	218267	191893	371620	229923	203327	280401
SPRV	2357	2840	2394	2414	2387	2758	2386
SPDV	53	54	48	53	45	51	50
SPAV	103	97	94	91	91	109	100
ICV	26	33	33	37	31	45	31
STRV	16	19	9	13	14	13	14
SG	128	134	118	129	132	136	119
SL	17	20	16	19	13	15	16
DInfI	2	3	2	1	1	6	4
DInfD	4	10	4	5	8	7	4
InfI	206	241	215	205	199	234	204
InfD	233	282	233	202	236	264	252
FV	382	406	393	358	344	357	369
MV	12130	13959	12292	12134	11934	14025	12012
SV	12287	15018	12544	12396	12284	14674	12417

Table 16: Variants found in five control samples with filter

Variants	C1	C2	C3	C4	C5
TVAF	125801	113590	126625	130539	135616
3PUTRV	3866	3525	4141	4178	4379
5PUTRV	2698	2332	2581	2798	2734
5UTRSGV	108	103	102	108	114
UGV	NA	NA	1	1	1
DGV	2	1	2	3	4
SO(C)	1	1	1	1	1
IntV	63908	56099	65011	66646	69974
IntgV	27323	23731	27344	28778	30845
SPRV	2295	2148	2228	2329	2313
SPDV	41	44	44	49	48
SPAV	88	79	89	92	85
ICV	38	34	32	26	34
STRV	9	12	12	11	14
SG	124	121	120	112	120
SL	15	16	17	12	16
DInfI	3	2	2	1	2
DInfD	6	5	8	4	4
InfI	185	172	183	180	188
InfD	207	193	186	206	211
FV	320	322	339	333	372
MV	11953	12039	11838	12200	11982
SV	12599	12569	12344	12465	12175

Table 17: Variants found in ten mosaic samples with filter

Variants	M1	M2	M3	M4	M5
TVAF	124589	145730	141964	147852	116806
3PUTRV	3695	4700	4625	4764	3476
5PUTRV	2490	2941	2796	2927	2205
5UTRSGV	98	98	112	106	79
UGV	1	NA	NA	NA	NA
DGV	6	3	3	3	3
SO(C)	1	1	1	1	1
IntV	59409	74773	74219	77414	55437
IntgV	32058	34413	31392	34101	28389
SPRV	2163	2522	2333	2474	2158
SPDV	43	49	41	41	43
SPAV	87	97	97	99	84
ICV	28	40	26	35	26
STRV	7	12	15	10	12
SG	116	124	110	125	122
SL	18	17	17	16	19
DInfI	3	4	3	2	2
DInfD	5	6	8	6	6
InfI	177	216	180	179	191
InfD	209	240	213	229	208
FV	330	370	333	378	338
MV	11605	12302	12420	12285	11844
SV	12034	12784	13002	12633	12157

Table 17: Variants found in ten mosaic samples with filter (continued)

Variants	M6	M7	M8	M9	M10
TVAF	113563	132314	116699	118752	124619
3PUTRV	3433	4181	3556	3510	3819
5PUTRV	2437	2725	2508	2410	2516
5UTRSGV	103	100	93	95	106
UGV	NA	1	NA	NA	NA
DGV	3	1	4	1	2
SO(C)	1	1	1	1	1
IntV	53960	66573	55261	56175	59904
IntgV	26685	31209	28663	29614	28090
SPRV	2123	2344	2117	2084	2416
SPDV	34	38	42	41	44
SPAV	89	97	97	89	93
ICV	32	33	29	26	27
STRV	13	10	12	14	12
SG	121	132	110	124	134
SL	12	16	15	11	16
DInfI	2	3	3	2	3
DInfD	8	4	6	7	7
InfI	191	186	190	193	190
InfD	229	201	230	249	237
FV	346	357	344	340	354
MV	11707	11810	11571	11715	13034
SV	12028	12280	11841	12027	13608

Table 18: Variants found in thirteen full mutation samples with filter

Variants	F1	F2	F3	F4	F5	F6
TVAF	127190	122231	122231	132443	160488	126344
3PUTRV	3936	8299	3855	4120	5137	3841
5PUTRV	2659	3998	2528	2724	3333	2419
5UTRSGV	104	135	110	109	111	115
UGV	1	N/A	NA	NA	NA	1
DGV	3	4	2	2	5	3
SO(C)	1	1	1	1	1	1
IntV	64018	205765	61594	67501	83179	61823
IntgV	29133	243218	26582	30924	37774	30845
SPRV	2314	2381	2216	2271	2604	2133
SPDV	38	44	48	47	44	44
SPAV	93	102	88	88	91	85
ICV	26	31	29	35	34	31
STRV	9	9	12	13	12	13
SG	119	119	122	112	120	125
SL	13	15	18	13	18	12
DInfI	2	2	2	3	5	2
DInfD	6	6	6	6	6	8
InfI	191	204	195	191	212	193
InfD	204	206	209	220	233	245
FV	317	335	311	324	391	361
MV	11815	11886	11977	11587	13256	11897
SV	12182	12153	12260	12134	13910	12141

Table 18: Variants found in thirteen full mutation samples with filter (continued)

Variants	F7	F8	F9	F10	F11	F12	F13
TVAF	110033	129488	117630	127923	118378	124778	125581
3PUTRV	3303	3932	3603	3948	3489	3778	3878
5PUTRV	2384	2825	2457	2638	2558	2784	2631
5UTRSGV	83	103	96	100	90	100	101
UGV	2	NA	NA	NA	NA	NA	2
DGV	2	2	3	3	1	3	2
SO(C)	1	1	1	1	1	1	1
IntV	51862	61444	56703	63702	56292	59119	60146
IntgV	25714	29561	27411	30061	29120	27689	31642
SPRV	2012	2448	2114	2240	2127	2388	2153
SPDV	47	46	41	49	42	45	43
SPAV	89	81	84	82	85	93	94
ICV	25	29	29	36	30	43	31
STRV	15	19	9	12	13	12	14
SG	119	131	118	127	131	126	116
SL	15	18	15	18	13	15	15
DInfI	2	3	2	1	1	5	4
DInfD	4	10	4	5	7	7	4
InfI	181	223	202	193	185	223	187
InfD	223	255	209	192	231	247	237
FV	355	383	364	340	329	326	356
MV	11667	13424	11915	11927	11615	13548	11729
SV	11910	14538	12232	12248	12012	14220	12183

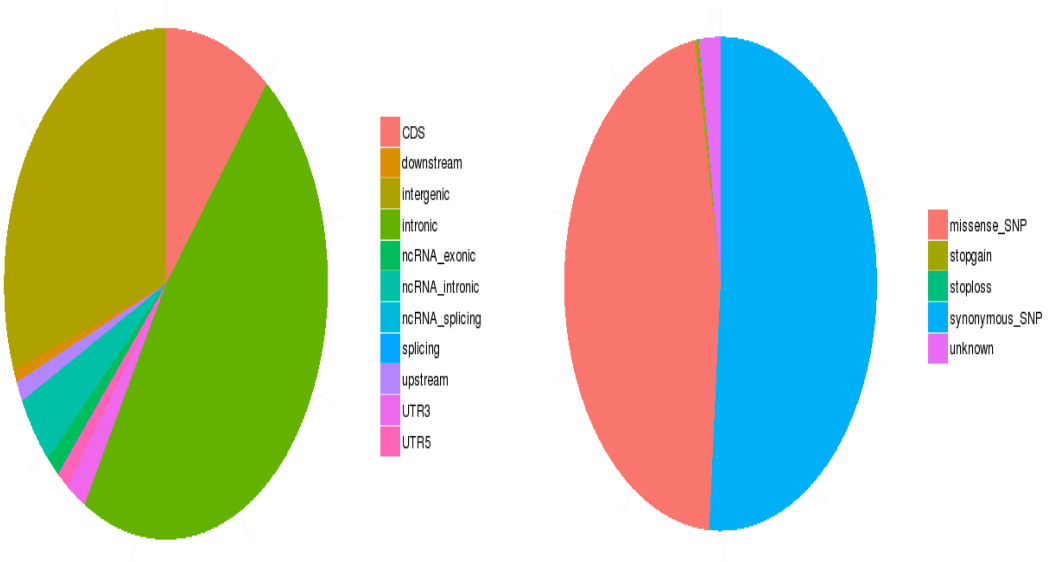

Figure 14: SNPs and other types of variants in mosaic samples

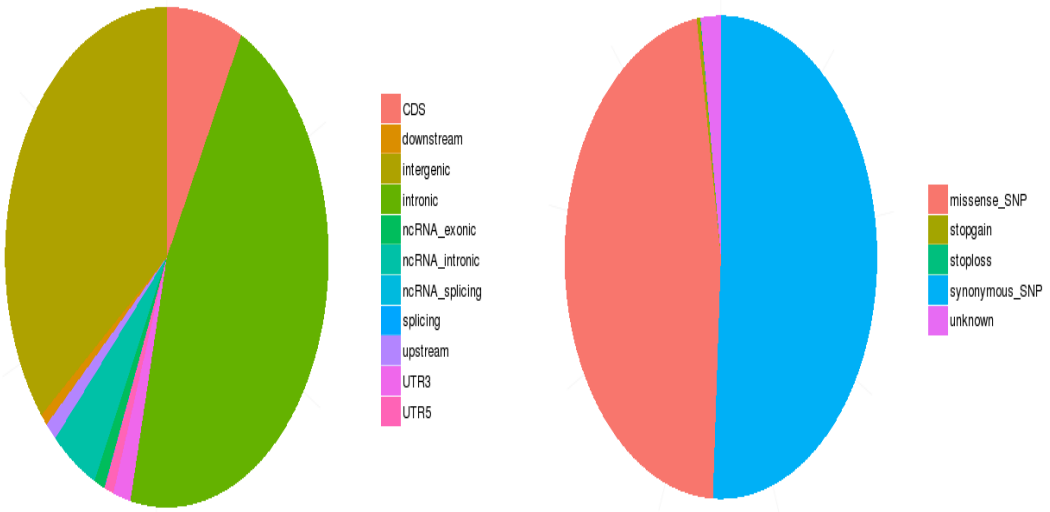

Figure 15: SNPs and other types of variants in full mutation samples

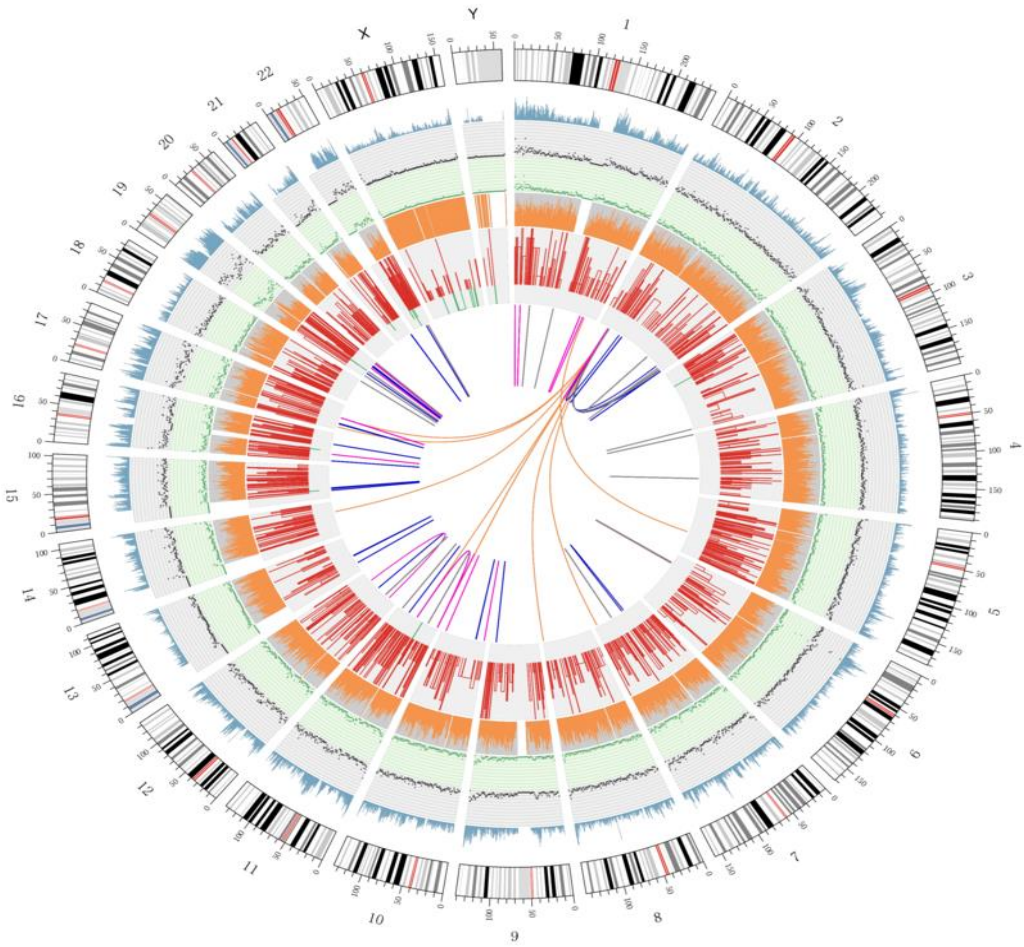

Figure 16: The whole genomic results of mosaic samples. It consists of seven rings. (a) The first (outer) ring has the chromosome information. (b) The second ring demonstrates the coverage of samples. (c) The Third ring represents the indels. (d) The fourth circle has SNPs information. The fifth circle represents homozygous SNP (orange) and heterozygous SNP (grey). The sixth circle represents CNV. The last circle demonstrates TRA (orange), INS (green), DEL (grey), DUP (pink) and INV (blue).

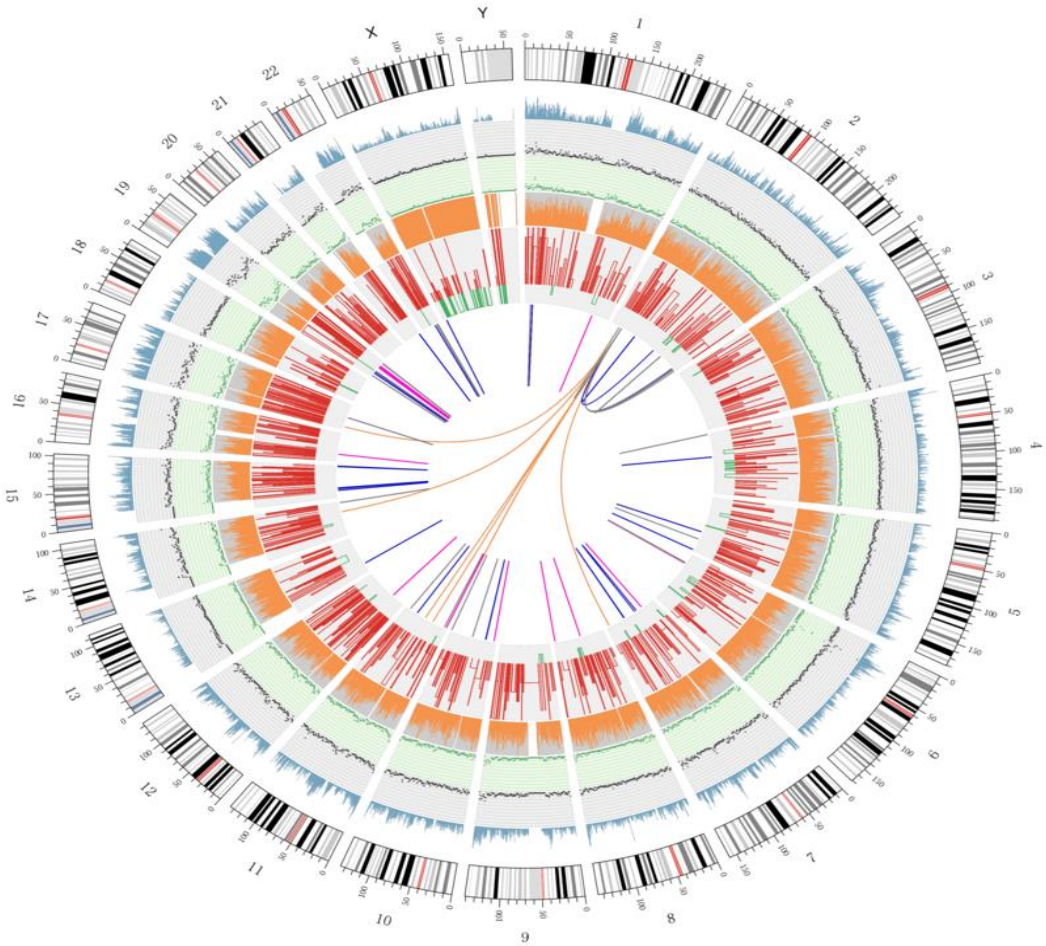

Figure 17: The whole genomic results of full mutation samples. It consists of seven rings. (a) The first (outer) ring has the chromosome information. (b) The second ring demonstrates the coverage of samples. (c) The Third ring represents the indels. (d) The fourth circle has SNPs information. The fifth circle represents homozygous SNP (orange) and heterozygous SNP (grey). The sixth circle represents CNV.The last circle demonstrates TRA (orange), INS (green), DEL (grey), DUP (pink) and INV (blue).

3.5.1 DNA and Histone Methyltransferase Genes Variants Analysis

From the whole exome variant analysis result, we filtered the variants which are present in the methyltransferase exome region, while filtering the frequency of the variant greater than 35% found in the mosaic and full mutation samples and not found in the control samples were considered as the significant mutations. Totally 7 significant variants were identified in the histone methyltransferase gene region (KMT2C and SMYD) (Table 19). Additional we found two more variants (found in EMHT1 and DOT1L gene), which are present in only mosaic samples (Table 20). All the significant variations were mapped in different chromosomal level in Figure 18.

1. Control (0%), Mosaic and Full mutation (>35%)

Table 14: Variant analysis results of histone methyltransferase genes (KMT2C) and (SMYD3).

Gene Name	Chr	Position	Ref	Alt	dbSNP	Variation type	Control % N=5	Mosaic% N=10	Full Mutation % N=13
SMYD3	1	246670298	CTT	-	N/A	Intron	0	6(60)	5(38.5)
KMT2C	7	151932747	C	T	N/A	Intron	0	5(50)	7(53.8)
KMT2C	7	151932748	A	G	N/A	Intron	0	5(50)	7(53.8)
KMT2C	7	151932756	A	T	N/A	Intron	0	7(70)	7(53.8)
KMT2C	7	151932774	A	G	N/A	Intron	0	7(70)	7(53.8)
KMT2C	7	151932824	A	T	N/A	Intron	0	7(70)	8(61.4)
KMT2C	7	151932876	G	T	N/A	Intron	0	5(50)	6(46.2)

2. Control (0%), Mosaic (≥50%) and Full mutation (0%)

Table 15: Variant analysis results of histone methyltransferase genes (EHMT1 and DOT1L).

Gene Name	Chr	Position	Ref	Alt	dbSNP	Variation Type	Control % N=5	Mosaic % N=10	Full Mutation % N=13
EHMT1	9	140611672	A	G	rs72766927	Intron	0	5(50)	0
DOT1L	19	2194661	-	TGTTGGC ACATGGC	N/A	Intron	0	5(50)	0

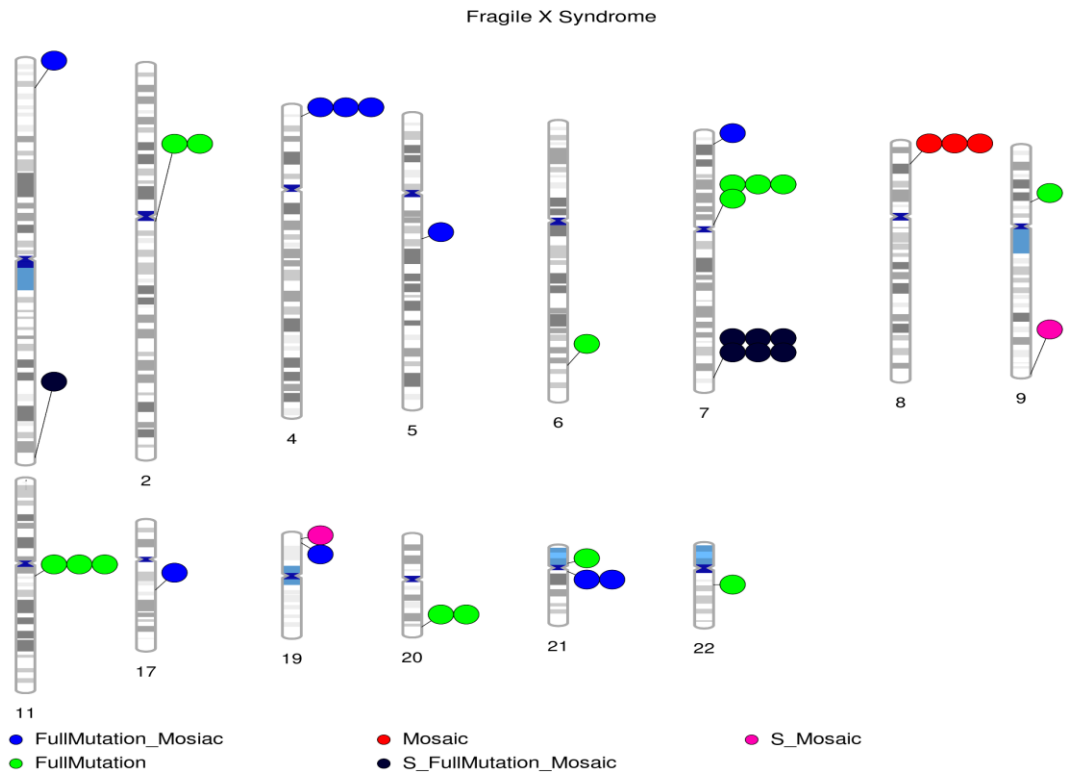

Figure 18: The position of the variation on genes in different chromosomes. a) Blue indicates the full mutation and mosaic variation in whole exome genes. b) Green the variation in the full mutation only. c) Red is the variation of the mosaic variation in whole exome genes. d) Black is the variation of the mosaic and full mutation samples in the histone methyltransferase genes. e) Pink is the variation in histone methyltransferase genes of the mosaic samples.

3.5.2 Other Genes Variant Analysis

From the whole exome variant results, we filtered the unique variations specific to mosaic and full mutation samples. During filtration, a variant occurred in more than 70% of mosaic samples, occurred more than 65% in full mutation samples and not found in the control samples were considered as significant mutations. Eleven significant mutations were identified in all studied samples; from that 6 mutations are already reported in the dbSNP database and 5 novel variations were found (Table 21). Likewise, Table 22 describes the significant unique variation found in the mosaic samples (not found in control and full mutation sample) and Table 23 describes the unique significant mutation found in the full mutation samples.

1. Control (0%), Mosaic and Full mutation (>69%)

Table 16: Whole exome sequencing results of Intergenic region, and two different genes.

Gene Name	Chr	Position	Ref	Alt	dbSNP	Variation Type	Control % N=5	Mosaic %N=10	Full mutation% N=13
N/A	1	16952703	C	T	N/A	Intergenic	0	9(90)	11(84.6)
EVC2	4	5617295	T	C	rs10025164	Intron	0	9(90)	12(92.3)
EVC2	4	5617369	G	T	rs10032860	Intron	0	9(90)	12(92.3)
EVC2	4	5624670	T	C	rs730469	Missense	0	9(90)	13(100)
N/A	5	76442651	G	A	rs6863608	Intergenic	0	9(90)	9(69.2)
N/A	7	6971266	A	G	N/A	Intergenic	0	9(90)	9(69.2)
N/A	17	43679861	TTTC	-	rs555779317	Intergenic	0	10(100)	9(69.2)
UHRF1	19	4945914	A	C	rs2250982	Synonymous	0	9(90)	9(69.2)
N/A	21	15281827	G	A	N/A	Intergenic	0	9(90)	10(76.9)
N/A	21	15281829	TG	-	N/A	Intergenic	0	9(90)	11(84.6)

2. Control (0%), Mosaic (≥70%) and Full mutation (0%)

Table 17: Whole exome sequencing results of KIAA1456 gene

Gene Name	Chr	position	Ref	Alt	dbSNP	Variation Type	Control% N=5	Mosaic% N=10	Full mutation% N=13
KIAA1456	8	12848221	T	C	rs36056654	Intron	0	7(70)	0
KIAA1456	8	12863700	G	C	rs35757493	Intron	0	7(70)	0
KIAA1456	8	12870186	C	G	rs12156420	Splice region	0	7(70)	0

3. Control (0%), Mosaic (0%) and Full mutation (>50%)

Table 18: Whole exome sequencing results of several genes and intergenic regions

Gene Name	Chr	position	Ref	Alt	dbSNP	Variation Type	Control% N=5	Mosaic% N=10	Full mutation% N=13
ANKRD36C	2	96585703	A	G	N/A	Intron	0	0	7(53.8)
FAM124B	2	225244923	A	G	rs3738953	Synonymous	0	0	7(53.8)
RAET1G	6	150244217	GTCTGAATGCAGCCC	-	rs71656790	5PUTRV	0	0	7(53.8)
N/A	7	56893989	C	G	rs372462579	Intergenic	0	0	7(53.8)
N/A	7	63041333	G	A	N/A	Intergenic	0	0	8(61.4)
SAMD9L	7	92762681	A	G	rs1029357	Synonymous	0	0	7(53.8)
EPHB6	7	142567942	A	G	rs4987691	Intron	0	0	7(53.8)
PRSS3	9	33796927	A	G	N/A	Intron	0	0	7(53.8)
DAGLA	11	61490880	C	A	rs9735635	Intron	0	0	7(53.8)
DAGLA	11	61505583	G	A	rs2240287	Intron	0	0	7(53.8)
CEP295	11	93454832	-	GT	N/A	Intron	0	0	7(53.8)
CDH26	20	58581863	G	C	rs195004	Intron	0	0	7(53.8)
CDH26	20	58581873	A	G	rs195005	Intron	0	0	7(53.8)
N/A	21	9911892	T	C	N/A	Intergenic	0	0	7(53.8)
SGSM1	22	25289335	G	C	rs3765480	Intron	0	0	8(61.4)

Chapter 4: Discussion

4.1 DNA and Histone Methyltransferase Genes Variants Analysis

4.1.1 Control (0%), Mosaic and Full mutation (>35%)

The DNA and histone methyltransferase genes are known to be involved in gene regulation. There were observable variations found in >35% mosaic and full mutation samples in two genes. The genes are lysine methyltransferase 2C (KMT2C) and SET and MYND domain-containing protein 3 (SMYD3) genes. Six variants were identified in the intron region of KMT2C genr. These variants are novel, and not reported in any of the databases. KMT2C (MLL3) is a member of the myeloid/lymphoid or mixed-lineage leukemia family (Chen et al., 2019). It's involved in monomethylation of H3K4 at cell type specific distal enhancers, acts as tumor repressor and regulates gene expression by modifying chromatin structure. KMT2C gene mutations are associated with multiple human cancer such as breast, endometrial, lung, large intestine and bladder carcinoma (Rao and Dou, 2015). A de novo mutations in KMT2C were found to be associated with intellectual disability and autism spectrum disorder, having the same clinical features and phenotype that resembles other disorders such as, Kleefstra syndrome, which is caused by EHMT1 mutations (Koemans et al., 2017). Moreover, there is one novel intron variation (deletion) in SMYD3 gene. SMYD3 protein form a transcriptional complex with RNA polymerase 2, that acts as a transcriptional factor by regulating the downstream genes and its suppression inhibits the growth of colorectal and hepatocellular carcinoma (Hamamoto et al., 2004).

4.1.2 Control (0%), Mosaic (≥50%) and Full mutation (0%)

There are two genes of lysine methyltransferase genes family (EHMT1 and DOT1L) contains an intron variant that only observed in 50% of Mosaic individuals. EHMT1gene is a Euchromatic Histone Lysine Methyltransferase 1 that regulates gene expression, important for normal neural development and growth. Alteration in EHMT1 gene results in Kleefstra syndrome which described previously (Koemans et al., 2017) and loss of function of EHMT1 results in 9q subtelomeric deletion syndrome which exhibits physical and behavior features such as, heart defects, flat face and mental retardation (Kleefstra et al., 2006). This gene has reported intron deletion has been reported (rs72766927) with no publication on the variant. DOT1L gene has a novel intron variant. It's involved in DNA damage response, gene regulation, cell progression and in embryonic development. Alteration in the gene is associated with leukemia, cartilage thickness and hip osteoarthritis (Betancourt et al., 2012).

4.2 Other Genes Variant Analysis

After analyzing the DNA and histone methylation for gene variation differences, we had analyzed the whole exome genes for each group (control, mosaic, and full). We included variants that fit the following criteria (Table 22 - Table 23).

4.2.1 Control (0%), Mosaic and Full mutation (>69%)

A. EVC2 and UHRF1 genes

EvC ciliary complex subunit 2 (EVC2) gene produce cilia proteins that have an N-terminal anchored transmembrane protein and a coiled structure. EVC and EVC2 genes form a protein complex at the base of the primary cilium which is necessary for

ciliary localization (Caparrós-Martín et al., 2012). These two genes considered to be the cause of Ellis-van Creveld syndrome (EvC). The clinical features of EvC patients are dwarfism, polydactyly, cardiovascular malformations, shorter limbs and ribs, hypomorphic nails, abnormal tooth and craniofacial development (Baujat et al., 2007; Kwon et al., 2018). It is responsible for bone development. Three variants were found in EVC2 gene, two reported in intron variants (rs10025164 & rs10032860) and one reported as missense variant (rs730469) in dbSNP (Table 22). We have observed variations in introns and coding regions in ≥ 90% in mosaic and full mutation. These finding requires further analysis.

UHRF1 (Ubiquitin-like with PHD and Ring Finger domains 1) gene is a multi-domain nuclear protein which regulates the epigenetic modification by histone markers recognition, heterochromatin formation and in maintenance of DNA methylation, and facilities the binding of DNMT1 to the new synthesized DNA strands to carry its function of transmitting the epigenetic information from cell to cell during replication. It also has a role in DNA damage repair (Kim et al., 2018; Hahm et al., 2018). A synonymous variation was found in the gene and is reported (rs2250982) in dbSNP (Table 22)

B. Intergenic regions

We observed seven intergenic variants were found, two are already reported and the others are novel SNPs (Table 22). The intergenic variant in chromosome 5 is reported (rs6863608) in dbSNP. According to dbSNP, this is believed to be ZBED3-AS1 which is long noncoding RNA has a role in regulating the chondrogenic differentiation in early stages (Wang et al., 2015). Other intergenic variant in

chromosome 17 (rs555779317) reports in dbSNP and gene consequence is Mitogen-Activated Protein Kinase 8 Interacting Protein 1 Pseudogene 2 (MAPK8IP1P2) with no publication on its function.

4.2.2 Control (0%), Mosaic (≥70%) and Full mutation (0%)

There are three intron variants were found in KIAA1456 (TRMT9B), which are reported in dbSNP, only one had a splice region variant. However, these variant were only present in more than 70% of mosaic and are not exist in control and full mutation. This gene is tRNA methyltransferase gene, which has a potential role in tumor repressing and in the stress signaling pathway.

4.2.3 Control (0%), Mosaic (0%) and Full mutation (>50%)

A. Multiple gene variations

Variations in twelve genes were found, nine of which were reported in dbSNP. The rest had novel variations Table 23. A novel intron variant in Ankyrin Repeat Domain 36C (ANKRD36C) gene was found. This gene has an unknown function, although it is associated with cancer. A synonymous in FAM124B gene variation already reported SNP (rs3738953) was also found in our samples. FAM124B gene is nuclear protein, found to be interacting and serving as a binding factor to two chromodomain helicase DNA binding proteins CHD7 and CHD8 which function in multi-protein complex that controls the gene expression by its association chromatin remolding (Batsukh et al., 2012). Mutations or malformation in CHD7 and CHD8 gene or proteins respectively are assumed to be involved in CHARGE Syndrome, neurodevelopmental (NDD) and autism spectrum disorders (ASD) (Zahir et al., 2007; Talkowski et al, 2012). The CHARGE syndrome individuals exhibit different clinical

features and behavior such as scoliosis, intellectual disability, ears abnormalities, heart defect, and cleft palate. etc (Sanlaville and Verloes, 2007; Blake and Prasad, 2006). We also found reported variants in retinoic acid early transcript 1G (RAET1G) gene in chromosome 6. It produces RAET1G protein which natural killer group 2, member D (NKG2D) receptor's ligand that initiates the immune response (innate and adaptive immunity) (Ohashi et al., 2010).

On another note, sterile alpha motif domain containing 9 like SAMD9L gene is involved in innate pathogen response (Lemos et al., 2013). Mutation in this gene is associated with several syndromes such as ataxia–pancytopenia (ATXPC) syndrome, MIRAGE syndrome myelodysplastic syndrome and leukemia syndrome with monosomy 7 syndrome (Davidsson et al., 2018). A synonymous variation was found during analysis and found to be reported in dbSNP (rs1029357). EPH Receptor B6 (EPHB6) is the largest tyrosine kinases family in humans, it has an affinity to ephrin ligand and it is involved in angiogenesis, axon guidance and hindbrain patterning. It is highly expressed in an advanced stage of tongue squamous cell carcinoma (Dong et al., 2015). An intron variant was found and it is reported previously(rs4987691). In addition, we found variation in serine protease3 which is an isoform of trypsinogen, it is secreted by pancreatic acinar cells into the small intestine to induce the digestion process (Qian et al., 2017). It's up-regulated in a different type of cancer, promotes their metastasis and growth (Wang et al., 2019). A novel SNP variation was found in the intron region of the gene. We also report two intron variants already reported (rs9735635 and rs2240287) in diacylglycerol lipase alpha (DAGLA) gene which regulates the central nervous system by promoting the axonal growth and the migration of new neurons (Reisenberg et al., 2012).

Centrosomal protein 295 (CEP295) gene has critical role in cell progression, conversion of centriole to during mitosis, generation of the distal half of new centriole, the assembly of centriolar proteins and in centriole elongation (Chang et al., 2016). Novel intron variant was found in the (CEP295) gene. The Cadherin 26CDH2 (CDH26) gene found to have two reported introns (rs195004 and s195005) variants. Its protein is localized in the stomach, epithelial cells and in the irritated esophagus. It regulates the immune activity and required for calcium-dependent cell adhesion (Caldwell et al., 2017). The Small G Protein Signaling Modulator 1 (SGSM1 gene) is localized in trans-Golgi network in neurons cells of the central nervous system (Yang et al., 2007). It acts as a modulator in two associated pathway of different G proteins a) intracellular signal transduction such as, regulating cell differentiation polarity, proliferation, secretion, movements and adhesion which are important for synaptic plasticity, neuron migrations and growth (Gloerich and Bos, 2011; Spilker et al., 2010), and b) vesicle transportation by RAP family and RAB family respectively in the brain (Yang et al., 2007). In all genes described previously, 50% of full mutation individuals were found to have the variants in those genes. Interestingly, these genes are involved in physical or behavioral issues not unknown in Fragile X.

B. Intergenic Region

There are three variants with unknown function or name, one is reported (rs3765480) and the others are novels having 70 to 80% of mosaic individuals.

DNA methyltransferase and arginine methyltransferase genes weren’t shown to be associated with fragile X syndrome (no variations were found between groups). Two histone lysine methyltransferase (KMT2C and SMYD3) intron variants existed in more than 35% of mosaic and full mutation individuals which means the variant

could affect the expression of these genes manifesting an affected individual. Additional genes were analyzed for their significant variants in both mosaic and full mutation. ECV2 gene with missense variants had 100% prevalence in full mutation and 90% in mosaic individuals. This gene is involved in bone formation and is associated with EvC syndrome. These variants might have other effect that might not involve the methylation levels in mosaic and full mutation group. Other intronic and intergenic variations were significant too in both individuals. Total variations we observed in different locations on the genome (intron, intergenic or coding region) in both mosaic and full mutation could be associated with different methylation levels in fragile X syndrome and other background genes might be involved in determining other phenotypes associated with FXS. Most genetic variations were in introns and intergenic variation which shows that the introns and intergenic region might have a significant role regulating the expression of the methylation levels and other biological function in fragile X syndrome. Intronic studies have shown the role of the introns in enhancing the gene expression (Chorev and Carmel, 2012; Jo and Choi, 2015), splicing, mRNA transport, and as genome protector against random mutations (Jo and Choi, 2015). Intergenic regions, contains many of noncoding RNAs that function in regulating the gene expression, protein biosynthesis, and act as catalytic molecules. Pseudogene is part of intergenic region that also acts as a regulator of gene expression and their deregulation could contribute to a disease. More studies are needed to identify the intergenic variations that are present in fragile X syndrome in both individuals as most of the intergenic region in this study was unknown except for two intergenic regions that had been identified in dbSNP; (ZBED3-AS1 long non coding RNA regulates the differentiation of chondrogenic during embryogenesis and MAPK8IP1P2

which has unknown function. More samples are needed to confirm the results and genome sequencing should be conducted to look for the intergenic and the intronic regions due to whole exome sequencing limitation which is only and more efficient for the coding regions (exons). Then, the studies of variations presented in both individuals are carried out to demonstrate its effect.

Chapter 5: Conclusion

Fragile X syndrome is a genetically inherited, they express different behavior, clinical and physical features. The cause of different methylation in fragile X syndrome is not yet well understood. Accordingly, we hypothesized that DNA and histone methyltransferase genes could be associated with different methylation levels in fragile X syndrome individuals and we believe other background genes are also involved in the syndrome. In this study, we identified genetic variation in DNA and histone methyltransferase genes among other genes. We presume that introns and intergenic regions has major role in gene expression such as, methylation levels related to fragile X syndrome as most of the variations were in the intronic and intergenic regions. More studies must be done on the intron and intergenic regions to discover their involvement in the methylation levels in FXS. This preliminary study, will help the researchers to understand more about the genetic variation associated with different fragile X syndrome condition that might explain the variation in symptoms within FXS individuals. In the future, the whole genome-based genetic analysis approach will pave the path towards more understanding about the methylation process in fragile X syndrome condition.

References

Ascano, M., Mukherjee, N., Bandaru, P., Miller, J. B., Nusbaum, J., Corcoran, D. L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M., Williams, Z., & Tuschl, T. (2012). FMR1 targets distinct mRNA sequence elements to regulate protein expression. Nature, 492(7429), 382-386.

Barasoain, M., Barrenetxea, G., Huerta, I., Télez, M., Criado, B., & Arrieta, I. (2016). Study of the Genetic Etiology of Primary Ovarian Insufficiency: FMR1 Gene. Genes, 7(12), 123.

Batsukh, T., Schulz, Y., Wolf, S., Rabe, T. I., Oellerich, T., Urlaub, H., Schaefer, I.M. & Pauli, S. (2012). Identification and characterization of FAM124B as a novel component of a CHD7 and CHD8 containing complex. PloS One, 7(12), e52640.

Baujat, G., Le Merrer, M. (2007). Ellis-Van Creveld syndrome. Orphanet J Rare Dis 2, 27, 1750-1172

Bedford, M. T., & Clarke, S. G. (2009). Protein arginine methylation in mammals: who, what, and why. Molecular Cell, 33(1), 1–13.

Bedford, M. T., & Richard, S. (2005). Arginine methylation: an emerging regulatorof protein function. Molecular Cell, 18(3), 263-272.

Begley, U., Sosa, M. S., Avivar-Valderas, A., Patil, A., Endres, L., Estrada, Y., Chan, C.T., Su, D., Dedon, P.C., Aguirre-Ghiso, J.A. & Begley, T. (2013). A human tRNA methyltransferase 9-like protein prevents tumour growth by regulating LIN9 and HIF1-α. EMBO Molecular Medicine, 5(3), 366-383.

Bestor, T. H. (2000). The DNA methyltransferases of mammals. Human molecular Genetics, 9 (16), 2395-2402.

Betancourt, M. C. C., Cailotto, F., Kerkhof, H. J., Cornelis, F. M., Doherty, S. A., Hart, D. J., Hofman, A., Luyten, F.P., Maciewicz, R.A., Mangino, M. and Metrustry, S.& Metrustry, S. (2012). Genome-wide association and functional studies identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. Proceedings of the National Academy of Sciences, 109(21), 8218-8223.

Blackwell, E., Zhang, X., & Ceman, S. (2010). Arginines of the RGG box regulate FMRP association with polyribosomes and mRNA. Human Molecular Genetics, 19(7), 1314-1323.

Blake, K. D., & Prasad, C. (2006). CHARGE syndrome. Orphanet journal of rare diseases, 1, 34, 1750-1172.

Brown, S. S., & Stanfield, A. C. (2015). Fragile X premutation carriers: a systematic review of neuroimaging findings. Journal of the Neurological Sciences, 352(1-2), 19-28.

Budworth, H., & McMurray, C. T. (2013). Bidirectional transcription of trinucleotide repeats: roles for excision repair. DNA Repair, 12(8), 672-684.

Caldwell, J. M., Collins, M. H., Kemme, K. A., Sherrill, J. D., Wen, T., Rochman, M., Rothenberg, M. E. (2017). Cadherin 26 is an alpha integrin-binding epithelial receptor regulated during allergic inflammation. Mucosal Immunology, 10(5), 1190–1201.

Caparrós-Martín, J. A., Valencia, M., Reytor, E., Pacheco, M., Fernandez, M., Perez-Aytes, A., Gean, E., Lapunzina, P., Peters, H., Goodship, J.A. & Ruiz-Perez, V. L. (2012). The ciliary Evc/Evc2 complex interacts with Smo and controls Hedgehog pathway activity in chondrocytes by regulating Sufu/Gli3 dissociation and Gli3 trafficking in primary cilia. Human Molecular Genetics, 22(1), 124-139.

Chang, C. W., Hsu, W. B., Tsai, J. J., Tang, C. J. C., & Tang, T. K. (2016). CEP295 interacts with microtubules and is required for centriole elongation. J Cell Sci, 129(13), 2501-2513.

Chen, T., & Li, E. (2004). Structure and function of eukaryotic DNA methyltransferases. In Current Topics in Developmental Biology, 60, 55-89.

Chen, X., Zhang, G., Chen, B., Wang, Y., Guo, L., Cao, L., Ren, C., Wen, L. & Liao, N. (2019). Association between histone lysine methyltransferase KMT2C mutation and clinicopathological factors in breast cancer. Biomedicine & Pharmacotherapy, 116, 108997.

Cheng, X., & Blumenthal, R. M. (2008). Mammalian DNA methyltransferases: a structural perspective. Structure, 16(3), 341-350.

Chorev, M., & Carmel, L. (2012). The function of introns. Frontiers in genetics, 3, 55.

Cong, T., Liu, G. X., Cui, J. X., Zhang, K. C., Chen, Z. D., Chen, L., Wei, B. & Huang, X. H. (2018). Exome sequencing of gastric cancers screened the differences of clinicopathological phenotypes between the mutant and the wide-type of frequently mutated genes. Zhonghua yi xue za zhi, 98(28), 2242-2245.

Davidsson, J., Puschmann, A., Tedgård, U., Bryder, D., Nilsson, L., & Cammenga, J. (2018). SAMD9 and SAMD9L in inherited predisposition to ataxia, pancytopenia, and myeloid malignancies. Leukemia, 32(5), 1106-1115.

Davis, J. K., & Broadie, K. (2017). Multifarious Functions of the Fragile X Mental Retardation Protein. Trends in Genetics: TIG, 33(10), 703-714.

Dillon, S. C., Zhang, X., Trievel, R. C., & Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biology, 6(8), 227.

Dockendorff, T. C., & Labrador, M. (2019). The Fragile X protein and genome function. Molecular Neurobiology, 56(1), 711-721.

Dong, Y., Pan, J., Ni, Y., Huang, X., Chen, X., & Wang, J. (2015). High expression of EphB6 protein in tongue squamous cell carcinoma is associated with a poor outcome. International Journal of Clinical and Experimental Pathology, 8(9), 11428-11433.

Fatemi, S. H., & Folsom, T. D. (2011). The role of fragile X mental retardation protein in major mental disorders. Neuropharmacology, 60(7-8), 1221-1226.

Fernandez-Carvajal, I., Lopez Posadas, B., Pan, R., Raske, C., Hagerman, P. J., & Tassone, F. (2009). Expansion of an FMR1 Grey-Zone Allele to a Full Mutation in Two Generations. The Journal of Molecular Diagnostics: JMD, 11(4), 306-310.

Garber, K. B., Visootsak, J., & Warren, S. T. (2008). Fragile X syndrome. European Journal of Human Genetics: EJHG, 16(6), 666-672.

Gloerich, M., & Bos, J. L. (2011). Regulating Rap Small G-proteins in time and space. Trends in Cell Biology, 21(10), 615-623.

Godler, D. E., Tassone, F., Loesch, D. Z., Taylor, A. K., Gehling, F., Hagerman, R. J., Burgess, T., Ganesamoorthy, D., Hennerich, D., Gordon, L. & Evans, A. (2010). Methylation of novel markers of fragile X alleles is inversely correlated with FMRP expression and FMR1 activation ratio. Human Molecular Genetics, 19(8), 1618-1632.

Hagerman, P. J., & Hagerman, R. J. (2015). Fragile X-associated tremor/ataxia syndrome. Annals of the New York Academy of Sciences, 1338(1), 58-70.

Hagerman, R. J., Berry-Kravis, E., Hazlett, H. C., Bailey Jr, D. B., Moine, H., Kooy, R. F., Tassone, F., Gantois, I., Sonenberg, N., Mandel, J.L. & Hagerman, P. J. (2017). Fragile X syndrome. Nature reviews Disease Primers, 3, 17065.

Hagerman, R. J., Berry-Kravis, E., Kaufmann, W. E., Ono, M. Y., Tartaglia, N., Lachiewicz, A., Tranfaglia, M. (2009). Advances in the Treatment of Fragile X Syndrome. Pediatrics, 123(1), 378-390.

Hahm, J. Y., Kim, J. Y., Park, J. W., Kang, J. Y., Kim, K. B., Kim, S. R., Cho, H. & Seo, S. B. (2018). Methylation of UHRF1 by SET7 is essential for DNA double-strand break repair. Nucleic Acids Research, 47(1), 184-196.

Halevy, T., Czech, C., & Benvenisty, N. (2015). Molecular Mechanisms Regulating the Defects in Fragile X Syndrome Neurons Derived from Human Pluripotent Stem Cells. Stem Cell Reports, 4(1), 37-46.

Hall, D. A., Birch, R. C., Anheim, M., Jønch, A. E., Pintado, E., O'Keefe, J., Leehey, M. A. (2014). Emerging topics in FXTAS. Journal of Neurodevelopmental Disorders, 6(31), 1866-1955.

Hamamoto, R., Furukawa, Y., Morita, M. et al. (2004) SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. Nat Cell Biol 6, 731-740.

Handt, M., Epplen, A., Hoffjan, S., Mese, K., Epplen, J. T., & Dekomien, G. (2014). Point mutation frequency in the FMR1 gene as revealed by fragile X syndrome screening. Molecular and Cellular Probes, 28(5-6), 279-283.

He, R. Q., Wei, Q. J., Tang, R. X., Chen, W. J., Yang, X., Peng, Z. G., … Chen, G. (2017). Prediction of clinical outcome and survival in soft-tissue sarcoma using a ten-lncRNA signature. Oncotarget, 8(46), 80336-80347.

Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: overviews and challenges. BioTechniques, 56(2), 61-77

Houtgast, E. J., Sima, V. M., Bertels, K., & Al-Ars, Z. (2015). An FPGA-based systolic array to accelerate the BWA-MEM genomic mapping algorithm. International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS) 221-227.

Hu, K, Jiang, W, Sun, H, Li, Z, Rong, G, Yin, Z. (2019). Long noncoding RNA ZBED3-AS1 induces the differentiation of mesenchymal stem cells and enhances bone regeneration by repressing IL-1β via Wnt/β-catenin signaling pathway. J Cell Physiol. 234: 17863-17875.

Jeltsch, A., & Jurkowska, R. Z. (2016). Allosteric control of mammalian DNA methyltransferases–a new regulatory paradigm. Nucleic acids research, 44(18), 8556-8575.

Jin, B., & Robertson, K. D. (2013). DNA Methyltransferases (DNMTs), DNA Damage Repair, and Cancer. Advances in Experimental Medicine and Biology, 754, 3-29.

Jin, B., Li, Y., & Robertson, K. D. (2011). DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy? Genes & Cancer, 2(6), 607-617.

Jiraanont, P., Kumar, M., Tang, H. T., Espinal, G., Hagerman, P. J., Hagerman, R. J., Chutabhakdikul, N., Tassone, F. (2017). Size and methylation mosaicism in males with Fragile X syndrome. Expert Review of Molecular Diagnostics, 17(11), 1023-1032.

Jo, B. S., & Choi, S. S. (2015). Introns: The Functional Benefits of Introns in Genomes. Genomics & Informatics, 13(4), 112-118.

Kar, S., Deb, M., Sengupta, D., Shilpi, A., Parbin, S., Torrisani, J., … Patra, S. (2012). An insight into the various regulatory mechanisms modulating human DNA methyltransferase 1 stability and function. Epigenetics, 7(9), 994-1007.

Kim, K. Y., Tanaka, Y., Su, J., Cakir, B., Xiang, Y., Patterson, B., Ding, J., Jung, Y.W., Kim, J.H., Hysolli, E. & Lee, H. (2018). Uhrf1 regulates active transcriptional marks at bivalent domains in pluripotent stem cells through Setd1a. Nature Communications, 9(1), 2583.

Kleefstra, T., Brunner, H. G., Amiel, J., Oudakker, A. R., Nillesen, W. M., Magee, A., Geneviève, D., Cormier-Daire, V., Van Esch, H., Fryns, J.P. & Hamel, B. C. (2006). Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. The American Journal of Human Genetics, 79(2), 370-377.

Koemans, T. S., Kleefstra, T., Chubak, M. C., Stone, M. H., Reijnders, M. R., de Munnik, S., Willemsen, M.H., Fenckova, M., Stumpel, C.T., Bok, L.A. & Saenz, M. S. (2017). Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder. PLoS Genetics, 13(10), e1006864.

Kraan, C. M., Godler, D. E., & Amor, D. J. (2019). Epigenetics of fragile X syndrome and fragile X-related disorders. Developmental Medicine & Child Neurology, 61(2), 121-127.

Kwon, E. K., Louie, K. A., Kulkarni, A., Yatabe, M., Ruellas, A. C. D. O., Snider, T. N., Mochida, Y., Cevidanes, L.H., Mishina, Y. & Zhang, H. (2018). The Role of Ellis-Van Creveld 2 (EVC2) in Mice During Cranial Bone Development. The Anatomical Record, 301(1), 46-55.

LaFauci, G., Adayev, T., Kascsak, R., & Brown, W. T. (2016). Detection and Quantification of the Fragile X Mental Retardation Protein 1 (FMRP). Genes, 7(12), 121.

Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J. & Jang, W. (2015). ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Research, 44(D1), D862-D868.

Lei, L., Lin, H., Zhong, S., Zhang, Z., Chen, J., Yu, X., Liu, X., Zhang, C., Nie, Z. & Zhuang, J. (2017). DNA methyltransferase 1 rs16999593 genetic polymorphism decreases risk in patients with transposition of great arteries. Gene, 615, 50-56.

Lemos de Matos, A., Liu, J., McFadden, G., & Esteves, P. J. (2013). Evolution and divergence of the mammalian SAMD9/SAMD9L gene family. BMC Evolutionary Biology, 13(121), 1471-2148.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics, 25(16), 2078-2079.

Li, H., Liu, J. W., Sun, L. P., & Yuan, Y. (2017). A meta-analysis of the association between DNMT1 polymorphisms and cancer risk. BioMed Research International, 2017, 3971259.

Loesch, D. Z., Godler, D. E., Evans, A., Bui, Q. M., Gehling, F., Kotschet, K. E., Horne, M. (2011). Evidence for the toxicity of bidirectional transcripts and mitochondrial dysfunction in blood associated with small CGG expansions in the FMR1 gene in patients with parkinsonism. Genetics in medicine: Official Journal of the American College of Medical Genetics, 13(5), 392-399.

Lozano, R., Azarang, A., Wilaisakditipakorn, T., & Hagerman, R. J. (2016). Fragile X syndrome: A review of clinical management. Intractable & Rare Diseases Research, 5(3), 145-157.

Lozano, R., Rosero, C. A., & Hagerman, R. J. (2014). Fragile X spectrum disorders. Intractable & Rare Diseases Research, 3(4), 134-146.

McLean, C. M., Karemaker, I. D., & Van Leeuwen, F. (2014). The emerging roles of DOT1L in leukemia and normal development. Leukemia, 28(11), 2131-2138.

Myrick, L. K., Hashimoto, H., Cheng, X., & Warren, S. T. (2014 ((a)). Human FMRP contains an integral tandem Agenet (Tudor) and KH motif in the amino terminal domain. Human Molecular Genetics, 24(6), 1733-1740.

Myrick, L. K., Nakamoto-Kinoshita, M., Lindor, N. M., Kirmani, S., Cheng, X., & Warren, S. T. (2014 (b)). Fragile X syndrome due to a missense mutation. European Journal of Human Genetics: EJHG, 22(10), 1185-1189.

Nolin, S. L., Brown, W. T., Glicksman, A., Houck Jr, G. E., Gargano, A. D., Sullivan, A., & Kooy, F. (2003). Expansion of the fragile X CGG repeat in females with premutation or intermediate alleles. The American Journal of Human Genetics, 72(2), 454-464.

Ohashi, M., Eagle, R. A., & Trowsdale, J. (2010). Post-translational modification of the NKG2D ligand RAET1G leads to cell surface expression of a glycosylphosphatidylinositol-linked isoform. Journal of Biological Chemistry, 285(22), 16408-16415.

Ou, F., Su, K., Sun, J., Liao, W., Yao, Y., Zheng, Y., & Zhang, Z. (2017). The LncRNA ZBED3-AS1 induces chondrogenesis of human synovial fluid mesenchymal stem cells. Biochemical and Biophysical Research Communications, 487(2), 457-463.

Pastori, C., Peschansky, V. J., Barbouth, D., Mehta, A., Silva, J. P., & Wahlestedt, C. (2014). Comprehensive analysis of the transcriptional landscape of the human FMR1 gene reveals two new long noncoding RNAs differentially expressed in Fragile X syndrome and Fragile X-associated tremor/ataxia syndrome. Human Genetics, 133(1), 59-67.

Pink, R. C., Wicks, K., Caley, D. P., Punch, E. K., Jacobs, L., & Carter, D. R. (2011). Pseudogenes: pseudo-functional or key regulators in health and disease? RNA, 17(5), 792-798.

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research, 33, D501-D504.

Qian, C., & Zhou, M. M. (2006). SET domain protein lysine methyltransferases: Structure, specificity and catalysis. Cellular and Molecular Life Sciences CMLS, 63(23), 2755-2763.

Qian, L., Gao, X., Huang, H., Lu, S., Cai, Y., Hua, Y., Zhang, J. (2017). PRSS3 is a prognostic marker in invasive ductal carcinoma of the breast. Oncotarget, 8(13), 21444-21453.

Quan, F., Zonana, J., Gunter, K., Peterson, K. L., Magenis, R. E., & Popovich, B. W. (1995). An atypical case of fragile X syndrome caused by a deletion that includes the FMR1 gene. American Journal of Human Genetics, 56(5), 1042-1051.

Rajaratnam, A., Shergill, J., Salcedo-Arellano, M., Saldarriaga, W., Duan, X., & Hagerman, R. (2017). Fragile X syndrome and fragile X-associated disorders. F1000 Research, 6, 2112.

Rao, R. C., & Dou, Y. (2015). Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. Nature Reviews. Cancer, 15(6), 334-346.

Ravichandran, M., Jurkowska, R. Z., & Jurkowski, T. P. (2018). Target specificity of mammalian DNA methylation and demethylation machinery. Organic & Biomolecular Chemistry, 16(9), 1419-1435.

Reisenberg, M., Singh, P. K., Williams, G., & Doherty, P. (2012). The diacylglycerol lipases: structure, regulation and roles in and beyond endocannabinoid signalling. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1607), 3264-3275.

Rosenberg, T., Gal-Ben-Ari, S., Dieterich, D. C., Kreutz, M. R., Ziv, N. E., Gundelfinger, E. D., & Rosenblum, K. (2014). The roles of protein expression in synaptic plasticity and memory consolidation. Frontiers in Molecular Neuroscience, 7, 86.

Saldarriaga, W., Tassone, F., González-Teshima, L. Y., Forero-Forero, J. V., Ayala-Zapata, S., & Hagerman, R. (2014). Fragile X Syndrome. Colombia Médica : CM, 45(4), 190-198.

Sanlaville, D., Verloes, A. (2007) CHARGE syndrome: an update. Eur J Hum Genet 15, 389-399.

Schapira, M., & de Freitas, R. F. (2014). Structural biology and chemistry of protein arginine methyltransferases. MedChemComm, 5(12), 1779-1788.

Shendure, J., Ji, H. (2008) Next-generation DNA sequencing. Nat Biotechnol 26, 1135-1145.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Research, 29(1), 308-311.

Sikkema-Raddatz, B., Johansson, L. F., de Boer, E. N., Almomani, R., Boven, L. G., van den Berg, M. P., van Spaendonck-Zwarts, K.Y., van Tintelen, J.P., Sijmons, R.H., Jongbloed, J.D. & Sinke, R. J. (2013). Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. Human Mutation, 34(7), 1035-1042.

Spilker, C., & Kreutz, M. R. (2010). RapGAPs in brain: multipurpose players in neuronal Rap signalling. European Journal of Neuroscience, 32(1), 1-9.

Tajima, S., Suetake, I., Takeshita, K., Nakagawa, A., & Kimura, H. (2016). Domain structure of the Dnmt1, Dnmt3a, and Dnmt3b DNA methyltransferases. In DNA Methyltransferases-Role and Function, Advances in Experimental Medicine and Biology, Springer, Cham, 945, 63-86.

Talkowski, M. E., Rosenfeld, J. A., Blumenthal, I., Pillalamarri, V., Chiang, C., Heilbut, A., Ernst, C., Hanscom, C., Rossin, E., Lindgren, A.M. & Pereira, S. (2012). Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell, 149(3), 525-537.

Valverde, R., Edwards, L., & Regan, L. (2008). Structure and function of KH domains. The FEBS Journal, 275(11), 2712-2726.

Wang, F., Hu, Y. L., Feng, Y., Guo, Y. B., Liu, Y. F., Mao, Q. S., & Xue, W. J. (2019). High-level expression of PRSS3 correlates with metastasis and poor prognosis in patients with gastric cancer. Journal of Surgical Oncology, 119, 1108-1121.

Wang, L., Li, Z., Li, Z., Yu, B., & Wang, Y. (2015). Long noncoding RNAs expression signatures in chondrogenic differentiation of human bone marrow mesenchymal stem cells. Biochemical and Biophysical Research Communications, 456(1), 459-464.

Yang, H., Sasaki, T., Minoshima, S., & Shimizu, N. (2007). Identification of three novel proteins (SGSM1, 2, 3) which modulate small G protein (RAP and RAB)-mediated signaling pathway. Genomics, 90(2), 249-260.

Yang, Q., Yang, Y., Zhou, N., Tang, K., Lau, W. B., Lau, B., Zhou, S. (2018). Epigenetics in ovarian cancer: premise, properties, and perspectives. Molecular Cancer, 17(1), 109.

Yang, Y., Bedford, M. (2013) Protein arginine methyltransferases and cancer. Nat. Rev. Cancer 13, 37-50.

Yıldız, K. Ş., Durmuş, K., Dönmez, G., Arslan, S., & Altuntaş, E. E. (2017). Studying the Association between Sudden Hearing Loss and DNA N-Methyltransferase 1 (DNMT1) Genetic Polymorphism. The Journal of International Advanced Otology, 13(3), 313-317.

Zahir, F., Firth, H. V., Baross, A., Delaney, A. D., Eydoux, P., Gibson, W. T., Langlois, S., Martin, H., Willatt, L., Marra, M.A. & Friedman, J. M. (2007). Novel deletions of 14q11.2 associated with developmental delay, cognitive impairment and similar minor anomalies in three children. Journal of Medical Genetics, 44(9), 556-561.

Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., & Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. Molecular Cell, 12(1), 177-185.

Zhong, X., Peng, Y., Yao, C., Qing, Y., Yang, Q., Guo, X., Xie, W., Zhao, M., Cai, X. & Zhou, J. G. (2016). Association of DNA methyltransferase polymorphisms with susceptibility to primary gouty arthritis. Biomedical Reports, 5(4), 467-472.