Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science

1 – DNA methylation quantification analysis in different human soft tussues for forensic purposes
2 – Whole exome sequencing data analysis for geneic alteration in heart malfunctioning

Ahmed Arslan

Examensarbete / Swedish University of Agricultural Sciences
Department of Animal Breeding and Genetics
426
Uppsala 2014

Master Thesis, 30 hp

Master's Programme
– Animal Science

Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

Part 1 – DNA methylation quantification analysis in different human soft tussues for forensic purposes
Part 2 – Whole exome sequencing data analysis for geneic alteration in heart malfunctioning

Ahmed Arslan

Supervisor:
Maria Allen, UU, Department of Immunology, Genetics and Pathology
Anna Maria Johansson, SLU, Department of Animal Breeding and Genetics

Examiner:
Sofia Mikko, SLU, Department of Animal Breeding and Genetics

Credits: 30 hp
Course title: Degree project in Animal Science
Course code: EX0578
Programme: Master's Programme – Animal Science
Level: Advanced, A2E

Place of publication: Uppsala
Year of publication: 2014
Name of series: Examensarbete / Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, 426
On-line publicering: http://epsilon.slu.se

Nyckelord: DNA metylering, epigenetik, SNP, genetiska tester
Key words: DNA methylation, epigenetics, SNP, genetic tests

Table of contents

Analysis of human epigenetic regions for Forensics: The Azlan model

Student: **Ahmed Arslan**[1]

Supervisor: **Marie Allen**[2]

Abstract

DNA methylation plays quite an important role in the mammalian genomes absolute functions. With the advancement in the epigenetics, it has been shown that the DNA methylation is a variable factor with respect to the age, differentially present in different human tissue types and may also be influenced by different life styles. We designed a project of forensic genetics, to identify the human soft tissues, based on the quantification of DNA methylation at specific genomic regions of different human tissues. We used bisulphite (pyro) sequencing to quantify the DNA methylation and confirmed that the regions are different with respect to their DNA methylation content. This approach may serve in the future, as a powerful forensic platform not only as an identification tool for tissues of unknown nature but also to predict age, gender of the donor and time of the crime scene.

Keywords: DNA, DNA methylation, mammalian genome, epigenetics, pyrosequencing, forensic.

Background

A heritable change in the gene function that is not controlled by DNA sequence change is called epigenetic change or simply epigenetics. All somatic cells from an individual are identical genetically (except B and T cells*) but still all of them are distinctly different and purposely differentiate to perform a particular physiological function (Christenson et al. 2008). The vast variety of the cellular phenotypes are only made possible by the epigenetically controlled gene expression, which also plays an important role in the cellular differentiation. DNA methylation is a property of higher eukaryotes that has profound effects on a number of life processes. In methylation of DNA, the methyl- group is attached at C5 of a cytosine at a given double stranded DNA fragment. The methylation of DNA is the most studied epigenetic modification because of its important role in embryonic development, genomic imprinting, X-chromosome inactivation. Moreover, the epimutations have been linked to a number of human diseases as well (Webar et al. 2007).

In an epigenome most of the methylation presents at specific sites rich in C and G contents. These unique DNA segments are present at transcription factor binding sites (promoters) known as CpG island (CGIs). These CGIs are evolutionary conserved at promoter regions whereas the rest of a mammalian genome is depleted from these regions due to frequent CpG to TpG transitions.

*The B cells and T-cells are important for their role in the immune system particularly in the production of antibodies. For this reason, these cells are able to manipulate their genetic makeup. They rearrange different functional gene segments to produce a functional antibody (Faber 2013).

1 - Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden.
2 - Department of Immunology, Genetics and Pathology, Uppsala University, Sweden.

The two important enzymes (TDG and MBD4) selectively removes the T:G mismatches resulted from the deamination of methylated cytosine from C to T.

Under normal conditions it is thought that cells maintain CGIs in an unmethylated state prior to the DNA transcription (Jones et al 2002). Since it has been established that the methylation of CGIs is incompatible with the gene activity, still there are contradictory claims of higher DNA methylation at promoter regions. It has also been suggested that the promoter CpG methylation is a complex phenomenon when coupled with human aging.

The DNA methylation has been proposed as a biomarker for the disease susceptibility, as the promoter hypermethylation has been observed in cancer cases (Christensen et al 2009). A cancer is an aging disease with the age dependent increase in the DNA methylation levels has been observed. So it has motivated the researchers to study the DNA methylation of the cancer cells. The locus specific increase in the methylation levels were observed in different cell types irrespective of their type and origin. This could also lead to a common hypothesis that there must be a similar mechanism governing this DNA methylation change in the different cell types. There are two possible explanations for it, one is the reduced maintenance fidelity of methyltransferases with aging and the second is the environmental affects on methylation over the time.

Recently genome-wide DNA methylation studies have shown tissue specific patterns of DNA methylation (Christensen et al 2009). These studies precisely define the tissue specific methylation patterns but are unable to show the mechanism by which these epigenetic variations resulted (Christensen et al 2009).

In forensic medicine, the tissue identification is very important to establish the case with respect to the donor of the sample of interest. For this purpose, there are a number of biological procedures (to process the information from the biological samples recovered from the crime scene) established. These procedures are based on either analysis of the DNA, RNA or protein. The DNA based identification procedures use highly polymorphic autosomal STR (short tandem repeat), SNP (single nucleotide polymorphism) and Y-STR profiling to identify the DNA donor. These procedures are advantageous and give precise forensic clues but there are also some drawbacks in the use of these procedures, for example the information must be compared with the DNA profiles already present in the investigative databases. If the profiles are not already present in the relevant database than these procedures are not very useful (Kayser and de Knijff 2011). The procedures, which are based on the identification of RNA, aim to identify the differentially expressed mRNA in different cells. These procedures suffer heavily due to the fact that mRNA is a very unstable molecule that degrades with time and may thus not be helpful for old samples (Frumkin et al. 2010). The protein detection based assays have some advantages over the routinely used lab procedures but still they are not ideal for forensic analyses. This is because they are based on the enzymatic or immunological detection and proteins are less stable then DNA and tend to denature with time and temperature. These procedures are less informative when used for the detection of old or degraded samples.

With the advancement in the -omic era, new data has been generated regarding different aspects of the human genome including the epigenome. As described earlier we have now established data that describes the presence of different DNA methylation patterns in different tissues. Based on this idea we have designed a study where we could be able to identify different human tissues while studying DNA methylation of a certain CpG regions. For this purpose, we used pyrosequencing technique to

quantify the DNA methylation. The pyrosequencing may also help in quantifying the DNA methylation of degraded DNA as conventionally pyrosequencing has been used for the sequencing of short DNA strands (50 to 200bp).

The aim of this study was to evaluate the presence of different DNA methylation patterns present at specific genomic regions in different (studied) human tissues, namely blood, semen, saliva and epithelium. The reason to perform these analyses was to make a new platform where we could confidently distinguish between different unknown tissues based on the epigenetic patterns. This procedure might also help in forensic tissue identification of unknown origin in future.

Material and methods:

DNA sampling and bisulphite conversion:

To analyse the DNA methylation levels of different tissues, we collected the samples of different human tissues including blood, semen, saliva and epithelium from a number of individuals of different age groups, gender and ethnic backgrounds.

The DNA was extracted by using Wizard DNA purification kit (Promega) according to the manufacturer's recommendation, and the concentrations were standardised with nanodrop.

The DNA was processed by bisulphite conversion with a commercially available kit (Epitech, Qiagen, Germany), which was reported elsewhere (Fraga and Esteller 2002), according to the manufacturers recommendations. Briefly, this procedure converts the unmethylated cytosine to uracil and the methylcytosine remains unchanged. So this change introduced by the bisulphite treatment in the DNA sequence helps in precise detection and quantification of the methylation status of a region at a single base pair resolution.

The methylated DNA was amplified by using pyromark PCR kit (Qiagen) (Table 1). This kit consists of a pyromark mix with a tailor-made HotStart Polymerase that is recommended for the bisulphite treated DNA's PCR and pyrosequencing procedures. The PCR conditions were provided by the manufacturer's directions with annealing temperature of 56°C. The amplified products were run on a 2% agarose gel for 25min at 90V.

Table 1. Polymerase chain reaction ingredients used for a reaction setup.

Ingredients	Concentration	Quantity
PyroMark PCR Mix#	10x	12.5ul
CoralLoad¤	1x	2.5ul
Primers (each)	10mM	0.5ul
MQ water		8ul
DNA	+/- 50ng/ul	1ul

PyroMark PCR Mix contains the Hotstart Polymerase, dNTPs and $MgCl_2$ (3mM). ¤ CoralLoad is a combination of different dyes namely gel loading, red and orange dye.

The pyrosequencing was performed using the Q24 pyrosequencing machine (Pyrosequencing AB, Uppsala). Briefly, pyrosequencing is a sequencing procedure that is known as "sequencing by synthesis" which depends on the detection of a pyrophosphate by the enzyme (ATP sulfurylase) and subsequent production of an ATP when a nucleotide is added to a growing sequence (Ahmadian et al 2006). Another enzyme converts the ATP into a light signal that is recorded by computer and hence the sequencing (Figure 1). During the methylation analysis, the pyrosequencing machine measures the ratio between the C without methyl-group to the C with methyl-group present in the studied sequence. A Cytosine nucleotide that is not followed by G is unmethylated and converted to T after the PCR of bisulfide treatment of DNA. This data depict as a pyrogram with the percentage of methylation present at a certain CpG position. A typical pyrogram is shown in figure 2, the height of

the peak is proportional to the number of the nucleotides added. We stored and analysed all the data on an excel sheet.

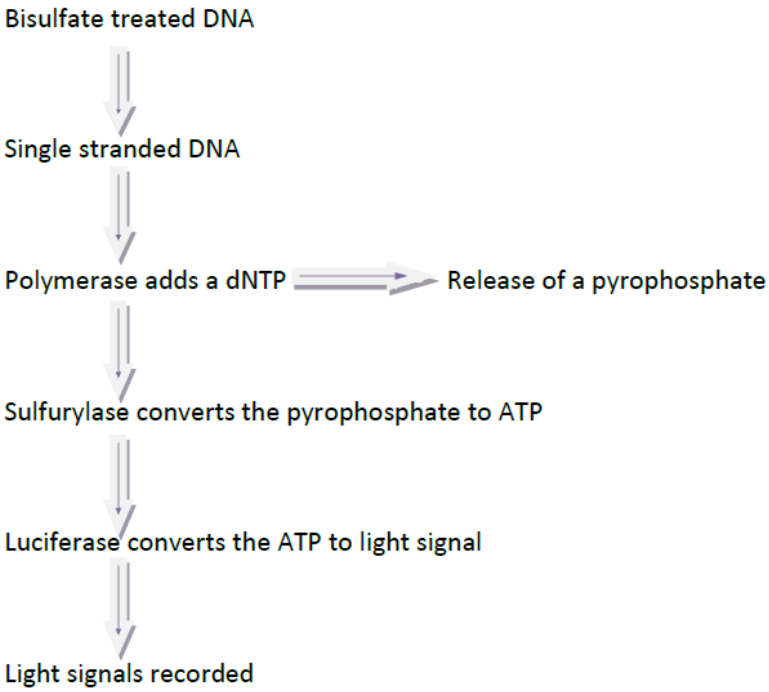


Figure 1: Pyrosequencing principle displayed. Pyrosequencing depends on a protein complex, which consists of DNA polymerase, Apyrase, sulfurylase and Luciferease. Polymerase incorporates the dNTPs to the growing DNA strand, as a nucleotide added to the growing strand it releases a pyrophosphate. This pyrophosphate converts to an ATP molecule by Sulfurylase enzyme, another enzyme, luciferase, helps to convert this ATP molecule to a light signal. In this manner a sequence of light signals with different wave lengths is recorded in real time as the DNA strand.

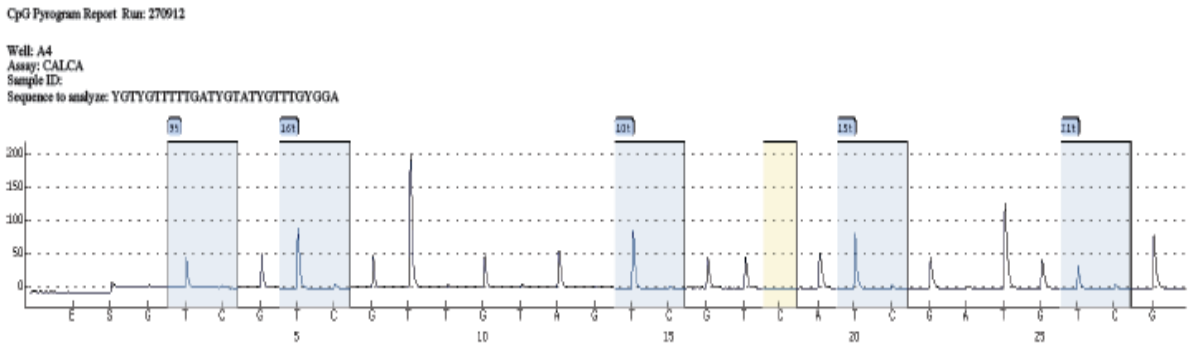


Figure 2: A typical pyrogram is shown here, with the horizontal axis showing the sequencing segment whereas the vertical axis showing the strength of the signals.

Results:

To identify the different unknown tissues on the basis of the DNA methylation, we selected a panel of markers that have previously been reported for their tissue, age, and gender methylation discrimination (Table 2). The table shows a panel of DNA markers with the DNA methylation patterns observed in different tissues and other aspects i.e. age, gender and population. We tested these markers for a number of (different) tissues and the results have shown the interesting findings for each of the locus studied. We have observed a marked different in the DNA methylation in a locus specific manner in a particular tissue type for each marker that we have used. For example, we have used PLSCR2 marker to observed blood specific increase in the DNA methylation as reported elsewhere (Fraser et al 2012). Likewise, we have used already reported TLE1, CALCA, LAMB1, HENMT1 markers for the saliva, age, blood and sperm specific DNA methylation fluctuation receptively, and observed a specific pattern for each marker. The results were analysed by plotting the DNA methylation quantity against a specific methylation site (figure 3-7) and complied into the excel charts.

Table 2: The panel of selected genes previously reported for their differential DNA methylation behaviour for different aspects. 'YES' shows the trait for which a particular marker was used / reported.

Gene/Locus	Blood	Saliva	Sperm	Age	Gender	Population
PLSCR2 (Fraser et al 2012)	YES			YES		
TLE1 (Liu et al. 2010)	YES	YES			YES	
CALCA (Sarter et al. 2005)				YES	YES	YES
LAMB1 (Christensen et al 2009)	YES			YES		
HENMT1 (Molaro et al. 2011)			YES			

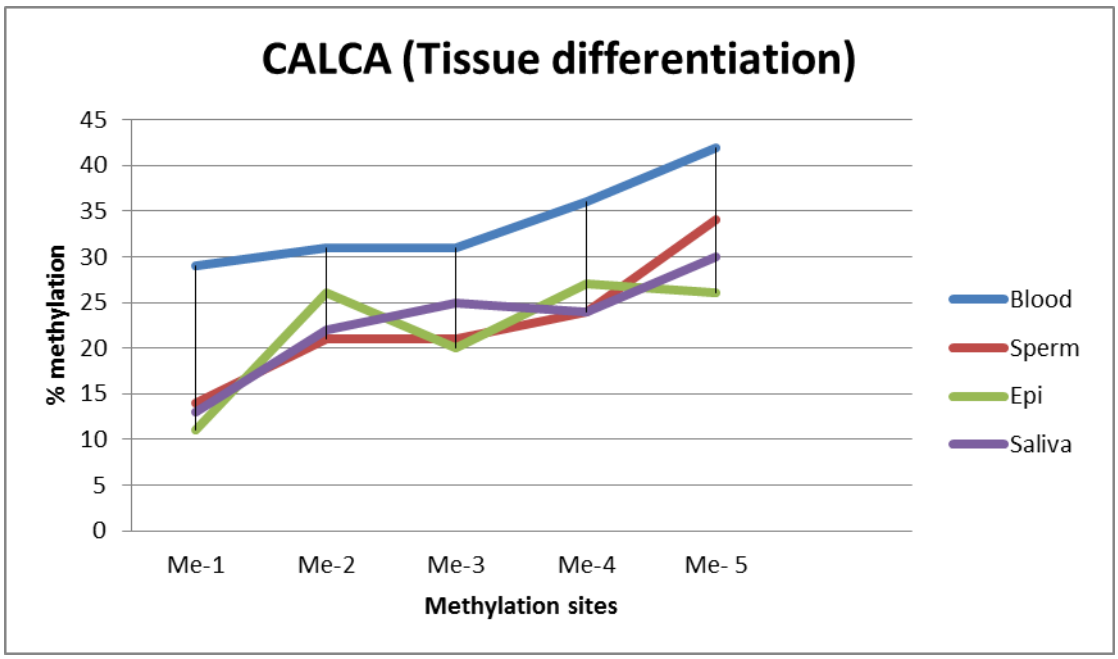
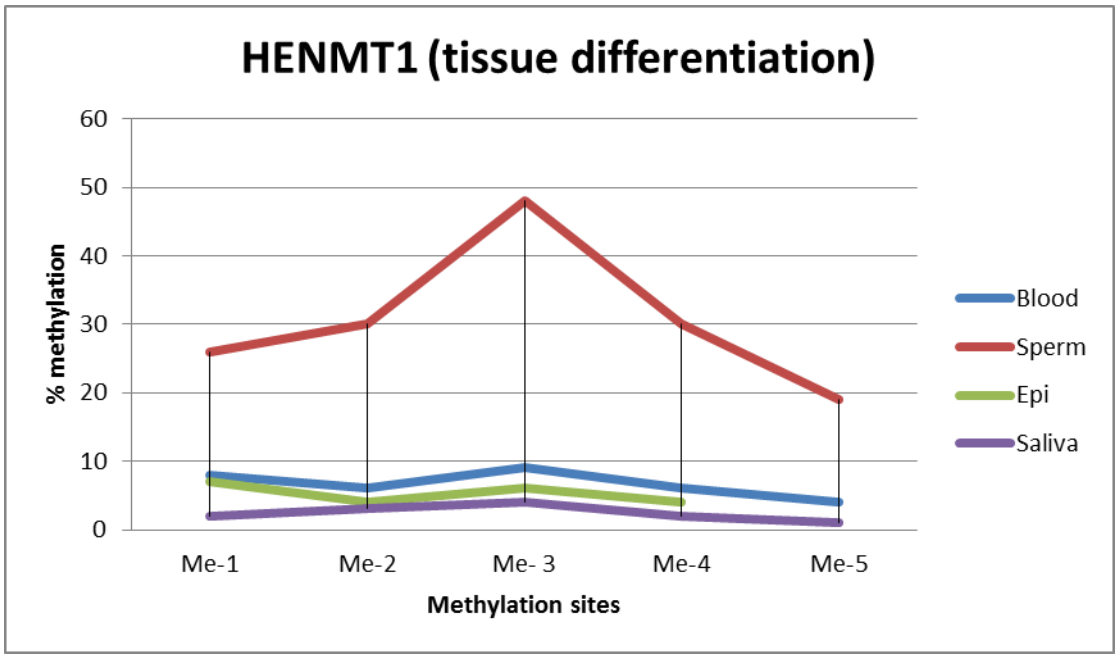


Figure 3: The tissue specific methylation observed at the CACLA gene region, the blood samples showing more methylation at this site when compared to other studied sample types. (The x-axis using the number of methylation observed sites in the given sequence whereas the y-axis showing the %age methylation at each site).

(A)



(B)

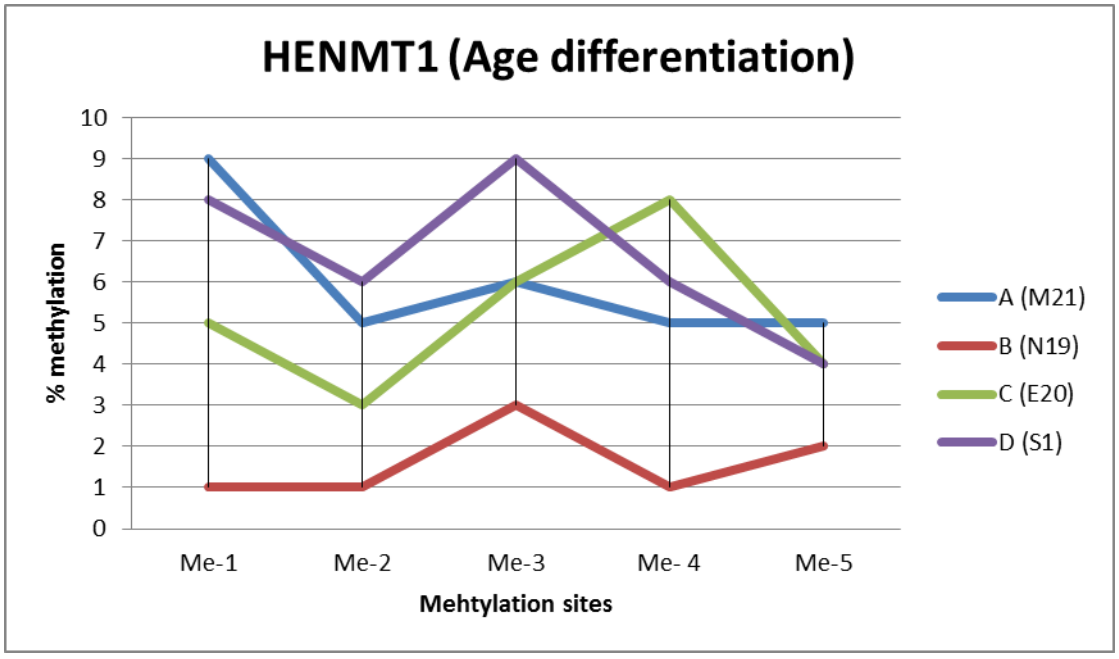
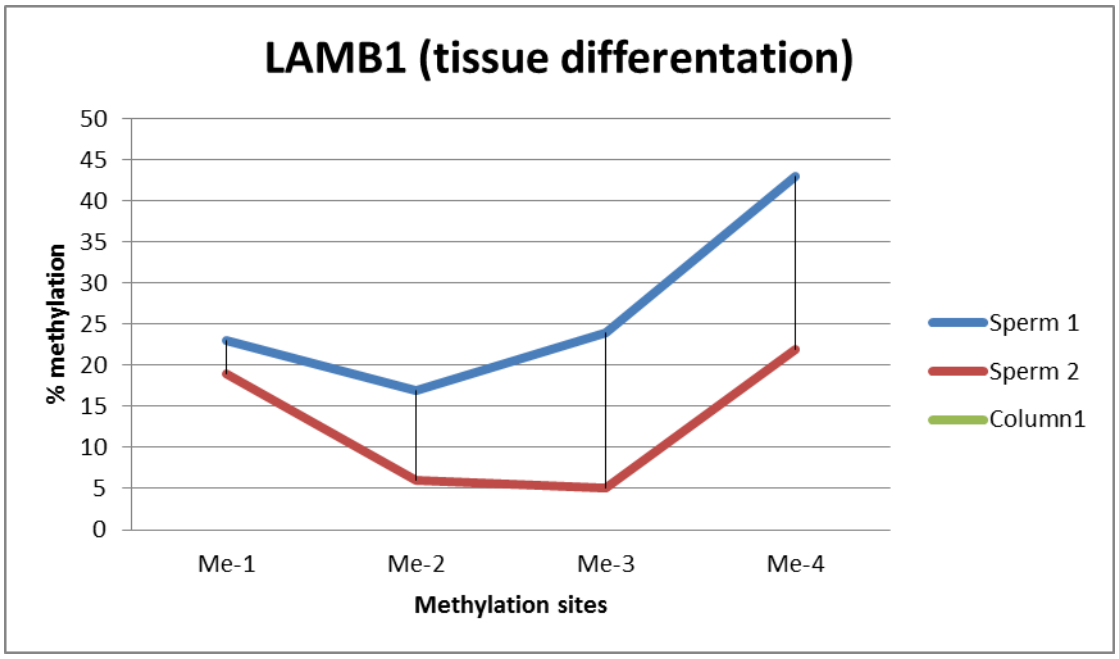


Figure 4: (A) shows the inter-tissue DNA methylation differences, the blood samples have much more quantities of methylation at this gene (HENMT1). (B) The methylation pattern observed at this locus in different age groups. (*A* is a 50 years old Swedish male, *B* is 50 years old English female, *C* 30 years old fin lady and *D* is a 30 years old Swedish man).

(A)



(B)

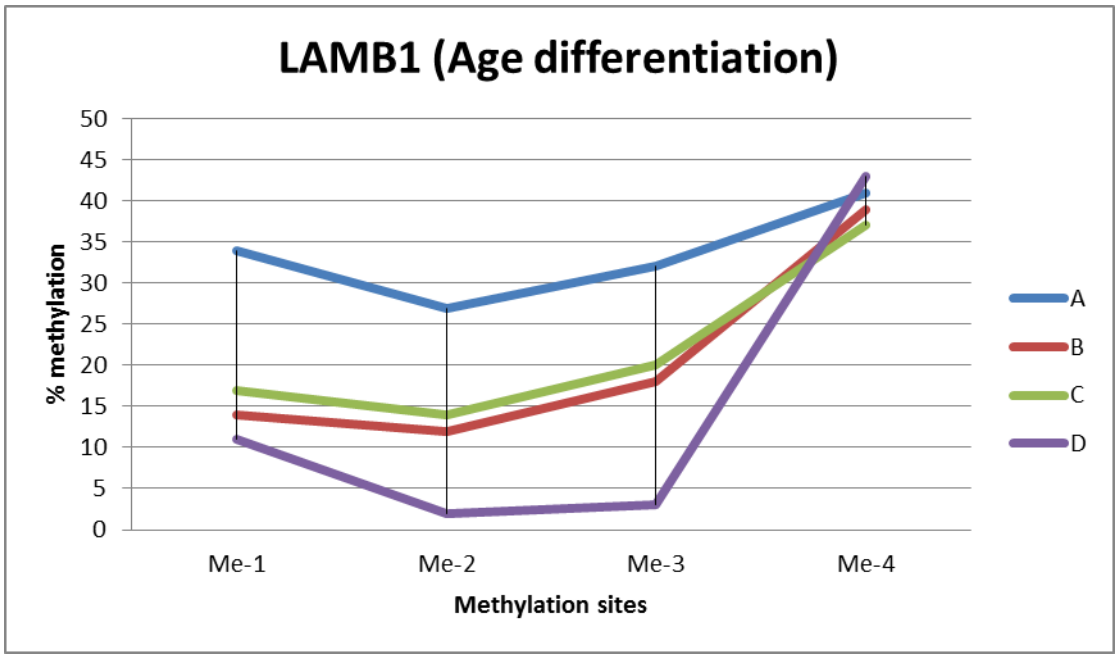
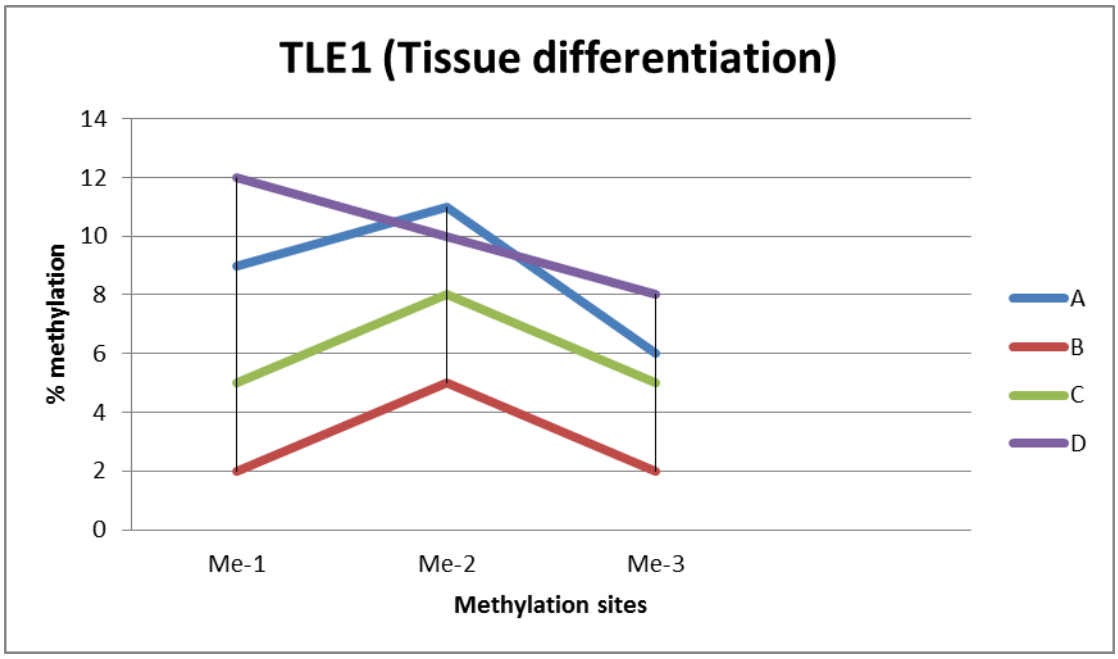


Figure 5: (A) Shows the DNA methylation difference at the LAMB1 locus when observed in two sperm samples from two individuals. (B) Shows the difference in the methylation of blood samples at this locus from different age and gander. (*A* is a 50 years old Swedish male, *B* is 50 years old English female, *C* 30 years old fin lady and *D* is a 30 years old Swedish man).

(A)



(B)

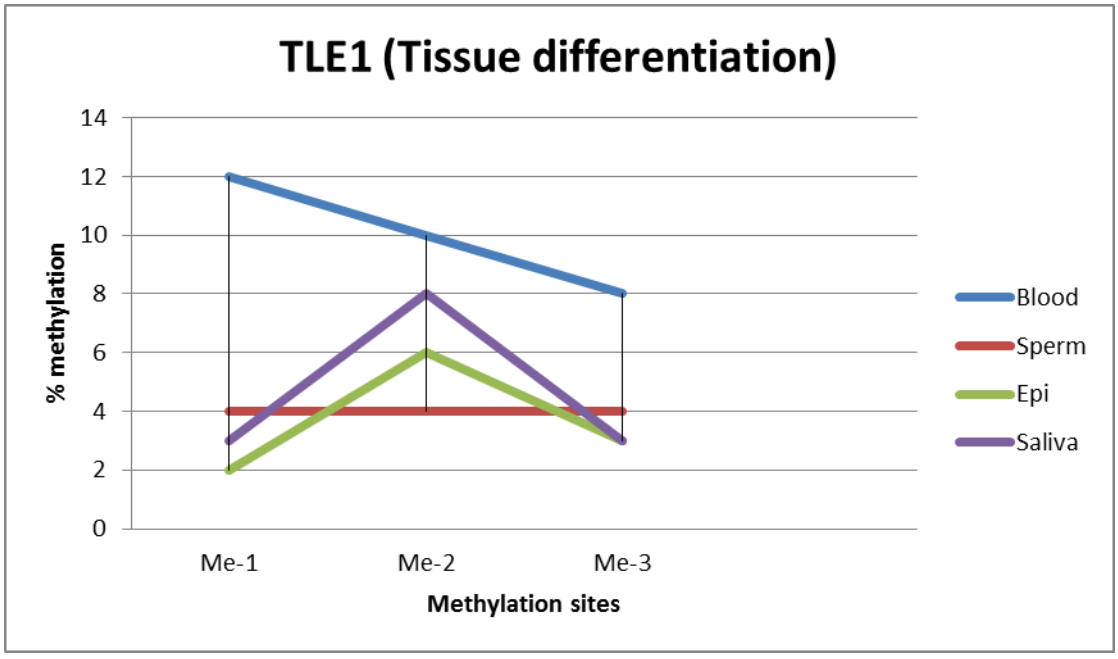


Figure 6: (A) The observed DNA methylation at TLE1 gene shown here in different blood samples from different age groups. (*A* is a 50 years old Swedish male, *B* is 50 years old English female, *C* 30 years old fin lady and *D* is a 30 years old Swedish man). (B) The inter-tissue methylation patterns observed at this locus shown with blood having higher levels of methylation when compared to other sample types.

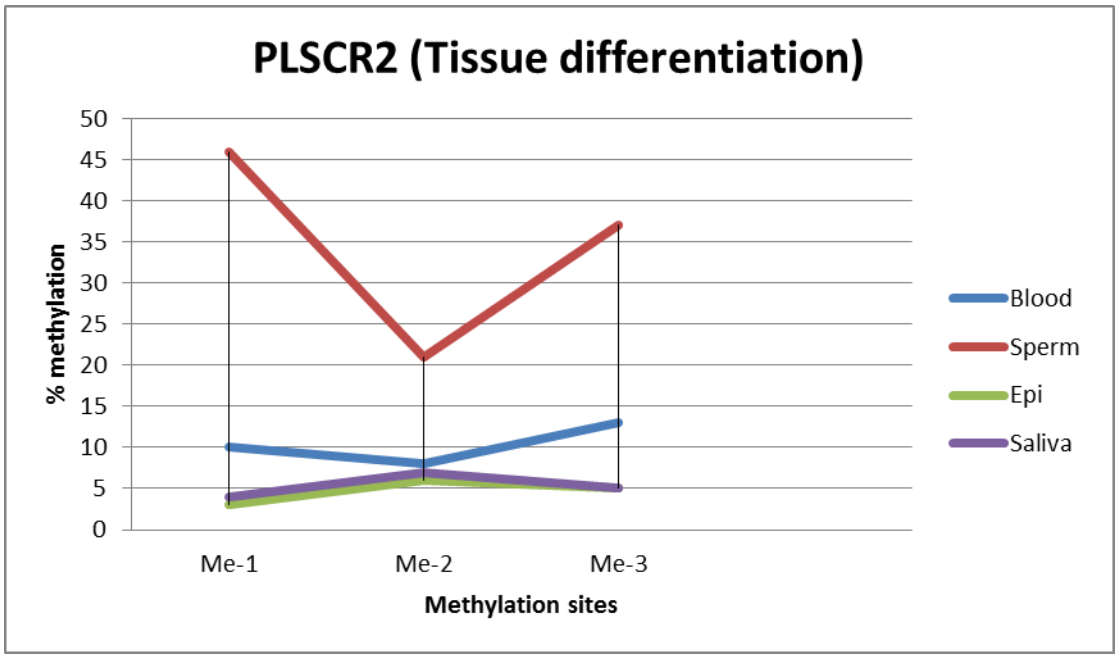


Figure 7: The locus (PLSCR2) specific methylation observed in different tissue samples. The sperm sample clearly shows a higher methylation trend at this studied locus.

Discussion:

The forensic medicine has advanced over the last few decades. Many new and advanced molecule detection procedures have been created. Some of them have advantages over other in tissue identification but still not a single procedure could be considered as an ideal procedure to detect all required data from a given forensic sample (Frumkin et al 2011). The protein detection based assays might not work properly well when performed on old or degraded samples, due to the difficulty in detection of the denatured protein(s) of interest. A few recently developed forensic systems are based on the detection of the RNA from the samples but these assays are also targeting a relatively unstable molecule (mRNA) (Haas et al. 2009). We have considered the DNA methylation as a source of information to detect the age, gender and tissue types from a crime scene sample. Our method has two important aspects; one is to detect the variations of the locus specific DNA methylation of different tissues. The second edge of our procedure is the use of pyrosequencing in methylation profiling. The methylation results from the covalently bonded chemical (methyl-) groups to the DNA, which is not degraded with time as in the case of RNA and protein. Hence DNA methylation tends to remain constant over a period of time. The pyrosequencing as described earlier is working best when used to sequence a segment with a length less than 500bp. So if we have degraded forensic samples our procedure of tissue detection may still works.

Also we have shown that a specific region of the human genome could have the methylation fluctuation for the age, gender or ethnicity, hence these regions could be used for the forensic analysis. We have used the DNA markers which were previously reported for a certain trait for a range of different traits, like for example, if a region is reported for an age specific methylation difference we tested it for the gender discrimination at the same time along with age specific methylation and observed if we could find any difference for this trait as well (Azlan model). The aim was to use less number of markers to get more types of information. The gene PLSCR2 is previously reported for its differential methylation with aging (Fraser et al. 2012), but when we have used this marker for tissue detection, we have found really interesting results (Figure 7). In this chart the sperm samples have more methylation patterns than rest of the studied tissues. Likewise, TLE1 locus specific methylation is reported for gender discrimination (Liu et al 2010), and we have used it for different age groups and gender as well. (Figure 6-A) For different tissue samples we can clearly see a difference of methylation among them (Figure 6-B). The LAMB1 marker is reported for its methylation differentiation with age (Christensen et al 2009), whereas we used it for both blood and sperm samples and showed a difference of methylation exist with age in both type of samples (Figure 5A-B). The HENMT1 is reported for its higher methylation trends in sperm tissue (Molaro et al 2011), but we have observed a different pattern of DNA methylation at this locus of different tissue types (Figure 4-A). The male individuals have more methylation content at this locus when compared to female samples with their respective age groups (Figure 4-B). The CALCA is shown to have a difference in the methylation with respect to the age and gender (Sarter et al 2005) but confirmed a tissue level difference in the methylation as well in our results (Figure 3).

The quality of this type of research is largely dependent on finding regions of methylation differentiation and testing them on a range of samples. The aim of this study was to further formulate a panel of markers that may be able to give a more comprehensive set of information by using less of them and help in in-depth forensic analysis. For the future, the need is to more refine these results by

gathering more information about the methylation differential regions. And we also wanted to sample different tissues from people who are alcoholic and smoking addicts, to investigate if we can be able to detect the different methylation patterns which are specific to these life styles?

References:

Ahmadian A, Ehn M, Hober S. 2006. Pyrosequencing: history, biochemistry and future. Clin Chim Acta. 363(1-2):83-94.

Sarter B, Long TI, Tsong WH, Koh WP, Yu MC, Laird PW. 2005. Sex differential in methylation patterns of selected genes in Singapore Chinese.Hum Genet. 117(4):402-3.

Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, Sugarbaker DJ,Yeh RF, Wiencke JK, Kelsey KT. 2009. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet. 5(8):e1000602.

Faber J. 2013. Transcription factor EBF1 reminds cells who they are. http://www.mpg.de/7472572/EBF1-cells1.

Fraga MF and Esteller M. 2002. DNA methylation: a profile of methods and applications.

Biotechniques. 33(3):632, 634, 636-49.

Fraser HB, Lam LL, Neumann SM, Kobor MS. 2012. Population-specificity of human DNA methylation. Genome Biol. 9;13(2):R8.

Frumkin D, Wasserstrom A, Budowle B, Davidson A. 2011. DNA methylation-based forensic tissue identification. Forensic Sci Int Genet. 5(5):517-24.

Haas C, Klesser B, Maake C, Bär W, Kratzer A. 2009. mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. Forensic Sci Int Genet. 3(2):80-8.

Jose Luis Royo, Manuel Hidalgo & Agustin Ruiz. 2007. Pyrosequencing protocol using a universal biotinylated primer for mutation detection and SNP genotyping. Nat. Prot.2: 1754-2189

Jones PA, Baylin SB. 2002.The fundamental role of epigenetic events in cancer. Nat Rev Genet. 3(6):415-28.

Kayser M, de Knijff P. 2011.Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet. 12(3):179-92.

Liu J, Hutchison K, Perrone-Bizzozero N, Morgan M, Sui J, Calhoun V. 2010. Identification of genetic and epigenetic marks involved in population structure. PLoS One. 7;5(10):e13209.

Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD . 2011. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. Cell. 146(6):1029-41.

Royo JL, Hidalgo M, Ruiz A. 2007. Pyrosequencing protocol using a universal biotinylated primer for mutation detection and SNP genotyping. Nat Protoc. 2(7):1734-9.

Sarter B, Long TI, Tsong WH, Koh WP, Yu MC, Laird PW. 2005. Sex differential in methylation patterns of selected genes in Singapore Chinese. Hum Genet.117(4):402-3.

Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 39(4):457-66.

Molecular Genomic explanation of human sudden cardiac arrest

Ahmed Arslan [1]

Marie Allen [2]

Abstract

Sudden cardiac death (SCD) is an unexpected causality resulted by cardiac malfunctioning over a short stretch of time (<1 hour). SCD accounts for nearly 325,000 deaths per year in the United States only; more deaths are attributable to SCD than any other kind of natural death. The development and function of the human heart is under strict control of gene networks. Any change in the functional sequence of these genes could potentially be fatal. We have observed recently two of such cases where abnormal human heart function leads to sudden cardiac arrest. We planned to conduct a study, based on the exome sequencing and analysis of DNA samples taken from the cases. The aim of this project was to detect the sequence variations (SNP and Indels) that are only present in the cases but not in the population; and also interpret the possible role of these variants in these deaths. We detected around 21 new SNPs and 17 Indels that might possibly be the cause of the SCD. We want to continue analysing this data to detect more previously unknown/unrelated cardiac loci for their potential role in these causalities and help in formulating genetic tests for SCD.

Keywords: Sudden cardiac death, gene, exome sequencing, sequence variations, SNP, Indels, genetics tests.

Background

Sudden cardiac death (SCD) is an unexpected causality resulted by cardiac malfunctioning occurring for a short period (generally within 1 h of symptom onset) in a person with known or unknown cardiac disease (in our case its was known). Most of the times the SCD cases are related to cardiac arrhythmias (Sovari et al 2012).

The coronary heart disease often leads to the sudden cardiac death and in developed world including the USA the SCD responsible of 50% of deaths by cardiovascular diseases (Douglas et al. 2012).

SCD accounts for nearly 325,000 deaths per year in the United States only this means only 0.1% to 0.2% of all US population is dead of SCD but still these numbers are more than lung cancer, breast cancer, or AIDS. The frequency of SCD in Western industrialized nations is similar to that in the United States. This represents an incidence of 0.1-0.2% per year in the adult population. SCD is often the first expression of CAD and is responsible for approximately 50% of deaths from CAD. The trend towards expending the numbers of SCD events in developing nations of the world is thought to reflect a change in their dietary and lifestyle habits in these nations. It has been estimated that SCD claims more than 7,000,000 lives per year worldwide (Sovari et al 2012) (Figure 1-A).

1 – *Student*, Department of Genetics and Bioinformatics, Swedish University of Agricultural Sciences, Uppsala, Sweden.
2 - *Professor*, Department of Immunology, Genetics and Pathology, Uppsala University, Sweden.

The sudden cardiac events might relate with age and also vary in gender in overall populations of the world. The SCD incidents among the coronary disease patients decreased with the age (Figure 1-B) and men suffering from coronary heart diseases are more prone to die than the women. This ratio is as high as 75% where women died 3 to 4 times less than men from sudden cardiac death. As the figure 1-B, depicts the mortality rate is high in the child with the age of 0 to 6 months, for that of sudden infant death syndrome. The condition of unexpected death in infants with no medical history that could also includes SCD.

The initial studies on SCD have shown that the most persons affected by the sudden cardiac events are those with young age from 1 to 13 and 14 to 21 years with percentages of 19% and 30% respectively (Neuspiel et al 1985).

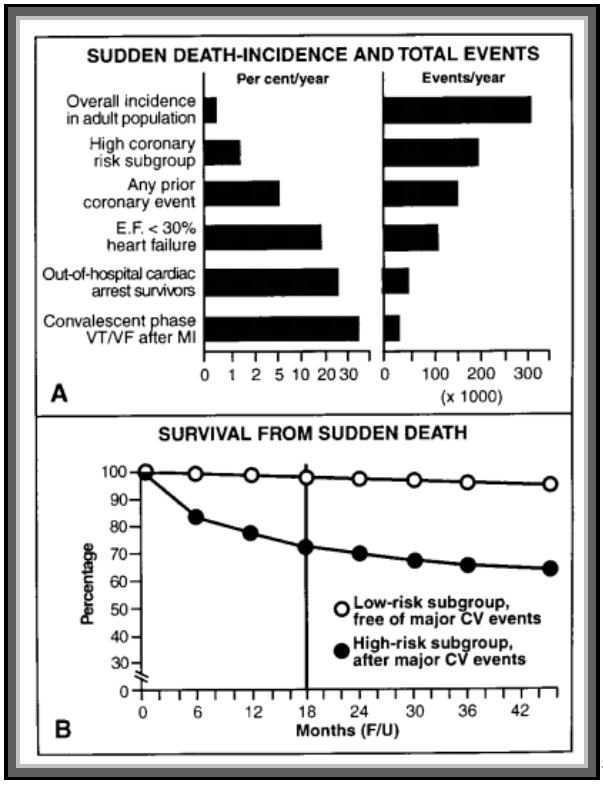


Figure 1: (A) Sudden cardiac deaths' relation with the sudden cardiac events (B) The relation of age with the sudden cardiac arrest shown.

The sudden cardiac arrests often related to the anatomical changes in the heart or in the heart related structures observed in the autopsy of the hearts of subjects. In nearly all of the cases it has been observed that the coronary plague morphology change and/or myocardial scars present. In the cases where heart conduction disturbances have seen the role of apoptosis also may participate in these SCD (James et al 1996). But these changes are almost the same as those with coronary heart disease so they might not safely represent the risk factor for the SCD (Douglas et al 2012).

Regular exercise can decrease the events of cardiac arrest in experiment models but vigorous exercise in the athletes on the other hand produces deaths at a rate of 1 in 12000 individuals (Mittleman et al 1993). Annually 20 to 25 deaths reported in US athletes due to sudden cardiac events. But still the overall impact of physical activity on the sudden cardiac arrest is somewhat controversial; as in one such study that shown the 67% of the deaths results from the SCD when the subjects were in their inactive physical states (de Vreede et al 1997).

The genetics play the major role in the development and properly functionality of body physiology. A strong relationship exists between DNA sequence alterations and disease aetiology. The stable genome is essential to produce a healthy physiology of a cell. Any change in the heritable information might lead to a disease condition.

The development and function of the human heart is under strict control of gene networks. A change in the functional sequence of these genes could potentially lead to an individual's death. In last four decades workers all over the world have stalked genetics knowledge of cardiovascular diseases (CVD). Both familial and non-familial approaches conducted to analyse the genetics bases of CVD that provided clues of genetic alterations in the patients with respect to the healthy populations (Morita et al 2008). These genetic alternations included the single nucleotide changes, telomere shorting, and copy number variation. Most of these CVD are termed as "complex phenotype" resulted from the interplay of environment and genes.

Genetic variants those are deleterious are often disease causing and remain rare in a population hence the disease. Whereas the variants those beneficial and lead to an advantage are established rapidly in a population. There are two schools of thoughts regarding the disease causing ability of these variants. One is called common disease common variants; stands for low penetrance with high frequency of the common variants lay the genetic bases of a common disease in a given population. On the other hand the rare variants involved in the pathology of a common disease with their strong penetrance. These rare variants are being quantified for common cardiac phenotypes with the emergence of next generation sequencing techniques these days. This lead to the discovery of previous unknown rare variants data that could help in understands the genetic complexity in the background of cardiac diseases like CAD (Novelli et al 2010).

The common cardiac abnormal phenotypes in patients are related with the presented genetic data for the cardiac diseases (Table 1). This data showing that there are more then one genes that have been involved in a particular disease; the reason for this fact might be that the complex genetics background of a particular disease. One gene could be involved in more then one disease types or one disease could be a result of variations in different genes. For example, according to this data, NXK2.5, associated with disease conditions like DiGeorge syndrome, HOS, Allagille syndrome where as many genes on the other hand are involved in the aetiology of the DiGeorge syndrome (Table 1).

We have observed recently two of such cases of sudden cardiac arrest registered at Uppsala University hospital where abnormal human heart function leads to sudden cardiac arrest. Both cases had some type of heart problems in their lifetime and none of them have them in their family. At the time of their death they were normal and feeling no signs of disease. Their ECG (electrocardiogram) recorded at the time of their death shown abnormally elevated ST-segment that helped in understanding the possible reason of their death. And also help in planning a genetics study to further evaluate the genetic bases of their death.

Table 1: Summery of genes involved in the cardiovascularpathy with respective phenotypes (Novelli et al 2010)

	Phenotype	Involved genes	Associated diseases
Congenital heart disease			
Cyanotic heart disease	Transposition of the great arteries	NKX2-5, THRAP2	
	Tetralogy of fallot	NKX2-5, NOTCH1, TBX1, JAG1, NOTCH2	DiGeorge syndrome, alagille syndrome
	Tricuspid atresia	NKX2-5	
	Pulmonary atresia	PTPN11, JAG1, NOTCH2	Alagille syndrome
	Ebstein's anomaly of the tricuspid valve	NKX2-5	
	Double outlet right ventricle	NKX2-5, THRAP2	
	Persistent truncus arteriosus	TBX1	DiGeorge syndrome
	Anomalous pulmonary venous connection		
Left-sided obstruction defects	Hypoplastic left heart syndrome	NOTCH1	
	Mitral stenosis		
	Aortic stenosis	NOTCH1, PTPN11	
	Aortic coarctation	NOTCH1, PTPN11	
	Interrupted aortic arch	TBX1	DiGeorge syndrome
Septation defects	Atrial septation defects	NKX2-5, GATA4, TBX20, MYH6, TBX5	HOS
	Ventricular septal defects	NKX2-5, GATA4, TBX20, TBX1, TBX5	HOS
	Atrioventricular septal defects	PTPN11, KRAS, SOS1, RAF1, CRELD1	Noonan syndrome
Other congenital heart defects	Bicuspid aortic valve	NOTCH1	
	Patent ductus arteriosus	TFAP2B	Char syndrome
Non ischemic cardiopathies			
Structural defects	CMH	MYH7, TNNT2, TPM1, MYBPC3, PRKAG2, TNNI3, MYL3, TTN, MYL2, ACTC1, CSRP3, LAMP2	CMH1, CMH2, CMH3, CMH4, CMH5, CMH6, CMH7, CMH8, CMH9, CMH10, CMH11, CMH12, Danon disease
	Dilated cardiomyopathy	ACTC, DES, SGCD, MYH7, TNNT2, TPM1, TTN, VCL, MYBPC, MLP, ACTN2, PLN, ZASP, MYH6, ABCC, TNNC1, TCAP, EYA4, LMNA, SCN5A, DMD, TAZ, TNNI3	Laminopathies, hypertension, ischemic disease
	Arrhythmogenic right ventricular dysplasia/cardiomyopathy	JUP, DSP, PKP2, DSG2, DSC2, RYR2, TGFB3	Naxos disease, Carvajal disease
Channelopathies	Long QT syndrome	SCN5A, SCN4B, KCNQ1, KCNH2, KNE1, KNE2, KCNJ2, ANK2, CAV3	Romano-Ward syndrome, Jervell Lange-Nielsen syndrome, Andersen-Tawil syndrome, Timothy syndrome
	Brugada syndrome	SCN5A, SCN1B, GPD1L, CACNA1C, CACNB2b	
	Sindrome di Lev-Lenègre	SCN5A	
	Short QT syndrome	KCNH2, KCNQ1, KCNJ2	
	Sindrome di Wolff-Parkinson-White	AMPK	
	Tachicardia ventricolare	ADRB1, ADRB2, ADRB3	
	Tachicardia ventricolare polimorfica catecolaminergica	RYR2, CASQ2	
	Atrial fibrillation	KCNQ1, KCNE2, KCNJ2, KCNH2	
Ischemic cardiopathy			
Coronary artery disease, myocardial infarction	Mendelian inheritance	LDLR, APOB, ABCG5, ABCG8, APOA1, ABCA1, CBS	Familial hypercholesterolemia
	Complex disease	9p21, SH2B3, MRPS6-SLC5A3-KCNE, PHACTR1, CELSR2-PSRC1-SORT, CXCL12, MIA3, PCSK9	

CMH: Hypertrophic cardiomyopathy; HOS: Holt-Oram syndrome.

Aim of the study:

The aim was to establish better understanding towards the role of genetic variations in the SCD cases. We decided to perform a study in collaboration with a group of physicians at Uppsala University hospital; the study was based on the exome sequencing analysis of DNA samples taken from the two index cases and parents of the third case. The objective was to detect and interpret the sequence variations (SNP and Indels) that are only present in the cases but not in other population.

Material and methods:

The DNA was extracted by whole blood sampling of cases using commercially available kit (Wizard DNA purification kit, Promega). The whole exome sequencing was performed on AB SOLiD™ (Applied Biosystems) at Uppsala Genome Centre (UGC) by their skilled workers. The sequencing data was analysed by custom made Perl program scripting at UGC, further SNP and data analysis was performed using publically available data bases like UCSC genome browser (http://genome.ucsc.edu/), Polyphen 2.0 (http://genetics.bwh.harvard.edu/pph2/), Sift database (http://sift.jcvi.org/), String (http://string-db.org/), SMART (http://smart.embl-heidelberg.de/) and softwares like IPA (http://www.ingenuity.com/products/pathways_analysis.html), Alamut (http://www.interactive-biosoftware.com/), MUpro (http://mupro.proteomics.ics.uci.edu/).

A thorough literature survey was performed on genes those are recently published for their role in heart diseases. And this step was further extended systematically to include more genes by doing gene-gene, protein-protein *in-silico* interaction analysis. This work plan helped us not only in generating a comprehensive panel of genes (Table 2) but also to filter the genes with their more disease causing ability.

Results:

The results are saved in the exile sheet (supplementary data file), the sheet 1 contains all the genes those were found with interesting variants with possible substitutions in DNA level and protein levels. The disease causing ability is shown under the heading Polyphen and Sift, these headings represent the data collected from these databases for individual SNP/Variant. The genes functions, individuals in which these are present and the publications in which these genes reported were also noted. The rs numbers of all individual SNP were also noted (where available) with the exon number they are present and their prevalence was also noted.

After that these variants were furthered filtered, and categorised according to their ability to cause disease and also those are only present in our cases but not in the rest of the population (Table 2/supplementary data sheet 5). Hence, by doing this type of filtering we omitted all the SNPs that are present in the general population and also those show doubtful *in silico* results regarding to the ability to cause disease. The focus was to generate as much data as possible from the seq data of our cases with deleterious affects. The table 2 (below) shows genes that are previously reported for their role in the heart dysfunctions with nucleotide changes for each of them. These nucleotide changes are not previously reported and are only found in our data from these individuals.

The sheet 2 of supplementary data contains the possible exonic deletions/insertions in the genes those were previously noted for their ability to cause heart defects. This sheet represents the names of the genes, the type of Indel (deletion or insertion), also the functions and disease association of these genes, the microRNA non-mediated decay (NMD); either the Indel responsible for NMD or not, subjects, exon number and so on. The possible affects of a amino acid variant on the protein stability was analysed by using MUpro database (Table 2), this platform known for predicting how single site amino acid change could affect the overall dynamics and stability of a protein.

Table 2: Newly discovered data set of single nucleotide variants in respective genes with *in-silico* results of their disease causing ability. (Symbols: 1 = index patient 1, 2 = index patient 2, 3 = father of the index patient 3, 4 = mother of the index patient 3)

Gene	Nucleotide Change	Amino acid change (Amalut prediction)	Protein Stability prediction (MUpro - results)	Individual	Mutation Taster - prediction	Polyphen / Sift prediction	Reference
ILF3	A1378G	Highly conserved	No results	**1**	Disease causing	Possibly damaging	Associated with MI (Yoshida et al. 2011)
CUL7	T674G	Highly conserved	Decrease	**1**	Disease causing	Probably damaging	Association with experimental MI (Hassnik et al 2009)
SRFBP1	C1244A	Moderately conserved	No results	**1**	Disease causing	Probably damaging	Cardiac gene regulator (Zhang X et al 2004)
MAPK14	G557C	Highly conserved	Decrease	**1**	Disease causing	Probably damaging	Cardiac therapeutic potential (Yeghiazarians Y et al 2012)
AKT1	A1255C	Highly conserved	Decrease	**1**	Disease causing	Possibly damaging	Developmental heart defects (Chang et al. 2011)
ALOX12	A563C	Highly conserved	Decrease	**1**	Disease causing	Probably damaging	Risk factors for atherosclerotic plaque burden and CHD events (Burdon et al 2010)
MMP2	T1676A	Highly conserved	Decrease	**2**	Disease causing	Probably damaging	Possible association with cardiac malfunctions (Mishra et

							al. 2012)
THBS4	T272G	Highly conserved	Decrease	**2**	Disease causing	Probably damaging	A direct influence heart function and to identify (Frolova et al 2012)
ACACB	T3514G	Highly conserved	Decrease	**2**	Disease causing	Probably damaging	SNP marker for LDLC (Ruano et al 2010)
ACVR2B	A767T	Moderately conserved	No results	**3**	Disease causing	Possibly damaging	Associated with atrial, ventricular septal defects; (Ma L et al 2012)
CASP3	A697G	Moderately conserved	Decrease	**3**	Disease causing	Possibly damaging	KD susceptibilit-y (Tremoulet et al 2012)
SHROOM3	G5035C	Moderately conserved	Decrease	**3**	Disease causing	Probably damaging	Heterotaxy-spectrum cardiovascular disorders (Tariq et al. 2011)
PTPLAD1	A350C	Highly conserved	Decrease	**3**	Disease causing	Probably damaging	Red blood cell differentiation and other traits (Ding et al 2012)
LMCD1	T818G	Highly conserved	Decrease	**3**	Disease causing	Probably damaging	Co-regulator for GATA4/ regulate hypertrophic process (Luosujärvi et al. 2012)

RYR3	G5963C	Highly conserved	Decrease	**3**	Disease causing	Possibly damaging	SNP associated with HF (Lynch et al 2012)
MYL2	T53G	Highly conserved	Decrease	**3**	Disease causing	Probably damaging	Role in ASDII (Posch et al 2011)
PTK2	G3C	No data	Decrease	**4**	No data	Possibly damaging	Contribute to sarcomere disorganization in DCM (Chu et al. 2011)
ZFHX3	G9453T	Highly conserved	No results	**4**	Disease causing	Probably damaging	Associated with atrial fibrillation (Onouchi et al. 2012)
HIF1A	A930C	Highly conserved	Decrease	**4**	Disease causing	Probably damaging	Association studied with ischemic stroke (Avdonina et al 2012)
GNAI1	T380C	Moderately conserved	Decrease	**4**	Disease causing	Possibly damaging	Signalling in cardiomyopathy pathway (Arejian et al 2009)

Discussion:

The role of genome stability in proper functioning of a cell is well established now. In last decade, the -omic research has advanced a lot with respect to the finding of genetic origins of diseases. With the advancement of next generation sequencing techniques it is now possible to sequence whole genome or a particular region of interest. The new data on the disease genetics has revealed previously unknown disease associated gene variants.

Our result showing quite a few interesting new findings in the exome-seq data from these cases DNA. Our main aim was to generate variants data for individual cases (Table 2). This genes set previously shown to be either involved or associated with heart malfunctions or development, as presented in the given table 2. These genes recently studied for their number of disease implications towards human heart. We now reporting the genetic variations that are not previously reported, these gene variants are new and unique to these cases only. Our analysis shows a substantial amount of evidence that these variants are quite deleterious.

We used publically available databases, SNP predictors; SNP variants disease causing ability predictors, different gene-gene interaction software, and new variants affects prediction software (see material and methods). All these available bioinformatics tools utilised and results detected. These tools precisely predicted the affects of these variants with respect to these individual genes (Supplementary data). Recently we reported the two new variants in the coding region of the PKP2 gene (Dahlgren et al 2011). This gene has previously published for its SNP association with the Arrhythmogenic right ventricle cardiomyopathy. This finding suggested the L366P and L560P variants of PKP2 gene in exon 4 and exon 7 (respectively) are pathogenic.

The new variants found in our data are rare gene variants and with low penetrance; these variants are only present in the our data when compared with SNP data at NCBI, since these genes played role in different types of heart diseases (Table 2) and known to take part in the human heart development and proper functioning in healthy individuals, we think they might work in concert to play an important role in the aetiology of sudden cardiac arrest.

Our results show different genes that involved in the heart pathology have new variants present only in our studied cases. And this is only possible with the inventions and gradual improvements in the sequencing platforms. This also shows the importance of a continuous approach to explore the complex genetic bases of SCD. Our data is novel, unique and credible with respect to the disease aetiology and addition to the role of rare genetic variants in the pathology. We have evaluated this data at number of stages (DNA/Protein) and at number of bioinformatics platforms, and then suggested that this data might be important with reference to the SCD. And if evaluated further we might confirm the possible role of these variants in actual generation of sudden cardiac arrest events in studied cases. This is also possible to predict a genetic evaluation test for patients with cardiac diseases for their possibility of going to SCD based on these findings.

Most of these variants could potentially affect the stability of their respective protein, as predicted by MUpro (online protein stability platform), and it would be very interesting if these finding confirmed on protein level studies, that should help in better understand towards the role of these variants in aetiology of SCD. Presently we are working to sequence the regions containing these variants to establish their presence in the cases, and to role out the possibility of a bioinformatics artefact.

References:

Arejian M, Li Y, Anand-Srivastava MB.2009. Nitric oxide attenuates the expression of natriuretic peptide receptor C and associated adenylyl cyclase signaling in aortic vascular smooth muscle cells: role of MAPK. Am J Physiol Heart Circ Physiol. 296(6):H1859-67.

Avdonina MA, Nasedkina TV, Ikonnikova AIu, Bondarenko EV, Slominskiĭ PA, Shamalov NA, Shetova IM, Limborskaia SA, Zasedatelev AS,Skvortsova VI. 2012. Association study of polymorphic markers of F12, PON1, PON2, NOS2, PDE4D, HIF1a, GPIba, CYP11B2 genes with ischemic stroke in Russian patients
Zh Nevrol Psikhiatr Im S S Korsakova. 2012;112(2):51-4

Burdon KP, Rudock ME, Lehtinen AB, Langefeld CD, Bowden DW, Register TC, Liu Y, Freedman BI, Carr JJ, Hedrick CC, Rich SS.2010. Human lipoxygenase pathway gene variation and association with markers of subclinical atherosclerosis in the diabetes heart study. Mediators Inflamm. 2010;2010:170153.

Chang Z, Zhang Q, Feng Q, Xu J, Teng T, Luan Q, Shan C, Hu Y, Hemmings BA, Gao X, Yang Z. 2011. Deletion of Akt1 causes heart defects and abnormal cardiomyocyte proliferation. Dev Biol. 347(2):384-91.

Chu M, Iyengar R, Koshman YE, Kim T, Russell B, Martin JL, Heroux AL, Robia SL, Samarel AM. 2011. Serine-910 phosphorylation of focal adhesion kinase is critical for sarcomere reorganization in cardiomyocyte hypertrophy. Cardiovasc Res. 1;92(3):409-19.

de Vreede Swagemakers JJM, Gorgels APM, Dubois-Arbouw WI, van Ree JW, Daemen MJAP, Houben LGE, Wellens HJJ. 1997. Out-of-hospital cardiac arrest in the 1990's: a population-based study in the Maastricht area on incidence, characteristics and survival. J Am Coll Cardiol. 30:1500–1505.

Ding K, Shameer K, Jouni H, Masys DR, Jarvik GP, Kho AN, Ritchie MD, McCarty CA, Chute CG, Manolio TA, Kullo IJ. 2012. Genetic Loci implicated in erythroid differentiation and cell cycle regulation are associated with red blood cell traits. Mayo Clin Proc. 87(5):461-74.

Douglas P. Zipes and Hein J. J. Wellens. 2012. Sudden Cardiac Death. http://circ.ahajournals.org/content/98/21/2334.

Frolova EG, Sopko N, Blech L, Popovic ZB, Li J, Vasanji A, Drumm C, Krukovets I, Jain MK, Penn MS, Plow EF, Stenina OI. 2012. Thrombospondin-4 regulates fibrosis and remodeling of the myocardium in response to pressure overload. FASEB J. 26(6):2363-73.

Giuseppe Novelli, Irene M Predazzi, Ruggiero Mango, Francesco Romeo, Jawahar L Mehta. 2010. Role of genomics in cardiovascular medicine. http://www.wjgnet.com/1949-8462office.

Hassink RJ, Nakajima H, Nakajima HO, Doevendans PA, Field LJ. 2009. Expression of a transgene encoding mutant p193/CUL7 preserves cardiac function and limits infarct expansion after myocardial infarction. Heart. 95(14):1159-64.

James TN, St Martin E, Willis PW III, Lohr TO. Apoptosis as a possible cause of gradual development of complete heart block and fatal arrhythmias associated with absence of the AV node, sinus node, and intermodal pathways. Circulation. 1996;93:1424–1438.

Luosujärvi H, Aro J, Tokola H, Leskinen H, Tenhunen O, Skoumal R, Szokodi I, Ruskoaho H, Rysä J. 2010. A novel p38 MAPK target dyxin is rapidly induced by mechanical load in the heart. Blood Press. 19(1):54-63.

Lynch AI, Irvin MR, Boerwinkle E, Davis BR, Vaughan LK, Ford CE, Aissani B, Eckfeldt JH, Arnett DK, Shrestha S. 2012. RYR3 gene polymorphisms and cardiovascular disease outcomes in the context of antihypertensive treatment. Pharmacogenomics J. doi: 10.1038/tpj.2012.22

Ma L, Selamet Tierney ES, Lee T, Lanzano P, Chung WK. 2012. Mutations in ZIC3 and ACVR2B are a common cause of heterotaxy and associated cardiovascular anomalies. Cardiol Young. 22(2):194-201

Mishra A, Srivastava A, Mittal T, Garg N, Mittal B. 2012. Association of matrix metalloproteinases (MMP2, MMP7 and MMP9) genetic variants with left ventricular dysfunction in coronary artery disease patients. Clin Chim Acta. 413(19-20):1668-74.

Morita H, Rehm HL, Menesses A, McDonough B, Roberts AE, Kucherlapati R, Towbin JA, Seidman JG, Seidman CE. Shared genetic causes of cardiac hypertrophy in children and adults. *N Engl J Med* 2008; 358: 1899-1908.

Mittleman MA, Maclure M, Tofler GH, Sherwood JB, Goldberg RJ, Muller JE, Determinants of Myocardial Infarction Onset Study Investigators. Triggering of acute myocardial infarction by heavy physical exertion: protection against triggering by regular exertion. N Engl J Med. 1993;329:1677–1683.

Neuspiel DR, Kuller LH. Sudden and unexpected natural death in childhood and adolescence. JAMA. 1985;254:1321–1325.

Onouchi Y. 2012. Genetics of Kawasaki disease: what we know and don't know. Circ J. 76(7):1581-6.

Posch MG, Waldmuller S, Müller M, Scheffold T, Fournier D, Andrade-Navarro MA, De Geeter B, Guillaumont S, Dauphin C, Yousseff D, Schmitt KR, Perrot A, Berger F, Hetzer R, Bouvagnet P, Özcelik C. 2011. Cardiac alpha-myosin (MYH6) is the predominant sarcomeric disease gene for familial atrial septal defects. PLoS One. 2011;6(12):e28872.

Ruaño G, Thompson PD, Kane JP, Pullinger CR, Windemuth A, Seip RL, Kocherla M, Holford TR, Wu AH. 2010. Physiogenomic analysis of statin-treated patients: domain-specific counter effects within the ACACB gene on low-density lipoprotein cholesterol? Pharmacogenomics. 2010 Jul;11(7):959-71.

Sovari. 2012. Sudden Cardiac death. http://emedicine.medscape.com/article/151907-overview.

Tariq M, Belmont JW, Lalani S, Smolarek T, Ware SM.2012. SHROOM3 is a novel candidate for heterotaxy identified by whole exome sequencing. Genome Biol. 21;12(9):R91.

Tremoulet AH, Pancoast P, Franco A, Bujold M, Shimizu C, Onouchi Y, Tamamoto A, Erdem G, Dodd D, Burns JC. 2012. Calcineurin inhibitor treatment of intravenous immunoglobulin-resistant kawasaki disease. J Pediatr. 161(3):506-512.e.

Yeghiazarians Y, Gaur M, Zhang Y, Sievers RE, Ritner C, Prasad M, Boyle A, Bernstein HS. 2012. Myocardial improvement with human embryonic stem cell-derived cardiomyocytes enriched by p38MAPK inhibition. Cytotherapy. 14(2):223-31.

Yoshida T, Kato K, Oguri M, Horibe H, Kawamiya T, Yokoi K, Fujimaki T, Watanabe S, Satoh K, Aoyagi Y, Tanaka M, Yoshida H, Shinkai S,Nozawa Y, Yamada Y. 2011. Association of polymorphisms of BTN2A1 and ILF3 with myocardial infarction in Japanese individuals with or without hypertension, diabetes mellitus or chronic kidney disease. Int J Mol Med. 27(5):745-52.

Zhang X, Azhar G, Zhong Y, Wei JY. 2004. Identification of a novel serum response factor cofactor in cardiac gene regulation. J Biol Chem. 279(53):55626-32