CYTOGENETIC PROFILING OF B-CELL LYMPHOMAS

MASTER THESIS

SOPHIE BECH RASMUSSEN

Supervisors:

Mette Nyegaard

Michael Toft Overgaard

AALBORG UNIVERSITY

Number of copies: 7

Pages: 79

Appendix number and kind: Appendix A-D are enclosed in the report, Appendix E–F is enclosed on a data disk.

ABSTRACT

The aim of this thesis was to set up a pilot study in regard to screening mutations in Diffuse Large B-cell Lymphoma with the goal to elucidate the prognostic and predictive impact as well as pathogenetic understanding. This was conducted with focus on a particular gene *EZH2*, as a novel somatic mutation was identified in the GCB-subtype of Diffuse Large B-cell Lymphoma, by Morin *et al.* in 2010. The somatic mutation, referred to as EZH2(Y641) was the only one identified across 19 exons and arose only in a codon, coding for Tyrosine 641, placed in the catalytic site of EZH2. To obtain the aims of examination, a PCR-based method, known as High Resolution Melting was applied for mutation screening, followed by sequencing. Samples screened were a broad cohort of lymphomas, but mainly Diffuse Large B-cell Lymphoma, from the retrospective collaborative translational trial termed CHEPRETRO. Samples found mutated was further analysed in regard to genomic alterations, by examining genomic tumour DNA with Cyto2.7M array from Affymetrix and gene profiling based tumor RNA by Affymetrix U_133_plus2 micro array. Lastly a clinical outcome analysis was conducted, comparing mutated EZH2 samples to those with EZH2 wild type, within the first five years following diagnosis.

The detection of the EZH2(Y641) mutation through high resolution melting analysis, was considered successful as all the samples displaying a heteroduplex melting curve, similar to that of the known EZH2(Y641) mutated cell line DB, had their mutational status verified through sequencing. This was achieved using three different templates during the analysis, the different templates were genomic DNA from snap-frozen OCT tissue and from formalin fixed paraffin embedded tissue along with cDNA, synthesized from RNA. This proved the method applicable for mutation screening across various templates.

It was not possible to detect a genomic profile distinct to EZH2(Y641) mutants using the data utilized from the Cyto2.7M array.

Gene expression analysis by the U_133_plus2 microarray platform, revealed 25 differentially expressed genes for patients with the identified EZH2(Y641) mutation when compared to EZH2(wild type) of the GCB-subtype of Diffuse Large B-cell Lymphoma.

The clinical outcome analyses performed in the available cohort of patients showed no significant difference, between patients mutated in EZH2(Y641) compared to EZH2(wt), in regard to the GCB-subtype, treatment with R-CHOP and in general to patients with Diffuse Large B-cell Lymphoma, disregarding subtype.

ABSTRAKT

Formålet med dette speciale var at opsætte et pilotstudie for mutations screeninger in Diffust Storcellet B-celle Lymfom, med det endelig mål at belyse den prognostiske og prædiktive effekt såvel som patogenetiske forståelse. For pilotstudiet blev et enkelt gen udvalgt til mutation screening, dette gen, EZH2, blev udvalgt på baggrund af et studie af Morin *et al*. i 2010, hvor en somatisk punkt mutation blev identificeret. Den somatiske punkt mutation opstod kun i det samme codon, kodende for tyrosine641, placeret i det katalytiske site i EZH2. For at mutations screeene danske patienter diagnostiseret med lymfomer, dog hovedsageligt diffust storcellet b-celle lymfom, blev en PCR-baseret metode (High Resolution Melting) opsat efterfulgt af sekventering. Prøverne benyttet til mutationsscreeningen, hører under den retrospektive kollaborative translationelle undersøgelse, CHEPRETRO. Efterfølgende mutationsscreeningen undersøges muterede EZH2 prøver med henblik på genomiske ændringer, ved at undersøge den genomiske tumor DNA med Cyto2.7M micro array fra Affymetrix. Yderligere undersøgelser blev foretaget, her i blandt blev en genprofil af påvirkede gener udarbejdet baseret på analyse af tumor RNA på Affymetrix U_133_plus2 mikro-array. En overlevelsesanalyse blev udarbejdet for at få indblik i den prognostiske værdi af EZH2(Y641) mutationen, inden for de første fem år efter diagnosen.

Mutationsscreeningen af EZH2 med High Resolution Melting, blev betragtet som vellykket, da samtlige prøver som indikerede en mutation, gennem en anderledes smeltekurve, fik deres mutation status verificeret ved sekventering. Tre forskellige forsøgsopsæt blev opsat indenfor rammen af mutationsscreeningen, således at forskellige startmaterialer, som genomisk DNA fra lyn frosset tumor biopsier og fra formalin fikseret paraffin indstøbt tumor biopsier, sammen med cDNA syntetiseret fra RNA, kan benyttes.

Det var ikke muligt at påvise en genomisk profil begrænset til EZH2 (Y641) mutanter ved hjælp af de data, produceret ved Cyto2.7M array. Derimod afslørede genekspressionsanalysen af data fra U_133_plus2 microarray platformen, 25 differentielt udtrykte gener, mellem muterede EZH2(Y641) og EZH2(vildtype), af GCB-typen af diffust storcellet B-celle lymfom.

Overlevelsesanalyserne udført for de tilgængelige patientprøver viste ingen signifikant forskel mellem patienter muterede i EZH2(Y641) sammenlignet med EZH2 (vildtype), med hensyn til GCB-subtype, behandling med R-CHOP og generelt til patienter med diffust storcellet B-celle lymfom, hvor subtypen tilsidesættes.

PREFACE

This Master Thesis has been conducted by Sophie Bech Rasmussen and was execute on the 3-4th semester of the Master in Medical Biotechnology Engineering at the Department of Biotechnology, Chemistry and Environmental Engineering at Aalborg University in 2010/2011.
The thesis was conducted in collaboration with the Department of Hematology Research at Aalborg Hospital.
The title of the thesis is Cytogenetic Profiling of B-cell Lymphomas and focuses on a somatic mutation in *EZH2* and its role in Diffuse Large B-cell Lymphoma. This is done in order to gain a more solid foundation in the knowledge of the genetic alterations found in Diffuse Large B-cell Lymphoma.
The Thesis is divided into chapters, sections and under sections. References are based on the Harvard method: [Author(s) surname, published year]. Undated references are shortened Undat. If references have the same surname and year, the references will be differentiated via capital letters e.g. [surname A, year]. Appendix A-D are found in the report while Appendix E&F are enclosed on a data disc and contain raw data, Excel data sheets, etc.
The project addresses fellow students, supervisors and others with interest in the investigated subjects.
Lastly acknowledgements towards the staff of the Department of Hematology Research at Aalborg Hospital should be placed for their kind assistance throughout the scope of this thesis.

Aalborg University, December 6th 2011.

_____

Sophie Bech Rasmussen

ABBREVIATIONS

ABC	– Activated B-Cell like
bp	– base pair
CN-LOH	– Copy Neutral Loss Of Heterozygosity
CNV	– Copy Number Variation
DLBCL	– Diffuse Large B-Cell Lymphoma
dsDNA	– double stranded DNA
EZH	– Enhancer of Zester Homolog
FL	– Follicular Lymphoma
GCB	– Germinal Center B-cell like
HRM	– High Resolution Melting
LOH	– Loss Of Heterozygosity
NHL	– Non-Hodgekins Lymphoma
Q-PCR	– Quantitative Polymerase Chain Reaction
PMBL	– Primary Mediastinal B-cell Lymphoma
PCR	– Polymerase Chain Reaction
PRC	– Polycomb Repressive Complex
R-CHOP	– Rituximab, Cyclophosphamide, Doxorubicin, Vincristine and Prednisolone
SNP	– Single Nucleotide Polymorphism
ssDNA	–single stranded DNA
Tm	– melting Temperature
wt	– wild type

TABLE OF CONTENTS

Introduction 11
Lymphomas 11
Molecular Basis of Cancer 13
Epigenetics 15
Polycomb Repressive Complex Proteins 16
Enhancer of Zester Homolog 2 (EZH2) 17
Enhancer of Zester Homolog 2 and Cancer 20
Thesis Statement 23
Clinical Protocol, Materials and Experimental Set Up 23
Methodology 25
Patient and Tissue Characteristics 25
Sample Preparation 25
Isolation of Genomic dna from Paraffin Embedded Tissues 27
cDNA Syntesis 27
Array Analysis 27
Affymetrix® Cytogenetic Assay 28
ABC/GCB Classification 28
Mutation Screening 28
Primer Design 28
High Resolution Melting Analysis 30
High Resolution Melting Analysis – Paraffin Embedded Tissue Samples 32
High Resolution Melting Analysis – cDNA 32
Statistical Analysis 32
Geneexpression Analysis Using Partek 32
Clinical Outcome – Kaplan and Meyer 33
Results 34
Theoretical Occurrence of Y641 Mutation 34
High Resolution Melting 35
High Resolution Melting – Sample Cohort 1 and 2(Genomic DNA) 37
Sample Cohort 4 (Genomic DNA From Paraffin Samples) 38
Sample Cohort 3 (cDNA) 41
Sequencing 43
Affymetrix® Cytogenetics Assay Results 45
Gene Expression 46

Clinical Outcome 52
Discussion 55
Detection of EZH2(Y641) in Danish Lymphoma Patients 55
Correlation of Genomic Alterations 56
Gene Profile Based on Gene Expression 56
Clinical Outcome Analysis 58
Conclusion 59
Future perspectives 59
References 61
Appendix – A 65
High Resolution Melting 65
Appendix – B 67
Microarray 67
Appendix – C 69
R-Script for Clinical Outcome Analysis 69
Appendix –D 74
High Resolution Melting Results 74

INTRODUCTION

In Denmark, approximately 16000 people are diagnosed with cancer every year. Their relative survival percent reach around 70% if the cancer is treated within the first year. [Kræftens Bekæmpelse, 2009] By understanding the molecular basis of cancer, it is possible to further improve the overall survival among cancer patients, by creating a platform for more target specific treatments.

The foundation for cancer is genetic changes, which will allow excessive and unregulated proliferation that is independent of regulated growth stimuli driving normal cells into malignant cancer cells. Cancer is therefore not a defined singular disease but a high variety of diseases, with a profound dysregulation as a common denominator, which can occur in any part of the human body. [Kumar *et al.*, 2007] This thesis will focus on B-cell lymphomas and more specifically Diffuse Large B-Cell Lymphomas (DLBCL) and related genetic changes. The following section of the thesis will contain an introduction to lymphomas and DLBCL along with the genomic alterations, which can contribute to tumorgenesis.

LYMPHOMAS

Lymphoid neoplasms, tumors of the white blood cells, can be divided into non-Hodgkins lymphomas, Hodgkins lymphomas, lymphocytic leukemias, plasma cell dyscrasias and other related disorders. The lymphoid neoplasms are a group with a wide variation amongst their clinical presentations and behavior and based on these, divided into two main subtypes; Leukemias and Lymphomas. Lymphomas are divided further into two types; non-Hodgkins lymphomas (NHLs) and Hodgkins lymphomas. [Kumar *et al.*, 2007] Of these two subtypes, NHL's are the most common, with 900 new cases every year in Denmark [Kræftens Bekæmpelse, 2011]. It is through the many steps of B-cell differentiation that lymphomas occur, creating several different types of NHL, where the most frequent is DLBCL. Diffuse Large B-cell lymphomas are characterized by their B-cell phenotype, a diffuse growth pattern and an aggressive clinical history, it can evolve *de novo* or transform from Follicular Lymphoma (FL) [Kumar *et al*. 2007, Swerdlow *et al.,* 2008]. Diffuse Large B-cell Lymphoma can occur in patients of any age, but often presents at the median age of 60 years [Kumar *et al.*, 2007]. Diffuse Large B-cell Lymphomas are classified into three molecular subtypes, based on gene-expression profiling. The three molecular subtypes differ in their gene expressions as they arise from different stages in the B-cell differentiation, see Figure 1

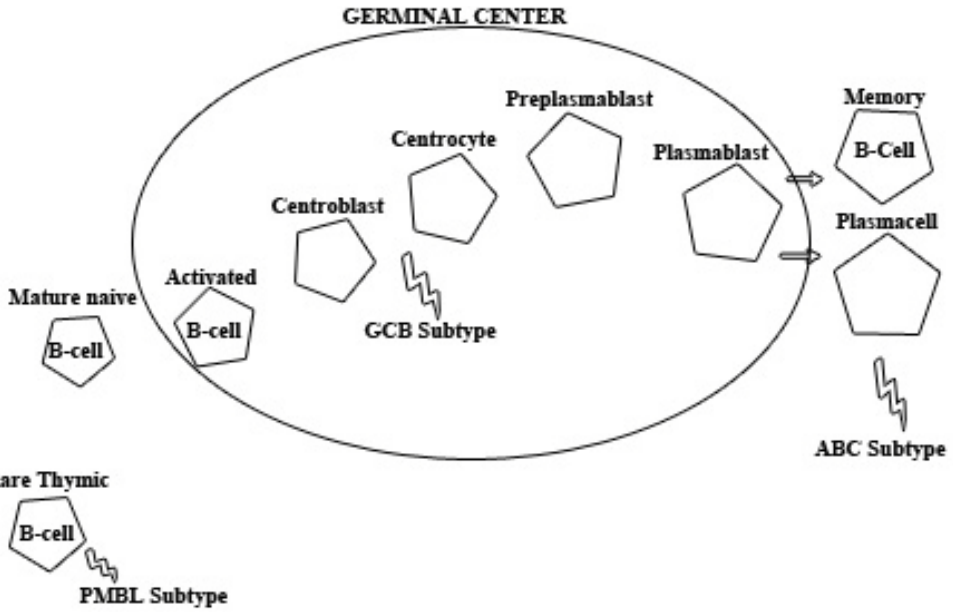

Figure 1 illustrates a simplified schematic of stages of B-cell differentiation, and the basis for the GCB, ABC and PMBL subtype of DLBCL. The ABC-subtype is characterized by plasma-cell gene expression and the GCB-subtype characteristics are an expression of genes, which define germinal-center B-cells. The third subtype PMBL, originates from a rare thymic B-cell. Figure simplified from Lenz and Staudt, 2010.

The three subtypes have a distinct variation in their clinical presentation, clinical outcome and responsiveness in regard to chemotherapy. The three subtypes are; Activated B-cell like (ABC), Germinal Center B-cell like (GCB) and Primary Mediastinal B-cell Lymphoma (PMBL). The ABC-subtype is characterized by plasma-cell gene expression and expression of the transcription factor IRF4, which pushes plasmablasts into differentiation of plasma-cells [Lenz and Staudt, 2010] giving the ABC subtype, a gene profile corresponding to activated peripheral B-cells. The GCB-subtype characteristics are an expression of genes, which define germinal-center B-cells. [Swerdlow *et al.,* 2008]

The third subtype PMBL, originates from a rare thymic B-cell. [Lenz and Staudt, 2010] Due to its origin, PMBL is not in focus of this thesis, and will therefore not be elaborated on further.

One of the distinct differences between the ABC and GCB subtype is that patients with a molecular defined subtype GCB in DLBCL, show a radically improvement in their clinical outcome compared to patients with the ABC subtypes. [Swerdlow *et al.,* 2008]

MOLECULAR BASIS OF CANCER

When tumorgenesis evolves, it is often based on one or several types of genetic aberrations leading to what is known as one or more of the six hallmarks of cancer. The six hallmarks of cancer are the different capabilities, which a tumor will acquire in order to obtain excessive and unregulated proliferation. [Hanahan&Weinberg, 2011] As the tumor progresses, so will the genetic aberrations, increasing the capabilities of the tumor to evade control from the host. [Kumar *et al.*, 2007] The six hallmarks of cancer can be seen in Figure 2. The different hallmarks and the pathways tumor cells use to obtain these capabilities are highly complex and diverse mechanistic strategies, which vary in each cancer, yet somehow result in a general framework. [Hanahan&Weinberg, 2011] To elaborate on each of the six hallmarks and known genetic aberrations leading to these, is beyond the scope of this thesis.

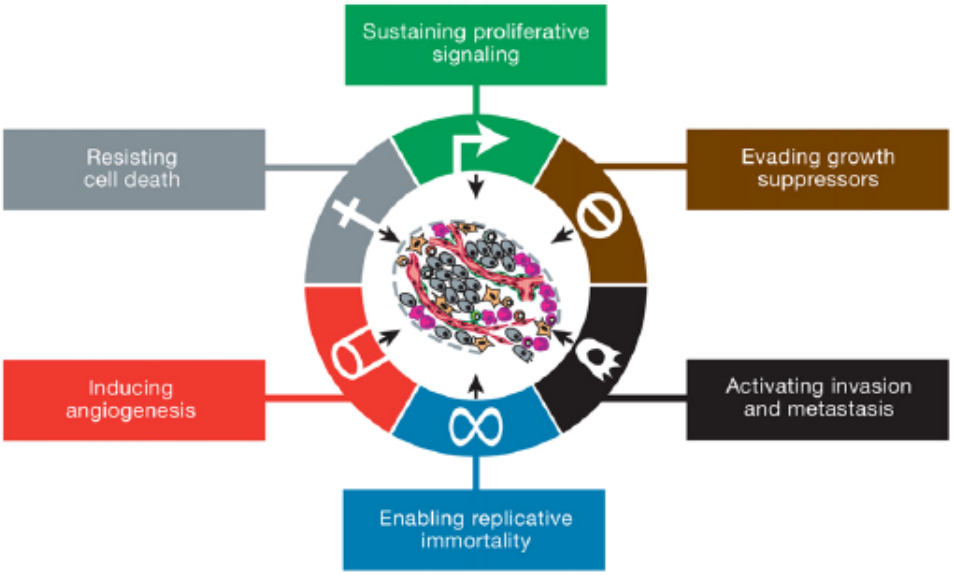

Figure 2 illustrates the six hallmarks of cancer. The capabilities obtained through tumorgenesis to ensure excessive and unregulated proliferation. [Hanahan&Weinberg, 2011]

As a cancer cell often contain a broad array of genetic aberrations, some of the most prominent types will be elaborated on here. Yet it is important to note that not all the genetic aberrations found in cancer cells are associated with tumorigenesis - these genetic aberrations are known as passenger aberrations, while the genetic changes involved in tumorigenesis are called driver aberrations [Buntz, 2008] A genetic aberration type that is common in cancer is mutations, which is a permanent change in DNA. There are up to several different types of mutations that can occur, with different properties. A germline mutation will be inherited while somatic mutation is found in somatic cells and therefore cannot be inherited by progeny. Somatic mutations are often found to be actuators in cancer.[Kumar *et*

al., 2007] If a mutation is not inherited through germline cells or occurring in somatic cells, viral infections have also been found to be actuators towards tumorigenesis, yet this is less common. [Bunz, 2008]

There are different variants of mutations, for example a point mutation is a substitution of a single nucleotide, resulting in a codon alternation. Another type of mutations that can occur is frameshift mutations, which are deletions/insertions of one to two basepairs (bp) resulting in a change in the reading frame of the DNA. [Kumar *et al.*, 2007] In general it is only mutations occurring in exons, which are considered to have an impact on gene function and protein structure, as in comparision of mutations occurring in introns, however exceptions have been known to occur. [Buntz, 2008]

There is also genetic variation of 1000 bp or larger, this is defined as copy number variation. Copy number variation (CNV) is a change in the genomic DNA, consisting of polymorphisms in the number of copies of chromosomal segments and the number of genes in those segments. This involves either a gain of a segment or gene (addition) or a loss (deletion), changing the original diploid status in the affected gene segments. [Shlien and Malkin, 2010]

A loss of heterozygosity (LOH) can occur through a loss of chromosomal segment in a somatic cell in areas were proportions of the genetic markers in the human genome are heterozygous. [Lodish *et al.*, 2008] When LOH is followed by duplication of the remaining gene marker, it is referred to as copy neutral loss of heterozygosity (CN-LOH). [O'Keefe *et al*., 2010] The changes that affect the diploid status are referred to as numeric abnormalities, while changes that affect the structure of chromosomes are referred to as structural abnormalities. An example of a structural abnormality is chromosomal translocations, which in brief are shuffling of genes or segments between chromosomes resulting from chromosomal breakage. [Kumar *et al.* 2007]

Mosaicism is a term used to describe the co-exsistance of two different cell populations in a sample. In cancer samples, mosaicism can indicate either the presence of two or more populations of cells, for example tumor cells and normal cells. It indicates a genomic alteration, and depending on the distrubution of cell populations, it can be indicated together with a gain or loss, however the average copy number between cell populations yield an intermediate.[Kumar *et al*., 2007]

The two subgroups of DLBCL, GCB and ABC, also are differentiated through genetic alterations. Some of the known alterations are different occurring chromosomal alterations, with gains occurring on chromosome 12q12 for the GCB subgroup and 3q, 18q21-22 for the ABC subgroup. The ABC subgroup is also known to show loss on chr6q21-22. [Swerdlow *et al.,* 2008] On a gene level, a preferential association of alterations in *BLC2* and *MYC* has been linked to the GCB-subtype, while for the ABC-subtype alterations in the *NF-Kβ, BCL6-BLIMP1* axis has been observed. [Pasqualucci *et al.* 2011]

EPIGENETICS

Genetic alterations are not the only effectors towards cancer initiation and progression, as epigenetic aberrations also are found to be influential [Chi *et al.*, 2010]. Epigenetic inheritance is inheritance through modifications of chromatin structure instead of DNA sequence modification, meaning that a regulation of gene expression occurs, which is independent of DNA sequence changes, yet still inherited by daughter cells. [Lodish *et al.*, 2008] The modification of chromatin structure consists of several different mechanisms that center around post-translational modifications of histone tails, DNA methylation, histone variants and chromatin compaction [Vaillant and Paskowski, 2007].

The two regulatory systems of DNA methylation and histone modifications occur in context to each other and whereas DNA methylation is a direct modification of DNA while histone modifications are post-translational modifications (PTMs) of proteins, located in the chromatin. Histones are known as DNA-packing proteins and are the most abundant protein found in chromatin. They are divided into five major types; H1, H2A, H2B, H3 and H4. Histones are small basic proteins that can interact with negatively charged phosphate groups in DNA, due to the high amount of positively charged basic amino acids found in their structure. [Lodish *et al.*, 2008] Histone modifications are utilized to modify chromatin structure along with having a controlling impact in DNA repair, replication and recombination. The effect of histone modifications depend on the utilized posttranslational modification. [Bannister and Kouzarides, 2011]

The posttranslational modifications of histones can mainly be differentiated between phosphorylation, methylation, acelyation and other modifications such as ubiquityaltion and sumoylation, among others. The most renowned histone modification is methylation, which occurs on side chains of lysines and arginines. The degree of the methylation depends on the amino acid, as lysine can be mono-trimethylated and arginine mono-dimethylated. Despite the varying degrees of methylation, there is no effect the original charge of the histone. Methylation is regulated by methyltransferases and demethylases, however it should be noted that depending on the amino acid, the two groups consist of different groups of enzymes. Lysine methytransferases transfer a methyl group from S-adenosylmethionine (AdoMet) to the lysine ε-amino group, while arginine methyltransferases facilitates the methyl group to the ω-guanidino group of arginine. The remaining type of histone modifications will not be elaborated on except for a short notation that ubiquitylation is the addition of ubiquitin to histone lysines. [Bannister and Kouzarides, 2011]

POLYCOMB REPRESSIVE COMPLEX PROTEINS

Polycomb repressive complex proteins (PRC1 and PRC2) are two known multi protein complexes, involved in gene silencing, through posttranslational modifications of histones.

The two complexes have very different compositions, PRC1 is found only to have two core components, RING1A/B found with either, BMI1, MEL18 or NSPC1 respectively, leaving the rest of the multi protein complex to be variable. Polycomb repressive complex 2 on the other hand has four core components; EZH1/EZH2, SUZ12, EED and RbAp46/48. Among these, EZH1/2 is the catalytic subunits of PRC2. Polycomb Repressive Complex 2 with the EZH1 subuint is found in divided and differentiated cells, while PRC2-EZH2 is found only in actively dividing cells. [van Kemenade, 2001] The two enzymes also have difference in their methyltransferase activity, with PRC2-EZH1 having the lowest activity, indicating that the PRC2-EZH2 complex generates histone methylation, while the EZH1 restores methylation that might have been lost through demethylation. [Magueron *et al.,* 2008] Notably EZH1/2 lacks its enzymatic properties, when separated from the PRC2 complex [Simon & Lange, 2008].

In the remainder of this Master Thesis, the focus will be on the PRC2-EZH2 complex and will be referred to as PRC2. Mainly PRC2 is found to be responsible for mono-, di- and tri-methylation of lysine 27 in histone H3 (H3K27me-3), while PRC1 monoubiquilates lysine119 on histone H2A, along with being able to bind to the H3K27me3 mark left by PRC2. [Margueron and Reinberg, 2011] Upon PRC1 binding to H3K27me3, PRC1 will block initiation of gene activation by RNA polymerase II, thereby maintaining gene repression, see Figure 3. [Lodish *et al.*, 2008] PRC1 is not the only complex that can bind to the trimethylation of H3K27, as another subunit in the PRC2 complex, EED, also can bind PRC2 to H3K27me3, this helps maintain the repressive mark through allosteric activation of PRC2 methylation activity. [Margueron *et al*, 2009] Upon binding by EED a positive feedback loop is created. The initial recruitment of PRC2 remains to be elucidated. [Margueron and Reinberg, 2011]

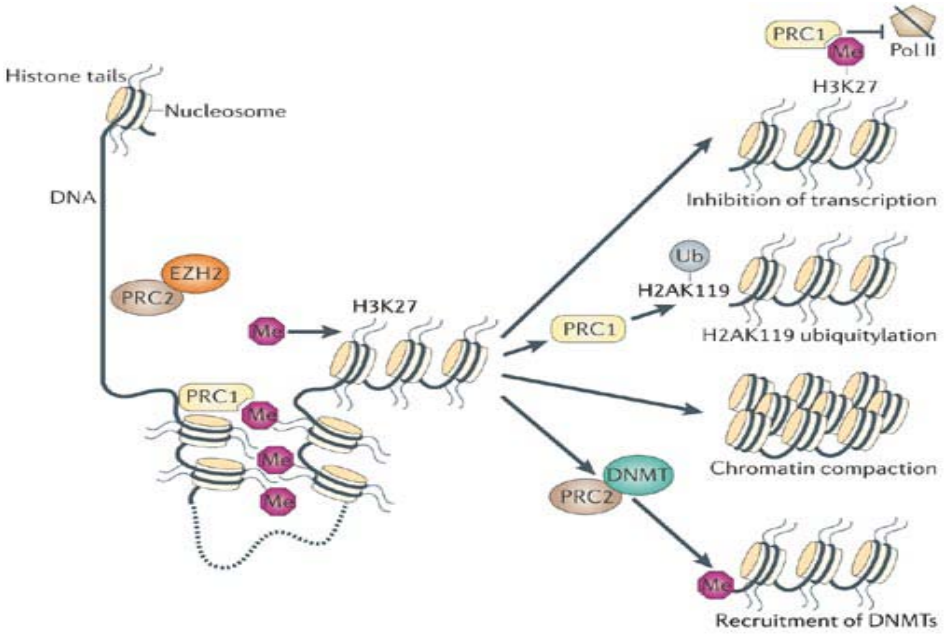

Figure 3 displays the possible procession of H3K27me3, PRC1 can block gene activation by RNA polymerase II or ubiquitylate H2AK119, while PRC2 can aid in the recruitment of DNAmethyltransferases. Both complexes are suggested to be involved in chromatin compaction.[Sparmann and Lohuizen, 2006] Not shown in this figure, is the binding of EED in PRC2 that creates an allosteric activation in PRC2 that increases the complex' methylation activity. [Margueron *et al*, 2009]

Enhancer of Zester Homolog 2 (EZH2)

Enhancer of Zester Homolog 2 and its methylation activity has by Velichutina *et al*. been hypothesized to have several functional roles in normal germinal center B-cells. The suggested functional roles include favoritism of cellular proliferation by repressing tumor suppressor genes, creation of a repression state resembling a stem-cell to foster self-renewal and prevention of premature differentiation and lastly to maintain a repression, already utilized in the previous state of B-cells – naïve B-cells. [Velichutina *et al*. 2010]

The gene encoding EZH2 is located on chromosome 7q36.1 and is 40kb bases long, stretched across 19 exons. The EZH2 protein consists of 746 amino acids and belongs to a highly conserved family of SET-domain proteins, that all but one, are known for histone methylation. [Dillion *et al,* 2005]

The SET-domain residing in EZH2 has yet to be structurally characterized. As the catalytic and Adomet site is highly conserved in SET-domain protein super family, see Figure 4, a general consensus for the catalytic activity of the SET-domain is widely accepted. [Schapira, 2011; Dillion *et al,* 2005]

The general consensus for the conformation of the catalytic site and facilitation of methylation by the SET-domain protein super family will be explained in the following.

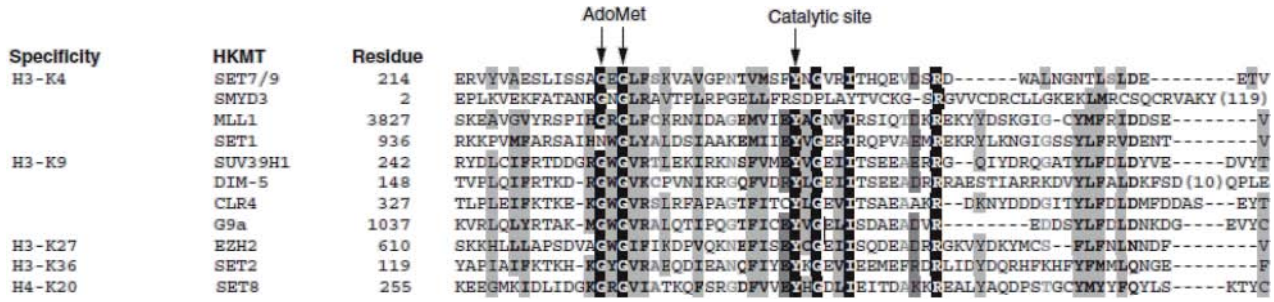

Figure 4 shows the highly conserved AdoMet and Catalytic site across various histone lysine methyltransferases within the SET-domain superfamily. [Dillion *et al,* 2005]

In the general consensus for the SET-domain, there are two binding sites, which are located at opposite ends of the domain. Between the two binding site-clefts is a channel running through the core of the SET-domain, where the catalysis occurs. This channel allows transferral of the AdoMet from the first binding site, to the lysine ε-amino group, at the second binding site. This conformation is suggested to allow mono-trimethylation of lysine before the protein substrate dissociates from the SET-domain. [Dillion *et al,* 2005] The electropositive histone tails and an electronegative binding grove of the SET-domain are attracted to each other through long range electrostatics. The histone lysine methyltransferase (HKMT) will then travel along the histone tail, held in place by the electrostatic connection, until it reaches a specific sequence of histone sidechains within the histone tail. Upon which, the lysine will deprotonate and lock the complex into a catalytic conformation, where a catalytic tyrosine is located at the C-terminal of the SET-domain. This catalytic conformation will complete the channel towards lysine, projecting towards the active site. The catalysis then occurs, with the methyl group being in close proximity of the deprotonated lysine ε-amino group. The hydroxyl end of tyrosine, along with main chain carbonyl oxygens will enhance the nucleophilicity of the methyl group while another tyrosine form a hydrogen bond with lysine, aligning the deprotonated ε-amino group with a methyl sulfur scissile bond. Once this conformation is obtained, a nucleophilic attack will follow, leaving lysine to be methylated and the release of S-adenosyl-homocysteine(AdoHcy), as a byproduct. See Figure 5, for a schematic overview. [Schapira, 2011; Dillion *et al,* 2005]

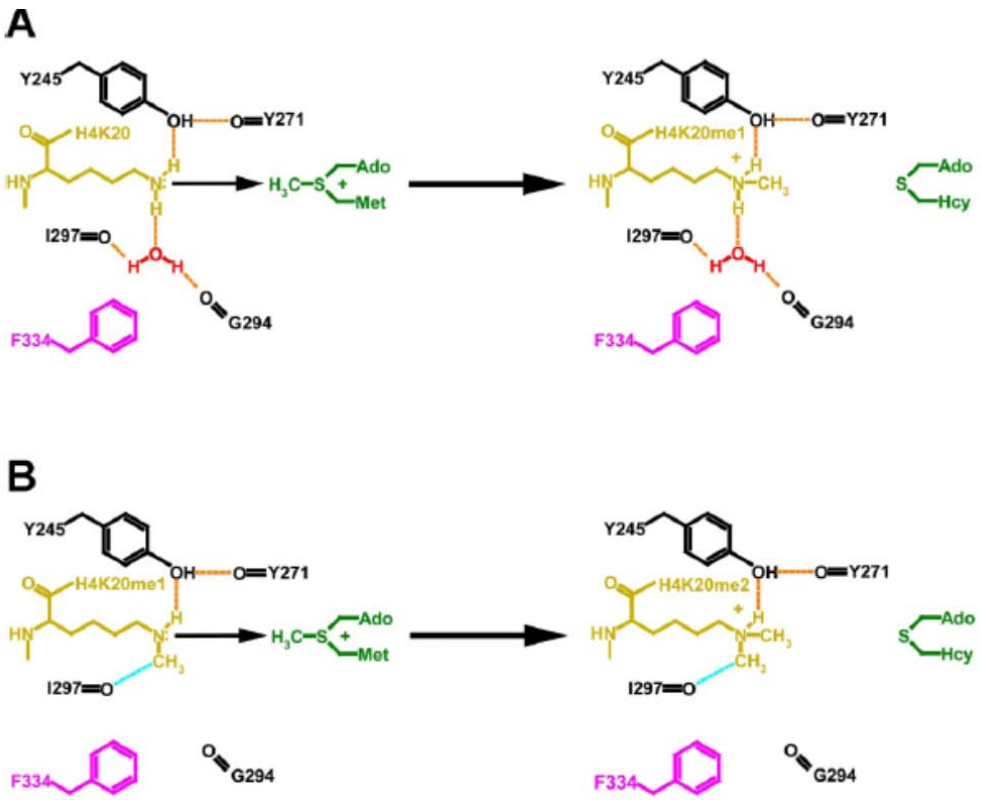

Figure 5 gives a schematic representation of histone methylation from me0 to me2. As a model for the mechanism H4K20 and the active site of SET8 is used. A) show monomethylation, where the hydroxyl group of tyrosine and the active water molecule in the enzyme functions as hydrogen bond acceptors that align the deprotonated ε-amino group with a methyl sulfur scissile bond. B) show mono to dimethylation. Here the monomethylated sidechain is bound in an alternative conformation with the methyl group in the binding pocket, forming a CH-O bond with isoleucine in the enzyme. By facilitating this conformation the ε-amino group is oriented for another methyltransferase from the AdoMet and hydroxyl bond from tyrosine. Modified from [Couture *et al.*, 2008]

A Phe/Tyr switch in SET domain has been identified, which governs the product specificity of the given lysine methyltransferase. The model suggests, that a SET domain containing a tyrosine in the catalytic site are partial to facilitating monomethylation, while a phenylalanine or another hydrophobic residue are partial to facilitate di- or trimetylation. A recent study shows that this mechanism can be modulated by binding a water molecule at the active site. [Del Rizzo *et al.*, 2010; Schapira, 2011]

Enhancer of Zester Homolog 2 and Cancer

After aberrant histone modifications were linked to cancer, EZH2 was hypothesized to be an enzyme capable to alter the properties of cancer cells. This hypothesis was confirmed in prostate cancer, where EZH2 gene expression was significantly up-regulated in metastatic prostate cancer compared to clinically localized prostate cancer. The study, one of the earliest of EZH2's relation to cancer, also showed that EZH2 levels of gene expression could be used as a predictor in regard to clinical outcome. Following this discovery, several other cancer-types, such as breast cancer, have also been found to have over-expression of EZH2 and that the level of expression can be linked to poor clinical outcome. [Simon&Lange, 2008]

In the search for unidentified mutations that could potentially contribute to NHL, a somatic mutation in exon15 of EZH2, was identified in a patient with Follicular Lymphoma (FL) by Morin *et al*. in 2010. Inquisitive by this finding, 672 NHL samples were screened for this mutation, revealing that the mutation was present in 7,2% in FL and 21,7% of DLBCL –GCB subtype. The mutation was absent in DLBCL samples of the ABC subtype and in the other types of lymphoma screened (Mantle-cell, small lymphocytic and peripheral T-cell lymphoma). [Morin *et al*., 2010]

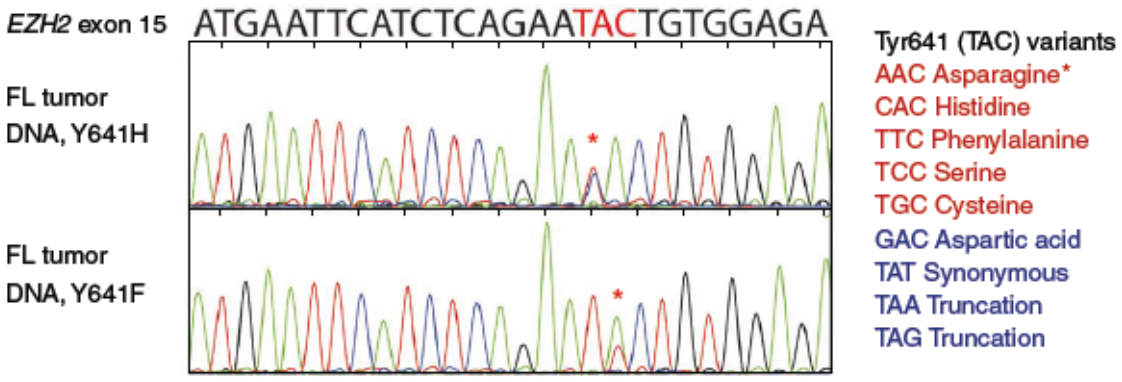

Figure 6 illustrates the sequencing results from Morin *et al*, 2010 and the possible variants of Y641, based on a single point mutation.

As EZH2 has been found to be involved in B-cell differentiation [Velichutina *et al*. 2010], it is naturally abundant in B-cells, which indicates that EZH2 involvement in lymphoma, constitutes of a different mechanism than of those previously seen in prostate and breast cancer [Morin *et al*., 2010].

The amino acid affected by the somatic mutation was tyrosine (Y641) in the catalytic site of EZH2. The amino acid mutation variants of Y641 discovered by Morin *et al*. were; Tyrosine to Phenylalanine (49%), Serine (21%), Asparagine (15%), Histidine (13%), and Cysteine (1%). Every gene with an Y641 variant was heterozygous and each variant can be incorporated into the catalytic site of EZH2. [Morin *et al.,* 2010] In the remainder of this thesis, mutated variants of EZH2, will be referred to as EZH2(Y641). Even though each variant of Y641 can be incorporated into the catalytic site, they show

little to no ability to perform the first methylation but in contrast, have enhanced catalytic efficiency for di- and trimethylation of H3K27, see Figure 7. The fact that EZH2(Y641) has little to no capability of facilitating the first methylation, but has enhanced efficiency for di- and trimethylation, indicates a need for coordinated enzymatic activity between the wild-type and mutant, making a heterozygous mutation, the only functional possibility. [Sneeringer *et al.,* 2010] Enhancer of Zeste Homolog 2(Y641) has also been shown to have increased levels of trimethylation, when compared to EZH2. [Yap *et al.*2011]

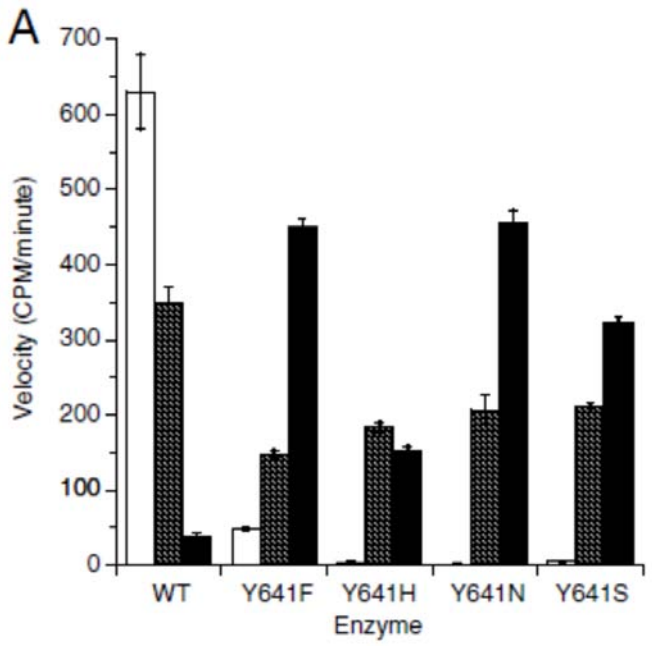

Figure 7 shows the methyltransferase activity of EZH2(wt) and EZH2(Y641) mutant complexes. From left to right, it is unmethylated, mono- and dimethylation (white, grey, black). The diagram illustrates how the wt has enhanced catalytic activity for un-monomethylation, but little to none with dimethylation where as the Y641 variants are opposite showing little to no capability to facilitate the first methylation. [Sneeringer *et al.,* 2010] Cystein is not shown in this schematic, but it also changes the substrate specificity of EZH2, in accordance to the other variations of EZH2(Y641) [Wigle *et al,*. 2011]

In the heterozygous state of EZH2 and EZH2(Y641) an allelic imbalance, leading to a preferential in gene expression for the mutant allele, has been observed. [Morin *et al*., 2011] On a protein level, this was verified by Yap *et al.* as the protein ratio of EZH2(Y641)/EZH2, in a cell line (WSU-DLCL2 (Y641F) was determined to be 60:40, this was determined by mass spectrometry techniques. Together with the fact that EZH2(Y641) shown increased trimethylation activity, it indicates that the EZH2(Y641) is a dominant gain of function mutation. However the targets of the increased trimethylation remain to be elucidated, as non canonical PRC2 targets also may be affected. [Yap *et al.*, 2011]
The mechanism behind the prominent occurrence of the EZH2 mutation, has been hypothesized by Morin *et al.* (2011) to be appointed to a mutational hot spot, which spans from codon 602-646 [Morin *et al*., 2011, sup]. A mutational hot spot is a DNA sequence where mutations occur with a higher

frequency than other areas in the DNA strand. The molecular basis for mutational hot spots and thereby the mechanism for the EZH2(Y641) mutation remains unclear. [Glazko *et al.,*2006] With the frequency of EZH2(Y641) mutation and the fact that it located in a mutational hot spot, it raised the question whether or not EZH2(Y641) is a candidate for a driver genetic aberration. This was addressed through statistical analysis by Morin *et al*.(2011) using a method proposed by Greenman *et al.*. The results of the analysis showed a positive selection for the somatic point mutation, suggesting that the somatic mutation in EZH2 creates a candidate cancer driver mutation. [Morin *et al.* 2011]
As EZH2 previously have been linked to poor prognosis due to over expression, in prostate and breast cancer, Morin *et al.* (2011) analysed 199 DLBCL patients treated with R-CHOP (Rituximab, Cyclophosphamide, Doxorubicin, Vincristine and Prednisolone), a standardized chemotherapy type used for NHL, See Figure 8. In contrast to previous findings of poor prognosis, the difference in clinical outcome for DLBCL with EZH2(Y641) was not significant when compared to DLBCL patients with EZH2(wt). Diffuse Large B-cell Lymphoma patients with EZH2(wt), consisted of both GCB and ABC subtype.

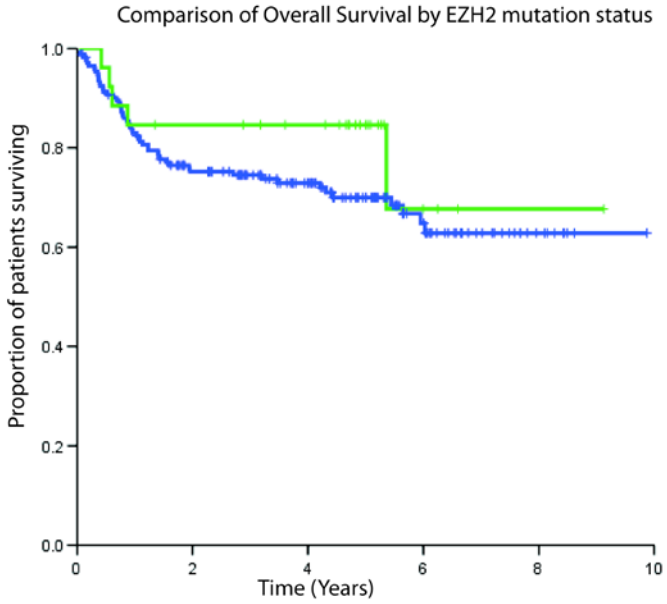

Figure 8 display the overall survival by EZH2(Y641) compared to GCB and ABC DLBCL patients with a EZH2(wt). n=199, p-value= 0,279 (log rank test). The p-value indicates no significant difference in survival outcome for DLBCL patients treated with R-CHOP. [Morin *et al*., 2011, sup]

THESIS STATEMENT

As the EZH2(Y641) mutation discovery is still novel, unanswered questions remains, as in regard to which genes are affected by the elevated trimethylation and gene repression, genomic profiling and clinical outcome. In order to obtain a broader understanding of the EZH2(Y641) mutation and its possible relevance in a clinical aspect in future studies on a bigger patient cohort, the following hypothesis were set up for this pilot study:

To determine if:

- the EZH2(Y641) mutation is present in Danish DLBCL patients, and found exclusively in GCB type of DLBCL and with the same frequency as previously reported (21% of GCB DLBCL), by utilizing a High Resolution Melting analysis approach.
- The GCB(EZH2(Y641)) tumors have a different genomic DNA profile compared to tumors with GCB(EZH2wt) regarding *EZH2* and at karyoview, using Cytogenetic microarray technology.
- The GCB(EZH2(Y641)) tumors have a different gene expression profile compared to GCB(EZH2wt), using gene expression microarray technology
- The GCB(EZH2(Y641)) tumors have a different prognosis in clinical outcome compared to non-EZH2(Y641)mutated DLBCL, through a statistical approach.

CLINICAL PROTOCOL, MATERIALS AND EXPERIMENTAL SET UP

This retrospective pilot study was initiated to prepare estimates used in planning an international prospective validation trial in DLBCL. The study was based on tissue material from the CHEPRETRO project approved by the local Scientific Etich Committee (Region North Jutland # N-20100059 dec 2010) and under the umbrella of the "Region Nordjyllands anmeldelse ved Datatilsynet – Sundhedsvidenskabelig forskning i Region Nordjylland (2008-58-0028)" and granted by The Danish Council for Strategic Research, Programme Commission on Health, Food and Welfare (FøSu),(Lead by Ass. Professor Karen Dybkær).

The experimental set up includes four tissue cohorts consisting of biopsies from lymphoma patients in CHEPRETRO. An overview of the experimental setup is schematically illustrated in Figure 9.

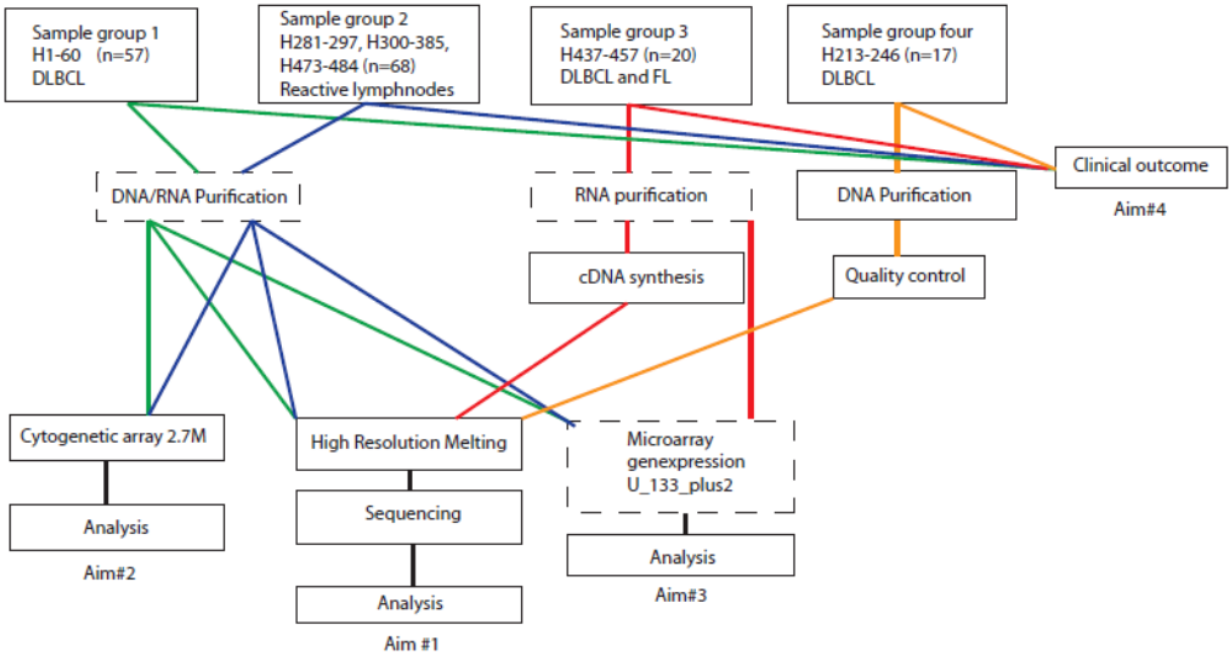

Figure 9 illustrates the experimental setup for this thesis and methods used in order to test the hypothesis in his thesis. The boxes with dashed lines indicate experimental work that was already completed by lab technicians at the Hematology Research Department, Aalborg hospital prior to this thesis.

METHODOLOGY

The methodology of the experiments conducted in this Master Thesis will be described in the following sections.

PATIENT AND TISSUE CHARACTERISTICS

All 162 lymphoma biopsies were diagnostic samples from the Department of Pathology in Odense or Aalborg in the retrospective collaborative translational trial termed CHEPRETRO (Lead Ass Professor PhD Karen Dybkær): A Retrospective validation of a new "Gene test for Chemotherapy Prediction of Resistance" on archived tissue from patients with Malignant Lymphomas. The patients were diagnosed between 1990 and 2008 and followed until 2011, by the National Clinical Quality Database called "LYFO registret" including diagnostic characteristics (prognostic IPI score), treatment and survival outcome..

Patients diagnosed with DLBCL ranged from 20 to 91 years old at the time of diagnosis and had a gender ratio of 49% female vs. 51% male. Characteristics of the 162 patient cohort, which was the basis for this thesis, are listed in appendix E on the supplied CD-rom.

The cell lines used were

OCI-LY7, which was kindly provided by Andreas Rosenwald, Germany.

OCI-LY3, which was kindly provided by Jose A. Martinez-Climent from the Molecular Oncology Laboratory Division of Oncology Center for Applied Medical Research University of Navarra Pamplona, Spain. OCI-LY cell lines are not commercially available and established by The Ontario Cancer Institute (OCI).

The DB cell line was purchased from the German Cell Bank DSMZ.

SAMPLE PREPARATION

The 112 patient samples were stored at primary diagnosis of Diffuse large B Cell Lymphoma (DLBCL) and divided into four sequentially groups (Table 1). Sample cohort 1, 2 and 4 were provided from the Institute of Pathology, Aalborg University Hospital and sample cohort 3 from Odense University Hospital.

Table 1 illustrates the number of samples in each group, used in this thesis

Samples:	Samples in total	DLBCL
Sample cohort 1 (H1-H60)	57	57
Sample cohort 2 (H281-297, H300-385, H473-484)	68	20
Sample cohort 3 (H437-457)	20	18
Sample cohort 4 (H213-246)	17	17

Cohort 1 originated solely from patients diagnosed with DLBCL up till 2006, Cohort 2 from a mixture of patients with abnormal lymph nodes, including patients diagnosed with DLBCL in 2007-2008. Cohort 3 from patients diagnosed with DLBCL and FL in Odense (1998-2005) and lastly cohort 4, from patients with DLBCL (2002-2008).

The samples were numbered H1-H60 (cohort 1), H281-H297, H300-H385 H473-H484 (cohort 2), H437- H457 (cohort 3) and H213-H246 (cohort 4). It should be noted that in sample cohort 2, there are two samples which originate from the same patient, which has been diagnosed with DLBCL twice, as the patient relapsed with DLBCL after complete remission. The two samples are H302 (first diagnose) and H385 (relaps sample), respectively.

The biopsies were received at the Institute for Patology and if tissue material were available following conventional diagnsotic histopathological procedures, the diagnostic biopsies were snap-frozen in OCT from which the DNA/RNA has been extracted. DNA/RNA from sample cohort 1-3 was extracted by a technician at the Research Laboratory Department of Haematology. In short total RNA was extracted using Invitrogen TRIzol Reagent combined with Qiagen RNeasy Mini kit. The quality was checked by Agilent 2100 Bioanalyzer. DNA was extracted using DNeasy Blood and Tissue kit from Qiagen. The quality was checked using Nanodrop spectrophotometer.

Isolation of Genomic dna from Paraffin Embedded Tissues

As part of the thesis 17 samples of formalin fixed paraffin embedded (FFPE) tissues from patients with DLBCL (sample cohort 4) were cut from the original embedded tissue samples using a microtom from micro M. The DNA was then isolated using a QIAMP® FFPE tissue kit. The protocol was followed as directed by the manufacturer, except for an increase in incubation time at 56˚C due to slow lysation of the samples. The incubation was increased with 30 min. Also RNase was not added. Lastly the incubation of the QIAamp Min Elute column loaded with milliQ water, for five minutes to ensure an increase in DNA yield, was performed prior to centrifugation. The DNA concentration was then measured using a Nanodrop spectrophotometer. Twelve of the samples were diluted to a template concentration of 10ng/µL, while concentration of the remaining samples were kept as stock concentration due to low DNA concentration (1,06 -5,09 ng/µL) and elution volume after isolation of gDNA.

The samples are denoted from H213-H246

The samples with stock concentration was; H230, H222, H232, H238, H221.

H246 was excluded based on results from the Nanodrop spectrophotometer.

cDNA Syntesis

20 mRNA samples purified from tissue embedded in OCT from DLBCL patients (sample cohort 3) were converted into cDNA using a First Strand cDNA synthesis approach, prior to high resolution melting. For the cDNA synthesis, SuperScript® III First Strand Synthesis Supermix from Invitrogen was used. Primers used for the synthesis was a 1:1 mix of Oligo$(dT)_{20}$ and random hexamer primers. 2 µL total RNA was added (ranging from 38ng/µL to 261ng/µL) and the protocol supplied by the manufacturer was followed. A thermal cycler from G-storm was used.

The samples are noted as; H437-H457

Array Analysis

Gene expression data based on Affymetrix U_133_plus2 platform were available for all samples in sample cohort 1&3; and for DLBCL samples in sample cohort 2.

All GEPs were performed using the Affymetrix microarray platform and standard procedures. The samples were prepared for hybridization to Affymetrix GeneChip HG-U133 Plus 2.0 arrays after the manufacturer's instruction and .CEL-files were generated by Affymetrix GeneChip Command Console Software (AGCC).

It should be noted that during the gene expression experiments, a change in labeling kit occurred as the initial kit was discontinued from the manufacturer Affymetrix. The first labeling kit used was

Affymetrix Genechip expression 3´ amplification one cycle target labeling and the second was Affymetrix Genechip 3'IVT EXPRESS Kit.

AFFYMETRIX® CYTOGENETIC ASSAY

75 DLBCL samples from sample cohort 1 and 2 were analysed with Affymetrix® Cytogenetic assay (cyto2.7M array). In brief, for each sample 100 ng of genomic DNA was labelled, fragmented, and hybridized to the Cyto2.7M array (Affymetrix) according to the manufacturer's protocol. Staining and washing of the arrays was carried out on a Fluidics Station 450 (Affymetrix) and scanned with a GeneChip scanner 3000 7G system (Affymetrix). CEL files were generated by Affymetrix GeneChip Operating Software (GCOS). CEL files from the Cyto2.7M array were analyzed in Chromosome Analysis Suite version 1.1 (ChAS)(Affymetrix) using single analysis manager with default parameter settings to generate CYCHP-files. All genomic positions are from the hg19 version of the human genome.
The analytic parameters used in ChAS, was a maker count of 50 and segment size of 200 kbp.

ABC/GCB CLASSIFICATION

All samples from sample cohort 1,2 and 3 from DLBCL and FL patients were classified into ABC/GCB subtypes using gene expression array data and a classifier based on an bayesian compound covariate classification [Wright *et al.,* 2003] with probeset list, weights and prior probabilities as described by Lenz et al. (specific details obtained by personal communication with George Wright). In addition to this, the probesets were brought to the same scale as Lenz et al.'s probesets by a rescaling of the probeset-wise standard deviation. This classification was conducted by Martin Bøgsted and Steffen Fallgren, biostaticians at the Haematological Research department of Aalborg Hospital.

Sample cohort four, was classified using immunohistochemistry, by Ken Young, USA (manuscript in prep) according to World Health Organisation criteria of 2008.

MUTATION SCREENING

PRIMER DESIGN

Primers for High Resolution Melting analysis (HRM) where designed with the online tool Primer3 and ordered from Eurofins, MWG operon. The first primer set was designed for EZH2, targeting the entire Exon15. See Figure 10.

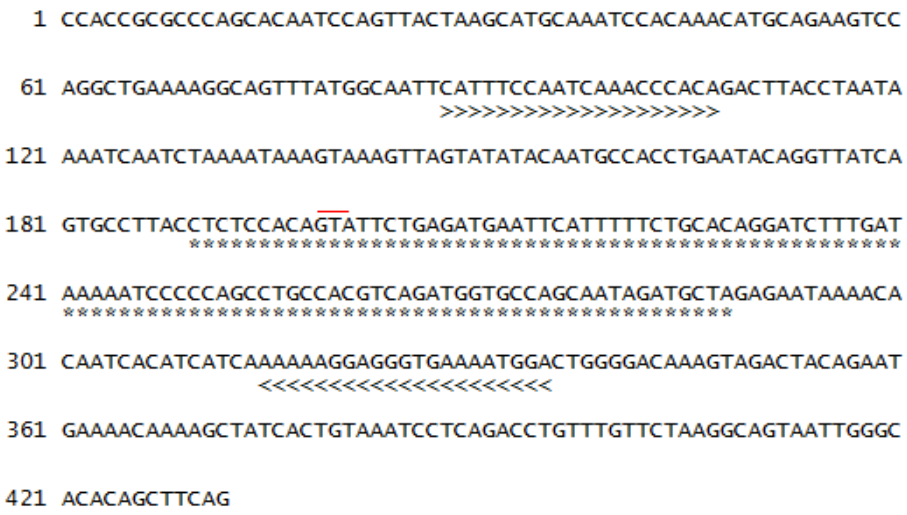

Figure 10 illustrates exon 15 of EZH2, ** indicates exon 15, while >> indicates the left primer position and << the right primer position, respectively, GTA is the codon for tyrosine 641. The sequence shown is the sense strand of DNA, for protein translation, the strand is read antisense and the target sequence for Y641 is therefore TAC.

The primers were used for High Resolution Melting analysis, for sample cohort 1 and 2. See Table 2, for the two primers designed.

Table 2 shows the two primers designed for determining mutations in EZH2, exon 15, by High Resolution Melting analysis. Amplicon size is 248 bp. These primers were used for HRM of sample cohort 1 and 2.

OLIGO	start	length	Tm	gc%	3' seq
F-Exon15 PRIMER	88	20	59.22	40.00	CATTTCCAATCAAACCCACA
R-Exon15 PRIMER	335	21	59.80	38.10	TCCATTTTCACCCTCCTTTTT

For HRM based on the genomic DNA from paraffin embedded tissues (sample cohort 4), new primers were designed to yield a smaller PCR-product. Three primers were designed using same approach as mentioned in the previous section. See Table 3, for the two primers designed.

Table 3 shows the two new primers designed for HRM assay with gDNA from paraffin embedded tissues. The F(111 bp) primer paired with the R(111 bp) primer yields a product of 111 bp. Their targets starts at bp 79 and has a length of 3 bp (GTA=Y641).

OLIGO	start	length	tm	gc%	3' seq
F(111 bp) PRIMER	59	20	58.88	55.00	CAGTGCCTTACCTCTCCACA
R(111 bp) PRIMER	169	23	62.65	47.83	CTAGCATCTATTGCTGGCACCAT

The first and second primer design and their location in regard to exon 15 can be seen in Figure 11.

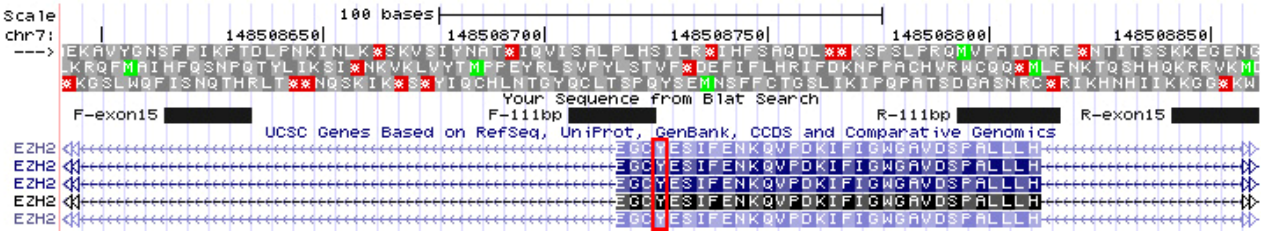

Figure 11 illustrates the location and target sequence of the primers mentioned in Mutation screening. Y641 is marked with a red box.

A third primer set was developed for screening of cDNA , sample cohort 3, with the same *Modus operandi* as with the two previous primer designs. See Table 4 for information on the two primers.

Table 4 shows the two primers designed for HRM with cDNA from sample cohort 3. The amplicon size is 168 bp.

OLIGO	start	length	Tm	gc%	3' seq
cDNA(ex14)F PRIMER	30	20	60.63	45.00	AGGCTGGGGGATTTTTATCA
cDNA(ex16)R PRIMER	197	20	60.62	45.00	GGTTGCATCCACCACAAAAT

High Resolution Melting Analysis

Genomic DNA of 125 samples from sample cohort 1 and 2 were analysed by High Resolution Melting to determine mutations in EZH2, Exon 15. For each High Resolution Melting analysis a Roche® Lightcycler 480 along with Lightcycler 480 High Resolution Melting Master kit was used. For each sample the following master mix was prepared; 5.0µL Lightcycler 480 High Resolution Melting Master x2, 0.2µM forward primer, 0.2µM reverse primer, 3mM MgCl, 1.6µL H_2O, 20ng DNA. The Master mix and template was loaded onto a white 96 well plate, which was sealed prior to analysis. Each sample

was analysed in duplicates. For each analysis, three cell lines were used as control, DB (EZH2(Y641) mut), OCI LY 3 (ABC WT), OCI LY 7 (GCB WT).
Cycling and melting conditions were the following: one cycle at 95°C for 10 minutes, 40 cycles of 95°C for 10 seconds, 58°C for 10 seconds, and 72°C for 10 seconds, a heteroduplex step of 95°C for 1 minute and 40°C for 1 minute, and a melt from 65°C to 90°C with 25 acquisitions per °C. All samples were run in duplicate, and melt curve analysis was performed by the LightCycler 480 Software (Roche Diagnostics), version 1.5.0.39
Normalization areas were 77-78°C before melt and 83-84°C after melt. Amplification plots were used to analyze quality of the DNA samples, and only those with a Cycle Threshold (Ct) of <30 and a sigmoid curve were considered.
The samples are denoted as; H1-H60 (group 1), H281-297, H300-385 & H473-484 (group 2).

Agarose Gels (3.5%)

The specificity of all primer sets were checked by running the amplicons from the High Resolution Melting analysis on an agarose gel. For each gel, 60mL TBE buffer (1x), 2.1 g Agarose, 240 µL ethitiumbromide were used. AmpliSize™ Molecular Ruler, 50-2000 bp Ladder by BioRad was used. For each sample 2.5µL Nucleic acid sample loading buffer (5x) from BioRad and 2.5µL sample was mixed and loaded on to the gel. The gels ran at 100 V for 45 minutes. The gels were then photographed upon UV transillumination using a Chemi-doc from BioRad.

PCR Purification

For PCR Purification prior to sequencing, PCR Purification kit from Qiagen was used. Purification was performed after protocol supplied by Qiagen. Concentration of the purified product was measured with a Nanodrop spectrophotometer.
Prior to sequencing, the samples were diluted in milliQ-water to 2ng/µL.

Sequencing

Selected samples based on High Resolution Melting analysis were sequenced by Eurofins, MWG Operon.

High Resolution Melting Analysis – Paraffin Embedded Tissue Samples

Based on an optimization of the two revised primer sets, the F(start59) and R(150bp) primers were chosen for the High Resolution Melting assays on FFPE tissues. 17 samples with gDNA from the paraffin embedded tissue samples were analysed by High Resolution Melting, following the approach mentioned in High Resolution Melting analysis.

Due to the DNA quality of FPPE tissue samples, the previously set parameter for quality control of the HRM assay was evaluated to consider only those with a sigmoid curve as the samples loaded with stock concentration would have a Ct <30.

High Resolution Melting Analysis – cDNA

High Resolution Melting analysis on cDNA synthesized from mRNA (sample cohort 3) was performed with the cDNA(ex14)-F and cDNA(ex16)-R primers, using the same approach as mentioned in High Resolution Melting analysis.

Statistical Analysis

Geneexpression Analysis Using Partek

Gene expression arrays were analysed using the software PARTEK - genomic suite from Partek incorporated. Prior to analysis samples were selected based on a quality control pipeline created by Philip de Groot [De Groot, 2006]. After all data files for GCB classified samples were imported and RMA normalized, histograms were used to evaluate each sample. If a sample had a distinctive different histogram compared to the majority of samples, it was excluded from the analysis. After determining which samples should be excluded, the remaining was imported and RMA normalized for the final analysis. A table of the samples included in the gene expression analysis, can be seen in Table 5.

Table 5 shows the samples used for the gene expression analysis in PARTEK – genomic suite. All samples originate from patients with DLBCL and are classified as the GCB-subtype. Excluded samples were H12, H14, H16, H21, H31, H39, H40, H47.

Sample cohort 1 (n=18)		**Sample cohort 2 (n=11)**		**Sample cohort 3 (n=8)**
H4	H43	H284	H385	H437
H6	H46	H285		H439
H7	H49	H295		H443
H9	H50	H296		H444
H11	H51	H302		H446
H15	H53	H318		H448
H19	H54	H330		H454
H37	H57	H344		H457
H42	H60	H380		

Due to batch effect from the change in labeling kits, cf. Array analysis, the samples was divided into two groups, based on labeling and analyzed individually for differentially expressed genes and then compared for overlap of genes, using a Venn diagram.

For each analysis, samples were divided into a category of GCB/GCB(Y641) and wt/mut. Sample cohort 1 analysis consisted of 18 samples, sample cohort 2 of 11 and lastly group 3 of 10 samples.

For each detection of differentially expressed genes, following criteria were set, an unadjusted p-value og 0.05 and a foldchange of +/-2.

Clinical Outcome – Kaplan and Meyer

Clinical outcome analysis was performed using Kaplan-Meier survival plots, generated in the statistical program R. The R-script can be seen in Appendix C.

The clinical outcome analysis was based on information recorded in the LYFO-database. The survival time in weeks for each patient was calculated from the date of diagnosis to the date of death, within the first five years from the diagnosis date. Surviving patients were censored.

The clinical outcome analysis was limited to the first five years after diagnosis.

Three different analysis were made, GCB(EZH2(Y641)) vs. GCB(wt), in regard to treatment; R-CHOP(EZH2(Y641)) vs R-CHOP and lastly EZH2(Y641) vs. All EZH2(wt) patients with DLBCL.

RESULTS

THEORETICAL OCCURRENCE OF Y641 MUTATION

Prior to determination of the presence of the EZH2(Y641) mutation in Danish lymphoma patients, a theoretical assessment of the possible mutations and their occurrence were made. The possible codons, resulting from a single nucleotide substitution in the Y641 site, can be seen in Table 6.

Table 6 shows the possible codons substitutions, due to a single point mutation in the coding sequence for Y641, TAC. The TAC sequence is wt and boxed in.

Position / Nucleotide	1	2	3
T	**TAC**	TTC	TAT
A	AAC	**TAC**	TAA
C	CAC	TCC	**TAC**
G	GAC	TGC	TAG

The various codons from Table 6 and their corresponding amino acid occurrence percentage are displayed in Table 7 along with the percentage of mutations identified by Morin *et al.*

Table 7 displays the theoretical occurrence of amino acids, based on a single point mutation in the Y641 codon, TAC and the percentage of the mutations found by Morin *et al.* The percentage calculated by Morin *et al.* is based on the specific amino acid in all EZH2(Y641).

Theoretical occurrence		**Identified by**
		Morin *et al.*
Tyrosine (TAT)	11,11%	0%
Aparagine (AAC)	11,11%	15%
Histidine (CAC)	11,11%	13%
Aspartic acid (GAC)	11,11%	0%
Phenylalanine (TTC)	11,11%	49%
Serine (TCC)	11,11%	21%
Stop codons (TAA, TAG)	22,22%	0%
Cysteine (TGC)	11,11%	1%

HIGH RESOLUTION MELTING

Prior to HRM for EZH2(Y641) mutation analysis in Danish lymphoma patients, optimization was conducted in regard to annealing temperature and overall reaction mix volume for HRM. The annealing temperature was tested at 55, 57, 58 and 59°C and the HRM reaction mix was tested at 10 and 20 µL, respectively (data not shown). Optimization was conducted with three cell linies; DB (EZH2(Y641)), OCI LY 7 (GCB EZH2(wt)), OCI LY 3 (ABC EZH2(wt)), which were also used as controls in the EZH2(Y641) HRM analysis. The primer set used was the Exon 15 primer set.

The quality controls parameters set for the HRM analysis was a sigmoid curve and a Ct value around 30. The parameter of a sigmoid curve indicates the amplification of nucleic acids within the range of exponential growth while the parameter of a Ct around 30, is used as an indicator for the quality and quantity of the DNA used in the analysis. If the amplification occurred at later Ct values, the distance to the negative control, was taken into consideration.

The results of the optimization for 10 and 20µL HRM reaction mix were equal in clearing the quality control parameters, so the 10µL HRM reaction mix was subsequently used for the HRM experiments. Optimization results for annealing temperature at 58 °C and HRM reaction mix volume of 10µL can be seen in Figure 12. These parameters generated a sigmoid like curve and the amplification occurred before the 30 Ct while the negative control (H20), did not amplify. The specificity of the primer set were checked by running the amplicons from the HRM analysis on an agarose gel, see Figure 13. The size of the amplicon was 248bp, which was verified by the agarose gel. Consequently no other amplicon was detected in the gel, demonstrating the specificity of the primer set. Based on these results, the 58°C annealing temperature and the 10µL reaction mix volume was subsequently used for the HRM experiments. For the following HRM experiments, further optimization was disregarded if the samples amplified in a respectable distance from the negative control and produced single amplicon band, detected by agarose gels, cf. the quality control parameters.

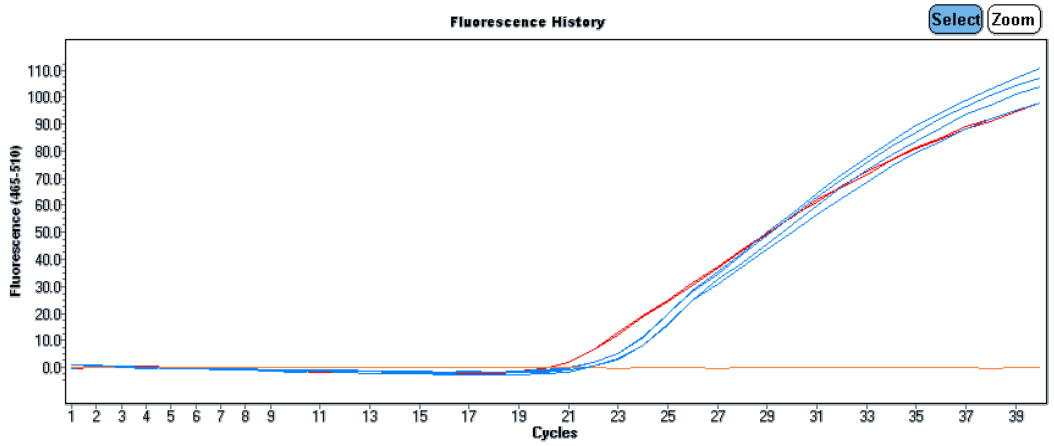

Figure 12 shows the amplification curve, based on the fluorescence signal, with an annealing temperature at 58°C and a mix volume of 10µL. the parameters generated a sigmoid curve and was amplified prior to the 30 Ct. The red amplification curves indicate the EZH2(Y641) cell line DB, the two blue amplification curves signify the two EZH2(wt) cell linies, OCI-Ly 7 and OCI Ly 3. The orange is the negative control (H2O).

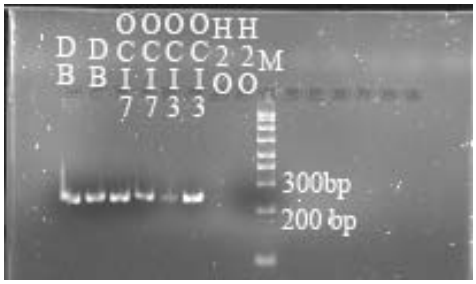

Figure 13 shows an agarose gel (3.5%) with the amplicons from the HRM analysis with an annealing temperature of 58°C and 10µL reaction mix. The agarose gel was used to check primer specificity of primer set Exon15. The amplicon size is 248bp. No other amplicons are observed.

HIGH RESOLUTION MELTING – SAMPLE COHORT 1 AND 2 (GENOMIC DNA)

For sample cohort 1 and 2, a total of 125 samples was analysed by HRM, using the parameters determined through optimization. The mutational status is examined based on a difference in amplicon melting curves as heteroduplexes dissociate more easily than homoduplexes, due to a decrease in stability. To further simplify the detection, positive (EZH2(Y641) and negative controls (EZH2(wt)) were also used.

The shifted and normalized melting curve for the analysis of H1-H8 (sample cohort 1), can be seen in Figure 14.A, here H7 demonstrates a difference in melting curve, resembling the melting curve of the cell line DB(positive control), displayed in red, while the wt type amplification curves (homoduplexes) are blue. The relative signal difference (difference plot) is seen in Figure 14.B. The difference plot, is derived from the normalized and shifted melting curves (Figure 14.A), which clusters samples with similar melting curves into groups.

The data indicate that H7 has a nucleotide change within the amplified region. All samples, with the same indications for a different melting curve, as H7, were selected for sequencing. The remaining HRM results for sample cohort 1 and 2 can be seen in Appendix D.

Samples in the intermediate area, between the melting curves of the EZH2(wt) cell lines and the EZH2(Y641), were also selected for sequencing along with OCI-Ly7 (EZH2wt).

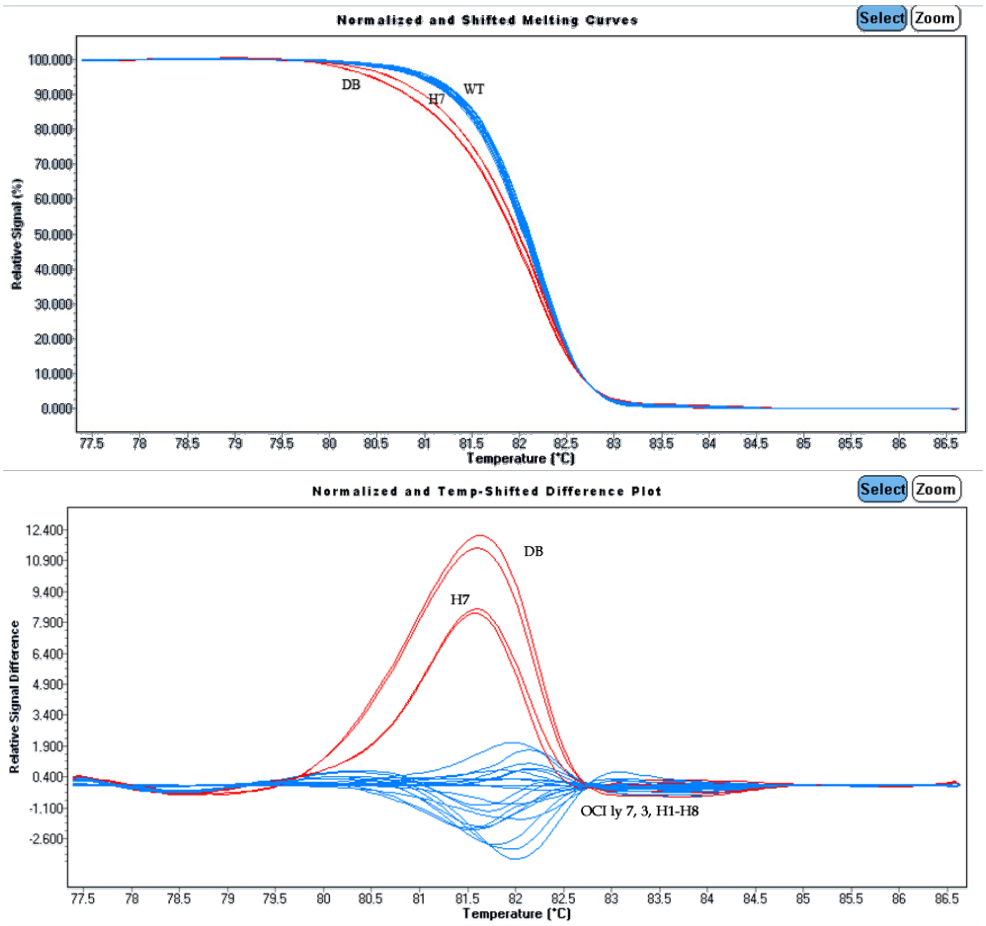

Figure 14.A(top picture) illustrates the shifted and normalized melting curve of sample H1-H8 (sample cohort 1). Here H7 is observed to have melting curve similar to DB, positive control, indicating that H7 has a nucleotide change within the amplified region. Sequencing revealed that this sample was mutated in the Y641 codon. **14.B**(bottom picture) show the shifted and normalized difference plot of sample H1-H8 (sample cohort 1).

In sample cohort 1 and 2, six samples had different melting curves, which were similar to the heteroduplex melting curve of DB. The samples were H7, H9, H15, H37, H60 and H302.

Sample Cohort 4 (Genomic DNA From Paraffin Samples)

Sample cohort 4, consisting of 17 samples, were analysed using the same approach as for sample cohort 1 and 2, however none of the samples cleared the quality control parameters, as they all amplified after the 30 Ct, see Figure 15. The specificity of the primer set (Exon15) were checked by running the amplicons from the HRM analysis on an agarose gel, see Figure 16. The gel only shows clear 248bp amplicons for the three cell lines used as control, the samples from sample cohort 4 shows

weak (if any) amplicons of variable size, indicating that the DNA from the samples are to degraded from the formalin-fixed paraffin embedding, to be targeted by the primer set(Exon15).

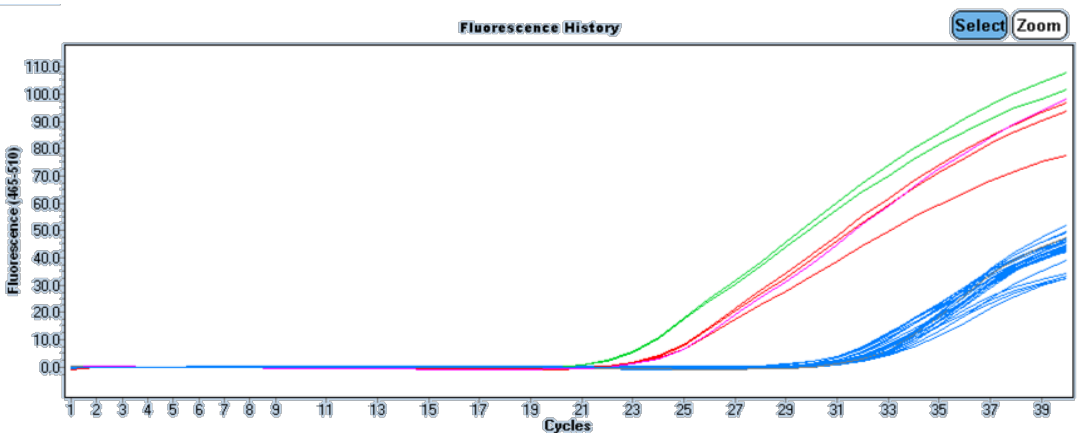

Figure 15 shows the amplification curve, based on the fluorescence signal, of sample cohort 4. The only samples that cleared the quality control parameter around 30 Ct was the controls (red & green). The samples from sample cohort 4 amplified, along with the negative control, H2O.

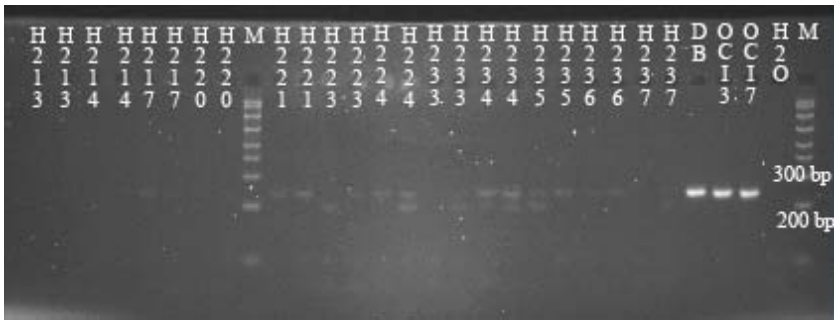

Figure 16 is the agarose gel used to check the specificity of the exon 15 primer set for sample cohort 4. The only clear amplicons at 248bp are the cell lines used as controls. The amplicons for the samples are weak, if present and of variable size. This indicates that the DNA is degraded from the formalin-fixed paraffin embedding.

Based on these results, a new primer set was designed and tested. The amplicon size was minimized to 111 bp, moving the primers closer to the codon encoding Y641. With the 111 bp primer set, the assay only targets part of exon 15. Figure 17 shows the amplification curve, based on the fluorescence signal, of sample cohort 4. The samples that start the amplification at 20-23 Ct are the cell line controls, the samples that amplify at 25-29 Ct are the twelve samples with a template concentration of 10 ng/µL, while the samples amplifying at 35-39 was the remaining samples with stock concentration(1,06-5,09 ng/µL). The samples amplifying at 35-39 Ct, was excluded from analysis, (H230, H222, H232, H238) despite that amplicons of 111 bp was detected by an agarose gel, see Figure 18. The exclusion was based on the observation, that they amplified along with the negative control, H2O.

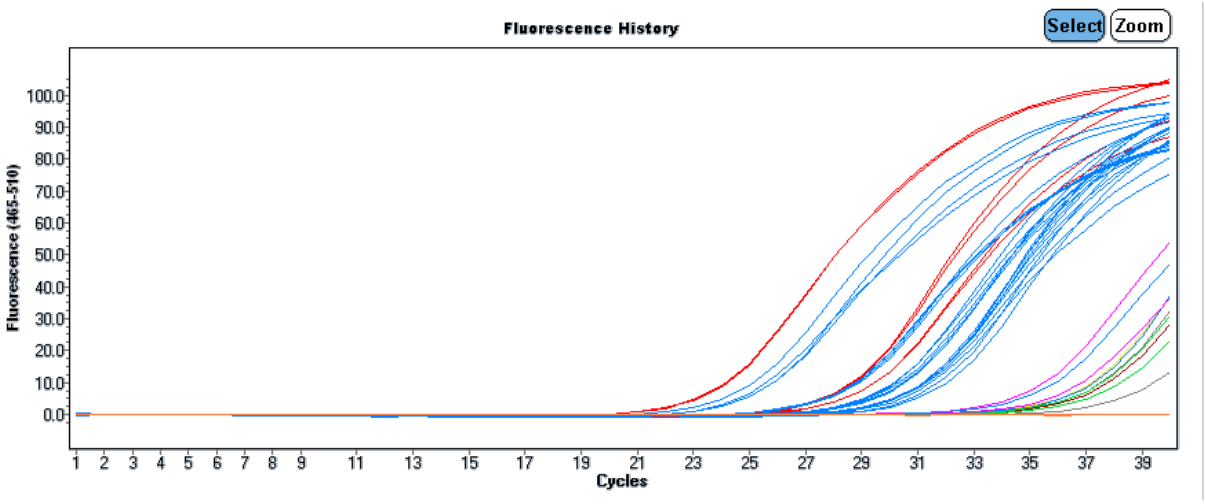

Figure 17 shows the shows the amplification curve, based on the fluorescence signal, of sample cohort 4. The samples that amplify at 20-23 Ct are the cell line controls, the samples at 25-29 Ct are H213, H214, H217, H220, H221, H223, H224, H233, H234, H235, H236 and H237 (20ng), while the samples amplifying at 35-39 Ct was the remaining samples with stock concentration(<20ng).

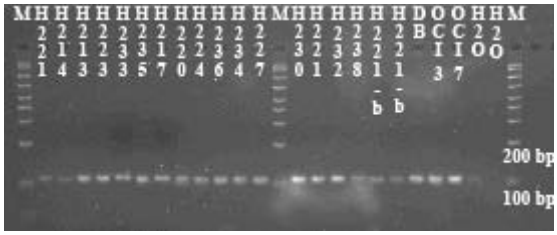

Figure 18 displays the agarose gel used to check the specificity for the 111bp primerset. Only the expected amplicon is observed, verifying the specificity of the primerset.

The difference plot can be seen in Figure 19. In sample cohort 4, H213 and H214 displayed heteroduplex melting curves similar to DB and was selected for sequencing.

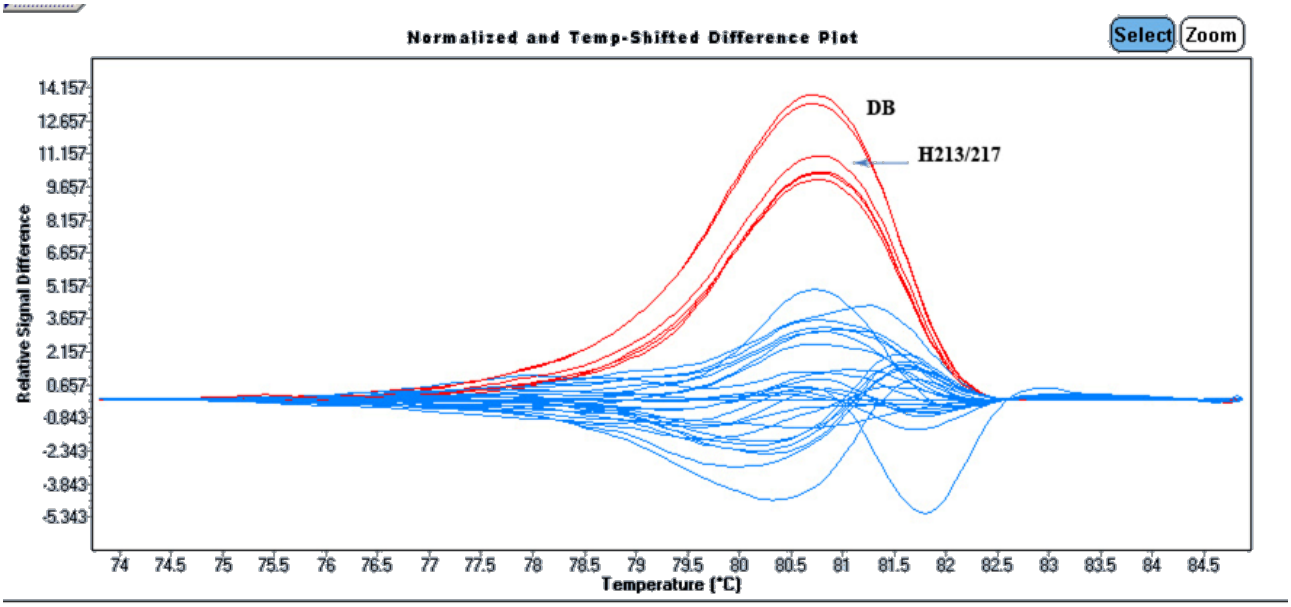

Figure 19 illustrates the shifted and normalized difference plot of sample cohort 4. H213 and H217 were selected for sequencing based on their melting curve.

Sample Cohort 3 (cDNA)

Because the previous two assays are based on primers that either partly or entirely anneal to intronic regions, a new primer set was designed to analyse cDNA samples. The primer set for this third assay was designed to target exon 14-16. The result of HRM with the cDNA primerset, can be seen in Figure 20. H457 displayed similar melting curve to DB. H455 displayed a melting curve different from both H457 and DB. Both samples were selected for sequencing, along with H437 and H439.

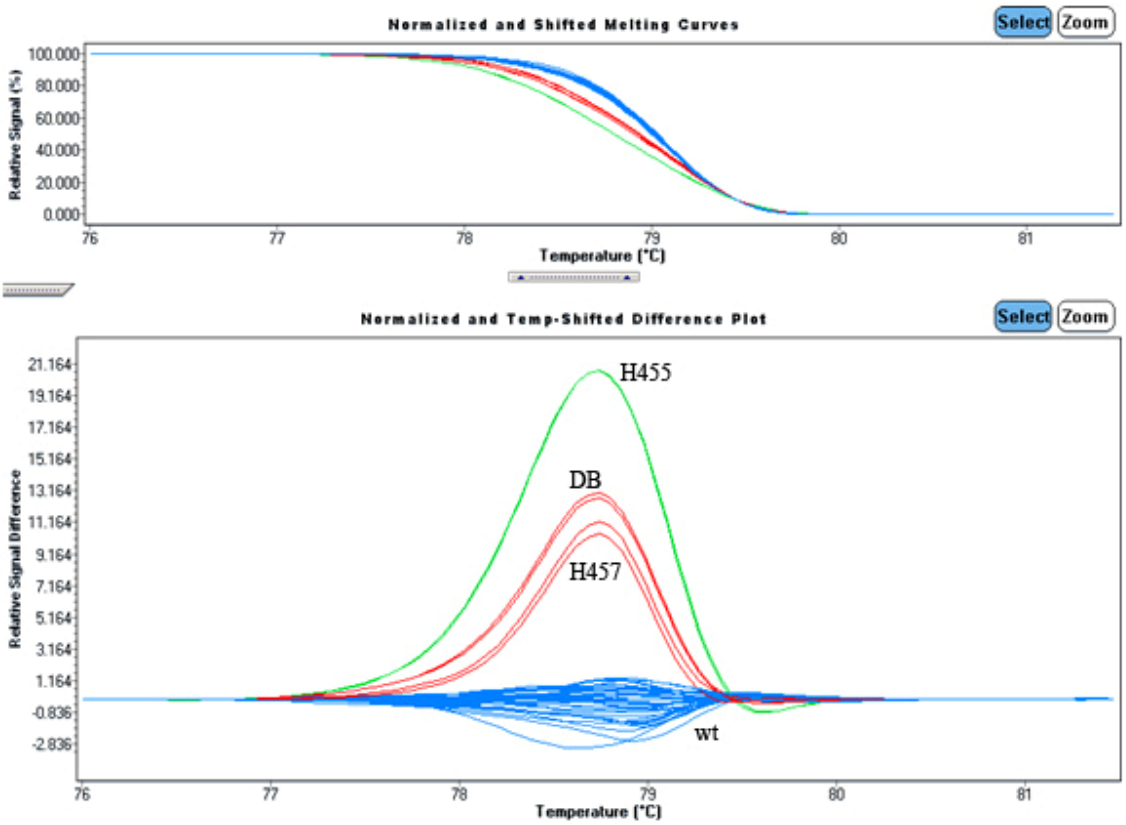

Figure 20 displays the result of the HRM with the cDNA primerset. . H457 display similar melting curve as DB. H455 displays a melting curve different from both H457 and DB.

The specificity of the cDNA primer set was verified using an agarose gel to visualize the PCR amplicons, see Figure 21.

Figure 21 illustrates the verification of the cDNA primerset, with an amplicon at 168 bp.

SEQUENCING

In total 20 samples were selected for sequencing based on their aberrant or intermediate melting curve, see Table 8. In addition the two control cell lines were also sequenced. Ten of the 20 clinical samples were found mutated; these were H7, H9, H15, H37, H60, H302, H213, H217, H455 and H457. The sequence for each of the mutated samples along with the sequence for DB and OCI-Ly 7 can be seen in Figure 22. Four different types of EZH2(Y641) mutations were identified, these were Asparagine, Serine, Phenylalanine and Cysteine.

Table 8 sequenced samples, their diagnosis, sample cohort, sequence for Y641 codon and the correlating Amino Acid.

H-sample	Diagnosis	Sample cohort	Seq at codon 641	Amino Acid
H7	DLBCL	1	TAC/TCC	Serine
H9	DLBCL	1	TAC/AAC	Asparagine
H14	DLBCL	1	TAC/TAC	wt
H15	DLBCL	1	TAC/AAC	Asparagine
H25	DLBCL	1	TAC/TAC	wt
H26	DLBCL	1	TAC/TAC	wt
H37	DLBCL	1	TAC/TCC	Serine
H40	DLBCL	1	TAC/TAC	wt
H60	DLBCL	1	TAC/AAC	Asparagine
H302	DLBCL	2	TAC/AAC	Asparagine
H345	DLBCL	2	TAC/TAC	wt
H347	Unspecified reactive lymphnode	2	TAC/TAC	wt
H361	B-CLL	2	TAC/TAC	wt
H385	DLBCL	2	TAC/TAC	wt
H213	DLBCL	4	TAC/AAC	Asparagine
H217	DLBCL	4	TAC/TTC	Phenylalanine
H437	DLBCL	3	TAC/TAC	wt
H439	DLBCL	3	TAC/TAC	wt
H455	Follicular lymphoma	3	TAC/TGC	Cysteine
H457	DLBCL	3	TAC/AAC	Asparagine

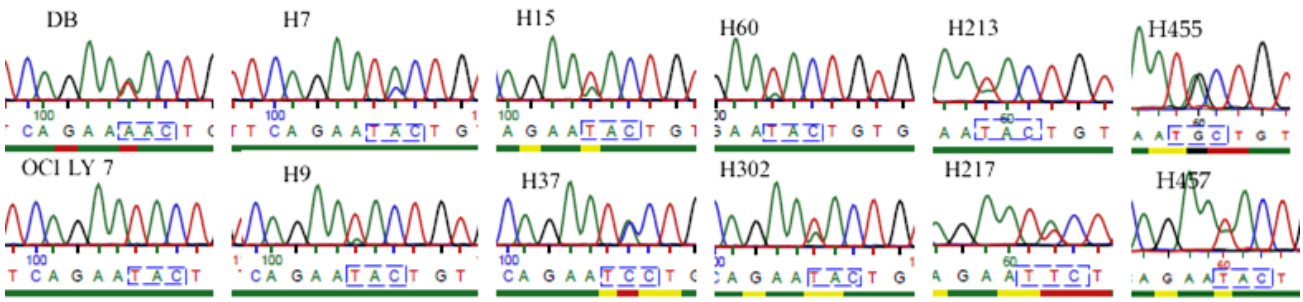

Figure 22 displays the sequences for the Y641 codon, in the mutated samples, DB and the EZH2 (wt) cell line OCI LY 7.

In summary of the results from HRM analysis and sequencing, the frequency of EZH2(Y641) in Danish DLBCL patients is listed in Table 9.

Table 9 Quantity of samples tested by HRM and their subtype along with the frequency of EZH2(Y641) for each sample cohort. The GCB, ABC and NC columns are DLBCL subtypes; GCB- germinal center B-cell like, ABC, Activated B-cell like and NC- not classified.

Samples:	Samples in total	DLBCL	GCB	ABC	NC	Y641	Y641/GCB
H1-H60,	57	57	26	24	6	5	19,2 %
H281-297, H300-385, H473-484	68	20	11	8	1	1	9,1 %
H213-H246	13	13	8	5	0	2	25,0 %
H437-H457	20	18	8	9	1	2	12,5 %

The theoretical expectancy of Y641 mutations and the experimentally frequency found by Morin *et al.* and in this thesis can be seen in Table 10.

Table 10 illustrates the theoretical expectancy of the possible single point mutations of Y641, the frequency of which they were discovered by Morin and in this Thesis.

Expected		Identified	
		Morin	This thesis
Tyrosine (TAT)	11,11%	0%	0%
Aparagine (AAC)	11,11%	15%	67%
Histidine (CAC)	11,11%	13%	0%
Aspartic acid (GAC)	11,11%	0%	0%
Phenylalanine (TTC)	11,11%	49%	11%
Serine (TCC)	11,11%	21%	22%
Stop codons (TAA, TAG)	22,22%	0%	0%
Cysteine (TGC)	11,11%	1%	11%

No other mutations were observed in (codon 602-646), in Danish lymphoma patients, a region that has been reported to habour mutations in DLBCL [Morin *et al.* 2011]

The complete sequences can be found in appendix F on the Cd-rom.

AFFYMETRIX® CYTOGENETICS ASSAY RESULTS

The global cytogenetic profile of EZH2(Y641) samples from sample cohort 1 and 2 were assessed using the affymetrix 2.7M cyto arrays and the software program ChaS. The analysed sampes were H7, H9, H15, H37 and H302. In regard to genomic alterations in a 200Kb region around the *EZH2* gene with a marker count of 50, H9, H15 and H302 showed no genomic abnormalities, while H7 and H60 displayed LOH. Lastly H37 had a gain, with a copy number of 3 (CN3).

The six EZH2(Y641) samples were analysed across all chromosomes, in order to investigate the presence of correlating genomic alterations. The six EZH2(Y641) samples was analysed for gain, loss, mosaicism and LOH (data not shown), see Figure 23. H7 and H37, both heterozygous for Tyr/Ser had increasing gain on chr1-q-arm. All six samples had LOH on chr14q23.3-24.1, however this was not exclusive to EZH2(Y641) as this also was seen in other GCB patients and in healthy DNA patient samples. As the samples were very heterogeneous regarding genomic alterations, the sample was too small to detect a possible correlation of genomic alterations was observed across the chromosomes in EZH2(Y641).

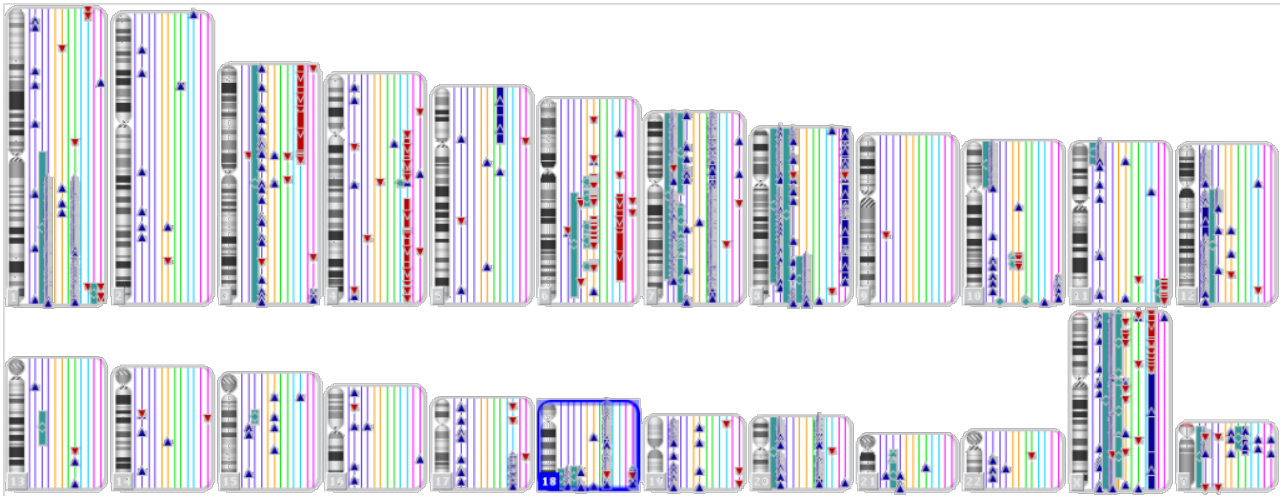

Figure 23 illustrates the genomics alterations found in H7, H9, H15, H37 and H302 in a karyoview. Red bloks indicate loss while blue indicate gain. Mosaicism is indicated with teal blocks. H7 and H37 (Y/S) both have increased gain on q-arm on chr1.

GENE EXPRESSION

Gene expression arrays were performed on all DLBCL samples in sample cohort 1, 2 and 3. An assessment of batch effect was conducted, as to two different labeling kits was used during the micro array labeling procedure. This was done in order to minimize false positives generated by batch effect. Figure 24 illustrates the grounds for the batch effect assessment.

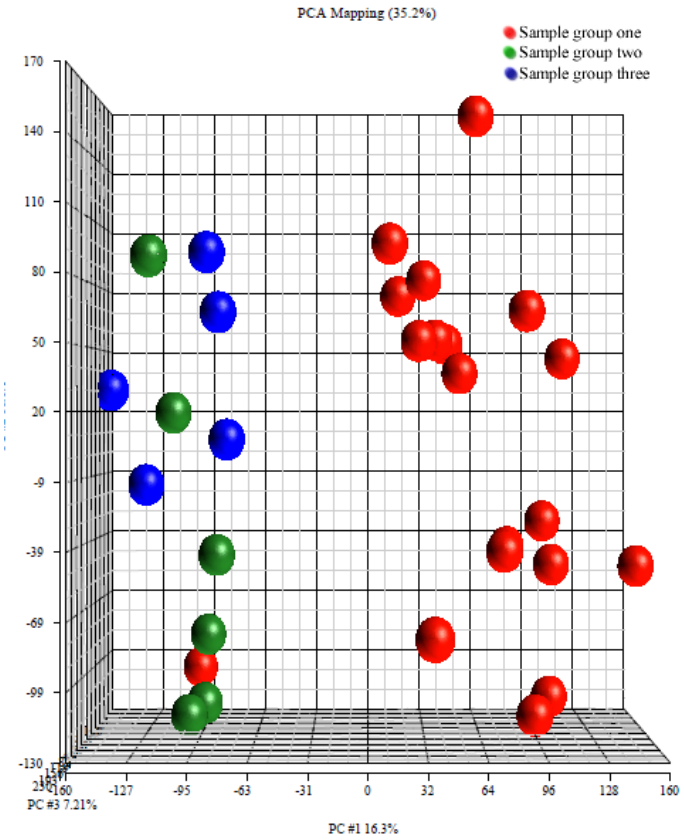

Figure 24 illustrates the batch effect, were sample cohort 1 (red) clusters individually. The single sample from group 1 that clusters with sample cohort 2&3 is labeled with the same kit, hence explaining the location in the PCA plot.

To evaluate the "batch effect removal function" available in the analytical program Partek, six samples labeled with each kit was used. Gene expression for a random gene was selected and plotted in a dot plot before and after the batch effect removal was conducted in Partek, see Figure 25.

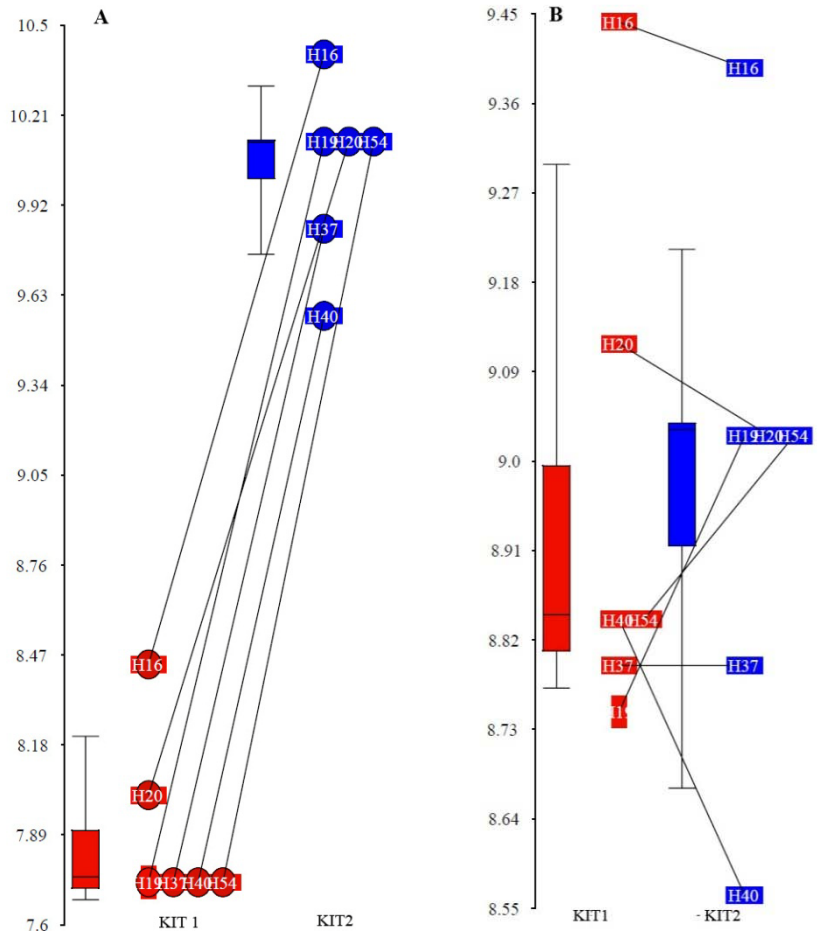

Figure 25 illustrates the batch effect on a random selected gene, prior (A) and after (B) applying the batch effect removal function.

The difference in batch effect between the two types of labeling is approximately a 4fold change prior to batch effect removal. After applying the batch effect removal function, it was expected that the duplicated samples would have been brought to the same level of expression, however this was not the case. The difference between duplicates was minimized but not extinguished, see Figure 25. Based on these results, the samples were grouped together due to their labeling and analysed for differentially expressed genes between EZH2(Y641) and EZH2(wt) DLBCL GCB-subtype samples. The two lists of differentially expressed genes were then compared for correlating genes using a venn diagram, see Figure 26. In sample cohort 1, there is 546 differentially expressed probesets with an unadjusted p-value 0.05 and a fold change of +/- 2. In sample cohort 2 and 3 there is 625 differentially expressed probes. Overlapping, between the two analyses, is 33 differentially expressed probes.

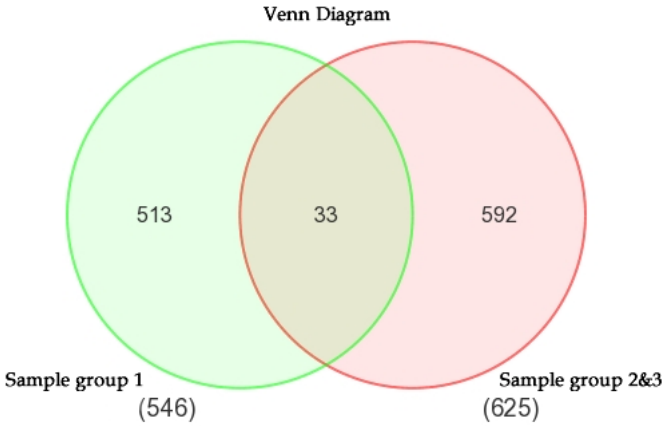

Figure 26 illustrates the venn diagram of differentially expressed genes between EZH2(Y641) and GCB(EZH2(wt)).

In Figure 27 a hierachial clustering of the 33 probeset for sample cohort 1 is illustrated. Amongst the 33 probesets in the list, there are 25 different genes, as some of the genes have more than one probeset on the array.

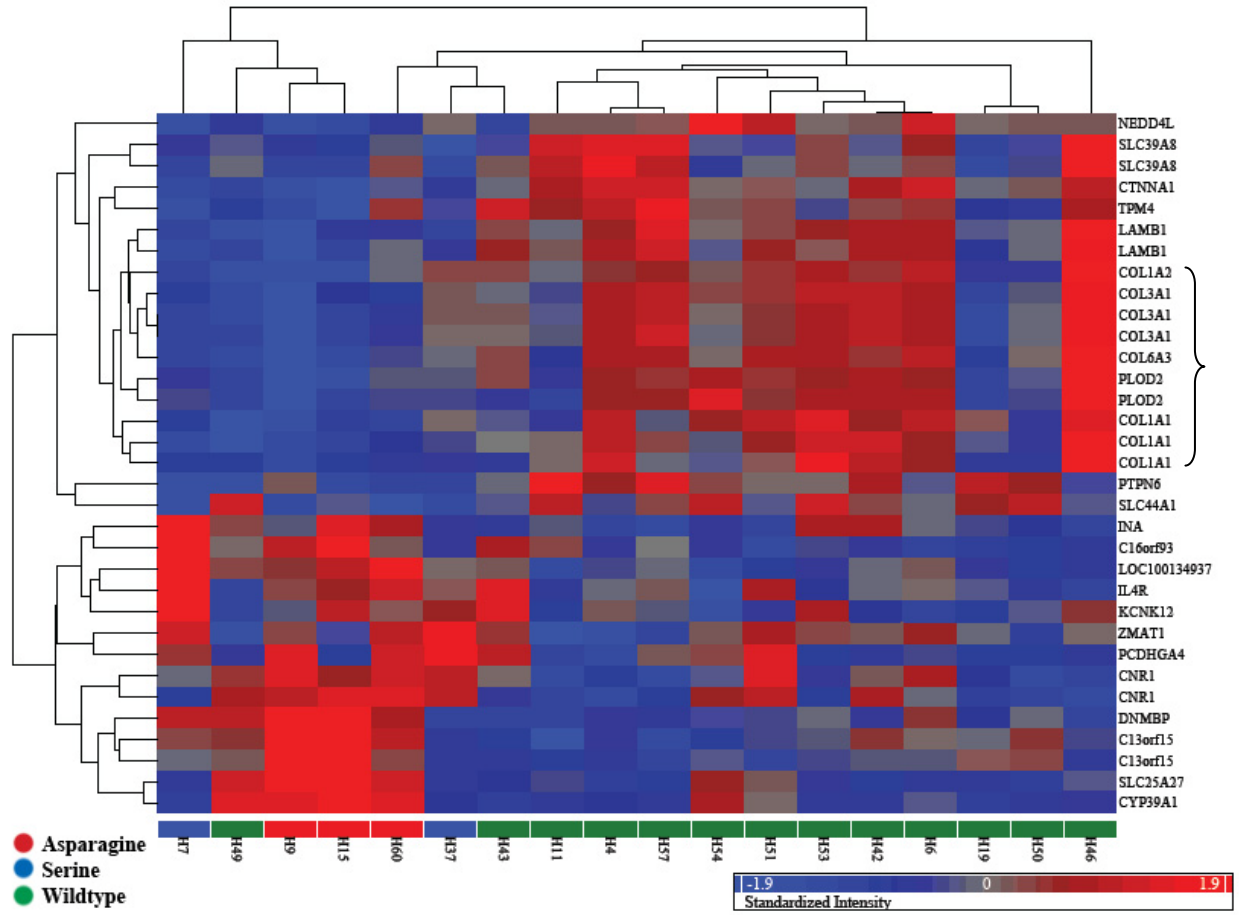

Figure 27 displays a standardized hierachial cluster of the 25 differentially expressed genes in sample cohort one. The mutated samples are indicated with blue and red boxes, respectively. The blue box indicates samples mutated to Asparagine while red indicates Serine. Green is wt. Of the 25 genes five are related to collagen. These genes are indicated by a bracket while CNR1 is marked by a star.

The hierachial clustering for sample cohort 2 and 3 is illustrated in Figure 28. Out of the 25 different genes, four are collagen - alpha receptor types and one is procollagen-lysine. The collagen related genes are also down regulated in the EZH2(wt) samples, H49 and H385.

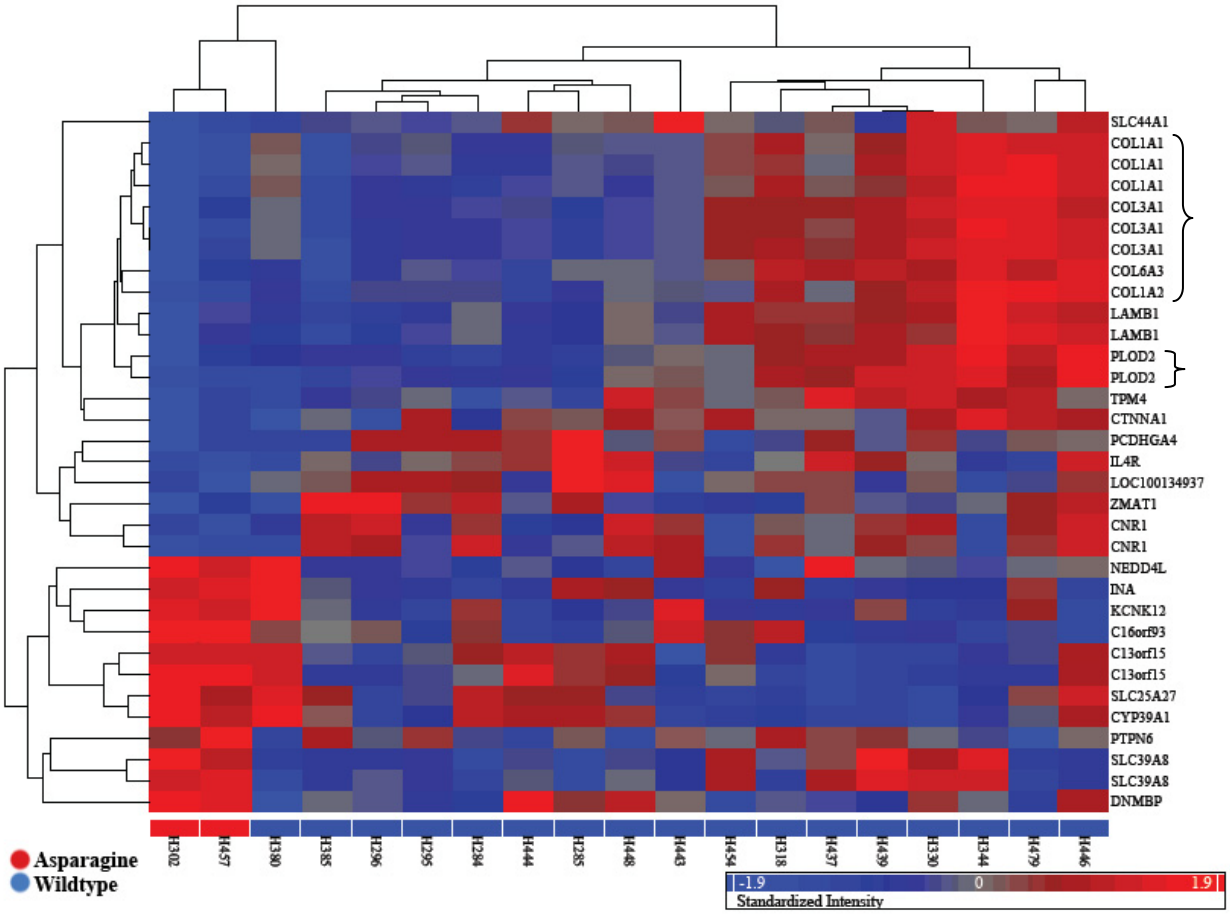

Figure 28 displays a standardized hierachial cluster of the 25 differentially expressed genes in sample cohort 2&3. The mutated samples are indicated with blue and red boxes. The red box indicates samples mutated to Asparagine while red indicates wt. The collagen related genes are indicated by a bracket. It should be noted that H302 and H385 are from the same patient, but from the primary and the relaps tumor, respectively.

In order to illustrate the difference in gene expression of the collagen related genes a dot plot of the expression of Collagen type 1- alpha 1(*COL1A1*), was constructed for each of the two analyses, see Figure 29.

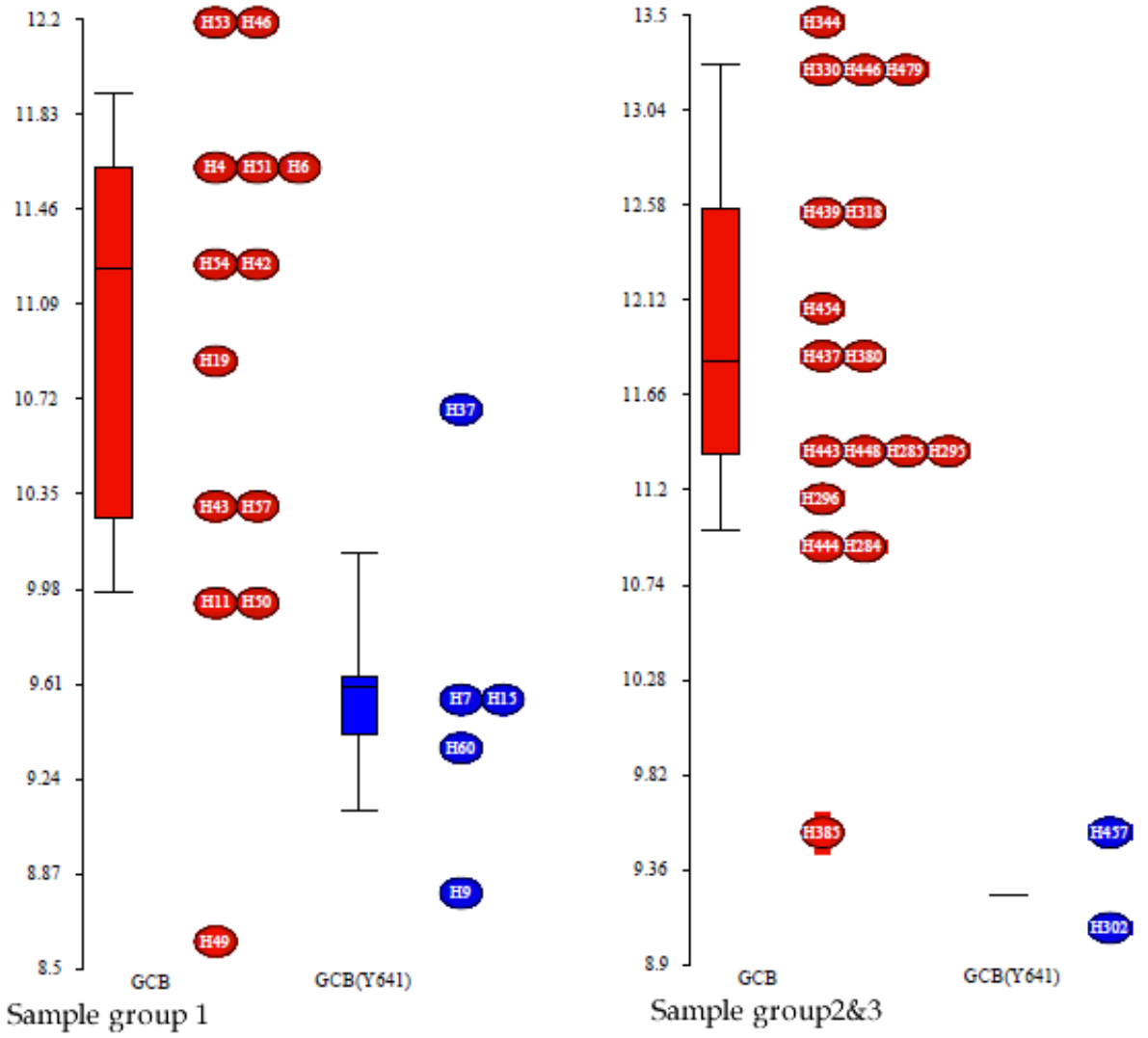

Figure 29 illustrates the two dot plots for COL1A1, from the differentially expressed gene list. The scale of the dot plot is a Log2 scale.

The *COL1A1* gene is down regulated in the EZH2(Y641) samples when comparing to GCB(EZH2(wt)) The difference is approximately 4fold in both plots.

CLINICAL OUTCOME

Three different approaches to assess the effect on clinical outcome for EZH2(Y641) was made. Each analysis was limited to the first five years after diagnosis.

The first analysis was based on DLBCL patients harboring the EZH2(Y641) mutation compared to patients with DLBCL of the GCB subtype (EZH2(wt)) samples, see Figure 30.

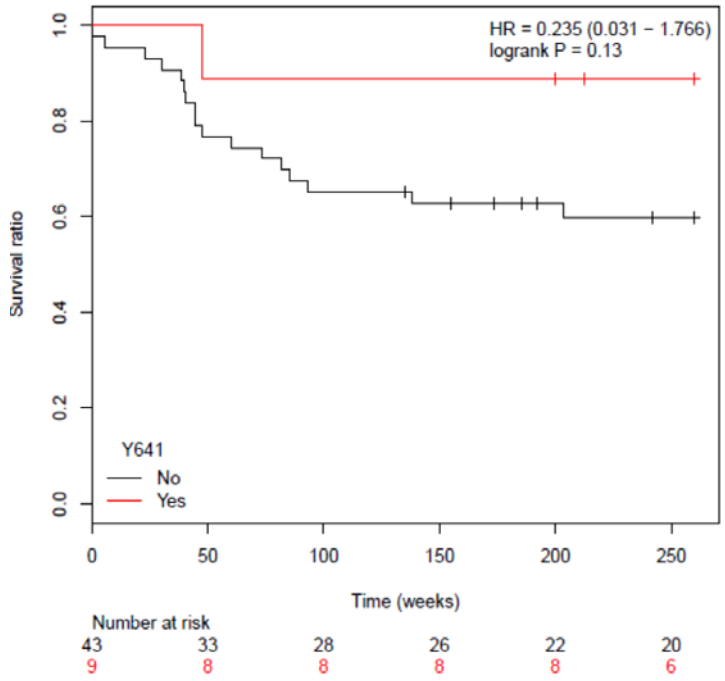

Figure 30 illustrates the clinical outcome for patients with a tumour of the GCB type with and without a EZH2 Y641 mutation within the first five years of diagnosis. The red graph represents patients harboring the EZH2(Y641), while black represents EZH2(wt). All patients were classified as the GCB-subtype of DLBCL.

The results for the first assessment, did not render significant in terms of clinical outcome between DLBCL patients with EZH2(Y641), as the P-value was 0.13.

The second clinical outcome analysis was based on EZH2(Y641) patients treated with R-CHOP compared to all DLBCL patients treated with R-CHOP(ABC, GCB and non classified subtypes), see Figure 31.

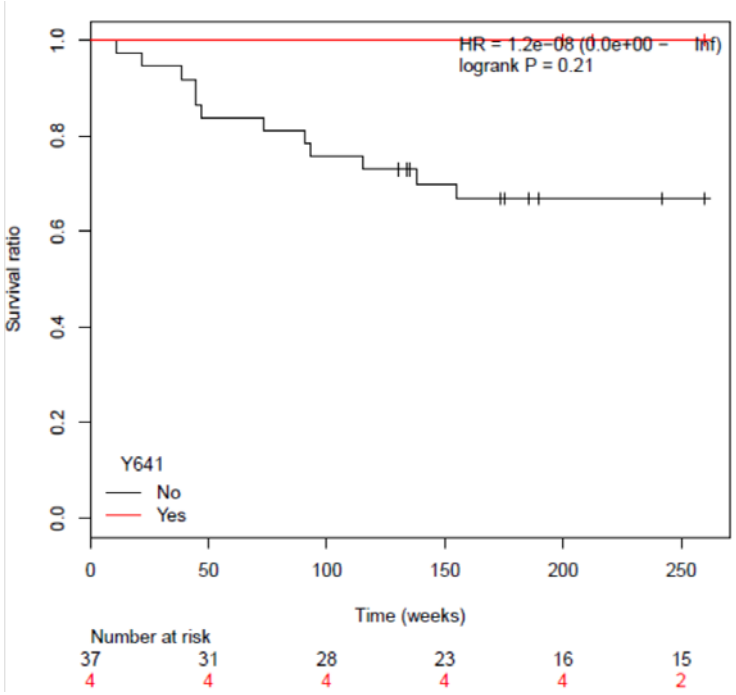

Figure 31 shows the clinical outcome for patients with EZH2(Y641) treated with CHOP-R when compared to all DLBCL patients treated with R-CHOP within the first five years from diagnosis The red graph represents patients harboring the EZH2(Y641) All patients with EZH2(Y641) were classified as the GCB-subtype of DLBCL. The black graph represents EZH2(wt) of all classifications of DLBCL (ABC, GCB and Non Classified).

In regard to the assessment of difference in clinical outcome between EZH2(Y641) patients and all other DLBCL patients treated with R-CHOP, the P-value of 0.21 signifies no distinction between the two groups.

The last clinical outcome analysis was in regard to patients with EZH2(Y641) compared to all DLBCL patients(ABC, GCB, Non Classified), within the first five years from diagnosis, see Figure 32. This analysis showed that there is a tendency towards an improved clinical outcome for patients with EZH2(Y641) when compared to all subtypes of DLBCL (p= 0.054).

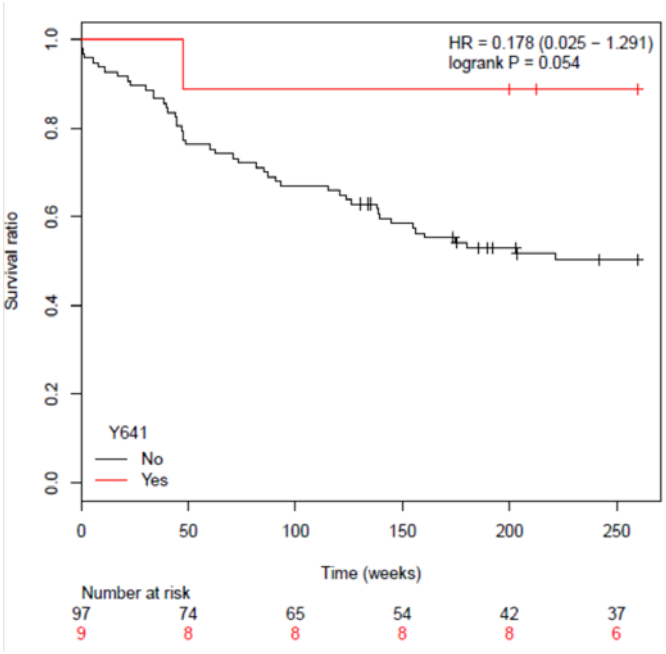

Figure 32 display the clinical outcome for patients with EZH2(Y641) compared to all patients diagnosed with DLBCL, within the first five years from diagnosis. The red graph represents patients harboring the EZH2(Y641) All patients with EZH2(Y641) were classified as the GCB-subtype of DLBCL. The black graph represents EZH2(wt) of all classifications of DLBCL (ABC, GCB and Non Classified).

DISCUSSION

The discussion is divided into sections, with respect to the statements made for this thesis.

DETECTION OF EZH2(Y641) IN DANISH LYMPHOMA PATIENTS

The detection of the EZH2(Y641) mutation through high resolution melting analysis, was considered successful as all the samples displaying a heteroduplex melting curve, similar to that of the EZH2(Y641) mutated cell line DB, had their mutational status verified through sequencing.
This was achieved using three different templates, genomic DNA from snap-frozen OCT tissue, genomic DNA from paraffin embedded tissue and cDNA, synthesized from RNA, proving the method applicable for mutation screening across various templates.
The overall EZH2(Y641) frequency in the GCB-subtype, in this thesis was found with an average frequency across the four sample groups at 18,9%, which closely matched the frequency found by Morin *et al.*(21,7%). Another study conducted by Park *et al.* on lymphoma patients of Korean heritage, found the EZH2(Y641) in 1.4% of DLBCL patients. This could be influenced by genetic variation across ethnical groups, with lower mutation frequency among DLBCL patients of asian origin.
Another possibility is that the assay developed by Park et al has low sensitivity and that some mutations have been missed, but screening of a well defined set of controls is needed to determine this.
When comparing the theoretical expected occurrence of EZH2(Y641) in this thesis and and the frequency found experimentally by Morin *et al.*, there is little correlation as the most prominent difference is a 37 % higher occurrence for phenylalanine than expected. For the results of this thesis the most significant difference in amino acid occurrence compared to the theoretical, is 66 % for asparagines compared to the expected 11 %. Phenylalanine and cysteine, have the same occurrence rate as theoretically expected The frequency of Serine observed by Morin *et al*. and observed in this thesis are very similar. Aspartic acid and stop codons are not at all detected, in either experimental setup, which on a protein level would result in a protein truncation or the insertion of a negatively charged aspartic acid. None the less, all possible codons should be detected on DNA level in case of no selection and as this is not the case, it strongly suggest positive selection toward codons translating into asparagine, histidine, phenylalanine, serine and cysteine. A positive selection of codons that yield a gain of function on a protein level, substantiates EZH2(Y641) as a candidate cancer driver gene.
This correlates with the findings of Morin *et al.* (2011), where a statistical analysis, also suggested a positive selection for the somatic point mutation. The functional role of EZH2 in normal b-cell differentiation, suggested to include favoritism of celluar proliferation by repressing tumor suppressor

genes [Velichutina *et al.*, 2010], would further substantiate the idea of EZH2(Y641) as a cancer driver candidate.

The gain of function on protein level resulting from the mutated codons, is an increased ability to facilitate di-trimethylation [Sneeringer *et al.,*2010; Wigle *et al.,*2011], something that previously have been identified in the SET-domain of lysine methyltransferases to be facilitated through a Phe/Tyr switch. Hence it could be speculated that the EZH2(Y641) facilitate the principle of Phe/Tyr switch in order to become more progressive in tumorgenesis. If this is the case, it would be the first Phe/Tyr switch linked to malignancy.

Correlation of Genomic Alterations

For genomic profiling by cyto2.7M array, different aspects were analysed for the six EZH2(Y641) samples and compared with the remaining non mutated GCB samples. All GCB samples were subjected to analyses for genomic alterations in the genomic region harbouring the *EZH2* gene. The result revealed that H37 had a CNV, with a gain (CN3) and H7 had mosaicism with a CN value of 2.4, both samples also had LOH for *EZH2*. The remaining samples showed no genomic alterations for *EZH2*. As the Y641 mutation is a single point mutation with a gain of function with the function af both alleles needed for tri methylation, it is not expected to have any effect on *EZH2* trough genomic alterations. The findings of few genomic alterations of *EZH2* among the mutated samples, is there for consistent with the current understanding of the EZH2 mutation function.

A complete overall assessment of the genomic profile among the mutated EZH2 samples, showed that H7 and H37 (both EZH2(Y641/S641)) had gain across the q-arm of chromosome 1, however this was not significantly different from the remaining unmutated GCB samples, due to the relatively low number of samples available for analysis. To determine if the gain is significant, more samples are need. No other correlation with CNV and mutation status was detected across the mutated EZH2 samples. With an suggested average of >100 genomic alterations pr DLBCL case [Pasqualucci *et al.* 2011] combined with the knowledge that different pathways in a cancer cell is used to obtain the same goals, [Hanahan&Weinberg, 2011] the difficulties in identifying overall correlating genomic alterations significant for six EZH2(Y641) samples, is exemplified.

Gene Profile Based on Gene Expression

The micro array results for gene expression identified 25 differentially expressed genes between patients with EZH2(Y641) and DLBCL GCB-subtype patients. Out of the 25 genes, focus was placed on collagen related genes. The collagen related genes were highlighted from the list, as they had a prominent downregulated gene expression for EZH2(Y641) and constituted of 20% of the differentially expressed genes. A litterature search of the collagen genes connection to cancer was

conducted. This revealed, one of the collagen genes *COL3A1* to be of particular interest, as it appears in several different gene expression profiling papers. Blenk *et al.* determined that *COL3A1* was expressed higher in the GCB-subtype, compared to the ABC-subtype and subsequently listed among genes, used to differentiate between the ABC and GCB-subtype by gene expression profiling. The gene expression results generated in this thesis, correlate with a high expression of *COL3A1* in GCB, but in comparison the EZH2(Y641) GCB samples shows down regulation of *COL3A1*. In the gene expression analysis, the ABC tumours were not included, to avoid pulling out genes differentially expressed genes between ABC and GCB subtype, as the mutation is only found among GCB. However it could be of interest to include ABC in the analysis, in order to determine if the expression level of COL3A1in EZH2(Y641) was lower than ABC or an intermediate.

Even though down regulation of *COL A*- genes is not a stranger to DLBCL, it should be noted that they are associated with the extracelluar matrix and there may not have any direct connection to the EZH2(Y641) lymphogenesis, but instead to the origin of the biopsy site. To validate that the identified genes are true target genes for downregulation by tri methylated EZH2 a validation should be carried out in a larger sample set.

When analysing the gene cluster sets, H49, H385 and H380 display a gene profile similar to the mutated EZH2 samples. Here it should be noted that H302 and H385 are a paired sample set, consisting of DNA from the primary tumor and the relaps tumor, which occurred six months after the patient was declared in complete remission. There are several different observations, which are interesting in regard to this paired sample set, the first and most significant for the scope of this thesis, is that H302 (primary tumour) harbours the EZH2(Y641) mutation, while H385 (relaps tumour) is EZH2(wt). This indicates that the EZH2(Y641) mutation is *de novo* and acquired through the development of tumorgenesis. This is consistent with the finding by Morin *et al.*, where paired samples set of normal and tumor DNA was examined, only to identify the EZH2(Y641) mutation in GCB-DLBCL.

In analysis of the gene cluster profile for H385, when compared to H302 and H457(EZH2(Y641)), it is noteworthy to mention that they share a similiar gene down regulation profile for the collagen related genes, this could indicate the different pathways applied in tumorgenesis to obtain a general framework, this however can only be considered valid, if the collagen genes can be linked to tumorgenesis. The assumption that the different pathways is applied to reach the same end goal, is also considered plausible for H49 and H380, explaining that their gene profile is similar to those with EZH2(Y641) mutation.

Clinical Outcome Analysis

Three different clinical outcome analyses were made, in order to determine if patients harbouring the EZH2(Y641) mutation had a difference in clinical outcome when compared to patients of the GCB subtype, all DBLCL patients treated with R-CHOP and lastly with all DLBCL patients, within the first five years following diagnosis. For the first two analyses, the P-value signified no difference in clinical outcome. The analysis comparing difference in clinical outcome in regard to treatment with R-CHOP, had a P-value of 0.21, which is in consistency with the P-value (0.27) determined by Morin *et al.*(2011) on a larger data set (n=199). This suggests that the treatment combination of R-CHOP, has no improved effect on patients harbouring the EZH2(Y641) mutation.

The third analysis comparing EZH2(Y641) DLBCL patients to patients diagnosed with DLBCL regardless of subtype, rendered a P-value of 0.054, which can almost be considered significant. None the less it should be speculated whether or not the difference is based on the significant difference known to dissociate the GCB and ABC subtypes of DLBCL, as the clinical outcome analysis comparing the EZH2(Y641) patients to GCB subtype patients did not result in a significant p-value.

CONCLUSION

A mutation assay was developed to screen patient samples for the presence of EZH2 Y641 mutations, and with the design of three different primer pairs, mutational screening can be performed using both high quality genomic DNA, compromised DNA from paraffin samples, and cDNA from samples where RNA is available. The presence of the EZH2(Y641) mutation was determined in Danish DLBCL patients and found exclusively in GCB-subtype, exclusively in a heterozygous state and with approximately the same frequency as previously determined by Morin *et al*. It was not possible to detect a genomic profile differentiating EZH2(Y641) patients from EZH2(wt), using cyto2.7M array. However a gene list of 25 differentially expressed genes was identified based on gene expression results from Ug_133_plus2 array, indicating possible target genes resulting from the increased trimethylation. The clinical outcome analyses performed showed no significant difference, between patients mutated in EZH2(Y641) and EZH2(wt).

FUTURE PERSPECTIVES

The scope of this thesis was to develop a mutation detection assay and to elucidate on the novel EZH2(Y641) mutation, within the prospects of genomic profilling, gene expression profiling and clinical outcome. In this paragraph suggestions for further elucidating the role of EZH2(Y641) in DLBCL is made.

For the aspect of gene expression, further analysis is required to validate the 25 differentially expressed genes between EZH2(Y641) and GCB-DLBCL patients. Firstly Q-PCR should be applied, to technically validate the differences found in gene expression by micro array. Secondly a larger data set should be examined with the same parameters, to biologically verify that the gene profile is distinctive for EZH2(Y641) patients. Also a more thorough literature study of the relation between collagen genes and DLBCL, should be conducted. Lastly, an chromatin immunoprecipitation analysis could be made, to determine the targets of EZH2(Y641) and verify the findings with the genes determined by microarray.

Lastly to use flow cytometry to sort cells from a biopsy sample, to determine in which step of the B-cell differentiation, that the EZH2(Y641) originates, could aid in the understanding of EZH2(Y641) in DLBCL.

REFERENCES

Kumar, V., Abbas, A.K., Fausto, N., Mitchell, R.N; "Robbins – Basic Pathology", 8th edition, 2007, Saunders –Elsvier.
ISBN: 978-0-8089-2366-4

Swerdlow, S.H., Campo, E., Harris, N.L., Jaffe, E.S. Pileri, S.A, Stein, H., Thiele, J., Vardiman, J.W.; "WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues" 4th edition, 2008, IARC – international Agency for Research on Cancer.
ISBN: 978-92-832-2431-0

Lenz, G., Staudt, M.L.; "Aggressive Lymphomas", 2010, The New England Journal of Medicine; N Engl J Med 2010;362;1417-29

Velichutina, I., Shaknovich, R., Huimin, G., Johnson, N.A, Gascoyne, R.D., Melnick, A.M., Elemento,O. ; "EZH2 mediated epigenetic silencing in germinal center B-cells contributes to proliferation and lymphomagenesis", 2010, Blood, 2010 116:5247-5255

Gascoyne, R.D., Rosenwald, A., Poppema, S., Lenz, G.; "Prognostic biomarkers in malignant lymphomas", 2010, Leukemia & Lymphoma, 2010; Early Online, 1-9

Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A., Ploegh,H., Matsudaira, P.; Molecular cell biology", 2008, sixth edition, W.H. Freemand and Company, New York,
ISBN: 978-0-7167-7601-7

Morin, R.D, Johnson N.A, Severson T.M, Mungall A.J. An, JGoya, R.,. Paul J.E, Boyle M., Woolcock B.W, Kuchenbauer F., Yap D., Humphries R K., Griffith O.L, Shah S., Zhu H., Kimbara M., Shashkin P., Charlot J. F, Tcherpakov M., Corbett R., Tam A., Varhol R., Smailus D., Moksa M., Zhao Y., Delaney A., Qian H., Birol I. , Schein J., Moore R., Holt R., Horsman D. E, Connors J. M, Jones S., Aparicio S., Hirst M., Gascoyne R. D & Marra M. A; "Somatic mutations altering EZH2(Tyr641) in follicular and diffuse large B-cell lymphomas of germinal center origin", 2010, Nature Genetics, vol 42, no. 2, feb. 2010

Sneeringer, C.J, Scott, M.P, Kuntz, K.W, Knutson, S.K, Pollock, R.M, Richon, V.M, Copeland R.A; Coordinated activites of wild type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human b-cell lymphomas", 2010, PNAS, Dec 7, 2010 vol 107, no 49

Simon, J.A. and Lange,C.A; "Roles of EZH2 histone methyltransferase in cancer epigenetics", 2008, Mutation Research 647 (2008) 21-29

Chi, P., Allis, D., Wang,G.G; "Covalent histone modifications – miswritten, misinterpreted and miserased in human cancers", Nature reviews – Cancer, vol 10, july 2010

Bunz, F. "Principles of Cancer Genetics", Springer, 2008

Shlien, A., Malkin, D.; “Copy Number variations and cancer susceptibility”, 2010, Current Opinion in Oncology 2010, 22:55-63

Margueron, R., Reinberg, D. “The polycomb complex PRC2 and its mark in life”, 2011, Nature Vol 469, January 2011,

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury III, W.J., Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., Pirrota, V., Reinberg,D., Gamblin, S.J. ; “Role of the polycomb protein EED in the propagation of repressive histone marks”, 2009 Nature Vol 461, October 2009

Chase, A., Cross, C.P.N.; “Abberrations of EZH2 in Cancer”, 2011, Clinical Cancer Research 2011;17:2613.2618

Bödör C, O'Riain C, Wrench D, Matthews J, Iyengar S, Tayyib H, Calaminici M, Clear A, Iqbal S, Quentmeier H, Drexler HG, Montoto S, Lister AT, Gribben JG, Matolcsy A, Fitzgibbon J; “EZH2 Y641 mutations in follicular lymphoma.” Leukemia. 2011 Apr;25(4):726-9.

Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Johnson NA, Severson TM, Chiu R, Field M, Jackman S, Krzywinski M, Scott DW, Trinh DL, Tamura-Wells J, Li S, Firme MR, Rogic S, Griffith M, Chan S, Yakovenko O, Meyer IM, Zhao EY, Smailus D, Moksa M, Chittaranjan S, Rimsza L, Brooks-Wilson A, Spinelli JJ, Ben-Neriah S, Meissner B, Woolcock B, Boyle M, McDonald H, Tam A, Zhao Y, Delaney A, Zeng T, Tse K, Butterfield Y, Birol I, Holt R, Schein J, Horsman DE, Moore R, Jones SJ, Connors JM, Hirst M, Gascoyne RD, Marra MA; “Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma.” Nature. 2011 Jul 27;476(7360):298-303. doi: 10.1038/nature10351. [Morin et al_2]

Vaillant, I. and Paszkwoski, J.; “Role of histone and DNA methylation in gene regulation”, current opinion in plant biology, 2007, 10:528-533

Bannister, J.A. and Kouzarides, T.; “Regulation of chromatin by histone modifications”, Cell research, 2011, 21:381-395 ‘

Sparmann, A. and Lohuizen, M.v.; “Polycomb silencers control cell fate, development and cancer” Nature Reviews Cancer 6, 846–856 (November 2006) |10.1038/nrc1991

Dillion, S.C., Zhang, X., Trivel, R.C., Cheng, X.;” The SET-domain protein superfamily: Protein lysine methyltransferases”, 2005, Genome Biology 6:227

Schapira, M; “Structural Chemistry of the Human SET Domain Protein Methyltransferases”, 2011, Current Chemical Genomics, 5, 85-94

Del Rizzo, P.A., Couture, J-F., Dirk, L.M.A., Strunk, B.S., Roiko,M.S., Brunzelle,J.S., Houtz,R.L, Trivel, R.C.;” SET7/9 Catalytic Mutants Reveal the Role of Active Site Water Molecules in Lysine Multiple Methylation”, 2010, The Journal of Biological Chemistry, vol 285 #41, 31849-31858.

Couture, J-F., Dirk, L.M.A., Brunzelle, J.S., Houtz, R.I., Trivel, R.C.;"Structural origins for the product specificity of the SET domain protein methyltransferases", 2008, PNAS, dec. 30 2008, #52, 20659-20664

Rush, M., Appanah, R., Lee, S., Lam,L.L., Goyal, P. Lorincz,M.C.; Targeting of EZH2 to a defined genomic site is sufficient for recruitment of Dnmt3a but not de novo DNA methylation", 2009, Epigenetics 4:6, 404-414, aug 16
Hanahan, D., Weinberg R.A., Hallmarks of Cancer: The Next Generation, Cell, Volume 144, Issue 5, 4 March 2011, Pages 646-674, ISSN 0092-8674, 10.1016/j.cell.2011

Lenz G et al. "Stromal gene signatures in large-B-cell lymphomas". 2008, *New England Journal of Medicine,* 359:2313-23.

de Groot, P. https://madmax.bioinformatics.nl, April 2006

O'Keefe,C., McDevitt,M.A. and Maciejewski,J.P. "Copy Neutral Loss of Heterozygosity, 2010, Blood", 8 April 2010 _ Volume 115, Number 14

Kræftens bekæmpelse, 2009,
http://www.cancer.dk/Hjaelp+viden/fakta+om+kraeft/kraeft+i+tal/kraeft+i+tal.html
Last updated 2011, last visited november 2011

Kræftens bekæmpelse, 2011,
http://www.cancer.dk/Hjaelp+viden/kraeftformer/kraeftsygdomme/nonhodgkinlymfom/
last updated 02-02 2011, last visited November 2011

van Kemenade, F.J., Raaphorst, F.M., Blokzijl, T., Fieret, E., Hamer, K.M., Satijn, D.P.E., Otte, A.P., and Meijer, C.J. L.M., "Coexpression of BMI-1 and EZH2 polycomb-group proteins is associated with cycling cells and degree of malignancy in B-cell non-Hodgkin lymphoma" Blood 2001 97:3896-3901; doi:10.1182/blood.V97.12.3896

Margueron, R.,Li,G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., Reinberg, D., "Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms", Molecular Cell, Volume 32, Issue 4, 21 November 2008, Pages 503-518, ISSN 1097-2765, 10.1016/j.molcel.2008.11.004.

Wright G et al. "A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma". 2003, *Proceedings of the National Academy of Sciences of the United States of America* 100:9991-6.

Wigle, T.J., Knutson, S.K., Jin, L., Kuntz, K.W., Pollock, R.M., Richon, V.M., Copeland, R.A., Scott, M.P.," The Y641C mutation of EZH2 alters substrate specificity for histone H3 lysine 27 methylation states", FEBS Letters, Volume 585, Issue 19, 3 October 2011, Pages 3011-3014, ISSN 0014-5793, 10.1016/j.febslet.2011.08.018.

Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Johnson NA, Severson TM, Chiu R, Field M, Jackman S, Krzywinski M, Scott DW, Trinh DL, Tamura-Wells J, Li S, Firme MR, Rogic S, Griffith M, Chan S, Yakovenko O, Meyer IM, Zhao EY, Smailus D, Moksa M, Chittaranjan S, Rimsza L, Brooks-Wilson A, Spinelli JJ, Ben-Neriah S, Meissner B, Woolcock B, Boyle M, McDonald H, Tam A, Zhao Y, Delaney A, Zeng T, Tse K, Butterfield Y, Birol I, Holt R, Schein J, Horsman DE, Moore R, Jones SJ, Connors JM, Hirst M, Gascoyne RD, Marra MA. "Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma." Nature. 2011 Jul 27;476(7360):298-303. doi: 10.1038/nature10351.

Greenman, C., Wooster, R., Futreal, P.A., Stratton, M.R. and Easton, D.F.; "Statistical Analysis of Pathogenicity of Somatic Mutations in Cancer" Genetics August 2006 173:2187-2198; published ahead of print June 18, 2006, doi:10.1534/genetics.105.044677

Glazko, G.V., Babenko, V.N., Koonin, E.V. and Rogozin, I.B.; "Mutational hotspots in the TP53 gene and, possibly, other tumor suppressors evolve by positive selection"
Biology Direct 2006, 1:4 doi:10.1186/1745-6150-1-4

Taylor, C.F. "Mutation scanning using high-resolution melting" 2009, Biochem Soc Trans 37:433-7. 2009

Reed, G.H., O Kent, J., and Wittwer, C.T. "High-resolution DNA melting analysis for simple and efficient molecular diagnostics" 2007, Pharmacogenomics 2007 8:6, 597-608

NCBI "Micro array fact sheet" ,2007, http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html last updated 27/7-2007, last visited 25/11/11

Shuga, J., Zeng, Y., Novak, R., Mathies, R. A., Hainaut, P. and Smith, M. T. (2010), Selected technologies for measuring acquired genetic damage in humans. Environmental and Molecular Mutagenesis, 51: 851–870. doi: 10.1002/em.20630

Heinrichs, S., Li, C. & Look, A.T., 2010. SNP array analysis in hematologic malignancies: avoiding false discoveries. *Blood*, 115(21), p.4157-4161.

Park, S.W., Chung, N.G., Eom, H.S., Yoo, N.J., Lee, S.H.; "Mutational analysis of EZH2 codon 641 in non-Hodgkin lymphomas and leukemias", Leukemia Research, Volume 35, Issue 1, January 2011, Pages e6-e7, ISSN 0145-2126, 10.1016/j.leukres.2010.08.018.

Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J-M., Bollen, M., Esteller, M., Di Croce, L., de Launoit, Y. and Fuks, F.; "The Polycomb group protein EZH2 directly controls DNA methylation" *Nature* 439, 871-874 (16 February 2006) | doi:10.1038/nature04431

S. Blenk, J. Engelmann, M. Weniger, J. Schultz, M. Dittrich, A. Rosenwald, H.K. Müller-Hermelink, T. Müller and T. Dandekar "Germinal Center B Cell-Like (GCB) and Activated B Cell-Like (ABC) Type of Diffuse Large B Cell Lymphoma (DLBCL): Analysis of Molecular Predictors, Signatures, Cell Cycle State and Patient Survival" *Cancer Informatics* 2007:3 399-420

APPENDIX – A

HIGH RESOLUTION MELTING

High Resolution Melting analysis (HRM) is an analytical method used for genotyping, mutation scanning and sequence matching. It is a simple PCR based method, detecting DNA sequence variation through measurement of changes in the melting profile of a DNA duplex. The PCR is carried out in the presence of a suitable dye, which can bind to double stranded DNA (dsDNA), but not single stranded DNA(ssDNA). Following the PCR, temperature rises and the DNA duplex will separate into ssDNA and the melting profile will be generated based on the reduction in the fluorescence generated by the dyes incorporated into the dsDNA, see Figure 33.A. [Taylor, 2009] The melting temperature of the DNA duplex is influenced by different factors, such as GC content, amplicon length and sequence. [Taylor, 2009, Reed *et al.*, 2007]

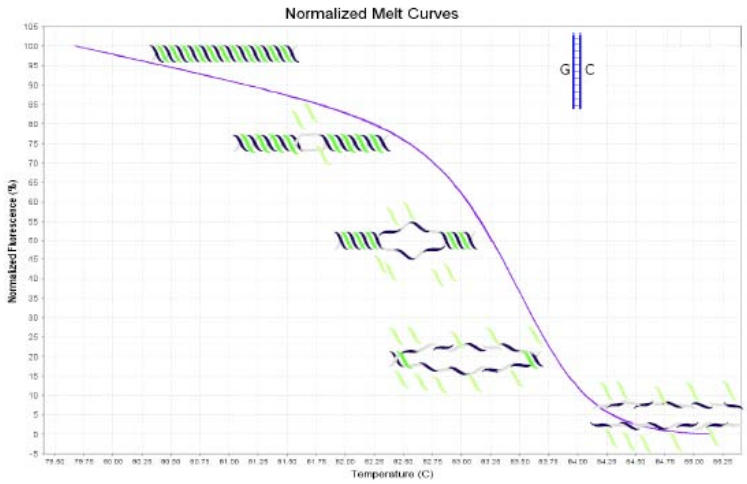

Figure 33.A illustrates the principle behind HRM analysis.the dissociation of dsDNA as the temperature increases, releasing the dye and creating a decrease in fluroresence intensity [Tavaria, 2006].

The basis for detection of mutations by HRM is created by the change in melting temperature (Tm) that occurs when different duplexes are formed from a heterozygous sample. The different duplexes created are a wt, a homozygous and heterozygous. The wt melting curve will have a different Tm than the homozygous duplex, while the heterozygous duplex will have a melting curve, which differs from the first two types of duplexes; see Figure 34.A [Taylor, 2009] For HRM analysis mutation scanning, the amplicon size of the PCR product, must be within a range of 50bp 1kbp, however it is recommended that amplicon size is less than 400bp, as the sensitivity and specificity of the method is a 100%. For higher amplicons, the specificity and sensitivity declines. It should be noted that different experimentally set ups, may have more specific limitations to the amplicon size. [Reed *et al.,* 2007]

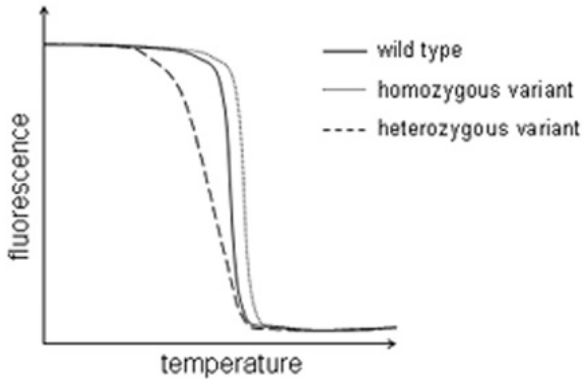

Figure 34.A. illustrates the difference in melting curves between the three different duplexes. The wild type duplex and homozygout will variate in Tm, while the heterozygous will variate in melting curve. [Taylor, 2009]

APPENDIX – B

MICROARRAY

Microarray is a technology with many applications in regard to gene analysis. The most prominent applications are gene expression (mRNA based) and genomic DNA based methods such as cytogenetic arrays, used for detecting genomic alterations. The microarray techniques is chip based, using hybridization probes which are fluorescently labeled. In general the probes will hybridize with complementary sequences found in the test sample. After the hybridization has been completed, the micro array can be scanned, using a complex consisting of lasers, a specialized microscope and a camera. The function of the laser is to excite the fluorescent probes, so that a digital image of the array can be captured by the microscope and camera. Then the data is ready for analysis by a specialized program, designed for the specific type of array. See figure 1.B for a schematic overview of micro array analysis for gene expression. Micro array yield a picture of the genomic state of the sample at a given time and location. [NCBI, 2007]

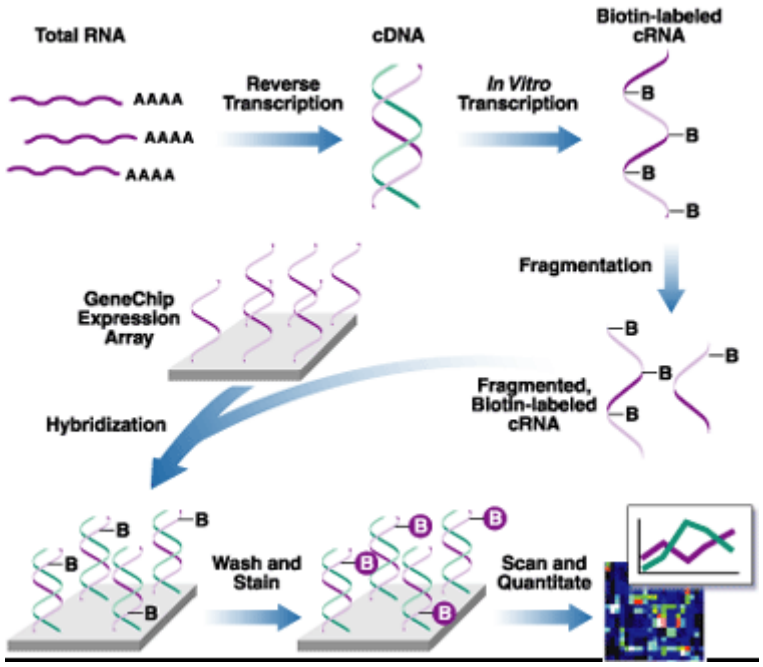

Figure 1.B illustrates the technique behind the micro array platform for gene expression, based on RNA. [Affymetrix, 2002]

Microarray used for determining the level of expression of a specific gene is a method based cDNA derived from mRNA, and is commonly known as micro array expression analysis [NCBI, 2007]

DNA micro array are used for determining genomic alterations, such as Copy Number Variations (CNV), Single Nucleotide Polymorphisms (SNPs) and Loss Of Heterozygosity (LOH). Those used for detecting SNPS, are referred to as SNP micro array and can also determine copy number neutral loss of heterozygosity. [Schuga *et al.*, 2010] Current SNP arrays can analyse <900.000 loci, yielding data for genotyping and copy number. [Heinrichs *et al.*, 2010]

APPENDIX – C

R-SCRIPT FOR CLINICAL OUTCOME ANALYSIS

#Generate data

```
Y641 <- read.csv("Ekstern/Overlevelses analyse.csv", header = TRUE,
            sep = ";")
```

#Clinical outcome analysis

```
data.end     <- as.Date(Y641$Dødsdato.OS.opfølgning, format = "%d-%m-%Y")
data.start   <- as.Date(Y641$Dato.For.Diagnostisk.Biopsi, format = "%d-%m-%Y")

Y641$OS.weeks <- as.numeric(data.end - data.start) / 7

Y641.5year   <- Y641
Y641.5year$OS.Status[Y641.5year$OS.weeks > 5*52] <- 0

Y641.5year$OS.weeks[Y641.5year$OS.weeks > 5*52] <- 5*52
```

##

GCB vs GCB(Y641)

##

#All data

```
pdf(paste("Output", "/GCBvsGCB(Y641)all.pdf", sep = ""),width = 7,
  height = 7)

par(mfrow = c( 1, 1))

Y641.new <- Y641[Y641$Subtype == "GCB", ]

survival.data <- Surv(Y641.new$OS.weeks,as.numeric(Y641.new$OS.Status))
survival.obj <- survfit(survival.data ~ Y641.new$Mutation)
```

Log-rank test for forskel mellem alle tre grupper

```
survdiff(survival.data ~ Y641.new$Mutation)

survplot(survival.data ~ Y641.new$Mutation,
      stitle = "Y641",
      xlab = "Time (weeks)",
      ylab = "Survival ratio")

dev.off()
```

5 year survival

```
pdf(paste("Output","/GCBvsGCB(Y641)5years.pdf", sep = ""),width=7,height=7)

par(mfrow=c(1,1))

Y641.new <- Y641.5year[Y641.5year$Subtype == "GCB",]

survival.data <- Surv(Y641.new$OS.weeks,as.numeric(Y641.new$OS.Status))
survival.obj <- survfit(survival.data ~ Y641.new$Mutation)
```

Log-rank test for forskel mellem alle tre grupper

```
survdiff(survival.data ~ Y641.new$Mutation)

survplot(survival.data ~ Y641.new$Mutation,
      stitle = "Y641",
      xlab   = "Time (weeks)",
      ylab   = "Survival ratio")

dev.off()
```

##

All vs All(Y641)

##

All data

```
pdf(paste("Output","/AllvsAll(Y641)all.pdf", sep = ""),width = 7,height = 7)
```

```
par(mfrow = c(1, 1))

survival.data <- Surv(Y641$OS.weeks, as.numeric(Y641$OS.Status))
survival.obj <- survfit(survival.data ~ Y641$Mutation)
```

Log-rank test for forskel mellem alle tre grupper

```
survdiff(survival.data ~ Y641$Mutation)

survplot(survival.data ~ Y641$Mutation,
      stitle = "Y641",
      xlab = "Time (weeks)",
      ylab = "Survival ratio")

dev.off()
```

5 year survival

```
pdf(paste("Output","/AllvsAll(Y641)5years.pdf", sep = ""), width = 7,
  height = 7)

par(mfrow = c(1, 1))

survival.data <- Surv(Y641.5year$OS.weeks,as.numeric(Y641.5year$OS.Status))
survival.obj <- survfit(survival.data ~ Y641.5year$Mutation)
```

Log-rank test for difference of all three subtypes

```
survdiff(survival.data ~ Y641.5year$Mutation)

survplot(survival.data ~ Y641.5year$Mutation,
      stitle = "Y641",
      xlab  = "Time (weeks)",
      ylab  = "Survival ratio")

dev.off()
```

##

CHOP-R vs CHOP-R(Y641)

##

All data

```
pdf(paste("Output","/CHOP-RvsCHOP-R(Y641)all.pdf", sep = ""), width = 7,height = 7)

par(mfrow = c(1, 1))

Y641.new <- Y641[Y641$Behandling == "CHOP-R",]

survival.data <- Surv(Y641.new$OS.weeks, as.numeric(Y641.new$OS.Status))
survival.obj <- survfit(survival.data ~ Y641.new$Mutation)
```

Log-rank test for forskel mellem alle tre grupper

```
survdiff(survival.data ~ Y641.new$Mutation)

survplot(survival.data ~ Y641.new$Mutation,
      stitle = "Y641",
      xlab   = "Time (weeks)",
      ylab   = "Survival ratio")

dev.off()
```

5 year survival

```
pdf(paste("Output","/CHOP-RvsCHOP-R(Y641)5years.pdf", sep = ""),
  width=7,height=7)

par(mfrow = c(1,1))

Y641.new <- Y641.5year[Y641.5year$Behandling == "CHOP-R",]

survival.data <- Surv(Y641.new$OS.weeks, as.numeric(Y641.new$OS.Status))
survival.obj <- survfit(survival.data ~ Y641.new$Mutation)
```

Log-rank test for difference of all three subtypes

```
survdiff(survival.data ~ Y641.new$Mutation)

survplot(survival.data ~ Y641.new$Mutation,
```

```
                stitle = "Y641",
                xlab   = "Time (weeks)",
                ylab   = "Survival ratio")

dev.off()
```

APPENDIX –D

HIGH RESOLUTION MELTING RESULTS

The remaining melting curves and difference plots will be shown in this appendix.

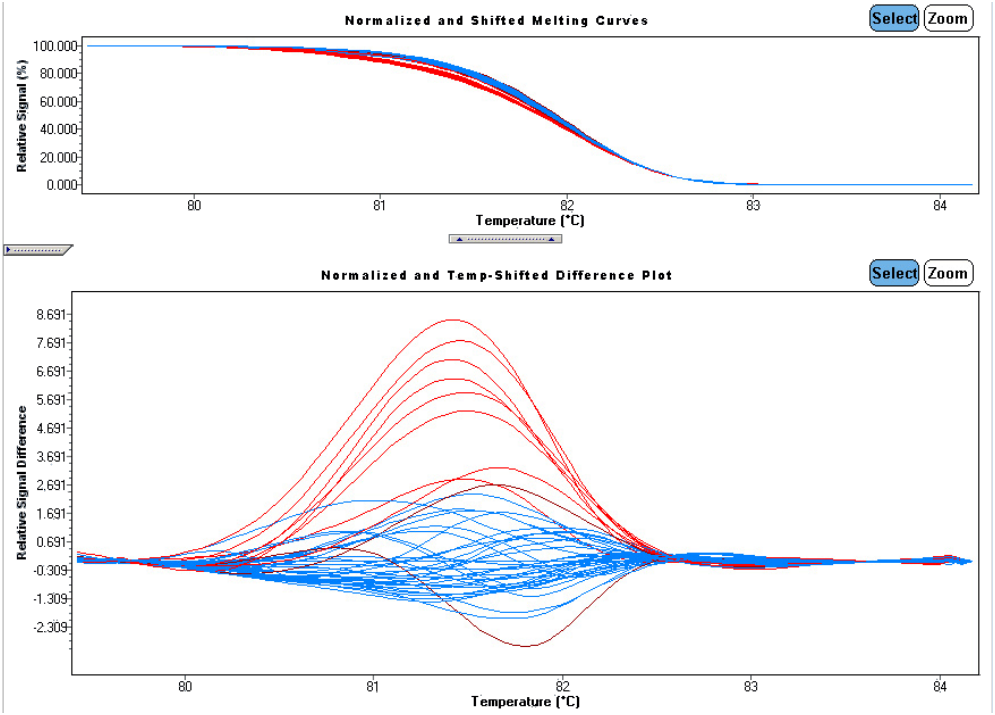

Figure 1.D shows sample H9-H30, sample group 1. Here H9, H15 displayed heteroduplex melting curves similar to DB.

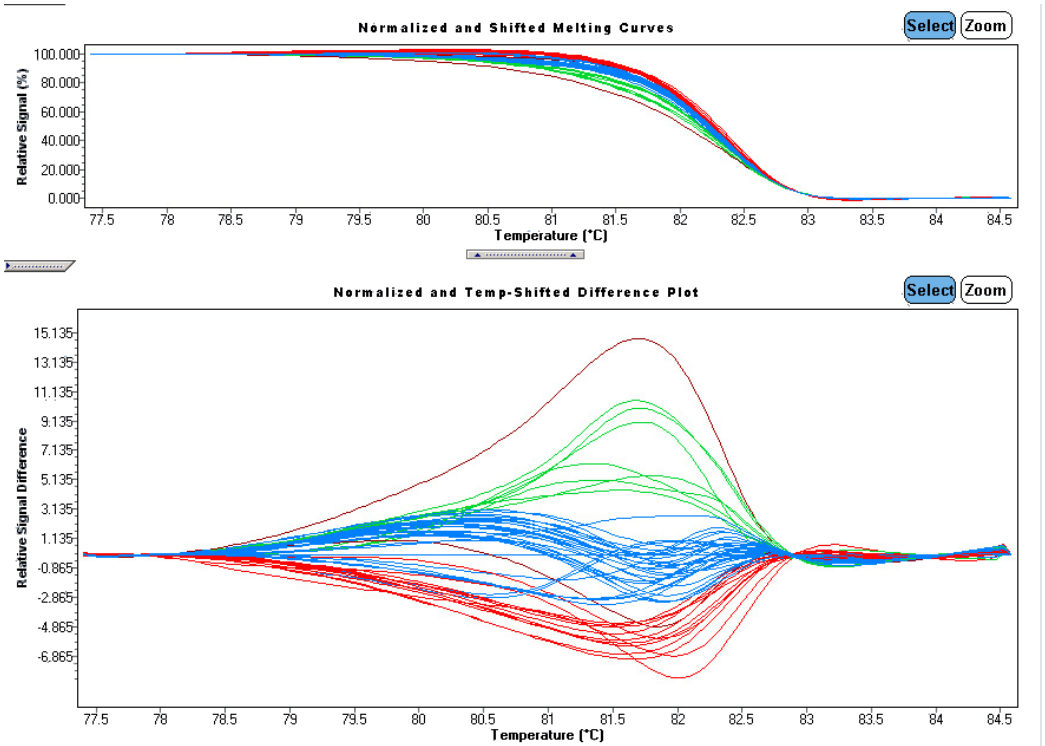

Figure 2.D show H31-H55. H37 has similar heteroduplex melting curve to DB (green).

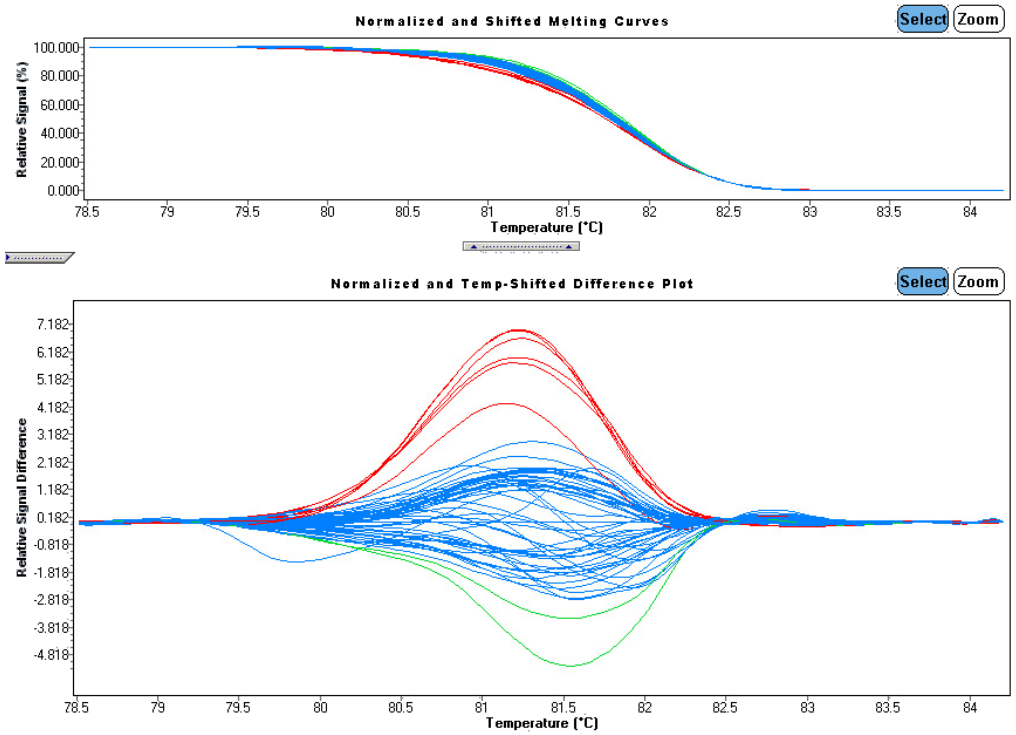

Figure 3.D show H56-60, sample group 1, along with H281-H319, sample group 2. H60 and H302 displayed a heteroduplex melting curve similar to DB (red).

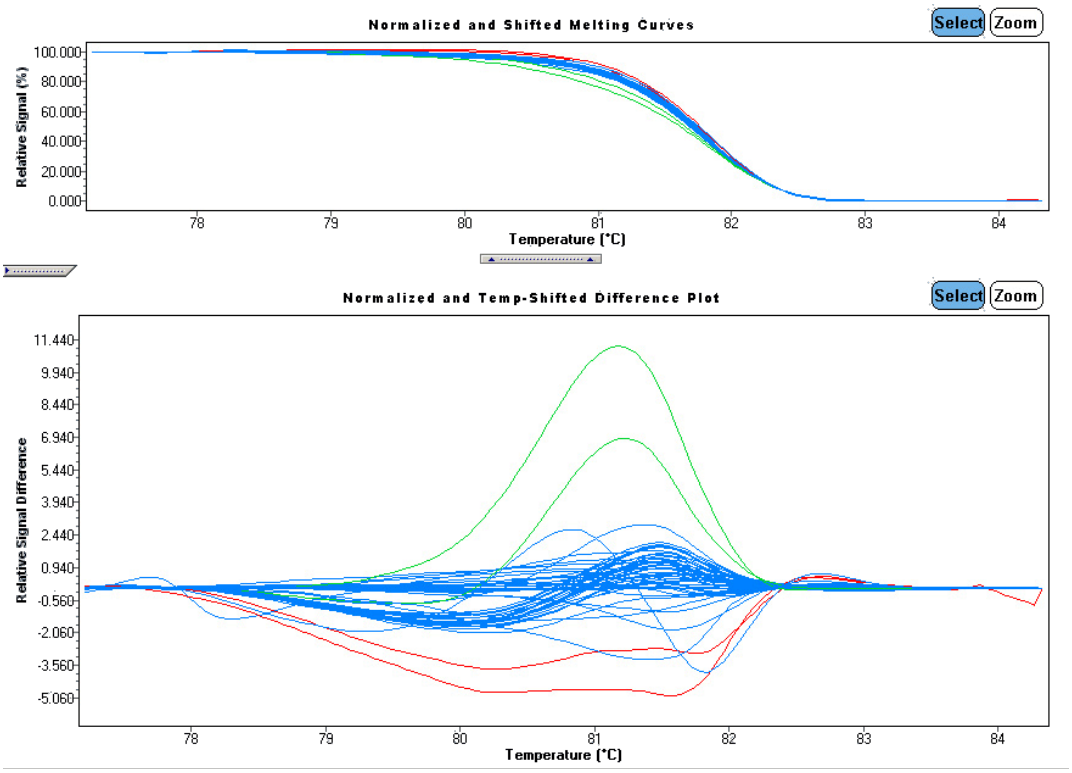

Figure 4.D H285,H320-H342, sample group two. None of the analysed samples displayed at heteroduplex melting curve similar to DB (green).

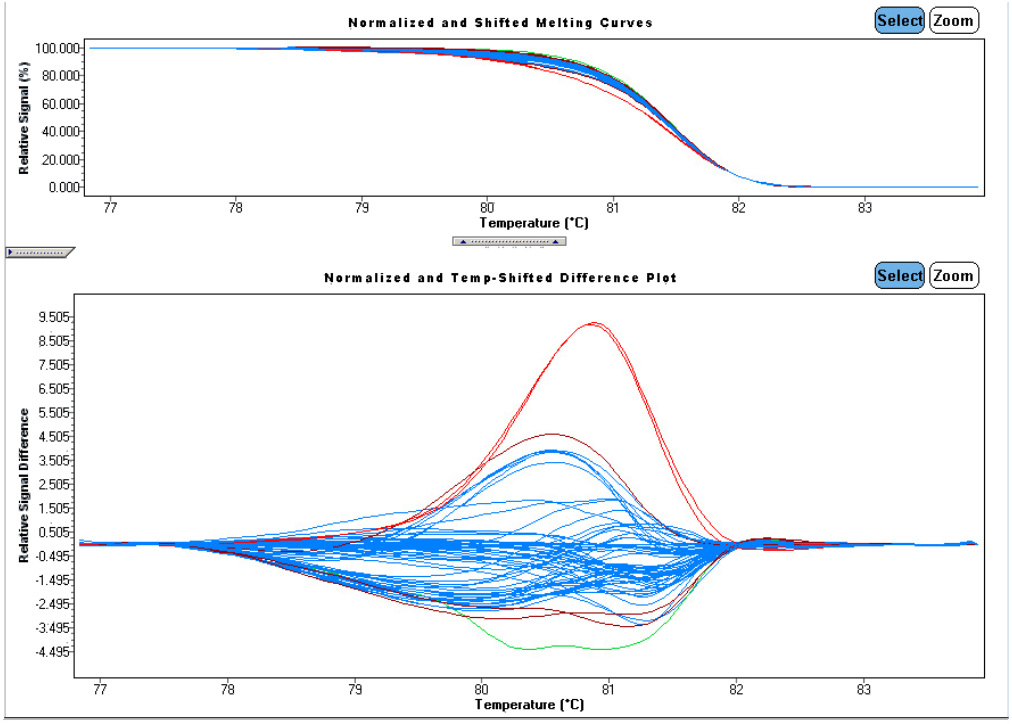

Figure 5.D show H348-H484, sample group 2. None of the samples displayed the same distinct heteroduplex melting curve as DB (red), but the intermediates were still selected for sequencing.

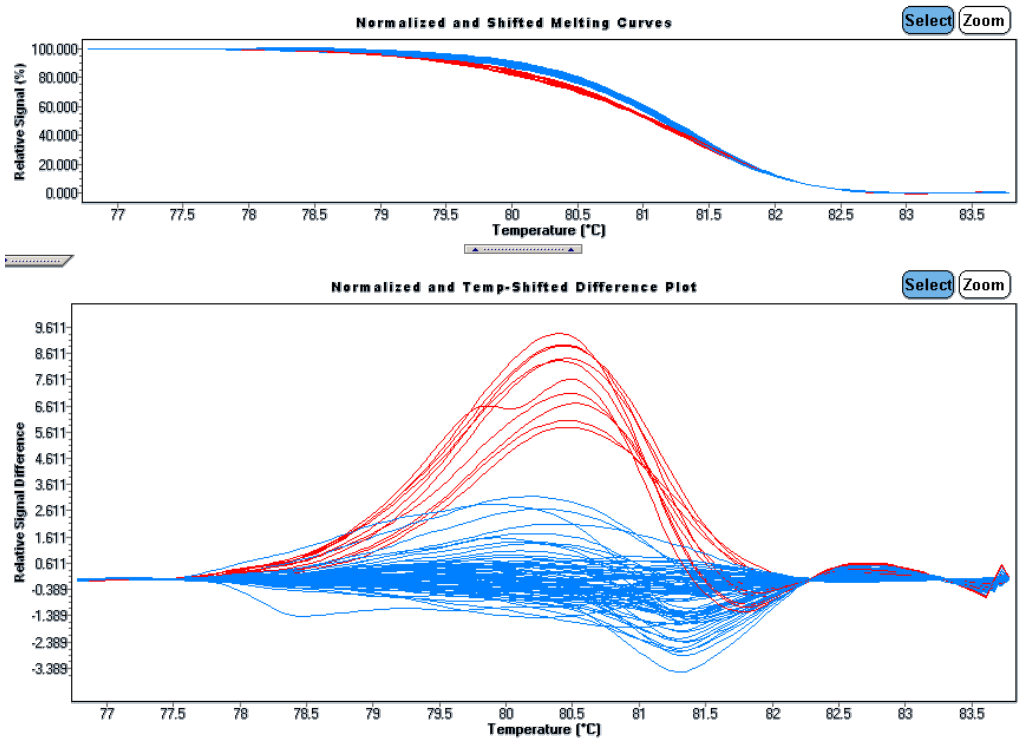

Figure 6.D show H1-H45, sample group 1, validated with 111bp primer set. H7,H9,H15,H37 show heteroduplex melting curves, as expected from the first analysis.

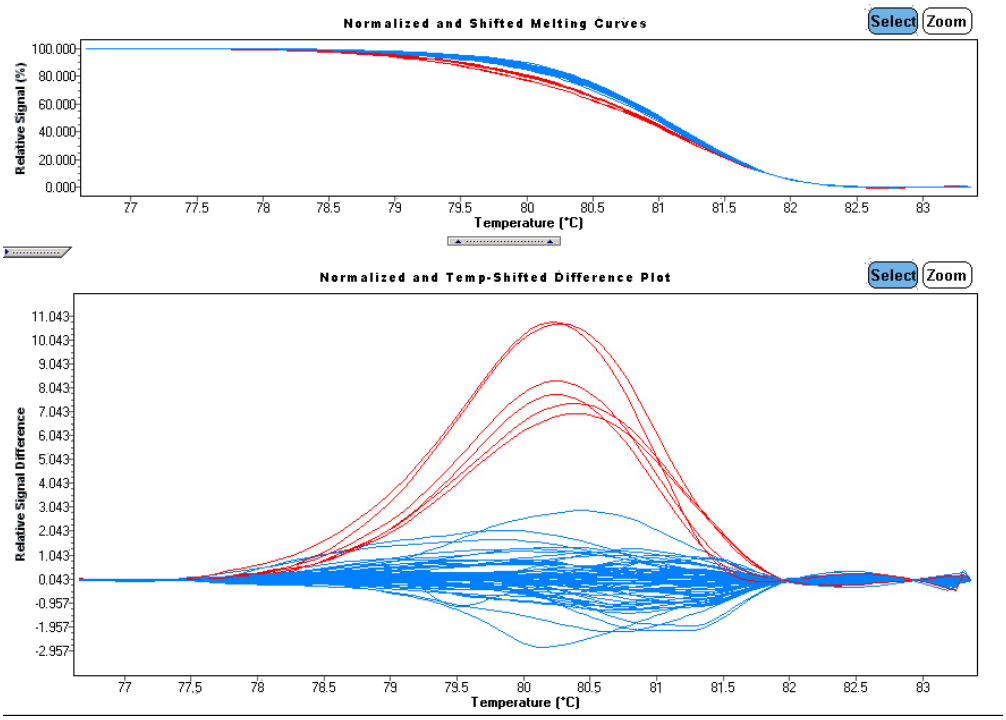

Figure 7.D show H46-H60, sample group 1 and H284-H484, sample group 2, validated with 111bp primer set. H60 and H302 show heteroduplex melting curves, as expected from the first analysis.

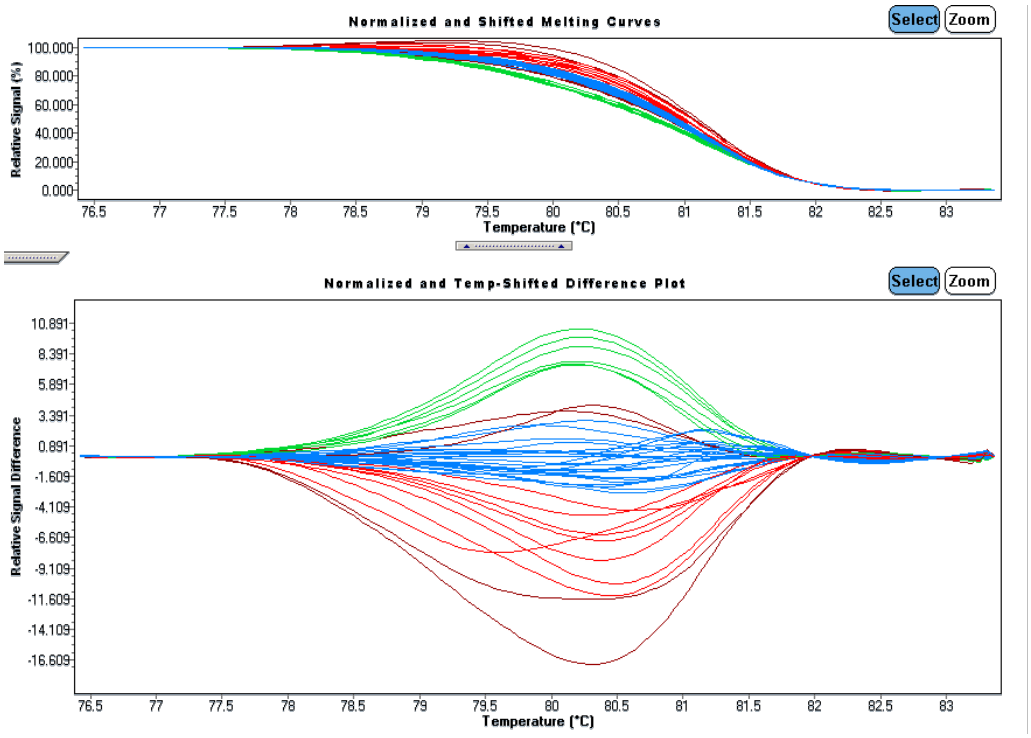

Figure 8.D show validation of H213-H238, sample group 4, where H213 and H217 diplay heteroduplex melting curves similar to DB, as expected from the first analysis.