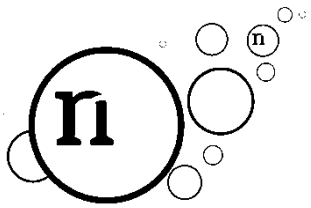
Cisplatin Sensitivity in Diffuse Large B-cell Lymphoma and the Influence of Mutations in DNA Repair Genes

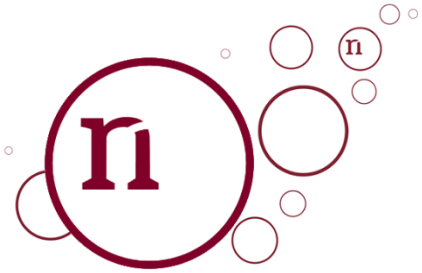
by Ann-Katrine Hyttel Rom Pedersen

AALBORG UNIVERSITY
DENMARK



Master Thesis in Biomedicine
Medicine with Industrial specialization
Aalborg University

In collaboration with the Department of Hematology
Aalborg University Hospital



CISPLATIN SENSITIVITY IN DIFFUSE LARGE B-CELL LYMPHOMA AND THE INFLUENCE OF MUTATIONS IN DNA REPAIR GENES

Master Thesis

9th and 10th semester

Medicine with Industrial Specialization, Biomedicine

Author

Ann-Katrine Hyttel Rom Pedersen

Stud.cand.scient.med

Aalborg University

Supervisors

Karen Dybkær Sørensen
Professor
Department of Hematology
Aalborg University Hospital

Julie Støve Bødker
Senior Researcher
Department of Hematology
Aalborg University Hospital

Hanne Due Rasmussen
Ph.D. student
Department of Hematology
Aalborg University Hospital

Specifications

Number of pages:	66 pages
Keystrokes total:	22.477
Separate appendix:	83 pages
Project period:	September 3rd 2018 – May 31st 2019

School of Medicine and Health
Aalborg University
May 31st 2019

ABSTRACT

Background

Diffuse large B-cell lymphoma (DLBCL) is the most common type of non-Hodgkin lymphoma. However, it is a heterogenous cancer both between and within classification with a variable clinical course and prognosis: 30-40 % of DLBCL patients relapse. Platinum-based chemotherapy, such as cisplatin, is often administered in the relapse situation. Treatment resistance remains a major hurdle in DLBCL. DNA repair mechanisms, such as nucleotide excision repair and mismatch repair are important for cisplatin's mechanism of action, as they are activated by cisplatin-induced crosslinks and mismatches and are therefore implicated in cisplatin resistance. Impaired mismatch repair has been proposed to lead to resistance, whereas impaired nucleotide excision repair is linked to increased sensitivity. Therefore, the overall aim of this thesis was to investigate the DNA repair related mechanism behind cisplatin resistance.

Methods

Mutations in DNA repair related genes, was investigated 14 DLBCL cell lines, a clinical cohort of 1,001 de novo DLBCL patients, kindly provided by DaveLab, Durham, USA, and another clinical cohort of Hematology Aalborg 28 relapsed DLBCL and 44 de novo DLBCL patients (in-house), and evaluated using a developed systematic assessment tool.
17-(Allylamino)-17-demethoxygeldanamycin (17AAG) was identified as an MSH2 inhibitor and utilized to investigate the effect on DLBCL cell lines' cisplatin response in vitro. Likewise, Comet Assay® was performed on cells affected by cisplatin and/or 17AAG to evaluate on the formation of DNA fragments. Lastly, graphical protein variant views were computed, illustration mutation variants of *XPF*, *EXO1*, and *MSH2* found in the 14 DLBCL cell lines or DLBCL clinical cohort.

Results

Mutation of DNA repair related genes were predominantly found in cisplatin sensitive DLBCL cell lines and in mismatch repair, nucleotide excision repair, and homologous enjoining. 17AAG was found to decrease cisplatin sensitivity of DLBCL cell lines, and decrease the amount of free DNA fragments, implicating MSH2 as a player in cisplatin resistance. *XPF*, *EXO1*, and *MSH2* was also found to harbor mutations in the clinical cohort.

Conclusion

17AAG inhibited MSH2, impairing mismatch repair, and leading to cisplatin resistance. Additionally, *MSH2* was found to be fairly mutated in the clinical cohort, with a slight accumulation of mutations in the MutS_I domain. From this thesis, mismatch repair seems an important player in mediating cisplatin resistance. However, the possible link between impaired MSH2 and cisplatin resistance would need to be investigated further in a large clinical cohort of relapsed DLBCL patients.

DANSK RESUMÉ

Baggrund

Diffus storcellet b-celle lymfom (DLBCL) repræsenterer den største andel af non-Hodgkin lymfom. Det er dog en heterogen cancer både inden for og imellem forskellige klassifikationer. Der ses et varierende klinisk billede og prognose, hvor 30-40 % af DLBCL-patienter får relaps. Platin-baseret kemoterapi, fx cisplatin, gives ofte til relaps-patienter, men resistens over for behandlingen er stadig et stort problem. DNA-reparationsmekanismer, fx nucleotide excision repair og mismatch repair, er vigtige for cisplatins virkningsmekanisme, eftersom de aktiveres af cisplatin-inducerede krydsbindinger og fejlparringer. Derfor bliver især de to mekanismer forbundet med cisplatin-resistens; forringet mismatch repair er blevet foreslået at lede til resistens, hvorimod forringet nucleotide excision repair er forbundet med øget sensitivitet. Derfor var det overordnede mål af dette Speciale at undersøge mekanismen bag cisplatin resistens relateret til DNA reparation.

Metoder

Mutationer i gener relateret til DNA reparation blev undersøgt i 14 DLBCL cellelinjer, en klinisk kohorte på 1.001 de novo DLBCL-patienter, fra DaveLab, Durham, USA, og endnu en klinisk kohorte på 28 relaps-DLBCL og 44 de novo DLBCL-patienter (in-house) fra Hæmatologi Aalborg. Varianterne blev evalueret ud fra et udviklet systematisk vurderingsværktøj.
17-(Allylamino)-17-demethoxygeldanamycin (17AAG) er blevet identificeret som en MSH2 inhibitor. Den blev brugt i dette studie til at undersøge, hvilken effekt det ville have på cellelinjernes cisplatin-respons in vitro. Comet Assay® er udarbejdet på celler, der er blevet påvirket af cisplatin og/eller 17AAG, for at vurdere mængden af DNA-fragmenter.
Grafiske afbildninger af protein varianter fundet i *XPF*, *EXO1*, and *MSH2* er blevet udformet for de varianter fundet i 14 DLBCL cellelinjer og den kliniske DLBCL-kohorte.

Resultater

Mutationer i gener relateret til DNA reparation blev hovedsageligt fundet i cisplatin-sensitive DLBCL cellelinjer og i mismatch repair, nucleotide excision repair og homologous recombination. 17AAG nedsatte cisplatin sensitivitet for DLBCL cellelinjer og nedsatte mængden af frie DNA-fragmenter. Dette hentyder til, at MSH2 spiller en rolle i cisplatin resistens. Der blev fundet mutationer i både *XPF*, *EXO1*, og *MSH2* i den kliniske kohorte.

Konklusion

17AAG inhiberede MSH2, derigennem forringede mismatch repair, og ledte til cisplatin resistens. Mutationer i *MSH2* var ligeledes identificeret i den kliniske kohorte, med en mindre akkumulering af mutationer i MutS_I domænet. Ud fra dette Speciale, virker MMR som en vigtig faktor i mediering af cisplatin resistens. Den mulige forbindelse imellem påvirket MSH2 og cisplatin resistens skal undersøges i en stor klinisk kohorte bestående af DLBCL-patienter i en relaps situation.

TABLE OF CONTENT

ABBREVIATIONS

* any letter

17AAG	17-(Allylamino)-17-demethoxygeldanamycin
ABS	Absolute Quantification of Target
AP site	Apurinic/apyrimidinic site
APE1	Apurinic/apyrimidinic endonuclease 1
ATM	Ataxia-telangiectasia Mutated Kinase
ATR	Ataxia-telangiectasia and Rad3-related
ATRIP	ATR-interacting protein
AUC	Area under curve
BER	Base excision repair
BRCA*	Brest cancer gene
CCLE	Cancer Cell Line Encyclopedia
CHK1	Checkpoint kinase 1
CS*	Cockayne syndrome WD repeat protein A
CtIP	CtBP-interacting protein
DDB	DNA damage-binding protein
ddPCR™	Droplet Digital™ Polymerase Chain Reaction
DDR	DNA damage response
DLBCL	Diffuse Large B-cell Lymphoma
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
DR	Direct repair
DSB	Double stranded break
dsDNA	Double stranded DNA
EGCG	Polyphenol epigallocatechin-3-gallate
ERCC*	Excision repair cross-complementation group
EXO1	Exonuclease 1
FBS	Fetal bovine serum
GG-NER	Global genome nucleotide excision repair
GI_{50}	Drug concentration resulting in 50% growth inhibition
HR	Homologous recombination
Hsp90	Heat shock protein 90
LIG*	DNA ligase 1
MMR	Mismatch repair
MSH*	MutS protein homolog

NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
NSCLC	Non-small cell lung cancer
NTC	No-template control
OGG1	8-oxoguanine DNA glycosylase
PARP*	Poly(ADP-ribose) polymerase
PCNA	Proliferating cell nuclear antigen
Pen/Strep	Penicillin-streptomycin
POL	DNA polymerase
RAD23B	RAD23 homolog B
RAD*	Role of Radiation gene
RFC	Replication factor C
RPA2	Replication protein A
SD	Standard division
ssDNA	Single stranded DNA
TC-NER	Transcription-coupled nucleotide excision repair
UNG	Uracil-DNA glycosylase
XLF	XRCC4-like factor
XLF	XRCC4-like factor
XP*	Xeroderma pigmentosum group *-complementing
XRCC*	X-ray repair cross-complementing protein

1 INTRODUCTION

1.1 DIFFUSE LARGE B-CELL LYMPHOMA

Diffuse large B-cell lymphoma (DLBCL) is a neoplasm of large B-cell lymphocytes.[1–3] It is the most common type of non-Hodgkin lymphoma, representing approximately 40% of non-Hodgkin lymphoma in Denmark, with 450 new cases each year and a median age of 67 years at diagnosis.[4] It presents as a rapidly growing primary tumor with a diffuse growth pattern.[1,2] As the mitotic rate is high, patients often show involvement of one or multiple lymph nodes or extranodal sites.[1,2] Half of the patients are diagnosed at advanced stages, grade III or IV.[1,2] DLBCL can be subcategorized according to origin of the B-cells: germinal center B-cell line (GCB), representing 40-50% of DLBCLs and activated B-cell like (ABC), composing 50-60% of DLBCLs.[1,2] 10-15% are unclassifiable.[1,2] Some DLBCL dominant acting mutations are more prominent in one subgroup. For example, mutations in *EZH2* (involved in histone methylation) and *GNA13* (involved in B-cell homing) are more associated with GCB, whereas mutations in *MYD88* (involved in Toll-like receptor signaling and interleukin-1 signaling) and *CD79B* (involved in B-cell receptor signaling) are more prevalent in ABC.[1] Rearrangements in *MYC* with *BCL2* and/or *BCL6* are shared between GCB and ABC, and are known as either double hit or triple hit lymphoma, respectively.[1,2]

Another subcategorization of DLBCL is according to tumor morphology. The most common variants include the centroblastic variant, the immunoblastic variant, and the anaplastic variant.

The centroblastic variant comprises 80 % of all DLBCL tumors, and contains predominantly centroblasts, large cells with a vesicular nucleus.[1,2] This subtype is most frequent in GCB.[2] The immunoblastic variant represents 8-10 % of DLBCL tumors, and is defined by the clear dominance (≥90 %) of immunoblasts, large lymphoid cells with a central trapezoid nucleolus.[1,2] The anaplastic variant encompasses only 3 % of DLBCL tumors, and is characterized by the dominance of large lymphoma cells with bizarre a nucleus.[1,2]

A key player in the pathogenesis of DLBCL is gene mutations and deletions, which can occur at any point in the clonal expansion of the cancerous B-cells.[1] A DLBCL neoplasm presents with an average of 75-90 mutations, some being driver mutations that arise early in the lymphomagenesis with a great involvement in the pathogenesis, and others being passenger mutations that occur later in the progress with no apparent effect on the pathogenesis.[1] With further progression, the individual DLBCL undergoes clonal selection, increasing survival abilities.[1] Several cellular processes and pathways may be influenced by these DLBCL-related mutations, e.g. histone modification, cell growth, DNA damage response, B-cell receptor signaling, Toll-like receptor signaling, and angiogenesis.[1] *TP53*, a known tumor suppressor gene, is mutated in 20 % of DLBCL cases.[1] The encoded protein, p53, is involved in many cellular functions, such as the cell cycle arrest, DNA repair, and apoptosis. In a healthy cell, p53 assist in the control of mutations that would otherwise have a great impact on the cell function.[1] Mutations in TP53 usually results in loss-of-functions and is an independent predictor of poor prognosis, especially if the mutation is located in the DNA-binding domain.[1]

However, it is a heterogenous cancer both between and within classification with a variable clinical course and prognosis.[1–3] Cytological findings are diverse and a high genetic diversity is observed.[1–3] This have impeded the advances in standardization and effective treatment.

1.1.1 Treatment of diffuse large B-cell lymphoma

The standard first-line treatment of DLBCL in Denmark is a combination regimen of cyclophosphamide, doxorubicin, vincristine, and prednisone given in combination with rituximab, a CD20 antibody (R-CHOP). The treatment strategy varies between 3 to 8 cycles of R-CHOP administered with an interval of either 14 days (R-CHOP14) or 21 days (R-CHOP21).[1,5,6] The DLBCL 5-year progression-free survival is approximately 60 %, whereas the overall survival is 65 %, when treating with R-CHOP.[1,2] 30-40% of DLBCL patients relapse and a small subset present with an R-CHOP resistant DLBCL.[1] At the moment, there is not enough evidence to recommend a specific regiment as a second line treatment of DLBCL in Denmark.[5,6] The most commonly used combination regimens are R-DHAP (dexamethasone, cytarabine, and cisplatin in combination with rituximab), R-ICE (ifosfamide, carboplatin, etoposide in combination with rituximab), and R-GDP (gemcitabine, dexamethasone, cisplatin in combination with rituximab).[5,6] In Denmark, R-ICE is recommended over R-DHAP due to the lower toxicity. Likewise, R-GDP is recommended for et beneficial improvements in a patient's quality of life, fewer hospitalizations and significantly fewer side effects, compared with R-DHAP.[5,6]

The research into the treatment of relapsed/resistant DLBCL focuses on 1) risk stratification of DLBCL patients to predict their response to R-CHOP and map the benefit from more aggressive combination regimes, 2) developing novel and less toxic treatment regiments for these patients, and 3) utilizing the molecular knowledge of DLBCL to advance the field of precision medicine in DLBCL.[1]

1.2 DNA DAMAGE REPAIR

Every day, our cells are exposed to a wide range of endogenous and environmental factors capable of introducing a variety of DNA lesions and thereby altering our DNA. These factors include e.g. carcinogens, radicals, radiation, and pharmaceuticals[7,8]. Such alterations to the DNA can be catastrophic; in worst case, it could lead to the development of malignancy or tumor progression. To limit genomic instability, caused by the aforementioned factors, cells have incorporated checkpoints into the cell cycle where the DNA damage response (DDR) is initiated if a lesion is recognized (cf. Figure 1). This lesion will either be tolerated or repaired. There are six main DDR pathways: direct repair (DR), base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), homologous recombination (HR), and non-homologous end joining (NHEJ). For simplicity, abbreviations of each protein involved in the DDR pathways is stated (cf. Abbreviations for the full form).

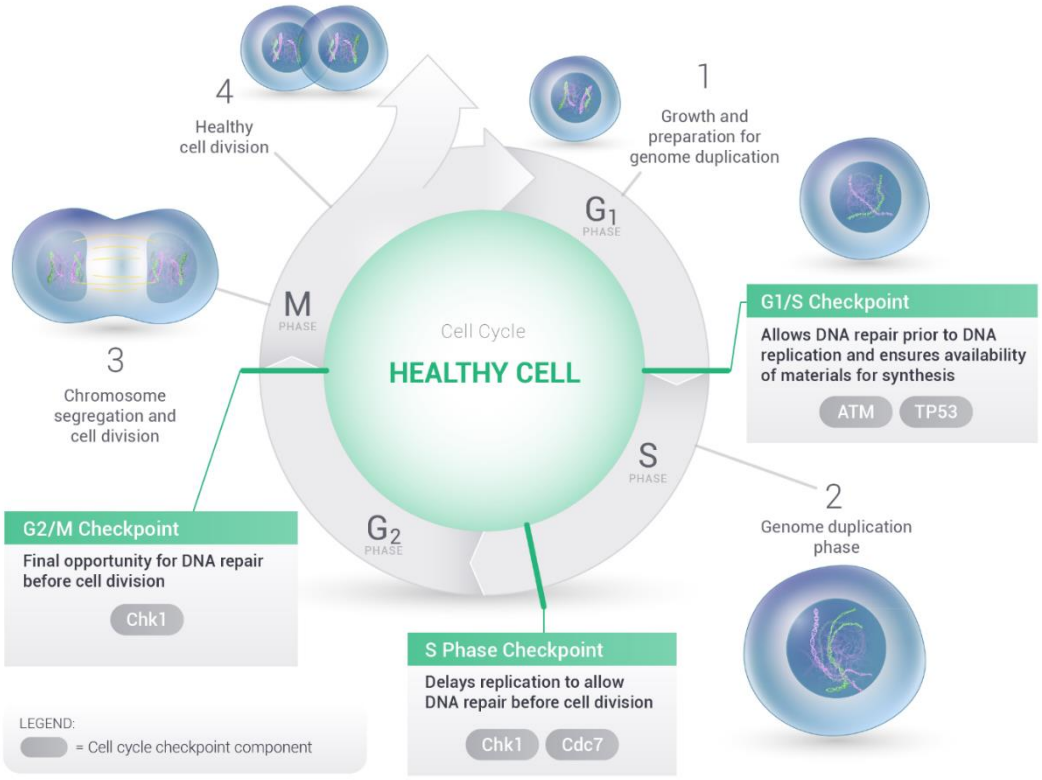


Figure 1: Overview of the four phases and three checkpoints of a healthy cell. Figure borrowed from Sierra Oncology[9].

DR is the simplest form of DNA repair; it is a direct reversal of the damage. Erroneous methylation and other forms of alkylation at the O^6-position of guanine is demethylated by O^6-methylguanine DNA methyltransferase (cf. Figure 4B).[8] Chemical modifications to a specific DNA base is repaired by BER[7] (cf. Figure 2) in the G1 phase of the cell cycle[10] (cf. Figure 1). DNA glycosylases, such as UNG and OGG1, scan the DNA by slightly pulling the nucleotide strand and thereby detecting any distortion caused by the lack of hydrogen bonds between Watson–Crick base pairs.[7] The damaged base is excised by the enzymes, creating an abasic site, also known as an apurinic- or apyrimidinic (AP) site, which is subsequently processed by APE1 that creates a single stranded break.[7,8] DNA POLβ inserts the appropriate base and DNA LIG1 or LIG3 ligate the strand.[7]

PARP1 and PARP2 facilitates the process of BER by recruiting XRCC1 and other BER proteins important for initialization.[8] The pathway is subdivided in short patch BER, where a single nucleotide is repaired, and long patch BER, where two to 13 nucleotides are repaired.[8]

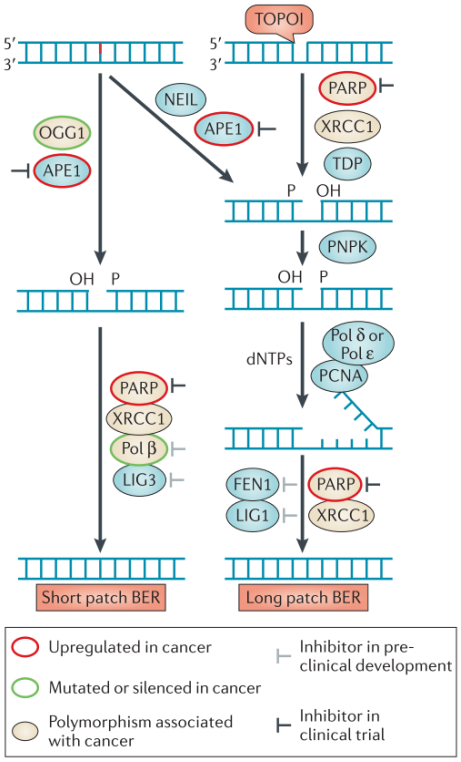


Figure 2: Schematic overview of base excision repair (BER) of chemical modifications to DNA bases with involved proteins. Illustration borrowed from Curtin (2012)[8].

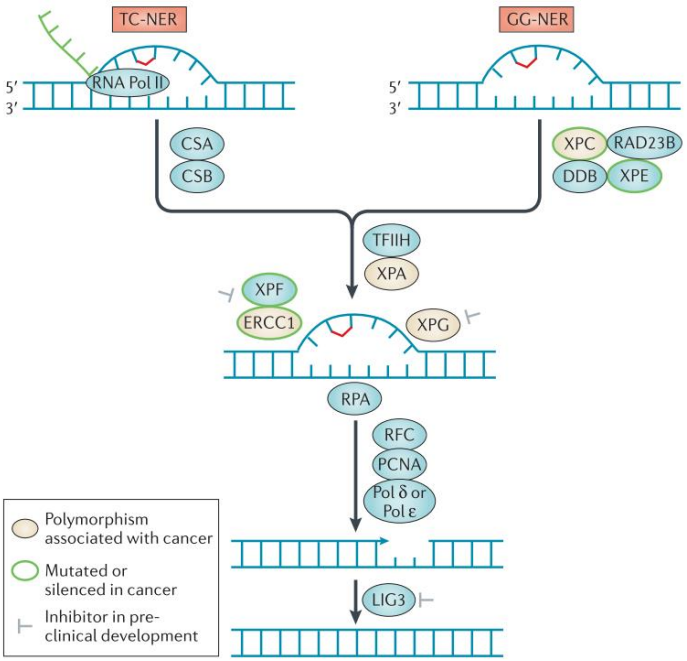


Figure 3: Schematic overview of nucleotide excision repair (NER) with involved proteins. Recognition of crosslinks is performed either through transcription-coupled nucleotide excision repair (TC-NER) in actively transcribing genes or global genome nucleotide excision repair (GG-NER) throughout the genome. Repair of the lesion occurs through a shared pathway. Illustration borrowed from Curtin (2012)[8].

Lesions that are more complex are repaired through NER. These lesions include intrastrand crosslinks and interstrand crosslinks, helix-distorting adducts, resulting in bulky formations (cf. Figure 3).[7,8] Repair of an actively transcribing gene is attempted by the subpathway transcription-coupled nucleotide excision repair (TC-NER) during the G1 phase of the cell cycle (cf. Figure 1), whereas the subpathway global genome nucleotide excision repair (GG-NER) recognizes and repairs lesions throughout the genome and independent of the point in cell cycle.[10,11] These subpathways differ in the initial steps. In TC-NER, a lesion is recognized by the blockade of RNA polymerase II elongation along with CSA and CSB, whereas a complex consisting of XPC, RAD23B, and DDB recognizes lesions in GG-NER.[8,11] Following lesion identification, the pathway is shared. The double helix is separated by the helicases XPB and XPD. ERCC1-XPF endonuclease-complex cleaves the phosphodiester bond a few nucleotides away from the lesion at the 5' end and XPG endonuclease cleaves the phosphodiester bond at the 3' end.[11] POLδ or POLε resynthesize the strand and finally, ligation is accomplished by LIG3.[8,11]

Unrepaired DNA modifications might result in DNA mismatches, where a nucleotide is paired with a non-complementary nucleotide during replication.[7,11] These single stranded DNA (ssDNA) lesions along with small insertions and deletions are recognized by MMR[7,8,11] (cf. Figure 4A) during the S phase of the cell cycle[10] (cf. Figure 1). Two heterodimers named MutSα, consisting of MSH2 and MSH6, and MutSβ, consisting of MSH2 and MSH3, can form a complex with MutL.[11] MutSα recognizes mismatches and MutSβ recognizes small insertions or deletions by sliding up and down the double stranded DNA until they encounter PCNA and RFC.[7,11] The mismatched nucleotide on the daughter strand is cleaved by EXO1 and the correct bases are incorporated by DNA polymerases (POLε and POLδ).[7,11]. The strand is ligated by LIG1.[7,8,11]

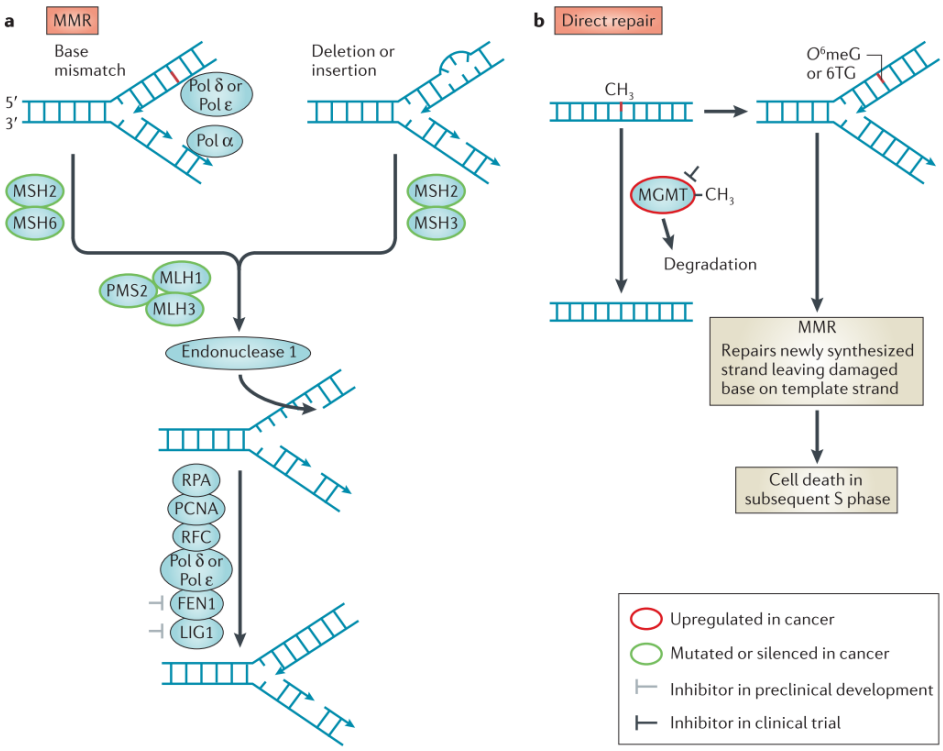


Figure 4: Schematic overview of **a**: Mismatch repair (MMR), recognizing and repairing mismatches and small insertions or deletions, and **b**: Direct repair, responsible for the direct reversal of alkylation. Illustration borrowed from Curtin (2012)[8].

Double stranded break (DSB) are the most complex DNA lesion to repair and the most cytotoxic[7,8], because the repair machinery is no longer guided by a complementary DNA strand.[11] These lesions can be repaired through two pathways: HR and NHEJ (cf. Figure 5).

HR occurs in the S and G2 phase of the cell cycle (cf. Figure 1).[7,10] The MRN nuclease complex comprising MRE11, RAD50 and NBS1, is recruited by BRCA1 and PARP1 to the DSB site, which in turn activates ATM kinase and recruits CtIP along with EXO1.[8] ATM facilitates phosphorylation of MRE11, NBS1, CtIP and EXO1, where CtIP and EXO1 subsequently resects the 3' ends.[7,8] A complex consisting of ATR and ATRIP is also recruited, which through phosphorylation of CHK1 results in S and G2 cell cycle arrest.[8] With the ATR-mediated phosphorylation of CHK1 and RPA2, RAD51 is phosphorylated. This phosphorylation along with BRCA2 leads to invasion of a homologous dsDNA and forms a Holiday junction.[8] The strand is extended by repair enzymes and ligated at the end of

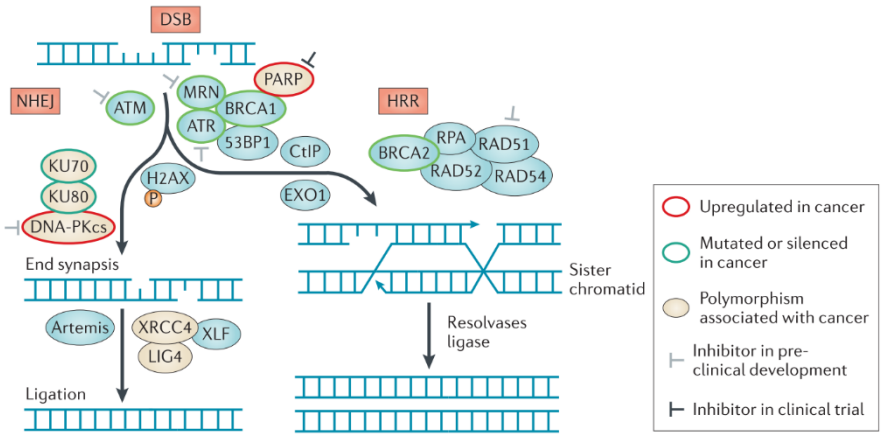


Figure 5: Schematic overview of non-homologous end joining (NHEJ) and homologous recombination (HR) repair of double stranded breaks (DSB) with involved proteins. Illustration borrowed from Curtin (2012)[8].

the DSB. The junction is resolved with or without the cross-over of genetic information.[7,8]
NHEJ is an error-prone repair mechanism mostly active in the G0 and G1 phase of the cell cycle (cf. Figure 1). It is initiated when there is no homologous dsDNA available (cf. Figure 5). Like with HR, the MRN complex is recruited to the site of DSB.[8] Subsequently the heterodimer comprising Ku70 (XRCC6) and Ku80 (XRCC5), and DNA-PKcs bind to the DNA and forms a complex with Artemis.[7,8,11] The DNA-PKcs-Artemis complex dissociates as a consequence of phosphorylation, and acts as an endonuclease, processing the DNA ends.[8,11] Then, LIG4 joins the phosphodiester backbone of the DNA ends.[7,11] This process is stabilized by a complex of XRCC4 and XLF.[7,8] POLμ and POLλ completes the resynthesis.[11]

1.3 CISPLATIN MECHANISM OF ACTION

Common for each second line treatment of DLBCL recommended in Denmark is a platinum-based combination regiment. Cisplatin and carboplatin have a similar mechanism of action; however, carboplatin differs from cisplatin in structure and toxicity. For simplicity, cisplatin will be the focus of this project. Cisplatin is a small platinum-based compound which primarily targets DNA, more specifically the N^7-position of purines.[11,12] Here, it covalently binds to mostly guanine and forms 1,2- or 1,3-intrastrand crosslinks and a few number of interstrand crosslinks.[12] This leads to the distortion of the double helix, which is registered by a wide range of proteins, including DNA repair proteins.[12,13] DNA repair plays a key role in mediating cisplatin cytotoxcicity.[12] Cisplatin induces DNA damage, which is registered at the cell-cycle check points.[12] Cisplatin is thought to be cell-cycle non-specific[14], however, it has been found to induce G2 phase arrest, where lesions of the G2 phase and late S phase is repaired.[12]
NER is the major pathway that is activated by cisplatin.[11,12] Here, the crosslink of DNA by cisplatin is removed through recognition, excision of 24-32 oligonucleotides containing the lesion, resynthesis, and ligation.[11,12] This induces cell-cycle arrest, ensuring no replication of the tumor cell.[5] Additionally, the large number of cisplatin-induced DNA crosslinks signals apoptosis of the cancerous cell.[5]
Binding of cisplatin to bases can cause mismatches, most often guanine-thymine mismatches.[11,12] This initiates MMR with recognition, excision of the wrong base, elongation and ligation.[11,12] The 'problem' is that the mismatched nucleotide that is excised is on the daughter strand, leaving the source intact.[11,12] So when elongation is finished, another mismatch will be present, re-initiating the MMR pathway. This creates a futile loop of repair, which retains the cell-cycle arrest.[11,12] The loop, and the possibly consequential DSBs, signals apoptosis.[5] Likewise, the many DNA-strand segments, cut off by EXO1, also signals apoptosis.[15] Furthermore, MMR proteins have been proposed to inhibit NER proteins, thereby impairing the repair of crosslinks and increasing cisplatin lethality.[5]
So basically: cisplatin takes advantage of the cells' natural DNA repair system to halter replication and eventually lead to apoptosis and tumor reduction. This affects all cell-cycle active cells of the body, however, to a greater extent fast proliferating cells, such as cancer cells.

1.3.1 DNA repair defects affecting nucleotide excision repair and mismatch repair

Cisplatin is a drug with two response poles. It is found to be very effective in testicular cancer and solid tumors.[11] However, resistance is still a big problem in these and other cancers: both intrinsic and acquired resistance is common.[11,13]

An increased capacity for NER and cisplatin-adduct repair could lead to cisplatin resistance.[16,17] Whereas, low expression of any key NER protein, such as XPD, XPF, XPG and ERCC1, could lead to decreased capacity of NER, increasing sensitivity to cisplatin and cell death.[11,13] A low expression of ERCC1-XPF is associated with increased cisplatin sensitivity of testicular cancer cells and a better survival in patients with cisplatin-treated non-small cell lung cancer (NSCLC), proposing the ERCC1-XPF complex as a limiting factor of cisplatin response.[17,18]

The MMR pathway also plays a role in mediating cisplatin's cytotoxicity. B-cell lymphomas often present with recurrent alterations in MMR-related genes, such as *EXO1*, *MSH2* and *MSH6*.[19] An impaired MMR pathway could lead to cisplatin resistance, where unrepaired base pair mismatches accumulate and lead to microsatellite instability.[11,13] *MSH2* or *MLH1* deficiency leads to decreased apoptosis of cancer cells and lower survival of patients in some studies, and a prolonged survival of patients in others. These conflicting results necessitates further studies into the involvement of *MSH2* in cisplatin resistance. Additionally, the ERCC1-XPF complex seems to play a role in the mechanism of cisplatin resistance.[20–23]

1.4 Combination treatment with inhibitors of DNA repair

The pressing problem with cisplatin resistance has led to the research into combination treatment with chemotherapy and inhibitors of DNA repair.

17-(Allylamino)-17-demethoxygeldanamycin (17AAG) is an inhibitor of heat shock protein 90 (Hsp90).[24,25] Hsp90 controls the folding of proteins, particularly those involved in signal transduction and cell cycle progression and is associated with proteins of DNA repair.[24–26] 17AAG inhibits Hsp90 by binding to the ATP interaction pocket in the Hsp90 N-terminal domain.[24,25] Additionally, 17AAG downregulates the expression of MSH2 and HR related renes (BRCA1 and RAD51), and consequently affect cells' response to chemotherapy.[24,25] At present, 53 clinical trials are listed on ClinicalTrials.gov[27] with the subjects of 17AAG within cancer research, none is currently active. The focus spans evaluation of its toxicity as a monotherapy, with multiple phase 1 and 2 trials describing a tolerable toxicity with a moderate to no effect on the cancer, to elucidating its effect in combination with chemotherapy, where phase 1 and 2 trials describe an acceptable toxicity but a conflicting tumor response.[27,28]

Another target for chemoresistance is the ERCC1-XPF complex. Polyphenol epigallocatechin-3-gallate (EGCG), a natural compound found in green tea, is a partially reversible inhibitor of ERCC1/XPF activity in vitro and affects interstrand crosslink repair, leading to increased cisplatin sensitivity.[29] Likewise, combination treatment results in a significant tumor reduction in tumor xenografts in vivo.[29] There are currently 34 clinical trials utilizing EGCG in cancer on ClinicalTrials.gov[27], of which six are active. The application of EGCG greatly varies and spans chemoprevention, monotherapy, and combination treatment with chemotherapy.[27] Phase 1 trials show EGCG to

be well tolerated with low toxicity and phase 2 trials report EGCG as a significant chemosensitizer.[27,30]

1.5 AIM AND HYPOTHESES

Although the distinction between GCB and ABC, their respective prognosis, and their molecular differences, DLBCL is still highly heterogeneous; patients present with an individual course of disease and 30-40% of patients relapse. Cisplatin, one of the chemotherapies administered as second line, presents a greatly variating response in the clinic: some patients being highly sensitive, others resistant or only partially sensitive. Treatment resistance is a huge challenge that complicates and prolongs treatment regimes, leads to a decreased Quality of Life for patients, and increases the risk of death. One explanation for resistance is affected DNA repair, specifically increased NER and decreased MMR in relation to cisplatin resistance. DNA repair pathways hold a great potential for inducing chemosensitivity, where compounds such as EGCG and 17AAG is being investigated in different applications and cancers. 17AAG decreases the expression of MSH2 but displays conflicting effects in combination treatment and EGCG shows great promise as a chemosensitizer but sparse research has been made in other than solid tumors.

Therefore, the overall aim of this thesis was to investigate the DNA repair related mechanism behind cisplatin resistance, with the following two research questions:

- How does the DNA repair related mutation profile of cisplatin-sensitive DLBCL compare to cisplatin-resistant DLBCLs?
- How does impaired DNA repair affect the cisplatin response and is it possible to manipulate the cisplatin response of DLBCL cell lines?

The research hypotheses comprised:

- Cisplatin-resistant DLBCLs present with mutations in MMR-related genes and none in other DNA repair mechanisms, whereas cisplatin-sensitive DLBCLs show mutations in DNA repair genes, especially NER-related genes.
- Affected NER increases cisplatin-sensitivity, whereas affected MMR decreases cisplatin-sensitivity.
- The addition of EGCG sensitize DLBCL cells to cisplatin, whereas 17AAG decrease cisplatin-sensitivity of DLBCL cells.

The research objectives included:

- To investigate the occurrence of mutations in selected DNA repair genes in DLBCL cell lines, defined as cisplatin-sensitive or -resistant, and clinical DLBCL samples.
- To investigate the effect of DNA repair inhibitors on the cisplatin response and the degree of generated DNA fragments (as a measure of DNA damage) in DLBCL cell lines.

2 PREVIOUS RESULTS OF IMPORTANCE IN THIS THESIS

2.1 CISPLATIN DOSE-RESPONSE SCREENING OF CELL LINES

Prior to initiation of this project, a systematic dose-response screening using cisplatin was carried out in our laboratory on a panel of DLBCL and multiple myeloma cell lines as described previously.[31] Cells at an optimized seeding concentration were transferred to 96-well plates, cultivated for 24 hours, and exposed to cisplatin as a 2-fold dilution with 18 different concentration. Cell growth was evaluated using CellTiter 96® MTS (3-[4,5-dimethylthiazol-2-yl]-5-[3-carboxymethoxyphenyl]-2-[4-sulfophenyl]-2H-tetrazolium) assay (CellTiter 96 Aqueous One Solution Reagent, Promega, Madison, WI) at 0 hours and 48 hours post drug application. Absorbance was measured by Optima Fluostar plate reader (BMG LABTECH, Ortenberg, Germany) at an absorbance of 492 nm. The dose-response curves are depicted in Figure 6A.

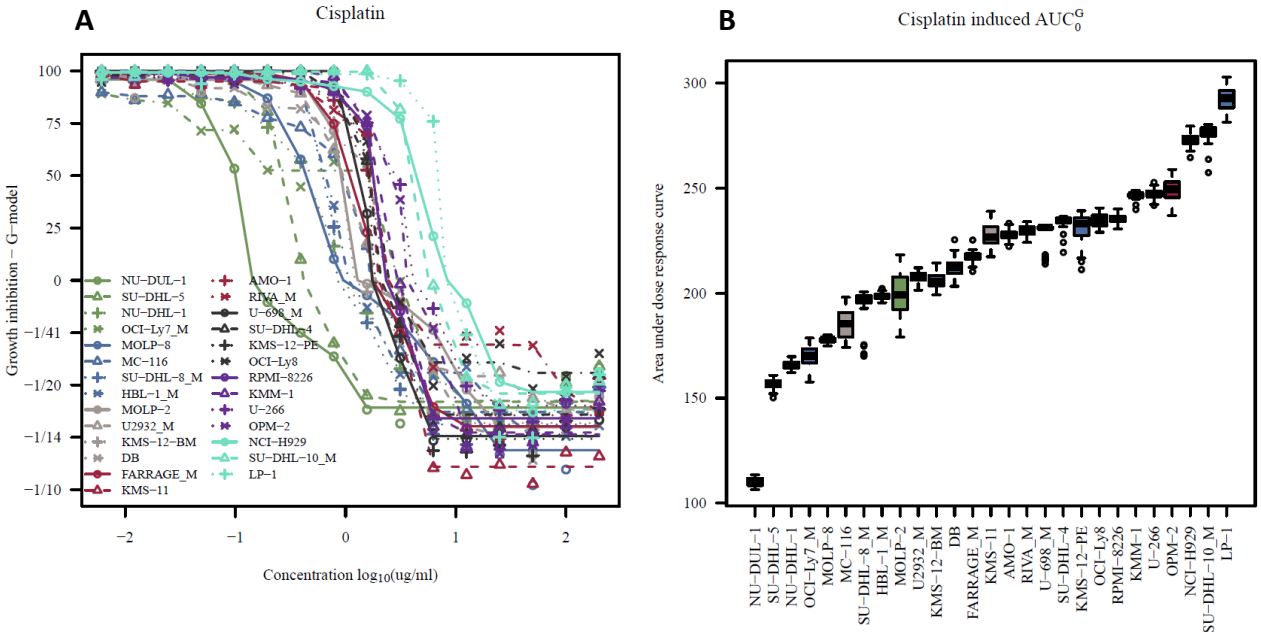


Figure 6: A systematic dose-response screening of cell lines using cisplatin (in-house graphs). Results are depicted as **A)** a dose-response curve for each cell line, and **B)** boxplots of stratified individual area under curve (AUC)-values of the cisplatin induces dose-response curve.

Descriptive values were calculated using the dose-response G-model, a model independent of exposure time and cell lines' individual doubling time.[32] These included GI_{50}, describing the concentration of cisplatin that resulted in 50% growth inhibition at 48 hours, TGI, denoting the concentration of cisplatin at which a total growth inhibition is observed, LC_{48}, translating to the concentration of cisplatin that is lethal to 50% of the cells at 48 hours, and AUC_0, meaning the area of cell growth between the dose-response curve and a line at growth 0.[32,33] Each of the values have a coupled 95% confidence interval (CI). GI_{50} have also been converted from $\log_{10}(\mu g/mL)$ to µg/mL. Results of the G Model are listed in Appendix 1: Dose-Response Screening.

In the present study, these results were used to stratify the 14 DLBCL cell lines according to AUC_0 and subsequently divide them into Sensitive and Resistant, like observed in Figure 6B. AUC_0 was used instead of GI_{50}, because AUC_0 provides the cumulative effect of cisplatin in the given cell line, with doubling time taken into account, and is not dependent on only the point on the curve, where 50% growth inhibition is obtained.

2.2 1,001 DE NOVO DLBCL PATIENT COHORT

Reddy et al. (2017)[34] investigated the impact of genetic drivers of DLBCL on functional and clinical outcomes in a cohort of 1,001 DLBCL patients.[34] Patients were required to be de novo (no relapse or transformed DLBCL) and treated with rituximab-containing standard regimen.[34] Clinical information was obtained to nearly every patient, comprising gender, age, DLBCL stage, performance status, degree of extranodal involvement, initial response to therapy and overall survival.[34] Included was also 400 paired-normal tissue.[34] Through whole-exome sequencing, joint variant calling and somatic variant calling, ca. 1.1 million mutation variants were identified in the cohort, and the accuracy was verified through Sanger sequencing.[34] Using three tools (MutsigCV, Hotnet2 and their own) 150 driver genes were identified by evaluation against background mutation rate, identifying groups of genes by protein interactions, and factoring in copy number variations, gene variability and functional variants.[34] An interactive webtool, consisting of patients mutational profile, risk modelling (Kaplan-Meier survival) and clinical data was made available at https://dlbcl.davelab.org[35]. The variants were filtered before being included in the interactive webtool. The variant had to be in an exon, it had to be somatic variant, and it had to be a rare and non-synonymous damaging variant.[34] The variant must not be in a repetitive sequence or in a region with poor coverage: the reading depth had to be >5.[34] The genome quality score had to be >30.[34] The gene had to be affected in ≥20 patients, to perform survival analysis.[34]
Of interest in this thesis, only seven DNA repair related genes remained post variant filtering (*ATM*, *ATR*, *BTG1*, *MSH2*, *MSH6*, *TP53* and *UBE2A*).[35]

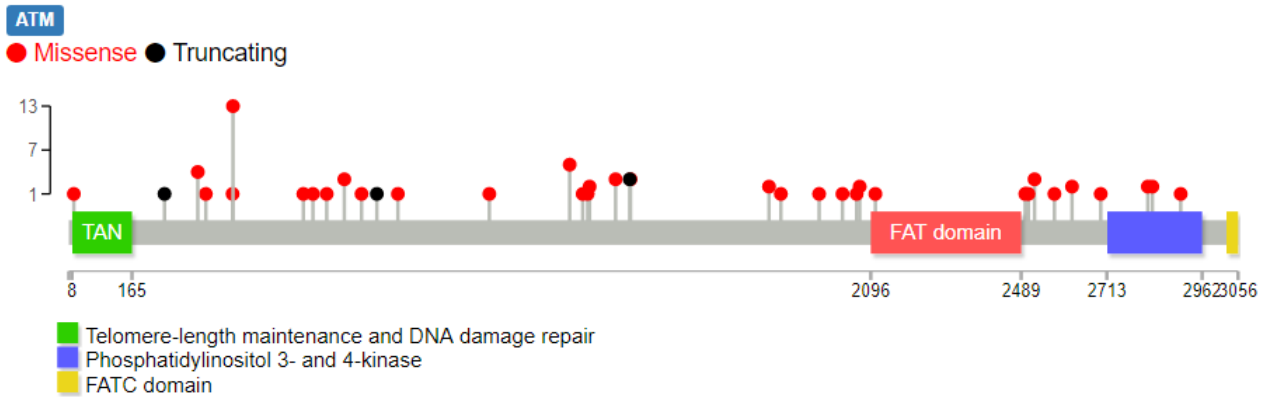


Figure 7: Plotted protein change variants of *ATM* found trough whole-exome sequencing of 1,001 de novo DLBCL patients by Reddy et al. (2017)[34]. Protein position and domains are visualized on the horizontal axis. Prevalence in the cohort (n=1,001) is indicated on the vertical axis. Black heads show truncating variants. Red heads show missense variants. Copied from https://dlbcl.davelab.org[35].

ATM is recruited to DSBs, e.g. generated through continuous repair of cisplatin-related lesions, along with the MRN complex and is involved in G1 phase arrest.[7,8,11,36] ATM deficient cells present with impairment of the G1-, Intra-S-, and G2-checkpoint for DNA damage.[36] Few *ATM* variants have previously been described in relation to DLBCL.[19,37] Likewise, the protein is found to be up-regulated in cisplatin-resistant NSCLC, whereto inhibition increase cisplatin sensitivity.[38]

Data Set	Sample Size	3-Year Survival Rate	0.95 Confidence Interval
Affected Patients	72	78.76 %	[66.61%, 86.91%]
Remaining Patients	864	73.07 %	[69.80%, 76.06%]

p-value: 0.157

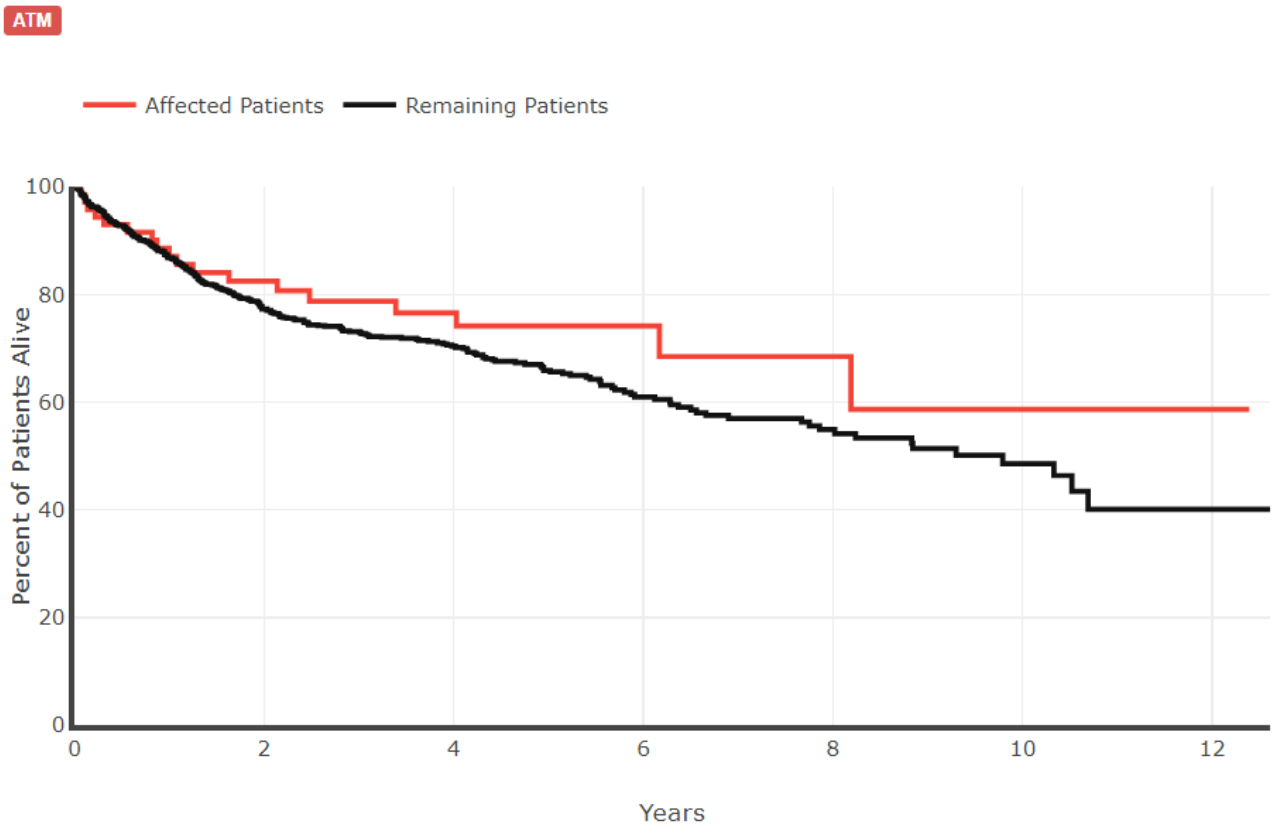


Figure 8: Survival Estimate of de novo DLBCL patients with (red) and without (black) *ATM* mutation (n=1,001), performed by Reddy et al. (2017)[34]. The table shows the survival statistics, and below is the Kaplan-Meier survival curve.[34] Copied from https://dlbcl.davelab.org.[35]

Reddy et al. (2017)[34] identified 72 patients in their cohort of 1,001 de novo DLBCL to have a mutation in *ATM*. The *ATM* variants were plotted on a protein view and published on https://dlbcl.davelab.org[35]. From this website, Figure 7 originates. No hotspots for mutation can be identified in the protein. In fact, the highlighted protein domains seem unaffected. The survival estimate of patients with *ATM* mutations and the patients without *ATM* mutations are shown in Figure 8.[34] A better survival can be observed for patients with an *ATM* mutation, however, there

was no significant difference (p=0.157) in the 3-year survival between the two datasets. The increased survival corresponds with the empiri, where impaired ATM increase chemosensitivity.

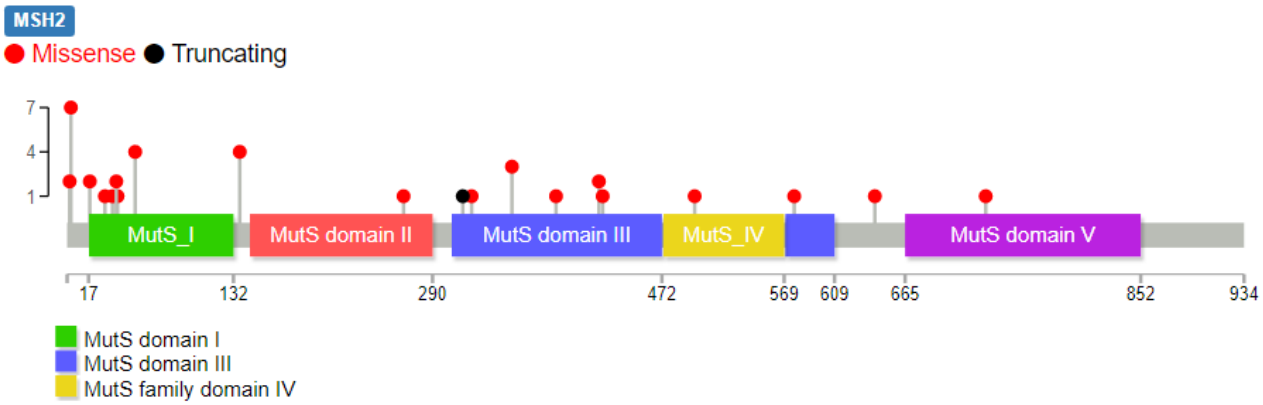


Figure 9: Plotted protein change variants of *MSH2* found trough whole-exome sequencing of 1,001 de novo DLBCL patients by Reddy et al. (2017)[34]. Protein position and domains are visualized on the horizontal axis. Prevalence in the cohort (n=1,001) is indicated on the vertical axis. Black heads show truncating variants. Red heads show missense variants. Copied from https://dlbcl.davelab.org.[35]

MSH2 (MutS protein homolog 2) is a key protein involved in the recognition of DNA damage in MMR.[8,13] Mismatched bases are recognized by MSH2-MSH6 heterodimers, whereas deletion and insertions are recognized by MSH2-MSH3 heterodimers.[8] Reddy et al. (2017)[34] found 36 patients with an *MSH2* mutation on their cohort. *MSH2* variants are plotted on the protein in Figure 9, copied from https://dlbcl.davelab.org[35]. Like with *ATM*, no hotspot is evident. However, most variants are in a protein domain. *MSH2* (p.A54T) was found in four patients.[34] It is located in the MutS_I domain, the DNA-binding domain[39,40] and described to have an uncertain significance.[41] *MSH2* (p.Q353K), found in three patients[34], is located in MutS_III domain, also known as the Lever domain, which folds in two areas to form a lever for MutS_V domain.[39,40] The domain is not described in relation to mutation-induced *MSH2* impairment.[39,40] *MSH2* (p.P729T), found in one patient[34], is located in MutS_V domain, also known as the ATPase domain.[39,40] The MutS_V domain is often mentioned as somewhere a mutation could have large effect of the proteins function and thereby on the function of MMR.[39,40] With impaired MMR, mismatched nucleotides accumulate and microsatellite instability might follow.[11] An endometrial cancer cell line (HEC59), deficient in MSH2, had a 1.8-fold increase in cisplatin resistance.[11] This corresponds well with Reddy et al. (2017) reporting decreased survival of patients with *MSH2* mutation, as seen in Figure 10.[34] Although not significant (p=0.339), the lower 3-year survival rate is indicative of increased chemoresistance in patients with *MSH2* mutations, compared to patients without mutaiton.[34]

Data Set	Sample Size	3-Year Survival Rate	0.95 Confidence Interval
Affected Patients	36	63.30 %	[44.30%, 77.35%]
Remaining Patients	900	73.90 %	[70.72%, 76.79%]

p-value: 0.339

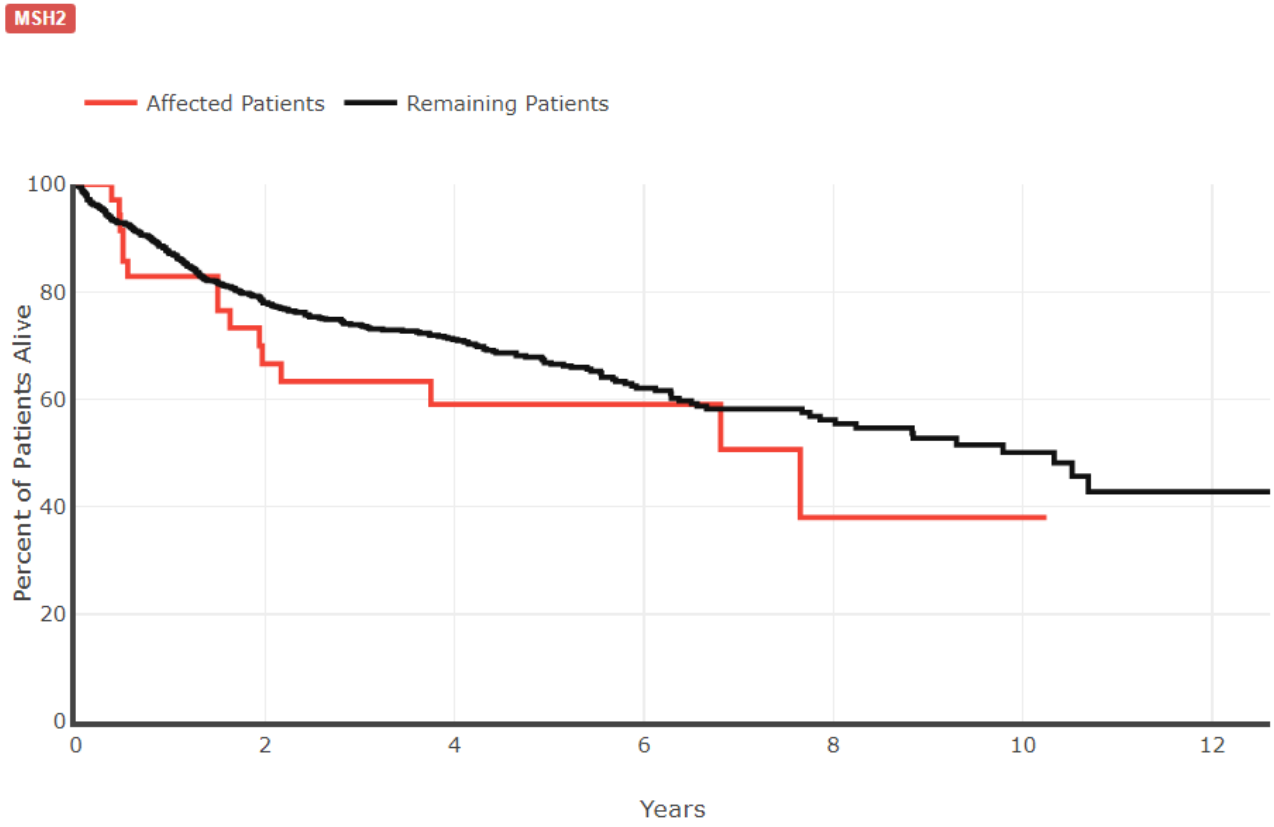


Figure 10: Survival Estimate of de novo DLBCL patients with (red) and without (black) *MSH2* mutation (n=1,001), performed by Reddy et al. (2017)[34]. The table shows the survival statistics, and below is the Kaplan-Meier survival curve.[34] Copied from https://dlbcl.davelab.org.[35]

The filtered annotated variants, along with patients' clinical information is available as Supplementary Information "Table S1" to Reddy et al. (2017).[34] The protein variant plots and survival analyses of the five remaining DNA repair related genes included in the webtool, but not mentioned here, can be found in Appendix 2: Reddy et al. (2017) Survival Estimate.
Based on variants published in the interactive webtool (https://dlbcl.davelab.org)[35], a specific variant was chosen for further analysis in the present study. Additionally, raw in silico data from the whole-exome sequencing of 1,001 de novo DLBCL patients, provided by DaveLab[34], was analyzed using an in-house pipeline and used to evaluate the mutation burden in clinical DLBCL through graphical protein variant views of specific genes.

2.3 72 Hematology Aalborg DLBCL patient cohort

A cutting-edge project is under way at the Department of Hematology, Aalborg University Hospital. Researchers, Bioinformaticians, Datamanegers, Clinical Research Nurses, Professors and other

staff are working on developing a hematological precision medicine pipeline, which will simplify the identification of cancer-driving molecular variants. Included in this project are sequencing data from three projects: RetroGen (N-20140099), ProGen (N-20150042) and ProSeq (N-20160089). ProGen and ProSeq are still including patients. RetroGen is a retrospective study that encompasses already obtained tumor samples from 73 DLBCL patients (≥18 years) from the Department of Hematology, Aalborg University Hospital. Most of the patients (n=69) had received standard R-CHOP treatment. The majority were de novo DLBCL at time of sample collection, few were in a relapse situation. More specific patient characteristics can be found in the Supplementary Table 2 of Due et al. (2019)[42]. Patients from the Department of Hematology, Aalborg University Hospital, with suspected relapse disease progression of hematological cancer is offered participation in ProGen. After relapse/progression verification, the patients are enrolled into ProSeq. These are prospective non-interventional studies. Patients must be ≥18 years and with no record of relapse treatment at time of inclusion. In ProGen, either blood, bone marrow (BM) or lymph node biopsy (determined by simultaneous clinical tests) is donated, whereas saliva or a mouth swab is donated in ProSeq. Biological samples are processed and analyzed using whole-exome sequencing. Somatic variants are detected using Mutect2 and varscan. Variants annotated “PASS” (Mutect2) and “high confidence” (varscan) are filtered according to 1) a QSS (quality score for somatic single nucleotide variants) score of ≥25, 2) an allele frequency of ≥0.02, and 3) an allele ratio (tumor:normal) of ≥4. Depending on the use, these datasets were subsequently filtered accordingly.
The information on RetroGen, ProGen and ProSeq comes from a personal communication with J.S. Bødker and Bødker et al. in preparation.
In the present study 51 of the RetroGen- and 21 of the ProSeq-included in silico patient sequencing data, was used, resulting in a cohort of 72 Hematology Aalborg DLBCL patients, of which 28 are from patients with relapse DLBCL and 44 are from patients with de novo DLBCL.

3 METHODS

The study sought to investigate the occurrence of mutations in selected DNA repair genes in DLBCL cell lines, defined as cisplatin-sensitive or -resistant, and clinical DLBCL samples, as well as examine the effect of DNA repair inhibitors on the cisplatin response and the degree of generated DNA fragments (as a measure of DNA damage) in DLBCL cell lines.

3.1 SAMPLES INCLUDED IN THE PROJECT

Fourteen cell lines of human DLBCL were included in the study. They are listed in Table 1, along with disease, origin, Epstein-Barr virus status, subtype, sex of the patient, and year of establishment. All originate from patients with DLBCL and were established between 1975 to 1990. One is from a patient with a secondary DLBCL and two are from relapsed DLBCL patients. Five is of the ABC subtype, nine are GCB subtype. The cell lines were established from a range of different sources: some from lymph nodes others from fluids, such as pleural effusion, peripheral blood, or ascites. There are ten from male and four from female patients.

Table 1: Information on the 14 cell lines included in this study. EBV: Epstein-Barr Virus, ABC: activated B-cell like, GCB: germinal center B-cell like

Cell line	Established from…	Disease	EBV +/-	Subtype	Sex	Year
DB	Ascites	Human B-cell lymphoma	EBV-	GCB	Male	~1988
FARAGE	Lymph node	Human B-cell lymphoma	EBV+	GCB	Female	~1990
HBL-1	Pleural effusion	Human B-cell lymphoma	EBV-	ABC	Male	1984
MC-116	Pleural effusion	Human B-cell lymphoma	EBV-	GCB	Male	1980
NU-DHL-1	Inguinal lymph node	Human B-cell lymphoma	EBV-	ABC	Male	1982
NU-DUL-1	Cerebrospinal fluid	Human B-cell lymphoma	EBV-	ABC	Male	1982
OCI-Ly7	Peripheral blood	Relapsed human B-cell lymphoma	EBV-	GCB	Male	1987
OCI-Ly8	Bone marrow and peripheral blood	Relapsed human B-cell lymphoma	EBV-	GCB	Male	1987

RIVA	Peripheral blood	Human B-cell lymphoma	EBV-	ABC	Female	1977
SU-DHL-10	Pleural effusion	Human B-cell lymphoma	EBV-	GCB	Male	1975
SU-DHL-4	Ascites	Human B-cell lymphoma	EBV-	GCB	Male	1975
SU-DHL-5	Lymph node	Human B-cell lymphoma	EBV-	GCB	Female	1975
SU-DHL-8	Pleural effusion	Human B-cell lymphoma	EBV-	GCB	Male	1975
U-2932	Ascites	Secondary human B-cell lymphoma	EBV-	ABC	Female	1996

Included in the project is also clinical samples, consisting of in silico data from whole exome sequencing of 1,001 de novo DLBCL patients, kindly provided by DaveLab[34] , Durham, USA, and whole exome sequencing of 21 ProSeq and 51 RetroGen DLBCL patients (in-house, gathered as of the 5th of January 2019) (cf. Table 2). DaveLab's 1,001 de novo DLBCL patients received a rituximab-containing standard regimen. 765 (76%) patients had a complete response to therapy, 88 (9%) patients had a partial response, and 69 (7%) patients showed no response, the therapy response for 72 (7%) patients were not available. Most of DaveLab's' de novo DLBCL patients have a GCB subtype, of those that have been classified.

The 72 Hematology Aalborg DLBCL patients consisted of 28 relapsed DLBCL and 44 de novo DLBCL, collected as part of three protocols: RetroGen (N-20140099), ProGen (N-20150042) and ProSeq (N-20160089). Most of Hematology Aalborg 72 DLBCL patients show an ABC subtype, of those that have been classified.

Table 2: The sample population of this project consisted of three groups: 14 DLBCL cell lines, patient data from 1,001 de novo DLBCL patients including DNA sequencing, provided by DaveLab[34], Durham, USA, and DNA sequencing from 72 de novo or relapsed DLBCL patients (in-house data). NC: non-classified, NA: not available.

	Cell lines	**DaveLab**[34]	**Hematology Aalborg**
Disease	DLBCL	De novo DLBCL	28 relapsed DLBCL, 44 de novo DLBCL
n=	14	1,001	72
ABC and GCB subtype	ABC: 5 (36%) GCB: 9 (64%) NC: 0 NA: 0	ABC: 313 (31%) GCB: 331 (33%) NC: 131 (13%) NA: 226 (23%)	ABC: 30 (41%) GCB: 20 (28%) NC: 20 (28%) NA: 2 (3%)

3.2 DNA REPAIR RELATED MUTATION PROFILE

To evaluate the DNA repair gene mutation profile of the 14 cell lines, DaveLab's 1,001 de novo DLBCL patients, and 72 de novo and relapsed DLBCL Hematology Aalborg patients, a comprehensive list of DNA repair genes was made by a merge of Wood, Mitchell, Sgouros, and Lindahl (2001)[43] and a list from MDAndersen Cancer Center[44], last updated in 2014, as well as single genes found through searches on https://www.genecards.org/. The DNA repair genes were searched for each of the 14 cell lines on the Cancer Cell Line Encyclopedia (CCLE), a database from Broad Institute, compiling sequencing data, gene expression, and chromosomal copy number from 1,457 human cancer cell lines.[45] The identification of variants in DaveLab's 1,001 de novo DLBCL patients and 72 Hematology Aalborg DLBCL patients is described in later sections.
The reference genome is GRCh37.

3.3 GENE MUTATION VARIANTS

Interesting gene variants were chosen from cell lines' mutation profile, found on the CCLE database[45], and the mutation profile of 1,001 de novo DLBCL patients from DaveLab[34], found on https://dlbcl.davelab.org[35], both filtered using the list of DNA repair genes. A quantitative assessment tool was developed to provide a systematic method for the variant selection (cf. Table 3).

Table 3: The eight assessment criteria used for the quantitative evaluation of variants found in the 14 DLBCL cell lines and Davelab's[34] 1,001 de novo DLBCL patients. A weighted score was attributed to each variant within each aspect. A weighted mean was calculated.

	Weight	Score: 5	Score: 4	Score: 3	Score: 2	Score: 1
Gene function in DNA repair	x1	Key	Important	Support		Unknown
DNA repair mechanism	x2	NER	MMR	HR	NHEJ	BER and DR
Hallmark gene	x0.5	YES	-	-	-	NO
Previous annotations: gene	x0.5	YES: damaging	YES: probably damaging	YES: vaguely described	YES: unknown effect	NO
Previous annotation: variant	x3	YES: damaging	YES: probably damaging	YES: (probably) benign	YES: unknown effect	NO
Type of mutation variant	x2	VEP impact: HIGH	VEP impact: MODERATE	-	VEP impact: LOW	VEP impact: MODIFIER
Location in protein	x3	Important domain	Less important domain	Domain proximity (±5)	Domain proximity (±20)	No domain
Sample prevalence	x3	>4 %	>3 % ≤4 %	>2 % ≤3 %	>1 % ≤2 %	≤1 %

Each variant was evaluated in regard to eight assessment criteria and assigned a weighted score between 5 to 1. These eight assessment criteria included: 1) the general function of the gene, 2) which DNA repair mechanisms it plays a role in and the importance in relation to cisplatin's mechanism of action, 3) whether the gene was included in the Hallmark DNA Repair Gene Set[46,47], 4) any previous annotations of the gene, 5) any previous annotations of the variant, 6) which type of mutation the variant is (ranked according to VEP Impact by Ensembl[48]), 7) the location of the variant in the protein, and 8) the prevalence in the respective samples.

Table 4: Conversion table for the weighted mean, calculated to each variant from the scores assigned to the eight aspects of the quantitative assessment.

Weighted average		Final score
≤3.4	=	1
>3.4 ≤4.9	=	2
>4.9 ≤6.4	=	3
>6.4 ≤7.9	=	4
>7.9	=	5

Table 3 describes the aspects within each assessment criteria that gives a score from 5 to 1 (5 being the best), and the weight of the respective scores are listed for each assessment criteria. A weighted mean was calculated from the assigned scores. The weighted mean was translated into a final score for the specific variant (cf. Table 4), ranging from 5 to 1 (5: highly interesting, 1: not interesting). Based on these final scores, variants for further research in this project was chosen. If there are multiple mutation variants in the highest final score, the variant with the highest weighted mean was chosen.

3.4 AllPrep DNA extraction

From all cell lines, DNA was purified using AllPrep® DNA/RNA/miRNA Universal Kit (Qiagen®, no. 80224), Buffers were prepared as stated in the kit protocol. Buffer RLT Plus was added to thawed pelleted cells, with a concentration of 5-10×10^6, and the lysate was passed through a 20-gouge needle, fitted to an RNase-Free Syringe, 5-7 times. The homogenized lysate was added to an AllPrep® DNA Mini Spin Column fitted in a collection tube and centrifuged for 30 seconds at full speed, repeating until all lysate had passed through the membrane. Buffer AW1 was added and the column was centrifuged for 15 seconds at full speed. Leftover RNA was digested by adding a solution of Proteinase K diluted 1:4 in Buffer AW1 onto the membrane and letting it incubate for 5 minutes at room temperature. Buffer AW1 was added to the column before centrifuging at 15 seconds at full speed. A double wash was performed with Buffer AW2 and centrifugation for 2 minutes at full speed. To elute DNA, Buffer EB was added directly to the AllPrep® DNA Mini Spin Column membrane, incubated for 1 minutes at room temperature, and subsequently centrifuged for 1 minute 10,000 *g*. The elution was repeated using the flow-through.

3.4.1 Qubit™ dsDNA Broad Range Assay

Qubit™ dsDNA Broad Range Assay Kit (Invitrogen™, no. Q32853) (assay range: 100 pg/µL-1,000 ng/µL) was used to determine the DNA concentration. Samples with DNA concentrations above

1,000 ng/µL were diluted. The dilution was accounted for in the analysis. The system was calibrated using two Standards specific for DNA analysis. 10 µL of Standard or 2 µL of sample was mixed with Qubit™ reagent diluted 1:200 in Qubit™ buffer, resulting in a 200 µL working solution, which incubated for 2 minutes prior to analysis in Qubit™ 3 Fluorometer (Invitrogen™).

3.5 Droplet Digital™ PCR

Droplet Digital™ Polymerase Chain Reaction (ddPCR™) is a method with reduced error rates and reduced signal-to-noise, compared with conventional PCR, which was beneficial in this assay with a rare mutation variant. By generating a water-oil emulsion of the sample, PCR amplification is completed and analyzed in each droplet, increasing sensitivity and statistical power. Two assays, one targeting *XPF* (p.A596V) and one targeting *MSH2* (p.A54T), were planned. At the time of writing this paper, ddPCR™ Rare Mutation Detection Assay (FAM+HEX) TaqMan primers-probe mix for *MSH2* (p.A54T) has been ordered, however *MSH2* (p.A54T) ddPCR™ Rare Mutation Detection Assay has not been run.

3.5.1 Primers

ddPCR™ Rare Mutation Detection Assay (FAM+HEX) TaqMan primers-probe mix for *XPF* (p.A596V) was designed and ordered from Bio-Rad (no. 10049047), containing both HEX-targeted wild-type allele (C) and FAM-targeted mutant target allele (T), with the sequence: 5'-AGATACGTGGTTCTTTATGACGCAGAGCTAACCTTTGTTCGGCAGCTTGAAATTTACAGGG[C/T]GAGTAGGCCTGGGAAACCTCTGAGGCAAGTTATAAAGAATCACAGCTTTCAGTTGCACAGT-3'.

Likewise, ddPCR™ Mutation Detection Assay (FAM+HEX) TaqMan primers-probe mix for *MSH2* (p.A54T) was designed and ordered from Bio-Rad (no. 10049047). It contained both wild-type allele (G), targeted by a HEX probe, and mutant target allele (A), targeted by a FAM probe. They were designed based on the following sequence: 5'-AGTGCGCCTTTTCGACCGGGGCGACTTCTATACGGCGCACGGCGAGGACGCGCTGCTGGCC[G/A]CCCGGGAGGTGTTCAAGACCCAGGGGGTGATCAAGTACATGGGGCCGGCAGGTGAGGGCCG-3'. A positive control was needed, since no cell line presumably presents with the *MSH2* variant of interest. gBlocks® Gene Fragments, containing the above sequence, were ordered from IDT, and supplied as a powder.

3.5.2 Determination of Optimal Temperature Gradient

To determine the optimal annealing/extension temperature, the designed *XPF* (p.A596V) primer-probe mix was run across a thermal gradient between 50-60°C. The optimal annealing/extension temperature was determined at a level where positive and negative droplets were sufficiently separated while still minimizing rain, describing droplets between the major positive and major negative droplet population.

gDNA, purified from FARAGE (presumably *XPF* (p.A596V) positive) and U2932 (unknown profile), were diluted to a concentration of 33 ng DNA/µL in nuclease-free water. ddPCR Master Mix was prepared with the following per well: 10 µL 2x ddPCR Supermix for Probes (no dUTP) (Bio-Rad, no. 1863024), 1 µL 20x *XPF* p.A596V target (FAM) and wild-type (HEX) primers/probe mix, and 7 µL nuclease-free water.

In strips of 8, ddPCR Reaction Mix was prepared by adding 2 µL of either gDNA or nuclease-free water, as a no-template control (NTC), to 18 µL Reaction Mix and incubated at room temperature for 3 minutes. 66 ng gDNA per 20 µL Reaction Mix was added. 66ng is the highest concentration before restriction digestion is necessary. 66 ng of gDNA, translating to approximately 18,000 copies, lies within the dynamic range of the QX200™ ddPCR™ System at 1-120,000 copies per 20 µL Reaction Mix.
Using a multipipette, the Reaction Mix was transferred to the middle wells of the DG8™ Cartridge (Bio-Rad, no. 1864008), placed in the DG8™ Cartridge Holder (Bio-Rad, no. 1863051). 70 µL of Droplet Generation Oil for Probes (Bio-Rad, no. 1863005) was aliquoted to the bottom wells of the DG8™ Cartridge. A DG8™ Gasket (Bio-Rad, no. 1863009) was placed over the DG8™ Cartridge and the DG8™ Cartridge Holder was transferred to QX200™ Droplet Generator (Bio-Rad, no. 1864002). Holding the multipipette at a 30° angle, droplets were slowly aspirated and transferred to a ddPCR™ 96-Well Plate (Bio-Rad, no. 12001925) by repelling them slowly along the side of the well. To limit evaporation of the oil, tape was placed over the wells as the next strips were completed. The ddPCR™ 96-Well Plate was sealed using PCR Plate Heat Seal Foil (Bio-Rad, no. 1814040) and PX1 PCR Plate Sealer (Bio-Rad, no. 1814000). On the C1000 Touch™ Thermal Cycler (Bio-Rad, no. 1851197), the cycles were defined as described in Table 5, the lid was preheated to 105°C, and the ddPCR™ 96-Well Plate was run.

Table 5: Droplet Digital™ PCR Temperature Gradient cycle conditions with a annealing/extension temperature gradient across rows A to H with temperatures of 60°C, 59.4°C, 58.1°C, 56.2°C, 54°C, 52.1°C, 50.7°C, 50°C, respectively.

Cycle step	Temperature	Time (hh:mm:ss)	Number of cycles	Ramp rate
Enzyme activation	95°C	00:10:00	1	2°C/sec
Denaturation	94°C	00:00:30	40	
Annealing/extension	*50-60*°C	00:01:00		
Enzyme deactivation	98°C	00:10:00	1	
Hold (cooling)	12°C	Infinite	1	1°C/sec

QX200™ Droplet Reader (Bio-Rad, no. 1864003) was used to read the ddPCR™ 96-Well Plate post ddPCR. In the QuantaSoft Software, the experiment was chosen as ABS (Absolute Quantification of Target) and the template was defined. Having placed the ddPCR™ 96-Well Plate in the holder, the system was run. A detailed protocol of the ddPCR™ Temperature Optimization is available in Appendix 4: Droplet Digital™ PCR Temperature Gradient.

3.5.3 *XPF* (p.A596V) Rare Mutations Variant Detection

ddPCR™ Rare Mutation Detection was performed on gDNA purified from all cell lines (cf. Table 1) to validate the presence of variants of interest. Each DNA sample was diluted to a concentration of 33 ng/µL in nuclease-free water (cf. Table 1 in Appendix 6: Droplet Digital™ PCR Mutation Variant

Detection). The same procedure was followed as with ddPCR™ PCR Temperature Gradient. However, the cycles on the C1000 Touch™ Thermal Cycler were defined as listed in Table 6. An Annealing/extension temperature of 54°C was chosen based on the prior temperature optimization.

Table 6: Droplet Digital™ PCR Rare Mutation Detection cycle conditions.

Cycle step	Temperature	Time	Number of cycles	Ramp rate
Enzyme activation	95°C	00:10:00	1	2°C/sec
Denaturation	94°C	00:00:30	40	
Annealing/extension	54°C	00:01:00		
Enzyme deactivation	98°C	00:10:00	1	
Hold (cooling)	12°C	Infinite	1	1°C/sec

A detailed protocol of the ddPCR™ Rare Mutation Detection is available in Appendix 6: Droplet Digital™ PCR Rare Mutation Detection.

3.6 CELL CULTURE

Four cell lines were cultured using the aseptic technique for the use in drug screenings and challenge assays. These included OCI-Ly7, DB, FARAGE and SU-DHL-4 (cf. Table 7). A detailed description of the cell line start-up from mother stocks is described in Appendix 8: Cell Line Start-Up [Danish]. OCI-Ly7, FARAGE and SU-DHL-4 were grown in RPMI 1640 (Gibco®, no. 52400-025) with 10 % fetal bovine serum (FBS) and 1 % penicillin-streptomycin (Pen/Strep), providing essential nutrients and reducing the risk of bacterial contamination. During the first week of culturing, OCI-Ly7 and SU-DHL-4 were grown in RPMI 1640 with 20 % FBS and 1 % Pen/Strep, and kept in 24-well plates, in order to provide them with the optimal physical environment. DB was grown in RPMI 1640 with 20 % FBS and 1 % Pen/Strep and was kept in 6-well plates during the first two weeks of culturing. The cell lines were kept in the incubator at 37°C with 5 % CO_2 and maintained at a concentration of 0.5×10^6 cells/mL by splitting every 2-3 days, using Trypan Blue stain (Gibco®, no. 15250-061) and Bürker-Türk hemocytometer (Assistant, Germany). They were cultured for no more than 30 passages. A detailed protocol for cell line split can be found in Appendix 8: Cell Line Maintenance in Cell Lab [Danish].

Mother stocks were made by isolating 5×10^6 cells/vial in RPMI 1640 with 20 % FBS, 1 % Pen/Strep, and 10 % dimethyl sulfoxide (DMSO), and frozen in a Styrofoam box placed in -80°C, which gradually reduced the temperature. After 48 hours the vials were transferred to -180°C. For a detailed description of the preparation of mother stocks, refer to Appendix 11: Cell Line Mother Stocks [Danish].

Before making mother stocks, discarding any cells or starting an experiment, the cell lines were authenticated using Barcoding and checked for mycoplasma contamination. For a detailed description of the preparation of material used for Barcoding and Mycoplasma Test see Appendix 10: Preparation of Material for Barcoding and Mycoplasma Test.

Table 7: Specific culture conditions for each of the four cell lines utilized in this project. FBS: Fetal Bovine Serum; Pen/Strep: Penicillin and Streptomycin; DMSO: Dimethyl sulfoxide.

Cell line	T2	Culture media	Split	Startup	Maintenance	Incubation	Freezing
OCI-Ly7	20-24 hours	RPMI 1640 + 10 % FBS + 1 % Pen/Strep	Every 2-3 days	During the first week, the cells are kept in 20 % FBS and grown in 24-well plates	0.5-1.0×10^6 cells/mL	37°C 5 % CO_2	RPMI 1640 + 20 % FBS + 1 % Pen/Strep + 10 % DMSO 5×10^6 cells/vial
DB	~50 hours	RPMI 1640 + 20 % FBS + 1 % Pen/Strep	Every 2-3 days	During the first two weeks, the cells are kept in 20 % FBS and grown in 6-well plates	0.5-2.0×10^6 cells/mL	37°C 5 % CO_2	RPMI 1640 + 20 % FBS + 1 % Pen/Strep + 10 % DMSO 5×10^6 cells/vial
FARAGE	48 hours	RPMI 1640 + 10 % FBS + 1 % Pen/Strep	Every 2-3 days	From the start, the cells are kept in 20% FBS and grown in a small culture flask	0.5×10^6 cells/mL	37°C 5 % CO_2	RPMI 1640 + 20 % FBS + 1 % Pen/Strep + 10 % DMSO 5×10^6 cells/vial
SU-DHL-4	20-30 hours	RPMI 1640 + 10 % FBS + 1 % Pen/Strep	Every 2-3 days	During the first week, the cells are kept in 20 % FBS and grown in 24-well plates	0.5-1.5×10^6 cells/mL	37°C 5 % CO_2	RPMI 1640 + 20 % FBS + 1 % Pen/Strep + 10 % DMSO 5×10^6 cells/vial

3.7 DRUG SCREENING

A cell count based drug screening is carried out with cisplatin to evaluate the cell lines response to the chemotherapy. Cells of OCI-Ly7, DB, FARAGE and SU-DHL-4 at an optimized seeding concentration determined using Trypan Blue stain was seeded into 12-well plates, incubated for 24 hours at 37°C with 5 % CO_2, and affected with 2-fold dilutions of cisplatin with three concentrations.
The concentrations were cell line specific and chosen based on the previous systematic dose-response screening (cf. Table 8). The plates incubated for 48 hours at 37°C with 5 % CO_2, and cell count was determined using Trypan Blue stain, completing one plate before starting another. Appendix 12: Drug Screening using Cisplatin include a detailed description of the procedure.

Table 8: Overview of cisplatin concentrations used in drug screening.

	Cisplatin concentration (µg/mL)		
OCI-Ly7	0.8	1.6	3.2
DB	0.8	1.6	3.2
FARAGE	0.55	1.1	2.2
SU-DHL-4	0.8	1.6	3.2

3.8 Inhibitors of DNA Repair

3.8.1 Green tea polyphenol epigallocatechin-3-gallate

At the time of writing this paper, EGCG has been ordered from Sigma-Aldrich (no. E4143) with a purity ≥95%, supplied as 50 mg and stored protected from light at 4°C. It is known for its inhibition of the ERCC1-XPF interaction. However, the assay has not yet been run.

3.8.2 17-(Allylamino)-17-demethoxygeldanamycin

17AAG is used to inhibit MSH2 and evaluate the consequential effect on cell lines' cisplatin response. A viability assay was carried out to determine the highest possible concentration of 17AAG, with a minimal cell growth inhibition, to be used in a challenge assay with cisplatin. 17AAG was purchased from Sigma-Aldrich (no. A8476), with a purity of ≥98%, supplied as 500 μg lyophilized powder, stored desiccated and protected from light at -20°C.
Cells of OCI-Ly7 and SU-DHL-4 were seeded in 24-well plates at a concentration of 0.25×10^6 cells/mL and cultivated for 24 hours at 37°C with 5 % CO_2. 500 μg 17AAG powder was dissolved in 100 μL DMSO and subsequently diluted in 1900 μL PBS, resulting in a stock concentration of 250 μg/mL. A 2-fold dilution of 17AAG was made ranging from an in-well concentration of 0.15 μg/mL to 2.4 μg/mL, yielding in a maximum of 0.048% DMSO in a well (cf. Appendix 14. Dilution of 17AAG). These in-well concentrations were inspired from Choi et al. (2014)[25] and scaled up according to cell-count. A 2-fold dilution of DMSO was made ranging from an in-well concentration of 0.003% to 0.048%. Cells were subjected to either diluted 17AAG or DMSO, incubated for 48 hours at 37°C with 5 % CO_2, and counted using Trypan Blue stain. A detailed protocol of the viability assay is available in Appendix 13: Viability Assay using 17AAG.

Based on the viability assay, the optimal concentration of 17AAG and corresponding concentration of DMSO was used in a challenge assay with cisplatin to evaluate the effect inhibition of MSH2 would have on the cisplatin response. Cells of OCI-Ly7 and SU-DHL-4 were seeded out in 24-well plates at a concentration of 0.25×10^6 cells/mL, incubated at 37°C with 5 % CO_2 for 24 hours, and affected by different combinations of cisplatin at in-well concentrations previously described, 17AAG at the optimized concentration, DMSO at corresponding concentration, or isotonic saline. (cf. the plate layout in Appendix 15: Challenge Assay using Cisplatin and 17AAG). Cells subjected to a 17AAG concentration with high cell lethality was added as a positive control for the function of the inhibitor. The cells incubated at 37°C with 5 % CO_2, and cell count was evaluated using Trypan Blue satin. A detailed protocol of the challenge assay is available in Appendix 15: Challenge Assay Using Cisplatin and 17AAG.

3.9 Comet Assay® Single Cell Gel Electrophoresis

Comet assay is a method to evaluate the degree of DNA damage. The CometAssay Kit® was purchased from Trevigen® (no. 4250-050-K). Cells of OCI-Ly7 and SU-DHL-4 were seeded in a 12-well plate at a concentration of 0.25×10^6 cells/mL, cultivated for 24 hours at 37°C with 5 % CO_2 for 24

hours, and subjected to different combinations of 17AAG at the optimized concentration, DMSO at corresponding concentration, cisplatin at an in-well concentration of 1.6 μg/mL, and isotonic saline (cf. the plate layout in Appendix 17: Comet Assay® – Preparation of Cells) for four hours. Cells were washed in chilled 1X PBS (Ca^{++} and Mg^{++} free), diluted in chilled 1X PBS (Ca^{++} and Mg^{++} free) at a concentration of 1×10^5 cells/mL, added 1:10 in molten Comet Assay® LMAgarose (Trevigen® no. 4250-050-02), and spread evenly on a well of the CometSlides™ (Trevigen® no. 4250-050-03), incubated 10 minutes at 4°C in the dark, and immerse in chilled Comet Assay® Lysis Solution (Trevigen® no. 4250-050-01) over night at 4°C. The CometSlides™ were drained and immersed in Alkaline Unwinding Solution (300mM NaOH [pellets: Merck, no. 106498] and 1mM EDTA [Trevigen® no. 4250-050-04] in dH_2O; pH>13, chilled to room temperature) for 20 minutes at room temperature. Electrophoresis was performed in a horizontal electrophoresis system (Bio-Rad). Alkaline Electrophoresis Solution (300 mM NaOH [pellets] and 1 mM EDTA [500mM: Invitrogen™, no. 15575-038] in dH_2O; pH >13, cooled to 4°C) was added and electrophoresis was run with 1 volt/cm for 40 minutes. Post electrophoresis, the CometSlides™ were immersed in dH_2O for 5 minutes two times, then 70 % ethanol for 5 minutes one time, dried for 10-15 minutes in 37°C, and subsequently stained with a dilution of 1 μL 10,000X SYBR® Gold (DMSO) (Thermo Fischer Scientific, Invitrogen™, nr. S11494) in 30 mL TE Buffer (10 mM Tris-HCl and 1 mM EDTA: Sigma-Aldrich, no. 93283) for 30 minutes at room temperature in the dark. The CometSlides™ were rinsed with water and allowed to dry completely at 37°C. The cells were visualized using Primovert Epifluorescence Microscope (Carl Zeiss) equipped with a fluorescein filter and pictures were taken with a cellphone through the ocular.
A detailed description of the preparation of cells are described in Appendix 17: Comet Assay® – Preparation of Cells. A detailed protocol for Comet Assay is available at Appendix 18: Comet Assay® – Alkaline Assay Reagent Kit.

3.10 VARIANT ANALYSIS OF DIFFUSE LARGE B-CELL LYMPHOMA PATIENT

The raw DNA sequencing data from DaveLab[34] was made available. From this, Mads Sønderkær, Senior Bioinformatician at The Department of Hematology, Aalborg University Hospital, conducted a variant analysis using Computerome. The input data consisted of: 1) the .vcf file with the raw DNA sequencing data, 2) a list of 35 important DNA repair related genes in .bed format, 3) an excel sheet with the sample information, supplied as Supplementary Information to Reddy et al. (2017)[34], and 4) a .txt file with sample IDs vs. sample names. The output data comprised variants found in the 35 genes, annotated using Variant Effect Predictor (VEP) impact (SIFT and Polyphen) from ENSEMBL. Further annotations consisted of information from OncoKB (gene information and both general and specific variant information), COSMIC Cancer Gene Census (gene information), CIViC (gene information), and ONgene (gene information) databases.

Additionally, DNA sequencing data from 28 relapsed and 44 de novo Hematology Aalborg DLBCL patients (gathered as of the 5th of January 2019) were analyzed using Computerome by Mads Sønderkær. The input data comprised 1) the .vcf files for ProSeq and RetroGen DNA samples and 2)

the list of 35 important DNA repair related genes in a .bed format. The output data comprised variants found in the 35 genes, annotated as the variant analysis on DaveLab's 1,001 de novo DLBCL patients.

These two variant analyses constitute a great analytical potential. In the scope of this thesis, the analyses were used to generate a graphical protein variant view of three chosen genes, to visualize the difference in mutation in cell lines and clinical patients. Variants had to be non-synonymous, in an exon, and annotated a protein change. Variants with VEP Impact annotations of both "tolerated" (SIFT) and "benign" (Polyphen) were excluded.

4 RESULTS

The present study hypothesized that affected NER increases cisplatin-sensitivity, whereas affected MMR decreases cisplatin-sensitivity. This was explored by investigating the occurrence of mutations in selected DNA repair genes in cisplatin-sensitive and -resistant DLBCL cell lines and clinical DLBCL samples, as well as investigating the effect of DNA repair inhibitors on the cisplatin response and the degree of generated DNA fragments in DLBCL cell lines.

4.1 DIVISION OF CELL LINES INTO SENSITIVE AND RESISTANT

The cisplatin derived dose-response AUC_0 computed from a systematic dose-response screening on the 14 DLBCL cell lines, performed prior to initiation of this project, and the mutations identified in most of the screened cell lines were used to generate three different cisplatin Sensitive:Resistant divisions. First, the stratification of AUC_0-values will be described, then the mutation profile of the cell lines will be presented, and finally, three division strategies are evaluated.

4.2 CISPLATIN DOSE-RESPONSE SCREENING

Based on results from the previously described cisplatin dose-response screening, the cell lines were stratified according to AUC_0 and plotted on a graph along with 95 % confidence intervals (cf. Figure 11). AUC_0 is used, instead of GI_{50}, because AUC_0 provides the cumulative effect of cisplatin in the given cell line and is not dependent on only the point on the curve where 50 % growth inhibition is obtained. A high AUC_0 indicate resistance to cisplatin, and a low AUC_0 indicates sensitivity.

Table 9: The difference in AUC_0 between each adjacent cell lines. The left cell line is more sensitive to cisplatin than the right. FARAGE vs. RIVA is twice as high, as DB vs. FARAGE and RIVA vs. SU-DHL-4.

Cell lines			Difference
NU-DUL-1	**vs.**	**SU-DHL-5**	47,28
SU-DHL-5	**vs.**	**NU-DHL-1**	8,68
NU-DHL-1	**vs.**	**OCI-Ly7**	4,71
OCI-Ly7	**vs.**	**MC-116**	14,03
MC-116	**vs.**	**SU-DHL-8**	12,17
SU-DHL-8	**vs.**	**HBL-1**	3,70
HBL-1	**vs.**	**U2932**	7,85
U2932	**vs.**	**DB**	3,76
DB	**vs.**	**FARAGE**	6,37
FARAGE	**vs.**	**RIVA**	10,99
RIVA	**vs.**	**SU-DHL-4**	5,31
SU-DHL-4	**vs.**	**OCI-Ly8**	0,78
OCI-Ly8	**vs.**	**SU-DHL-10**	41,97

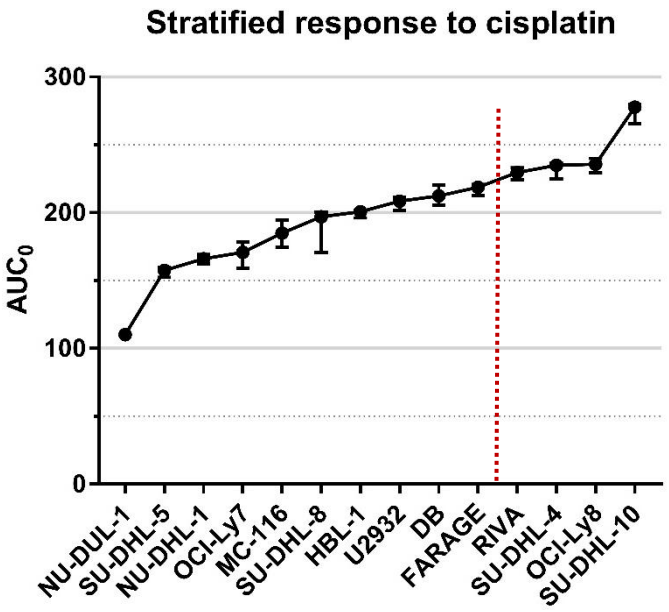


Figure 11: Cell lines' response to cisplatin including 95% confidence interval (in-house dose-response assay), stratified according to area under the curve and divided into two groups, illustrated by the dotted red line. Cisplatin-sensitive cell lines are on the left, cisplatin-resistant on the right.

From the stratified AUC_0 in Figure 11 no obvious split is identified. The difference in AUC_0 between each two adjacent cell lines was calculated and listed in Table 9. When looking at the bottom half, representing cell lines with the highest AUC_0 (lowest sensitivity), the largest difference in sensitivity is observed between FARAGE and RIVA, suggesting this as a possible split between sensitive and resistant cell lines.

4.2.1 Mutational profile of cell lines

A list of 177 DNA repair related genes (cf. Appendix 3: DNA Repair Genes), was used to filter mutation variants detected in the cell lines in the CCLE project[45], where sequencing of DLBCL cell lines are publicly available. Four of this project's cell lines were not registered in the CCLE database[45], these comprised HBL-1, U2932, RIVA, and OCI-Ly8. A total of 58 mutation variants distributed across 38 different genes was identified in 10 DLBCL cell lines, as seen in Figure 13.

4.2.2 Venn diagram

To evaluate the distribution of mutations between cisplatin resistant and sensitive cell lines, a Venn diagram was made (cf. Figure 12). Three Sensitive:Resistant division of 7:7, 10:4 and 9:5 were evaluated. Twenty-four and two mutation variants are maintained in Sensitive and Resistant, respectively, independent of the division. Depending on where the split is set, different mutation variants are identified in the Resistant and Sensitive group. Specifically, DB and FARAGE impact the numbers since they contain 8 and 9 mutation variants, respectively (cf. Figure 13), thereby skewing the distribution depending on which group they are classified into.

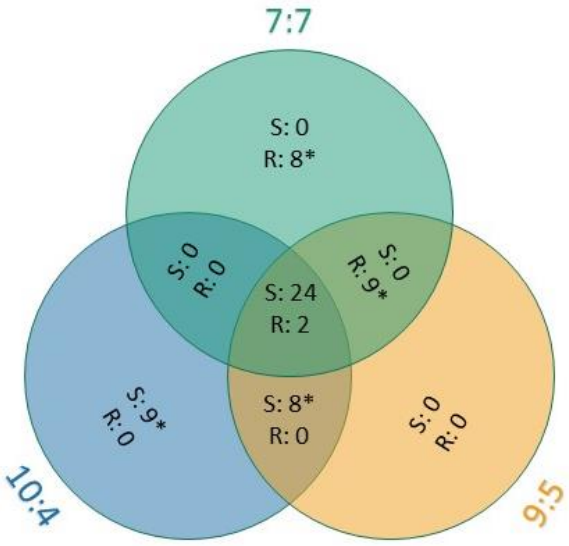


Figure 12: Venn diagram illustrating mutation variants, excluding silent mutations. The asterisk illustrates how eight and nine mutation variants shift position, depending on which cut-off split strategy is being compared.

The number of mutation variants in Sensitive and Resistant are not obvious and depends greatly on how the split is performed. However, to identify possible mutations of importance in cisplatin response and based on the largest AUC_0 ratio difference, the 10:4 cut-off split strategy was chosen. This resulted in only few mutation variants in Resistant compared with Sensitive. As a result, the cell lines are divided into two groups: one group that comprises 10 cell lines sensitive or semi-sensitive to cisplatin, named Sensitive, and a Resistant group, comprising the 4 cell lines (RIVA, SU-DHL-4, OCI-Ly8, and SU-DHL-10) with the highest AUC_0. The dotted red line in Figure 11 illustrates this division between Sensitive and Resistant.

4.3 Mutation profile of Sensitive and Resistant

With the division of 10:4, Sensitive present with a total of 54 mutation variants and Resistant present with a total of 4 mutation variants, of which two are silent mutations, 1 is a missense mutation, and one is a splice site mutation (cf. Figure 13). Besides 13 silent mutation variants, Sensitive encompasses 29 missense mutations, three nonsense mutations, six frame shift mutations and three splice site mutations. The pie charts in Figure 13 illustrate the number of mutation variants (excluding silent mutations) in each cell line. The color intensity increases clockwise, describing the increase in AUC_0, i.e. the decrease in sensitivity to cisplatin. The four cell lines whose mutation profile was not included in CCLE (Sensitive: HBL-1 and U2932, Resistant: RIVA and OCI-Ly8) is excluded from the flow diagram and pie charts. In Sensitive, increasing AUC_0 comes with an increase in the amount of mutation variants. However, in cell lines with the highest AUC_0, the Resistant group, no mutation variants are seen.

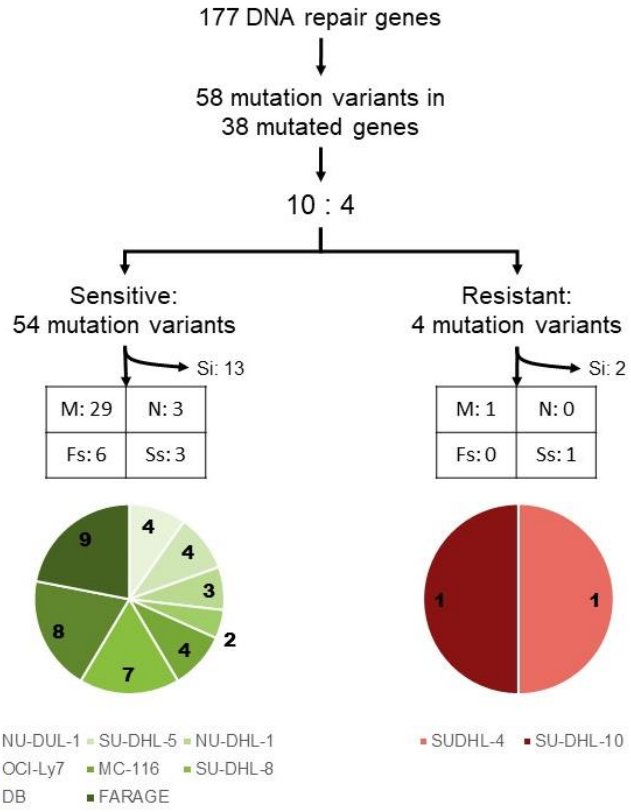


Figure 13: A flow diagram showing the distribution of the number and type of mutation variants in the cell lines. Data found on Cancer Cell Line Encyclopedia, a database from Broad Institute. Excluding HBL-1, U2932, RIVA, and OCI-Ly8 of which no data was available. Si: silent mutation; M: missense mutation; N: nonsense mutation; Fs: frame shift mutation; Ss: splice site mutation.

Table 10: A list of each mutation variant in Sensitive and Resistant. Synonymous variants have been excluded.

Sensitive	Resistant
ATRIP (p.Q118R), *EXO1* (p.Q774L), *FANCA* (p.R1400H), *FANCB* (p.N99fs), *FEN1* (p.K377*), *LIG4* (p.G446A), *MDC1* (p.T1615I), *MSH4* (p.Y705H), *MSH5* (p.S3F), *POLE1* (p.Y473H), *POLE2* (p.S432R), *POLI* (p.C387F), *POLN* (p.A352S), *POLQ* (p.ILL1421fs), *POLZ* (p.V1004E), *POLZ* (p.F301fs), *PRKDC* (p.Y788C), *PRPF19* (p.V438G), *RAD9A* (p.GP330fs), *RAD50* (p.C681F), *RAD51AP1* (p.W184*), *RAD54L* (p.F656fs), *RNF4* (p.69_70insF), *RNF4* (g.2502467_2502468insTGACGA), *SHPRH* (p.S104F), *SLX4* (p.E49K), *TOPBP1* (p.W1145*), *TP53* (p.V143M), *TP53* (p.Y234N), *TP53* (p.C238Y), *TP53* (p.G245D), *TP53* (p.R248Q), *TP53* (p.R248Q), *TP53* (p.R248Q), *TP53* (p.R248W), *TP53* (p.R248W), *TP53* (p.R249G), *WRN* (p.L363P), *WRN* (p.S1333C), *XPF* (p.K353fs), *XPF* (p.A596V)	*TP53* (g.7574034C>T), *TP53* (p.R273C)

Each variant is listed in Table 10, where they are pooled within Sensitive and Resistant, respectively, illustrating the markedly higher number of mutation variants in Sensitive, compared to Resistant. Sensitive encompasses a total of 41 mutation variants in 28 different genes, whereas Resistant presents with two mutation variants in one gene.

4.4 MUTATION VARIANTS AND DNA REPAIR MECHANISMS

To evaluate which mechanisms of DNA repair are affected, the mutated genes and mutation variants were grouped into the six different DNA repair mechanisms: DR, BER, NER, MMR, HR, NHEJ, or as Other. Other comprises a group of genes with an unknown effect in DNA repair, genes with a poorly described function, genes who have been associated with diseases affecting DNA repair function, and genes with a complex multifunctional interaction with DNA repair, such as the known tumor suppressor *TP53*.

Four bar charts have been drawn: two on a gene-level, and two on variance-level. Some genes play a role in more than one DNA repair mechanism. In this case, the mutated gene is included in each involved DNA repair mechanism, to fully evaluate how affected the given mechanism is. In this study, this applied to three of the mutated genes: *POLE1*, involved in both NER and MMR, *POLE2*, involved in both NER and MMR, and *EXO1*, involved in both MMR and HR.

Figure 14 shows the difference between the number of mutated genes, not influenced by mutation variants, in each DNA repair mechanism, without (Figure 14A) and with (Figure 14B) regard to frequency within each DNA repair group. Sensitive, shown with green bars, present with a higher number of mutated genes than Resistant, shown with red bars.

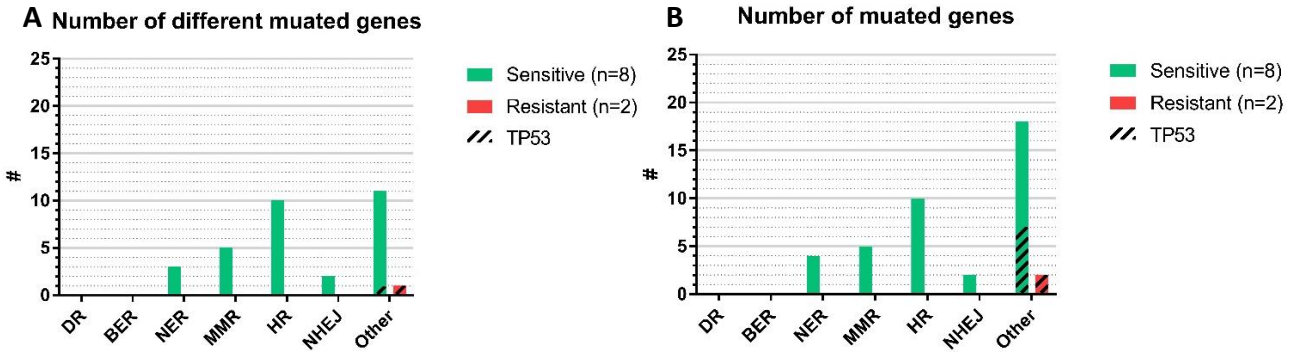


Figure 14: Bar charts illustrating the difference between the number of genes in each DNA repair mechanism. Genes are included in each of the repair mechanism in which they play a role. n equals the number of cell lines in each group to which data was available in the CCLE database. The gene, *TP53*, is marked with a striped column, to visualize its dominance. **A)** Mutated genes are shown without regard to frequency within each DNA repair group. The gene is counted as one within the DNA repair group, regardless of there being more than one cell line, within the DNA repair group, with a mutation in this gene, and regardless of there being different mutation variants in the gene. **B)** Genes are counted as one, if a mutation is present. If there are multiple mutation variants in the same gene, the gene is still only counted as one. This is done within each cell line. Therefore, the gene is counted as one again, if another cell line within the same DNA repair group also presents with a mutation in this gene, regardless of it being the same mutation variant as another cell line.

When including frequency, only NER and Other increases, visualizing that one gene in NER, namely *XPF* and two genes in Other, *TP53*, shown as a striped overlaid bar, and *WRN*, is mutated in more

than one cell line. Figure 15 focuses on variance-level and shows the mutation variants sorted into each of the six DNA repair mechanism in which the respective gene plays a role in, or Others. With Figure 15A, each unique mutation variant is included, regardless of multiple variants in the same gene. The frequency of mutation variants has been included in Figure 15B, where each variant is counted as one, even if another cell line presents with the same variant. When including frequency, *TP53* is the only gene with the same variant in more than one cell line, namely *TP53* (p.R248Q) and *TP53* (p.R248W) (cf. Table 10)

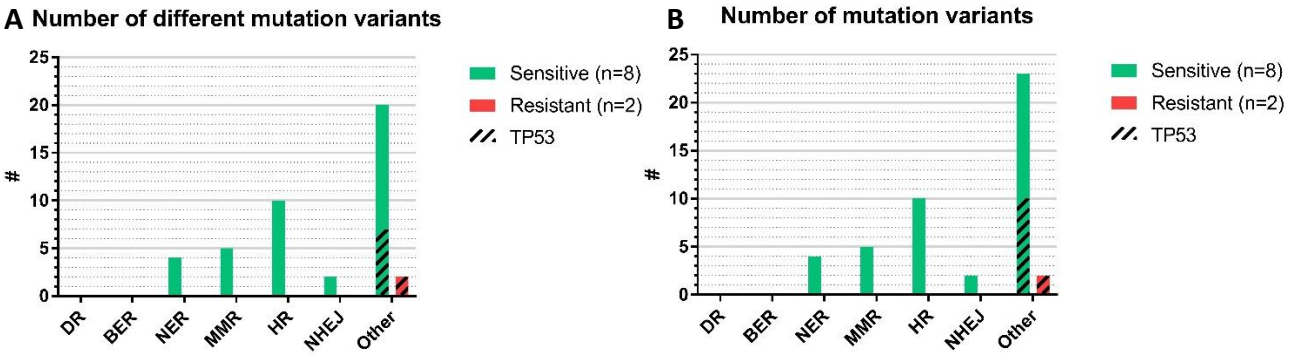


Figure 15: Bar charts illustrating the amount of mutation variants in each DNA repair mechanism. Mutation variants are included in each of the repair mechanism in which the gene plays a role. n equals the number of cell lines in each group, whose mutation profile was found in the CCLE database. Variants in *TP53* is indicated with a striped column, to visualize its dominance. **A)** Mutations are included on variance level, meaning each unique mutation variant is counted as one within each DNA repair group and within Sensitive and Resistant. Therefore, the same mutation variant might be included in both Sensitive and Resistant. Each different mutation variant is counted as one, despite of them being in the same gene. **B)** Mutations are included on variance level with regard to frequency, meaning each mutation variant, regardless of it being the same variant another cell line presented with, is counted as one within each cell line. Therefore, the same mutation variant might be included multiple times within each DNA repair group. If one gene encompasses multiple mutation variants, these are all counted as one.

The mutation profile is heterogenous between the cell lines and no hotspots can be observed, despite *TP53*. When looking at the six DNA repair mechanisms, NER, MMR, and HR are mostly affected by mutations. The Other category seems highly mutated as well. However, *TP53*, a known tumor suppressor with a complicated and intertwined role in DNA repair signaling, dominates this group. This indicates that there could be a link between mutations in these three DNA repair mechanisms and sensitivity towards cisplatin, NER and MMR being the most important in its mechanism of action.

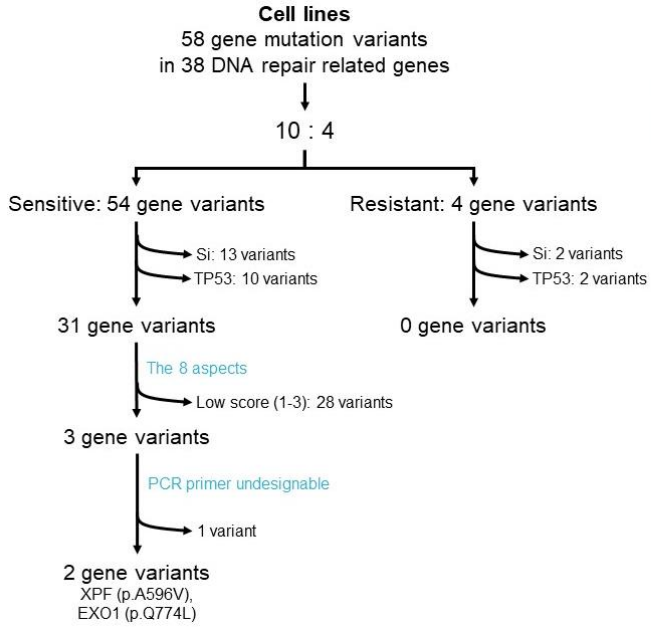
4.5 CHOOSING GENE MUTATION VARIANTS

To investigate the link between mutations in certain DNA repair genes and the sensitivity to cisplatin, specific mutation variants were chosen for further analysis. A systematic score was developed, comprising eight assessment criteria: 1) the general function of the gene, 2) which DNA repair mechanisms it plays a role in and the importance in relation to cisplatin's mechanism of action, 3) if the gene is part of the Hallmark DNA Repair Gene Set[46,47], 4) any previous annotations of either

the gene or 5) the variant, 6) the type of mutation variant (ranked according to VEP Impact by Ensembl[48]), 7) the location of the variant in the protein, and 8) the prevalence of the specific variant in the sample population.

4.5.1 Variants found in cell lines

Mutation variants in 10 out of 14 cell lines were found in the CCLE database[45] (cf. Figure 16).



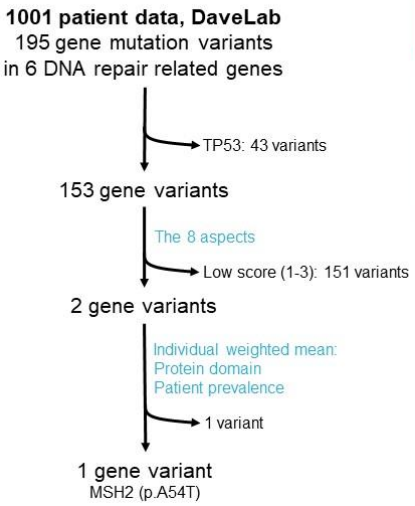
Final Score	n	Variants
5	0	
4	3	XPF (p.K353fs), XPF (p.A596V), EXO1 (p.Q774L)
3	3	FANCA (p.R1400H), POLZ (p.F301fs), RAD50 (p.C681F),
2	12	ATRIP (p.Q118R), POLZ (p.V1004E), RAD54L (p.F656fs), WRN (p.L363P), WRN (p. p.S1333C), MDC1 (p.T1615I), POLE1 (p.Y473H), POLE2 (p.S432R), FANCB (p.N99fs), POLN (p.A352S), POLQ (p.ILL1421fs), PRKDC (p.Y788C)
1	13	SLX4 (p.E49K), POLI (p.C387F), RAD51AP1 (p.W184*), TOPBP1 (p.W1145*), FEN1 (p.K377*), LIG4 (p.G446A), MSH4 (p.Y705H), MSH5 (p.S3F), PRPF19 (p.V438G), RAD9A (p.GP330fs), RNF4 (p.69_70insF), RNF (g.2502467_2502468insTGACGA), SHPRH (p.S104F)

Figure 16: Flow diagram showing the process of choosing two gene mutation variants for further analysis on the basis of cell lines' DNA profile, found on the CCLE database[45]. All variants were assigned a weighted score within eight assessment criteria, which included gene function, DNA repair mechanism, inclusion in the Hallmark DNA repair Gene Set[46,47], previous annotations of the gene or the variant, type of mutation, protein change, and the prevalence. A weighted average was calculated and translated into a final score for the variant (depicted in the table). *XPF* (p.K353s) was excluded because primers for ddPCR were undesignable.

After having removed silent mutation variants and variants in *TP53*, Sensitive presents with 31 gene variants, whereas Resistant has none left (cf. Figure 16). The 31 variants were assigned weighted scores from 1 to 5 within the eight assessment criteria. From the weighted scores, a weighted mean was calculated and translated to final scores of 5 to 1; 1 was given to variants of no interest, 5 was given to highly interesting variants. The table in Figure 16 shows the variants, their final scores, and the total number of variants (n) within the group. The lowest scored variants (1-3) were excluded, leaving three variants: *XPF* (p.K353fs), *XPF* (p.A596V), and *EXO1* (p.Q774L). *XPF* (p.K353fs) was found in NU-DHL-1, *XPF* (p.A596V) in FARAGE, and *EXO1* (p.Q774L) in DB.

4.5.2 Variants found in DaveLab's 1,001 de novo DLBCL

The list of 177 DNA repair genes was used to filter the mutations profile of 1,001 de novo DLBCL patients from DaveLab[34], publicly available at https://dlbcl.davelab.org[35]. This identified 195 mutation variants in 7 DNA repair related genes (cf. Figure 17), whereto silent mutation variants had already been excluded.



Final Score	n	Variants
5	0	
4	2	MSH2 (p.A54T), MSH6 (p.R1242H)
3	23	ATM (p.Q2108R), ATR (p.S1142G), ATR (p.S1875N), ATR (p.A1972T), ATR (p.D2055N), ATR (p.D2063G), ATR (p.R2089Q), ATR (p.R2363Q), ATR (p.F2483L), ATR (p.H2517R), ATR (p.E2626K), MSH2 (p.R30H), MSH2 (p.L36P), MSH2 (p.R39Q), MSH2 (p.R40K), MSH2 (p.L314*), MSH2 (p.P729T), MSH6 (p.S459C), MSH6 (p.Q889H), MSH6 (p.F916L), MSH6 (p.T1225M), MSH6 (p.T1225M), MSH6 (p.G1265E)
2	119	ATM (p.R13H), ATM (p.R250*), ATM (p.R337H), ATM (p.N358S), ATM (p.P428S), ATM (p.N429S), ATM (p.V613L), ATM (p.K638Q), ATM (p.H674R), ATM (p.R720C), ATM (p.N765T), ATM (p.R805*), ATM (p.D860G), ATM (p.D1099N), ATM (p.A1309T), ATM (p.I1343S), ATM (p.N1356D), ATM (p.Q1361R), ATM (p.T1429I), ATM (p.R1466*), ATM (p.M1830T), ATM (p.L1861M), ATM (p.Y1961C), ATM (p.G2022R), ATM (p.T2059I), ATM (p.A2067), ATM (p.N2501S), ATM (p.G2508R) , ATM (p.A2524P) , ATM (p.E2576K) , ATM (p.A2622G) , ATM (p.N2697H) , ATM (p.F2820L) , ATM (p.R2832H) , ATM (p.V2906A), ATR (p.A9T), ATR (p.I97F), ATR (p.R215C), ATR (p.R668W), ATR (p.S1102F), ATR (p.D1298N), ATR (p.K1379E), ATR (p.R141W), ATR (p.Y1497C), ATR (p.Y1626C), ATR (p.D2111N), ATR (p.R2148*), ATR (p.A2181T), ATR (p.R2190C), ATR (p.L2306R), ATR (p.I2323), ATR (p.T2605A), ATR (p.R2606Q), BTG1 (p.M11I), BTG1 (p.E14D), BTG1 (p.A18V), BTG1 (p.V19M), BTG1 (p.L26P), BTG1 (p.G30E), BTG1 (p.G30R), BTG1 (p.L31F), BTG1 (p.S33N), BTG1 (p.T39I), BTG1 (p.S43N), BTG1 (p.Q45P), BTG1 (p.H51P), BTG1 (p.H54Y), BTG1 (p.W56*), BTG1 (p.P61S), BTG1 (p.G64R), BTG1 (p.G66D), BTG1 (p.Y67H), BTG1 (p.C69Y), BTG1 (p.I70V), BTG1 (p.A84V), BTG1 (p.L104H), BTG1 (p.V106I), BTG1 (p.I115F), BTG1 (p.I121T), UBE2A (p.S29P), UBE2A (p.I33T), UBE2A (p.W36*), UBE2A (p.P43S), UBE2A (p.E44*), UBE2A (p.E61K), UBE2A (p.P64L), UBE2A (p.M76K), UBE2A (p.Y82D), UBE2A (p.I91K), UBE2A (p.S103T), UBE2A (p.I109K), UBE2A (p.L12P), UBE2A (p.L113W), UBE2A (p.D114V), UBE2A (p.N117D), UBE2A (p.S120R), UBE2A (p.S120G), UBE2A (p.Q125*), MSH2 (p.A2V), MSH2 (p.V3M), MSH2 (p.G18D), MSH2 (p.G137R), MSH2 (p.C267R), MSH2 (p.L321R), MSH2 (p.Q353K), MSH2 (p.I388T), MSH2 (p.L422V), MSH2 (p.K425M), MSH2 (p.T498A), MSH2 (p.E577K), MSH2 (p.C641Y), MSH6 (p.W50C), MSH6 (p.D257G), MSH6 (p.T269S), MSH6 (p.E318G), MSH6 (p.R1076H), MSH6 (p.P1082L), MSH6 (p.P1087H)
1	9	BTG1 (p.H2Y), BTG1 (p.P3L), BTG1 (p.S133N), BTG1 (p.T138S), BTG1 (p.D144Y), BTG1 (p.T158A), BTG1 (p.M166L), UBE2A (p.M1T), UBE2A (p.A5V)

Figure 17: Flow diagram showing the process of choosing one gene mutation variants for further analysis on the basis of the DNA profile of 1,001 de novo DLBCL patient from DaveLab[34], publicly available at https://dlbcl.davelab.org[35]. All variants were assigned a weighted score within to eight assessment criteria, which included gene function, DNA repair mechanism, inclusion in the Hallmark DNA repair Gene Set[46,47], previous annotations of the gene or the variant, type of mutation, protein change, and the prevalence. A weighted average was calculated and translated into a final score for the variant (depicted in the table). One variant was chosen from the two variants with a final score of 4 by evaluating their individual weighted mean.

After having removed variants in *TP53*, 153 variants were scored in regard to the eight assessment criteria of the systematic scoring system. The weighted averages were translated to final scores between 1 to 5. The table in Figure 17 lists each variant, their final score, and the total number of variants (n) within that score. When 151 variants had been removed because of a low score (1-3), one variant in *MSH2* and one variant in *MSH6* remained. *MSH2* (p.A54T) is located in MutS_I domain, the DNA binding domain and found in four out of the 1,001 de novo DLBCL patients, whereas *MSH6* (p.R1242H) is located in MutS_V domain, the ATPase domain, and only found in one out of 1,001 de novo DLBCL patients (cf. Appendix 2: Reddy et al. (2017) Survival Estimate). Therefore, *MSH2* (p.A54T) was scored with the highest weighted mean and was chosen for further analysis.

4.5.3 The three variants

The information on the three chosen mutations variants is listed in Table 11. The variants found in cell lines, *XPF* (p.A596V) and *EXO1* (p.Q774L), were not found in DaveLab's cohort, and the variant, *MSH2* (p.A54T) found in 4 out of 1,001 de novo DLBCL patients from DaveLab[34] was not identified in the cell lines.

Table 11: Information on the three mutation variants that were chosen based on the DNA profile of cell lines and 1,001 de novo DLBCL patients, provided by DaveLab[34]. NER: nucleotide excision repair; MMR: mismatch repair; HR: homologous recombination; Chr: chromosome; ICL: interstrand crosslinks.

Protein change	***XPF* (p.A596V)**	***MSH2* (p.A54T)**	***EXO1* (p.Q774L)**
DNA repair mechanism	NER	MMR	MMR & HR
Genomic change	g.14029576C>T	g.47630490G>A	g.242048725A>T
Chromosome location	Chr16	Chr2	Chr1
Exon number	Exon 8	Exon 1	Exon 15
Protein domain	-	MutS_I domain	Interaction with MSH2 domain

XPF encodes a protein that is involved in the NER mechanism, where it is responsible for resecting the damaged section of the DNA strand. *XPF* (p.A596V) is found in the coding region of the gene, but in no functional domain of the protein. It is, however, in proximity to the Nuclease domain, which starts at position 658 in the XPF protein.
MSH2 encodes a protein involved in damage recognition of the MMR pathway, more specifically recognizing mismatches in complex with MSH6 and small insertions and deletions in complex with MSH3 and is therefore particular important for the normal function of MMR. *MSH2* (p.A54T) is situated in the MutS_I domain, which is the DNA binding domain.
EXO1 encodes a protein involved in both MMR and HR. EXO1 is responsible for cleaving the daughter strand harboring mismatches in MMR and resecting the 3' ends of damaged DNA strands alongside CtIP in HR. *EXO1* (p.Q774L) is in the 'Interaction with MSH2' domain. Due to limited time, EXO1 was not prioritized as a mutation candidate for further analysis in this project.

4.6 *XPF* (P.A596V) DROPLET DIGITAL™ RARE MUTATION DETECTION

ddPCR™ was performed on DNA extracted from the cell lines by AllPrep as a proof of principle for validating the CCLE database mutations in the in-house cell lines. The presence of *XPF* (p.A596V) was validated in the 14 in-house cell lines, including HBL-1, U2932, RIVA, and OCI-Ly8 whose mutation profile was not available in the CCLE database[45]. ddPCR™ primer-probe mix for *XPF* (p.K353fs), a variant deemed interesting, was found to be undesignable, so it was excluded.

First, an annealing/extension temperature optimization was performed on the *XPF* (p.A596V) primer-probe mix (cf. Figure 18). Only wells with a droplet count above 12,000 are included. No positive signal was observed in the NTC wells. FARAGE was included as expected *XPF* (p.A596V) positive, based on the CCLE database[45], which was verified by p.A596V positive droplets. When looking both at the wild type amplitude (cf. Figure 18A) and the mutant amplitude (cf. Figure 18BFigure 18), the well with an annealing/extension temperature of 54°C displayed the best separation between positive and negative droplets, whilst minimizing rain.

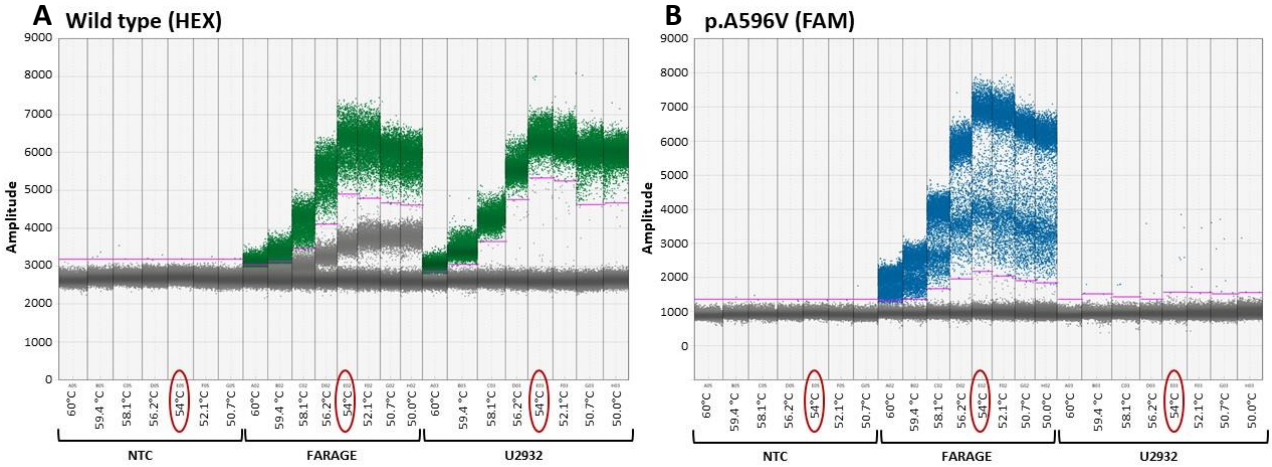


Figure 18: 1D plots from Droplet Digital™ PCR temperature gradient on *XPF* (p.A596V) primer-probe mix. Wells with a droplet count below 12,000 have been removed from the plot. **A)** Shows the amplitude signal on wild type (HEX robe). **B)** Shows the amplitude signal on mutant p.A596V (FAM probe).

Supplementary figures from this ddPCR™ optimization assay can be found in Appendix 5: Supplementary figures to ddPCR™ Temp. Gradient, these comprise droplet count, 1D and 2D plots of 54°C, as well as concentration plot. To evaluate the optimal amplitude threshold, wells with an annealing/extension temperature of 54°C were used to generate a 2D plot (cf. Figure 19). Four clusters are formed: a cluster of double negative droplets, a cluster of mutant positive droplets, a cluster of wild type positive droplets, and a cluster of double positive droplets. The fluorescent threshold for wild type signal is set at 4343 Channel 2 amplitude, whereas the p.A596V signal is set at 2929 Channel 1 amplitude.

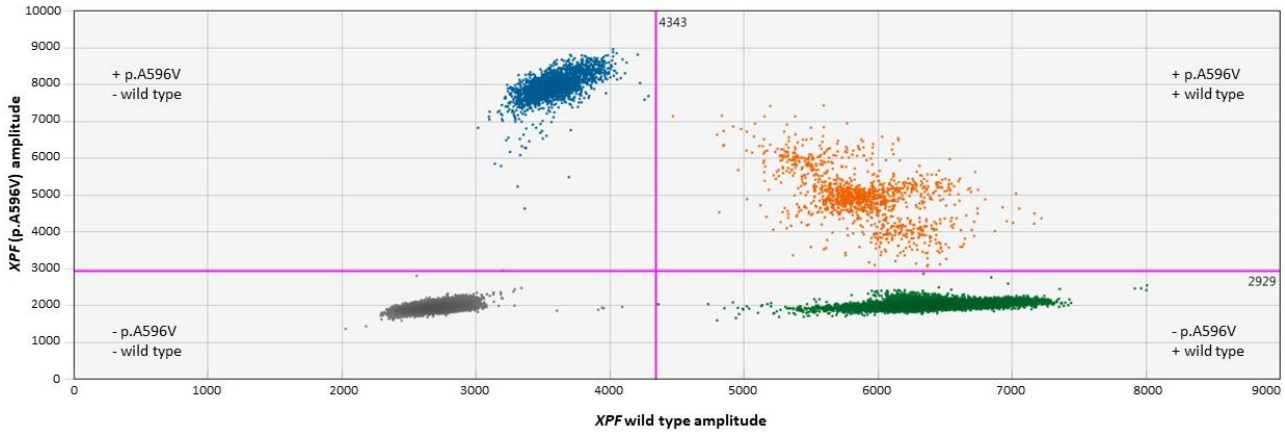


Figure 19: 2D plot of the from Droplet Digital™ PCR temperature gradient on *XPF* (p.A596V). Depicted are wells with an annealing/extension temperature of 54°C, used to set the amplitude threshold for the wild type probe (HEX) and the mutant p.A596V probe (FAM).

Using *XPF* (p.A596V) Rare Mutations Variant ddPCR™ Detection each cell line was run (cf. Figure 20). There were no positive droplets in NTC wells. All cell lines were positive for wildtype (cf. Figure 20A), whereas only FARAGE was positive for p.A596V (cf. Figure 20B). Supplementary figures

are listed in Appendix 8: Supplementary figures to ddPCR™ Rare Mut. Detection, containing droplet count, 1D and 2D plots of *XPF* (p.A596V) negative wells, 1D and 2D plots of *XPF* (p.A596V) positive wells, as well as 2D plots of NTC.

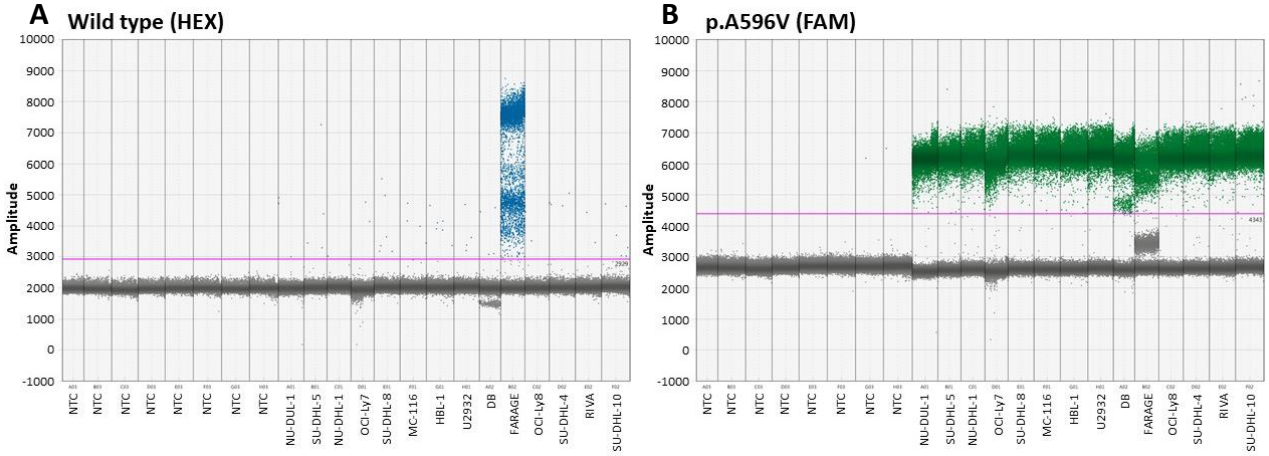


Figure 20: 1D plots illustrating each cell line run on Rare Mutations Variant ddPCR™ Detection using *XPF* (p.A596V) primer-probe mix. **A)** Shows the amplitude signal on wild type (HEX robe) in each cell line and no-template control (NTC). **B)** Shows the amplitude signal on mutant p.A596V (FAM probe) in each cell line and NTC.

On the 2D plot, amplitude thresholds were set as determined in the ddPCR™ Temperature Gradient Assay (cf. Figure 21). The four clusters are distinguishable. Like previously, the p.A596V positive cluster is skewed towards the double positive cluster and the double positive cluster is stretched between the p.A596V positive cluster and wild type positive cluster, rather than being confined to its own quadrant.

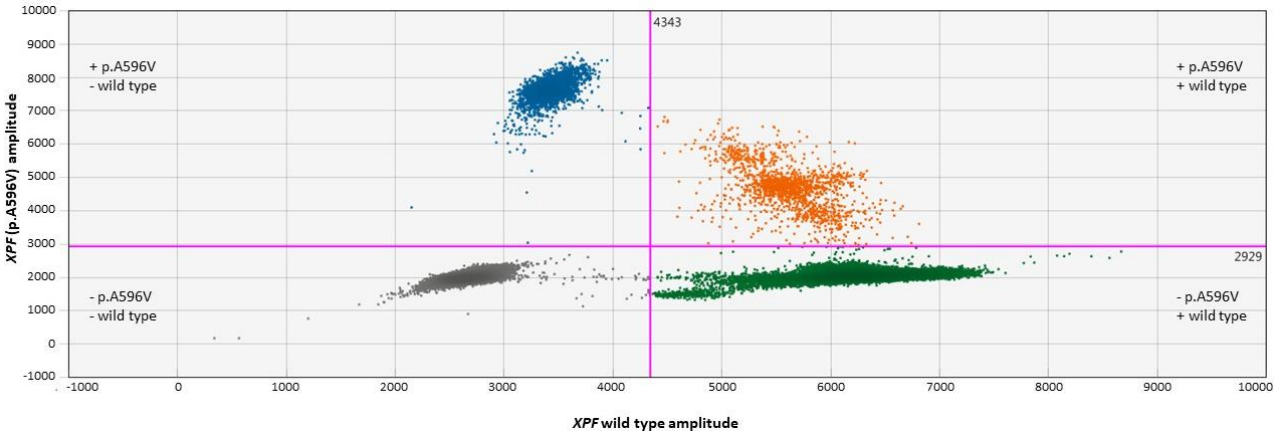


Figure 21: 2D plot illustrating the four clusters from Rare Mutations Variant ddPCR™ Detection using *XPF* (p.A596V) primer-probe mix. Each cell lines and no-template control are included. The grey cluster contains double negative droplets, the blue cluster comprises p.A596V positive droplets, the green cluster contains wild type positive clusters, and the double positive droplets are gathered in the orange cluster.

With the two thresholds defined it is possible to calculate the concentration in copies/µL of *XPF* (p.A596V) and wild type. Figure 22 illustrates a plot of these concentrations. A few of the NTC wells present a small positive result. The well containing FARAGE (B02) showed an approximately

equal concertation of p.A596V (490 copies/µL) and wild type (468 copies/µL). The rest of the cell lines were all positive for the wild type, with concentrations between 1353 copies/µL and 899 copies/µL. Most had a total concentration in the range of 900 copies/µL to 1200 copies/µL, meaning 18,000-24,000 copies/20 µL. This corresponded well with the calculated amount of DNA sample added: 66 ng of DNA, translating to 18,000 copies/20 µL.

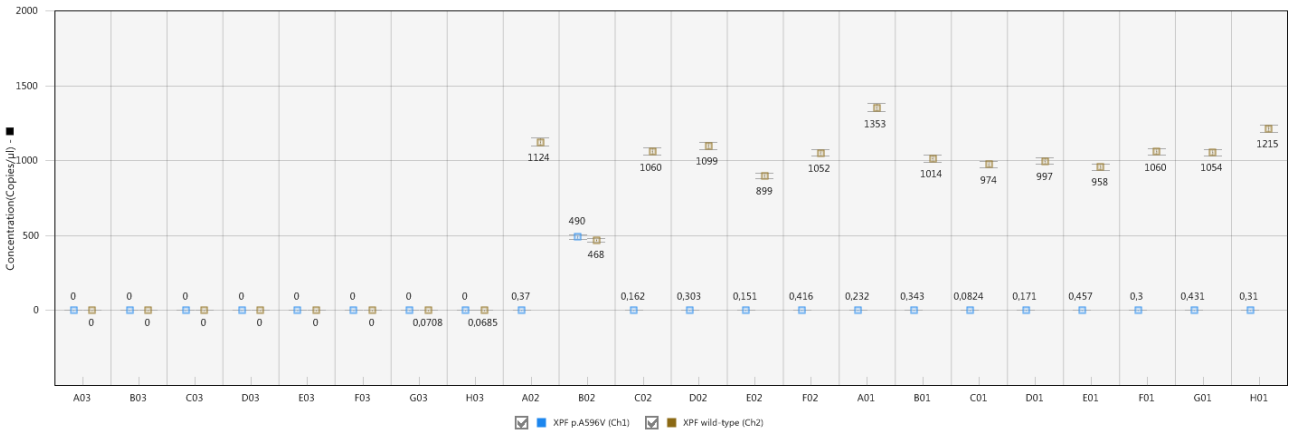


Figure 22: Concentration (copies/µL) of both *XPF* (p.A596V), shown as blue dots, and wild type, shown as brown dots. Well A03-H03 are no-template control. The remaining wells are respectively as follows: NU-DUL-1, SU-DHL-5, NU-DHL-1, OCI-Ly7, SU-DHL-8, MC-116, HBL-1, U2932, DB, FARAGE, OCI-Ly8, SU-DHL-4, RIVA, and SU-DHL-10.

4.7 DRUG SCREENING OF FOUR CELL LINES

As a validation of the cell lines' cisplatin response, observed in the previous dose-response screening, four cell lines were chosen for drug screening with cisplatin based on their cisplatin sensitivity and their general mutation profile, considering the presence of a mutation candidate and mutations that could interfere with results. OCI-Ly7 was chosen because it only has two variants, *PRKDC* (p.Y788C) and *TP53* (p.G245D), and a relatively sensitive response to cisplatin (cf. Figure 11). Likewise, SU-DHL-4 was chosen because it only had a *TP53* (p.R273C) variant but presented with a relatively resistant response to cisplatin (cf. Figure 11). DB and FARAGE had a semi-sensitive response to cisplatin and were chosen because of their mutation in *EXO1* (p.Q774L) and *XPF* (p.A596V), respectively.

The drug screening was a cell count based assay. Results are depicted in Figure 23 as a within cell line relative cell count to 0 µg cisplatin per mL and plotted as a mean of 2-9 assays with SDs. Each assay included three technical replicates, additionally OCI-Ly7 (cf. Figure 23A) and SU-DHL-4 (cf. Figure 23B) have been carried out as two and three independent biological replicates, respectively.

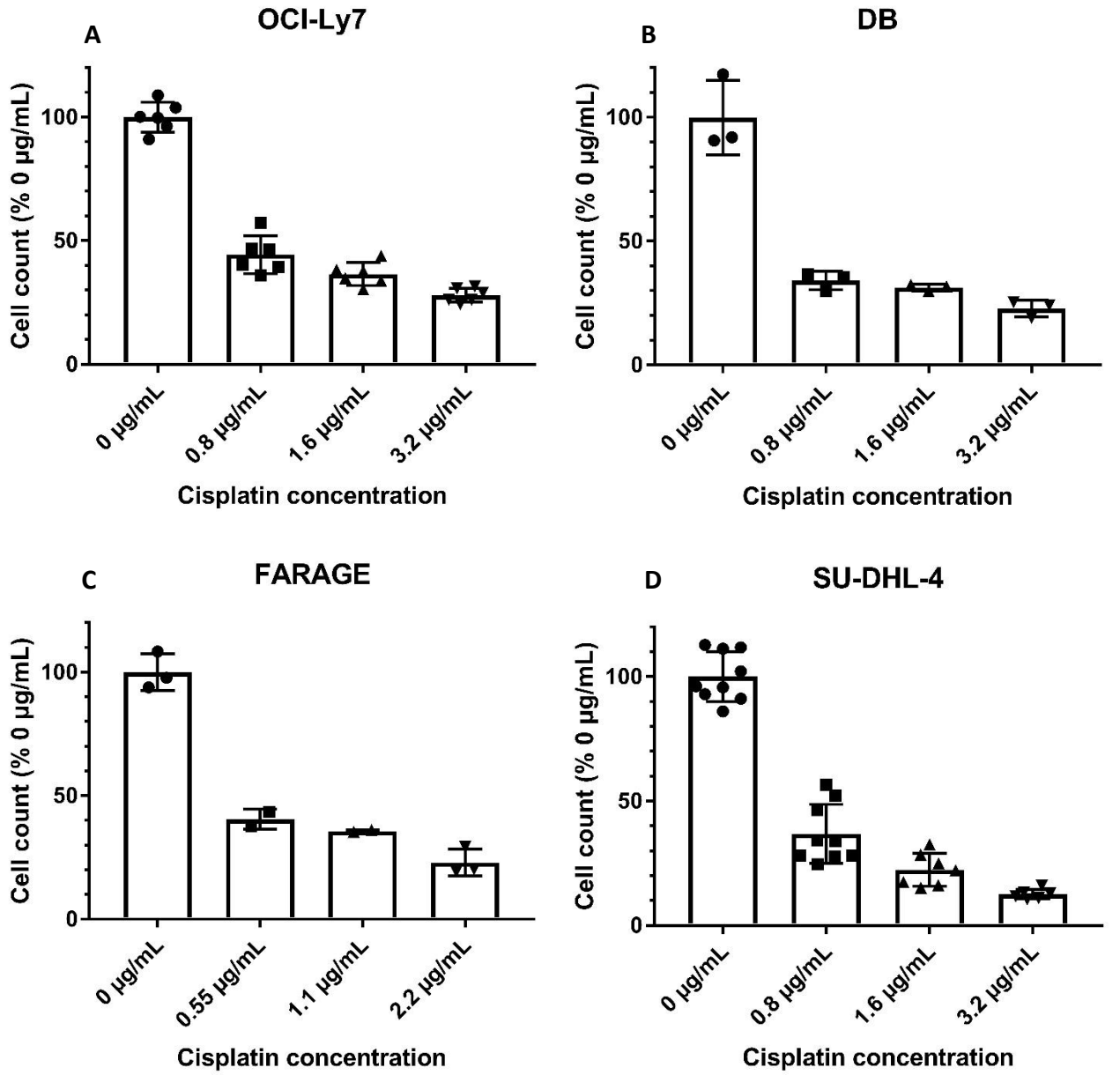


Figure 23: Cell count based drug screening of four DLBCL cell lines (**A**: OCI-Ly7, **B**: DB, **C**: FARAGE, and **D**: SU-DHL-4) using cisplatin at four cell line specific concentrations, determined on the basis of a dose-response screening performed prior to initiation of this project. Cisplatin concentrations are in-well concentrations. Cell count is depicted as a relative to a cisplatin concentration of 0 µg/mL with respective standard deviations.

Three cisplatin concentrations were chosen based on cell line specific GI_{50} from the systematic dose-response screening. For each cell line, it is evident that the cell count decreases with an increasing cisplatin concentration. A large drop in cell count is observed between 0 µg of cisplatin per mL and the lowest supplemented cisplatin concentration. The following decline in cell count is moderate and varies in between cell lines, as well as the size of the SDs varies.

4.8 Inhibition of MSH2

Mutations in *MSH2* showed a decreased survival in DaveLab's cohort of 1,001 de novo DLBCL patients, although not significant (cf. Figure 10). Additionally, in the empiri *MSH2* is implicated in the resistance to cisplatin. The Hsp90 inhibitor, 17AAG, which has been found to decrease the amount of MSH2 and affect cells response to cisplatin, was chosen as an MSH2 inhibitor in this project, for the purpose of investigating the effect of inhibition of MSH2 on the cisplatin response. Two cell lines were included in this challenge assay. OCI-Ly7 and SU-DHL-4 were chosen due to their lack of mutations in DNA repair genes, which could otherwise interfere with the results, and their opposing response to cisplatin: OCI-Ly7 being sensitive and SU-DHL-4 being relatively resistant.
First, a viability assay was carried out, to determine the highest possible concentration of 17AAG that would not inhibit cell growth. With inspiration from a Choi et al. (2014)[25], five concentrations of 17AAG were chosen and scaled up in relation to cell count. Results are depicted in Figure 24.

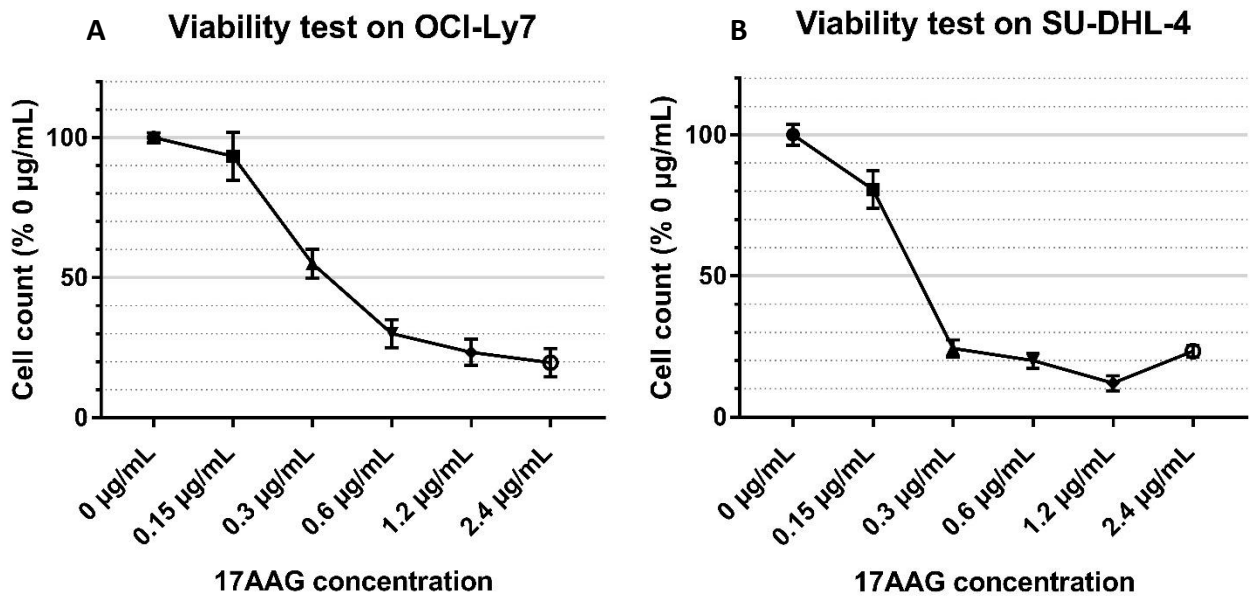


Figure 24: A viability assay on OCI-Ly7 **(A)** and SU-DHL-4 **(B)** using six concentrations of 17-(Allylamino)-17-demethoxygeldanamycin (17AAG). Results are shown as a main of three technical replicates with corresponding standard divisions.

OCI-Ly7 presents with a classical S-shaped curve and a decrease in cell count even at the lowest concentration of 17AAG (cf. Figure 24A). The results from SU-DHL-4 also has an S-shape, although a large drop in cell count is observed between a 17AAG concentration of 0.15 µg/mL and 0.3 µg/mL (cf. Figure 24B). From this assay, a 17AAG concentration of 0.15 µg/mL was chosen for the challenge assay of both cell lines, noting however that a small drop in cell count is observed in both cell lines.

The challenge assay was carried out with the four cisplatin concentrations from the drug screening and with or without 0.15 µg 17AAG per mL, to visualize the effect of the addition of 17AAG, an MSH2 inhibitor, on cisplatin response. The results are depicted as a mean cell count, determined at 48 hours, of three technical replicates, relative to the cisplatin concentration of 0 µg/mL, and with respective SDs. The results of the challenge assay on OCI-Ly7 is shown in Figure 25. The cell count was generally higher, when 17AAG and cisplatin had been added consecutively, compared to cells that were only affected by cisplatin. This suggests a relationship between the function of MSH2 and the sensitivity towards cisplatin.

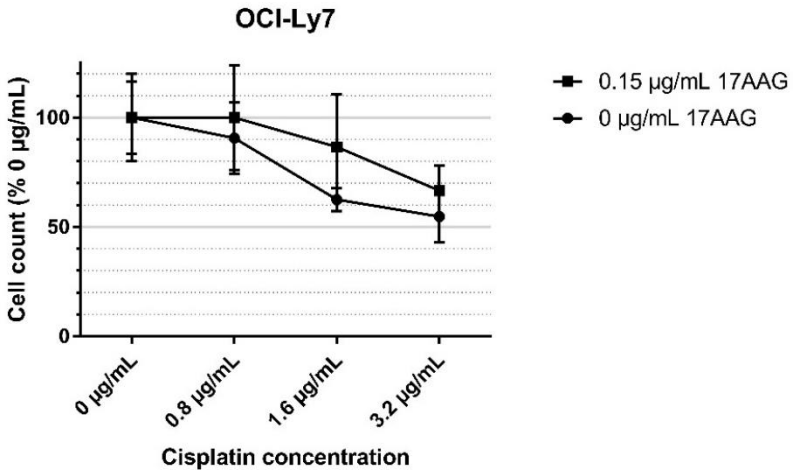


Figure 25: Challenge assay on OCI-Ly7 using four cisplatin concentrations with or without 0.15 µg 17AAG per mL. Results are shown as a relative cell count to cisplatin 0 µg/mL and with standard divisions. The curve with circled points represents the cells who only received cisplatin and the curve with squared points represents the cells who received both 17AAG and cisplatin.

The same tendency was observed with SU-DHL-4, as seen in Figure 26. A consistently higher cell count was observed for cells who had been affected by 17AAG in each cisplatin concentration, showing a lower cisplatin sensitivity because of 17AAG addition. To validate the stability of the inhibitor, a lethal dose of 2.1 µg 17AAG/mL was administered to the cell lines. OCI-Ly7 and SU-DHL-4 showed a 66% and 80% decrease in cell count, respectively, compared to cells only affected by DMSO, verifying the inhibitors functionality (cf. Appendix 16: Lethal Dose of 17AAG).

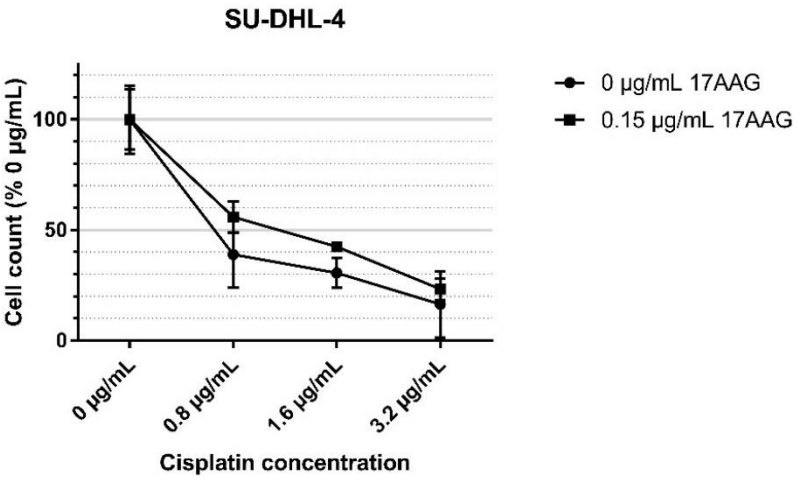


Figure 26: Challenge assay on SU-DHL-4 using four cisplatin concentrations with or without 0.15 µg 17AAG per mL. Results are shown as a relative cell count to cisplatin 0 µg/mL and with standard divisions. The curve with circled points shows the cells affected by cisplatin, whereas the curve with squared points shows the cells affected by both 17AAG and cisplatin.

These results indicate that there could be a correlation between the function of MSH2 and a decrease in the sensitivity to cisplatin.

4.9 Comet Assay® Single Cell Gel Electrophoresis

Since mutations in *MSH2* have been implicated to decrease the sensitivity of cells to cisplatin, by impairing MMR, Comet Assay® Single Cell Gel Electrophoresis was utilized. The difference in the

amount of DNA fragments, otherwise generated through DNA repair, was evaluated between wild type cells, cells who received cisplatin, cells who received 17AAG, and cells who received both cisplatin and 17AAG.
Denatured, cleaved DNA fragments migrates out of the nucleus of a lysed cell while under the influence of electrical current, whereas undamaged DNA migrates slower and remains within the area of the nucleus. This generates cell 'comets', where the nucleus constitutes the comet head and the comet tail composes damaged DNA migrated out of the cell. The tail moment (the distance between the center of the comet head to the center of the comet tail) allows for evaluation of the degree of DNA damage.
A cisplatin-sensitive cell line, OCI-Ly7, and a cisplatin-resistant cell line, SU-DHL-4, was investigated. The results are depicted in Figure 27, where cells are stained using SYBR® Gold DNA staining, viewed at a 40X magnification, and photographed with a phone.

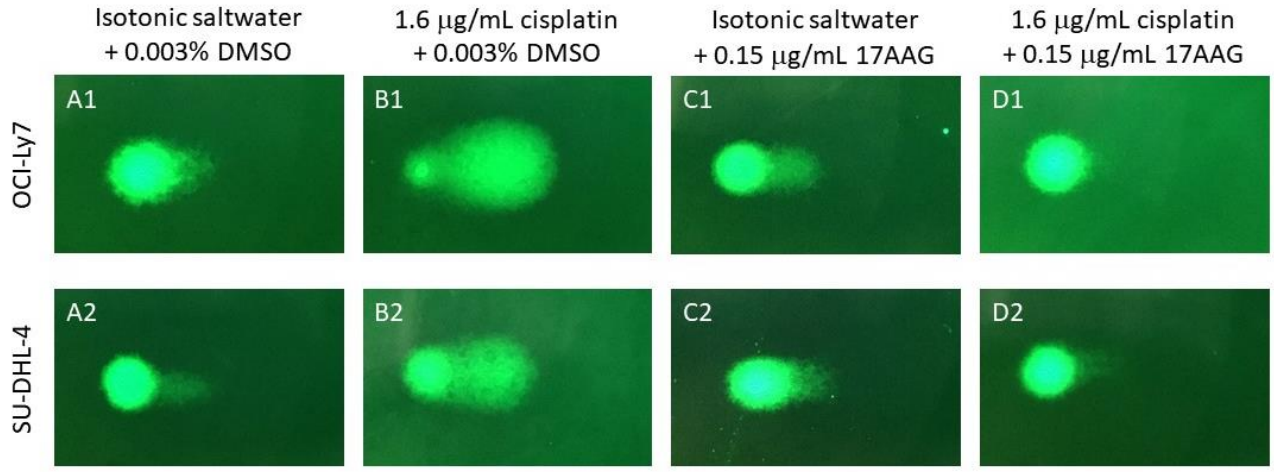


Figure 27: Comet Assay® Single Cell Electrophoresis Assay of OCI-Ly7 (cisplatin-sensitive) and SU-DHL-4 (cisplatin-resistant). Cells from each cell line have been subjected to either **A:** isotonic saltwater (solvent for cisplatin) and 0.003% DMSO (solvent of 17-(Allylamino)-17-demethoxygeldanamycin [17AAG]); **B:** cisplatin at a concentration of 1.6 µg/mL and 0.003% DMSO; **C:** isotonic saltwater and 17AAG at a concentration of 0.15 µg/mL; or **D:** cisplatin at a concentration of 1.6 µg/mL and 17AAG at a concentration of 0.15 µg/mL. Both cells from **1:** OCI-Ly7 and **2:** SU-DHL-4. Cells are stained with SYBR® Gold DNA staining and viewed at a 40X magnification. The comet tail composes damaged DNA migrated out of the cell under electrical current.

When looking at wild type cells, who received only solvents (cf. Figure 27A) i.e. isotonic saltwater and DMSO, a small comet tail is present, indicative of some DNA fragments having migrated. Cells affected by cisplatin at a concentration of 1.6 µg/mL (cf. Figure 27B), presented with the longest comet tail, and thereby the highest amount of migrated DNA fragments. Cells affected by 17AAG at a concentration of 0.15 µg/mL (cf. Figure 27C) showed a comet tail similar to wild type cells, albeit slightly longer. When looking at Figure 27D, showing cells affected by both cisplatin at a concentration of 1.6 µg/mL and 17AAG at a concentration of 0.15 µg/mL, almost no comet tail is visible, creating a large contrast to cells of Figure 27B. These results indicate that inhibition with 17AAG along with cisplatin, decreases the amount of migrating DNA fragments that would otherwise have been induced by cisplatin.

4.10 DIFFUSE LARGE B-CELL LYMPHOMA PATIENTS

Fourteen cell lines are not representative of the clinical population of DLBCL patients. In silico data from whole exome sequencing of 1,001 de novo DLBCL patients, kindly provided by DaveLab[34] , and whole exome sequencing of 28 Hematology Aalborg patients with relapsed DLBCL and 44 Hematology Aalborg patients with de novo DLBCL (gathered as of the 5th of January 2019), provided the opportunity to visualize the mutations found in DLBCL patients and elucidate possible mutation hotspots. The genes in focus is *XPF*, important in NER, *MSH2*, important in MMR, and *EXO1*, important in MMR and HR. Two variant analyses were carried out by Mads Sønderkær: one on DaveLab's 1,001 patient cohort and one on the Hematology Aalborg 72 patient cohort. From this, variants in *EXO1*, *XPF* and *MSH2* who were non-synonymous and in an exon was plotted on a graphical view of the respective protein, given that the variants could be annotated with a protein change. Variants annotated as both "tolerated" (VEP SIFT) and "benign" (VEP Polyphen) have been excluded.

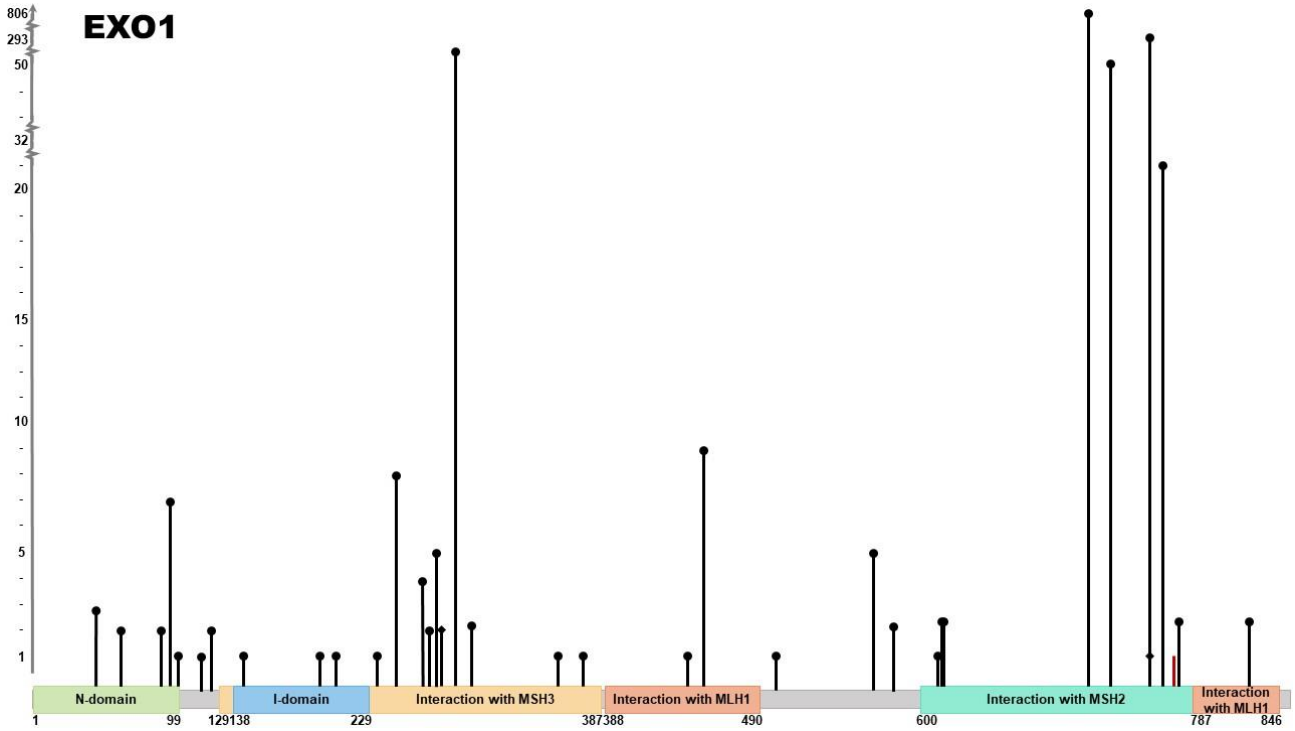


Figure 28: Graphical protein variant view of *EXO1* variants. Start and end of protein domains are marked on the horizontal axis, whereas the vertical axis annotates variant prevalence within the sample group. The spikes with a round head are variants found in the variant analysis on DaveLab's 1,001 de novo DLBCLs, whereas the spikes with a rhomb head are variants found in Hematology Aalborg 72 de novo or relapsed DLBCL patients. The spikes with no head are variants found in cell lines through the CCLE database[45]. The red color of a spike depicts the variant chosen for this study (p.Q774L). Inspired from https://dlbcl.davelab.org[35] and based on data from https://www.uniprot.org/[49].

The graphical protein variant view depicts the length of the protein and its domains, into which each mutation variant is plotted. Each spike represents a variant at the given position; the higher the spike, the more prevalent the variant is in the variant analysis. The spikes with a round head represent the variants found in the DaveLab's 1,001 de novo DLBCLs, whereas the spikes with a rhomb head symbolize the variants from the Hematology Aalborg 72 relapsed DLBCLs. Finally, the

spikes with no head mark the variants found in cell lines on the CCLE database. The red color of a spike marks a variant chosen for this study.

Figure 28 shows the protein EXO1 (exonuclease 1) and the related variants found in the two clinical cohorts and cell lines. Each domain of the protein is marked and named. DaveLab's 1,001 de novo DLBCLs present with a total of 33 unique mutation variants widely scattered across the protein. The most affected domains are the one with MSH3 interaction and the one with MSH2 interaction. Five spikes are present: *EXO1* (p.N279S) found in 54 patients, *EXO1* (p.Q718*) found in 806 patients, *EXO1* (p.D731Y) found in 50 patients, *EXO1* (p.G759E) found in 293 patients, and *EXO1* (p.K768E) found in 21 patients. This is a sharp contrast to the number of variants found in the Hematology Aalborg 72 DLBCL patient cohort; post exclusion, these presented with only *EXO1* (p.N279S) in the first domain with MSH3 interaction (n=2) and *EXO1* (p.G759E) in the domain with MSH2 interaction (n=1). Additionally, the variant initially chosen for this project, *EXO1* (p.Q774L), found in DB, is also found in the domain with MSH2 interaction.

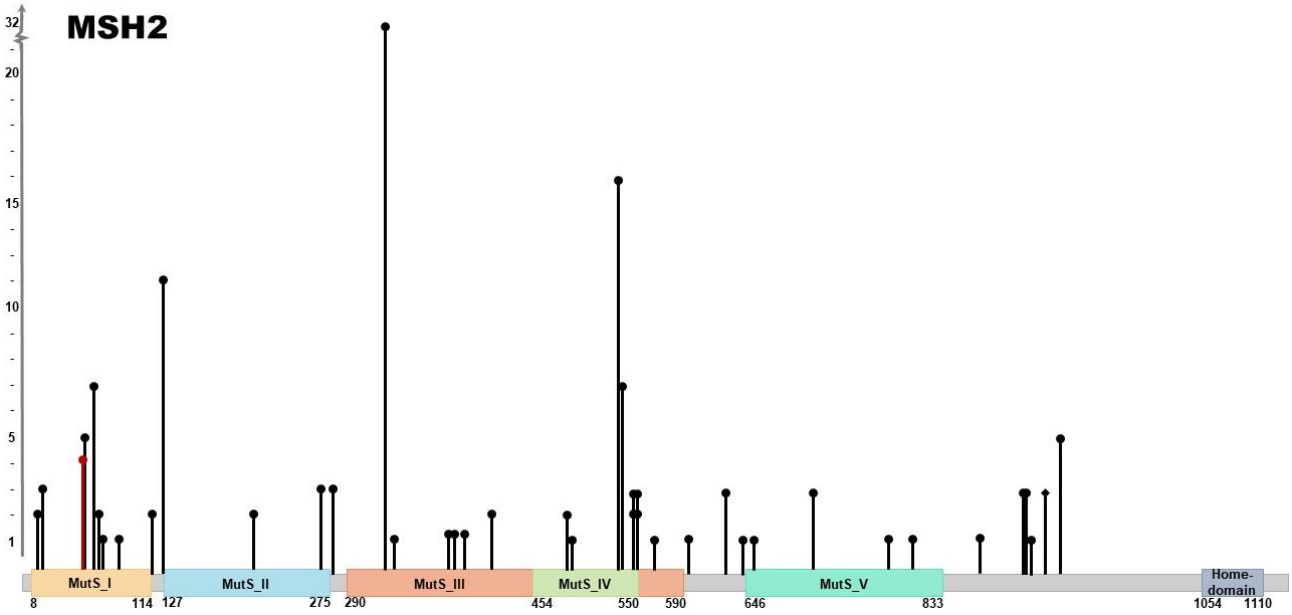


Figure 29: Graphical protein variant view of *MSH2* variants. Start and end of protein domains are marked on the horizontal axis, whereas the vertical axis annotates variant prevalence within the sample group. The spikes with a round head are variants found in the variant analysis on DaveLab's 1,001 de novo DLBCLs, whereas the spikes with a rhomb head are variants found in Hematology Aalborg 72 de novo or relapsed DLBCL patients. The spikes with no head are variants found in cell lines through the CCLE database[45]. The red color of a spike depicts the variant chosen for this study (p.A54T). Inspired from https://dlbcl.davelab.org[35] and based on data from http://pfam.xfam.org/[50].

This tendency is also observed in Figure 29, where DaveLab's 1,001 de novo DLBCLs present with 40 different mutation variants in total, dispersed across MSH2 (MutS protein homolog 2). Three spikes are visible, with 11 patients presenting with *MSH2* (p.N127S), 32 patients presenting with *MSH2* (p.G322D), and 16 patients presenting with *MSH2* (p.V532Kfs*13). The most affected domains are MutS_I and MutS_III. MutS_I is the DNA binding domain, and the domain in which the variant candidate of this project, *MSH2* (p.A54T), is. In contrast, only one variant, *MSH2* (p.Q915R),

remained of the Hematology Aalborg 72 DLBCL patient cohort after exclusion (n=3), located between domains. No variants of *MSH2* was found in the cell lines on the CCLE database[45].

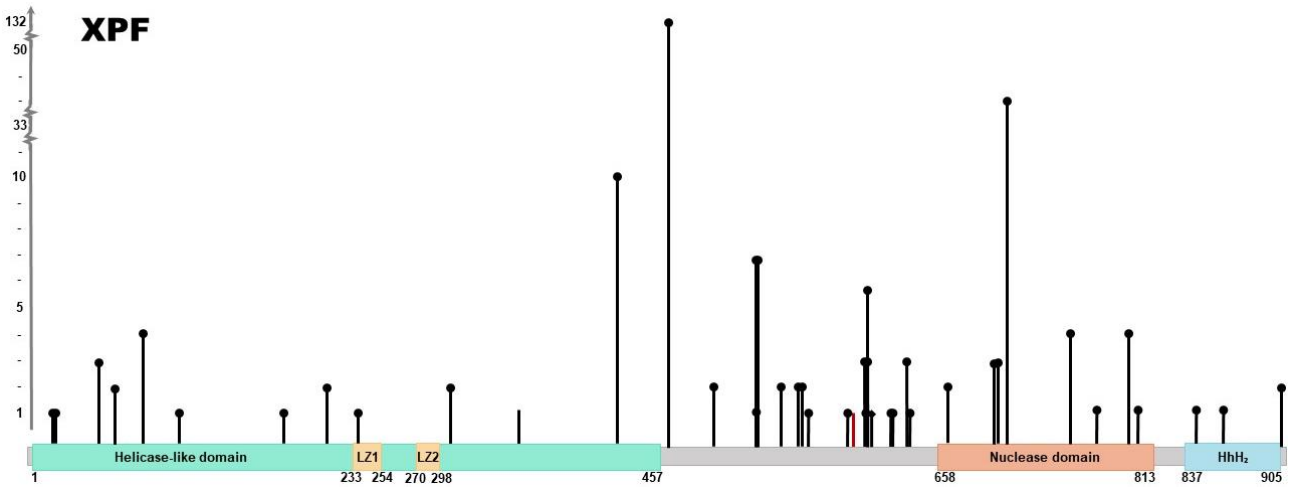


Figure 30: Graphical protein variant view of *XPF* variants. Start and end of protein domains are marked on the horizontal axis, whereas the vertical axis annotates variant prevalence within the sample group. The spikes with a round head are variants found in the variant analysis on DaveLab's 1,001 de novo DLBCLs, whereas the spikes with a rhomb head are variants found in Hematology Aalborg 72 de novo or relapsed DLBCL patients. The spikes with no head are variants found in cell lines through the CCLE database[45]. The red color of a spike depicts the variant chosen for this study (p.A596V). HhH_2: helix-hairpin-helix. LZ: Leucine-zipper. Inspired from https://dlbcl.davelab.org[35] and based on data from https://www.uniprot.org/[51].

The graphical protein variant view of XPF is depicted in Figure 30. A total of 33 unique mutation variants were found in the variant analysis of DaveLab's 1,001 de novo patients. These were spread across the protein, however with a visible accumulation of mutations in the segment between the helicase-like domain and nuclease domain. Within this intermediary segment is also the variant found in FARAGE (*XPF* p.A596V) and the one variant (XPF p.P609H) that remained after filtering the variant analysis of 72 Hematology Aalborg relapsed or de novo DLBCL patients (n=1). Three spikes can be identified of DaveLab's 1,001 de novo patients: p.P425S found in 10 patients, p.R461Q found in 132 patients, and p.S708P found in 48 patients.

4.11 Schematic overview of the project

To provide a schematic overview of both the experimental and the clinical track of this thesis, three diagrams have been drawn. The clinical mutation track is described in Figure 31. Mutations were identified in the 14 cell lines, DaveLab's 1,001 de novo DLBCL patients, and 72 de novo and relapse DLBCL Hematology Aalborg patients, mutation candidates were chosen and these were validated using either ddPCR™ Rare Mutation Detection or bioinformatic variant analysis, generously performed by Mads Sønderkær, Senior Bioinformatician at The Department of Hematology, Aalborg University Hospital.

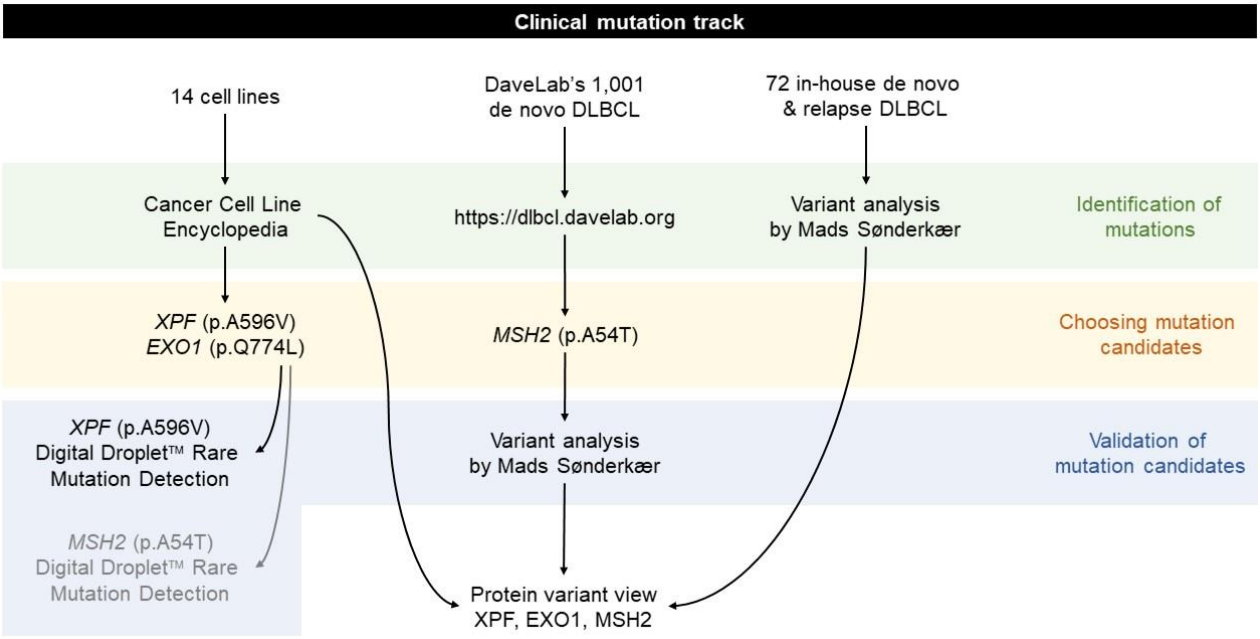


Figure 31: A flow diagram depicting the clinical mutation track of this thesis. Data from three samples were included: 14 cell lines (DB, FARAGE, HBL-1, MC-116, NU-DHL-1, NU-DUL-1, OCI-Ly7, OCI-Ly8, RIVA, SU-DHL-4, SU-DHL-5, SU-DHL-8, SU-DHL-10, and U-2932), 1,001 de novo DLBCL patients provided by DaveLab[34], and 72 de novo and relapsed Hematology Aalborg DLBCL patients. Green shows the three sources from which mutations were identifies. Yellow shows the chosen mutations variants. Blue shows the process for validation. Faded steps have not been performed.

MSH2 (p.A54T) ddPCR™ Rare Mutation Detection has not been performed due to limited time, and the low prevalence in DaveLab's cohort (4 out of 1,001), making it unlikely to be present in the cell lines. From the mutations found in the three populations, a graphical protein variant view was made for XPF, EXO1, and MSH2, visualizing potential hotspots and differences between the three populations.

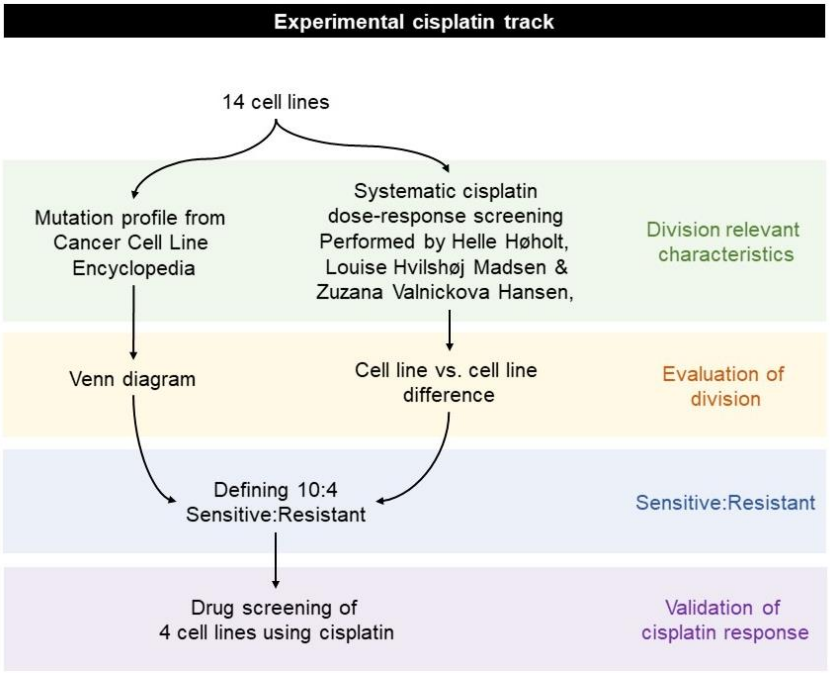


Figure 32: A flow diagram depicting the experimental cisplatin track of this thesis. Green shows the two data sources: the mutation profile of 14 cell lines (DB, FARAGE, HBL-1, MC-116, NU-DHL-1, NU-DUL-1, OCI-Ly7, OCI-Ly8, RIVA, SU-DHL-4, SU-DHL-5, SU-DHL-8, SU-DHL-10, and U-2932) and their cisplatin response from a dose-response screening. Yellow shows the tools used for evaluation of the divisions. Blue defines the chosen division. Purple shows the validation of DB, FARAGE, OCI-Ly7, and SU-DHL-4 cisplatin response.

The experimental cisplatin track is depicted in Figure 32. Using the mutation profile of cell lines found on CCLE[45] and results from a systematic cisplatin dose-response screening, different divisions of cisplatin-sensitive and cisplatin-resistant were made. The cisplatin dose-response assay had been carried out prior to initiation of this project by Helle Høholt, Medical Laboratory Technician at The Department of Hematology, Aalborg University Hospital, Louise Hvilshøj Madsen, Laboratory Technician at The Department of Hematology, Aalborg University Hospital, and Zuzana Valnickova Hansen, Laboratory Technician at The Department of Hematology, Aalborg University Hospital. The different divisions were evaluated using a Venn diagram and the calculated difference in cisplatin response between two adjacent cell lines. From this, a division of Sensitive:Resistant was chosen and the cisplatin response found in the dose-response screening was validated in a drug screening of DB, FARAGE, OCI-Ly7, and SU-DHL-4.

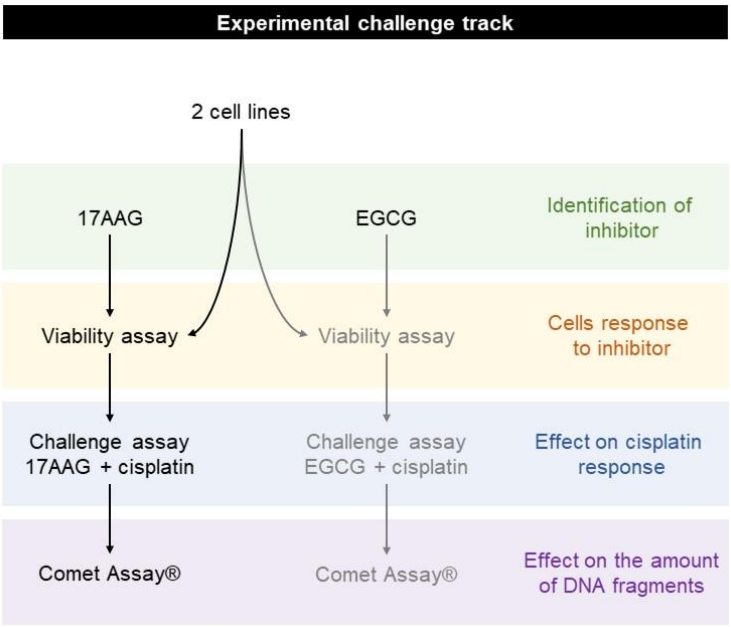


Figure 33: A flow diagram depicting the experimental challenge track of this thesis. Green shows the identification of 17-(Allylamino)-17-demethoxygeldanamycin (17AAG), used as an inhibitor for MSH2, and Green tea polyphenol epigallocatechin-3-gallate (EGCG), used as an inhibitor of ERCC1-XPF complex. Yellow shows the viability assay on OCI-Ly7 and SU-DHL-4. Blue shows the challenge assay. Purple shows the evaluation of the amount of DNA fragments. Faded tracks have not been performed.

The experimental challenge track is shown in Figure 33. 17AAG and EGCG was identified as inhibitors of respectively MSH2, a part of MMR, and ERCC1-XPF a part of NER. A validation assay was carried out on OCI-Ly7 and SU-DHL-4 and a subsequent challenge assay, using the inhibitor and cisplatin, was performed to evaluate the effect of DNA repair inhibition on the cisplatin response. Comet Assay® was used to evaluate the change in the amount of DNA fragments when cells were affected with a DNA inhibitor and cisplatin. At the time of writing this thesis, the steps of EGCG have not been performed, hence the faded track in Figure 33.

5 DISCUSSION

5.1 EXPERIMENTAL TRACK

One of the objectives of this thesis was to determine the frequency and investigate the role of mutations of selected DNA repair genes found in cell lines with cisplatin response varying from sensitive to resistant. No natural subdivision in cisplatin sensitivity was observed between the 14 tested DLBCL cell lines in systematic dose-response screenings. However, based on the difference in ratios between AUC_0 of neighboring cell lines, ranked according to increasing AUC_0, the largest difference was observed between FARAGE and RIVA, except for cell lines in the extremities. This split also separated cell lines with many mutations (Sensitive) from cell lines with extremely few mutations (Resistant) in DNA repair related genes and was therefore chosen as the best suggestion for dividing cell lines into Sensitive and Resistant, yielding the division of 10:4 (Sensitive:Resistant). When running cell count based drug screening on DB, FARAGE, OCI-Ly7, and SU-DHL-4 as validation of previously performed MTS based dose-response screening, unexpected observations was made: SU-DHL-4, classified as Resistant in MTS assays, showed a larger decrease in cell count during drug screening and challenge assay, compared to OCI-Ly7, classified as Sensitive by MTS, indicating a larger cisplatin-sensitivity in SU-DHL-4. In contrast to these results is the fact that when these two cell lines were affected with cisplatin, lysed, and subjected to electrical current in Comet Assay®, a lower amount of DNA fragments were identified in SU-DHL-4 cells, compared with cells of OCI-Ly7, indicating a cisplatin-resistant phenotype in SU-DHL-4, in accordance with the MTS based classification of SU-DHL-4 as Resistant. The previous systematic dose-response screenings on cell lines were MTS-based assays detecting the level of metabolic active cells, whereas drug screenings of this project were cell count-based assays detecting actual living cells, thereby providing different output parameters, based on different stages – and possibly different mechanisms – of growth inhibition: dormant (metabolic inactivity) and death.

DNA Barcoding have authenticated all cell lines' ID before and after assays, ruling out cell contamination or cell line swap. Therefore, the difference in output parameter and persons performing the analysis were considered as reasonable explanatory factors for the observed differences in the response assessment. With these arguments, the division of ten Sensitive and four Resistant cell lines were considered to be acceptable.

It is an interesting observation that cisplatin-resistant cell lines have very few mutations in DNA repair related genes (only two variants in one gene in total), whereas cisplatin-sensitive cell lines show a total of 41 mutation variants in 28 different genes. To the author's knowledge, no previous study has reported the difference in cumulative mutation landscape of cisplatin-sensitive and cisplatin-resistant cell lines of DLBCL. The cisplatin resistance could be due to different molecular mechanisms between the two groups. The markedly high AUC_0 of SU-DHL-10 conferring the most resistant profile, compared to the other cell lines, could be due to the splice site mutation in *TP53* (g.7574034C>T), a pathogenic variant located at the acceptor (3') splice site[52], or possible hypermethylation of the promotor region of e.g. *MSH1* or *MSH2*, described to have a great impact on

the MMR function.[53] This could also be a factor, for the observed relative cisplatin resistance in the other three cell lines with no mutations in DNA repair related genes.
A systematic assessment tool was developed, enabling quantitative evaluation of each variant found in cell lines on the CCLE database[45], and DaveLab's 1,001 de novo DLBCL patients on https://dlbcl.davelab.org[35]. The variants were filtered according to eight weighted assessment criteria, making it possible to assess a) the gene's function in DNA repair important for cisplatin's mechanism of action, b) the empiri concerning the gene or mutation variant, c) two aspects important for variant Impact, and d) the prevalence of the variant in its population-group (cell line, DaveLab[34] cohort, or Hematology Aalborg cohort). From this, three variants were chosen: *EXO1* (p.Q774L), *XPF* (p.A596V), and *MSH2* (p.A54T).

The presence of *XPF* (p.A596V) in FARAGE, was verified using ddPCR™. No other cell line had the mutation, including the four cell lines that were not included in the CCLE database[45]. The four clusters mapped out on the 2D plots of ddPCR™ (cf. Figure 19 and Figure 21) are skewed, rather than at right angles to each other: the p.A596V positive cluster is shifted towards the double positive cluster, and the double positive cluster is stretched out between the p.A596V positive cluster and wild type positive cluster. This is due to cross-reactivity of the probes that is to be expected in Rare Mutation Detection ddPCR™ Assay.[54] These skewed cluster appear as double bands in the 1D plots (cf. Figure 18 and Figure 20), as a result of viewing two-dimensional data in one dimension.[54] At the time of writing this paper, the presence of *MSH2* (p.A54T) in cell lines have not been tested, due to a limited time frame. However, the variant was only found in four out of 1,001 DLBCL patients (DaveLab's cohort) and was not mentioned in ten out of this study's 14 cell lines on CCLE[45], making it highly unlikely that either of the cell lines would be positive for the variant. Despite it being a theoretically interesting *MSH2* variant, it is likely not relevant in DLBCL and cisplatin sensitivity.

Another aim of this thesis was to evaluate how impaired DNA repair affects cell lines' response to cisplatin. The Hsp90 inhibitor, 17AAG, was chosen. Hsp90 (also known as HspC) is a chaperone involved in protein folding and is a positive regulator of e.g. MMR proteins.[24,26] 17AAG binds to the NH_2-terminal ATP-binding domain of Hsp90 and locks it in an ATP-bound formation.[24,26] Tung et al. (2014)[24] found that 17AAG significantly downregulated the pemetrexed-induced MSH2 expression in NSCLC cell lines, and that 17AAG sensitized NSCLC and lung adenocarcinoma cell lines to pemetrexed.[24] However, pemetrexed is an antifolate chemotherapy drug with a downstream inhibition of purine and pyrimidine formation.[24] Therefore, inhibition of MSH2 with 17AAG could be expected to have a different impact when focusing on DLBCL cell lines treated with cisplatin. In this study, administration of 17AAG resulted in a decreased cell count of two DLBCL cell lines (OCI-Ly7 and SU-DHL-4) when administered as a single intervention with increasing concentrations, even at the lowest concentration of 0.15 µg 17AAG per mL. This is observed in other studies as well, and an argument for it being tested in Phase 2 clinical trials as a treatment for cancer.[27,28] However in this thesis, the goal was to evaluate the effect in combination with cisplatin, and it can be argued whether a concentration >0.15 µg 17AAG per mL would be more appropriate in further

studies. 17AAG administered in combination with cisplatin resulted in an increased cell count of the two cell lines, compared to cells only affected by cisplatin. This indicates that administration of 17AAG decreases cisplatin sensitivity of DLBCL cell lines, and shows ineffective MMR, as a result of impaired MSH2, as a possible mechanism for cisplatin-resistance.

A different response was observed by Choi et al. (2014)[25], where a sublethal dose of 17AAG enhanced HR proficient ovarian cancer cell lines' sensitivity to carboplatin, a drug with relatively identical mechanism of action as cisplatin.[25] However, 17AAG could not sensitize a HR deficient ovarian cancer cell line to carboplatin.[25] Additionally, they identified a significantly decreased expression of HR related genes (BRCA1 and RAD51) when treating HR proficient ovarian cancer cell lines with 17AAG[25], in line previous descriptions of Hsp90 and its interaction with HR-proteins, such as MRN complex, BRCA1, BRCA2 and RAD51.[12]

However, in the present study, 17AAG resulted in a decreased cisplatin sensitivity. Furthermore, a noticeable smaller amount of DNA fragments was observed from cells treated with 17AAG and cisplatin, compared to cells affected only by cisplatin, indicating a decreased DNA repair. All supporting the hypothesis of 17AAG inhibiting MSH2 and impairing MMR, leading to a cisplatin-resistant phenotype. Moreover, SU-DHL-4 presented with a smaller amount of DNA fragments when treated with cisplatin, compared with OCI-Ly7 treated with cisplatin, advocating SU-DHL-4 as more cisplatin-resistant than OCI-Ly7 and supporting its classification into Resistant.

Thus, the DNA repair mechanism targeted by mutations is highly important for the effect in cisplatin sensitivity. DLBCL patients with impaired MSH2 function might be more resistant to cisplatin treatment in the relapsed setting and have a poor survival compared to relapse patients with a functional MSH2 – a hypothesis that needs testing in large clinical cohorts.

5.2 CLINICAL TRACK

To evaluate on the mutations in selected DNA repair genes found in cell lines, compared to a cohort of clinical DLBCL patients, in silico data from whole-exome sequencing of DaveLab's 1,001 de novo DLBCL patients and Hematology Aalborg 72 de novo and relapse DLBCL patients was analyzed using a pipeline developed by Mads Sønderkær, Senior Bioinformatician at The Department of Hematology, Aalborg University Hospital.

There was very little overlap in variants identified in cell lines, DaveLab cohort and Hematology Aalborg cohort. The variant analysis of DaveLab cohort showed a markedly higher number of variants in *EXO1*, *XPF*, and *MSH2*, than those found in the variant analysis of Hematology Aalborg cohort. Despite that, none of the CCLE-identified variants in the cell lines were found in either of the variant analyses, illustrating a cell line associated genetic profile, possibly selected through in vitro culturing and long-term passaging since the original isolation from primary patients.

Additionally, when focusing on *EXO1*, *XPF*, and *MSH2*, only one variant, *EXO1* (p.G759E), was found in both variant analyses; one patient out of 72 Hematology Aalborg DLBCL patients presented with the variant, as well as 293 patients out of DaveLab's 1,001 DLBCL patients. Eight variants was found in the variant analysis of DaveLab's cohort to have a high prevalence, these included EXO1 (p.Q718TER) (n=806), *EXO1* (p.G759E) (n=293), *EXO1* (p.N279S) (n=54), *XPF*

(p.R461Q) (n=132), *XPF* (p.S708P) (n=48), *XPF* (p.E921G) (n=33), and *MSH2* (p.G322D) (n=32). These variants are likely natural variants that would be filtered away in Reddy et al. (2017)[34], through their 400 paired-normal tissue. However, since VEP SIFT and Polyphen annotations were not applicable for these variants, they were still included in the graphical protein variant views of this thesis.

Increased sensitivity to cisplatin has been described in relation to NER proteins. Testicular cancer patients have shown a high response rate to cisplatin, leading to the investigation of this cancers molecular background.[11] Usanova et al. (2010)[18] found that a low level of the ERCC1-XPF complex in testis cells correlated with a decreased repair of interstrand crosslinks, and proposed that the ERCC1-XPF complex could be the limiting factor of cisplatin response.[18] Furthermore, Olaussen et al. (2006)[17] investigated 761 NSCLC patients and found that just over half of the patients presented with ERCC1-negative tumors. The patients were treated with cisplatin-based regimes, to which the patients with ERCC1-negative tumors showed a significantly better survival profile than patients with ERCC1-positive tumors.[17] In principle, low expression of any key NER protein, such as XPD, XPF, XPG and ERCC1, could lead to decreased capacity of NER, increasing sensitivity to cisplatin and cell death.[11,13] Likewise, an increased capacity for NER has been linked to cisplatin resistance.[16] Olaussen et al. (2006)[17] found that an overexpression of ERCC1-XPF increased interstrand crosslink repair and decreased tumors sensitivity towards cisplatin. Therefore, the ERCC1-XPF complex seems to play a role in the mechanism of resistance to cisplatin. Furthermore, Trisianes et al. (2005)[55] suggested that XPF supports the protein folding of ERCC1, since ERCC1 was deemed unable to fold correctly without XPF in vitro.[55] They also identified the XPF Phe905 residue as essential to the interaction with ERCC1, since this positions in the hydrophobic pocket of ERCC1.[55] The XPF Ala906 residue has also been linked to the interaction with ERCC1.[56] The HhH_2 (helix-hairpin-helix) domain (837-905) has also been demonstrated to bind both dsDNA and ssDNA.[57] The nuclease domain presents as the active site, and mutations of XPF Asp687, Asp715, Lys727, and Asp731 residues impacted the catalytic activity.[58] No essential residues have been described in the helicase-like domain of XPF.[56]

In the present study, mutations in *XPF* was also present. Especially, the segment between helicase-like domain and nuclease domain of the XPF protein seemed hold many mutations, including the one variant found in Hematology Aalborg 72 patient cohort and one of the two variants found in cell lines. As of yet, no residues in this segment have been highlighted; however, the apparent accumulation of mutation in this thesis' cohort, suggests this segment as a possible region for further inspection in future studies. One of the mentioned residues, Ala906, was found in the variant analysis to be mutated (XPF [p.A906T]) in two patients of DaveLab's 1,001 DLBCL patients. It can be speculated that this mutation would interfere with ERCC1-XPF interaction, thereby hampering the function of NER and result in increased cisplatin sensitivity.

The MMR pathway also plays a role in mediating cisplatin's cytotoxicity, where *MLH1* and *MSH2* have been highlighted, and impaired MMR has been implicated in relation to cisplatin resistance.[11]

Fink et al. (1996)[20] reported a 1.8-fold increased cisplatin resistance in an MSH2-deficient endometrial cancer cell line (HEC59).[20] Likewise, Lage et al. (1999)[21] found that melanoma cells with a reduced nuclear content of MSH2 had a decreased apoptosis and cisplatin resistance.[21] de Miranda et al. (2013)[19] analyzed 29 B-cell lymphomas for mutations in DNA repair related genes. They identified MMR-deficient tumors of DLBCL with recurrent alterations of MMR genes, such as EXO1, MSH2 and MSH6.[19] Hsu et al. (2005)[22] observed promoter hypermethylation of *MLH1* and *MSH2* in 66.7% and 34.3% of 105 NSCLC tumors, respectively, and consequential loss of protein expression.[22] Patients with affected *MSH2* was found to have significantly lower overall survival and cancer-specific survival.[22] In line with this, Reddy et al. (2017)[34] reported that DLBCL patients with mutations of *MSH2* (n=36) had a poor survival, compared to patients without *MSH2* mutation (n=900). Although not significant (p=0.339), the lower 3-year survival rate could indicate an increased chemoresistance in patients with *MSH2* mutations, compared to patients without mutaiton.[34] These observations are in accordance with the findings of this thesis, where a decreased cisplatin sensitivity was observed when inhibiting MSH2, using 17AAG on DLBCL cell lines.

Some studies have opposing results. Kamal et al. (2010)[23] investigated 673 lung cancer patients treated with cisplatin and found a low expression of MSH2 in 62% and a high MSH2 expression in 38%. Controversially, they identified a tendency of prolonged survival in patients with low expression of MSH2.[23] Ollila et al. (2018)[59] investigated the effect of 18 variants of *MSH2* on the MMR efficiency, using protein extracts from Spodoptera frugiperda 9, transfected with *MSH2* cDNA containing mutations of interest, in an MMR assay using a *MSH2*-deficient human sporadic CRC cell line. Two of the 18 variants mentioned in Ollila et al. (2018)[59] were also found in the variant analysis of DaveLab's 1,001 patients, namely *MSH2* (p.N127S) in 11 patients and *MSH2* (p.G322D) in 32 patients. Ollila et al. (2018)[59] also described *MSH2* (p.C333Y), where at the same position *MSH2* (p.C333R) was found in one patient of DaveLab's 1,001 patients. Ollila et al. (2018)[59] described 11 of the variants as MMR deficient, including *MSH2* (p.C333Y), and five as MMR proficient in vitro, including *MSH2* (p.N127S) and *MSH2* (p.G322D).[59] Most of the variants, leading to an unstable MSH2 was found in either the MutS_II domain (connector) or the MutS_III domain (lever).[59] *MSH2* (p.N127S) is in the MutS_II domain, and showed no protein instability.[59] In the MutS_III domain, which is suggested to transmit signals between the DNA-binding domain (MutS_I) and ATPase domain (MutS_V), are *MSH2* (p.G322D) and *MSH2* (p.C333Y). *MSH2* (p.G322D) was found to only slightly reduce MMR, whereas *MSH2* (p.C333Y) displayed lower MSH2 stability and induced defective MMR.[59]

These conflicting results necessitates further studies into the involvement of *MSH2* in cisplatin resistance.

For the variants found in DaveLab's cohort, it should be noted that there is not much agreement between the variants made publicly available as Supplementary Information to Reddy et al. (2017)[34] and the variants found through the in-house variant analysis of the raw data from DaveLab. Generally, the variant analysis found a higher number of variants, probably due to the less rigid variant filtering in this thesis, compared to Reddy et al. (2017)[34]. Variants of this thesis

were not filtered in regard to rarity, repetitive sequence, reading depth, or genome quality. Additionally, Reddy excluded genes in which <20 patients presented with a mutation to ensure possible survival analysis.

Another and more important disagreement was observed when looking at *MSH2:* ten *MSH2* variants included in the Supplementary Information to Reddy et al. (2017)[34] was not identified in the in-house variant analysis of their raw in silico data, including the variant *MSH2* (p.A54T) chosen in this thesis. The graphical protein variant views are based solely on the variant analysis and not the Supplementary Information to Reddy et al. (2017)[34]. However, the variant *MSH2* (p.A54T) has been added to Figure 29 since it was chosen at the beginning of the thesis.

The reason why 10 variants, and probably more if looking at all genes, are listed in the Supplementary Information to Reddy et al. (2017)[34] but could not be found through the in-house variant analysis, needs to be elucidated before any further analysis, and conclusions on these data should be made with caution.

5.3 Conclusion and future perspectives

In conclusion, this thesis found cisplatin-sensitive DLBCL cell lines to have a markedly higher number of mutations in DNA repair related genes compared to cisplatin-resistant cell lines. The important repair mechanisms for mediating cisplatin cytotoxicity (MMR, NER, and HR) was most affected by mutations in the 14 DLBCL cell lines. However, there was very little overlap between the mutations found in cell lines and the ones found in the clinical cohort, leading to the speculation of cell line specific mutations.

17AAG, an MSH2 inhibitor, was found to decrease cisplatin sensitivity of DLBCL cell lines, and decrease the amount of free DNA fragments, as hypothesized. This is different from some previous reports but follows the theory of MMR's role in the mechanism of action of cisplatin and signifies the possible role of MSH2 is cisplatin resistance. Additionally, *MSH2* was found to be fairly mutated in the clinical cohort, with a slight accumulation of mutations in the MutS_I domain. From this thesis, MMR seems an important player in mediating cisplatin resistance. However, the possible link between impaired MSH2 (and consequently MMR) and cisplatin resistance would need to be elucidated in a large clinical cohort of relapsed DLBCL patients. It would be interesting to test the effect of 17AAG on the cisplatin response in cell lines with many mutations, such as DB and FARAGE, and possibly evaluate whether DNA repair proficient mutations outweighs impairment of DNA repair through MSH2.

XPF and *EXO1* were also be mutated in the clinical cohort. Even through, *EXO1* was not prioritized in this study, it would still be interesting to investigate its relation to cisplatin resistance, since its association with cisplatin is sparsely explored, and it is involved in both MMR and HR, two important pathways in relation to cisplatin. Also, targeting EXO1 is associated with increased sensitivity to cisplatin. Likewise, the effect of XPF on the cisplatin resistance would be interesting to explore. EGCG, is a partially reversible inhibitor of ERCC1/XPF activity in vitro, which affects interstrand crosslink repair, and possibly leading to increased cisplatin sensitivity. A hypothesis, that would be interesting to investigate in the DLBCL cell lines of this study.

Lastly, the variant analyses of the clinical cohorts hold great potential, once the disagreements in variants have been clarified. The dataset could help in answering if DLBCL patients present with a different mutation landscape than solid tumors, of which most research into DNA repair mutations and cisplatin resistance is currently based on. It would be of great importance if a difference in mutation landscape was observed, since advances in the empiri of solid tumors (resistance and alternative treatments) would not be directly translatable to hematological cancers, and further studies into treatment and resistance in hematologic cancers, or specifically DLBCL, would have to be carried conducted.

6 ACKNOWLEDGEMENTS

This project was conducted and written by Ann-Katrine Hyttel Rom Pedersen as a Master in Biomedicine (9th and 10th semester), Medicine with Industrial Specialization at Aalborg University and completed in collaboration with the Department of Hematology, Aalborg University Hospital.

I would like to extend my appreciation to the staff at the Department of Hematology, Aalborg University Hospital, for the way they included me in their everyday routine and did not hesitate to ask questions or lend a helping hand. I want to send a special thanks to Helle Høholt and Louise Hvilshøj Madsen for their patience and impeccable guidance in the laboratories, as well as their genuine curiosity and interest in my project. Also, Mads Sønderkær, for generating the variant analyses on DNA sequencing of clinical DLBCL patients; I really appreciate it, I am sure it was not easy, as was answering all my questions that followed.

A sincere thanks to Karen Dybkær Sørensen, Julie Støve Bødker, and Hanne Due Rasmussen for their tenacity when we had to piece together a Master project in a matter of days, their continues guidance, rewarding discussions, and dedication from start to finish. You allowed me to control most of the project, and it has been a stressful, frustrating, educational, and wonderful experience that I am grateful for.

Last but not least, I send a heartfelt gratitude to the man who stood by me through all the frustrating moments, who listened when I chatted on about mutation variants, inhibitors, and cell line responses, and supported me when I just needed time to work. Lasse, you have been my rock this past year, thank you for all you have done for me and our son.

7 REFERENCES

1. Li S, Young KH, Medeiros LJ. Diffuse large B-cell lymphoma. *Pathology*. 2018;50(1):74-87. doi:10.1016/j.pathol.2017.09.006
2. Gascoyne RD, Campo E, Jaffe ES, et al. Diffuse large B-cell lymphoma, NOS. In: Swerdlow SH, Campo E, Harris NL, et al., eds. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. 4th ed. International Agency for Research on Cancer (IARC); 2017:291-297.
3. Morin R, Mungall K, Pleasance E. Mutational and structural analysis of diffuse large B-cell lymphoma using whole genome sequencing. *Blood*. 2013;122(7):1256-1265. doi:10.1182/blood-2013-02-483727.The
4. Afdeling for Klinisk Epidemiologi og Biostatistik, Den landsdækkende LYFO Database. *Malignt Lymfom Og CLL - National Årsrapport 2016*.; 2016.
5. Brown P de N, Madsen J, Jørgensen J, et al. *Diffust Storcellet B-Celle Lymfom*.; 2019.
6. Dansk Lymfomgruppe. *Retningslinjer for Diagnostik Og Behandling Af DIFFUST STORCELLET B-CELLE LYMFOM (DLBCL)*.; 2015.
7. Mikhed Y, Görlach A, Knaus UG, Daiber A. Redox regulation of genome stability by effects on gene expression, epigenetic pathways and DNA damage/repair. *Redox Biol*. 2015;5:275-289. doi:10.1016/j.redox.2015.05.008
8. Curtin NJ. DNA repair dysregulation from cancer driver to therapeutic target. *Nat Rev Cancer*. 2012;12(12):801-817. doi:10.1038/nrc3399
9. Sierra Oncology. DNA Damage Respons (DDR) Network. https://www.sierraoncology.com/science/ddr/. Accessed February 27, 2019.
10. Branzei D, Foiani M. Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol*. 2008;9(4):297-308. doi:10.1038/nrm2351
11. Rocha CRR, Silva MM, Quinet A, Cabral-neto JB, Menck FMC. DNA repair pathways and cisplatin resistance : an intimate relationship. *Clinics*. 2018;73(Sippl 1):e478s. doi:10.6061/clinics/2018/e478s
12. Wang D, Lippard SJ. Cellular processing of platinum anticancer drugs. *Nat Rev Drug Discov*. 2005;4(4):307-320. doi:10.1038/nrd1691
13. Martin LP, Hamilton TC, Schilder RJ. Platinum resistance: The role of DNA repair pathways. *Clin Cancer Res*. 2008;14(5):1291-1295. doi:10.1158/1078-0432.CCR-07-2238
14. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):D1074-D1082. doi:10.1093/nar/gkx1037
15. Sawant A, Kothandapani A, Zhitkovich A, Sobol RW, Patrick SM. Role of mismatch repair proteins in the processing of cisplatin interstrand cross-links. *DNA Repair*. 2015;35:126-136. doi:10.1016/j.dnarep.2015.10.003.
16. Stewart DJ. Mechanisms of resistance to cisplatin and carboplatin. *Crit Rev Oncol Hematol*. 2007;63(1):12-31. doi:10.1016/j.critrevonc.2007.02.001
17. Olaussen K, Dunant A, Fouret P, et al. DNA Repair by ERCC1 in Non–Small-Cell Lung Cancer and Cisplatin-Based Adjuvant Chemotherapy. *N Engl J Med*. 2006;355(10):983-992.
18. Usanova Svetlana, Piée-Staffa Andrea, Sied Ulrike, et al. Cisplatin sensitivity of testis tumour cells is due to deficiency in inters...: University of Liverpool Library. *Mol Cancer*. 2010;9(1):248. http://eds.a.ebscohost.com.liverpool.idm.oclc.org/eds/detail/detail?vid=7&sid=54640a51-2244-4607-8d7f-

1635c3de025b%40sessionmgr4010&hid=4113&bdata=JnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ%3D%3D#AN=99b4fc79a3b0b37677421e18b75f37ec&db=edsdoj.
19. de Miranda NF, Peng R, Georgiou K, et al. DNA repair genes are selectively mutated in diffuse large B cell lymphomas. *J Exp Med*. 2013;210(9):1729-1742. doi:10.1084/jem.20122842
20. Fink D, Nebel S, Aebi S, et al. The Role of DNA Mismatch Repair in Platinum Drug Resistance. 1996;(619):4881-4886.
21. Lage H, Christmann M, Kern MA, et al. EXPRESSION OF DNA REPAIR PROTEINS hMSH2 , hMSH6 , hMLH1 , O 6 -METHYLGUANINE-DNA METHYLTRANSFERASE AND N-METHYLPURINE-DNA GLYCOSYLASE IN MELANOMA CELLS WITH. *Int J Cancer*. 1999;750(September 1998):744-750.
22. Hsu H, Wen C, Tang Y, et al. Human Cancer Biology Promoter Hypermethylation Is the Predominant Mechanism in hMLH1 and hMSH2 Deregulation and Is a Poor Prognostic Factor in Nonsmoking Lung Cancer. 2005;11(15):5410-5417.
23. Kamal NS, Soria J, Mendiboure J, et al. MutS Homologue 2 and the Long-term Benefit of Adjuvant Chemotherapy in Lung Cancer. 2010;16(8):1206-1216. doi:10.1158/1078-0432.CCR-09-2204
24. Tung CL, Chiu HC, Jian YJ, et al. Down-regulation of MSH2 expression by an Hsp90 inhibitor enhances pemetrexed-induced cytotoxicity in human non-small-cell lung cancer cells. *Exp Cell Res*. 2014;322(2):345-354. doi:10.1016/j.yexcr.2014.02.002
25. Choi YE, Battelli C, Watson J, et al. Sublethal concentrations of 17-AAG suppress homologous recombination DNA repair and enhance sensitivity to carboplatin and olaparib in HR proficient ovarian cancer cells. *Oncotarget*. 2014;5(9):2678-2687. doi:10.18632/oncotarget.1929
26. Sottile ML, Nadin SB. Heat shock proteins and DNA repair mechanisms: an updated overview. *Cell Stress Chaperones*. 2018;23(3):303-315. doi:10.1007/s12192-017-0843-4
27. U.S. National Library of Medicine. ClinicalTrials.gov. NIH. https://clinicaltrials.gov/. Accessed May 25, 2019.
28. Holzbeierlein JM, Windsperger A, Vielhauer G. Hsp90 : A Drug Target ? 2010:95-101. doi:10.1007/s11912-010-0086-3
29. Heyza JR, Arora S, Zhang H, et al. Targeting the DNA Repair Endonuclease ERCC1-XPF Epigallocatechin-3-Gallate (EGCG) and Its Prodrug to Enhance Cisplatin Efficacy in Human Cancer Cells. 2018. doi:10.3390/nu10111644
30. Lecumberri E, Marc Y, Miralbell R, Pichard C. Green tea polyphenol epigallocatechin-3-gallate (EGCG) as adjuvant in cancer therapy. *Clin Nutr*. 2013;32(6):894-903. doi:10.1016/j.clnu.2013.03.008
31. Marques SC, Ranjbar B, Laursen MB, et al. High miR-34a expression improves response to doxorubicin in diffuse large B-cell lymphoma. *Exp Hematol*. 2016;44(4):238-246.e2. doi:10.1016/j.exphem.2015.12.007
32. Falgreen S, Laursen MB, Bødker JS, et al. Exposure time independent summary statistics for assessment of drug dependent cell line growth inhibition. 2014.
33. Developmental Therapeutics Program. COMPARE Introduction, Methodolegy. National Cancer Institute, Division of Cancer Treatment and Diagnosis. https://dtp.cancer.gov/databases_tools/docs/compare/compare_methodology.htm. Published 2015. Accessed April 10, 2019.
34. Reddy A, Zhang J, Davis NS, et al. Genetic and Functional Drivers of Diffuse Large B Cell

Lymphoma. *Cell*. 2017;171(2):481-494.e15. doi:10.1016/j.cell.2017.09.027
35. Reddy A, Zhang J, Davis NS, et al. 1K DLBCL Study. DaveLab. https://dlbcl.davelab.org/. Published 2017. Accessed May 13, 2019.
36. Smith J, Mun Tho L, Xu N, A. Gillespie D. *The ATM-Chk2 and ATR-Chk1 Pathways in DNA Damage Signaling and Cancer*. Vol 108. 1st ed. Elsevier Inc.; 2010. doi:10.1016/B978-0-12-380888-2.00003-0
37. Leeksma OC, De Miranda NF, Veelken H. Germline mutations predisposing to diffuse large B-cell lymphoma. *Blood Cancer J*. 2017;7(2):e532-9. doi:10.1038/bcj.2017.15
38. Zhang F, Shen M, Yang L, et al. Simultaneous targeting of ATM and Mcl-1 increases cisplatin sensitivity of cisplatin-resistant non-small cell lung cancer. *Cancer Biol Ther*. 2017;18(8):606-615. doi:10.1080/15384047.2017.1345391
39. Tomé S, Holt I, Edelmann W, et al. MSH2 ATPase domain mutation affects CTG•CAG repeat instability in transgenic mice. *PLoS Genet*. 2009;5(5). doi:10.1371/journal.pgen.1000482
40. Edelbrock MA, Kaliyaperumal S, Williams KJ. Structural, molecular and cellular functions of MSH2 and MSH6 during DNA mismatch repair, damage signaling and other noncanonical activities. *Mutat Res Mol Mech Mutagen*. 2013;743-744:53-66. doi:10.1016/j.mrfmmm.2012.12.008
41. National Center for Biotechnology Information. NM_000251.2(MSH2):c.160G>A (p.Ala54Thr). ClinVar; Variation ID 229782. https://preview.ncbi.nlm.nih.gov/clinvar/variation/229782/. Accessed January 26, 2019.
42. Due H, Schönherz AA, Ryø L, et al. MicroRNA-155 controls vincristine sensitivity and predicts superior clinical outcome in diffuse large B-cell lymphoma. *Blood Adv*. 2019;3(7):1185-1196. doi:10.1182/bloodadvances.2018029660
43. Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA Repair Genes. *Science (80-)*. 2001;291(5507):1284-1289. doi:10.1126/science.1056154
44. MDAnderson Cancer Center. Human DNA Repair Genes. https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html. Published 2014. Accessed April 9, 2019.
45. Barretina J, Caponigro G, Stransky N. The Cancer Cell Line Encyclopedia enables predictive modeling of anticancer drug sensitivity. *Nature*. 2012;483(7391):603-607. doi:10.1038/nature11003.The
46. Liberzon A. Gene Set: HALLMARK_DNA_REPAIR. Gene Set Enrichment Analysis. http://software.broadinstitute.org/gsea/msigdb/cards/HALLMARK_DNA_REPAIR.html. Published 2015. Accessed May 16, 2019.
47. Liberzon A, Birger C, Ghandi M, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2016;1(6):417-425. doi:10.1016/j.cels.2015.12.004.The
48. Ensemble. Ensembl Variation - Calculated variant consequences. https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html. Published 2019. Accessed May 1, 2019.
49. UniProtKB. UniProtKB - Q9UQ84 (EXO1_HUMAN). UniProt. https://www.uniprot.org/uniprot/Q9UQ84#family_and_domains. Published 2019. Accessed May 13, 2019.
50. El-Gebali S, Mistry J, Bateman A, et al. Protein: A0A0V0RV23_9BILA (A0A0V0RV23). Pfam. http://pfam.xfam.org/protein/A0A0V0RV23_9BILA. Published 2018. Accessed May 13, 2019.
51. UniProtKB. UniProtKB - Q92889 (XPF_HUMAN). UniProt.

https://www.uniprot.org/uniprot/Q92889#similar_proteins. Published 2019. Accessed May 13, 2019.
52. U.S. National Library of Medicine. rs587782272, build: 152. NIH. https://www.ncbi.nlm.nih.gov/snp/rs587782272?embedded=true&app_context=Variation_Viewer_2-1&id=NC_000017.11#variant_details. Published 2018. Accessed April 28, 2019.
53. Hsu HS, Wen CK, Tang YA, et al. Promoter hypermethylation is the predominant mechanism in hMLH1 and hMSH2 deregulation and is a poor prognostic factor in nonsmoking lung cancer. *Clin Cancer Res*. 2005;11(15):5410-5416. doi:10.1158/1078-0432.CCR-05-0601
54. Bio-Rad Laboratories Inc. *Droplet Digital™ PCR Apllications Guide*.
55. Tripsianes K, Folkers G, Ab E, et al. The Structure of the Human ERCC1 / XPF Interaction Domains Reveals a Complementary Role for the Two Proteins in Nucleotide Excision Repair. 2005;13:1849-1858. doi:10.1016/j.str.2005.08.014
56. McNeil EM, Melton DW. DNA repair endonuclease ERCC1-XPF as a novel therapeutic target to overcome chemoresistance in cancer therapy. *Nucleic Acids Res*. 2012;40(20):9990-10004. doi:10.1093/nar/gks818
57. Das D, Folkers GE, Dijk M Van, Jaspers NGJ, Hoeijmakers JHJ, Kaptein R. Article The Structure of the XPF-ssDNA Complex Underscores the Distinct Roles of the XPF and ERCC1 Helix-Hairpin-Helix Domains in ss / ds DNA Recognition. *Struct Des*. 2012;20(4):667-675. doi:10.1016/j.str.2012.02.009
58. Enzlin JH, Scha OD. The active site of the DNA repair endonuclease XPF ± ERCC1 forms a highly conserved nuclease motif. 2002;21(8).
59. Ollila S, Bebek DD, Jiricny J, Nyström M. Mechanisms of Pathogenicity in Human MSH2 Missense Mutants. 2008;29(110300):1355-1363. doi:10.1002/humu.20893