Generation of HOXB13^{G84E} cellular models with prime editing

Master's Thesis

University of Turku
Department of Life Technologies
Cell Biology

May 2021

Laura Laihonen

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

UNIVERSITY OF TURKU
Department of Life Technologies
LAURA LAIHONEN, Generation of HOXB13^{G84E} cellular models with prime editing
Master`s thesis, 58 p. Appendix 2 p.
Cell biology
May 2021

Prostate cancer (PCa) is the second most common cancer in men and the fifth leading cause of cancer-related death in men. PCa is a highly heritable disease and genetic factors explain 58 % of the risk. Studies on PCa have led to identification of germline mutations associated with PCa, and the G84E mutation in HOXB13 is one of these identified mutations. This germline mutation was identified almost a decade ago in 2012, however, its function and mechanism in PCa progression remain elusive.

Prime editing is a novel genome editing method where new genetic information is added into a specific DNA site with the prime editor (PE) containing catalytically impaired Cas9 endonuclease fused to reverse transcriptase (RT) and green fluorescent protein (GFP). An additional prime editing guide RNA (pegRNA) contains the template with the desired edit and targets the PE to the specific site in the genome.

The aim of the thesis was to generate the G84E mutation in HOXB13 with prime editing. Three different pegRNAs with different lengths of primer binding sites (PBS) were generated to see which is the most efficient. Single guide RNA (sgRNA) was used to nick the non-edited strand which favors the DNA repair of the non-edited strand forcing the cells to use the newly PE edited strand as a template. Protein lysates of PCa cell lines were prepared and analyzed with Western blot to see if they express HOXB13. LNCaP cells expressing HOXB13 were transfected with plasmid constructs and GFP-positive cells were sorted. Gene editing efficiency was measured with droplet digital PCR (ddPCR). pegRNA with a PBS of 16 nucleotides (nt) worked with approximately 10 % efficiency, but the efficiency of pegRNA with PBS of 13 nt remains unclear. PBS of 10 nt did not work at all suggesting that it is too short for proper binding to the target DNA. In conclusion, pegRNA with PBS16 can be used to generate HOXB13G84E cellular model for further studies of the mutation to understand its role in PCa progression.

Keywords: Prostate cancer, HOXB13, prime editing

I would like to thank Genetic cancer predisposition -research group, and especially my supervisor Ph.D. Christoffer Löf for the guidance, and M.Sc. Nasrin Sultana for the help. Special thanks also for Prof. Johanna Schleutker for the opportunity to do my Master's thesis in the group.

Contents

Abbreviations

A	adenine
ABE	adenine base editor
ABL	tyrosine-protein kinase ABL1
ADT	androgen deprivation therapy
AML	acute myeloid leukemia
APC	adenomatous polyposis coli
AR	androgen receptor
ARE	androgen response element
ATM	ATM Serine/Threonine Kinase
BCR	breakpoint cluster region protein
BH	bridge helix
bp	base pair
BRCA	breast cancer gene
C	cytosine
Cas	CRISPR associated endonuclease
CBE	cytosine base editor
CHEK2	checkpoint kinase 2
CIP2A	cancerous inhibitor of PP2A
co-IP	co-immunoprecipitation
CRISPR	clustered regularly interspaced short palindromic repeats
CRPC	castration-resistant prostate cancer
crRNA	CRISPR-RNA
CYP17A1	cytochrome P450 family 17 subfamily A member 1
dDNA	donor DNA

ddPCR	droplet digital polymerase chain reaction
DHT	dihydrotestosterone
DNA	deoxyribonucleic acid
DRE	digital rectal examinations
DSB	double strand break
EDTA	ethylenediaminetetraacetic acid
EMT	epithelial-mesenchymal transition
ER	estrogen receptor
FDA	U.S. Food and Drug Administration
G	guanine
GFP	green fluorescent protein
GM-CSF	granulocyte-macrophage colony stimulating factor
gRNA	guide RNA
HBB	hemoglobin subunit beta
HDR	homology directed repair
HEK293T	human embryonic kidney cells
HEXA	hexosaminidase subunit alpha
HOX	homeobox
iap	alkaline phosphatase
IGF1R	insulin-like growth factor 1 receptor
KLK3	kallikrein 3
mCRPC	metastatic castration resistant prostate cancer
MEIS	murine ectopic integration site
MGE	mobile genetic element
MRI	magnetic resonance imaging

MSH2	MutS homolog 2
NFW	nuclease free water
NHEJ	nonhomolohous end joining
nt	nucleotide
NUC	nuclease lobe
PAM	protospacer adjacent motif
PAP	prostatic acid phosphatase
PBS	primer binding site
PBX	pbx homeodomain protein
PCa	prostate cancer
PCR	polymerase chain reaction
PE	prime editor
pegRNA	prime editing guide RNA
PI	PAM-interacting domain
PIN	prostatic intraepithelial neoplasia
PKNOX	PBX/Knotted 1 homeobox
PP2A	protein phosphatase 2A
pRB	retinoblastoma protein
PSA	prostate specific antigen
PTEN	phosphate and tensin homolog
PVDF	polyvinylidene fluoride
REC	recognition lobe
RIPA	radioimmunoprecipitation assay
RNA	ribonucleic acid
RNASEL	ribonuclease L

RPKM	reads per kilobase million
RT	reverse transcriptase
SDS	sodium dodecyl sulfate
sgRNA	single guide RNA
SNV	single nucleotide variant
SRSR	short regularly spaced repeats
SSR	short-sequence DNA repeat
T	thymine
TALE	three amino acid loop extension
TALEN	transcription activator-like effector nuclease
TGIF	TGFB induced factor homeobox
TP53	tumour protein 53
tracrRNA	trans-activating crRNA
TSG	tumour suppressor gene
WT	wild type

Introduction

The development of cancer is a complex, multi-step process with different factors contributing to the progression. Deregulation of cellular genetics is one of the eight Hallmarks of cancer (Hanahan & Weinberg, 2011). Inactivating loss of function mutations in tumour suppressor genes (TSGs), which normally control cell cycle and promote apoptosis, and activating gain of function mutations in oncogenes, which promote cell cycle progression and inhibit apoptosis, are typically contributing to carcinogenesis (Strachan & Read, 2019). There are several factors increasing cancer risk, and genetic factors can be considered as one of these.

Prostate cancer (PCa) is the second most common cancer in men worldwide, and it is a highly heritable disease (Hjelmborg et al., 2014; Rawla, 2019). PCa research and has led to identification of several genetic factors contributing to PCa progression and these have benefited PCa patients and led to increased life expectancy due to improved screening. Despite the improvements, PCa still remains one of the most common causes of cancer-related deaths among men (International Agency for Research on Cancer (IARC), World Health Organization (WHO)).

Studying PCa genetics is crucial for developing treatments and for identification of new screening targets. Therefore, cellular models of the mutations predisposing to cancer are required for studying the function of these mutations in the progression of the disease. Different genetic tools have been used over the past decades with Clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9 endonuclease (CRISPR-Cas9) being among the most commonly used ones now, but development of new precise tools is required for better editing efficiencies and for minimizing undesired off-target effects. The aim of this thesis was to make tools for generating the G84E mutation in *HOXB13* with prime editing. These tools can later be used to generate cellular model of the mutation for further studies to reveal its function and mechanism in PCa progression.

1. Review of the Literature

1.1 Prostate cancer

1.1.1 Epidemiology

PCa is the most common cancer diagnosed among men in Finland with more than 5 000 cases diagnosed per year (The Finnish Cancer Registry, 2018 report). It is the second leading cause of cancer-related death in men with approximately 900 deaths per year even though the mortality rate is relatively low with five-year survival rate of 93 % (The Finnish Cancer Registry, 2018 report). Worldwide, PCa is the second most common cancer and the fifth leading cause of cancer-related death in men (IARC, WHO).

1.1.2 Risk factors

PCa risk increases with age and cases with early onset tend to be more aggressive than cases with late onset (Grönberg et al. 1999). PCa prevalence varies among groups with different ethnicity and the rates are the highest in African Americans. This is explained by both biological and socioeconomic factors: prostate specific antigen (PSA) levels were higher in African American compared to others, and they might not receive as high-quality healthcare and are less likely to undergo screening. (Kolonel et al., 2004; Vijayakumar et al., 1998.)

Family history is also one of the known risk factors of PCa. Men with a first-degree relative with PCa diagnosis at age <60 have over 2-fold risk of developing PCa. Familial PCa has also been associated with early onset and more aggressive cancer. (Chen et al. 2008.)

Migration studies have indicated that PCa risk increased among Asians who had moved to the United States (Lee et al., 2007). This suggests that dietary factors have an influence on PCa risk, because the Western countries tend to have a diet higher in fat and red meat. However, only a few dietary factors have been established so far to increase PCa risk. Dairy, red meat and saturated fats have been associated with increased risk (Aune et al., 2015; Chan et al., 2005). Other lifestyle factors that increase the PCa risk are low physical activity, smoking and obesity (Kaaks & Stattin, 2010).

1.1.3 Genetics

Hanahan and Weinberg described the Hallmarks of Cancer theory (Figure 1) consisting of properties needed for a normal cell to transform into a cancerous cell. Genetic alterations is one of these hallmarks. Other hallmarks are resisting cell death, deregulating cellular energetics, sustaining proliferative signaling, evading growth suppressors, avoiding immune destruction, enabling replicative immortality, tumour-promoting inflammation, activating invasion and metastasis, and inducing angiogenesis. (Hanahan & Weinberg, 2011.)

Figure 1 The hallmarks of cancer. *Different types of cancers have individual properties, and they develop through different mutations or alterations in epigenetic modifications. However, tumours possess similar properties that either can cause the development of cancer or are a consequence of cancer. These common properties are called hallmarks of cancer. Modified from Hanahan & Weinberg (2011).*

There are several types of genomic alterations that can affect entire chromosomes or distinct genes (Figure 2). Chromosomal rearrangements are insertions, deletions, inversions, duplications and translocations whereas alterations in DNA sequence are deletions, insertions and substitutions. These genetic alterations can be somatic or inherited in the germline.

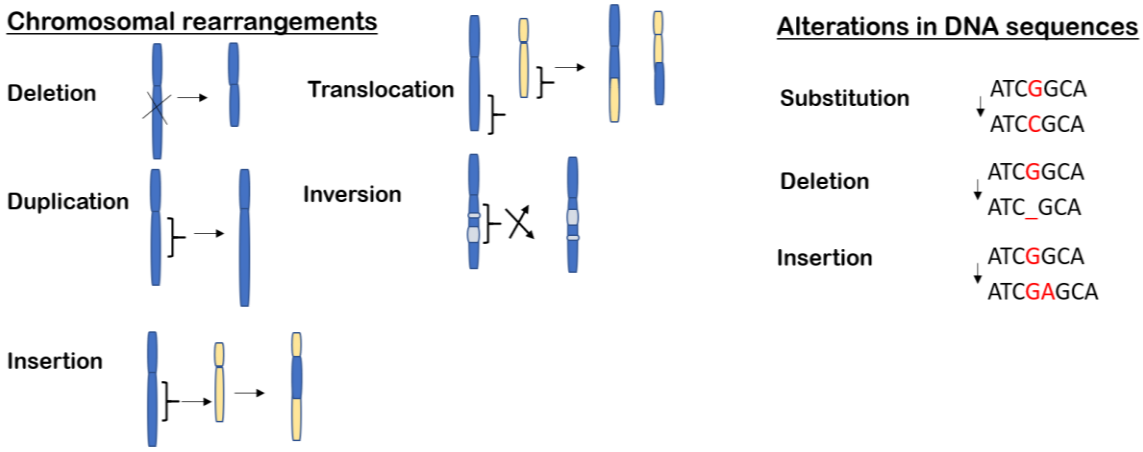


Figure 2 Different types of genetic alterations. *Genetic alterations can affect entire chromosomes, individual genes and other parts of DNA sequences. Deletion, duplication, insertion, translocation and inversion are types of chromosomal rearrangements, whereas substitution, deletion and insertion are types of alterations in DNA sequence.*

Genetic alterations enable the acquisition of many of the hallmarks described previously. Mutations may give cells a growth advantage or then increase the mutation rate by destabilizing the genome. Cancers develop in stages by accumulating random mutations and by generating a diverse heterogeneous cell population until ultimately, one cell acquires the capability to form a tumour. The mutations which are responsible for tumorigenesis are called "driver mutations" and the background mutations are called "passenger mutations". The genes that gain these driver mutations in oncogenesis are divided into TSGs and oncogenes. (Strachan & Read, 2019.)

Oncogenes are mutated proto-oncogenes with a natural function of promoting proliferation. In benign cells, proto-oncogenes are regulated and are activated in response to specific signals. However, if a gain-of-function mutation occurs, control over regulation is lost and this will lead to uncontrollable cell growth and proliferation. These activating gain-of-function mutations can be caused by point mutations or by amplification of an oncogene. (Strachan & Read, 2019.) Translocations can also lead to formation of a novel chimeric gene which can be more active than the original gene. The Philadelphia chromosome is the most well-known example of this situation. The Philadelphia chromosome is a result from translocation of chromosomes 22 and 9. Breakpoint cluster region (*BCR*) gene from chromosome 22 and *ABL,* a proto-oncogene encoding a tyrosine kinase from chromosome 9, are brought together by a translocation leading to formation of a chimeric *BCR-ABL1* fusion gene. (Strachan & Read, 2019.)

TSGs are the second major class of genes mutated in cancers. Their normal function is to control cells by suppressing cell division, maintaining genomic integrity and inducing apoptosis. Unlike oncogenes, TSGs promote cancer progression by acquiring loss-of-function mutations that inactivate them. Most TSGs require inactivating mutations in both alleles of the gene to lose their proper function (Figure 3). (Strachan & Read, 2019; Weinberg, 1991.)

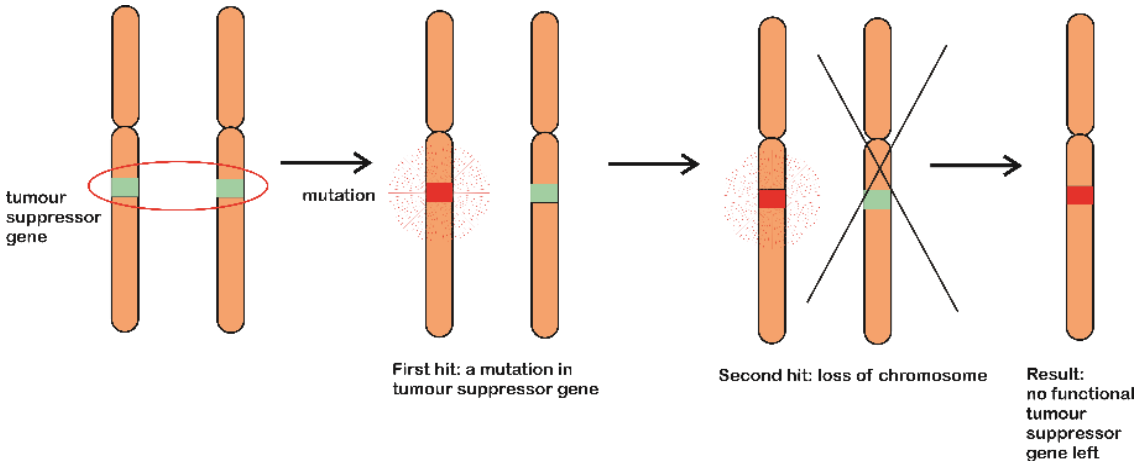


Figure 3 Knudson's two-hit hypothesis. *For a tumour suppressor gene (TSG) to promote carcinogenesis, usually two hits are required for both alleles of the gene in order for a gene to become nonfunctional. Loss of function may result from mutations or chromosome loss.*

Epigenetic modifications are reversible, and they do not alter the DNA sequence. The most common modifications are DNA methylation and post-translational modifications in histone proteins (Figure 4). Gene expression can be regulated by these epigenetic modifications and therefore, they can promote cancer progression by silencing TSGs or by upregulating oncogenes. (Sharma et al., 2010; Albany et al., 2011)

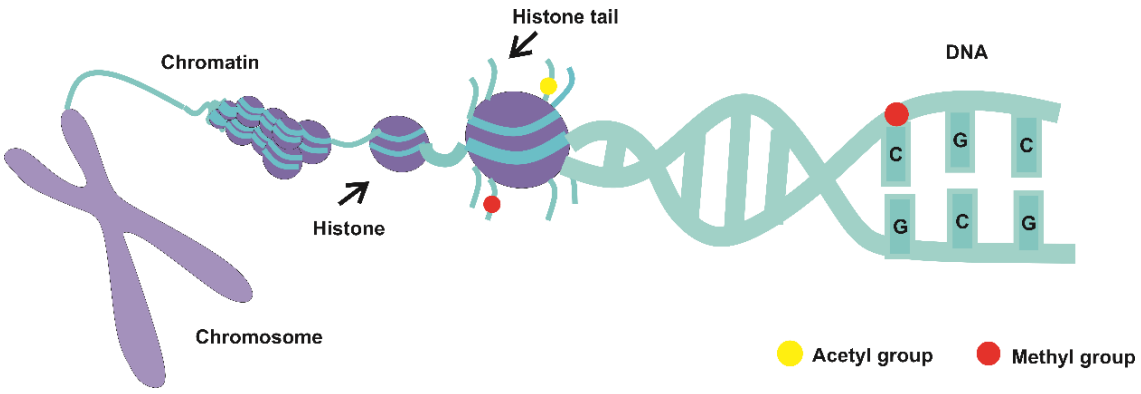


Figure 4 Epigenetic modifications. *Epigenetic modifications are heritable changes in the expression of genes, which do not involve alterations in the DNA sequence. Epigenetic modifications involve changes in the chromatin structure, posttranslational histone modifications, DNA methylation and noncoding RNAs.*

PCa is a highly heritable disease with genetic factors explaining 58 % of the risk (Hjelmborg et al., 2014). Germline mutations, which are mutations located in the germ cells and are inherited, have been associated with more aggressive disease and poorer outcome (Pritchard et al., 2016). Breast cancer genes *BRCA1* and *BRCA2* are both TSGs participating in the DNA repair process, and loss of their proper function leads to a defect in the repair process of double-strand breaks (DSB) called homology directed repair (HDR) process. Mutations in *BRCA1* and *BRCA2* have been associated with PCa susceptibility. The risk of PCa by the age of 80 years varies from 19 % to 61 % for mutated *BRCA2* carriers and from 7 % to 26 % for *BRCA1* (Lecarpentier et al., 2017). Germline mutations in *BRCA2* are also enriched in metastatic PCa cases. *CHEK2* (checkpoint kinase 2), which encodes a checkpoint kinase, *MSH2* (MutS homolog 2), which encodes a DNA mismatch repair protein, *ATM* (ATM serine/threonine kinase), which encodes a serine/threonine protein kinase, *RNASEL* (ribonuclease L), which encodes ribonuclease L inducing apoptosis, and *HOXB13* (homeobox b13), which encodes transcription factor needed in embryonic development, have also been associated with PCa susceptibility (Pritchard et al., 2016). Mutations in *BRCA2* do not appear to have as important role in the Finnish population (Ikonen et al., 2003). The most relevant PCa susceptibility genes in Finland are *RNASEL, CHEK2* and *HOXB13*, and the germline variant G84E in *HOXB13* has the strongest impact on PCa susceptibility (Laitinen et al., 2013).

Somatic mutations arise spontaneously and can be induced by different environmental factors such as chemicals or radiation. Somatic mutations driving PCa progression usually occur in genes participating in androgen signaling pathways and DNA repair pathways, however, mutations in epigenetic regulators and spliceosome pathway have also been identified in PCa cases (Armenia et al., 2018). Somatic mutations found in metastatic castration resistant prostate cancer (mCRPC) cases frequently occur in *AR* (androgen receptor), *TP53* (tumour protein 53), *PTEN* (phosphatase and tensin homolog), *APC* (adenomatous polyposis coli) and *BRCA2* (Robinson et al., 2015).

1.1.4 Prostate anatomy and tumorigenesis

The human prostate is located below the urinary bladder and it surrounds the urethra. The prostate consists of three different histological zones: peripheral zone, transition zone and central zone (Figure 5). The peripheral zone is the major component of prostate, constituting about 70 % of tissue, and it is the origin site of majority of prostate cancers. The transition zone constitutes approximately 5 % of the prostate and is the common site

where hyperplasia develops. The central zone is not a cancer origin site, but it can be secondarily involved in the disease progression. (McNeal, 1981; Ittmann, 2018.)

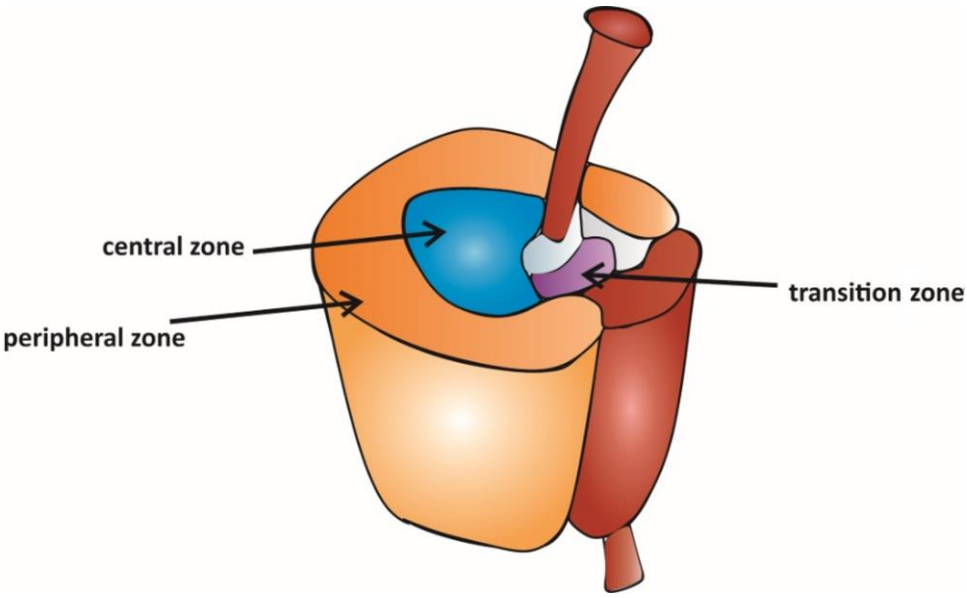


Figure 5 Histological zones of the prostate. *The prostate is divided into three zones: peripheral, transition and central zone. The peripheal zone is the major component and comprises approximately 70 % of the tissue, whereas the transition zone accounts for approximately 5 % of the prostate. Adapted from Hedayat & Lapraz (2019).*

The prostate glands consist of three different epithelial cell types: basal, luminal and neuroendocrine cells. The cell types are defined by their gene expression profiles, surface antigens, shape and position. Basal epithelial cells express cytokeratin 5 and p63, which is a transcription factor, and luminal epithelial cells express cytokeratin 8 and kallikrein 3 (KLK3), a secretory protein regulated by androgens. (Henry et al., 2018). The luminal cells have a secretory function and are responsible for producing mucus, which forms most of the semen. Neuroendocrine cells are rarer, and they express endocrine markers, but do not possess androgen receptors such as basal and luminal cells. (Shen & Abate-Shen, 2010.)

The development of PCa is a multi-step process. Cells start to proliferate in an uncontrolled manner inside the prostate gland and prostatic intraepithelial neoplasia (PIN) is formed. In this stage, the basement membrane remains uninvaded. More changes in the gene expression pattern occurs and the next step is the formation of low-grade carcinoma where the basement membrane is invaded by rapidly dividing luminal cells. The next step is high-grade carcinoma and ultimately, metastasis. (Shen & Abate-Shen, 2010.) Different molecular processes occur in each step of cancer progression. The typical

processes in initiation of PCa are inflammation, DNA damage and telomere shortening, and as it progresses, cellular senescence and re-activation of developmental signaling pathways occur in cancer cells. At the stage of metastasis, signaling processes allowing epithelial-mesenchymal transition (EMT) are activated, and this is the stage where the cancer usually becomes castration-resistant. During the whole process, different oncogenes are overexpressed and TSGs are inactivated or down-regulated. (Shen & Abate-Shen 2010.)

1.1.5 Diagnostics

Early detection of PCa is one of the key aspects to increase survival rate, but since PCa does not always cause any symptoms or then the symptoms can be similar with other urinary problems, it may lead to late detection of the disease. Measuring the PSA levels in blood is currently the most sensitive marker available for screening PCa. PSA is a serine protease encoded by the *KLK3* gene and secreted by prostate epithelial cells. (Schröder, 2010.) The normal PSA serum concentration for healthy men is 0 – 2,5 ng/ml, and the concentration is elevated in men suffering from PCa. PSA-level in blood is proportional to both the clinical stage of PCa and to the volume of the tumour within the prostate. (Stamey et al. 1987.) PSA-testing is also used in monitoring the treatment responses because it indicates the effectiveness of the treatment. Despite the fact that PSA-testing is widely used now, it causes controversy since the downside of it is overdiagnosis due to poor specificity, which results in aggressive treatments of cancer cases that would not be lethal. (Sharma et al., 2017)

The gold standard for PCa diagnosis is prostate biopsy. Magnetic resonance imaging (MRI) and ultrasound are used to target the biopsy to the cancer site. The biopsy samples are histologically evaluated with Gleason scale, where cancer grade is evaluated with numbers between three to five (Gleason, 1966). The Gleason score combines two numbers, the most common and the second most common Gleason grades, and it can vary between six and ten, and the higher the score is, the more aggressive the cancer is (Sharma et al., 2017).

PSA is currently the only biomarker used in PCa detection and disease monitoring, however systematic PSA-testing is recommended only for men with family history of PCa and who are therefore at a high risk of developing PCa. The main benefits of PCa-screening are the reduction of unnecessary and invasive biopsies and digital rectal examinations (DRE), and early detection of the disease. PSA-analysis however has

limitations due to its poor specificity, and this is why finding new cancer-specific biomarkers have been a great interest of researchers. (Sharma et al., 2017)

1.1.6 Treatment

In the case of localized PCa, monitoring, surgery and radiation are the primary treatment methods. In radical prostatectomy, the prostate and its surrounding tissues and seminal vesicles are removed. Prostatectomy is usually recommended for intermediate- and high-risk patients whereas radiation therapy for those with shorter life expectancy. Other treatments for local PCa cases are laser ablation, cryotherapy and high-intensity-focused ultrasound. (Litwin & Tan, 2017.)

In the case of metastatic PCa, patients usually receive androgen deprivation therapy (ADT) either by surgical or chemical castration to decrease the amount of circulating testosterone. In most of the cases where patient receives ADT, a resistance is developed, and this results in cancer progression towards castration-resistant prostate cancer (CRPC). (Nevedomskaya et al., 2018.)

Huggins and Hodges discovered the central role of androgen signaling in PCa in 1941, and it is now clear that the majority of prostate tumours express androgen receptor (AR) (Dai et al., 2017). The development of CRPC is usually result of restored androgen signaling and it is treated with AR-targeted drugs to achieve androgen deprivation with antagonists of the AR ligands, or by reducing androgen synthesis. Antagonists of dihydrotestosterone (DHT) are used for blocking the AR function, and androgen synthesis is usually reduced with Cytochrome P450 Family 17 Subfamily A Member 1 (CYP17A1) lyase inhibitors. If the disease continues to progress despite the medication, cytostatic drugs and second-generation AR antagonists are given to the patients. (Dai et al., 2017; Nevedomskaya et al., 2018.)

mCRPC is treated with additional therapies which include taxanes, radium-223 and sipuleucel-T. Radium-223 is used for patients with bone metastasis in combination with other drugs. (Nevedomskaya et al., 2018.)

Docotaxel and cabazitaxel are chemotherapeutic drugs approved by the U.S. Food and Drug Administration (FDA) in 2004 and 2010, respectively. They are used in the treatment of mCRPC and their mechanism of action is preventing mitotic cell division by inhibiting tubulin depolymerization thus resulting in cell death. Cabazitaxel is a second-line drug and is used after docotaxel treatment. Taxanes have been a gold standard in

mCRPC treatment, but second-generation antiandrogens are now preferred since taxanes have more side effects. (Nevedomskaya et al., 2018.)

Sipuleucel-T is used for autologous cellular immunotherapy and it was approved in 2010. In this immunotherapy, peripheral blood mononuclear cells are obtained from the patient and cultured *ex vivo* in the presence of prostatic acid phosphatase (PAP) fused to granulocyte-macrophage colony stimulating factor (GM-CSF) to achieve an immune response. PAP is almost exclusively expressed by prostate cells and therefore it is an excellent target. This activated sipuleucel-T product is then returned back into the patient intravenously. Clinical benefits and only few mild side effects have been reported, but the costs of this therapy are high, and it is also laborious and therefore its usage has been limited. (Nevedomskaya et al., 2018.)

Novel and effective drugs are still needed for the treatment of PCa. With personalized precision medicine paving the way, best treatment strategies for individuals are determined. The costs however are higher, and it is more laborious to perform individual treatments compared to conventional therapies. Identification of novel biomarkers also is one of the major aspects in the field of PCa diagnostics and treatment since they would provide aid for selecting optimal drug combinations for PCa patients by helping to predict the drug response and the disease progression. (Nevedomskaya et al., 2018; Rice et al., 2019.)

1.2 HOXB13 and PCa

1.2.1 The homeobox gene family

HOX genes are a highly conserved subgroup of the homeobox superfamily. HOX genes encode proteins that are transcription factors critical for normal embryonic development and master regulators which can either activate or repress other genes. In 1921, Morgan and Bridges were the first ones to propose existence of HOX genes. They observed changes in body structure development in *Drosophila melanogaster* mutants and therefore suggested that there are genes responsible for correct spatial body development in fruit flies. Despite the findings of Morgan and Bridges, HOX genes and their protein structure were confirmed in humans nearly 70 years later in 1990 (Nourse et al., 1990; Kamps et al., 1990).

The common characteristic of HOX genes is the 120 base pair (bp) DNA sequence, homeobox sequence, encoding the homeodomain. Homeodomain is the DNA-binding domain, and it is located in exon 2. (Bhatlekar et al., 2014.) HOX genes are located within clusters in vertebrates, and the amount of these clusters vary according to the anatomic complexity of the organism. There are four clusters in four different chromosomes and 39 identified HOX genes in humans (Figure 6). (Shah & Sukumar, 2010.)

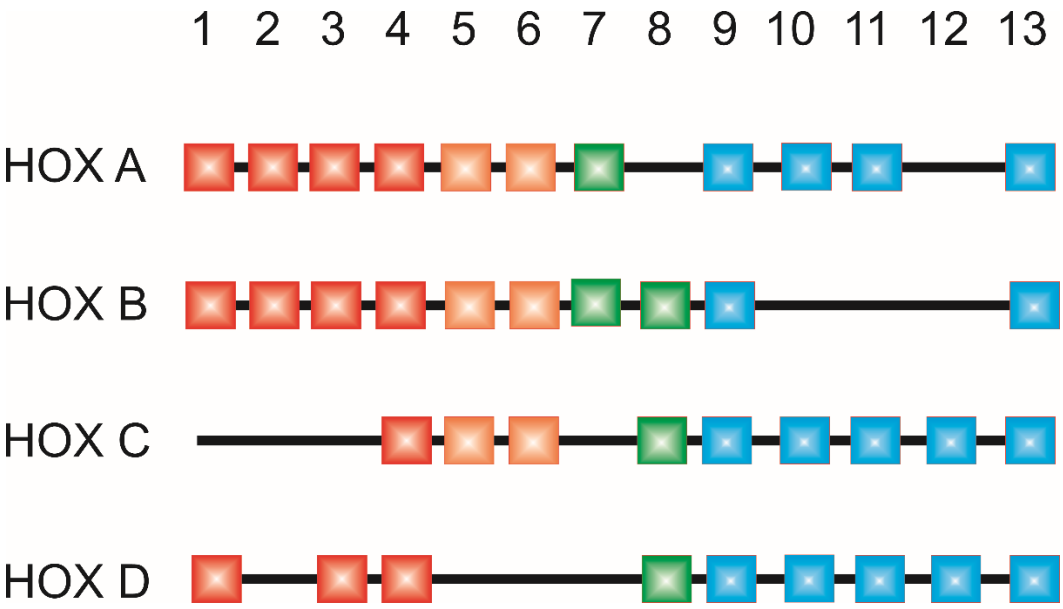


Figure 6 Classification of the 39 identified HOX genes in humans. *There are 39 identified HOX genes located in four different clusters and chromosomes in humans. Cluster A is in chromosome 7 and has 11 HOXA genes, cluster B is in chromosome 17 and has 10 HOXB genes, cluster C is in chromosome 12 and has 9 HOXC genes, and cluster D is in chromosome 2 and has 9 HOXD genes. The genes illustrated in red and orange (HOXA-D 1-6) are expressed in the most anterior parts of the human body such as face and neck, those illustrated in green (HOXA-D 7-8) in abdomen, and those illustrated in blue (HOXA-D 9-13) are expressed in the posterior parts of the body such as limbs and genitals. Adapted from Features of the animal kingdom: Figure 4, OpenStax College, Biology.*

The nomenclature of HOX genes is based on the cluster where it is located. The four clusters are A for chromosome 7, B for chromosome 17, C for chromosome 12 and D for chromosome 2. The genes are numbered between 1 - 13 according to their position, starting from the ends where genes expressed in anterior parts of the body are located. (Scott, 1992.)

There are three basic precepts in what mechanisms the HOX genes control and are controlled in normal development. First is the spatial collinearity, meaning that the

position of a gene 3' to 5' within a cluster corresponds to its expression along the anterior-posterior (A-P) axis. Genes located in 3' are generally expressed in anterior tissue whereas 5` genes in the posterior tissues. Second is the posterior prevalence, meaning that the HOX genes positioned in 5' cluster have a more dominant phenotype. Third is the temporal collinearity, meaning that the HOX genes are expressed in order corresponding to their position (3' to 5'). (Shah & Sukumar 2010.)

HOX genes have been linked to various cancers because their protein products promote carcinogenesis by activating oncogenes or by suppressing TSGs. This indicates that both up- and downregulation of HOX genes can be critical for cancer progression. (Bhatlekar et al., 2014.)

1.2.2 HOX genes in carcinogenesis

Many of the molecular pathways that are associated with normal growth and development are also associated with carcinogenesis. The first suggestions of HOX genes and their role in carcinogenesis were that they promote cancer progression by being upregulated in cancer cells, but now it is known that the actual mechanisms are far more complex. The aberrant expression of HOX genes in cancer can be classified into three different mechanisms. The first is that the HOX genes can be re-expressed in tumour cells originated from embryonic cells where the gene is normally expressed during development. The second is that the HOX genes can also be expressed in tumour cells originated from cells in which the HOX gene is not normally expressed, but these cases seem to be rarer compared to the previous mechanism. Third is that HOX genes are downregulated in those tumour cells that originate from a tissue where a HOX gene is expressed even in the differentiated state. (Abate-Shen, 2002.) In cancerous tissue, changes in gene expression patterns, especially increased expression levels, usually occur with those HOX genes that have oncogenic functions, whereas HOX genes that have tumour suppressing properties are usually silenced. (Shah & Sukumar, 2010.)

Epigenetic regulation of HOX genes have been shown in the case of primary squamous cell carcinoma of the lung. Rauch et al. reported that methylation analysis of *HOXA* genes (*HOXA7* and *HOXA9*) revealed highly methylated CpG islands within the gene promoters in stage 1 tumours. Cases where upregulation of HOX genes in lung carcinomas occur have also been reported. Abe et al. showed that the expression levels of *HOXA1, A5, A10* and *HOXC6* were higher in squamous cell carcinoma tissues when compared with normal tissue. They also suggested that the HOX genes are involved in the histological diversity

of lung carcinomas because they compared HOX gene expressions in squamous cell carcinoma and adenocarcinoma tissues, and the results showed elevated expression of *HOXA1, HOXD9, D10* and *D11* in squamous cell carcinoma.

Overexpression, downregulation and epigenetic silencing of HOX genes have also been reported in breast cancer. Wu et al. demonstrated that *HOXB7* overexpression occurs in primary breast tumours and bone metastases, and it was shown to induce EMT in epithelial cells. *HOXB13* overexpression has also been reported in breast cancer tissue compared to normal counterpart (Cantile et al., 2003). HOXA10 has been suggested to regulate p53 expression. Its downregulation results in downregulation of p53, a tumour suppressor protein, in estrogen receptor (ER) -positive tumours, implicating that downregulation of *HOXA10* promotes cancer progression (Chu et al. 2004). As well as in the case of lung cancer, epigenetic silencing of HOX genes have been shown to be present in breast cancer cases also. Reisman et al. suggested that methylation of *HOXA5* promoter would lead to downregulation of the gene expression and ultimately to loss of p53 expression in breast cancer.

Overexpression of HOXA9 has been associated with acute myeloid leukemia (AML) (Silverman et al., 2002; Faber et al., 2009). The exact mechanism how HOXA9 drives the proliferation of leukemia cells is not fully known. However, HOXA9 overexpression seems to upregulate insulin-like growth factor 1 receptor (IGF1R) in a B-lineage acute lymphoblastic leukemia cell line, which would indicate that the proliferation is a result from an autocrine cause (Whelan et al., 2008). *HOXA9* can also promote AML by forming chimeric fusions resulting in upregulation of genes needed for cell proliferation like murine ectopic integration site 1 (*MEIS1*), a binding partner of HOXA9 (De Braekeleer et al., 2014).

Elevated expression of HOXC genes has been reported in cervical cancer progression, and *HOXC10* seems to take part in the progression of a high grade squamous intraepithelial lesion into invasive carcinoma by increasing the motility and invasiveness of cancerous epithelial cells (Zhai et al., 2007). Other HOX genes have also been associated with cancer progression in normal cervical cells (Hung et al., 2005). The mechanism how HOX genes promote the progression of invasive carcinoma is not known and more research would be needed.

Several studies have implicated the role of HOX genes in ovarian carcinomas (Cheng et al., 2005; Miao et al., 2007; Idaikkadar et al., 2019). Upregulated expression *of HOXA5,*

A9, HOXB2, B5, B6, B7 and *HOXD1*, and downregulation of *HOXC6* was observed in microarray gene expression analysis where non-malignant and malignant ovarian tissues were analyzed (Bahrani-Mostafavi et al., 2008). Similar results indicating overexpression of HOX genes in ovarian cancer were obtained from a study where real-time quantitative PCR (RT-qPCR) assay was performed to create an expression profile of HOX genes using ovarian cancer cells and ovarian derived materials (Yamashita et al. 2006). Results from a study performed by Yamashita et al. revealed overexpression of 14 different HOX genes in the A and B clusters, and two genes in the C and D clusters. From these 16 genes, *HOXB7, HOXA13* and *HOXB13* were overexpressed almost exclusively in cancer cells and tissues, but only rarely or none in control cells and tissues.

Aberrant expression of HOX genes in the prostate has been suggested to result in loss of differentiation and tumorigenesis. Waltregny et al. showed that *HOXC8* is overexpressed in prostate adenocarcinoma and the levels correlate with loss of differentiation. *HOXC4, C5* and *C6* overexpression in malignant prostate cell lines and lymph node metastases has also been confirmed with RT-qPCR (Miller et al., 2003). *HOXB13* is needed in the normal development of the human prostate, and its aberrant expression is associated with tumorigenesis (Jung, Kim, Zhang et al., 2004; Jung, Kim, Lee et al., 2004). However, the biological function of the gene in normal prostate development and PCa progression remains elusive.

These findings indicate the numerous roles of HOX genes in cancer progression. Even though scientists have long been aware of HOX genes and their importance in normal embryonic development and in tumorigenesis, their complex mechanisms remain unclear. This indicates the importance of studying these genes in both normal and cancer tissue. Understanding downstream targets and regulators of HOX genes would help to understand their mechanism of action in normal development, and once their normal function is understood, their mechanisms in cancer progression can be elucidated.

1.2.3 HOXB13 and its role in cancer progression

HOXB13 is located in chromosome 17q21.32 and it consists of two exons (Figure 7). Since it is located at the 5' end of the HOXB cluster and is one of the posterior HOX genes, it is expressed in the most posterior systems of the human body, and the reproductive tract is one of those. In normal conditions, *HOXB13* expression is restricted to the prostate and colon tissue, however, aberrant expression is associated to different types of cancers. In addition to breast and ovarian cancer (see section 1.2.2), deregulated

expression of *HOXB13* has been reported in melanoma, colon and renal cancers (Brechka et al., 2017).

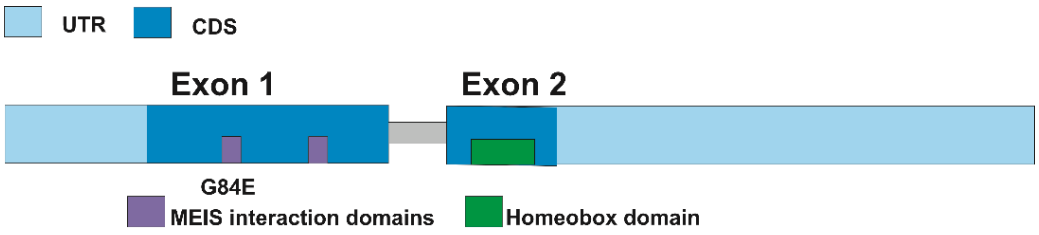


Figure 7 Illustration of HOXB13 gene*. The gene consists of two exons. Exon 1 is positioned at 157 – 757 base pairs (bp) and contains two MEIS interaction domains. Exon 2 is positioned at 1707 – 1960 bp and contains the DNA-binding homeobox domain. HOXB13 transcript is 3987 bp long and the protein is 284 amino acids long. The germline mutation G84E is located in the exon 1 at the first MEIS interaction domain at position 84 of the protein sequence (NP_006352.2, ProtID Q92826). UTR=untranslated region, CDS=codine sequence, MEIS=murine ectopic integration site. Adapted from Brechka et al. (2017).*

Unlike *HOXA13* and *HOXC13* that are expressed mostly during development in murine prostate, *HOXB13* is expressed into adulthood in multiple species (Brechka et al., 2017). Because *HOXB13* is highly expressed into adulthood and it is the most differentially expressed of the HOX proteins when compared between lobes of the rodent prostate, it is suggested that it could have more important function in determining prostatic identity and maintaining organ homeostasis (Huang et al., 2007). In human prostate, *HOXB13* is expressed in prostate luminal epithelial cells. In rodent models, it is expressed in the ventral prostate lobe and it is required for normal differentiation and secretory functions (Economides & Capecchi, 2003).

Homeodomains of HOX genes bind to AT-rich DNA sequences, but these motifs are very abundant across the genome and therefore HOX proteins require additional cofactors for specificity. (Mann et al., 2009.) The best-known cofactors of HOX proteins are three amino acid loop extension (TALE) homeodomain proteins. MEIS, PKNOX (PBX/Knotted 1 homeobox), TGIF (TGFB Induced Factor Homeobox) and PBX (Pbx homeodomain protein) proteins belong to TALE superfamily, and they contain three amino acid insertions allowing them to cooperatively bind to other cofactors, which is one of the characteristics of TALE proteins. (Brechka et al., 2017.) Many mutations in *HOXB13* that are associated with increased prostate cancer risk are located within the MEIS interaction domains. MEIS proteins interact with HOX proteins and recruit other

proteins and form multimeric complexes at gene promoters, thus regulating gene expression. However, TALE proteins also function independently from HOX proteins implicating that their role in disease development is extending beyond HOX protein regulation. (Mann et al., 2009.)

The relationship between HOXB13 and AR has been studied since they both are expressed in prostate cells. However, the data obtained from these studies is quite controversial. The expression of *HOXB13* is thought to be independent from androgen signaling as shown by studies made with mice, where *HOXB13* mRNA accumulation remained following castration (Sreenath et al., 1999). HOXB13 also regulates cellular responses to androgen by interacting with the DNA binding domain of AR, and this interaction leads to inhibition of transcription of the genes containing androgen response element (ARE). In addition to its inhibitory effects, HOXB13 and AR can form a complex and bind to promoters specific for HOXB13 (Norris et al., 2009). HOXB13 can also inhibit PCa cell growth by functioning as a repressor of AR signaling. Jung et al. reported that HOXB13 suppressed AR activity in a dose-responsive manner and that it also suppressed the growth of LNCaP cells, which is an androgen-sensitive cell line, whereas in AR-negative CV-1 cells, it did not have similar effects. *HOXB13* has also been shown to contribute to cell cycle arrest. Ectopic expression of *HOXB13* decreased Cyclin D1 levels by inducing its degradation by ubiquitination, resulting in cell growth inhibition due to reduced phosphorylation of the retinoblastoma protein (pRB) and stabilization of the pRB-E2F complex. Consistent with these results, depletion of *HOXB13* expression led to an increase in cyclin levels, stabilization of E2F and increase in pRB phosphorylation. (Hamid et al., 2014.) On the contrary, under androgen-free conditions, expression of *HOXB13* seems to inhibit the expression of p21, a TSG, leading to promotion of E2F signaling and cell cycle progression (Kim et al., 2010). Studies performed by Sipeky et al. have also demonstrated the synergistic effects of *HOXB13* and *CIP2A* (cancerous inhibitor of PP2A), inhibitor of tumour suppressor *PP2A* (Protein Phosphatase 2A). HOXB13 binds to *CIP2A* and promotes its transcription leading to PCa progression. These observations implicate that *HOXB13* has both oncogenic and tumour suppressing functions.

The germline mutation HOXB13(G84E) was identified by Ewing et al. in 2012 within familial prostate cancers. This mutation is located in exon 1 and is a result of guanine (G) to adenine (A) substitution at position 251 of the DNA sequence. This substitution leads to a missense mutation, where glycine is substituted with glutamic acid at position 84 of

the protein sequence. This mutation is in the first MEIS interaction domain, however, Johng et al. demonstrated in their studies, that this germline mutation does not interfere with the HOXB13-MEIS1 interaction. They also demonstrated with expression analyses and co-immunoprecipitation (co-IP) assays, that both *HOXB13* and *MEIS1* are expressed and form a stable complex in both normal and PCa epithelial cells.

The G84E mutation is a Finnish founder mutation, and it has been shown to significantly increase PCa risk (Ewing et al., 2012; Xu et al., 2013). The frequency in Finnish familial prostate cancer cases is 8,4 %, and it contributes to early onset of cancer and high PSA levels (Laitinen et al., 2013). The mutation has also been associated with early onset of familial PCa cases worldwide (Breyer et al., 2012). Analysis performed by Breyer et al. also revealed that the mutation carrier frequency was higher among cases with higher Gleason score, but the results were not statistically significant. Studies performed by Storebjerg et al. however concluded that the germline mutation carriers in Danish men had higher Gleason score and that the mutation may be associated with aggressive PCa.

Other *HOXB13* variants (Y88D, L144P, G216C, R217C, and R229G) that increase PCa risk have been identified in U.S. Caucasians (Ewing et al., 2012), and other germline mutations have been identified in non-Northern Europeans. Because the G84E mutation is mostly present in the European and Caucasian populations, studies about other *HOXB13* mutations predisposing to PCa in non-Caucasian populations are needed. (Brechka et al., 2017.)

Even though the *HOXB13(G84E)* variant was identified almost a decade ago, its function and mechanism in PCa development remains elusive. It has been suggested to result in increased HOXB13 protein stability which would increase the transcription of its downstream targets (Chandrasekaran et al., 2017). Cardoso et al. reported that the variant phenotype does not differ from wild type (WT) in apoptosis or proliferation rates, indicating that there is still controversy in the field regarding the molecular function of this variant.

This is why functional studies on the mutated *HOXB13* are needed to elucidate the transcriptional impacts of its target genes and how this leads to tumour initiation within the prostate.

1.3 CRISPR-Cas9

1.3.1 History

CRISPR-Cas9 is an adaptive immune system against invading viruses in bacteria and archaea. CRISPR are the most widely distributed family of short noncoding DNA repeats among prokaryotes. The common features of these repeats are that they are composed of direct repeats and are separated by nonrepetitive spacer sequences, which are often located adjacent to *cas* (CRISPR-associated) genes. (Treangen et al., 2009.)

The first findings of CRISPR-sequences were reported in 1987 in studies with the *iap* (alkaline phosphatase) gene of *Escherichia coli* (Ishino et al., 1987). Approximately ten years later, these repetitive sequences were described as Short-sequence DNA Repeat (SSR) (Van Belkum et al., 1998). The term CRISPR became generally accepted in 2002 among scientists working on the subject (Jansen et al., 2002).

In 2000, Mojica et al. suggested that these Short Regularly Spaced Repeats (SRSRs), as they described them, in bacteria and archaea are functionally related, but their physiological relevance was not known until in 2005 thanks to DNA sequencing. Mojica et al. showed that CRISPR spacers are derived from already existing sequences which are chromosomal or within transmissible genetic elements. These extrachromosomal elements showed a relationship between CRISPR and immune defense against specific, targeted DNA. Viruses containing this targeted DNA were shown to be unable to infect the cells containing the spacer sequence identical to the viral sequence. However, they did succeed to infect closely related strains lacking the specific CRISPR spacer. In addition, plasmids transferred among various species could not be maintained in members with the specific spacer even though the species belonged to same phylogenetic group. In 2007, the CRISPR-Cas system was proven to be a part of the adaptive immunity of prokaryotes. Barrangou et al. reported that the spacer derived from phage DNA that is integrated into specific host genomic loci called the CRISPR-array, provides resistance specific for the phage, and that the bacteria can quickly adapt to certain population of phages by inserting new spacers.

1.3.2 Classification of CRISPR-Cas systems

CRISPR-Cas systems can be divided into two classes which are further divided into subclasses. Class 1 systems contain multi-Cas protein complexes and consists of subtypes I, III and IV, whereas class 2 systems use only one effector protein in interference and

consists of subtypes II, V and VI. (Hille et al., 2018.) In addition to differences in Cas-proteins between class I and II, these classes and their subtypes differ in the way of processing CRISPR-RNA (crRNA), and their targets can be either DNA or RNA. Generic organization of CRISPR-Cas classes are presented in figure 8.

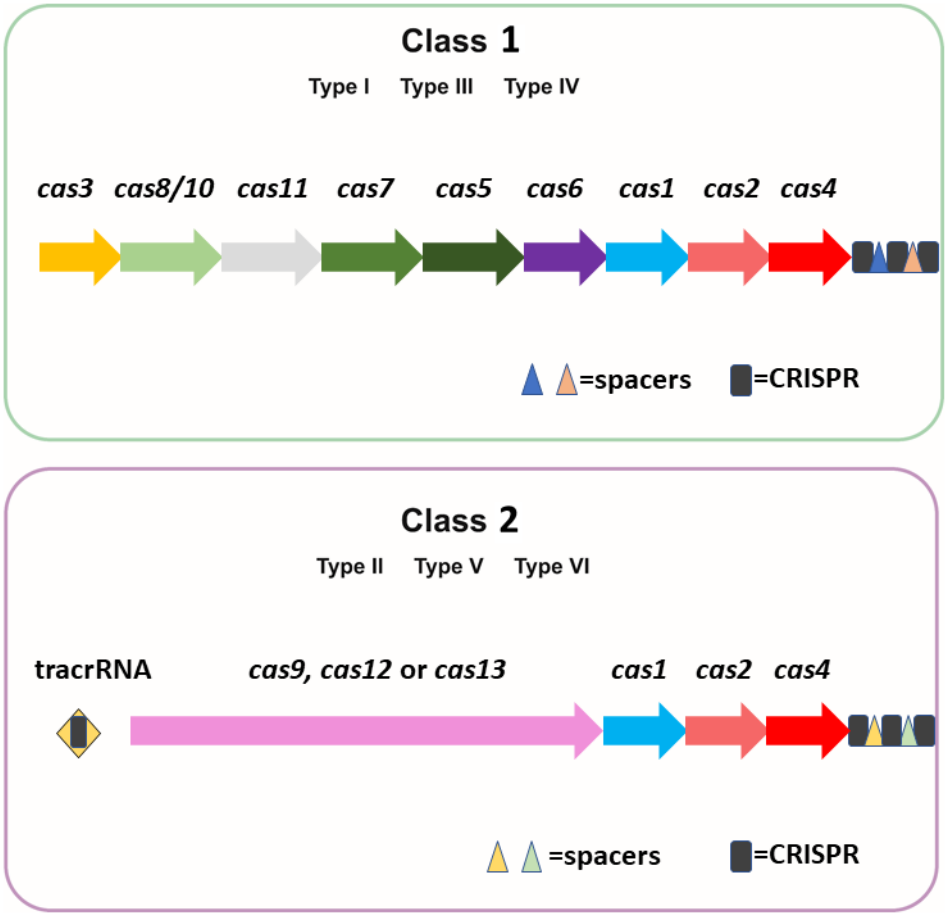


Figure 8 Classification of the CRISPR-Cas systems. *Systems are divided into two classes: class 1 and class 2. Class 1 has multiprotein Cas complexes and is further divided into types I, III and IV. Class 2 systems use only one effector Cas protein and is further divided into type II, V and VI. Adapted from Makarova et al. (2020).*

In class 1 systems, proteins needed in the interference consists of multiple Cas-proteins: Cas3, which can be sometimes fused to Cas2, Cas5, Cas8, Cas10 and Cas11. These proteins function in different combinations in different subtypes, whereas in class 2 systems, interference is performed by one single effector protein: Cas9 in subtype II, Cas12 in subtype V and Cas13 in subtype VI. (Makarova et al., 2020.) There are also common elements in different subtypes of CRISPR-Cas systems, such as Cas1 and Cas2, that are needed in the spacer integration into the CRISPR array.

There are clear trends in the distribution of different classes and their types. Class 1 is more abundant than class 2, whereas class 2 is almost exclusively present in bacteria and not in archaea, which can be explained by the RNase III needed in pre-crRNA processing that is not present in archaea. (Makarova et al. 2020.)

This thesis will mostly focus on the type II system and its applications.

1.3.3 Mechanism of action

The adaptive immunity of CRISPR-Cas9 is occurring in three stages: insertion of the spacer sequence into the CRISPR array, transcription of crRNA and mature crRNA-directed cleavage of foreign DNA sites complementary to the spacer sequence by Cas. These stages are called adaptation, expression and interference, and they are present in all types of systems of both class 1 and 2 (Figure 9). (Hille et al., 2018.)

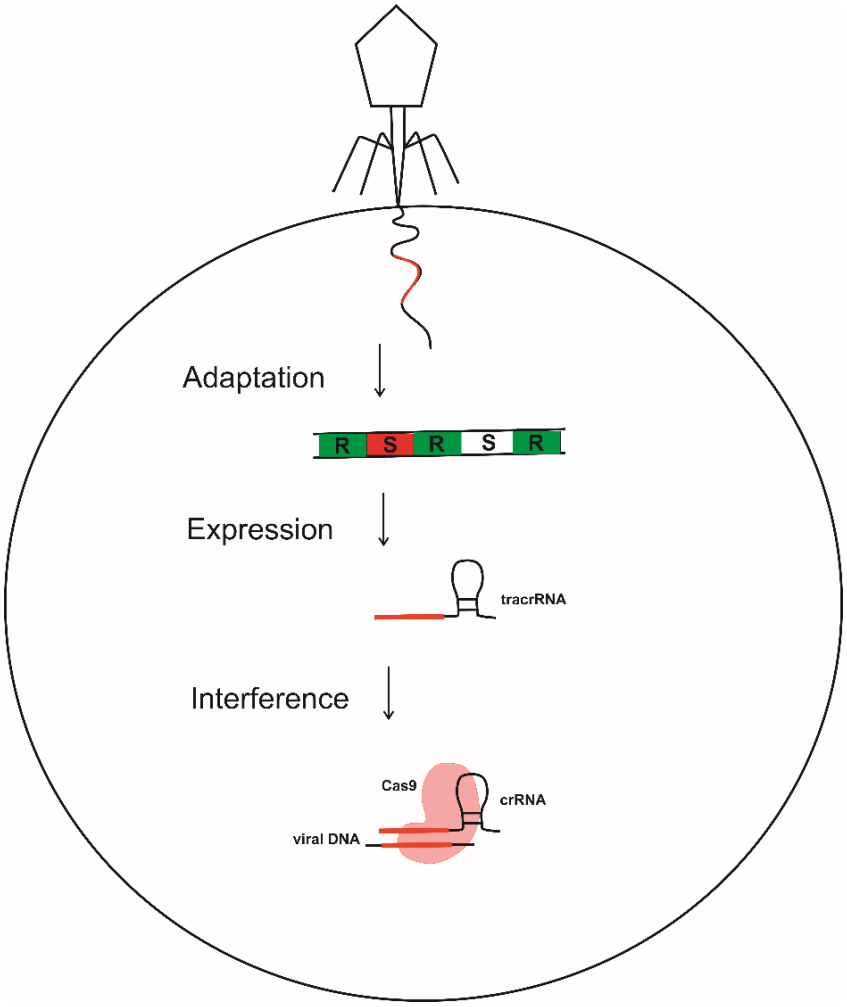


Figure 9 Mechanism of the CRISPR-Cas system. *The mechanism has three basic steps. The first step is adaptation, where new spacers are acquired into the CRISPR array. The second step is expression where pre-crRNA is transcribed and processed into a mature crRNA. The third and*

the final step is interference where invasive sequences are recognized and cleaved by Cas proteins. R=repeat, S=spacer. Adapted from Rath et al. (2015).

Adaptation, also known as the spacer acquisition, is the process in which bacteria or archaea gain specific genetic memory towards the infection, and the reason why CRISPR-Cas immunity is adaptive (Hille et al., 2018). Adaptation includes several steps which are detection of the mobile genetic element (MGE), selection of protospacer and processing it, and integration of the spacer into the CRISPR array. Protospacers are viral sequences which are located adjacent to protospacer adjacent motif (PAM) sequences. The adaptation machinery, consisting of Cas1 and Cas2, recognizes the PAM sequence located in the viral DNA. However, in type II systems, Cas9, Csn2 accessory protein and trans-activating crRNA (tracrRNA) are also essential components of the spacer acquisition (Heler et al., 2015). The PAM sequence recognition is essential for spacer acquisition, but also for interference. Protospacer is excised from the target DNA and inserted into the CRISPR array (Hille et al., 2018.)

The next step after spacer acquisition is expression of the crRNA and processing it. The newly acquired spacer is transcribed into pre-crRNA and processed by Cas proteins and additional RNAses. In order to process the pre-crRNA into mature crRNA, type II system requires RNase III, Cas9 and tracrRNA. (Deltcheva et al., 2011.) The tracrRNA is a noncoding RNA required for initiation of the pre-crRNA processing by activating RNase III (Jinek et al., 2012), and it is almost exactly complementary with the pre-crRNA. This enables RNase III to recognize it and process it into a mature crRNA needed in the interference (Deltcheva et al., 2011).

In interference, which is the final step of CRISPR adaptive immunity, foreign viral DNA is cleaved by Cas nuclease. Type II systems are simple in contrast to class 1 systems requiring cascades since they only use Cas9 endonuclease and its cofactors. crRNA and tracrRNA function as a guide RNA (gRNA) by base pairing with Cas9 as a RNA-duplex. This complex recognizes the target site where PAM sequence is located and cleaves both strands forming a DSB. (Jinek et al., 2012.)

Type II system is widely used in biotechnological applications (see section 1.3.5), but instead of a crRNA:tracrRNA complex, a synthetic single guide RNA (sgRNA) is used.

1.3.4 Cas9 structure and function

This chapter will focus on crystal structure of *Streptococcus pyogenes* Cas9. Cas9 contains two nuclease domains (HNH and RuvC), two recognition domains (REC1 and 2) and PAM-interacting domain (PI). It consists of two lobes which are recognition (REC) and nuclease lobe (NUC). (Nishimasu et al., 2014.) The two lobes are connected by a long α-helix called bridge helix (BH). The NUC lobe contains both nuclease domains, HNH and RuvC, needed for cleavage of both the complementary and noncomplementary strand, and it also contains the PI domain required for interaction with PAM sequence. The REC lobe is required for binding the sgRNA and the target DNA. The interaction between RuvC and PI domain forms a positively charged surface where negatively charged 3'-tail of the sgRNA can bind (Figure 10). (Nishimasu et al., 2014; Jinek et al., 2014.)

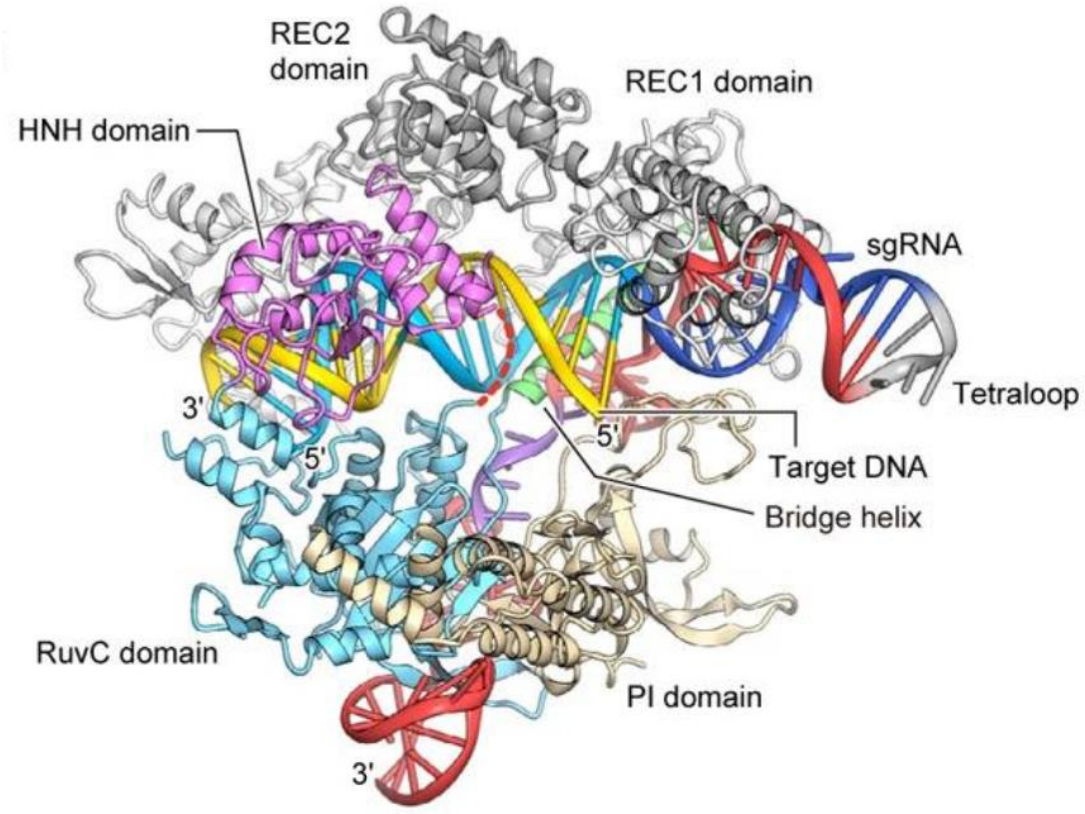


Figure 10 Structure of the Streptococcus pyogenes Cas9. *Figure illustrates the structure of SpCas9-single guide RNA (sgRNA)-DNA complex. It consists of two lobes: nuclease lobe (NUC) and recognition lobe (REC). PI=PAM interacting domain. Nishimasu et al. (2014)*

In this structure described by Nishimasu et al., SpCas9 binds sgRNA and target DNA, but it is in inactive state (apo state), and additional cofactors are not bound to it. Once the cofactors are bound, conformation change into active state occurs and the target DNA is cleaved (Jinek et al., 2014).

The REC lobe consists of REC1 and REC2 domains. REC1 contains elongated α-helical structure consisting of 25 α-helices and two β-sheets, whereas REC2 possess a six-helix bundle structure. The structure of REC lobe seems to be unique, indicating its specificity for Cas9. REC2 is not mandatory for DNA cleavage since a Cas9 mutant lacking it has been shown to retain Cas9 activity. However, REC1 is critical for DNA cleavage. (Nishimasu et al., 2014.)

HNH domain in the NUC lobe is responsible for cleavage of the DNA sequence complementary to crRNA and RuvC domain is responsible for cleaving the noncomplementary strand, resulting in DSB (Jinek et al., 2014). HNH domain consists of two-stranded antiparallel β-sheet and four α-helices. The active site has three catalytical amino acid residues: His41, Asp40 and Asn62, and additional Mg^{2+} for cleaving the substrates with a single-metal mechanism. (Nishimasu et al., 2014.) RuvC domain is formed by a mixed six-stranded β-sheet, seven α-helices and two two-stranded antiparallel β-sheet, and the active site contains four catalytical amino acid residues: Asp7, Glu70, His143 and Asp146, and two Mg^{2+}-ions for cleaving the noncomplementary strand with two-metal mechanism (Nishimasu et al., 2014).

The site specific DSB can be repaired by two different mechanisms: either by nonhomologous end joining (NHEJ) or homologous recombination. In NHEJ, broken ends of the DNA fragments are joined before they drift apart. However, NHEJ is more prone to errors and it usually can lead to undesired insertions or deletions of nucleotides. In homologous recombination, DNA helix is opened and the damaged strands invade into homologous DNA duplex. Homologous duplex can be for example from the sister chromatid or a vector. The homologous strand is then used as a template for repairing the DSB. Homologous recombination requires multiple proteins, and BRCA1 and BRCA2 are well-known examples of these. Because of the template homology required in the process, less errors usually occur compared to NHEJ. (Strachan & Read, 2019.)

The PI domain in the NUC lobe is also unique for Cas9 protein family, and it consists of seven α-helices, one three-stranded antiparallel β-sheet, one five-stranded antiparallel β-sheet and one two-stranded antiparallel β-sheet (Nishimasu et al., 2014). In addition to its

function in recognition of the PAM sequence, functional studies have implicated its importance in Cas9 nuclease activity since the deletion of the domain prevents Cas9 from cleaving the DNA (Nishimasu et al., 2014).

1.3.5 Applications

The discovery of CRISPR-Cas9 and its development into a genome editing tool has made it one of the most commonly used tools in genome editing. In 2012, Doudna and Charpentier showed that CRISPR-Cas9 introduces a DSB in the target DNA (Jinek et al., 2012), and later they showed that the tracrRNA:crRNA complex could be fused into a chimeric sgRNA, which made CRISPR-Cas9 even more convenient for genome editing, since it only required two components: Cas9 and a sgRNA. The HDR process can be used to introduce a desired edit into the target site by an additional exogenous donor DNA (dDNA) template.

Now CRISPR-Cas9 is widely used within different organisms because it has many advantages when compared with other gene editing technologies. It does not require as much labor and is not as expensive compared to for example endonuclease-based zinc fingers or transcription activator-like effector nucleases (TALENs), which demand re-engineering for every new target sequence. Cas9 is identical in all cases so only sgRNA sequences require editing if new target needs to be identified, also, CRISPR-Cas9 has the potential of editing multiple loci at the same time. (Lino et al., 2018.)

CRISPR-Cas systems have been widely used within eukaryotes, but not that much in bacteria due to their lack of capacity to fix DSBs. If they had the capacity to repair them after invader attack, the invader would persist. However, different techniques promoting homologous recombination of linear DNA fragments in bacteria have been developed (Luo et al., 2016).

CRISPR-Cas9 has provided quite a simple way to genetically engineer animals and thus, helped to understand molecular mechanisms underlying different diseases. It can also be exploited for the treatment of diseases, which have a genetic origin. The method is for example widely used in the research of cardiovascular, metabolic and neurodegenerative diseases, and cancer. (Lino et al., 2018.)

Despite the many advantages of CRISPR-Cas9 and its applications in genome engineering, there are limitations and challenges. DSB repair processes can easily lead to undesired byproducts such as insertions, deletions or mixtures of gene products. Editing

may also lead to off-target effects when highly homologous sequences can also be cleaved in addition to the desired target, and this can lead to chromosomal rearrangements and mutations. (Li et al., 2020.) HDR efficiencies also vary among different cell types, and it is active during only specific phases of the cell cycle. It also competes with NHEJ in the repair of DSBs and usually is outcompeted by it. (Porto et al. 2020.)

These limitations and challenges restrict the potential of CRISPR-Cas to be exploited for therapeutic purposes. Ideal therapeutical tools for genome editing would possess high editing efficiency for the target site and minimal off-target effects. Therefore, enhancing the efficiency and precision of CRISPR-Cas9 or developing new more precise tools is of interest for researchers.

By mutating either the HNH or the RuvC domain of the Cas9, researchers have created the Cas9 nickase which only cleaves one strand. Nicking only one strand does not make a DSB, and therefore these editing methods where Cas9 nickases are used have less undesired edits like insertions and deletions. Base editing and prime editing are examples of these tools, where only one strand is nicked instead of introducing DSBs and the desired edit is introduced without an additional dDNA template. (Anzalone et al. 2019; Porto et al. 2020.)

1.3.6 Base editing

Base editing is a genome editing method using CRISPR components and other enzymes to introduce mutations into DNA sequence without causing DSBs. Base editing does not cleave the nucleic acid backbone because it chemically alters the target site. Base editors can target both DNA and RNA. (Porto et al., 2020.)

DNA base editors can be divided into cytosine base editors (CBEs) and adenine base editors (ABEs), and they can introduce permanent transition mutations (C→T, T→C, A→G and G→A) with high efficiencies. CBEs use cytidine deaminase to convert cytosine into uracil resulting in conversion of G-C bp to A-T bp, whereas ABEs mediate editing with inosine-containing intermediate resulting in conversion of A-T bp to G-C bp. RNA base editors are divided according to the modification they introduce. (Porto et al., 2020.)

Base editing offers promising alternative for genome editing without introducing DSBs and it can efficiently install all four transition mutations. However, it lacks the ability to install transversion mutations (C→A, A→C, G→C. C→G, G→T, T→G, A→T and

T→A) and targeted deletions or insertions, and this reflects the need for alternative editing tools. (Anzalone et al., 2019; Porto et al., 2020.)

1.3.7 Prime editing

Prime editing is a novel genome editing method described by Anzalone et al. in 2019. Prime editing uses a catalytically impaired Cas9 endonuclease fused with reverse transcriptase (RT) and a prime editing guide RNA (pegRNA) containing the primer binding site (PBS) and template sequence for RT with the desired edit (Figure 11).

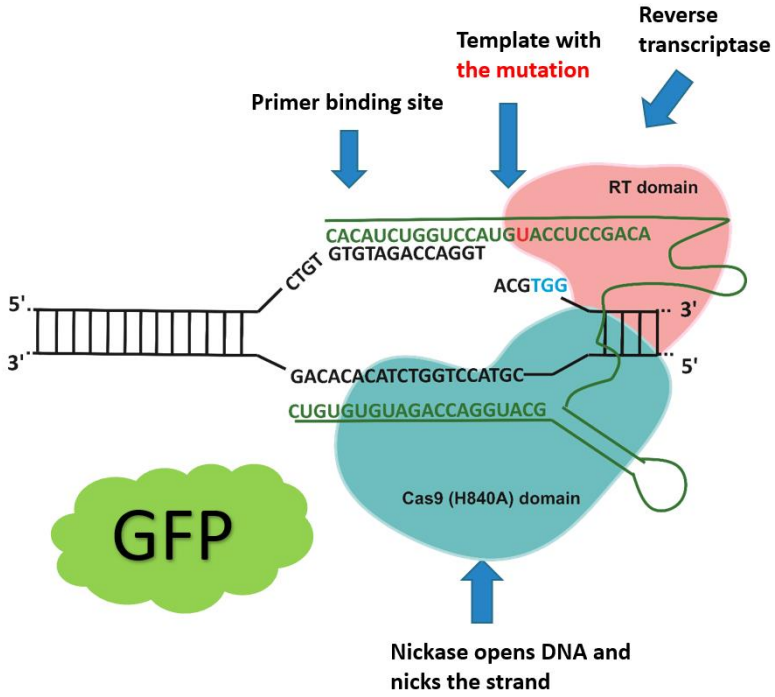


Figure 11 Illustration of the prime editing complex*. Prime editing guide RNA (pegRNA) is coloured in green. Prime editor (PE) consisting of reverse transcriptase (RT) and Cas9 domain is nicking the template strand and introducing the mutation by using the pegRNA as a template. After introducing the nick, mutation is introduced by RT using the pegRNA as a template. PE complex also contains green fluorescent protein (GFP) as a reporter. Adapted from Anzalone et al. (2019).*

The mechanism is that the prime editor (PE), a fusion protein, nicks the target DNA sequence and exposes a 3'-hydroxyl group. This hydroxyl group is then used to initiate reverse transcription of the pegRNA template. This results in an intermediate containing two flaps of DNA: a 3'-flap containing the desired edit and a 5'-flap containing the unedited DNA. Endogenous 5'-exo- or endonuclease is used to cleave this flap and 3'-flap is ligated into DNA. Cleavage of the unedited flap favors the DNA repair of the non-

edited strand and forces the cells to use the prime edited strand as a template (Figure 12). (Anzalone et al., 2019.)

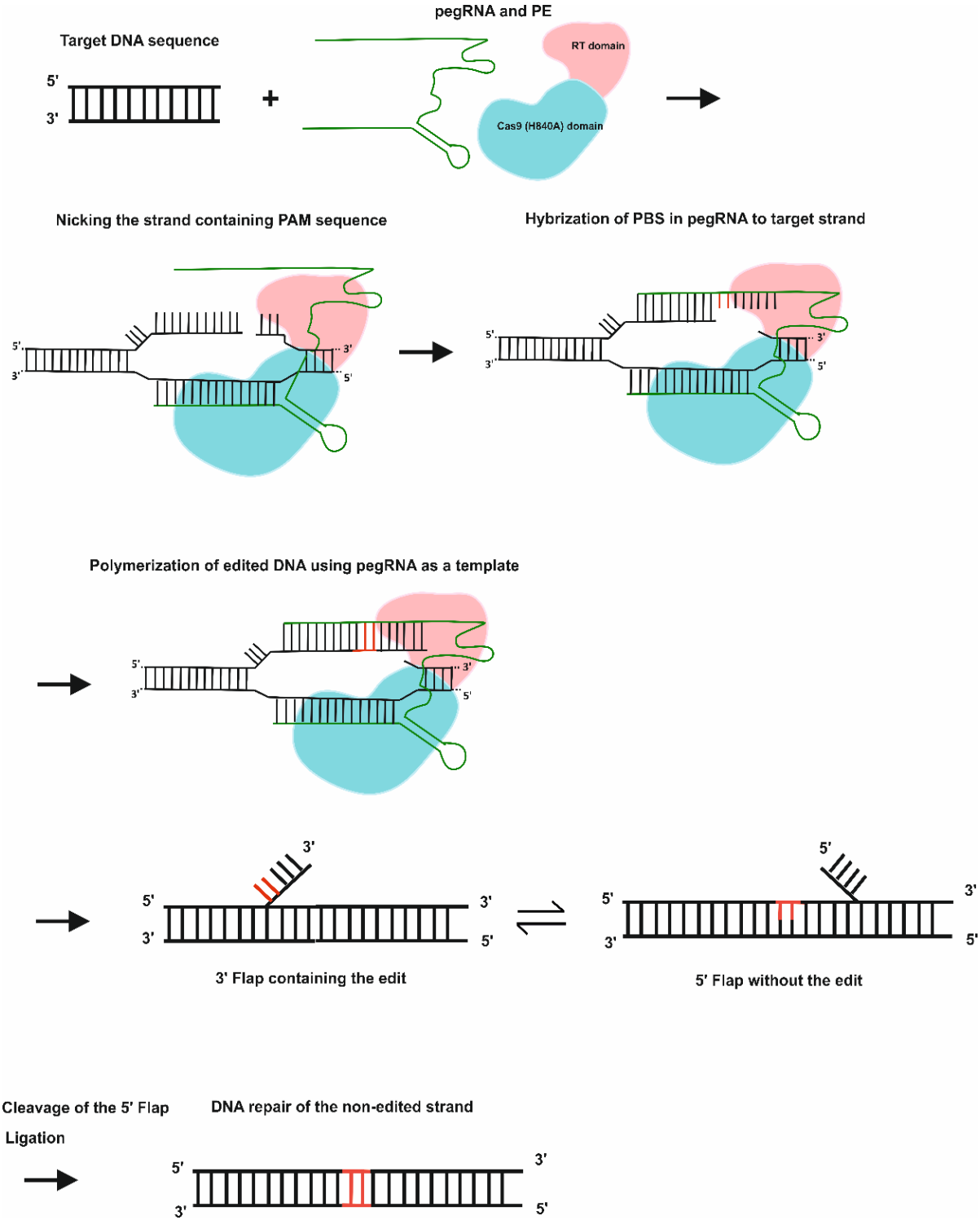


Figure 12 The mechanism of prime editing. *Prime editing guide RNA (pegRNA) and prime editor (PE) are introduced to target DNA and after recognition of the protospacer adjacent motif (PAM) a nick is introduced by Cas9 nickase. Primer binding site (PBS) in the pegRNA is hybridizated to the target strand and desired edit is introduced with reverse transcriptase (RT) using the pegRNA as a template. Ultimately, this results in equibilirium of two flaps competing with each other: a*

3' flap containing the edit and a 5' flap without the edit. After cleavage of the 5 flap, ligation and DNA repair of the non-edited strand in the target site, target DNA containing the desired edit is achieved. Adapted from Anzalone et al. (2019).

Anzalone et al. used prime editing to install and correct the primary genetic alterations causing sickle cell disease and Tay-Sachs disease. Sickle cell disease is usually caused by transversion resulting in E6V in hemoglobin subunit beta (*HBB*) whereas Tay-Sachs disease by a 4 bp insertion into hexosaminidase subunit alpha (*HEXA*). They installed the *HBB* E6V into human embryonic kidney cells (HEK293T cell line) resulting in 44 % efficiency and 4,8 % insertions and deletions (indels). Correcting this mutation back to WT resulted in 26 – 52 % efficiency with 2,8 ± 0,7 % indels. They installed the 4 bp insertion into *HEXA* with 31 % efficiency and 0,8 % indels and reverted this insertion back to WT with 33 % efficiency and 0,32 % indels.

In order for prime editing to work, pegRNA spacer complementarity with target DNA is required for the Cas9 domain binding and PBS of pegRNA needs to be complementary with the target DNA for the initiation of reverse transcription. The RT product also needs to be complementary with the target DNA for flap resolution (Anzalone et al., 2019). Anzalone et al. tested the ability of these hybridization steps to reduce off-target prime editing compared to conventional CRISPR-Cas9 editing methods requiring only gRNA complementarity. They concluded that prime editing introduces much lower off-target editing.

This method does not need additional donor DNA's or DSBs such as conventional CRISPR-Cas9. DSBs can easily introduce undesired byproducts such as translocations or mixtures of products, and it can also activate p53 (Haapaniemi et al., 2018). Even though DSB-mediated genome editing has a promising future especially in the therapeutic applications, alternative precision editing methods are constantly explored, and prime editing is a promising tool for this purpose.

Base editing introduces efficiently single-nucleotide variants into DNA without DSBs (Porto et al., 2020). However, it can currently perform only the four transition mutations (pyrimidine to pyrimidine or purine to purine) and not the transversion (purine to pyrimidine or pyrimidine to purine) mutations in mammals. Base editing also cannot perform targeted deletions or insertions whereas prime editing can. (Anzalone et al. 2019; Porto et al. 2020.) Anzalone et al. compared prime editing and base editing in three genomic loci containing multiple target cytosines in the base editing window and

depending on the desired locus, they functioned in a complementary fashion. Current base editors are more efficient and generate fewer indels. However, prime editing introduces less off-target editing and it can perform in the absence of suitably positioned PAM site.

Currently there are not many publications considering the method. Prime editing has been performed in plants such as rice and maize, but it has succeeded with only small efficiencies (Jiang et al., 2020; Lin et al., 2020), and in *Drosophila melanogaster* (Bosch et al., 2021) in addition to experiments performed by Anzalone et al.

Prime editing offers a promising addition to the precision genome editing toolbox by allowing DNA specificity and editing efficiency and by enabling all 12 possible base-to-base conversions and indels. Base editors can also install transition mutations and currently, they are more efficient and introduce fewer indels if the target locus is located within the canonical base editing window. However, base editing is limited to only transitions and it introduces unwanted bystander edits more than prime editing. Prime editing eliminated the need for DSBs and additional donor DNAs which are one of the challenges when using conventional CRISPR-Cas9 genome editing methods. Using prime editing still has many challenges and it may not be able to perform larger DNA insertions and deletions. Using RT might also incorporate unwanted, random cDNAs and pegRNAs and sgRNAs might confer problems when it comes to RNA-stability. However, more studies on prime editing within different cell lines are needed to optimize the method. Possibly, in the future it has the potential to be exploit for therapeutic applications in the treatment of genetic disease by enabling transitions, transversions and indels in mammalian genomes without DSBs. (Anzalone et al., 2019; Matsoukas, 2020.)

2. Aim of the thesis

The aim of the thesis was to generate tools to generate HOXB13^{G84E} cellular models with prime editing. Four different prime editing plasmids were generated: three pegRNAs with PBS lengths of 10 nt, 13 nt and 16 nt, and an additional sgRNA for nicking the non-edited strand. A plasmid encoding HOXB13 was also generated and the G84E mutation was introduced with site-directed mutagenesis by PCR, and this was used as a positive control in droplet digital PCR (ddPCR) assay. Western blot was performed to protein lysates of PCa and RWPE-1 cell lines to determine if they express HOXB13, and based on the results, a cell line for the experiments was chosen. The aim was to optimize the PBS of the pegRNA by determining which of the three was the most efficient, and if prime editing could be used for introducing the G84E mutation in the chosen cell line for further research. In the future, the most suitable construct can be used to generate cellular models of the germline mutation G84E in *HOXB13* for elucidating its function and mechanism in cancer development.

3. Materials and methods

3.1 Generation of the plasmid constructs

3.1.1 pegRNA plasmids

The pU6-pegRNA-GG-acceptor plasmid was a gift from David Liu (Addgene plasmid # 132777) (Anzalone et al., 2019). The plasmid was digested in a PCR tube containing 30 µl of reaction mix [2000 ng plasmid, 20 U Bsa1-HFv2 (NEB*),* 1X CutSmart buffer (NEB*,* 50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate, 100 µg/ml BSA, pH 7.9)]. The reaction mix was incubated for 5 h at 37 °C. Blue loading dye was added to sample (NEB, 2.5% Ficoll®-400, 11 mM ethylenediaminetetraacetic acid (EDTA), 3,3 mM Tris-HCl, 0,017 % sodium dodecyl sulfate (SDS), 0,015 % bromophenol blue, pH 8) and reaction products were separated on a 1 % agarose gel (100 V). The plasmid backbone was extracted from the gel with NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) according to its protocol for DNA extraction from agarose gels. The DNA concentration was measured with NanoDrop One/One^{C} Microvolume UV-Vis Spectrophotometer (ThermoFisher Scientific).

Oligonucleotides (Integrated DNA technologies) for the pegRNAs (Table 1, oligonucleotides 1 - 6) were annealed by combining them in a PCR tube with total volume of 25 µl [4 µM top oligonucleotide, 4 µM bottom oligonucleotide, annealing buffer (9,2 mM Tris-Cl pH 8.5, 46 mM NaCl)]. Reaction mixes were heated at 95 °C for 5 min in a thermocycler and gradually cooled to 22 °C (0,1 °C/s). Annealed oligonucleotides were diluted to 1 µM except for the sgRNA scaffold, which was phosphorylated in a PCR tube with a total volume of 25 µl [1 µM sgRNA scaffold oligonucleotide duplex, 5 U T4 PNK (NEB*),* 1X T4 DNA ligase buffer (70 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM dithiothreitol (DTT), pH 7.6, NEB)]. Phosphorylation was done by incubating the reaction mix at 37 °C for 1 h.

pegRNAs were assembled with Golden Gate cloning reaction. Reactions were performed in PCR tubes with total volumes of 10 µl [30 ng digested pU6-pegRNA-GG-acceptor plasmid vector, 1 µM annealed protospacer oligonucleotides, 1 µM annealed pegRNA 3´-extension oligonucleotides, 1 µM annealed sgRNA scaffold oligonucleotides, 5 U BsaI-HFv2 (NEB), 1000 U T4 DNA ligase (NEB), 1X T4 DNA ligase buffer]. Reaction was performed in a thermocycler with the following program: 5 min at 16 °C and 5 min at 37 °C for 8 cycles, 15 min at 37 °C and 15 min at 80 °C, hold at 12 °C.

3.1.2 gRNA plasmid for nicking the non-edited strand

The MLM3636 plasmid was a gift from Keith Joung (Addgene plasmid # 43860). This plasmid was used for creating the MLM3636-HOXB13-G84E-3GtoA+97nick plasmid, which encodes a gRNA targeting *HOXB13*. gRNA_HOXB13G84_3GtoA_+97nick forward (F) and reverse (R) oligos were annealed (0,5 µM of each oligo) in reaction volume of 20 µl by heating the mix to 95 °C for 5 min and cooling to 10 °C gradually (5 °C/min). The MLM3636 vector backbone had already been digested with Esp3I. The oligo duplex was ligated into the vector in reaction mix with total volume of 20 µl [150 ng digested MLM3636 plasmid, 1 µl of double-stranded annealed oligos, 1X T4 DNA, 5 U Esp3I (NEB), 1000 U T4 DNA ligase]. Reaction was performed in a thermocycler with the following program: 5 min at 16 °C and 5 min at 37 °C for 8 cycles, 15 min at 37 °C and 15 min at 80 °C, hold at 12 °C.

3.1.3 WT pcDNA3-HOXB13

cDNA synthesis was performed for RNA extracted from LNCaP native cells with iScript™ Select cDNA Synthesis Kit (Bio-Rad) with a 20 µl reaction mix [1 µg of total RNA, 1X iScript select reaction mix, 2 µl Oligo$(dT)_{20}$ primer, 1 µl iScript reverse transcriptase]. The reaction mix was incubated for 90 min at 42 °C and for 5 min at 85 °C. The PCR reaction for cloning the HOXB13 insert from cDNA was performed in a reaction mix with total volume of 25 µl [1X CloneAmp HiFi PCR Premix (Takara), 2 µl LNCaP cDNA, 0,3 µM primer HOXB13_In_F1, 0,3 µM primer HOXB13_In_R1 (Integrated DNA technologies)]. PCR reaction was performed with the following program: 98 °C 10s, 55 °C 15 s, 72 °C 5 min for 35 cycles and hold at 10 °C.

Linearization and removal of NLS-mCherry from pcDNA3-mCherry-NLS was made with PCR. The reaction mix was prepared into a total volume of 25 µl [1X CloneAmp HiFi PCR Premix, 10 ng template, 0,3 µM primer pcDNA3-Ano7_InF3, 0,3 µM primer pcDNA3_R2 (Integrated DNA technologies)]. The PCR reaction was run with the following program: 98 °C 10s, 55 °C 15 s, 72 °C 5 min for 35 cycles and hold at 10 °C.

Both PCR reaction products were separated on a 1 % agarose gel (100 V) and the correct size DNA was cut from the gel and extracted with the NucleoSpin Gel and PCR Clean-up kit according to its protocol for DNA extraction from agarose gels.

In-Fusion cloning reaction was done for HOXB13 insert and pcDNA3 backbone in a total reaction volume of 10 µl [1X In-Fusion HD Enzyme Premix, 150 ng pcDNA3 vector, 50 ng HOXB13 insert]. Reaction mix was incubated for 15 min at 50 °C and placed on ice.

Removing the intron from pcDNA3-HOXB13 was done by PCR in a reaction mix with total volume of 25 µl [10 ng template plasmid, 1X CloneAmp HiFi PCR Premix, 0,3 µM primer HOXB13intronDel_F1, 0,3 µM primer HOXB13intronDel_R1 (Integrated DNA technologies)]. The PCR reaction was run with the following program: 98 °C 10s, 55 °C 15 s, 72 °C 6 min for 35 cycles and hold at 10 °C. The PCR-product was run on a 1 % agarose gel (100 V) and PCR-clean-up was performed to the pcDNA3-HOXB13 plasmid without the intron. In-Fusion cloning was performed on the cleaned plasmid incubating the reaction mix with a total volume of 10 µl [1X In-Fusion HD Enzyme Premix, 50 ng pcDNA3-HOXB13] for 15 min at 50 °C following transformation to Stellar bacteria.

Undesired P144L SNP was removed with PCR-mutagenesis as described in section 3.1.4, but primers that were used were HOXB13_P144L_F1 and HOXB13_P144L_R1 (Integrated DNA technologies).

3.1.4 pcDNA3-HOXB13_G84E

Site-directed mutagenesis was performed on the WT pcDNA3-HOXB13 in a reaction mix of 50 µl [1X CloneAmp HiFi PCR mix, 0,2 µM primer HOXB13_G84Emut_F1, 0,2 µM primer HOXB13_G84Emut_R1 (Integrated DNA technologies), 10 ng template]. PCR reaction was run with the following program: 98 °C 10s, 65 °C 15 s, 68 °C 6 min for 18 cycles. 10 µl of PCR-product was run on 1 % agarose gel (100 V) and PCR-clean-up was performed to the rest of the product.

DpnI treatment was performed on the cleaned product in a 50 µl reaction mix [1 µg plasmid, 1X CutSmart buffer, 20 U DpnI]. The mix was incubated for 1 h at 37 °C and PCR-clean-up was performed.

3.1.5 Plasmid isolation and sequencing

Plasmids were transformed into 5α-competent *E.coli* (NEB). Transformation was performed according to the protocol of NEB. Transformed bacteria were plated on LB plates containing 100 µg/ml of ampicillin and grown overnight at 37 °C.

For plasmid isolation, a *E.coli* colony from the LB agar -plate was selected for liquid culture (LB + ampicillin 100 µg/ml). Liquid cultures were grown overnight at shaking

(300 rpm, 37 °C). The following day, plasmids were isolated with NucleoSpin kit for plasmid DNA purification (Macherey-Nagel) according to the manufacturer's instructions.

Glycerol stocks were made from each plasmid (1:1 liquid culture and 65 % glycerol) and stored at -80°C. Samples were prepared for sequencing according to instructions (Eurofins) and sent to Eurofins Genomics for Sanger sequencing.

Table 1 Primer/oligonucleotide table. *Primers/oligonucleotides used for generating the pegRNAs, sgRNa and pcDNA3-HOXB13.*

Primer name	Forward (F) 5'→3'	Reverse (R) 5'→3'
HOXB13G84E_3GtoA_PBS10_Ext13	GTGCGTACCCGCCTTCAAAGTAACCAT	AAAAATGGTTACTTTGAAGGCGGGTAC
HOXB13G84E_3GtoA_PBS13_Ext13	GTGCGTACCCGCCTTCAAAGTAACCATAAG	AAAACTTATGGTTACTTTGAAGGCGGGTAC
HOXB13G84E_3GtoA_PBS16_Ext13	GTGCGTACCCGCCTTCAAAGTAACCATAAGGCA	AAAATGCCTTATGGTTACTTTGAAGGCGGGTAC
pegRNA_scaffold	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAA CGGACTAGCCTTATTTTAACTTGCTATTTCTAG
HOXB13_6GtoA_pegRNAspacer	CACCGTGCCTTATGGTTACTTTGGGTTTT	CTCTAAAACCCAAAGTAACCATAAGGCAC
gRNA_HOXB13G84E_3GtoA_+97nick	ACACCGGGTACTCTTCCCCGGCCGTG	AAAACACGGCCGGGGAAGAGTACCCG
pcDNA3	GTTTAAACCCGCTGATCAGCCTC	GGTGGCCGTAAGCTTAACTAGCC
HOXB13_In_	AAGCTTACGGCCACCATGGAGCCCGGCAATTATG	TCAGCGGGTTTAAACTTAAGGGGTAGCGCTGTTC
HOXB13intronDel	ATTTGCAGACTCCAGCGGGCAGCACC	CTGGAGTCTGCAAATGCTGCCTTCCAAAAGG
HOXB13_G84Emut	CTTATGGTTACTTTGAAGGCGGGTACTACTC	GAGTAGTACCCGCCTTCAAAGTAACCATAAG
HOXB13_P144L	CTATGGCCAGTTACCTGGACGTGTCTGTGG	CCACAGACACGTCCAGGTAACTGGCCATAG

3.2 Cell lines and culture conditions

3.2.1 Cell lines and growth conditions

LNCaP (American Type Culture Collection, ATCC), RWPE-1 (ATCC) and PC-3 (ATCC) cells were cultured at +37 °C and 5 % CO_2- concentration. Trypsin-EDTA (0,25 %) (Gibco, Thermo Fisher) was used for detaching the cells, and phosphate buffer saline (PBS) was used for washing the cells.

LNCaP cells were cultured in RPMI-1640 medium (ATCC, 30-2001), PC-3 cells in F-12K medium (ATCC-30-2004). Both media contained 10 % fetal bovine serum (FBS) and 5 % penicillin and streptomycin (PenStrep). RWPE-1 cells were cultured in Keratinocyte Serum Free Medium (K-SFM) (GIBCO), containing 0,05 mg/ml bovine

pituitary extract (BPE), 5 ng/ml epidermal growth factor (EGF) and 5 % PenStrep. All cell lines were cultured in Corning® T-75 flasks.

3.2.2 Transfection of cell lines and cell sorting

Transfection was performed with Invitrogen Lipofectamine 3000 reagent (ThermoFisher Scientific) according to its protocol. Following amounts of plasmids per well of 12-well plate were used in the transfections: 3 µg of pCMV-PE2-P2A-GFP (a gift from David Liu (Addgene plasmid # 132776) (Anzalone et al., 2019), 1 µg of pU6-pegRNA-HOXB13-3GtoA-PBS10/13/16-Ext13, 0,33 µg MLM3636-HOXB13-G84E-3GtoA+97nick. GFP-positive cells were sorted with a Sony SH800 Cell sorter.

3.3 Determination of cell lines expressing HOXB13 with Western blot

3.3.1 Protein lysate preparation

Protein lysates were prepared by removing media from the cell culture, washing with PBS and by adding 200 µl of radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris-Cl, pH 8; 150 mM NaCl; 1 % Triton-X; 0,5 % Sodium deoxycholate; 0,1 % SDS; protease inhibitor (Roche)). Cells were scraped and collected into microcentrifuge tubes and incubated for 30 min on ice. After incubation, cells were centrifuged for 20 min at 13000 × g (+ 4 °C) and supernatant was collected.

Protein concentration determination was performed with Pierce™ BCA Protein Assay Kit (Thermo Scientific) according to kit protocol and BSA was used for making the standard curve.

3.3.2 Western blot

The protein amount of the cells was analyzed with Western blot. Protein samples were prepared by adding 30 µg of protein lysate, 1 × Laemmli sample buffer (65,8 mM Tris-HCl, pH 6.8; 2,1 % SDS; 26,3 % (w/v) glycerol; 0,01 % bromophenol blue; 5 % β-mercaptoethanol) and adjusting the volume to 20 µl. Samples were boiled at +95 °C for 5 min and centrifuged shortly. 20 µl of sample/well was loaded to 4 – 20 % gradient gel (Mini-PROTEAN TGX Precast Gels, #4561096, BIO-RAD) with protein standard

(Bluestar Prestained Protein Marker Plus, Nippon Genetics). Gel was run with 100 V in running buffer (0,025 M Tris; 12,5 mM glycine; 0,1 % (w/v) SDS) for 1,5 h.

After the gel run, proteins were transferred onto polyvinylidene fluoride (PVDF) membrane (Immobilon PVDF Membrane, Sigma Aldrich) with wet electroblotting system (Mini Trans-Blot Cell, BIO-RAD) with 70 V for 3 h (+4 °C) in transfer buffer (0,025 M Tris; 12,5 mM glycine; 0,1 % SDS; 20 % (v/v) ethanol). Membrane was blocked with 5 % milk in TBST for 1 h on a tube roller at room temperature (RT). The membrane was washed with TBST and incubated overnight with HOXB13 primary antibody (Cell signaling technology, D7N80 Rabbit mAb) diluted 1:1000 in TBST with 1 % BSA and 0,02 % Na-azide at +4 °C. The membrane was washed three times with TBST for 5 min and incubated with horseradish peroxidase conjugated secondary antibody (Abcam, Goat anti-Rabbit IgG) diluted 1:10000 in TBST with 5 % milk for 2 h at RT. The membrane was washed three times as above and incubated with Western Bright Quantum chemiluminescent substrate for two min. The (Advansta) chemiluminescence was detected with a ChemiDoc imaging system (BIO-RAD). The images and protein band intensities were quantified with ImageJ (Schindelin et al., 2012). Normalization of the samples was made using β-actin as a loading control.

3.4 Measuring prime editing efficiency with droplet digital PCR

3.4.1 DNA isolation

DNA from the transfected cells was isolated with Illustra blood genomicPrep Mini Spin Kit (GE Healthcare) according to the manufacturer's instructions, but cells were washed with PBS for one additional time after detaching them.

3.4.2 ddPCR assay

Prime editing efficiency was measured with ddPCR. The reaction mix was prepared and each well contained 1 × ddPCR Supermix for Probes (No dUTP), 250 nM probe FAM (WT HOXB13, BIO-RAD), 250 nM probe HEX (HOXB13 G84E, BIO-RAD), 5 U HindIII HF and 40 – 150 ng sample (total volume 22 µl/well). Droplets were generated with Automated Droplet Generator (#1864101, BIO-RAD).

PCR was performed for droplets with the following program: 95 °C 10 min, 94 °C 30 s, 56 °C 1 min, 98 °C 10 min and 4 °C hold. Steps 2 and 3 were repeated 40 times.

Droplets were analyzed with QX200 Droplet Reader and the QuantaSoft analysis software (BIO-RAD).

4. Results

4.1 Plasmid generation

4.1.1 Three different pegRNA-plasmids were generated for optimizing the PBS

The aim was to optimize the PBS length by generating three different pegRNA constructs with different lengths of PBS: one with 10 nt PBS, one with 13 nt PBS and one with 16 nt PBS (Figure 13A). Plasmids were successfully generated by digesting the pU6-pegRNA-GG-vector acceptor with BsaI (Figure 13A & B) and annealing the pegRNA spacer, pegRNA scaffold and pegRNA 3' extension (containing the PBS) (Table 1), and by assembling the annealed components with pU6-vector backbone with Golden Gate cloning.

After assembling the pegRNAs, they were grown in *E.coli* and extracted. Plasmids were sent for sequencing and the correct sequences of each pegRNA were verified with Sanger sequencing (Figure 13C).

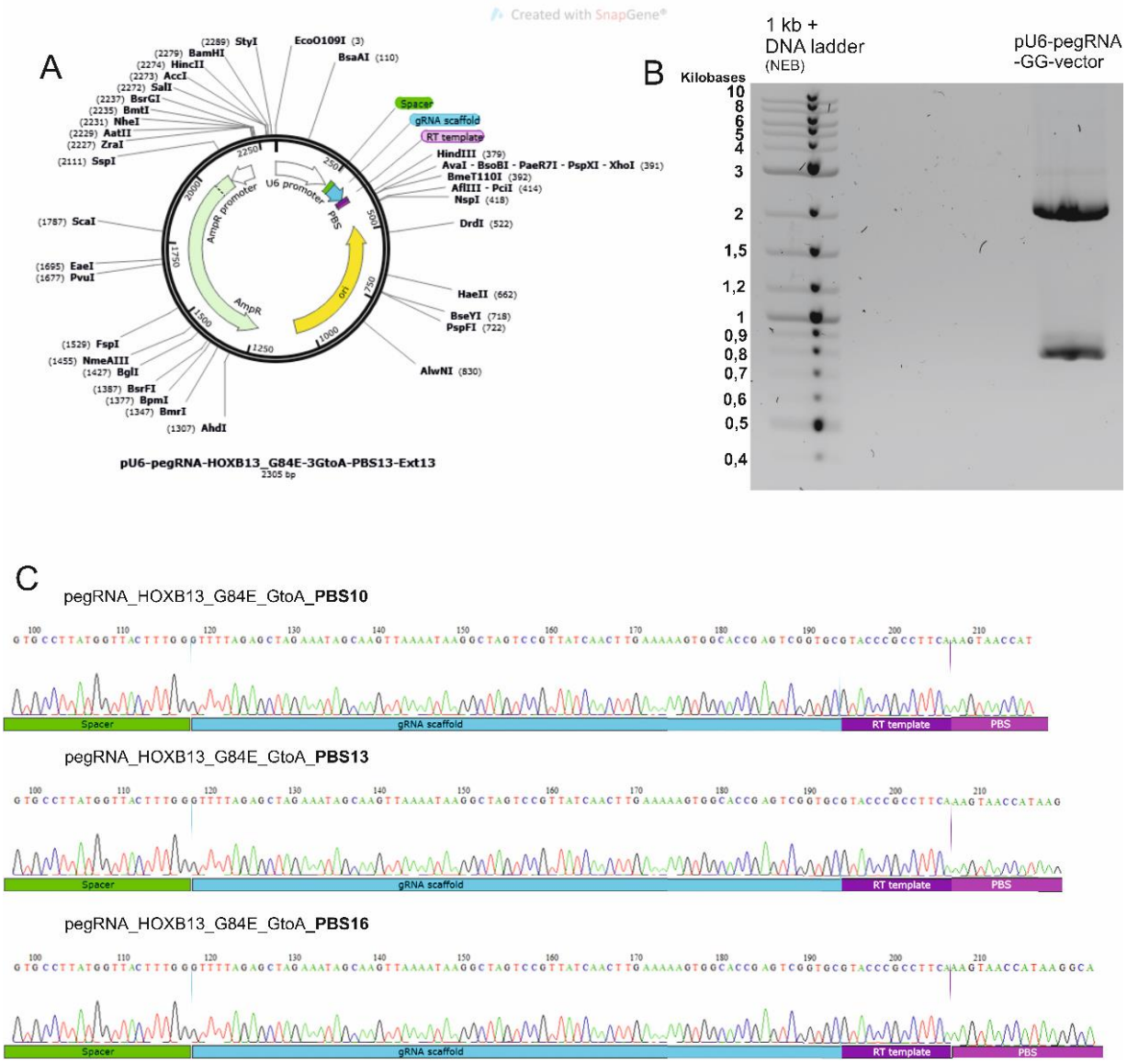


Figure 13 Plasmid map of pU6-pegRNA-HOXB13_G84E-3GtoA-PBS13-Ext13 and its generation process. *A= plasmid contains ampicillin resistance cassette (AmpR) and promoter for it, U6 promoter, origin of replication (ori), spacer, guide RNA (gRNA) scaffold, primer binding site (PBS) and reverse transcriptase (RT) template. Restriction sites for enzymes are shown in the picture. B= pU6 -GG-vector backbone was digested with BsaI and 2.2 kb pU6 fragment was cleaned from 1 % agarose gel and used as a plasmid backbone for three different pU6-pegRNA-HOXB13_G84E-3GtoA-PBS10/13/16-Ext13 plasmids. C= Sanger sequencing results showing correct sequences of spacer, gRNA scaffold, RT template and PBS of all three plasmids.*

4.1.2 sgRNA-plasmid was generated for nicking the non-edited strand at +97 bases from the prime editing site

A plasmid for sgRNA was generated for nicking the non-edited strand at +97 bp from prime edited site to favor the repair of non-edited strand in the target DNA by using the newly edited strand as a template. gRNA oligonucleotides (Table 1) were successfully annealed and ligated into Esp31 digested MLM3636 vector backbone resulting in MLM3636-HOXB13_G84E.3GtoA+97nick -plasmid (Figure 14A).

After transformation into *E.coli* for cloning and extraction, correct sequence were verified with Sanger sequencing (Figure 14B).

Figure 14 Plasmid map of the MLM3636-HOXB13_G84E-3GtoA+97nick and result from Sanger sequencing showing the correct crRNA and gRNA scaffold sequence. *A= The plasmid contains origin of replication (ori), ampicillin resistance cassette (AmpR) and its promoter, U6 promoter for transcription of the gRNA scaffold. Transcribed gRNA is used for nicking the nonedited strand at 97 bp from the prime edited site, and this favors the repair of the nonedited strand in the target DNA by using the prime edited strand as a template. B=Results from Sanger sequencing verified that the plasmid is correct. Sequences of crRNA and gRNA scaffold can be seen from the picture.*

4.1.3 The pcDNA3-HOXB13_G84E -plasmid was generated for a positive control for the mutation in the ddPCR assay

A plasmid containing the HOXB13 cDNA was generated and the G84E mutation was later introduced with site-directed mutagenesis by PCR, and this was used as a positive control for the mutation in the ddPCR assay. cDNA was synthetized from LNCaP RNA and HOXB13 cDNA was amplified with PCR. pcDNA3 was used as a backbone and it was linearized and the NLS-mCherry was removed with PCR and gel electrophoresis (Figure 15A). The pcDNA3 backbone was of correct size (5,5 kb) but the size of the HOXB13 cDNA insert was approximately 1,8 kb even though it should be 884 bp (Figure 15A). HOXB13 contains two exons, and they have one intron between them which is 949 bp, so the results implicated that the intron was retained. The plasmid was assembled with In Fusion -cloning reaction and after cloning and extraction, samples were sent for sequencing. Sanger sequence verified that the intron was retained in HOXB13 between the two exons (Figure 15B).

The intron was removed from pcDNA3-HOXB13 plasmid with PCR and same steps were repeated (Figure 15C). Sanger sequencing results showed that the intron was removed. However, the results revealed additional single nucleotide polymorphisms (SNP) in exon 2 of the HOXB13 which was not noticed previously. Ewing et al. has previously reported that LNCaP cells have the L144P mutation in one allele. The L144P SNP was corrected and G84E mutation was introduced with PCR, and ultimately WT-pcDNA3-HOXB13 and pcDNA3-HOXB13_G84E plasmids were generated (Figure 15D) and Sanger sequencing verified the correct sequences (Figure 15E).

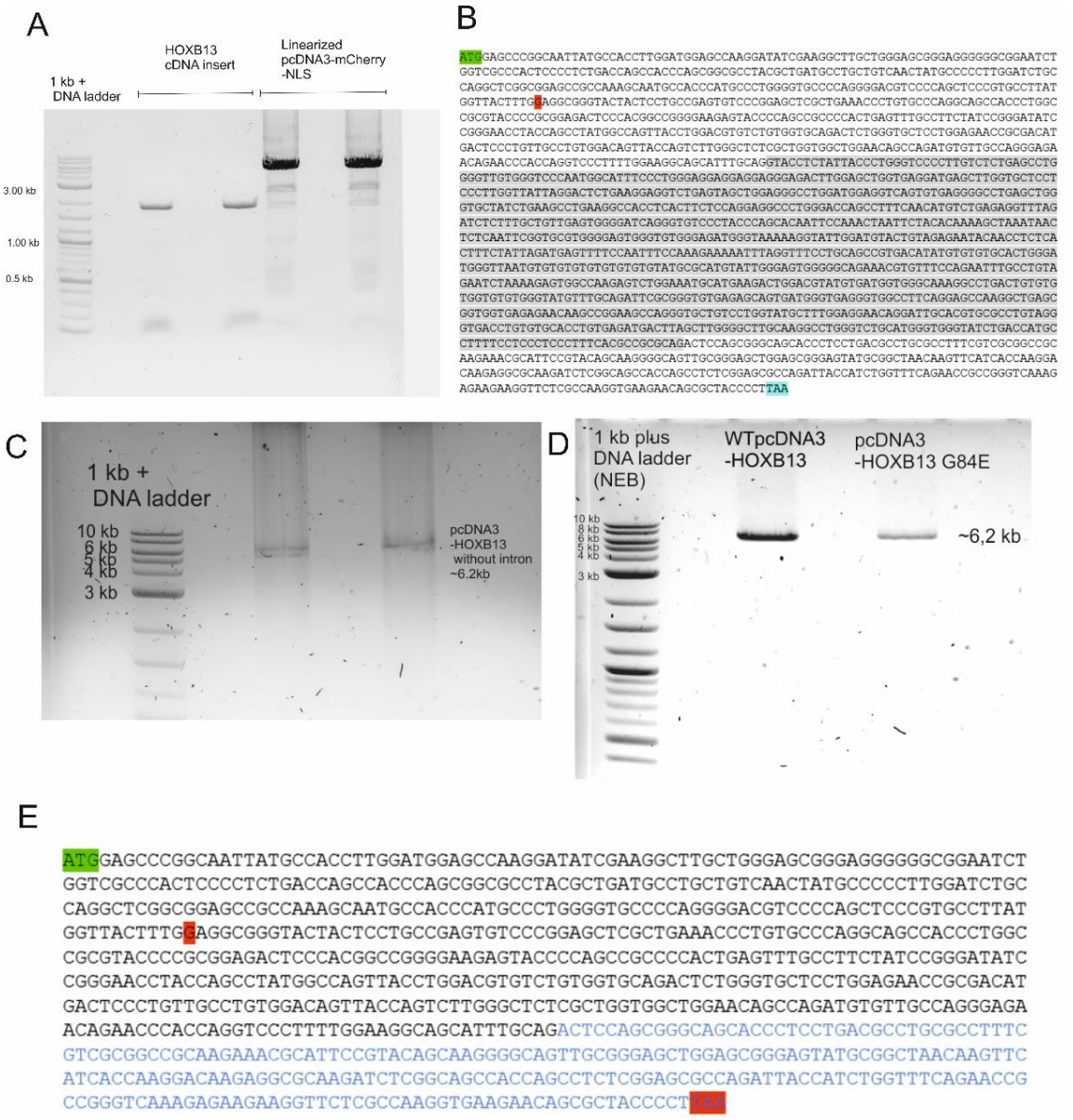


Figure 15 The pcDNA3-HOXB13 plasmid was generated and the G84E mutation was introduced so the plasmid could be used as a positive control of the mutation in the droplet digital PCR (ddPCR) assay. *A= mCherry-NLS was removed from the pcDNA3 plasmid backbone with PCR and the HOXB13 insert amplified from cDNA. PCR products were run on a 1 % agarose gel and combined with a In Fusion -cloning reaction. The HOXB13 insert was ~1,8 kb even though it should be 949 bp, and results from Sanger sequencing revealed that the intron was retained between the two exons. B= DNA sequence of the HOXB13 cDNA insert with retained intron between the two exons. Transcription start codon and termination codon are illustrated with green and blue, the desired mutation site (G) in red and intron sequence with grey background. C= the intron was removed with PCR and the correct size of the product (6,2 kb) was verified by running it on a 1 % agarose gel. D= Site-directed mutagenesis introduced the G84E mutation in the wild type (WT) pcDNA3-HOXB13 plasmid. Both plasmids containing the mutation and the WT plasmid were run on a 1 % agarose gel. E= Removal of the intron was*

verified with Sanger sequencing. Start and termination codons are illustrated in green and red and the mutation site with red. Exon 1 and 2 are illustrated with different colors.

4.2 All PCa cell lines express HOXB13

Western blot was performed on protein lysates of PCa cell lines (LNCaP, PC-3, vCaP, 22Rv1, MDA PCa 2b) and a normal immortalized prostate epithelial cell line RWPE-1 to determinate which of them express HOXB13. A protein lysate from LNCaP cells transfected with the WT-pcDNA3-HOXB13 construct was used as a positive control. β-actin was used as a loading control and the data was normalized with its signal. Results revealed that all PCa cell lines express HOXB13, whereas RWPE-1 does not (Figure 16). The MDA PCa 2b -cell line has the highest HOXB13 expression (2,5 ± 2,1), and PC-3 (0,3 ± 0,2) and 22Rv1 (0,2 ± 0,1) have the lowest. The results are presented as a relative HOXB13 expression and normalized with LNCaP (LNCaP relative expression = 1). The results are consistent with the Broad Institute HOXB13 RNASeq data (Appendix Figure 1) (https://portals.broadinstitute.org/ccle).

Based on the HOXB13 expression data, LNCaP cells were chosen for further experiments.

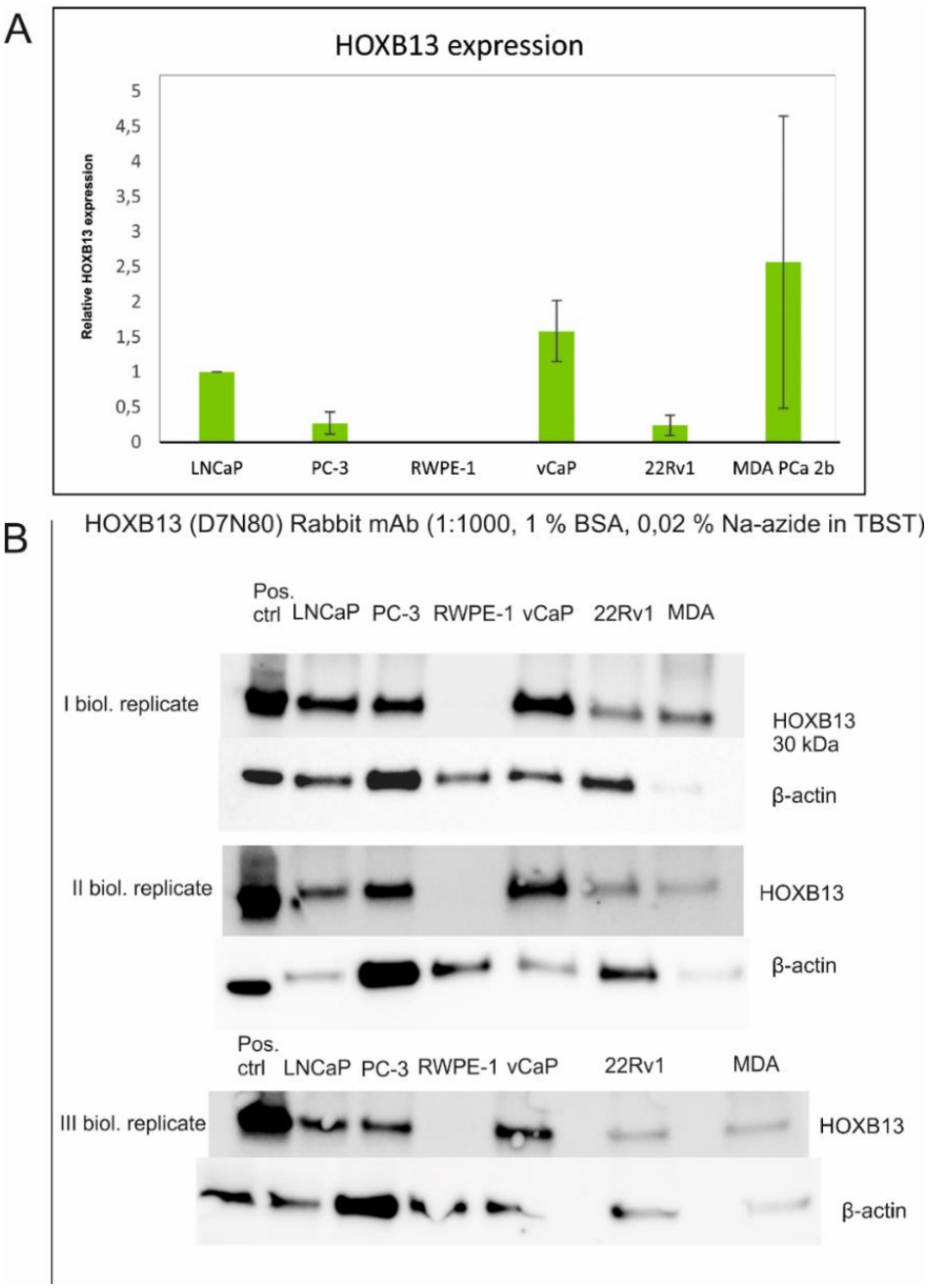


Figure 16 Prostate cancer cell lines express HOXB13 and RWPE-1 does not. *Western blot was performed on protein lysates of six different cell lines and for positive control of the protein (lysate extracted from LNCaP cells after 24 h transfecting them with pcDNA3-HOXB13 plasmid). β-actin was used as a loading control. The results from Western blot were normalized with loading control and graph A shows relative HOXB13 expression in all six cell lines.*

4.3 Prime editing of LNCaP cells

LNCaP cells were transfected with pegRNA-, sgRNA- and PE -plasmids. The PE and GFP is separated by a 2A self-cleaving peptide in the plasmid. This leads to separation of the PE and GFP proteins during translation and enables the enrichment of transfected cells with a cell sorter based on the GFP-signal.

Only a small amount of cells was obtained from the sorting (Neg C: 12000, PBS10: 3200, PBS13: 4000, PBS16: 12000), and the cells grew very slowly and many of them died when transferred to a 96-well plate, and therefore, only a small amount of DNA was gained from the isolation (concentration was 8 ng/µl for DNA extracted from cells transfected with pegRNA-PBS13 and 30 – 40 ng/µl for others).

4.4 Prime editing was successful in generating the HOXB13^{G84E} mutation

As shown by Anzalone et al., PBS length may have an impact on how efficiently prime editing occurs in the cells at different loci. The ddPCR assay was performed on DNA extracted from cells transfected with different pegRNA-constructs to measure if the PBS length affects prime editing efficiency at this specific locus. The assay was first performed for a positive control containing DNA isolated from LNCaP cells transfected with pcDNA3-HOXB13_G84E (Appendix Figure 2). The assay was done twice for the samples (two technical replicates) (Figure 17). The first time, smaller amount of template DNA was used (1 µl) but second time the amount was increased (5 µl).

Results from the first assay (Figure 17A), showed that samples transfected with pegRNA-PBS10 did not have the mutation (fractional abundance 0 %), and those transfected with pegRNA-PBS13 and pegRNA-PBS16 did (fractional abundance average 24,5 ± 10,4 % and 9,3 ± 1,0 %).

The second assay (Figure 17B) showed similar results with pegRNA-PBS10 and pegRNA-PBS 16 samples (fractional abundances 0 % and 11,1 ± 1,5 %, respectively), but the efficiency of pegRNA-PBS13 was 46,3 ± 2,1 % this time. Fractional abundance of the mutation was 99,8 % in the positive control and 0 % in the negative control in both assays.

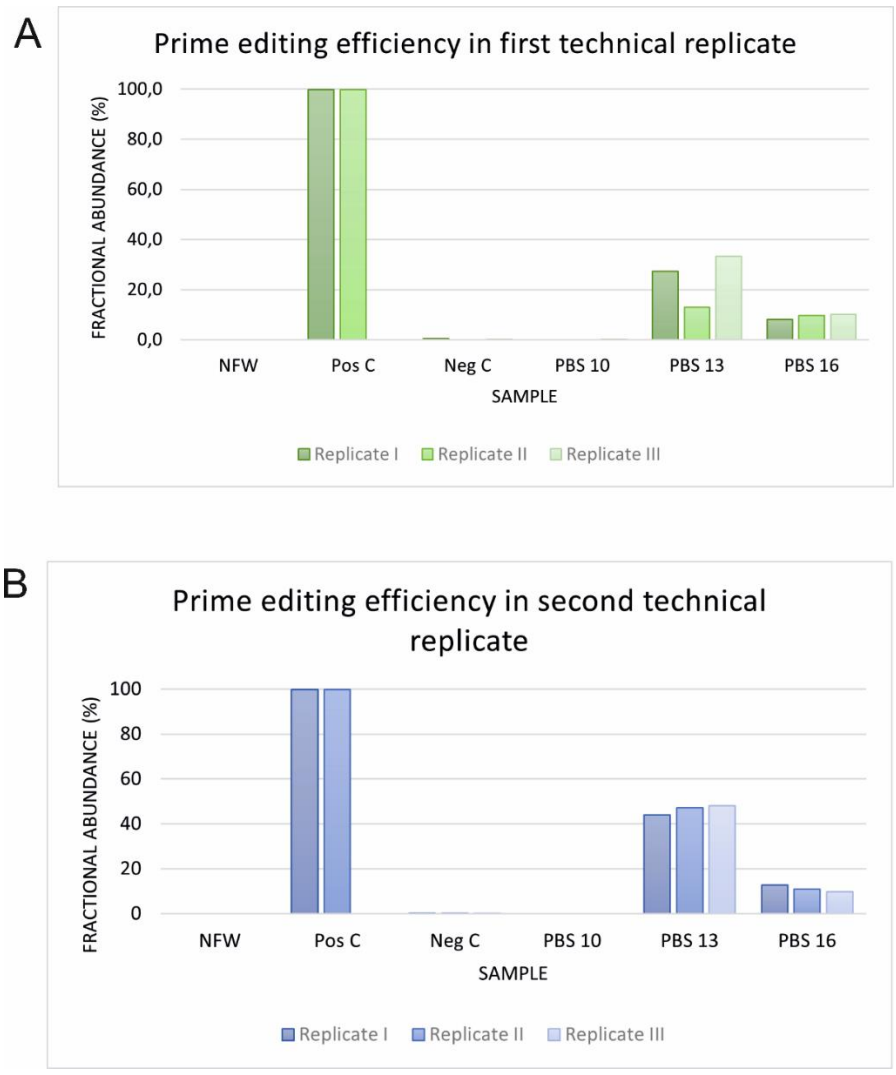


Figure 17 Prime editing efficiencies with three prime editing guide RNAs (pegRNAs) containing different length primer binding sites (PBS) were checked with droplet digital PCR (ddPCR). *A= results from first assay with smaller amount of template. B= results from second assay with five fold template amount. Assay contained two wells (replicates) of nuclease free water (NFW), positive control (Pos C, DNA from LNCaP cells transfected with pcDNA3-HOXB13_G84E) and negative control (Neg C, cells transfected with only prime editor and not with pegRNA) and three wells per DNA sample extracted from LNCaP cells transfected with pegRNA with PBS of 10, 13 or 16 nucleotides.*

The average of all results from both assays showed that the pegRNA with PBS16 seems to work with 10,2 ± 1,5 % efficiency. The pegRNA with PBS13 seems to be working also, since the average of fractional abundance was 35,4 %, but the standard deviation was very high (13,7). PBS10, however does not seem to be working at all, since the efficiency from all replicates is 0 % (Figure 18).

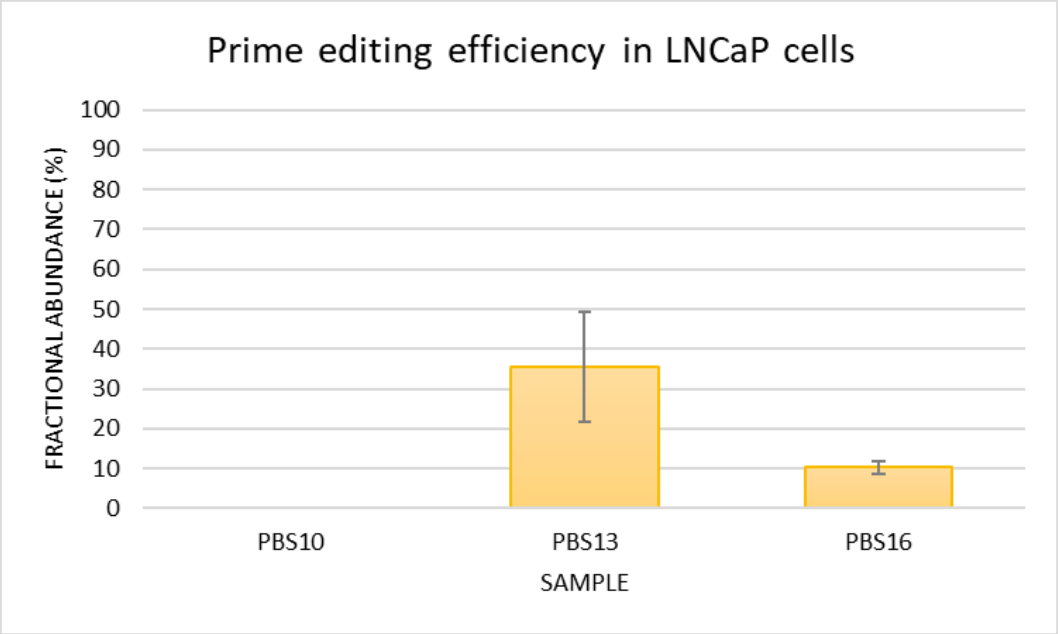


Figure 18 Prime editing worked with two of the three generated prime editing guide RNAs (pegRNAs). *Graph shows the average of fractional abundances of the G84E mutation in LNCaP cells transfected with prime editing plasmids. The pegRNA plasmids contained different length primer binding sites (PBS): 10, 13 or 16 nucleotides. Droplet digital PCR (ddPCR) assay was performed two times and both assays contained three technical replicates of the sample, and the average is from six technical replicates. Standard deviations can be seen in the graph.*

Because of the high variation in the results from pegRNA-PBS13 transfected cells, extracted DNA was amplified with PCR and sample was sent for Sanger sequencing. Results showed that the mutation was present, however, the WT was far more abundant compared to the mutation (Figure 19). The true efficiency of prime editing with pegRNA containing the 13 nt long PBS remains unclear.

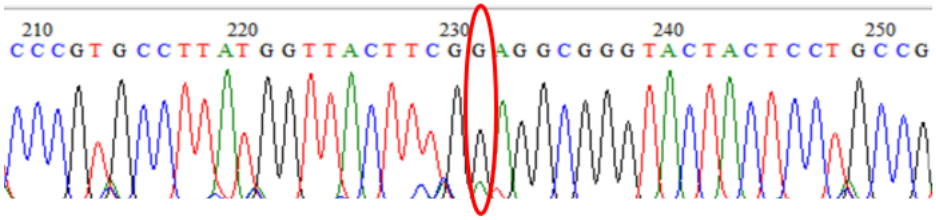


Figure 19 The wild type HOXB13 sequence is far more abundant than the mutation in LNCaP cells transfected with prime editing guide RNA (pegRNA) with primer binding site of 13 nucleotides. *DNA extracted from transfected LNCaP cells was amplified with PCR and a sample was sent for Sanger sequencing. The desired mutation site (G to A) is visualized in the graph.*

5. Discussion

5.1 HOXB13 protein expression in prostate cell lines

Western blot was performed on protein lysates and β-actin was used as a loading control for normalization of the signal. Results showed that HOXB13 is expressed in AR positive PCa cell lines, which have a luminal epithelial origin (LNCaP, vCaP, 22Rv1, MDA PCa 2b), but also in PC-3 cells that are AR negative and small cell origin. RWPE-1, a normal immortalized prostate epithelial cell line expressing minuscule amounts of AR, does not express HOXB13. Similar results have been previously reported with some of the cell lines used in this study (Yao et al., 2019).

MDA PCa 2b cell line has the highest expression level of HOXB13, but the standard deviation between the three biological replicates is very high as can be seen from the results (Figure 16). When these results were compared with Broad Institute RNASeq data, they were consistent, except that LNCaP had higher RPKM (Reads Per Kilobase Million) -value (155,5) than vCaP (136,8) whereas Western blot results indicated that LNCaP expressed less HOXB13. However, the standard deviation between the three biological replicates in Western blot results is quite high among all samples, and the loading is not evenly distributed as can be seen from the β-actin loading control. The loading pattern is however similar in all three replicates, which would suggest that there was probably something wrong with the protein concentration determination or the cell lines express β-actin quite differently. However, the results are normalized and the errors from loading differences should not affect the final results.

Based on these results, it was determined that the LNCaP cell line express HOXB13 and they could be used in the further experiments and were transfected with the prime editing plasmids.

5.2 Transfection and cell sorting affected LNCaP cell viability

The cells were transfected with prime editing constructs, but they were mostly growing in clumps and not as a monolayer, and therefore many of them were discarded in the sorting process. GFP-positive cells were sorted since the PE plasmid contained GFP, and

it functioned as a marker for identifying the transfected cells. The amount of the GFP-positive cells obtained from the sorting for all four differently transfected cells (pegRNA-PBS10, pegRNA-PBS13, pegRNA-PBS16 and negative control (PE only)) was very small and in addition to that, cells may have suffered from the transfection and from the sorting, and therefore they were growing very slowly and many of them died. Because of the small number of cells, the DNA concentration was very low, and the quality was not that good which affected the results obtained from ddPCR.

5.3 PBS length is important for prime editing efficiency

The fractional abundance of the G84E mutation compared to WT DNA was measured with ddPCR assay. DNA from LNCaP cells transfected with the pcDNA3-HOXB13_G84E plasmid was used as a positive control for the assay to verify the performance and for identifying the expected cluster positions for thresholding. The concentration of the mutation in the sample was very high compared to the WT DNA which could be seen from the ddPCR data (Appendix Figure 2). NFW was used to ensure there were not any contamination, and negative control with WT DNA was used to monitor false positives.

The assay was performed twice for the DNA from the prime edited cells. The first time the DNA amount was very low, especially for the PBS13 pegRNA (8 ng), and the amount of positive droplets in the assay was therefore very low and this caused high variation among the three wells containing the template (Figure 17A). The template amount was increased in the second assay (five-fold amount), and the efficiencies seemed to be even higher and with smaller standard deviations. However, the quality of the extracted DNA is not very good and the results from the assay do not seem to be that accurate when it comes to the PBS13 samples because the overall variation among samples was very high. Based on the results obtained from the second ddPCR assay, PCR was performed for the DNA extracted from cells transfected with pegRNA-PBS13 and a sample was sent for sequencing because if the efficiency would really be as high as the assay results indicated, the mutation would be more abundant. However, the WT DNA (G) was far more abundant than the mutation (A) and therefore, the efficiency of pegRNA with PBS13 remains unclear, but based on the ddPCR and the sequencing results, it looks like it would be working.

The average efficiency for PBS16 was 9,3 % in the first assay with small standard deviation, and the results from the second assay were consistent with the first assay

(average 11,1 %). Based on these results from both assays, it would seem like the pegRNA with PBS16 is working with 10 ,2 % efficiency.

The pegRNA with PBS10 did not work at all based on the results. The reason presumably being that the PBS is too short for proper binding to the target DNA.

These results show the importance of optimizing the PBS in the pegRNA since only a difference of three bases can lead to big differences in the editing efficiencies. The efficiency of PBS13 remains unknown, but the pegRNA with PBS16 can be used for further experiments.

It is also important to consider that prime editing is such a novel method and has not yet been used that much, and therefore optimization is needed in different cell types. Anzalone et al. used HEK293 cells which are easily transfected and reached 30 – 60 % efficiencies. However, when using HeLa cells, the prime editing efficiency was approximately 10 % and similar to the efficiency observed in this study. Studies where prime editing was used in plants also reported quite low efficiencies (<10 %) at most target sites (Jiang et al., 2020; Lin et al., 2020), so the efficiency of pegRNA with PBS16 used in here can be considered good.

Further studies for detecting unwanted byproducts would be needed in order to see indel frequencies.

6. Conclusions and future perspectives

The aim of the thesis was to generate tools for generating the G84E mutation in PCa cells with prime editing. For generating this mutation, the length of PBS of the pegRNA was optimized with three different constructs containing a PBS of 10, 13 or 16 nt, and their efficiencies were compared to one another. The pegRNA binds to the target strand where the mutation is introduced and therefore optimization is crucial for prime editing. Because the method is new and has not been used in many cell lines, it requires optimization of different components and editing conditions for achieving optimal efficiencies and for minimizing unwanted byproducts. Prime editing offers many possibilities but also challenges. They might not be capable of performing as large insertions and deletion as conventional CRISPR-Cas9 systems and because the edit is introduced with RT using RNA template, the stability of the RNA molecule might cause difficulties especially when it comes to long RNA molecules. Therefore, additional research is required. However, prime editing enables precise targeted editing with wider scope when compared to other gene editing tools and reduces the need for additional DNA templates and is not dependable on the HDR.

The germline mutation G84E in HOXB13 was chosen for the experiments because it has been associated with increased PCa risk and it is a Finnish founder mutation (Ewing et al., 2012; Xu et al., 2013). Even though its association in PCa susceptibility has been known since its discovery by Ewing et al. in 2012, its role and mechanism in cancer progression remains elusive. The mutation is located within the MEIS interaction domain, but it has been demonstrated that the interaction does not interfere with the interaction between HOXB13 and MEIS1 *in vitro* (Johng et al., 2019). HOXB13 has both oncogenic and tumour suppressing functions but further studies are needed to understand its complex mechanisms in both normal prostate development and in tumorigenesis. It is known that the gene is required for normal prostate development (Economides et al. 2003) and that it interacts with AR (Norris et al., 2009), but how this mutation manifests itself is not known. It is possible that it affects protein stability since glycine, a small hydrophobic amino acid is replaced with glutamate which is hydrophilic.

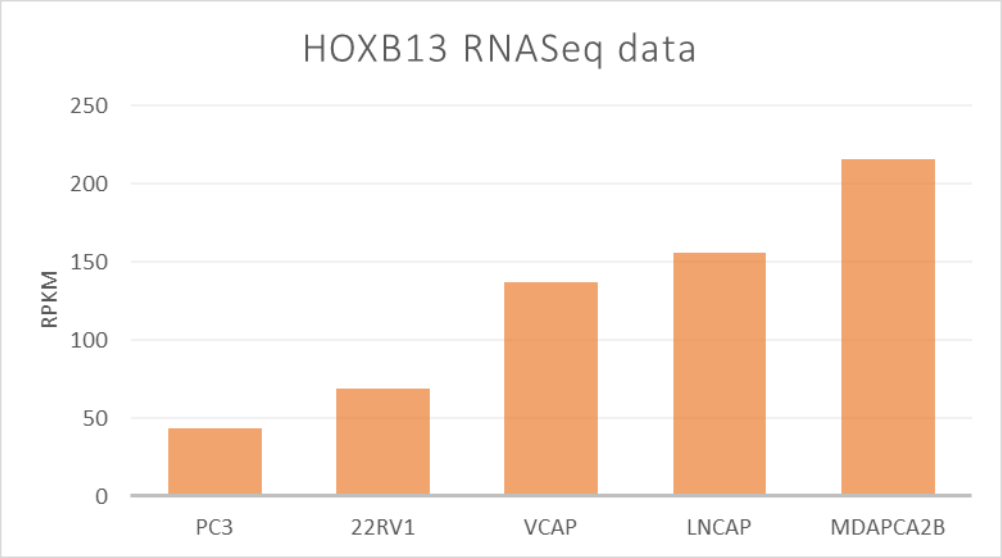
Cellular models of the G84E mutation in *HOXB13* are needed to further study its role and mechanistic function in cancer progression, and that is why this thesis focused on this topic. The results showed that LNCaP, PC-3, vCaP, 22Rv1 and MDA PCa 2b cell lines express *HOXB13* and that they can be used for studying this gene, whereas RWPE-1 did

not. LNCaP cells were chosen for prime editing of *HOXB13* to introduce this mutation, and from three different pegRNA constructs, the ones with PBS of 16 and 13 nt worked, and the one with PBS of 10 nt did not, suggesting that it may be too short for proper binding to the target site. The efficiency of PBS13 remains unclear, but pegRNA with PBS16 worked with 10 % efficiency and it can be used for generating cellular model of the mutation. Further experiments would include isolation of the cellular clones and sequencing them, and ultimately using these models for functional experiments to elucidate the mechanism in which mutated *HOXB13* promotes cancer progression. Checking the prime edited cells for unwanted byproducts would also be necessary for assessment of the off-target prime editing levels.

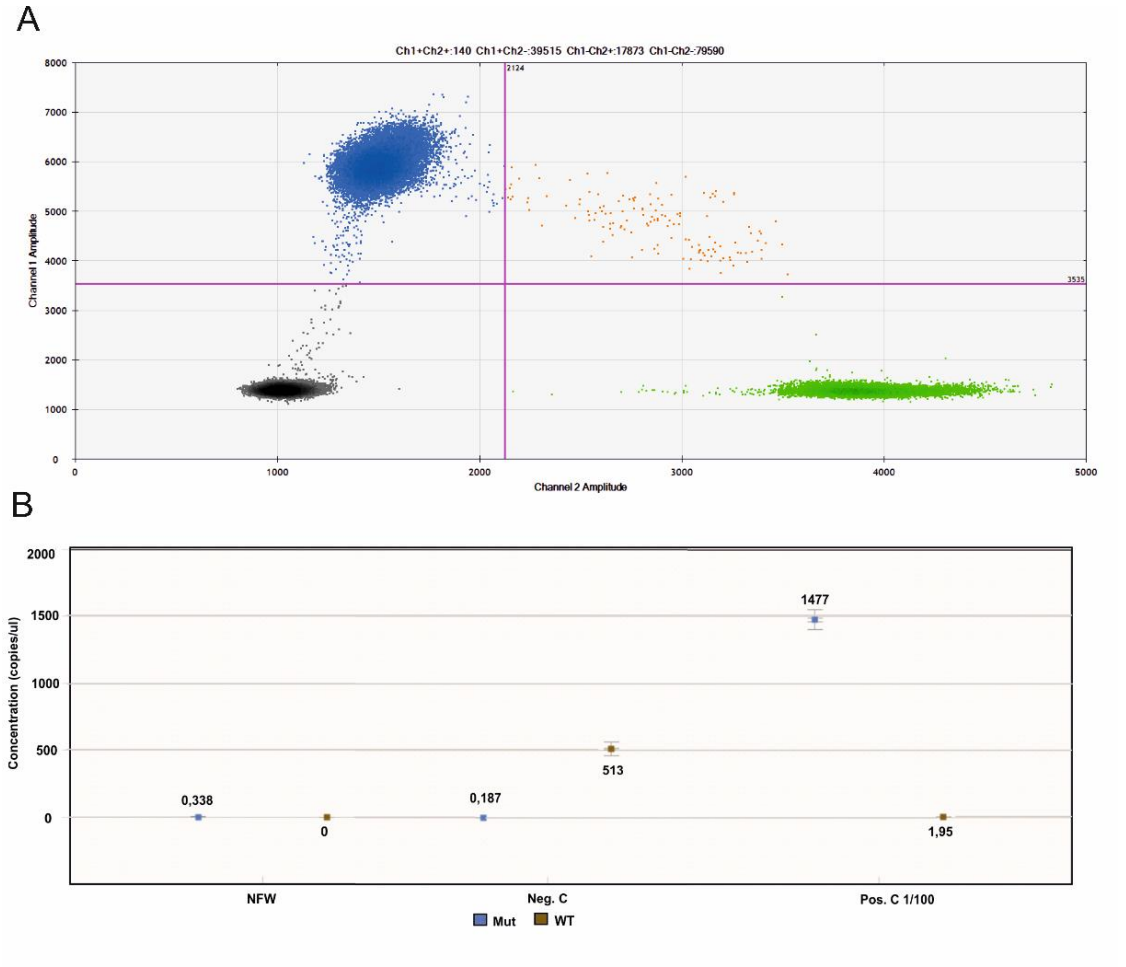
Understanding the role of this mutation in cancer progression is important for developing treatments for patients carrying the germline mutation especially when targeted therapies are paving the way in drug development. Since PCa is a highly heritable disease and many variants are a result of single nucleotide variants (SNVs), effective tools for correcting these SNVs would be required. Prime editing offers a promising alternative for this, but currently much more research remains before they can be used in the treatment of genetic disorders.

In summary, the results from this study implicate that prime editing can be efficiently used to introduce the G84E mutation into prostate cancer cells. Furthermore, the results show the importance of optimization the PBS length in the pegRNA and provides knowledge that can be exploited to generate cellular models of HOXB13 G84E with prime editing.

Appendix



Appendix Figure 1 RNASeq data of HOXB13 expression in PCa cell lines. *RNASeq data from Broad Institute Cancer cell line encyclopedia (https://portals.broadinstitute.org/ccle) was compared between five PCa cell lines and compared to protein expression levels of HOXB13. RPKM= Reads Per Kilobase Million.*



Appendix Figure 2 Droplet digital PCR (ddPCR) assay was performed to LNCaP cells transfected with pcDNA3-HOXB13_G84E plasmid. *A= DNA droplets are clustered into four groups: FAM-(mut) and HEX-(WT) negative (black spots), FAM-positive (blue spots), HEX-positive (green spots) and FAM- and HEX-positive (orange spots). B= Concentration results from ddPCR assay for positive control. Concentrations are plotted as copies per µl Three replicates per sample was used. Neg C= negative control, Pos C positive control diluted 1:100, NFW Nuclease free water.*

References

Abate-Shen, C. (2002) Deregulated homeobox gene expression in cancer: cause or consequence? *Nat Rev Cancer* **2**:777-785.

Abe, M., Hamada, J., Takahashi, O., Takahashi, Y., Tada, M., Miyamoto, M., Morikawa, T., Kondo, S. & Moriuchi, T. (2006) Disordered expression of HOX genes in human non-small cell lung cancer. *Oncol Rep* **15**:797–802.

Albany, C., Alva, A.S., Aparicio, A.M., Singal, R., Yellapragada, S., Sonpavde, G. & Hahn, N.M. (2011) Epigenetics in Prostate Cancer. *Prostate cancer* **2011**:580318. doi: 10.1155/2011/580318.

Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A. & Liu, D.R. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**:149-157.

Armenia, J., Wankowicz, S.A.M., Liu, D., Gao, J., Kundra, R., Reznik, E., Chatila, W.K., Chakravarty, D., Han, G.C., Coleman, I., Montgomery, B., Pritchard, C., Morrissey, C., Barbieri, C.E., Beltran, H., Sboner, A., Zafeiriou, Z., Miranda, S., Bielski, C.M., Penson, A.V., Tolonen, C., Huang, F.W., Robinson, D., Wu, Y.M., Lonigro, R., Garraway, L.A., Demichelis, F., Kantoff, P.W., Taplin, M.E., Abida, W., Taylor, B.S., Scher, H.I., Nelson, P.S., de Bono, J.S., Rubin, M.A., Sawyers, C.L., Chinnaiyan, A.M; PCF/SU2C International Prostate Cancer Dream Team, Schultz, N. & Van Allen, E.M. (2018) The long tail of oncogenic drivers in prostate cancer. *Nat Genet* **50**:645-651.

Aune, D., Navarro Rosenblatt, D.A., Chan, D.S., Vieira, A.R., Vieira, R., Greenwood, D.C., Vatten, L.J. & Norat, T. (2015) Dairy products, calcium, and prostate cancer risk: a systematic review and meta-analysis of cohort studies. *Am J Clin Nutr* **101:**87-117.

Bahrani-Mostafavi, Z., Tickle, T.L., Zhang, J., Bennett, K.E., Vachris, J.C., Spencer, M.D., Mostafavi, M.T. & Tait, D.L. (2008) Correlation Analysis of HOX, ErbB and IGFBP Family Gene Expression in Ovarian Cancer. *Cancer Invest* **26**:990-998.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. & Horvath, P. (2007) CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315:**1709–1712. doi:10.1126/science.1138140.

Bechis, S.K., Carroll, P.R., & Cooperberg, M.R. (2011) Impact of age at diagnosis on prostate cancer treatment and survival. *J Clin Oncol* **29:**235–241.

Bhatlekar, S., Fields, J.Z. & Boman, B.M. (2014) HOX genes and their role in the development of human cancers. *J Mol Med* **92**.

Bosch, J.A., Birchak, G. & Perrimon, N. (2021) Precise genome engineering in Drosophila using prime editing. *Proc Natl Acad Sci USA* **118:**e2021996118; DOI: 10.1073/pnas.2021996118

Brechka, H., Bhanvadia, R.R., VanOpstall, C. & Vander Griend, D.J. (2017) HOXB13 mutations and binding partners in prostate development and cancer: Function, clinical significance, and future directions. *Genes Dis* **4**:75-87.

Breyer, J.P., Avritt, T.G., McReynolds, K.M., Dupont, W.D. & Smith, J.R. (2012) Confirmation of the HOXB13 G84E germline mutation in familial prostate cancer. *Cancer Epidemiol Biomarkers Prev* **21**:1348-1353.

Bridges, C. B. (1921) Current maps of the location of the mutant genes of Drosophila Melanogaster. *Proc Natl Acad Sci USA* **7:** 127–132.

Cantile, M., Pettinato, G., Procino, A., Feliciello, I., Cindolo, L. & Cillo, C. (2003) In vivo expression of the whole HOX gene network in human breast cancer. *Eur J Cancer* **39**:257-264.

Cardoso, M., Maia, S., Paulo, P. & Teixeira, M.R. (2016) Oncogenic mechanisms of HOXB13 missense mutations in prostate carcinogenesis. *Oncoscience* **3**:288-296.

Chandrasekaran, G., Hwang, E.C., Kang, T.W., Kwon, D.D., Park, K., Lee, J. & Lakshmanan, V. (2017) Computational Modeling of complete HOXB13 protein for predicting the functional effect of SNPs and the associated role in hereditary prostate cancer. *Sci Rep* **7:**43830. doi: 10.1038/srep43830.

Chan, J.M., Gann, P.H. & Giovannucci, E.L. (2005) Role of diet in prostate cancer development and progression. *J Clin Oncol* **23:**8152-8160.

Cheng, W., Liu, J., Yoshida, H., Rosen, D. & Naora, H. (2005) Lineage infidelity of epithelial ovarian cancers is controlled by HOX genes that specify regional identity in the reproductive tract. *Nat Med* **11**:531-537.

Chen, Y.C., Page, J.H., Chen, R. & Giovannucci, E. (2008) Family history of prostate and breast cancer and the risk of prostate cancer in the PSA era. *Prostate* **68**:1582-1591

Chu, M. C., Selam, F. B. & Taylor, H. S. (2004) HOXA10 regulates p53 expression and matrigel invasion in human breast cancer cells. *Cancer Biol Ther* **3**:568–572.

Dai, C., Heemers, H. & Sharifi, N. (2017) Androgen Signaling in Prostate Cancer. *Cold Spring Harb Perspect Med* **7:**a030452. doi: 10.1101/cshperspect.a030452.

De Braekeleer, E., Douet-Guilbert, N., Basinko, A., Le Bris, M., Morel, F. & De Braekeleer, M. (2014) Hox gene dysregulation in acute myeloid leukemia. *Future Oncol* **10**:475-495.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**:602-607.

Economides, K. & Capecchi, M. (2003) Hoxb13 is required for normal differentiation and secretory function of the ventral prostate. *Development* **130**:2061-2069.

Ewing, C.M., Ray, A.M., Lange, E.M., Zuhlke, K.A., Robbins, C.M., Tembe, W.D., Wiley, K.E., Isaacs, S.D., Johng, D., Wang, Y., Bizon, C., Yan, G., Gielzak, M., Partin, A.W., Shanmugam, V., Izatt, T., Sinari, S., Craig, D.W., Zheng, S.L., Walsh, P.C., Montie, J.E., Xu, J., Carpten, J.D., Isaacs, W.B. & Cooney, K.A. (2012) Germline Mutations inHOXB13and Prostate-Cancer Risk. *New Engl J Med* **366**:141-149.

Faber, J., Krivtsov, A.V., Stubbs, M.C., Wright, R., Davis, T.N., Van Den Heuvel-Eibrink, M., Zwaan, C.M., Kung, A.L. & Armstrong, S.A. (2009) HOXA9 is required for survival in human MLL-rearranged acute leukemias. *Blood* **113:** 2375-2385.

Gleason, D.F. (1966) Classification of prostatic carcinoma. *Cancer Chemother Rep* **50:**125-128.

Grönberg, H. (1999). Prostate Cancer Susceptibility Genes: A Current Update. *The Prostate Journal* **1:**115–119. doi:10.1046/j.1525-1411.1999.09928.x

Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. (2018) CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med* **24**:927-930.

Hamid, S.M., Cicek, S., Karamil, S., Ozturk, M.B., Debelec-Butuner, B., Erbaykent-Tepedelen, B., Varisli, L., Gonen-Korkmaz, C., Yorukoglu, K. & Korkmaz, K.S. (2014) HOXB13 contributes to G1/S and G2/M checkpoint controls in prostate. *Mol Cell Endocrinol* **383**:38-47.

Hanahan, D. & Weinberg, R. (2011) Hallmarks of Cancer: The Next Generation. *Cell* **144**:646-674.

Hedayat, K.M. & Lapraz, J.C. (2019) The Theory of Endobiogeny. pp. 135-164, Academic Press, Massachusetts.

Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D. & Marraffini, L.A. (2015) Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature* **519:**199–202.

Henry, G.H., Malewska, A., Joseph, D.B., Malladi, V.S., Lee, J., Torrealba, J., Mauck, R.J., Gahan, J.C., Raj, G.V., Roehrborn, C.G., Hon, G.C., MacConmara, M.P., Reese, J.C., Hutchinson, R.C., Vezina, C.M. & Strand, D.W. (2018) A Cellular Anatomy of the Normal Adult Human Prostate and Prostatic Urethra. *Cell Rep* **25**:3530-3542.e5.

Hille, F., Richter, H., Wong, S.P., Bratovič, M., Ressel, S. & Charpentier, E. (2018) The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**:1239-1259.

Hjelmborg, J.B., Scheike, T., Holst, K., Skytthe, A., Penney, K.L., Graff, R.E., Pukkala, E., Christensen, K., Adami, H.O., Holm, N.V., Nuttall, E., Hansen, S., Hartman, M., Czene, K., Harris, J.R., Kaprio, J., Mucci, L.A. (2014) The heritability of prostate cancer in the Nordic Twin Study of Cancer. *Cancer Epidemiol Biomarkers Prev* **23**:2303-2310.

Huang, L., Pu, Y., Hepps, D., Danielpour, D. & Prins, G.S. (2007) Posterior Hox Gene Expression and Differential Androgen Regulation in the Developing and Adult Rat Prostate Lobes. *Endocrinology* **148**:1235-1245.

Huggins, C. & Hodges, C.V. (1941) Studies on prostatic cancer: I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. *Cancer Res* **1**:293–297.

Hung, Y.-C., Ueda, M., Terai, Y., Kumagai, K., Ueki, K., Kanda, K., Yamaguchi, H., Akise, D. and Ueki, M. (2003) *Homeobox* gene expression and mutation in cervical carcinoma cells. *Cancer Science* **94:**437-441.

Idaikkadar, P., Morgan, R. & Michael, A. (2019) HOX Genes in High Grade Ovarian Cancer. *Cancers* **11**:1107. doi:10.3390/cancers11081107.

Ikonen, T., Matikainen, M.P., Syrjakoski, K., Mononen, N., Koivisto, P.A., Rokman, A., Seppala, E.H., Kallioniemi, O.P., Tammela, T.L.J., & Schleutker, J. (2003). BRCA1 and BRCA2 mutations have no major role in predisposition to prostate cancer in Finland. *J Med Genet* **40**:e98. https://doi.org/10.1136/jmg.40.8.e98

International Agency for Research on Cancer, World Health Organization. Cancer Fact Sheets: Prostate Cancer, https://gco.iarc.fr/today/data/pdf/fact-sheets/cancers/cancer-fact-sheets-19.pdf

Ishino, Y., Shinagawa, H., Makino, K., Amemura M. & Nakata A. (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* **169**:5429-5433.

Ittmann, M. (2018) Anatomy and Histology of the Human and Murine Prostate. *Cold Spring Harbor Perspect Med* **8**.doi: 10.1101/cshperspect.a030346.

Jansen, R., van Embden, J.D.A., Gaastra, W. & Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**:1565-1575.

Jiang, Y., Chai, Y., Lu, M., Han, X., Lin, Q., Zhang, Y., Zhang, Q., Zhou, Y., Wang, X., Gao, C. & Chen, Q. (2020) Prime editing efficiently generates W542L and S621I double mutations in two ALS genes in maize. *Genome Biol* **21**:257. https://doi.org/10.1186/s13059-020-02170-5

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. & Charpentier, E. (2012) A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**:816-821.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A.T., Charpentier, E., Nogales, E. & Doudna, J.A. (2014) Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* **343**:1247997. doi:10.1126/science.1247997

Johng, D., Torga, G., Ewing, C.M., Jin, K., Norris, J.D., McDonnell, D.P. & Isaacs, W.B. (2019) HOXB13 interaction with MEIS1 modifies proliferation and gene expression in prostate cancer. *The Prostate* **79**:414-424.

Jung, C., Kim, R., Lee, S., Wang, C. & Jeng, M. (2004) HOXB13 Homeodomain Protein Suppresses the Growth of Prostate Cancer Cells by the Negative Regulation of T-Cell Factor 4. *Cancer Res* **64**:3046-3051.

Jung, C., Kim, R., Zhang, H., Lee, S. & Jeng, M. (2004) HOXB13 Induces Growth Suppression of Prostate Cancer Cells as a Repressor of Hormone-Activated Androgen Receptor Signaling. *Cancer Res* **64**:9185-9192.

Kaaks, R. & Stattin, P. (2010) Obesity, Endogenous Hormone Metabolism, and Prostate Cancer Risk: A Conundrum of "Highs" and "Lows". *Cancer Prev Res* **3:**259-262.

Kamps, M.P., Murre, C., Sun, X. & Baltimore, D. (1990) A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B all. *Cell* **60**:547-555.

Kim, Y., Oh, K., Park, R., Xuan, N.T., Kang, T., Kwon, D., Choi, C., Kim, M.S., Nam, K.I., Youn Ahn, K. & Jung, C. (2010) HOXB13 promotes androgen independent growth of LNCaP prostate cancer cells by the activation of E2F signaling. *Mol Cancer* **9**: 124. doi: 10.1186/1476-4598-9-124.

Kolonel, L., Altshuler, D. & Henderson, B. (2004) The multiethnic cohort study: exploring genes, lifestyle and cancer risk. *Nat Rev Cancer* **4:**519–527

Laitinen, V.H., Wahlfors, T., Saaristo, L., Rantapero, T., Pelttari, L.M., Kilpivaara, O., Laasanen, S.L., Kallioniemi, A., Nevanlinna, H., Aaltonen, L., Vessella, R.L., Auvinen, A., Visakorpi, T., Tammela, T.L., Schleutker, J. (2013) HOXB13 G84E Mutation in Finland: Population-Based Analysis of Prostate, Breast, and Colorectal Cancer Risk. *Cancer Epidemiol Biomarkers Prev* **22**:452-460.

Lecarpentier, J., Silvestri, V., Kuchenbaecker, K.B., Barrowdale, D., Dennis, J., Mcguffog, L., Soucy, P., Leslie, G., Rizzolo, P., Navazio, A.S., Valentini, V., Zelli, V., Lee, A., Amin, A., Olama, A., Tyrer, J.P., Southey, M., John, E.M., Conner, T.A., Goldgar, D.E., Buys, S.S., Janavicius, R., Steele, L., Ding, C., Neuhausen, S.L., Hansen, T.V.O., Osorio, A., Weitzel, J.N., Toss, A., Medici, V., Cortesi, L., Zanna, I., Palli, D., Radice, P., Manoukian, S., Peissel, B., Azzollini, J., Viel, A., Cini, G., Damante, G., Tommasi, S., Peterlongo, P., Fostira, F., Hamann, U., Gareth, D., Henderson, A., Brewer, C., Eccles, D., Cook, J., Ong, K., Walker, L., Side, L.E., Porteous, M.E., Davidson, R., Hodgson, S., Frost, D., Adlard, J., Izatt, L., Eeles, R., Ellis, S., Tischkowitz, M., Godwin,

A.K., Meindl, A., Gehrig, A., Dworniczak, B., Sutter, C., Engel, C., Niederacher, D., Steinemann, D., Hahnen, E., Hauke, J., Rhiem, K., Kast, K., Arnold, N., Ditsch, N., Wang-Gohrke, S., Wappenschmidt, B., Wand, D., Lasset, C., Stoppa-Lyonnet, D., Belotti, M., Damiola, F., Barjhoux, L., Mazoyer, S., Study, G., Van Heetvelde, M., Poppe, B., De Leeneer, K., Claes, K.B.M., De La Hoya, M., Garcia-Barberan, V., Caldes, T., Segura, P.P., Kiiski, J.I., Aittomäki, K., Khan, S., Nevanlinna, H., EMBRACE , GEMO Study Collaborators , HEBON & KConFab Investigators (2017) Prediction of Breast and Prostate Cancer Risks in Male BRCA1 and BRCA2 Mutation Carriers Using Polygenic Risk Scores. *J Clin Oncol* **35**. https://doi.org/10.1200/JCO.2016.69.4935

Lee, J., Demissie, K., Lu, S.E. and Rhoads, G.G. (2007) Cancer incidence among Korean-American immigrants in the United States and native Koreans in South Korea. *Cancer Control* **14**:78-85.

Li, H., Yang, Y., Hong, W., Huang, M., Wu, M. & Zhao, X. (2020) Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduct Target Ther* **5**:1. doi:10.1038/s41392-019-0089-y.

Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., Wang, Y., Anzalone, A.V., Raguram, A., Doman, J.L., Liu, D.R. & Gao, C. (2020) Prime genome editing in rice and wheat. *Nat Biotechnol* **38**:582-585.

Lino, C.A., Harper, J.C., Carney, J.P. & Timlin, J.A. (2018) Delivering CRISPR: a review of the challenges and approaches. *Drug Deliv* **25**:1234-1257.

Litwin, M.S. & Tan, H. (2017) The Diagnosis and Treatment of Prostate Cancer: A Review. *JAMA* **317**:2532-2542.

Luo, M.L., Leenay, R.T. & Beisel, C.L. (2016) Current and future prospects for CRISPR-based tools in bacteria. *Biotechnol Bioeng* **113**:930-943.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Scott, D., Shah, S.A., Siksnys, V., Terns, M.P., Venclovas, Č., White, M.F., Yakunin, A.F., Yan, W., Zhang, F., Garrett, R.A., Backofen, R., van der Oost, J., Barrangou, R., Koonin, E.V. (2020) Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* **18:** 67–83.

Mann, R.S., Lelli, K.M. & Joshi, R. (2009) Hox Specificity unique roles for cofactors and collaborators. *Curr Top Dev Biol* **88**:63-101

Matsoukas, I.G. (2020) Prime Editing: Genome Editing for Rare Genetic Diseases Without Double-Strand Breaks or Donor DNA. *Frontiers Genet* **11**:528. https://doi.org/10.3389/fgene.2020.00528

McNeal, J.E. (1968) Regional morphology and pathology of the prostate *Am J Clin Pathol* **49:**347-357

Miao, J., Wang, Z., Provencher, H., Muir, B., Dahiya, S., Carney, E., Leong, C.O., Sgroi, D.C. & Orsulic, S. (2007) HOXB13 Promotes Ovarian Cancer Progression. *PNAS* **104**:17093-17098.

Miller, G.J., Miller, H.L., Van Bokhoven, A., Lambert, J.R., Werahera, P.N., Schirripa, O., Lucia, M.S. & Nordeen, S.K. (2003) Aberrant HOXC Expression Accompanies the Malignant Phenotype in Human Prostate 1. *Cancer Research* **63**: 5879-5888.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**:174-182.

Mojica, F.J., Díez-Villaseñor, C., Soria, E. and Juez G. (2000) Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* **36**:244-6

Nevedomskaya, E., Baumgart, S.J. & Haendler, B. (2018) Recent advances in prostate cancer treatment and drug discovery. *Int J Mol Sci* **19**:1359.

Nishimasu, H., Ran, F., Hsu, P., Konermann, S., Shehata, S., Dohmae, N., Ishitani, R., Zhang, F. & Nureki, O. (2014) Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156**:935-949.

Norris, J.D., Chang, C., Wittmann, B.M., Kunder, R.S., Cui, H., Fan, D., Joseph, J.D. & McDonnell, D.P. (2009) The Homeodomain Protein HOXB13 Regulates the Cellular Response to Androgens. *Mol Cell* **36**:405-416.

Nourse, J., Mellentin, J.D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S.D. & Cleary, M.L. (1990) Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. *Cell* **60**:535-545.

Pritchard, C.C., Mateo, J., Walsh, M.F., Walsh, T., De Sarkar, N., Abida, W., Beltran, H., Garofalo, A., Gulati, R., Carreira, S., Eeles, R., Elemento, O., Rubin, M.A., Robinson, D., Lonigro, R., Hussain, M., Chinnaiyan, A., Vinson, J., Filipenko, J., Garraway, L., Taplin, M., AlDubayan, S., Han, G.C., Beightol, M., Morrissey, C., Nghiem, B., Cheng, H.H., Montgomery, B., Casadei, S., Berger, M., Zhang, L., Zehir, A., Vijai, J., Scher, H.I., Sawyers, C., Schultz, N., Kantoff, P.W., Solit, D., Robson, M., Van Allen, E.M., Offit, K., de Bono, J. & Nelson, P.S. (2016) Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *New Engl J Med* **375**:443-453.

Porto, E.M., Komor, A.C., Slaymaker, I.M. and Yeo, G.W. (2020) Base editing: advanced and therapeutic opportunities. *Nature Rev Drug Disc* **19:** 839 – 859.

Rath, D., Amlinger, L., Rath, A. & Lundgren, M. (2015) The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie* **117:**119–128.

Rauch, T., Wang, Z., Zhang, X., Zhong, X., Wu, X., Lau, S.K., Kernstine, K.H., Riggs, A.D. & Pfeifer, G.B. (2007) Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *PNAS* **104**:5527-5532.

Rawla, P. (2019) Epidemiology of Prostate Cancer. *World J Oncol* **10:**63–89.

Reisman, D., Evron, E., Sukumar, S., Raman, V., Odenwald, W.F., Martensen, S.A., Marks, J. & Jaffee, E. (2000) Compromised HOXA5 function can limit p53 expression in human breast tumours. *Nature* **405**:974-978.

Rice, M.A., Malhotra, S.V. & Stoyanova, T. (2019) Second-Generation Antiandrogens: From Discovery to Standard of Care in Castration Resistant Prostate Cancer. *Front Oncol* **9**:801. doi: 10.3389/fonc.2019.00801.

Robinson, D., Van Allen, E., Wu, Y., Schultz, N., Lonigro, R., Mosquera, J., Montgomery, B., Taplin, M., Pritchard, C., Attard, G., Beltran, H., Abida, W., Bradley, R., Vinson, J., Cao, X., Vats, P., Kunju, L., Hussain, M., Feng, F., Tomlins, S., Cooney, K., Smith, D., Brennan, C., Siddiqui, J., Mehra, R., Chen, Y., Rathkopf, D., Morris, M., Solomon, S., Durack, J., Reuter, V., Gopalan, A., Gao, J., Loda, M., Lis, R., Bowden, M., Balk, S., Gaviola, G., Sougnez, C., Gupta, M., Yu, E., Mostaghel, E., Cheng, H., Mulcahy, H., True, L., Plymate, S., Dvinge, H., Ferraldeschi, R., Flohr, P., Miranda, S., Zafeiriou, Z., Tunariu, N., Mateo, J., Perez-Lopez, R., Demichelis, F., Robinson, B.,

Schiffman, M., Nanus, D., Tagawa, S., Sigaras, A., Eng, K., Elemento, O., Sboner, A., Heath, E., Scher, H., Pienta, K., Kantoff, P., de Bono, J., Rubin, M., Nelson, P., Garraway, L., Sawyers, C. & Chinnaiyan, A. (2015) Integrative Clinical Genomics of Advanced Prostate Cancer. *Cell* **161**:1215-1228.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P. & Cardona, A. (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9:**676–682. https://doi.org/10.1038/nmeth.2019

Schröder, F.H., M.D (2010) Prostate cancer around the world. An overview. *Urol Oncol* **28**:663-667.

Scott, M.P. (1992) Vertebrate homeobox gene nomenclature. *Cell* **71**:551–553

Shah, N. & Sukumar, S. (2010) The Hox genes and their roles in oncogenesis. *Nature Rev* **10**:361-371.

Sharma, S., Zapatero-Rodríguez, J. & Kennedy, R. (2017) Prostate cancer diagnostics: clinical challenges and the ongoing need for disruptive and effective diagnostic tools. *Biotechnol Adv* **35**:135-149.

Sharma, S., Kelly, T.K. & Jones, P.A. (2010) Epigenetics in cancer. *Carcinogenesis (New York)* **31**:27-36.

Shen, M.M. & Abate-Shen, C. (2010) Molecular genetics of prostate cancer: new prospects for old challenges**.** *Genes Dev* **24:**1967-2000.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., Severinov, K., Zhang, F. & Koonin, E.V. (2017) Diversity and evolution of class 2 CRISPR–Cas systems. *Nat Rev Microbiol* **15**:169-182.

Silverman, L.B., Sallan, S.E., Lander, E.S., Pieters, R., Armstrong, S.A., Staunton, J.E., Golub, T.R., Korsmeyer, S.J., den Boer, M.L. & Minden, M.D. (2002) MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* **30**:41-47.

Sipeky, C., Gao, P., Zhang, Q., Wang, L., Ettala, O., Talala, K.M., Tammela, T.L.J., Auvinen, A., Wiklund, F., Wei, G. & Schleutker, J. (2018) Synergistic Interaction of HOXB13 and CIP2A Predisposes to Aggressive Prostate Cancer. *Clin Cancer Res* **24**:6265.

Sreenath, T., Orosz, A., Fujita, K. & Bieberich, C.J. (1999) Androgen-independent expression of hoxb-13 in the mouse prostate. *The Prostate* **41**:203-207.

Stamey, T.A., Yang, N., Hay, A.R., McNeal, J.E., Freiha, F.S. & Redwine, E. (1987) Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *N Engl J Med* **317**:909-916.

Strachan, T. & Read, A.P. (2019) Human molecular genetics, 5th edition, pp. 366-371 & 597-628. CRC Press, Taylor & Francis Group, Florida.

Storebjerg, T.M., Høyer, S., Kirkegaard, P., Bro, F., Ørntoft, T.F., Borre, M. & Sørensen, K.D. (2016) Prevalence of theHOXB13G84E mutation in Danish men undergoing radical prostatectomy and its correlations with prostate cancer risk and aggressiveness. *BJU Int* **118**: 646-653.

The Finnish Cancer Registry, 2018 report https://syoparekisteri.fi/assets/files/2020/05/Cancer_in_Finland_2018-report.pdf

Van Belkum, A., Scherer, S., van Alphen, L., & Verbrugh, H. (1998) Short-Sequence DNA Repeats in Prokaryotic Genomes. *Microbiol Mol Biol Rev* **62**:275-293.

Vijayakumar, S., Winter, K., Sause, W., Gallagher, M.J., Michalski, J., Roach, M., Porter, A. & Bondy, M. (1998) Prostate-specific antigen levels are higher in African-American than in white patients in a multicenter registration study: results of RTOG 94-12. *Int J Radiat Oncol Biol Phys* **40**:17-25.

Waltregny, D., Alami, Y., Clausse, N., Leval, J.D. & Castronovo, V. (2002) Overexpression of the homeobox geneHOXC8 in human prostate cancer correlates with loss of tumor differentiation. *The Prostate* **50**:162-169.

Weinberg, R.A. (1991) Tumor suppressor genes. *Science* **254**:1138-1146. doi: 10.1126/science.1659741.

Whelan, J.T., Ludwig, D.L. & Bertrand, F.E. (2008) HoxA9 induces insulin-like growth factor-1 receptor expression in B-lineage acute lymphoblastic leukemia. *Leukemia* **22**:1161-1169.

Wu, X., Chen, H., Parker, B., Rubin, E., Zhu, T., Lee, J.S., Argani, P. & Sukumar, S. (2006). HOXB7, a Homeodomain Protein, Is Overexpressed in Breast Cancer and Confers Epithelial-Mesenchymal Transition. *Cancer Res*, **66:** 9527–9534.

Xu, J., Lange, E.M., Lingyi, L., Zheng, S.L., Wang, Z., Thibodeau, S.N., Cannon-Albright, L.A., Teerlink, C.C., Camp, N.J., Johnson, A.M., Zuhlke, K.A., Stanford, J.L., Ostrander, E.A., Wiley, K.E., Isaacs, S.D., Walsh, P.C., Maier, C., Luedeke, M., Vogel, W., Schleutker, J., Wahlfors, T., Tammela, T., Schaid, D., Mcdonnell, S.K., Derycke, M.S., Cancel-Tassin, G., Cussenot, O., Wiklund, F., Grönberg, H., Eeles, R., Easton, D., Kote-Jarai, Z., Whittemore, A.S., Hsieh, C., Giles, G.G., Hopper, J.L., Severi, G., Catalona, W.J., Mandal, D., Ledet, E., Foulkes, W.D., Hamel, N., Mahle, L., Moller, P., Powell, I., Bailey-Wilson, J.E., Carpten, J.D., Seminara, D., Cooney, K.A., Isaacs, W.B.; International Consortium for Prostate Cancer Genetics. (2013) HOXB13 is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG). *Hum Genet* **132**:5-14.

Yamashita, T., Tazawa, S., Yawei, Z., Katayama, H., Kato, Y., Nishiwaki, K., … Ishikawa, M. (2006). Suppression of invasive characteristics by antisense introduction of overexpressed HOX genes in ovarian cancer cells. *Int J Oncol* **28:**931-938.

Yao, J., Chen, Y., Nguyen, D.T., Thompson, Z.J., Eroshkin, A.M., Nerlakanti, N., Patel, A.K., Agarwal, N., Teer, J.K., Dhillon, J., Coppola, D., Zhang, J., Perera, R., Kim, Y. & Mahajan, K. (2019) The Homeobox gene, HOXB13, Regulates a Mitotic Protein-Kinase Interaction Network in Metastatic Prostate Cancers. *Sci Rep* **9**:9715-18.

Zhai, Y., Kuick, R., Nan, B., Ota, I., Weiss, S.J., Trimble, C.L., Fearon, E.R. & Cho, K.R. (2007) Gene Expression Analysis of Preinvasive and Invasive Cervical Squamous Cell Carcinomas Identifies HOXC10 as a Key Mediator of Invasion. *Cancer Res* **67**:10163-10172.