The impact of CRISPR/Cas9 generated mutations in pioneer transcription factors FOXA1 and GATA3 on oestrogen receptor function

Miroslav Bobrik

Master's Degree Project, 60cr.

29th August 2018

International Agency for Research on Cancer

World Health Organization

LUND UNIVERSITY

Faculty of science

Department of biology

Molecular biology, specialization: Molecular Genetics and Biotechnology

Abstract

Breast cancer is the most common cancer among women and two thirds of all breast cancers are classified as oestrogen receptor positive.

ER is a transcription factor that regulates gene expression and promote cell proliferation that is critical in mammary gland development, and its aberrant activity is commonly present in breast cancer.

As a nuclear receptor, ER activity could be diverted by either its aberrant DNA-binding that enables a direct regulation of gene expression and/or its aberrant response to an activating ligand, oestrogen. The ability of ER transcription factors to bind DNA and activate gene transcription is deregulated in most breast cancers in the absence of mutations in this transcription factor, where ER is no longer under usual control mechanisms.

Because ERs drive most breast cancers in the absence of mutations in its coding region, we hypothesise that the ER function may be diverted in breast cancer through epigenetic mechanisms, thus aberrant ER activities becomes a key driving force in inducing uncontrolled cell proliferation that lead to development and progression of oestrogen-induced breast cancers. We also propose that lowering the expression of transcription factors FOXA1 and GATA3 may deregulate ER function through epigenetic mechanisms and revert methylation patterns. in ER binding sites, from hypo- to hypermethylated.

Key words: Breast cancer; epigenetics; methylation; CRISPR/Cas9; transcription factors; FOXA1; GATA3; MCF-7

Abbreviations

BrCa Breast cancer

ER Oestrogen receptor

FOXA1 Forkhead box protein A1

DMEM Dulbecco's Modified Eagle's Medium

MEM Minimal Essential Medium

TF Transcription factor

CRISPR Clustered regularly interspaced short palindromic repeats

Cas9 CRISPR associated protein 9

TALENs Transcription activator like effector nucleases

ZFNs Zinc-finger nucleases

crRNA CRISPR RNA

tracrRNA trans-activating crRNA

TCGA The Cancer Genome Atlas

Contents

1 Introduction 4
1.1 Breast cancer 4
1.2 Epigenetics of breast cancer 5
1.3 Role of Estrogen Receptors and its transcription factors in breast cancer 7
1.4 Genome editing 7
1.5 Aim of the project 8
2 Materials and Methods 9
2.1 gRNA construction 9
2.2 Cloning in *E. coli*. 9
2.3 Pyrosequencing 9
2.4 Plasmid purification 9
2.5 Cell culture 10
2.6 Transfections 10
2.7 Fluorescent microscopy 10
2.8 RNA Isolation and RT-qPCR 10
2.9 Flow Cytometry 11
2.10 FACS sorting 11
2.11 Western blot assays 11
2.12 RNAi 11
2.13 Methylation analysis 12
3 Results 13
3.1 Expression of FOXA1 and GATA3 13
3.2 CRISPR/Cas9 using transient plasmid expressing gRNA 13
3.3 Silencing of *FOXA1* 14
3.4 Expression of Cas9 in Stable cell lines 15
3.5 CRISPR/Cas9 using synthetic gRNA molecules 16
4 Discussion 18
4.1 Cell growth 18
4.2 Transcription factors of ER 18
4.3 Cas9 expression goes away in passaged subcultures 18
4.4 Pooled gRNA knock-out 19
4.5 Predicting the result and future prospective 20
References 21
Appendix 24

1 Introduction

1.1 Breast cancer

Despite the availability of new chemotherapeutic agents and progress in prevention and surveillance, breast cancer is still the most common cancer affecting women mostly in developed countries (Figure 1). Furthermore, the incidence of breast cancer has significantly increased in recent decades, a phenomenon partly attributed to the modern lifestyle (Ferlay et al. 2015).

Among established risk factors, the steroid hormones, notably oestrogens, have been recognized as the key players in majority of breast cancer cases and current chemo preventive strategies target hormonally responsive breast tumours (Williams & Lin 2013).

Mechanistic studies have revealed that the hormone oestrogen activates the oestrogen receptors, ligand-activated nuclear receptor family of transcription factors, to mediate cellular functions that are essential to the normal development and maintenance of tissue homeostasis (Ropero, Alonsomagdalena, Quesada & Nadal 2008). Because the unique transcriptional response to oestrogen hormone is partially mediated by the epigenome state, inappropriate activation of the oestrogen-mediated gene expression can deregulate epigenetic mechanisms and initiate neoplastic transformation in mammary epithelial cells (Hervouet, Cartron, Jouvenot & Delage-Mourroux 2013).

In addition, it is possible that breast cancer patients that commonly receive anti-oestrogen therapy may develop hormonal therapy resistance through epigenetic mechanisms (Abdel-Hafiz 2017).

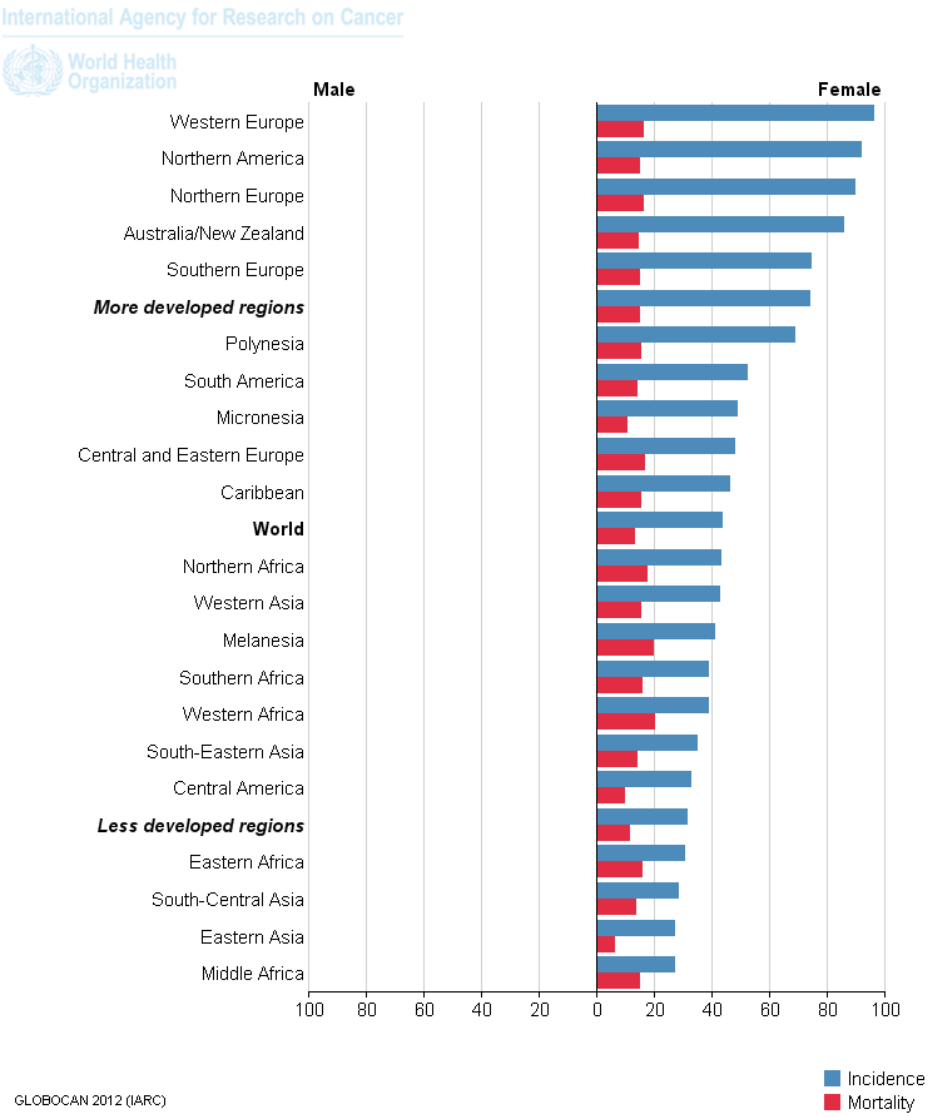


Figure 1: *The world-wide incidence and mortality of breast cancer by sex and geographical region. Grade shows cases per 100 000 people.*

1.2 Epigenetics of breast cancer

Epigenetic mechanisms are essential for development and cell differentiation. They seem to be critical for the integration of endogenous and environmental signals during the life of a cell or an organism (Herceg 2007; Feinberg, Koldobskiy & Göndör 2016).

Deregulation of epigenetic mechanisms has been associated with a variety of human malignancies, including breast cancer (Jovanovic, Rønneberg, Tost & Kristensen 2010). Technological advances in epigenomics have created the opportunity to comprehensively characterize the epigenetic landscape of breast tumours (Umer & Herceg 2013), and recent DNA methylome profiling of a large series of breast tumours suggested the existence of previously unrecognized breast cancer types beyond the currently known expression subtypes (Dedeurwaerder et al. 2011; Koboldt et al. 2012). This and other studies of breast cancer

samples demonstrated that DNA methylation changes are omnipresent in primary breast cancer cells and that the phenotypic diversity of breast cancers may reflect intra-tumour epigenetic heterogeneity.

The list of DNA methylation changes in breast cancer tissues has further increased with the completion of major international sequencing initiatives, however the underlying mechanisms and functional importance of DNA methylation changes in breast cancer have not been systematically evaluated (Mathot et al. 2017).

Although there is currently limited evidence linking epigenetic changes induced by different risk factors to breast cancer, the precise targets of epigenetic deregulation are unknown (Jovanovic et al. 2010; Terry, Delgado-Cruzata, Vin-Raviv, Wu & Santella 2011). DNA methylation has been widely considered as a regulator of gene expression. However, methylation of cytosines has also been implicated in the transcriptional silencing of transposons, imprinted genes, and genes on the inactive X chromosome (Bird 2002). Because DNA methylation has the potential to regulate gene expression program through its impact on the activity of most, if not all, transcription factors, this layer of epigenetic modifications is likely to be implicated in normal ER activity and aberrant events implicating ER pathway.

Despite the fact, that the predominant consequence of methylation is transcriptional repression, being well established, it is less clear if this is mediated directly or indirectly (Sawan, Vaissière, Murr & Herceg 2008). Direct inhibition of transcription may occur through blocking the binding of transcription factors to promoters containing methylated CpG sites (Iguchi-Ariga & Schaffner 1989), while indirect repression may involve proteins such as MePC2 that specifically bind to methylated DNA via a methyl-CpG-binding domain (MBD) (Jones et al. 1998; Nan et al. 1998; Bird 2002; Mutskov 2002). However, more recent evidence implies that DNA methylation may also be a secondary event in transcriptional silencing and that DNA methyltransferases (enzymes that mediate DNA methylation) may be triggered by histone modifications, which may serve as the primary epigenetic mark (Moore, Le & Fan 2013). In this regard, it is intriguing that some transcription factors have been shown to bind methylated specific regions in the genome and trigger their demethylation (Stadler et al. 2011). Furthermore, it has been suggested that binding and activity of some transcription factors that are sensitive to DNA methylation relies on additional determinants to induce local hypomethylation (Domcke et al. 2015). Among these additional factors is a subset of transcription factors that possess the remarkable ability to activate their target genes in heterochromatin, that is usually refractory to transcription factors, thus, behaving as pioneer transcription factors to initiate local demethylation in such chromatin configuration (Schübeler 2015).

The pioneer transcription factors appear to have the special ability to engage developmentally silenced genes (that are typically embedded in heterochromatin in differentiated cells), consistent with their highest reprogramming activity. Since pioneer transcription factors play key role in establishment and maintenance of gene expression, their deregulation could trigger aberrant cell reprogramming and human disease, including cancer (Iwafuchi-Doi & Zaret 2014).

1.3 Role of Estrogen Receptors and its transcription factors in breast cancer

Transcription factors FOXA1 and GATA3 are essential for ER binding and regulating expression of target genes (Albergaria et al. 2009, p. 3; Hurtado, Holmes, Ross-Innes, Schmidt & Carroll 2011, p. 1). This makes them interesting candidates to observe how deregulation of ER in ER+ cancer changes genome wide methylation patterns. Unlike *FOXA1*, *GATA3* is one of three genes with somatic mutations in more than 10% of all breast cancers (Koboldt et al. 2012).

ER, together with these two transcription factors, form the defining signature genes observed in ER+ breast cancers (Jozwik & Carroll, 2012) and all three proteins have been shown to be required for the function of the oestrogen responsive transcription complex (Hurtado, Holmes et al., 2011, Kong, Li et al., 2011).

In consistence with their role in estrogen responsive cellular processes, such as differentiation and tumorigenesis, deregulation of these factors in mice models results in mammary gland tumours (Kouros-Mehr, Slorach et al., 2006, Theodorou, Stark et al., 2013). These findings provide an important insight into the function of ER pathway, where FOXA1 and GATA3 may mediate cellular processes (such as proliferation and differentiation) in normal mammary gland independent of ER, but become an essential component within the ER complex during cancer development and progression (Carroll, 2016).

The ER genes are rarely mutated in primary breast tumours (Ciriello, Gatza et al., 2015, Ellis, Ding et al., 2012), suggesting that ER function in breast tumours may be deregulated through non-mutational processes. This notion is further supported by the cancer genome sequencing that found *FOXA1* mutated in a small subset of breast cancers (1.8%) while only a fraction of those breast cancers exhibit amplification of the genomic region of the *FOXA1* locus. In addition, there is evidence of mutations in the genome that are binding sites of FOXA1, which may impact indirectly FOXA1-mediated regulation of ER pathways. Interestingly, GATA3, another ER interacting protein, was found to be mutated in a marked subset of breast cancers (>10% of breast tumours in TCGA study), and mutations in *GATA3* and *FOXA1* seem to be mutually exclusive (Cancer Genome Atlas Network, 2012, Robinson, Holmes et al., 2013). The impact of deregulation of FOXA1 and GATA3, through mutational or non-mutational processes, on the function of these proteins and whether these alterations affect ER transcriptional activity is currently unknown.

1.4 Genome editing

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated protein 9 (Cas9) system has become popular in recent years due to many benefits it offers compared to zinc-finger nucleases (ZFNs) and transcription activator like effector nucleases (TALENs) genome editing approaches. Overall CRISPR/Cas9 is an easier to design and more flexible technique and can be used from inducing simple indel mutations to being coupled with other enzymes and direct their activity(Enríquez 2016).

1.5 Aim of the project

The scope of this master's thesis is to validate the hypothesis that TFs FOXA1 and GATA3 are altering the transcription pattern in ER+ breast cancers through altering the methylation patterns of these cancers. The aims of this project can be summarised in following steps:

1. Design and optimize CRISPR/Cas9 strategy to knock-out *FOXA1* and *GATA3* genes in ER+ breast cancer cell line and a normal epithelial cell line from breast
2. Sequence the genome-wide methylation patterns of knocked-out cells and parental population of cell to see if methylation changes are mostly correlated with ER related genes and regions

2 Materials and Methods

2.1 gRNA construction

Six gRNAs were designed (3 for *FOXA1* and 3 for *GATA3* genes) and ordered as simple DNA oligos alongside with their reverse complements. Every oligo contained adapters compatible with cleavage sites of two enzymes (BamH1 and BsmB1) used for cleaving out the spacer DNA of the backbone plasmid pGuide-EF1a-GFP by OriGene™ (Supplementary figure 1). At first, 6 oligos were annealed with their reverse complements. Then, the plasmid was cleaved with two enzymes one after another as the digestion condition varied. The order was chosen based on salt concentration required by the two enzymes (low salt conc. first and high salt conc. second). In six separate reactions the annealed inserts were ligated into the backbone, using T4 ligase by New England Biolabs.

2.2 Cloning in *E. coli*.

The constructed plasmids were transformed into DH5α *E. coli* strain using heat shock protocol. The cells were plated on LB plates with 100μg/mL ampicillin and incubated overnight at 37°C. Clones were selected by ampicillin resistance which was induced by the presence of the plasmid.

2.3 Pyrosequencing

This method was used to verify the presence of desired gRNA inserts after colony PCR with ampicillin resistant colonies.

Firstly, a larger fragment of pGuide plasmid has been amplified using PCR with following primers; pGuide_insert-F: GGCTGTTAGAGAGATAATTAG and pGuide_insert-R: ACTTGCTATTTCTAGCTCTAA, the reverse primer was also biotinylated for the amplicon to be able to bind to the beads used for sequencing. The Sequencing primer was designed to bind just six bases upstream from the potential gRNA sequence and had the following structure; TATCTTGTGGAAAGGACGCG. See pyrograms (Supplementary figures 4 and 5).

2.4 Plasmid purification

Purification on a column was performed using QIAprep Plasmid purification kit (Mini/Midi scale) and concentrations measured on Nanodrop. Plasmids were then cleared of endotoxins using Pierce™ High Capacity Endotoxin Removal Resin.

2.5 Cell culture

MCF-10A/Cas9 cell line was grown in DMEM/F:12 medium supplemented with 5% Fetal Bovine Serum (FBS), and MCF-7/Cas9 cell line was grown in standard DMEM supplemented with 10% FBS, Pen Strep and Insulin (400 mL DMEM, 40 mL FBS, 4 mL P/S, 400 µL Insulin) in a humidified incubator at 37 °C and 5% CO^2. Cells were transfected with the desired constructs at 70-80% confluency using Lipofectamine 3000 reagent (TS #L3000015).

2.6 Transfections

The plasmids were extracted form bacteria and purified as seen in chapter 2.4 and transfected into MCF-10A and MCF-7 cells using Lipofectamine 2000. The transfection efficiency was extremely low, even after attempts to optimise the transfection conditions by using different amounts of DNA and/or Lipofectamine 2000 reagent. Thus, latter transfections were conducted using Lipofectamine 3000 which has increased efficiency slightly.

Genome editing using TrueGuide Synthetic gRNAs. Six TrueGuide™ synthetic crRNA molecules were designed and annealed with scaffold tracrRNA. These RNA duplexes are ready to transfect, highly resistant against nucleases, gRNA molecules. The optimisation of the reagent amounts used has been performed on a 6-well plate format with different Lipofectamine transfection reagents and DNA amounts. Following the manufacturers recommendations for transfection.

2.7 Fluorescent microscopy

Transfection efficiency was estimated using Nikon eclipse Ti inverted microscope. Magnification used was 400X and exposure time from 5ms to 3s. Excitation light used was 475nm (FITC filter set).

2.8 RNA Isolation and RT-qPCR

Total cellular RNA was isolated using AllPrep DNA/RNA Mini Kit (50) from dry pellets stored in -80°C. Reverse transcription reaction was performed for cDNA conversion, using RevertAid reverse transcription kit (Thermo scientific #K1621). PCR reactions were carried out with 75 ng of cDNA as template, using PowerUp SYBR™ Green Master Mix (Thermo Scientific #A25742) on Real-Time PCR Detection System (CFX96 Touch™, Bio-Rad #1855195).

The calculations of the relative expression of *FOXA1* and *GATA3* genes were carried out as follows; the cycle threshold (Ct) mean was calculated from technical triplicates of each biological triplicate. Then, for each biological triplicate a difference between Ct of the *FOXA1* or *GATA3* compared to the *GAPDH* gene was calculated. By using the delta value as a negative exponent of two and then averaging these values from the three biological triplicates a value

was obtained and plotted in the chart below (Figure 2). The standard deviations were calculated from the results of two to the power of negative delta Ct by using STDEV function in Excel.

2.9 Flow Cytometry

Cells were harvested, washed with PBS and resuspended in 250µL PBS prior to measurement. Negative control (same cell line which was not transfected with pGuide plasmid) has been used to set up the measurement parameters. MCF-10A and MCF-7 transfected with pGuide plasmid (with different inserts) have been measured and fluorescence levels compared to negative control.

2.10 FACS sorting

Cells were detached using trypsin (0.05% for MCF-7 and 0.5% for MCF-10A) and resuspended in 500µL of complete DMEM medium with additional FBS. Sorting was carried out with settings shown in the Supplementary figure 3, on Bio-Rad S3e cell sorter.

2.11 Western blot assays

Total protein was isolated from freshly harvested cells on ice and quantified on Eppendorf Biophotometer Spectrophotometer UV/Vis Reader. 50µg of total protein were mixed with 6x loading buffer (containing mercaptoethanol) and denatured at 99°C for 5 minutes. Denatured samples were loaded onto ready to use polyacrylamide gel and protein fractions were left to separate as much as possible. The proteins were transferred onto Bio-Rad PVDF Mini format membrane using Bio-Rad Trans-Blot Turbo transfer system. Prepared membrane was blocked with 4% milk overnight in 4°C or for an hour in RT, after stripping the membrane.

Depending on the antibody (Ab) different dilutions were used. Anti-GATA3 monoclonal, mouse by Santa-Cruz Biotech was diluted 1:100 in 4% Milk (50uL in 5mL milk). Anti-Cas9 monoclonal, mouse by Abcam 1:200, anti-FOXA1 polyclonal, rabbit by Santa-Cruz Biotech 1:1000 and anti-β-Actin monoclonal, mouse by MP Biomedical was diluted 1:10 000). Incubation with primary Ab was overnight in 4°C or for two hours in RT. The secondary Ab were always diluted 10-times less than the primary Ab has been and incubated in the same way as primary Ab. The protein bands were visualised with Bio-Rad ChemiDoc™ Imager.

2.12 RNAi

Pre-designed small interacting RNA molecules by ThermoFisher have been transfected into MCF-7 cell line using RNAiMAX Lipofectamine. After 72 hours incubation media were changed to remove the residual Lipofectamine reagent. On 24-well plate groups of triplicates were collected at different time points and stored in -80°C until RNA extraction.

Two negative control groups were used one collected 24 hours after the end of incubation (together with first silenced group) and one was collected with the last silenced group 168 hours post incubation.

To calculate the relative silencing efficiency the cycle threshold (Ct) mean was calculated from technical triplicates of each biological triplicate. Then, for each biological triplicate a difference between Ct of the FOXA1 or GATA3 compared to a house-keeping gene was calculated. By using the delta value as a negative exponent of two and then averaging these values from the three biological triplicates a value was obtained and plotted in the chart below (Figure 3). The standard deviations were calculated from the results of two to the power of negative delta Ct by using STDEV function in Excel.

2.13 Methylation analysis

Cells detached using trypsin and frozen as dry pellet in -80°C have been used for later DNA extraction. QIAamp DNA Mini Kit have been used, following included protocol and eluted DNA was measured on Nanodrop. The samples with enough DNA were quantified on Qubit and diluted to 13.33ng/µL concentration in 50µL sample volume. 400ng DNA from each sample will be used for bisulphite conversion prior to MethylationEPIC BeadChip (Infinium) microarray (850K) analysis, which covers over 850 000 CpG methylation sites.

3 Results

3.1 Expression of FOXA1 and GATA3

Expression levels, of *FOXA1* and *GATA3*, have been found to be ten times higher in MCF-7 compared to the negative control cell line MCF-10A shown in a RT-qPCR assay results (Figure 2). It has been also observed that MCF-7 have more difficulties surviving after silencing or knocking out these two genes.

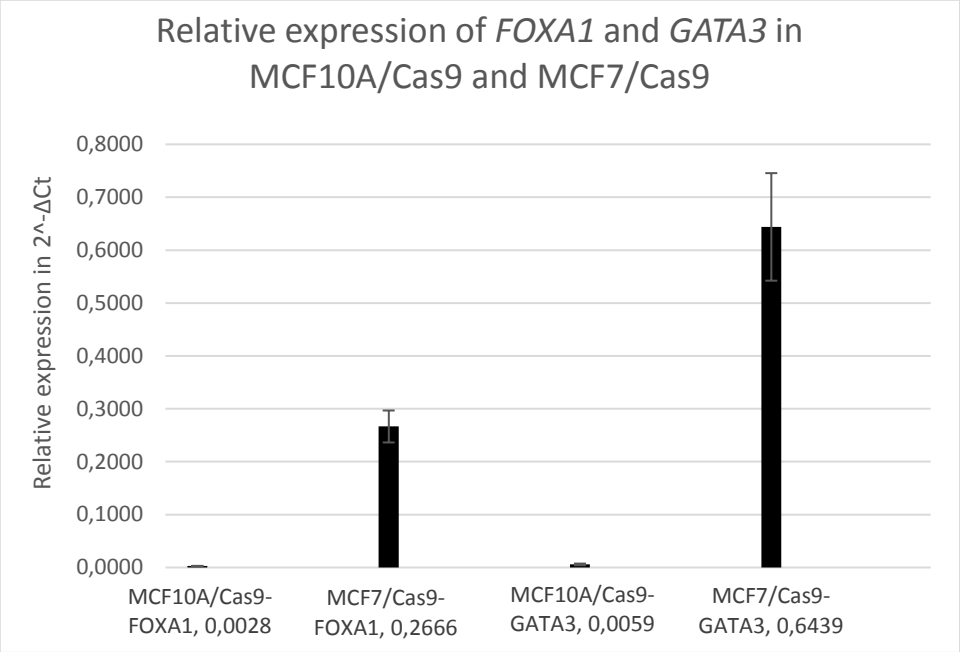


Figure 2: *Relative expression levels of target genes FOXA1 and GATA3 in cell lines used in this project (MCF-7 and MCF-10A). The samples were prepared for RT-qPCR run in three replicates and negative control for each sample.*

3.2 Introducing transient plasmid expressing gRNA into cell lines

MCF-7 and MCF-10A cells transfected by plasmids expressed GFP (Figures 3) which's gene is present in the plasmid (Supplementary figure 1). This selection marker was used to sort and isolate single cells and/or pools of transfected cells.

After many attempts of transfecting cells with pGuide plasmids and sorting out GFP positive cells it was repeatedly observed these cells struggle to survive and grow.

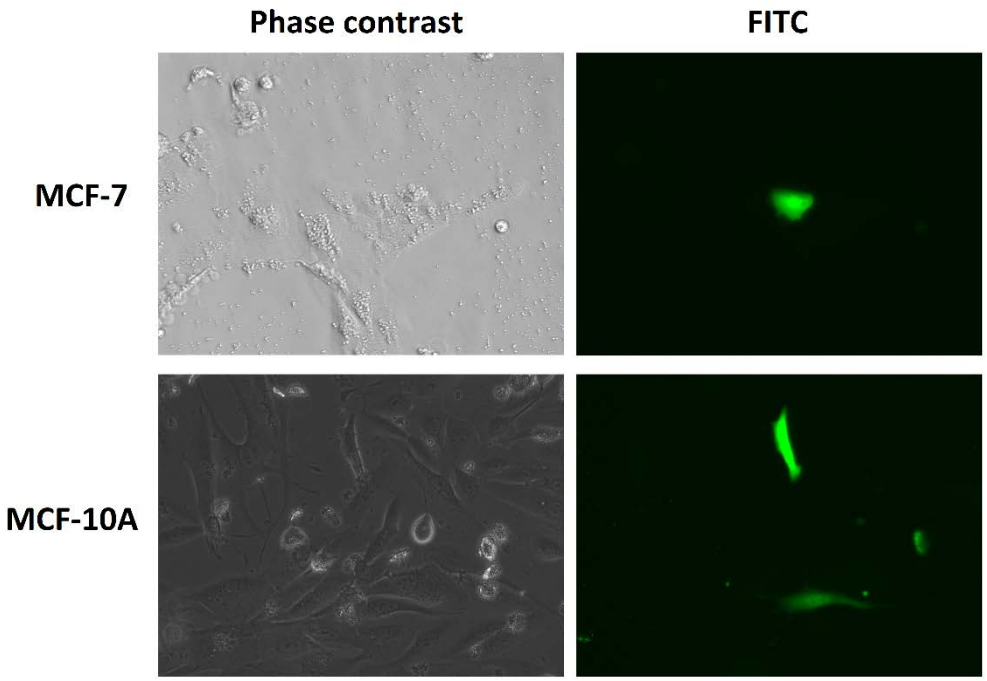


Figure 3: *Fluorescent MCF-10A (above) and MCF-7 (below) after transfections. Only two (MCF-10A) and one (MCF-7) fluorescent cells per approximately twenty cells in each group. The fluorescent cells are carrying the plasmid used in transfection and needed to be isolated from the non-transfected cells. In MCF-10A two very distinct intensities of fluorescence were observed on both a microscope and the FACS.*

3.3 Silencing of *FOXA1*

Silencing using siRNA was tried out as an alternative approach of removing FOXA1 protein from the MCF-7 cells with assumption the FOXA1 will be not present long enough for methylation patterns to change. On 24-well plate cells were plated and split into groups to be analysed in triplicates after different periods post incubation (Figure 4).

Our observations show that silencing is efficient to at least 50% and the effects are present at least 5 days after the three days of incubation (Figure 4).

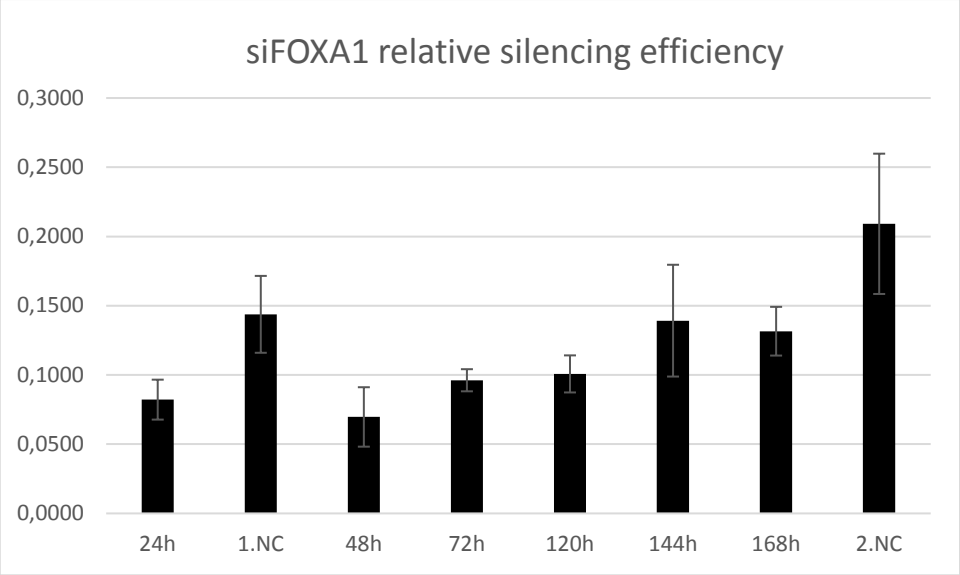


Figure 4: *Testing siRNA targeting FOXA1 gene to estimate the time of highest silencing potential by amount of FOXA1 mRNA present in the cells. The times represent time after 72h incubation. Negative controls (marked 1.- and 2.NC) were collected at different time points as well. 1.NC was collected with first group of silenced cells and 2.NC with the last group. This was done to be able to see how much can FOXA1 level fluctuate without siRNA affecting it.*

3.4 Expression of Cas9 in Stable cell lines

The Cas9 levels in stable cell lines were found to be very high in MCF-7 but absent in MCF-10A (Figure 5). In both cell lines expression disappears in around ten passages (Figure 5 lane Cas9 MB).

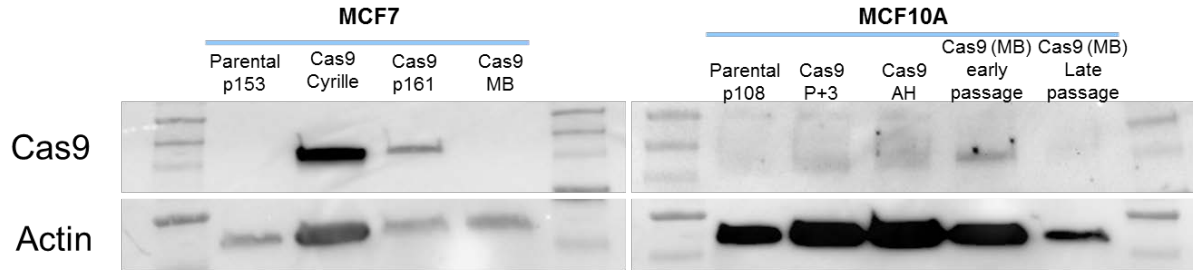


Figure 5: *Western blots showing levels of Cas9 in the total protein isolated from the MCF-7 and MCF-10A cells. The faint and blurry bands in MCF-10A samples suggest degradation of Cas9 which is completely gone in less than ten passages.*

3.5 CRISPR/Cas9 using synthetic gRNA molecules

Western blot of MCF-7 cells transfected with TrueGuide synthetic RNA molecules showed decrease in target protein in the transfected pool of cells. Especially when gRNA molecules targeting more upstream or when more than one target site has been cleaved (Figures 6 and 7).

The more upstream the gRNA targeted the sequence the more dramatic is the result (Figure 6). In 'F1-3' and in 'Parental' lanes we observe a band (approximately 72kDa) larger than expected FOXA1 size which is around 50kDa.

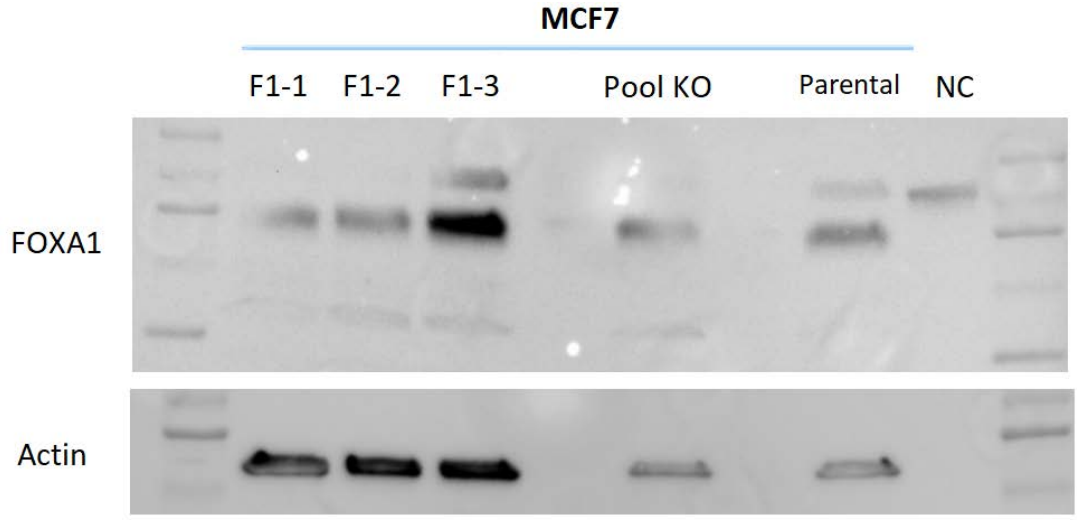


Figure 6: *Western blots showing levels of FOXA1 in relation to β-Actin after transfection with TrueGuide gRNAs. Lanes one, two and three show FOXA1 protein in single knock-out populations of MCF-7 cells, targeting sites are in order starting from the closest to the beginning of transcription site in exon one of FOXA1 gene. In lane five the protein fraction from cells knocked-out with cocktail of all six gRNAs is shown. The last two lanes, seven and eight, show controls. First, parental control which is the MCF-7 cells treated as knock-out populations, but no gRNA was used. Second is a protein fraction from MCF-10A cells which do not express FOXA1.*

The MCF-7 cells with edited *GATA3* in the start and middle of the sequence and the cells transfected with all six gRNA molecules targeting *FOXA1* and *GATA3* gene at three different loci each, have much lower levels of the target protein compared to cells in which the gene was cleaved further on in its reading frame (Figure 7).

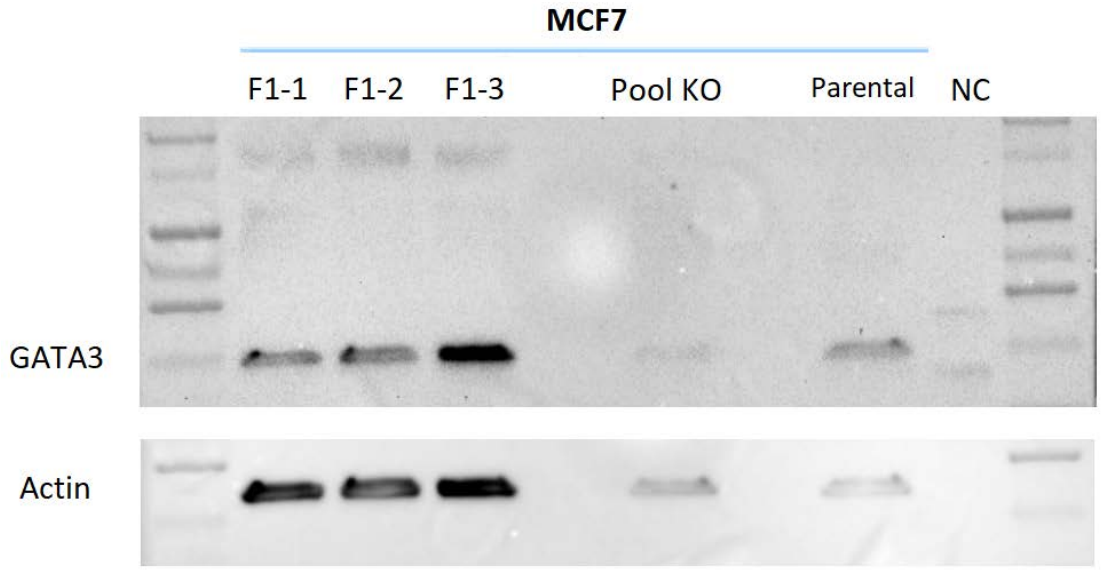


Figure 7: *Western blots showing level*s of GATA3 in relation to *β*-Actin after treatment with TrueGuide *gRNAs. Lanes one, two and three show GATA3 protein in single knock-out populations of MCF-7 cells, targeting sites are in order starting from the closest to the beginning of transcription site in exon one of GATA3 gene. In lane five the protein fraction from cells knocked-out with cocktail of all six gRNAs is shown. The last two lanes, seven and eight, show controls. First, parental control which is the MCF-7 cells treated as knock-out populations, but no gRNA was used. Second is a protein fraction from MCF-10A cells which do not express GATA3.*

4 Discussion

This project was focused on optimising CRISPR/Cas9 strategy, to be used in breast cancer cell lines. Main goal was to induce knock-out in two transcription factors and measure changes in global methylation of the genome. Along the timeline of this project several technical limitations have appeared that were not accounted for from the start. This fact forced us to reconsider the approach and come up with quicker and easier way to achieve our goal.

4.1 Cell growth

The retarded growth and increased mortality of cells with silenced or knocked-out FOXA1 and/or GATA3 suggests these two transcription factors are important for the homeostasis and cell proliferation in ER+ cancer tumours. The mortality of transfected cells, observed repeatedly after transfection with pGuide plasmid or siRNA, suggests the importance of FOXA1 and GATA3 for survival of MFC-7 cancer cells.

While culturing MCF-7 cells it has sometimes been observed that cells had issues attaching and surviving after being passed without an obvious cause. This could have been caused by overly stressful factor affecting the cells, such as longer trypsinisation or temperature shock maybe, but neither of this was ever noticed in correlation to the MCF-7 cells having issues attaching themselves.

4.2 Transcription factors of ER

In cancer versus normal epithelial cell lines we see 10-fold difference in FOXA1 and GATA3 expression. This has been previously shown, as the ER+ breast cancers are dependent on ER activity which is in turn coupled with other transcription factors (Albergaria et al. 2009). Thus, these interacting transcription factors in MCF-7 breast carcinoma cells are overexpressed together with ER. This observation is favouring the hypothesis that after removal of FOXA1 and GATA3 proteins from the cells the methylation patterns could be reverted to more original state, similar to healthy epithelial cells from breast. This does not mean the progression of ER+ cancers would be necessarily stopped as about 30% of breast cancers are driven by other mechanisms than ER (Chang 2012).

4.3 Cas9 expression goes away in passaged subcultures

The Cas9 levels in generated stable cell lines were found to be very high in MCF-7 but absent in MCF-10A (Figure 4). This could have happened for various reasons. For example a resistance to selection agent G418 (geneticin) is a common occurrence *in vitro* (Paludan, Duch, Jørgensen, Kjeldgaard & Pedersen 1989). Also, the smearing on the membrane where Cas9 should be suggests degradation in MCF-10A cells (Figure 5). Due to this complication the

negative control cell line MCF-10A could not be used for the new approach of CRISPR/Cas9 editing with synthetic crRNA molecules.

4.4 Pooled gRNA knock-out

Due to reoccurring failed attempts with cell sorting a more efficient and faster approach was chosen to knock-out the *FOXA1* and *GATA3* genes in the end of this project. The advantages of using synthetic gRNA molecules are the addition of 2'O-methyl analogues and 3' phosphorothioate internucleotide bonds (Supplementary figure 2). These modifications enhance editing efficiency by increasing binding to the target site and inhibiting nuclease degradation.

In knock-out experiment, where gRNA molecules used for CRISPR/Cas9 used were synthetic oligos ('no-plasmid' approach), we observed a better efficacy of knocking out the protein when several gRNA molecules, targeting the same gene at different sites, were used. This might be due to a fragment of gene being cleaved out rather than few unspecific indels being created at one site in the gene (Zhang, Tee, Wang, Huang & Yang 2015). When using more than one gRNA molecule in the same knock-out experiment the chances of successfully disabling the target are higher. This could be due to the fact that if the cleavage happens in about the same time in two different locations within the same gene it is likely that whole intermediate (cleaved out) fragment will be deleted rather than random indel mutation being induces at only one cleavage site when single gRNA is used (Zhang et al. 2015).

When using single gRNA molecules for transfection only the first two out of three seem to be lowering the gene amount in population of cells by approximately 50%. The other molecules are targeting the gene region more downstream from the transcription start site (TSS) and the induced mutation might not be dramatic enough to disable the protein. The cleavage sites are dispersed, but all are in the TSS proximity, except maybe the third cleavage site for both genes. This could explain the lower efficacy of knock-down of target proteins in the pool of cells when only the gRNA binging to the third cleavage site is used (Figure 8).

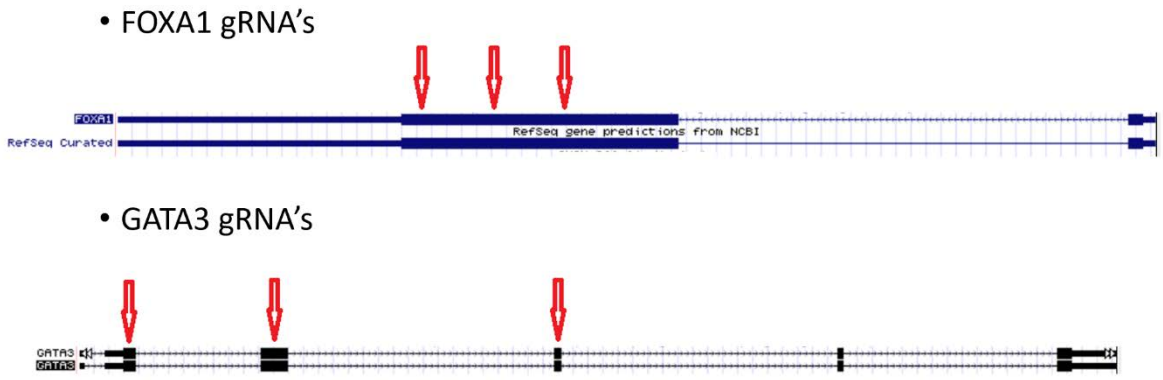


Figure 8: *Scheme showing the two target genes and homology regions to the gRNA molecules. The thick regions show exons. Two exons in FOXA1 and five exons in GATA3.*

After collecting total protein fraction from knock-out populations and conducting a western blot assay to visualise an approximate difference in FOXA1 and GATA3 proteins amounts in the control cells and the transfected groups we have observed a decrease in in the target proteins in both repeats of the western blot. However, this was just a partial decrease in some groups. Still, that could be enough to detect methylation changes in 850K methylome analysis (Moran, Arribas & Esteller 2016).

4.5 Predicting the result and future prospective

This project is not far from over, but the result can be only predicted based on other similar studies (Fleischer et al. 2017). Assuming that in the population of cells which either had or had not the FOXA1 and GATA3 knocked-out the methylation changes will be detected by 850k analysis I would expect the methylation to change form hypo- to hypermethylated in the knocked-out populations compared to parental MCF-7 that express FOXA1 and GATA3 fully. This result would confirm that FOXA1 and GATA3 are tightly connect to methylation patterns and can influence each other.

The technical restrictions, such as unavailability of FACS in the city of Lyon and me not receiving authorisation to operate MethylationEPIC BeadChip 'Infinium' microarray, did not allow the results to be available yet. However, I am confident my contribution will prove useful to the new investigator who will take over this project in the future.

References

Abdel-Hafiz, H. A. (2017). Epigenetic Mechanisms of Tamoxifen Resistance in Luminal Breast Cancer. *Diseases*, 5(3), doi:10.3390/diseases5030016.

Albergaria, A., Paredes, J., Sousa, B., Milanezi, F., Carneiro, V., Bastos, J., Costa, S., Vieira, D., Lopes, N., Lam, E. W., Lunet, N. & Schmitt, F. (2009). Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours. *Breast Cancer Research*, 11(3), doi:10.1186/bcr2327.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & Development*, 16(1), pp. 6–21, doi:10.1101/gad.947102.

Chang, M. (2012). Tamoxifen Resistance in Breast Cancer. *Biomolecules & Therapeutics*, 20(3), pp. 256–267, doi:10.4062/biomolther.2012.20.3.256.

Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C. & Fuks, F. (2011). Evaluation of the Infinium Methylation 450K technology. *Epigenomics*, 3(6), pp. 771–784, doi:10.2217/epi.11.105.

Domcke, S., Bardet, A. F., Adrian Ginno, P., Hartl, D., Burger, L. & Schübeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature*, 528(7583), pp. 575–579, doi:10.1038/nature16462.

Enríquez, P. (2016). Focus: Epigenetics: CRISPR-Mediated Epigenome Editing. *The Yale journal of biology and medicine*, 89(4), p. 471.

Feinberg, A. P., Koldobskiy, M. A. & Göndör, A. (2016). Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nature Reviews. Genetics*, 17(5), pp. 284–299, doi:10.1038/nrg.2016.13.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D. & Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*, 136(5), pp. E359-386, doi:10.1002/ijc.29210.

Fleischer, T., Tekpli, X., Mathelier, A., Wang, S., Nebdal, D., Dhakal, H. P., Sahlberg, K. K., Schlichting, E., Børresen-Dale, A.-L., Borgen, E., Naume, B., Eskeland, R., Frigessi, A., Tost, J., Hurtado, A. & Kristensen, V. N. (2017). DNA methylation at enhancers identifies distinct breast cancer lineages. *Nature Communications*, 8, doi:10.1038/s41467-017-00510-x.

Herceg, Z. (2007). Epigenetics and cancer: towards an evaluation of the impact of environmental and dietary factors. *Mutagenesis*, 22(2), pp. 91–103, doi:10.1093/mutage/gel068.

Hervouet, E., Cartron, P.-F., Jouvenot, M. & Delage-Mourroux, R. (2013). Epigenetic regulation of estrogen signaling in breast cancer. *Epigenetics*, 8(3), pp. 237–245, doi:10.4161/epi.23790.

Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nature Genetics*, 43(1), pp. 27–33, doi:10.1038/ng.730.

Iguchi-Ariga, S. M. & Schaffner, W. (1989). CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes & Development*, 3(5), pp. 612–619, doi:10.1101/gad.3.5.612.

Iwafuchi-Doi, M. & Zaret, K. S. (2014). Pioneer transcription factors in cell reprogramming. *Genes & Development*, 28(24), pp. 2679–2692, doi:10.1101/gad.253443.114.

Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J. & Wolffe, A. P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genetics*, 19(2), pp. 187–191, doi:10.1038/561.

Jovanovic, J., Rønneberg, J. A., Tost, J. & Kristensen, V. (2010). The epigenetics of breast cancer. *Molecular Oncology*, 4(3), pp. 242–254, doi:10.1016/j.molonc.2010.04.002.

Koboldt, D. C., Fulton, R. S., McLellan, M. D., Schmidt, H., Kalicki-Veizer, J., McMichael, J. F., Fulton, L. L., Dooling, D. J., Ding, L., Mardis, E. R., Wilson, R. K., Ally, A., Balasundaram, M., Butterfield, Y. S. N., Carlsen, R., Carter, C., Chu, A., Chuah, E., Chun, H.-J. E., Coope, R. J. N., Dhalla, N., Guin, R., Hirst, C., Hirst, M., Holt, R. A., Lee, D., Li, H. I., Mayo, M., Moore, R. A., Mungall, A. J., Pleasance, E., Gordon Robertson, A., Schein, J. E., Shafiei, A., Sipahimalani, P., Slobodan, J. R., Stoll, D., Tam, A., Thiessen, N., Varhol, R. J., Wye, N., Zeng, T., Zhao, Y., Birol, I., Jones, S. J. M., Marra, M. A., Cherniack, A. D., Saksena, G., Onofrio, R. C., Pho, N. H., Carter, S. L., Schumacher, S. E., Tabak, B., Hernandez, B., Gentry, J., Nguyen, H., Crenshaw, A., Ardlie, K., Beroukhim, R., Winckler, W., Getz, G., Gabriel, S. B., Meyerson, M., Chin, L., Park,

P. J., Kucherlapati, R., Hoadley, K. A., Todd Auman, J., Fan, C., Turman, Y. J., Shi, Y., Li, L., Topal, M. D., He, X., Chao, H.-H., Prat, A., Silva, G. O., Iglesia, M. D., Zhao, W., Usary, J., Berg, J. S., Adams, M., Booker, J., Wu, J., Gulabani, A., Bodenheimer, T., Hoyle, A. P., Simons, J. V., Soloway, M. G., Mose, L. E., Jefferys, S. R., Balu, S., Parker, J. S., Neil Hayes, D., Perou, C. M., Malik, S., Mahurkar, S., Shen, H., Weisenberger, D. J., Triche Jr, T., Lai, P. H., Bootwalla, M. S., Maglinte, D. T., Berman, B. P., Van Den Berg, D. J., Baylin, S. B., Laird, P. W., Creighton, C. J., Donehower, L. A., Getz, G., Noble, M., Voet, D., Saksena, G., Gehlenborg, N., DiCara, D., Zhang, J., Zhang, H., Wu, C.-J., Yingchun Liu, S., Lawrence, M. S., Zou, L., Sivachenko, A., Lin, P., Stojanov, P., Jing, R., Cho, J., Sinha, R., Park, R. W., Nazaire, M.-D., Robinson, J., Thorvaldsdottir, H., Mesirov, J., Park, P. J., Chin, L., Reynolds, S., Kreisberg, R. B., Bernard, B., Bressler, R., Erkkila, T., Lin, J., Thorsson, V., Zhang, W., Shmulevich, I., Ciriello, G., Weinhold, N., Schultz, N., Gao, J., Cerami, E., Gross, B., Jacobsen, A., Sinha, R., Arman Aksoy, B., Antipin, Y., Reva, B., Shen, R., Taylor, B. S., Ladanyi, M., Sander, C., Anur, P., Spellman, P. T., Lu, Y., Liu, W., Verhaak, R. R. G., Mills, G. B., Akbani, R., Zhang, N., Broom, B. M., Casasent, T. D., Wakefield, C., Unruh, A. K., Baggerly, K., Coombes, K., Weinstein, J. N., Haussler, D., Benz, C. C., Stuart, J. M., Benz, S. C., Zhu, J., Szeto, C. C., Scott, G. K., Yau, C., Paull, E. O., Carlin, D., Wong, C., Sokolov, A., Thusberg, J., Mooney, S., Ng, S., Goldstein, T. C., Ellrott, K., Grifford, M., Wilks, C., Ma, S., Craft, B., Yan, C., Hu, Y., Meerzaman, D., Gastier-Foster, J. M., Bowen, J., Ramirez, N. C., Black, A. D., XPATH ERROR: unknown variable "tname"., R. E., White, P., Zmuda, E. J., Frick, J., Lichtenberg, T. M., Brookens, R., George, M. M., Gerken, M. A., Harper, H. A., Leraas, K. M., Wise, L. J., Tabler, T. R., McAllister, C., Barr, T., Hart-Kothari, M., Tarvin, K., Saller, C., Sandusky, G., Mitchell, C., Iacocca, M. V., Brown, J., Rabeno, B., Czerwinski, C., Petrelli, N., Dolzhansky, O., Abramov, M., Voronina, O., Potapova, O., Marks, J. R., Suchorska, W. M., Murawa, D., Kycler, W., Ibbs, M., Korski, K., Spychała, A., Murawa, P., Brzeziński, J. J., Perz, H., Łaźniak, R., Teresiak, M., Tatka, H., Leporowska, E., Bogusz-Czerniewicz, M., Malicki, J., Mackiewicz, A., Wiznerowicz, M., Van Le, X., Kohl, B., Viet Tien, N., Thorp, R., Van Bang, N., Sussman, H., Duc Phu, B., Hajek, R., Phi Hung, N., Viet The Phuong, T., Quyet Thang, H., Zaki Khan, K., Penny, R., Mallery, D., Curley, E., Shelton, C., Yena, P., Ingle, J. N., Couch, F. J., Lingle, W. L., King, T. A., Maria Gonzalez-Angulo, A., Mills, G. B., Dyer, M. D., Liu, S., Meng, X., Patangan, M., Waldman, F., Stöppler, H., Kimryn Rathmell, W., Thorne, L., Huang, M., Boice, L., Hill, A., Morrison, C., Gaudioso, C., Bshara, W., Daily, K., Egea, S. C., Pegram, M. D., Gomez-Fernandez, C., Dhir, R., Bhargava, R., Brufsky, A., Shriver, C. D., Hooke, J. A., Leigh Campbell, J., Mural, R. J., Hu, H., Somiari, S., Larson, C., Deyarmin, B., Kvecher, L., Kovatich, A. J., Ellis, M. J., King, T. A., Hu, H., Couch, F. J., Mural, R. J., Stricker, T., White, K., Olopade, O., Ingle, J. N., Luo, C., Chen, Y., Marks, J. R., Waldman, F., Wiznerowicz, M., Bose, R., Chang, L.-W., Beck, A. H., Maria Gonzalez-Angulo, A., Pihl, T., Jensen, M., Sfeir, R., Kahn, A., Chu, A., Kothiyal, P., Wang, Z., Snyder, E., Pontius, J., Ayala, B., Backus, M., Walton, J., Baboud, J., Berton, D., Nicholls, M., Srinivasan, D., Raman, R., Girshik, S., Kigonya, P., Alonso, S., Sanbhadti, R., Barletta, S., Pot, D., Sheth, M., Demchok, J. A., Mills Shaw, K. R., Yang, L., Eley, G., Ferguson, M. L., Tarnuzzer, R. W., Zhang, J., Dillon, L. A. L., Buetow, K., Fielding, P., Ozenberger, B. A., Guyer, M. S., Sofia, H. J. & Palchik, J. D. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418), pp. 61–70, doi:10.1038/nature11412.

Mathot, P., Grandin, M., Devailly, G., Souaze, F., Cahais, V., Moran, S., Campone, M., Herceg, Z., Esteller, M., Juin, P., Mehlen, P. & Dante, R. (2017). DNA methylation signal has a major role in the response of human breast cancer cells to the microenvironment. *Oncogenesis*, 6(10), p. e390, doi:10.1038/oncsis.2017.88.

Moore, L. D., Le, T. & Fan, G. (2013). DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, 38(1), pp. 23–38, doi:10.1038/npp.2012.112.

Moran, S., Arribas, C. & Esteller, M. (2016). Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*, 8(3), pp. 389–399, doi:10.2217/epi.15.114.

Mutskov, V. J. (2002). The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. *Genes & Development*, 16(12), pp. 1540–1554, doi:10.1101/gad.988502.

Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N. & Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 393(6683), pp. 386–389, doi:10.1038/30764.

Paludan, K., Duch, M., Jørgensen, P., Kjeldgaard, N. O. & Pedersen, F. S. (1989). Graduated resistance to G418 leads to differential selection of cultured mammalian cells expressing the neo gene. *Gene*, 85(2), pp. 421–426.

Ropero, A., Alonsomagdalena, P., Quesada, I. & Nadal, A. (2008). The role of estrogen receptors in the control of energy and glucose homeostasis. *Steroids*, 73(9–10), pp. 874–879, doi:10.1016/j.steroids.2007.12.018.

Sawan, C., Vaissière, T., Murr, R. & Herceg, Z. (2008). Epigenetic drivers and genetic passengers on the road to cancer. *Mutation Research*, 642(1–2), pp. 1–13, doi:10.1016/j.mrfmmm.2008.03.002.

Schübeler, D. (2015). Function and information content of DNA methylation. *Nature*, 517(7534), pp. 321–326, doi:10.1038/nature14192.

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K. & Schübeler, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, 480(7378), pp. 490–495, doi:10.1038/nature10716.

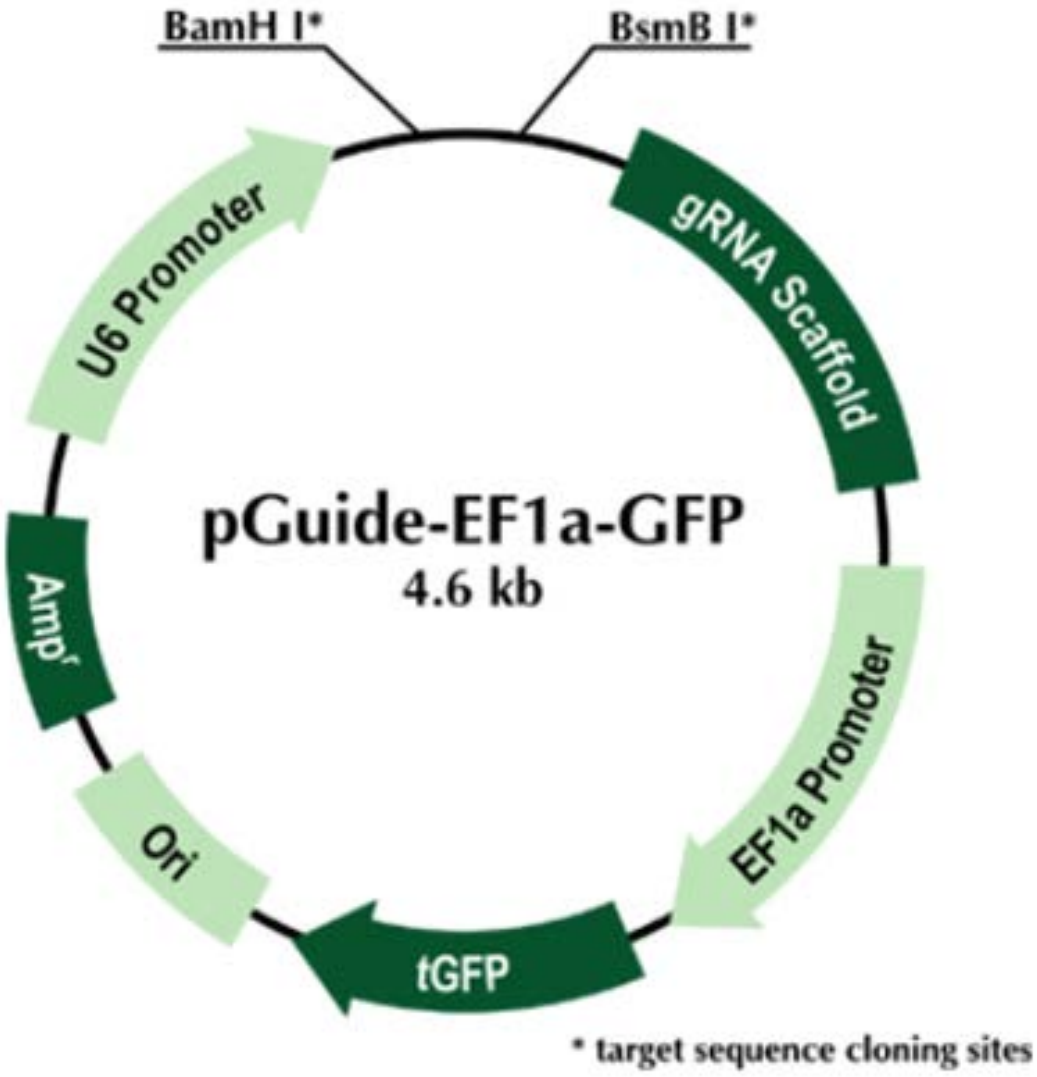
Terry, M. B., Delgado-Cruzata, L., Vin-Raviv, N., Wu, H. C. & Santella, R. M. (2011). DNA methylation in white blood cells: Association with risk factors in epidemiologic studies. *Epigenetics*, 6(7), pp. 828–837, doi:10.4161/epi.6.7.16500.

Umer, M. & Herceg, Z. (2013). Deciphering the epigenetic code: an overview of DNA methylation analysis methods. *Antioxidants & Redox Signaling*, 18(15), pp. 1972–1986, doi:10.1089/ars.2012.4923.

Williams, C. & Lin, C.-Y. (2013). Oestrogen receptors in breast cancer: basic mechanisms and clinical implications, p. 12.

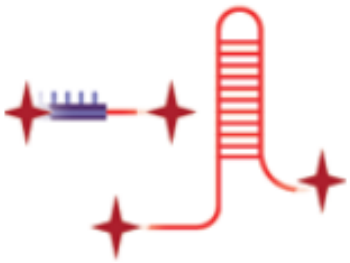
Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S. & Yang, S.-H. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids*, 4, p. e264, doi:10.1038/mtna.2015.37.

Appendix

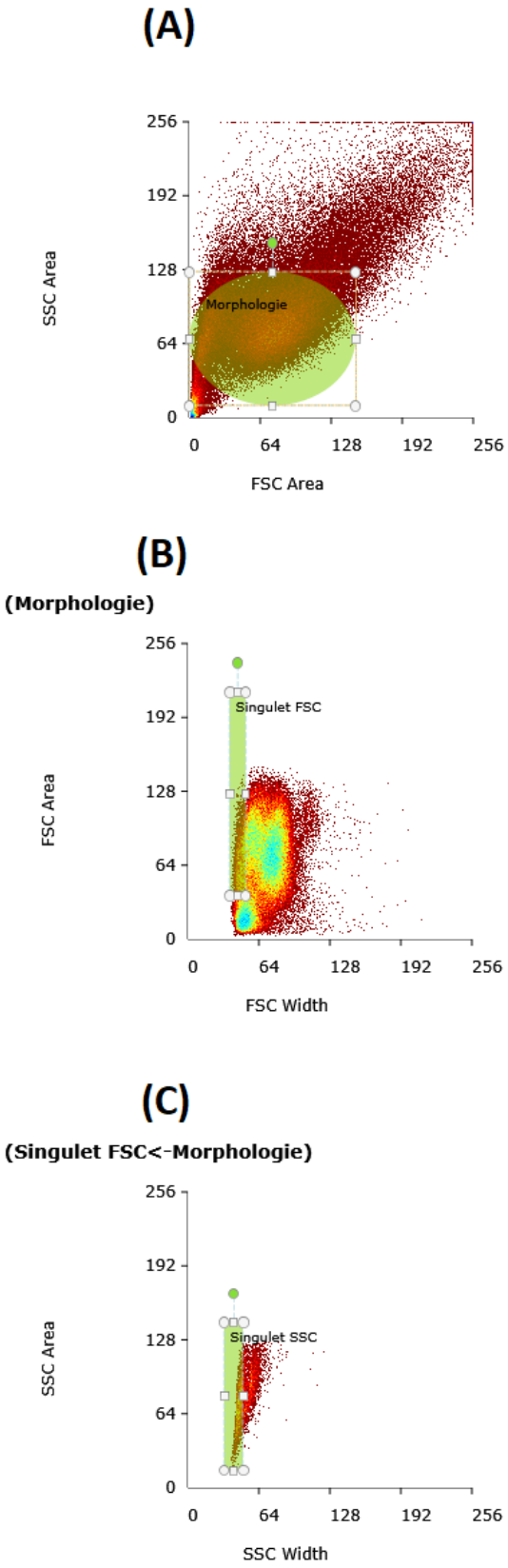


Supplementary figure 1: Scheme of pGuide plasmid's map. This plasmid was used for cloning and delivering of gRNA molecules inside the target cells. (from CRISPR manual by OriGene)

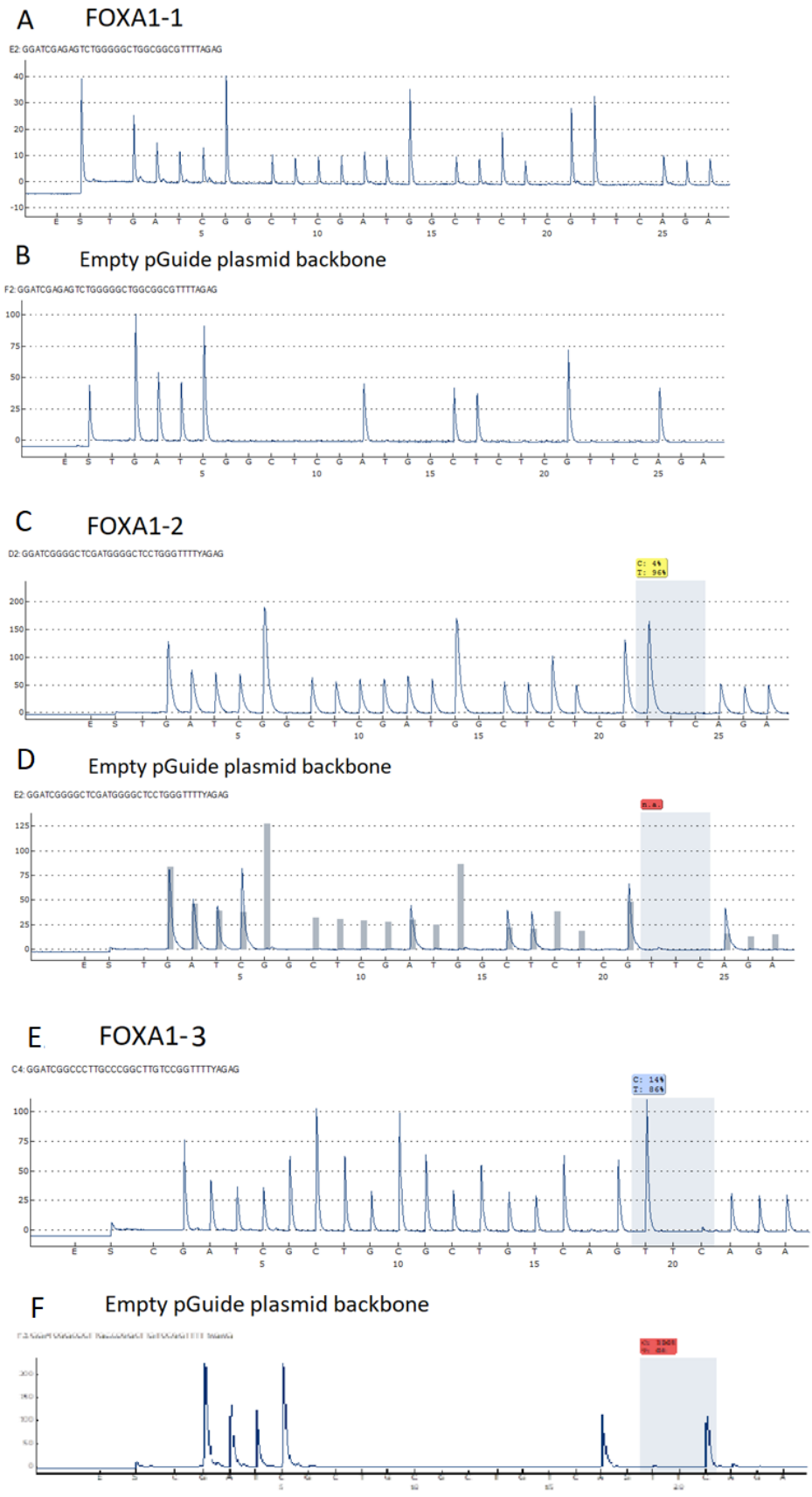
TrueGuide Synthetic gRNA formats

	 TrueGuide 2-piece modified Synthetic gRNA
Predesigned knockouts for mouse and human cells	✓
Custom sequences	✓
Validated controls available	✓
Available in plate format	✓
Ready-to-transfect	Annealing step required
Chemical modifications*	✓
Ideal for primary and stem cells	

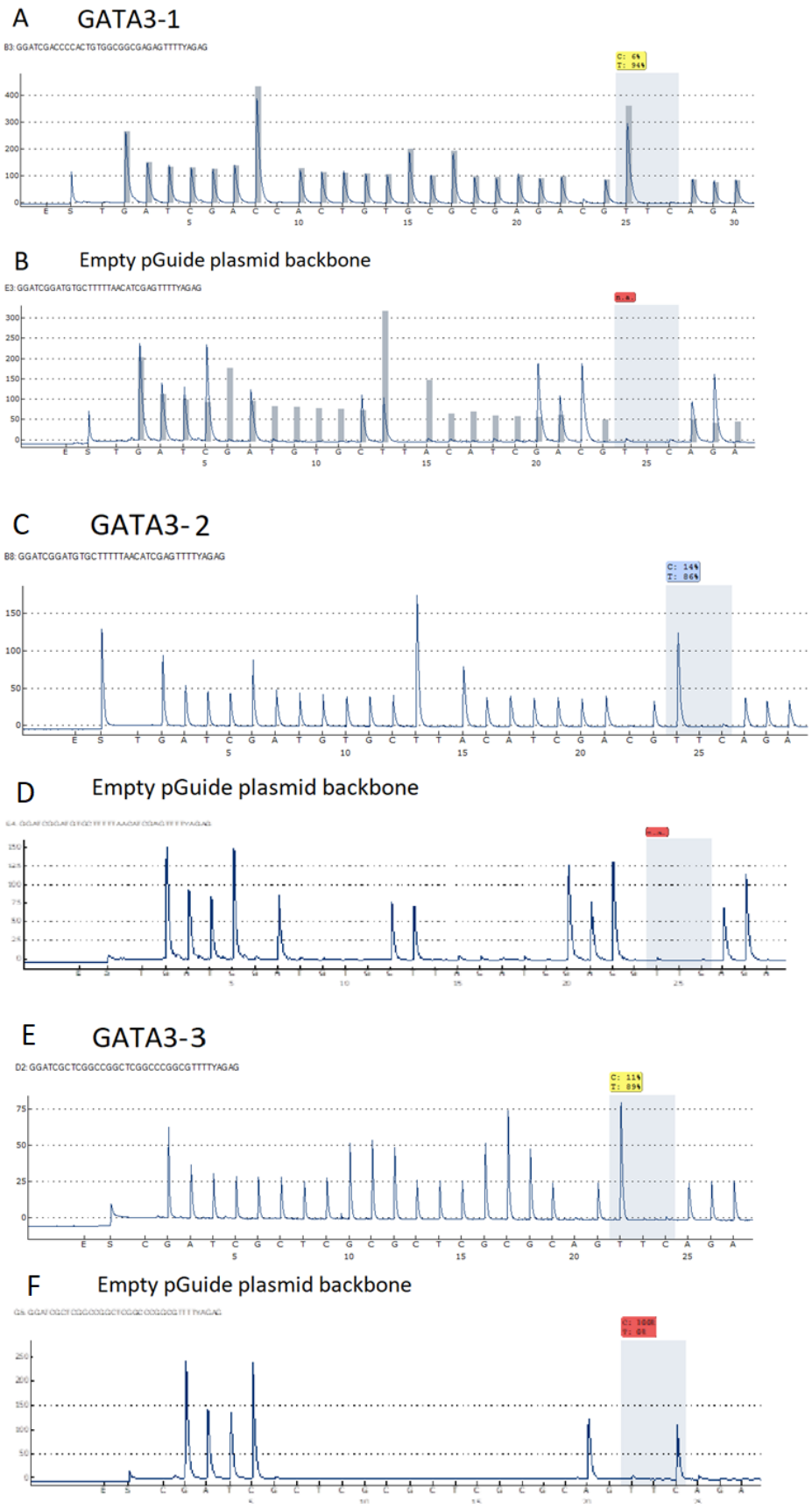
Supplementary figure 2: Scheme of TrueGuide two-piece gRNA synthetic oligos. Chemical modifications include 2´O-Methyl analogues and phosphorothioate linkages to increase editing efficiency and protect against nuclease degradation. (https://www.thermofisher.com/sk/en/home/life-science/genome-editing/geneart-crispr/crispr-libraries/trueguide-grnas.html accessed on: 24/08/2018, 16:06)



Supplementary figure 3: Sorting gates set up on Bio-Rad S3e cell sorter. Part A shows all cells passing through the sorter and the gate for the first selection, aiming for non-apoptotic cells. Parts B and C show cells after first and second gate selection and aim to select single cells only to avoid cells too close to each other and clumped cells.



Supplementary figure 4: Pyrograms of individual confirmed (A, C, E) and false positive (B, D, F) bacterial clones carry the pGuide plasmid. Three different gRNA sequences homologous to different *FOXA1* sites were screened for.



Supplementary figure 5: Pyrograms of individual confirmed (A, C, E) and false positive (B, D, F) bacterial clones carry the pGuide plasmid. Three different gRNA sequences homologous to different *GATA3* sites were screened for.