Toward Multiplex Genome Engineering in Mammalian Cells

Dissertation Advisor: Professor George Church Xavier Ríos Villanueva

Toward Multiplex Genome Engineering in Mammalian Cells

ABSTRACT

Given the explosion in human genetic data, new high-throughput genetic methods are necessary for studying variants and elucidating their role in human disease. In Chapter I, I will expand on this concept and describe current methods for genetically modifying human cells. In *E. coli,* Multiplex Automatable Genome Engineering (MAGE) is a powerful tool that enables the targeting of multiple genomic loci simultaneously with synthetic oligos that are recombined at high frequencies in an optimized strain. MAGE as a method has two components: organism-specific optimization of oligo recombination parameters and a protein capable of increasing recombination frequencies.

In Chapter II, I describe my work in determining and optimizing the parameters for oligo recombination in human cells using a HeLa-based *EGFP*-rescue reporter system. I show that modified base analogs increase oligo recombination frequencies by avoiding the mismatch repair machinery. With an optimized oligonucleotide design, stably *EGFP*-corrected cells were generated at a frequency of ~0.05%. In addition, I investigate oligonucleotide toxicity and provide evidence from comparative RNA-seq analysis suggesting cellular immunity induced by the oligonucleotides might contribute to the low viability of oligo-corrected cells. I tested several proteins that could potentially increase oligo recombination frequencies, but no such activity was found.

Chapter III describes my characterization of the single-strand annealing protein (SSAP) Beta, capable of enhancing oligo recombination frequencies 10,000-fold in *E. coli*, with the goal of translating its function to human cells. This protein is the most important factor mediating MAGE in *E. coli*, but currently not much is known about it beyond its biochemical DNA annealing activity. One fundamental question is how Beta interacts with the *E. coli* machinery. I show that Beta interacts with the *E. coli* Single-Strand Binding Protein (SSB) in an interaction that requires the C-termini of both proteins, further supporting a mechanism of oligo incorporation during DNA replication.

Finally, in Chapter IV I discuss future directions for this work, including further dissection of the biochemical function of Beta, and screens for SSAPs that enhance oligo recombination in any organism of interest.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION: WHY WE NEED MULTIPLEX GENOME ENGINEERING IN MAMMALIAN CELLS

Next-generation sequencing technologies have revolutionized biomedical research, decreasing the cost of sequencing projects several orders of magnitude (Metzker 2010). This has led to a dramatic increase in the amount of human genetic data available. Soon it will be possible to sequence full human genomes for $1,000(Mardis 2006; von Bubnoff 2008), further accelerating the growth of data. Already, exome sequencing projects are revealing many novel variants, including some responsible for rare Mendelian disorders (Bamshad et al. 2011).

Unfortunately, our current ability to sequence and identify genetic variants greatly outpaces our ability to characterize their functional significance. Genome-wide association studies (GWAS) have revealed many single-nucleotide polymorphisms (SNPs) linked to disease phenotypes, most of which are 'tag SNPs' in the context of large genomic regions where the actual causal variants remain unknown (Freedman et al. 2011). It has been suggested that rare variants with large effects could be significant contributors to common human diseases (McClellan & King 2010). Because they are rare, such variants would be underpowered in GWASs, although full genome sequencing of family cohorts can help narrow the possibilities.

Cancer genomics, another field producing huge amounts of sequencing data(Meyerson et al. 2010), would also greatly benefit from improved methods for functionally testing genetic

variants. Comparison of the genome sequence of a lung cancer tumor with adjacent normal tissue revealed 50,675 SNP differences between these(Lee et al. 2010). In addition, studies on tumor evolution and responsiveness to drug can reveal hundreds or thousands of gene copy number changes(Jones et al. 2010). Demonstrating the power direct functional studies can have in this field, deletion of an enhancer region of the MYC gene containing a SNP long associated with tumorigenesis proved to be protective for intestinal tumorigenesis in mice (Sur et al. 2012). However, this was a case where the variant had been heavily studied for years. It remains a daunting challenge to identify new functionally relevant cancer mutations, let alone mutations directly responsible for tumor growth and malignancy.

There are several complementary approaches used to sift through the data generated by GWAS studies. Computational models (Schaub et al. 2012) and other high-throughput technologies such as ChIP-Seq (McLean et al. 2010) can help prioritize SNP functional studies. In one case, an expression quantitative trait loci (eQTL) study of 95 primary human osteoblast cell lines combined with GWAS data on bone density helped narrow down SNPs to a specific eQTL in the *SRR* gene, considered a key player in chondrogenesis (Grundberg et al. 2009). However, this work failed to conclusively identify the causal SNP.

Regulatory region variants are of particular interest, since expression variability of ~5% of all human genes is linked to SNPs located within 200kb of the gene (Pickrell et al. 2010). Although they have smaller effects than protein-coding variants, these variants are important because they are more common (Vernot et al. 2012). Massively parallel reporter studies of cis-regulatory regions can provide a high-throughput way to study these variants (Patwardhan et al. 2012; Melnikov et al. 2012). However such studies still miss the effects of chromosome structure and context. Ultimately, in order to prove causality variants need to be experimentally tested, the

gold standard being traditional genetics, where variants are modified in their native locus to determine what effects, if any, the modification has.

Traditionally, targeted genetic modifications have been generated by selecting for rare homologous recombination (HR) events after introducing a donor DNA that carries the desired mutation along with a selectable gene marker. After selection, the marker may be removed using a targeted recombinase such as Cre. However, the throughput of this approach is limited by the extremely low frequency of HR and the high frequency of random integration events in human cells (Vasquez et al. 2001). In addition, selection markers or residual recombinase sequences could confound subtle phenotypes, such as those caused by variants in regulatory regions.

Custom nucleases such as zinc-finger nucleases (ZFNs) (Urnov et al. 2005), transcription activator-like effector nucleases(TALENs) (Hockemeyer et al. 2011), and more recently CRISPR/Cas9(Mali et al. 2013; Cong et al. 2013) can create site-specific DNA double strand breaks that increase homologous recombination rates several orders of magnitude, allowing screening of seamlessly-modified cells. These targeted nucleases can be used with synthetic single-stranded oligonucleotides as donor, which increases efficiencies and eliminates the donor DNA preparation step (Soldner et al. 2011). Showing their potential for both therapeutics and biomedical research applications, ZFNs have been used to study the effects of reverting Parkinson's disease in patient derived induced pluripotent stem cells (Reinhardt et al. 2013). Using CRISPR/Cas-mediated genome engineering, it is now possible to simultaneously target up to eight alleles in mouse embryonic stem cells, allowing the one-step generation of mouse models with homozygous changes(Wang et al. 2013).

However, production of custom nucleases for each new locus is dependent on knowledge of protein-DNA binding specificities, laborious design and optimization methods or expensive commercial licensing(Maeder et al. 2009; Chandrasekharan et al. 2009; Anon 2011). These limitations have been quickly eliminated, first by simplified TALEN construction methods (Reyon et al. 2012; Briggs et al. 2012) and more recently by adapting the bacterial CRISPR/Cas9 system to work in human cells, whose DNA biding specificity is provided by an easily synthesized guide RNAs complementary to the target site (Cong et al. 2013; Mali et al. 2013). However, the potential for multiplexing custom targeted nucleases may be limited since multiple double strand breaks could lead to gross genome rearrangements or deletions(Şöllü et al. 2010; Doyon et al. 2010) In addition, these targeted nucleases have specificity problems, with off-target binding leading to unwanted double-strand break-mediated mutagenesis(Pattanayak et al. 2011; Fu et al. 2013). Thus, there is still a need for improved genome engineering methods that can keep up with the impending demand for functional characterizations of genetic variants.

Oligonucleotide-mediated recombination is an attractive alternative strategy for genome engineering(Carr & Church 2009) in which mutation-encoding oligonucleotides modify the genome without the need for site-specific DNA-binding proteins. Oligo recombination methods have been developed in mammalian cells with reported gene targeting frequencies of up to 2% (M. Aarts & te Riele 2010; Olsen, Randøl, Luna, et al. 2005; Wu et al. 2005; Papaioannou et al. 2009). There is strong evidence suggesting a mechanism where the oligo is incorporated(Olsen, Randøl, Luna, et al. 2005) in a strand-biased manner(Olsen, Randøl & Krauss 2005; Radecke et al. 2006). The endogenous mismatch repair (MMR) machinery has been found to inhibit oligo incorporation in mammalian cells(Papaioannou et al. 2009; Dekker et al. 2003; Olsen et al. 2009) and oligos modified with phosphorothioate (PTO) bonds, which prevent degradation by

nucleases, have been found to increase recombination frequencies(Papaioannou et al. 2009). The main limitation of oligonucleotide-mediated recombinations in mammalian cells is the low survival of modified cells, with only a few studies having demonstrated long-term survival and proliferation of corrected clones. To increase survival, the use of unprotected oligos has been suggested(Olsen, Randøl, Luna, et al. 2005; Marieke Aarts & te Riele 2010), however this also makes recombination frequencies too low for the method to be useful.

In Chapter II, I evaluate oligo-mediated recombination in human cells in order to elucidate its specific requirements and limitations. Using a previously established broken EGFP reporter cell line, I verified that transfecting short, heavily phosphorothioated (PTO) oligos can achieve high (up to ~5%) recombination frequencies in this cell line. However, after culturing for a few days, the frequency of these modified cells quickly dropped to undetectable levels. Corrected EGFP+ cells were single-cell sorted and this revealed cells recombined using highly PTO-protected oligos were unable to proliferate and establish stably modified clonal populations. Allele-specific qPCR genotyping showed that the EGFP+ cells had at least one copy of the corrected EGFP ORF, suggesting the oligos had successfully modified their target.

Published results suggested PTO-protected oligos might induce DNA damage. I attempted to mitigate this potential DNA damage response using drugs such as ATR/ATM inhibitors and apoptosis inhibitors, but these had no effects on oligo correction frequencies or the proliferation of corrected cells. Eventually, I found that having a high number of PTO modifications in the targeting oligo led to high toxicity and poor proliferation of modified cells. I determined that there is an optimal level of PTO bonds that minimize oligo toxicity while still increasing stable oligo recombination frequencies.

Using this reporter system I was able to test whether chemically modified bases capable of avoiding mismatch repair in *E. coli* would work similarly in human cells. As in *E. coli,* I found that 2'-Fluorouracil produced a ~2-fold increase on oligo recombination frequencies, and this increase was significantly reduced by transiently silencing the mismatch repair machinery. This suggests a mechanism where modified bases increase oligo recombination frequencies at least in part by avoiding mismatch repair.

The main limitation of this method was generating stably modified cells. During the extensive troubleshooting I determined that DNA damage might not be the main cause of the toxicity seen in cells recombined with PTO-modified oligos. Although higher number of PTO bonds correlated with higher levels of DNA damage, corrected EGFP+ cells actually have lower gamma-H2AX phosphorylation levels than EGFP- cells. This finding was surprising, since most oligo toxicity studies have focused on the potential role of PTO modifications in DNA damage. To follow-up on this observation we performed RNA-seq and compared the transcription profiles of corrected and uncorrected cells. This showed a large number of differentially expressed genes, especially genes involved in cellular immunity and viral responses. This observation was validated by qPCR. Oligo-modified cells tend to have higher amounts of oligos in them. If the oligos are invoking cellular immunity and modified cells have higher amounts of oligo, then it is possible that the reason modified cells have a proliferation defect might be due to the aforementioned immune response. It is possible that decreasing the half-life of the oligo by reducing the number of PTO bonds would in turn decrease the signaling through cellular immune ssDNA sensors. Alternatively, the observed phenotype might be due to a cellular immune response induced by abnormal backbone of PTO-modified oligos. Regardless, these observations may explain why reducing the number of PTO bonds reduced their toxicity enough to allow

oligo-recombined cells to proliferate. Cellular immunity in the context of oligo recombination in human cells had not been previously described. It might be possible to use this observation to find ways of suppressing the toxicity of more heavily PTO-modified targeting oligos in order to increase oligo recombination frequencies.

The reporter system described in this chapter enabled some improvement in oligo recombination. However, the method achieves stable oligo correction frequencies of only ~1/1,000 in HeLa cells. Further improvements are necessary for this to be a useful method for characterizing human variants. In *E. coli,* the most important factor for high oligo recombination frequencies is the single-strand annealing protein Beta, increasing oligo recombination frequencies by four orders of magnitude. I tested if expressing Beta and similar proteins in human cells would have any effect in oligo-recombination frequencies. Even when trying different oligo sizes and verifying expression and nuclear localization of the protein, I was unable to detect any effect in oligo recombination frequencies.

This result led me to the study in Chapter III, where I characterize the Beta protein biochemically in order to elucidate how it enables multiplex genome engineering in *E. coli* (MAGE). Optimization of this method by knocking out the mismatch repair pathway and enhancing the stability of the ssDNA oligo with phosphothioate bonds led to oligo recombination frequencies as high as 20-50% (Wang et al. 2009). The inherent simplicity and scalability of this method makes it particularly well-suited for generating a large number of genetic variants. When performed with multiplexed degenerate oligos, MAGE can generate $>4x10^6$ combinatorial genomic variations.

Currently, the exact mechanism of Beta-mediated oligonucleotide recombination is not understood. It is known that an oligo that anneals with the lagging strand is more efficiently recombined than one targeting the leading strand(Ellis et al. 2001). This suggests a mechanism of oligo incorporation where the discontinuously synthesized lagging strand would be more accessible for the invading oligo, and after annealing to its target sequence the oligo gets incorporated as a pseudo-Okazaki fragment (Ellis et al. 2001; Court et al. 2002). Very little is known about Beta's biochemical ability beyond annealing, strand invasion and limited strand displacement activities (Mythili et al. 1996; Li et al. 1998; Rybalchenko et al. 2004).

The fact that I and others(van Kessel & Hatfull 2007; Swingle, Bao, et al. 2010; Binder et al. 2013) have observed that Beta is not capable of significantly enhancing oligo recombination in other organisms suggests that it may interact with a specific component of the host's machinery. In eukaryotic cells, the SSAP RAD52 is capable of DNA annealing in the presence of its single-strand binding protein RPA (Sung 1997; New et al. 1998; Shinohara & Ogawa 1998). Similarly, in *E. coli* and other bacteria RecO anneals DNA in the presence of SSB (Kantake et al. 2002).

I hypothesized that Beta acts in a similar fashion, removing the annealing inhibition of SSB in an interaction requiring the C-terminus of the Beta protein, a domain previously shown to be dispensable for Beta's DNA annealing activity (Wu et al. 2006). I show that deleting the C-terminus of the protein greatly reduces oligo recombination. Using a His-tag purification strategy I measured the effects of different Beta mutations in an *in vitro* DNA annealing assay. In this assay two complementary 90mer oligos, one with a 3'-Fluorescein and the other with a 5'-Iowa Black FQ dark quencher are mixed and the annealing reaction can be measured on a plate reader as the fluorescein is quenched. Using this assay I found that Beta is capable of annealing

complementary ssDNA oligos coated with SSB in a reaction requiring magnesium. Only full-length Beta is capable of annealing SSB-coated oligos, with deletions and single-amino acid substitutions abolishing function. When using a SSB with a C-terminal eight amino acid deletion Beta is no longer capable of annealing the oligos. There is growing evidence suggesting that the SSB C-terminus interaction is fairly species-specific(Manfredi et al. 2010). This terminal domain sequence is known to be required for SSB to interact with other cellular partners such as RecO(Hobbs et al. 2007), DNA polymerase IV(Furukohri et al. 2012) and ExoI(Lu & Keck 2008). Similarly, in the SV40 *in vitro* eukaryotic DNA replication assay only human and not yeast RPA is able mediate replication(Wang et al. 2000). Thus, one possible reason for Beta having no activity in human cells might be that it cannot interact with human RPA.

Future work will focus on expanding and validating these observations. I will compare Beta to other SSAPs using the *in vitro* annealing assay to search for any specific biochemical properties of Beta that might explain its high activity in oligo-mediated recombination. For example, it would be interesting to compare Beta with RecT, another *E. coli* phage-derived SSAP that is 1,000-fold less efficient than Beta in oligo-mediated recombination(Datta et al. 2008). In addition, I will expand this by creating a library of phage-derived SSAPs, select for the most active members in oligo recombination and characterizing them biochemically. Insights from this characterization may lead to a rational design of SSAPs capable of enhancing oligo recombination rates in human cells.

CHAPTER TWO

STABLE GENE TARGETING IN HUMAN CELLS USING SINGLE-STRAND OLIGONUCLEOTIDES WITH MODIFIED BASES

Xavier Rios[1], Adrian W. Briggs[1], Danos Chistodoulou[1], Josh M. Gorham[1], Jonathan G. Seidman[1], and George M. Church[1,2]

[1]Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston 02115 USA
[2]Corresponding author

Xavier Rios design the experiments, analyzed the data and wrote the manuscript, Adrian Briggs contributed with the allele-specific qPCR experiments and writing of the manuscript, Danos Chistodoulou analyzed the RNA-seq data, Josh M Gorman constructed the RNA-seq library. John Seidman and George Church helped plan the experiments and write the manuscript.

INTRODUCTION

As discussed in the previous chapter, there has been an explosion in the amount of human genetic data available (Mardis 2006; von Bubnoff 2008). Unfortunately, our current ability to sequence and identify genetic variants greatly outpaces our ability to characterize their functional consequences. Genome-wide association studies (GWAS) have revealed many single-nucleotide polymorphisms (SNPs) linked to disease phenotypes in the context of large genomic regions (Freedman et al. 2011). Ultimately, potential causal variants linked to these SNPs need to be experimentally tested. Traditionally, targeted genetic modifications are achieved by selecting for rare homologous recombination(HR) events using a donor DNA that carries the desired mutation as well as a selectable gene marker. However, the throughput of this approach is limited by the extremely low frequency of HR in human cells (Vasquez et al. 2001).

Custom targeted zinc-finger nucleases (ZFNs)(Urnov et al. 2005), and more recently transcription activator-like effector nucleases(TALENs)(Hockemeyer et al. 2011) and CRISPR/Cas9(Mali et al. 2013; Cong et al. 2013) have shown great promise by increasing homologous recombination rates, allowing screening of seamlessly-modified cells. However, the potential for multiplexing custom targeted nucleases is limited, since multiple double strand breaks could lead to gross genome rearrangements or deletions(Şöllü et al. 2010; Doyon et al. 2010). In addition, these proteins tend to have high off-target mutagenesis rates(Pattanayak et al. 2011; Fu et al. 2013), which would confound functional studies. Thus, a different genome engineering technology will be necessary to keep up with the demand for post-GWAS functional characterizations.

Oligonucleotide-mediated targeting is an alternative strategy for genome engineering(Carr & Church 2009), in which mutation-encoding oligonucleotides modify the genome without the need for custom-designed DNA-binding proteins. The inherent simplicity and scalability of this method makes it particularly well suited for generating a large number of genetic variants. In *E. coli* lambda Red-mediated oligonucleotide recombineering, an oligo preferentially anneals to the lagging strand of the genome during DNA replication and incorporates into the daughter strand(Ellis et al. 2001). This has been optimized to achieve targeting efficiencies as high as 20%, and when performed with multiplexed degenerate oligos can generate $>4x10^6$ combinatorial genomic variations(Wang et al. 2009). A similar method has been developed in mammalian cells, with reported gene targeting frequencies of up to 2%(M. Aarts & te Riele 2010; Olsen, Randøl, Luna, et al. 2005; Wu et al. 2005; Papaioannou et al. 2009). There is strong evidence showing oligo incorporation(Olsen, Randøl, Luna, et al. 2005) and strand bias(Olsen, Randøl & Krauss 2005; Radecke et al. 2006). Similarly to *E. coli,* the endogenous mismatch repair (MMR) machinery has been found to inhibit oligo incorporation in mammalian cells(Dekker et al. 2003; Papaioannou et al. 2009; Olsen et al. 2009), and oligos modified with phosphorothioate (PTO) bonds to prevent nuclease degradation have been found to increase targeting efficiencies(Papaioannou et al. 2009).

The main limitation of oligonucleotide-mediated targeting in mammalian cells is the low survival of modified cells, with only a few studies having demonstrated long-term survival and proliferation of corrected clones. To increase survival, the use of unprotected oligos has been suggested(Marieke Aarts & te Riele 2010; Olsen, Randøl, Luna, et al. 2005), however this also makes corrections too low for multiplexed genome engineering. Here I describe our progress towards surpassing these limitations. Using an *EGFP*-rescue based reporter system in HeLa

cells(Wu et al. 2005), I found that the use of commercially available chemically modified base analogs to encode the mismatch in the targeting oligo increase its frequency of incorporation, similar to results in *E. coli* (Wang et al. 2011). I also found that the number of PTO modifications correlate with both the toxicity of oligo transfection and the proliferation of oligo-modified cells, and having 3-5 PTO bonds 3' to the mismatch produced a >10-fold increase in stable targeting compared to an unmodified oligo. Applying these oligo design principles I was able to stably correct a mutant *EGFP* with an efficiency of ~0.05% in HeLa cells. Finally, by comparing the transcription profiles of *EGFP*-corrected and -uncorrected cells via RNAseq, I found upregulation of many genes associated with cellular immunity. I propose a new model explaining the toxicity in oligo-mediated recombination in human cells, where cells that uptake more oligo are more likely to become corrected, and likewise these cells are more likely to activate cellular immune responses. Future exploration of this cellular response to transfected oligos may lead to novel strategies for reducing their toxicity and eventually enable multiplexed genome engineering in human cells.

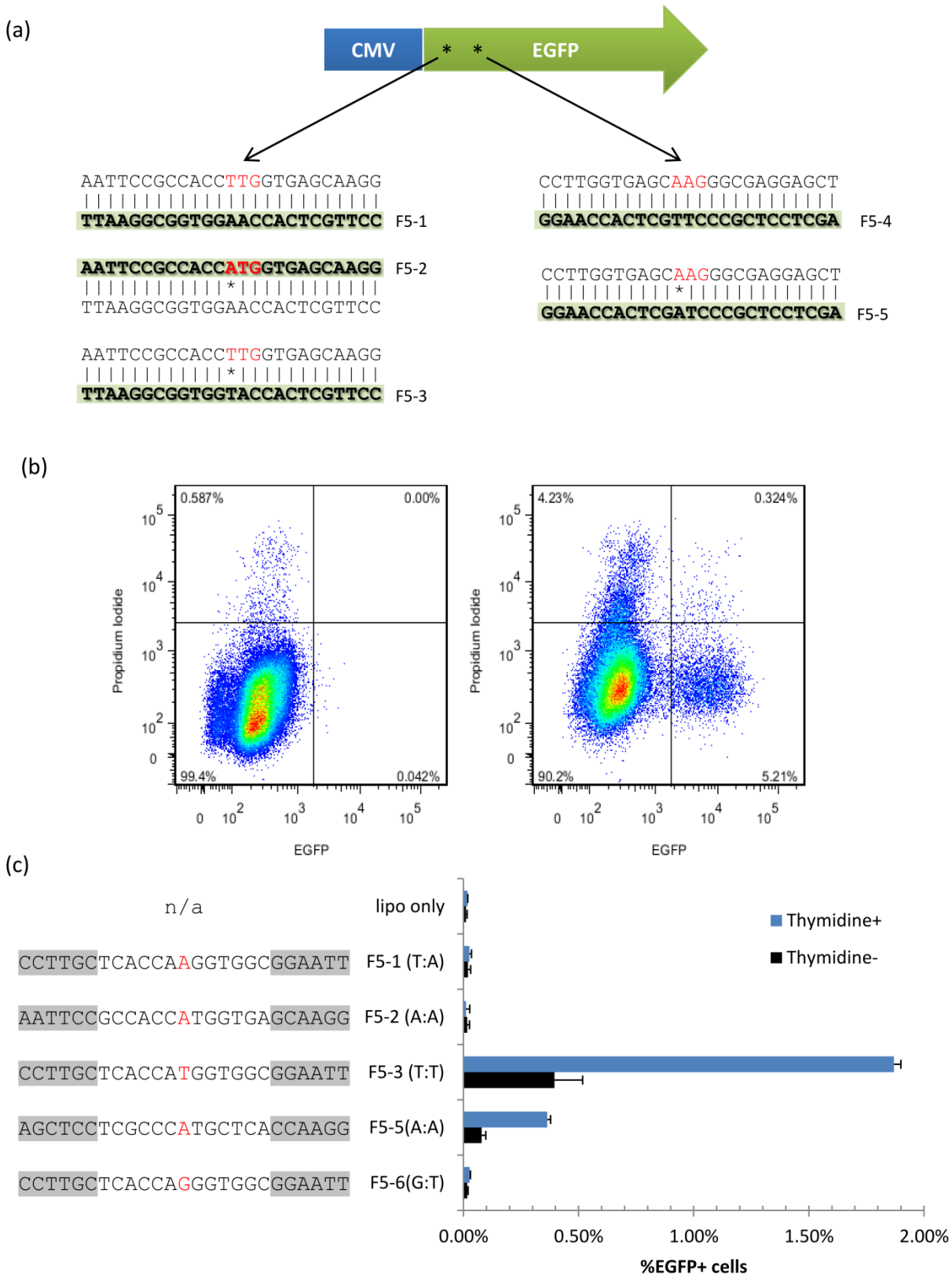
RESULTS

Mismatch-specific EGFP correction

To improve human genome engineering with oligonucleotides, I worked with a well-characterized reporter system(Wu et al. 2005) consisting of a HeLa cell line with two stably integrated copies of a modified *EGFP* gene (*mEGFP*). This version of *EGFP* has a non-functional start codon (TTG) which can be rescued by integration of targeting oligos (Figure

Figure 2.1.Oligo-mediated targeting reporter system. (a) Reporter consists of a HeLa cell line with two stably integrated copies of EGFP with a mutated TTG start codon and a second potential start codon (AAG) downstream, both shown in red. Each mutated start codon can be targeted by sense or antisense oligos. Representative oligos used in this study are highlighted, in duplex form, with the mismatch shown by an asterisk. Further detail can be found on Table 1. (b) Sample flow cytometry dot plot of cells with no oligo (left) or transfected with F5-8 (right, Table 1). The frequency of oligo-induced correction of a start codon can be estimated as the %EGFP+/Propidium Iodide- cells (c) %EGFP+ cells generated after different oligo transfections, with and without 20mM thymidine, assayed 36h post transfections. Oligo sequences are shown 5' to 3' , and the PTO bonds are highlighted in gray. The control oligos F5-1 and F5-6 did not produce any significant proportion of EGFP+ cells, neither did F5-2, which is complementary to the transcribed strand and encodes an ATG-restoring mutation. The oligos targeting the first and second potential start codons on the non-transcribed strand, F5-3 and F5-5 respectively, did produce EGFP+ cells. Lipo=lipofectamine only control. n=4 technical replicates

Figure 2.1 (Continued)



2.1a, Table 2.1) encoding for a functional ATG, thus the oligo-mediated targeting efficiency can be determined by flow cytometry as the percentage EGFP+ cells (Figure 2.1b). I confirmed previous findings(Yin et al. 2005) where targeting *mEGFP* with a 25bp long oligo complementary to the non-transcribed strand and carrying a centrally located mismatch and six PTO bonds at each end (F5-3, Figure 2.1c) delivered with Lipofectamine 2000 yielded a substantial proportion of EGFP+ cells after 48 hours (~0.5%), and this efficiency was further increased to ~2% by slowing down DNA replication with thymidine treatment. Other cationic lipid transfection reagents resulted in lower efficiencies, possibly due to low nuclear accumulation of the oligo (Supplementary Figure 2.1). Alternative delivery methods, such as electroporation and nucleofection failed to produce any significant proportion of EGFP+ cells (not shown). An oligo complementary to the transcribed strand, F5-2, did not produce any EGFP+ cells significantly different from background. Transfection of an oligo encoding an alternative ATG-restoring mutation 9bp away from the first site resulted in ~0.4% *EGFP*+ cells (F5-5, Figure 2.1a,c), whereas an oligo carrying a non-coding mismatch (F5-6, Figure 2.1c) did not produce EGFP+ cells, demonstrating that the expression of EGFP depends on the targeting oligo sequence restoring the ORF. To further verify that EFGP+ cells had undergone the desired genomic modification, EGFP+ and – cells were sorted out by fluorescence-activated cell sorting (FACS) and genotyped by allele-specific qPCR (AS-qPCR). The EGFP+ population was estimated by AS-qPCR to carry 11-13% converted DNA, which matches the 12.5% expected if these cells have undergone a single oligo incorporation at one of the two genomic *mEGFP* during DNA replication (1/8 strands at the end of S phase) but not yet proceeded through cell division. In the EGFP- population, the proportion of corrected alleles was estimated to be ~1%,

Table 2.1. Oligonucleotides used in this study.

Oligo name	Sequence	Corresponding Fig. 2.1 strand
F5-1	C*C*T*T*G*C*TCACCAAGGTGGC*G*G*A*A*T*T	F5-1
F5-2	A*A*T*T*C*C*GCCACCATGGTGA*G*C*A*A*G*G	F5-2
F5-3	C*C*T*T*G*C*TCACCATGGTGGC*G*G*A*A*T*T	F5-3
F5-4	A*G*C*T*C*C*TCGCCCTTGCTCA*C*C*A*a*G*G	F5-4
F5-5	A*G*C*T*C*C*TCGCCCATGCTCA*C*C*A*A*G*G	F5-5
F5-6	A*G*C*T*C*C*TCGCCCGTGCTCA*C*C*A*A*G*G	F5-3
F5-7	C*C*T*T*G*C*TCACCAdUGGTGGC*G*G*A*A*T*T	F5-3
F5-8	C*C*T*T*G*C*TCACCAFuGGTGGC*G*G*A*A*T*T	F5-3
F5-9	A*G*C*T*C*C*TCGCCCFaTGCTCA*C*C*A*A*G*G	F5-4
F5-10	A*G*C*T*C*C*TCGCCCAmTGCTCA*C*C*A*A*G*G	F5-4
F5-11	C*C*T*T*G*C*TCACCATGGTGGCGGAATT	F5-3
F5-12	CCTTGCTCACCATGGTGGC*G*G*A*A*T*T	F5-3
F5-13	C*C*T*TGCTCACCATGGTGGCGGA*A*T*T	F5-3
F5-14	C*C*T*T*G*CTCACCATGGTGGCG*G*A*A*T*T	F5-3
F5-15	C*C*T*T*G*C*T*CACCATGGTGG*C*G*G*A*A*T*T	F5-3
F5-16	C*C*T*T*G*C*T*C*A*CCATGGT*G*G*C*G*G*A*A*T*T	F5-3
F5-17	CCTTGCTCACCAFu*G*G*TGGCGGAATT	F5-3
F5-18	CCTTGCTCACCAFu*G*G*T*G*G*CGGAATT	F5-3
F5-19	C*C*T*T*G*C*TCACCAFuGGTGGC*G*G*A*A*T*T	F5-3
F5-20	CCTTGCTCACCAA*G*G*TGGCGGAATT	F5-1
F5-21	CCTTGCTCACCAFuGGTGGCGGAATT	F5-3
F5-22	CCTTGCTCACCAFu*GGTGGCGGAATT	F5-3
F5-23	CCTTGCTCACCAFu*G*GTGGCGGAATT	F5-3
F5-24	CCTTGCTCACCAFu*G*G*T*GGCGGAATT	F5-3
F5-25	CCTTGCTCACCAFu*G*G*T*G*GCGGAATT	F5-3
F5-26	CCTTGCTCACCAFu*G*G*T*G*G*C*GGAATT	F5-3
F5-27	CCTTGCTCACCAFu*G*G*T*G*G*C*G*GAATT	F5-3
F5-28	CCTTGCTCACCAFu*G*G*T*G*G*C*G*G*AATT	F5-3
F5-29	CCTTGCTCACCAT*G*G*TGGCGGAATT	F5-3
F5-30	CCTTGCTCACCAFuGGTGGCGGA*A*T*T	F5-3
F5-31	AATTCCGCCACCA*T*G*GTGAGCAAGG	F5-2
F5-32	AATTCCGCCACCAm*T*G*GTGAGCAAGG	F5-2
F5-33	CCTTGCTCACCAFu*A*G*TGGCGGAATT	F5-3
F5-34	CCTTGCTCACCAFu*A*G*GTGGCGGAAT	F5-3
F5-35	CCTTGCTCACCAT*A*G*TGGCGGAATT	F5-3
F5-36	CCTTGCTCACCAT*A*G*GTGGCGGAAT	F5-3
F5-37	Cy5-GGCTCCTTCAGCT*G*T*GAGACTACGT	n/a
F5-38	CCTTGCTCACCAFu*G*G*TGGmCGGAATT	F5-3

Sequences shown 5' to 3'. PTO bonds shown as asterisks (*). dU=deoxyUridine, Fu=2'-Fluorouracil, Fa= 2'-Fluoroadenine, Am=2-Aminopurine, mC= 5-Methyl deoxyCytidine

which may either represent corrected cells that had not yet produced functional EGFP protein or PCR artifacts caused by residual targeting oligonucleotides in the cells(Disterer et al. 2009).

Modified bases increase efficiency in part by avoiding mismatch repair in mammalian cells

Previous work has shown that the MMR machinery plays a significant role recognizing and removing the mutation caused by the oligo incorporation event. Currently, the only way around this is by either completely knocking down one of the main MMR proteins (e.g. MSH2) or by transient silencing with RNAi(Aarts et al. 2009). Our group recently showed an alternative strategy in *E. coli,* where the oligo contains chemically modified bases that avoid mismatch repair recognition(Wang et al. 2011). I tested oligos complementary to the non-transcribed strands of the two potential start codons of *mEGFP* while varying the mismatched base (Figure 2.2). The best replacement for the T-T mismatch on the first start codon was 2'-Fluorouracil (FU), while for the A-A mismatch on the second start codon 2-Aminopurine (AM) was best, each giving a ~2-fold increase in *mEGFP* correction efficiency (Figure 2.2a,b). To test whether the increased efficiency in *mEGFP* correction by the modified bases was due to avoidance of MMR, I transfected cells with validated shRNAs plasmids targeting *MSH2* and *MLH1*(Table2.2),then tested the *mEGFP* correction efficiency using oligos with the different modified bases.

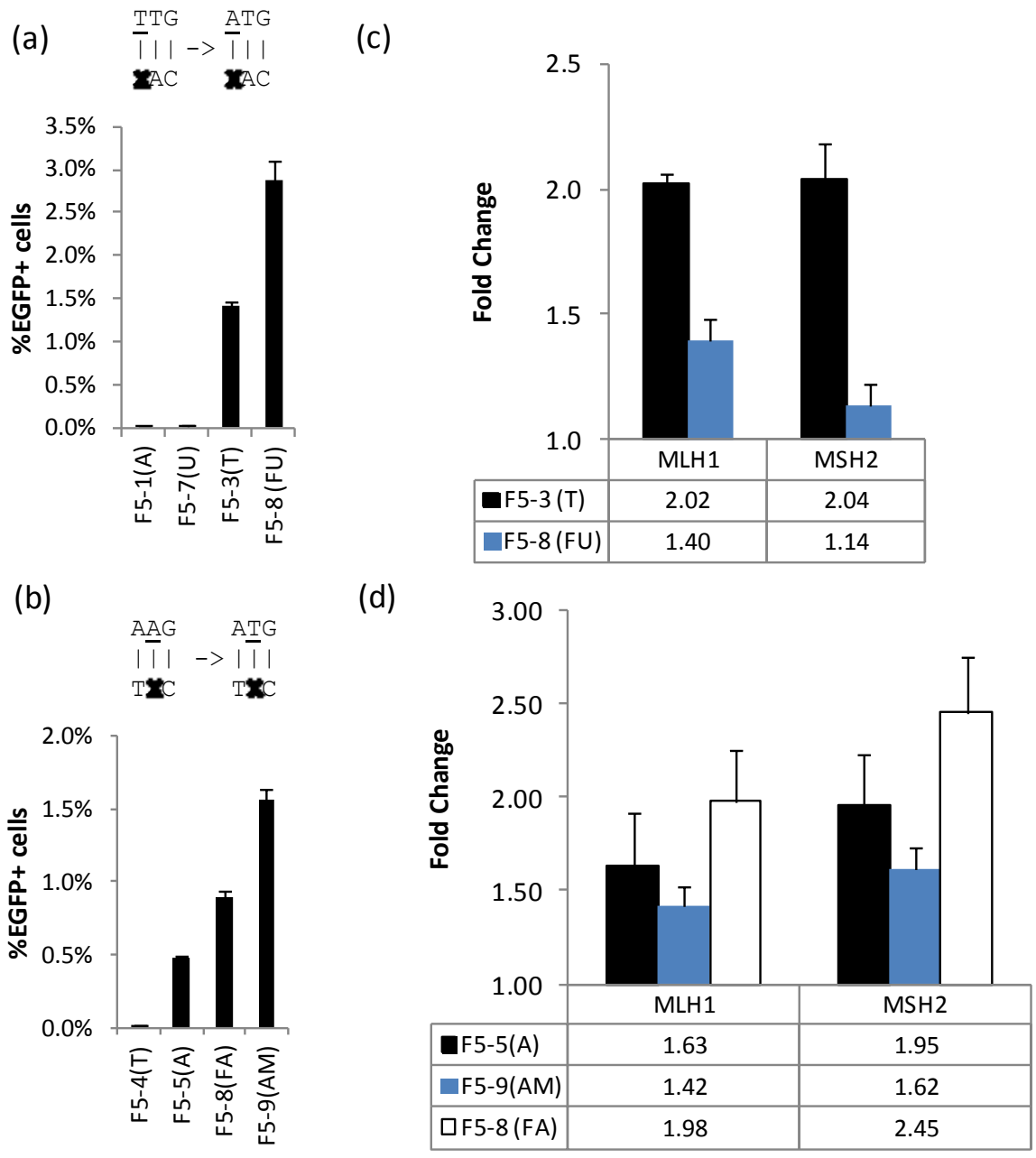


Figure 2.2 Chemically-modified base analogs. (a-b)Modified base-containing oligos complementary to the non-transcribed strand's (a)first potential start codon TTG and (b) second potential start codon AAG, where each mismatched base X in the targeting oligo is shown in parenthesis. (c-d) RNAi targeting key mismatch repair proteins MLH1 and MSH2 for the (a)TTG and (b)AAG start codon targeted by oligos containing modified bases. Data was normalized relative to scr shRNA, thus in (c) each MMR component silencing produced a 2-fold improvement for the natural T base, while the improvement for FU was reduced. This is seen to a lesser degree in (d) comparing A and AM, while FA was further improved, suggesting it is more strongly recognized by MMR. n=4 technical replicates

Table 2.2 shRNAs sequences

Name	Target	Sequence
TRCN0000040056	MLH1	CCGGCCAAGTGAAGAATATGGGAAACTCGAGTTTCCCATATTCTTCACTTGGTTTTTG
TRCN0000039670	MSH2	CCGGGCCTTGCTGAATAAGTGTAAACTCGAGTTTACACTTATTCAGCAAGGCTTTTTG
#1864 (Addgene)	scramble	CCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGG

Table 2.3 PCR primers sequences

PCR primers	
Name	**Sequence**
N1	TCAATGGGCGTGGATAGCGG
N3	CGTTGTGGCTGTTGTAGTTG
EGFP(-)-specific forward primer	CGCTGAATTCCGCCACCt
EGFP(+)-specific forward primer	CGCTGAATTCCGCCACCa
EGFP Reverse primer	CAGGGTCAGCTTGCCGTAGG
qPCR primers	
Name/Target	Sequence
POLR2B Fwd(housekeeping)	GCGGATGAGGATATGCAATATGA
POLR2B Rvs	ACCAAGCCTTTCTCGTCAAAA
IL32 Fwd	ATGTGCTTCCCGAAGGTCCTC
IL32 Rvs	TCATTTTGAGGATTGGGGTTC
HLAB Fwd	CAGTTCGTGAGGTTCGACAG
HLAB Rvs	CAGCCGTACATGCTCTGGA
OAS3 Fwd	TCTGAGACTCACGTTTCCTGA
OAS3 Rvs	CACTGTTGAGGAGGGTAGAGTA

Downregulating either MSH2 or MLH1, as confirmed by western blotting (Supplementary Figure 2.2a), led to a ~2-fold increase in *mEGFP* targeting efficiency in both the T-T and the A-A mismatched oligos (Figure 2.2c,d). However, this increase was significantly lower for FU-T (Student t-test p-val=1.26E-5 for *MLH1*, 3.42E-5 for *MSH2*) and somewhat lower for AM-A mismatch (p-val=0.07 for *MLH1*). Interestingly, FA seems to benefit more from MMR silencing than the natural base, suggesting an alternate mechanism is producing the increase in targeting efficiency observed. These results were verified by targeting MSH2 and MLH1 with siRNAs (Supplementary Figure 2.2b,c). Thus, silencing *MSH2* and *MLH1* decreased the gap in targeting frequencies seen between the natural and modified bases, suggesting that the increase in targeting efficiency by oligos containing modified bases can be explained in part due to avoidance of MMR.

Phosphorothioate modifications correlate with oligo transfection toxicity

By integrating an oligo into the transcribed strand of *mEGFP* during DNA replication, oligo incorporation can be quantified early, since transcription and translation of *EGFP* can already take place before cell division. It has been shown that post-replication/pre-division corrected cells each contain only one out of four strands corrected at the individual target locus, so upon two rounds of cell division only 25% of cells stemming from the original EGFP+ population will still be *EGFP*+ (Marieke Aarts & te Riele 2010). However, I tracked the percentage of EGFP+ cells generated with oligo F5-3 over time and found an even greater decrease than expected; after 5 days EGFP+ cells had dropped to undetectable levels (not shown). This indicates that the corrected cells were either proliferating slower than the non-corrected cells or dying. To distinguish these possibilities I sorted corrected single cells 48hrs

post F5-3 oligo transfection into 96-well plates. After multiple attempts, no *EGFP*+ clones were detected even when using oligos with the FU base analog, suggesting that corrected cells were unable to proliferate after oligo incorporation. This is in agreement with previous results(Liu et al. 2009), that were also unable to generate stably corrected clones with this HeLa F5 cell line.

PTO modifications are used to protect oligos from nuclease degradation, however this modification has been reported to be toxic to cells(Liu et al. 2009). To quantify this toxicity in relationship to *mEGFP* targeting efficiency I varied the number of PTO modifications in the targeting oligos and measured the effects on both cell survival and correction efficiency after 48hrs. I found a clear trend where increased number of PTO bonds results in lower survival of cells 48hrs after oligo transfection (Figure 2.3a), consistent with PTOs being toxic. In addition, an oligo with PTOs only at the 3' end (F5-12) led to higher correction efficiency than one with PTOs only at the 5' end (F5-11, Fig. 2.3a). Interestingly, *mEGFP* correction steadily decreased with additional PTOs when there were more than six at each end. These results suggested that finding a balance between toxicity and efficiency might be necessary for generating stably corrected cells.

Cells corrected with less-protected oligos are more likely to proliferate

I hypothesized that reducing the number of PTO bonds might increase the number of proliferating EGFP+ cells. However, since reducing the number of PTO bonds also decreases targeting efficiency, I decided to concentrate on protecting the 3' end. Furthermore, I decided to test internally protected oligos, since these have been found to lead to higher oligo survival rates(Papaioannou et al. 2009). I designed two new oligos with three or six PTO bonds 3' to the centrally located mismatch, all using a FU base analog (Table 2.1, Fig. 2.3b).

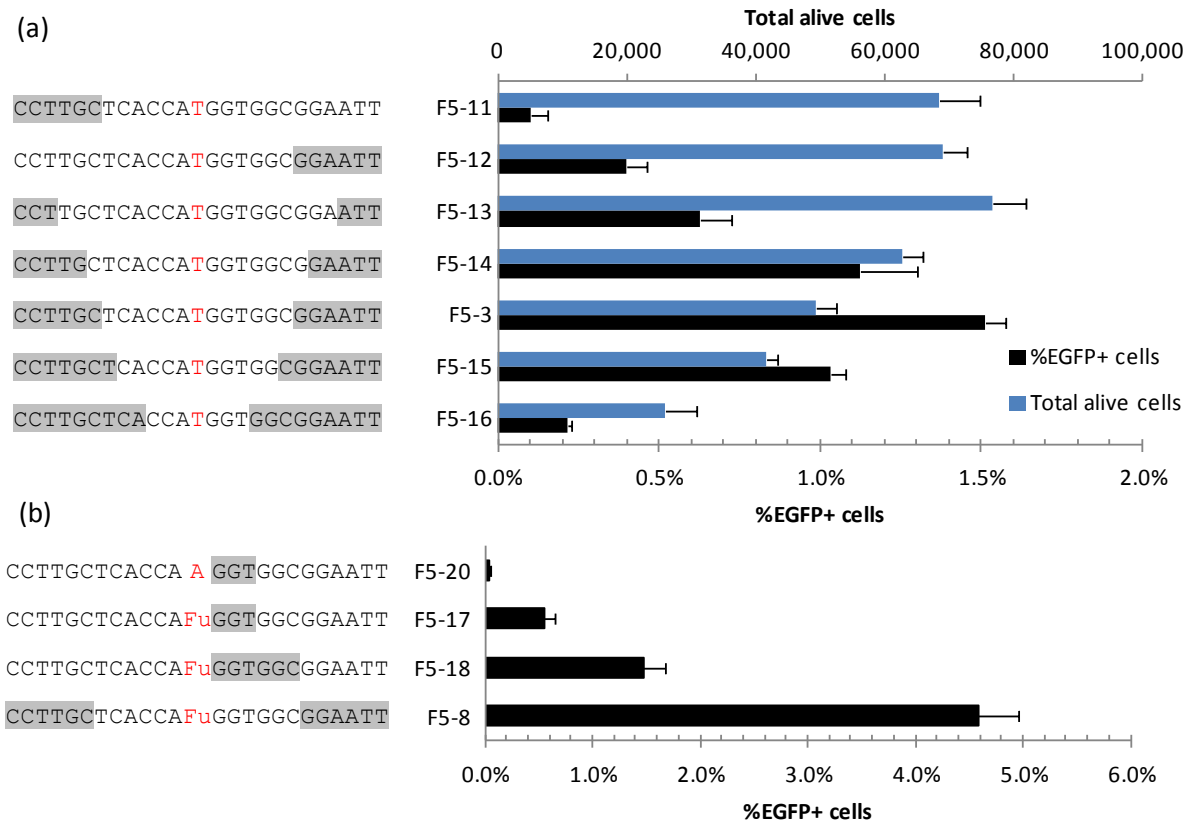
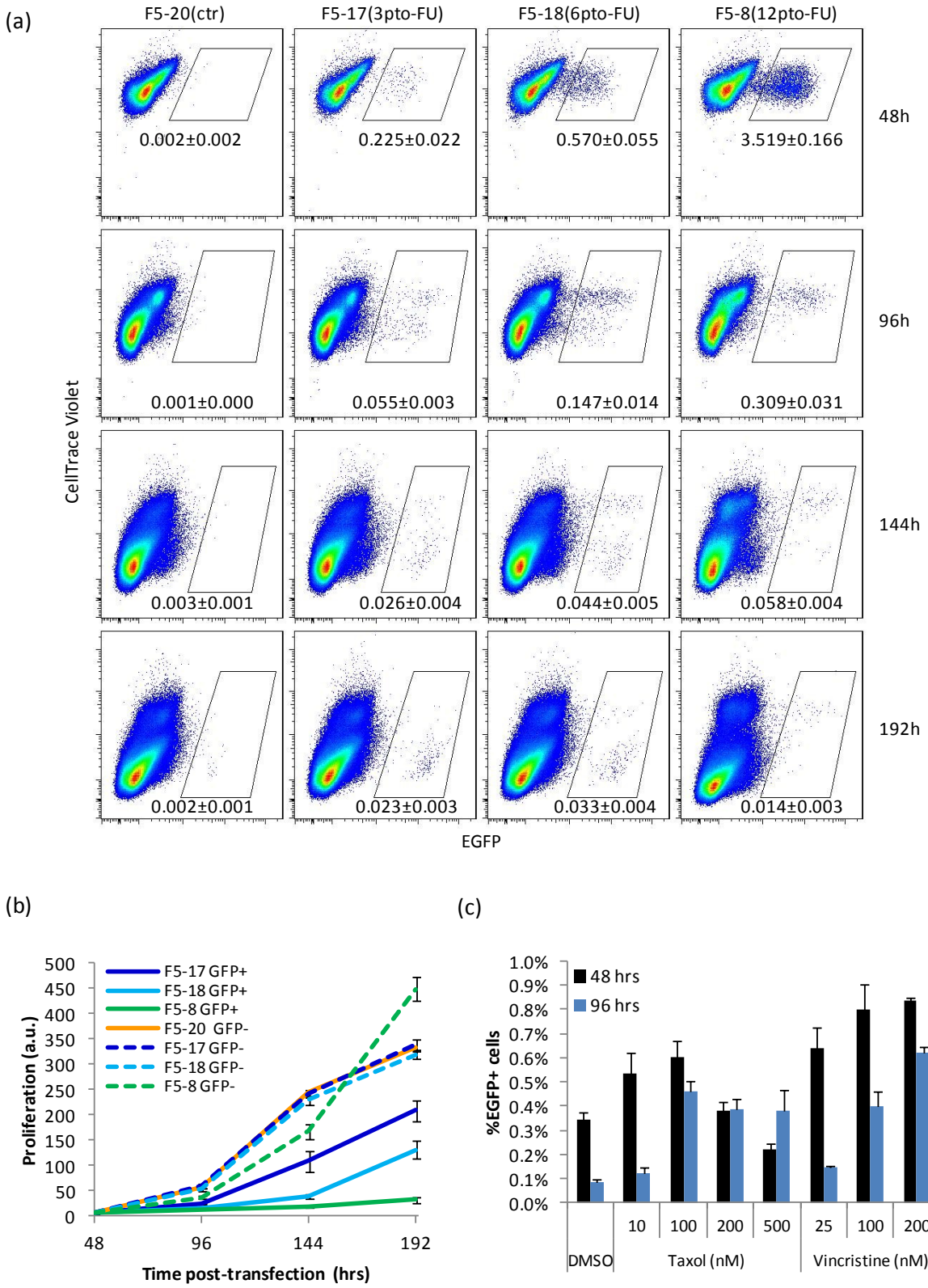


Figure 2.3 Varying PTOs in targeting oligos. All oligos in this figure are complementary to the non-transcribed strand's first potential start codon TTG, similar to F5-3 in Fig. 1. (a) Oligo toxicity and targeting efficiency as a function of PTO modifications. PTO bond position shown in gray. 'Total alive cells' was estimated as the total gated PI- cells in 200uL from a 24-well plate 48hrs post oligo transfection. (b) Correction efficiencies of new oligo designs with reduced PTO bonds and using FU modified base. n=4

Previous work has described the defect in oligo-corrected cells as a G2/M cell cycle arrest(Olsen, Randøl & Krauss 2005; Papaioannou et al. 2009; Olsen et al. 2009), with few cells moving past it. In order to generate a more detailed view of the effects of oligos on proliferation, I tracked cells using the CellTrace Violet dye(Efimova et al. 2003; Hawkins et al. 2007) (Figure 2.4a). This allowed following the relative proliferation rate of the corrected (EGFP+) and uncorrected (EGFP-) cell populations by flow cytometry as the inverse dilution rate of the CellTrace dye's Mean Fluorescent Intensity (MFI) (Figure 2.4b). First, we observe that in

Figure 2.4 Proliferation defect in corrected cells. Proliferation is inversely proportional to the MFI of the CellTrace Violet dye. Cells were treated with CellTrace Violet dye, then split and transfected with oligos 48hrs later. Cells were fixed at each time point and run together on a LSRfortessa cell analyzer at the end of the time course (a) Flow cytometry dot plots showing relative proliferation of EGFP+ and EGFP- cells generated with different targeting oligos, at four different time points post-transfection. %EGFP shown as average±s.d. n=6. (b) Proliferation index is calculated as the inverse MFI of CellTrace $x10^6$ for each population. (c) Cells treated with anti-microtubule drugs Taxol and Vincristine 24h after transfection with a 24h exposure time, checked at 48h and 96h post-transfection. n=4 technical replicates

Figure 2.4 (Continued)



general EGFP+ cells proliferate slower than corresponding EGFP- cells. Second, the proliferation of EGFP+ cells inversely correlates with the number of PTO bonds of the oligo used to generate them, with the *EGFP*+ cells from the F5-8 (12 PTO) oligo hardly proliferating as expected, and the F5-18(6 PTO) having an intermediate phenotype. The proliferation rate of corrected cells depended on the number of PTO bonds on the oligo, suggesting it is the correcting oligo itself and not the expression of *EGFP* the cause of their proliferation defect. To functionally validate this observation, I treated F5-17 transfected cells with anti-microtubule drugs, which are preferentially toxic to proliferating cells(Engstrom et al. 2009), 24hrs post-transfection (Fig. 2.6c). This produced an increase of EGFP+ cells at 48 and 96hrs, suggesting that non-corrected cells had been preferentially inhibited, and therefore that corrected cells experience delayed proliferation. To verify that the effects of the anti-microtubule drugs were not due to endosome destabilization leading to increased oligo release, cell were treated with chloroquine, which instead slightly decreased targeting frequencies (not shown).

In the case of the F5-17 oligo, the EGFP+ cells seemed to achieve a proliferation rate similar to the uncorrected *EGFP*- cells eight days post-transfection, since the percentage of positive cells remained constant between 144 and 192hrs (Figure. 2.4a) and when checked again a few weeks of passaging (not shown). By doubling the amount of lipofectamine:oligo complex I was able to double the frequency of corrected cells to ~0.05% eight days post-transfection (Supplementary Figure 2.3). I chose this time point to retry their single-cell sorting assay and found that EGFP+ cells generated with the F5-17 oligo easily generated clonal populations. These clones were verified by sequencing and found to have the desired genome modification (Supplementary Figure 2.4). To verify that this modification was not due just to spontaneous mutations, I repeated the experiment with oligos that introduced two adjacent nucleotide

changes, one of which would restore the ATG start codon of *mEGFP* (using either a T or FU base) and other of which would introduce either a second substitution or a single-base insertion (Figure 2.5). As expected, efficiency of incorporation of double-mutant oligos was significantly lower than for single-substitution oligos, with the double-substitution being more efficient than the single-substitution-and-insertion. Interestingly, using the FU base had an adverse effect in the double-substitution, suggesting a more complex mismatch recognition effect when more than one substitution is involved. Although the correction efficiencies were low, it was possible to sort EGFP+ cells generated with all four oligos, and all clones proliferated normally and were verified by sequencing to be correctly modified (Supplementary Figure 2.4b,c).

To further elucidate oligo design principles, I checked for long-term survival of EGFP+ cells while varying the number of PTO bonds 3' to the mismatch (Figure 2.6a). This showed 3-5PTO bonds as optimal, resulting in at least ~10-fold increase in targeting efficiency compared to unmodified oligos. I also varied the position of the PTO bonds, the strand polarity of the oligo used and the presence of modified bases (Figure 2.6b). When using an oligo complementary to the transcribed strand (F5-31, -32) I was able to detect some EGFP+ cells, but only slightly above the control oligo F5-20, suggesting the strand bias against the transcribed strand observed in this cell line is not just due to the additional DNA replication necessary to express the corrected gene when targeting this strand. Comparing the presence of 3PTO bonds at the 3' terminus (F5-30) vs. 3' internal to the mismatch (F5-17) produced a ~4-fold reduction

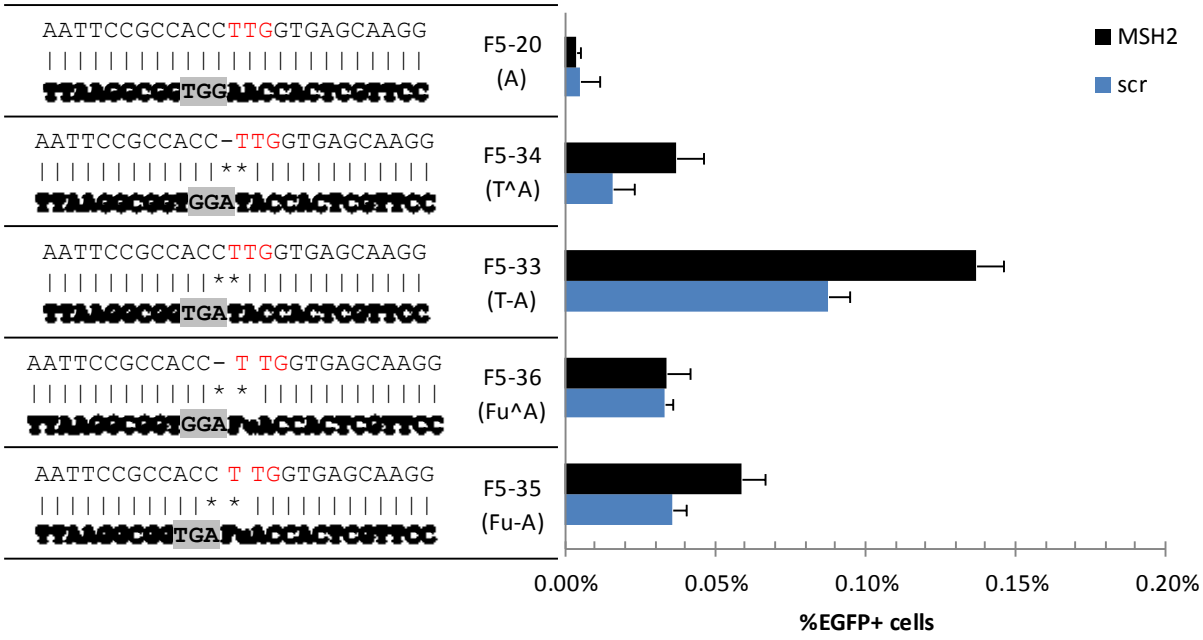


Figure 2.5. Generating double-mutants. Oligos complementary to the first start codon in the non-transcribed strand and containing a second mismatch, either an insertion (^A) or substitution (-A), with silenced MSH2 and assayed 36hrs after transfection. Targeting oligos shown in duplex 3-'to 5'-, with PTO highlighted in gray and mismatches shown with an asterisk. F5-20 was used as non-correcting control. n=4 technical replicates. Fu: 2'Fluorouracil.

in EGFP+ cells, confirming that internal PTO modifications result in higher stable targeting frequencies. Comparing the natural (F5-29) and modified (F5-17) base oligos resulted in a similar 2-fold increase for the modified base as when assayed 48hrs after transfection. Thus, optimizing the oligo design with modified base analogs and reducing the number of PTO bonds enabled the generation of stable, isogenic populations genetically modified with oligos.

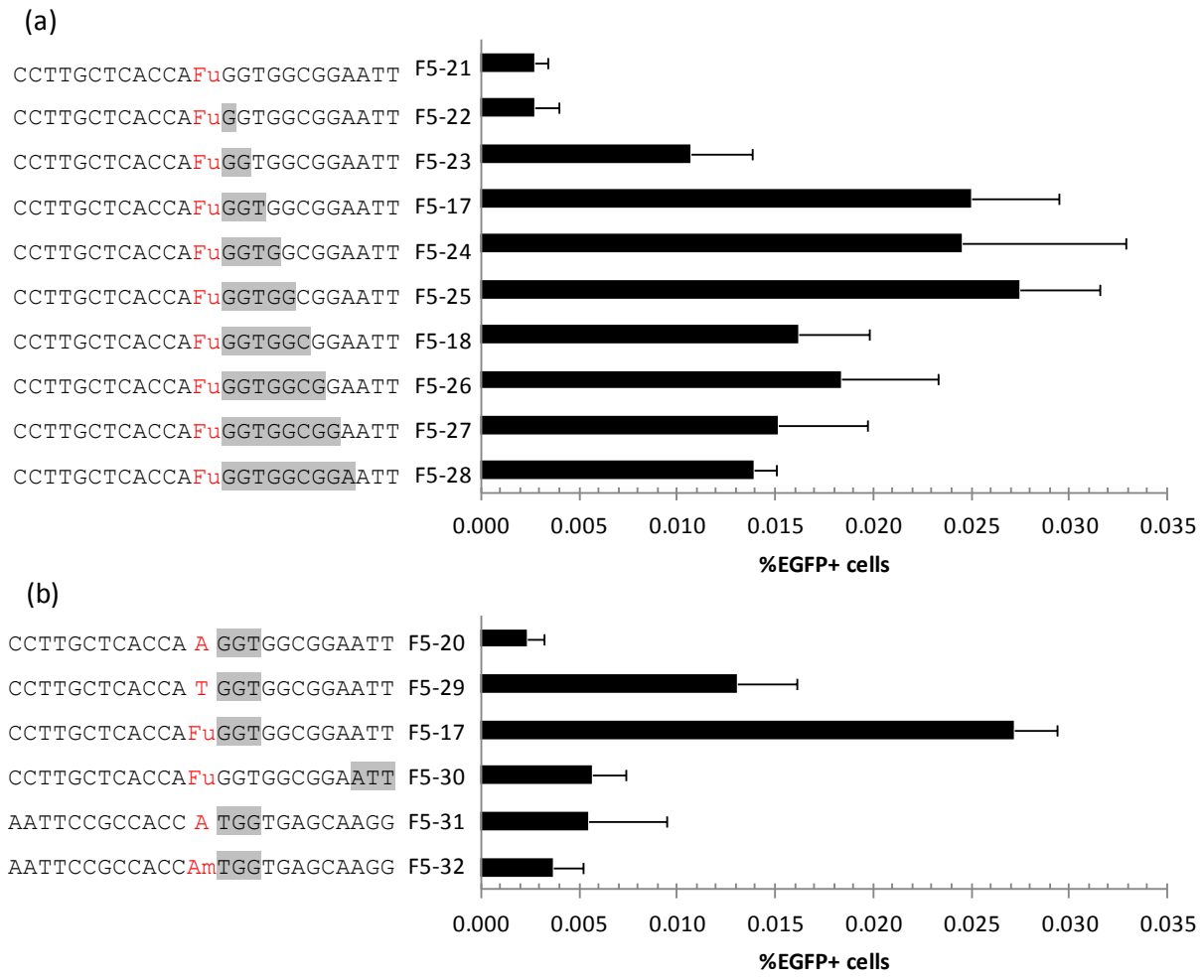


Figure 2.6. Effects of PTO in long-term survival of corrected cells. (a) Varying numbers of PTO bonds 3' to the mismatch, shown in gray, suggest 3-5 PTO bonds as optimal for balancing toxicity and stable targeting frequencies, assayed 14-days after oligo transfection. (b) Testing the position of PTO bonds, strand targeted and use of modified bases in long term survival. F5-20 used as non correcting control, F5-17, -29 and -30 are complementary to the non-transcribed strand, while F5-31, -32 are complementary to the transcribed. n=4 technical replicates

The proliferation defect in corrected cells may be associated to a cellular immune response to transfected oligos

These results suggest that the PTO modification in the oligos is the main reason corrected cells proliferate slower, rather than the mismatch generated by the oligo incorporation. Currently, the principal mechanism associated with PTO-modified oligo toxicity is the generation of DNA

double-strand breaks. To check this I measured the levels of γH2AX phosphorylation in oligo transfected cells by flow cytometry (Supplementary Figure 2.5), and find that reducing the number of PTO bonds reduced the levels of γH2AX phosphorylation. Interestingly, this also showed that EGFP+ cells had lower levels of γH2AX phosphorylation when compared with the non-corrected EGFP- cells treated with the same oligo. This suggests that although γH2AX phosphorylation may be a general marker of oligo toxicity, DNA double-strand breaks might not be the cause of decreased proliferation of oligo-modified GFP+ cells, or at least not the only cause.

To further explore the proliferation defect in corrected cells, I sorted EGFP+ and EGFP- cells 36hrs after transfection with the F5-17 oligo and performed RNA-seq to compare the two populations. This revealed a surprisingly high number of modestly differentially regulated genes (400 genes, $p<10^{-70}$) between corrected and uncorrected cells (Appendix 1). We analyzed this top 400 gene list with ToppFun(http://toppgene.cchmc.org) to look for enrichment on any Gene Ontology categories. Interestingly 'viral reproduction' and 'viral infectious cycle' were among the most highly represented (Table 2.4).

Table 2.4 Enriched GO categories. Top 400 genes based on p-value were analyzed with Toppgene, showing strong representation of viral process category.

GO ID	Name	P-value
GO:0006415	translational termination	5.45E-33
GO:0016032	**viral reproduction**	1.26E-30
GO:0019058	**viral infectious cycle**	1.43E-30
GO:0022415	**viral reproductive process**	2.85E-30
GO:0006414	translational elongation	2.98E-30
GO:0034623	cellular macromolecular complex disassembly	8.34E-30
GO:0019083	**viral transcription**	9.65E-30
GO:0019080	**viral genome expression**	9.65E-30
GO:0032984	macromolecular complex disassembly	2.01E-29
GO:0031018	endocrine pancreas development	7.36E-28
GO:0071845	cellular component disassembly at cellular level	1.56E-27

Table 2.4 (continued)

GO:0022411	cellular component disassembly	2.92E-27
GO:0043624	cellular protein complex disassembly	5.63E-27
GO:0043241	protein complex disassembly	1.35E-26
GO:0006412	translation	2.06E-26
GO:0031016	pancreas development	4.57E-25
GO:0034621	cellular macromolecular complex subunit organization	2.81E-21
GO:0035270	endocrine system development	3.43E-21
GO:0048610	cellular process involved in reproduction	1.10E-18
GO:0071822	protein complex subunit organization	5.58E-16
GO:0043933	macromolecular complex subunit organization	1.50E-15
GO:0022414	reproductive process	3.49E-08
GO:0000003	reproduction	4.03E-08
GO:0006396	RNA processing	3.82E-07
GO:0016071	mRNA metabolic process	5.80E-05
GO:0007093	mitotic cell cycle checkpoint	1.44E-03
GO:0008380	RNA splicing	2.34E-03
GO:0042274	ribosomal small subunit biogenesis	2.93E-03
GO:0012501	programmed cell death	7.88E-03
GO:0006364	rRNA processing	1.01E-02
GO:0006260	DNA replication	1.14E-02
GO:0002576	platelet degranulation	1.25E-02
GO:0016072	rRNA metabolic process	1.39E-02
GO:0031575	mitotic cell cycle G1/S transition checkpoint	1.50E-02
GO:0071779	G1/S transition checkpoint	1.50E-02
GO:0006090	pyruvate metabolic process	1.50E-02
GO:0046034	ATP metabolic process	1.82E-02
GO:2000045	regulation of G1/S transition of mitotic cell cycle	1.97E-02
GO:0042273	ribosomal large subunit biogenesis	2.02E-02
GO:0072431	signal transduction involved in mitotic cell cycle G1/S transition DNA damage checkpoint	2.14E-02
GO:0072395	signal transduction involved in cell cycle checkpoint	2.14E-02
GO:0072413	signal transduction involved in mitotic cell cycle checkpoint	2.14E-02
GO:0072404	signal transduction involved in G1/S transition checkpoint	2.14E-02
GO:0072474	signal transduction involved in mitotic cell cycle G1/S checkpoint	2.14E-02
GO:0006977	DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest	2.14E-02
GO:0072422	signal transduction involved in DNA damage checkpoint	2.14E-02
GO:0072401	signal transduction involved in DNA integrity checkpoint	2.14E-02
GO:0006096	glycolysis	2.38E-02
GO:0000075	cell cycle checkpoint	3.43E-02

Table 2.4 (continued)

GO:0071158	positive regulation of cell cycle arrest	3.59E-02
GO:0006457	protein folding	3.59E-02
GO:0006915	apoptosis	3.77E-02
GO:0006397	mRNA processing	4.25E-02
GO:0031571	mitotic cell cycle G1/S transition DNA damage checkpoint	4.79E-02

To further explore this observation, we choose three highly differentially expressed genes known to be involved in antiviral/cellular immune responses: *HLAB*, *IL32* and *OAS3*. I validated these genes by RT-qPCR, which allowed us to normalize the expression levels of corrected and uncorrected cells to untransfected cells (Figure 2.7a). This revealed much higher induction levels for both populations relative to untransfected cells. The difference between transfected and untransfected cells was much higher (8-30 fold) than between EGFP+ and EGFP- cells (2-4 fold), suggesting that both non-corrected and corrected cells share an 'oligo-transfected' transcription profile. To verify that the gene changes were not due to EGFP itself, I checked the expression levels of EGFP+ and EGFP- cells after transfection with the F5-18 oligo. Like F5-17, F5-18 also generates EGFP+ cells, but F5-18 is a more toxic oligo as it contains three additional PTO bonds. If the upregulation of the three immune response genes observed were due to the EGFP protein and not oligo toxicity, EGFP+ or EGFP- cells should have similar expression of these genes whether they had been transfected with F5-17 or F5-18. In contrast, however, the F5-18 transfected cells showed proportionally higher levels of induction of the three genes relative to F5-17, with EGFP+ cells still ~2-4 fold higher than EGFP- cells. One possible explanation for the difference between corrected and uncorrected cells would be that corrected cells got more oligos during transfection. I tested this by co-transfecting a targeting oligo along with a Cy5-labled one (F5-37). This showed that EGFP+ cells consistently had 2-5 times higher levels of

oligos at varying oligo concentrations (Figure 2.7b), suggesting that corrected cells may be proliferating less due to relatively higher oligo concentrations inducing a stronger immune response. Oligos with higher number of PTO modification have increased half-lives, which might lead to higher immune stimulation.

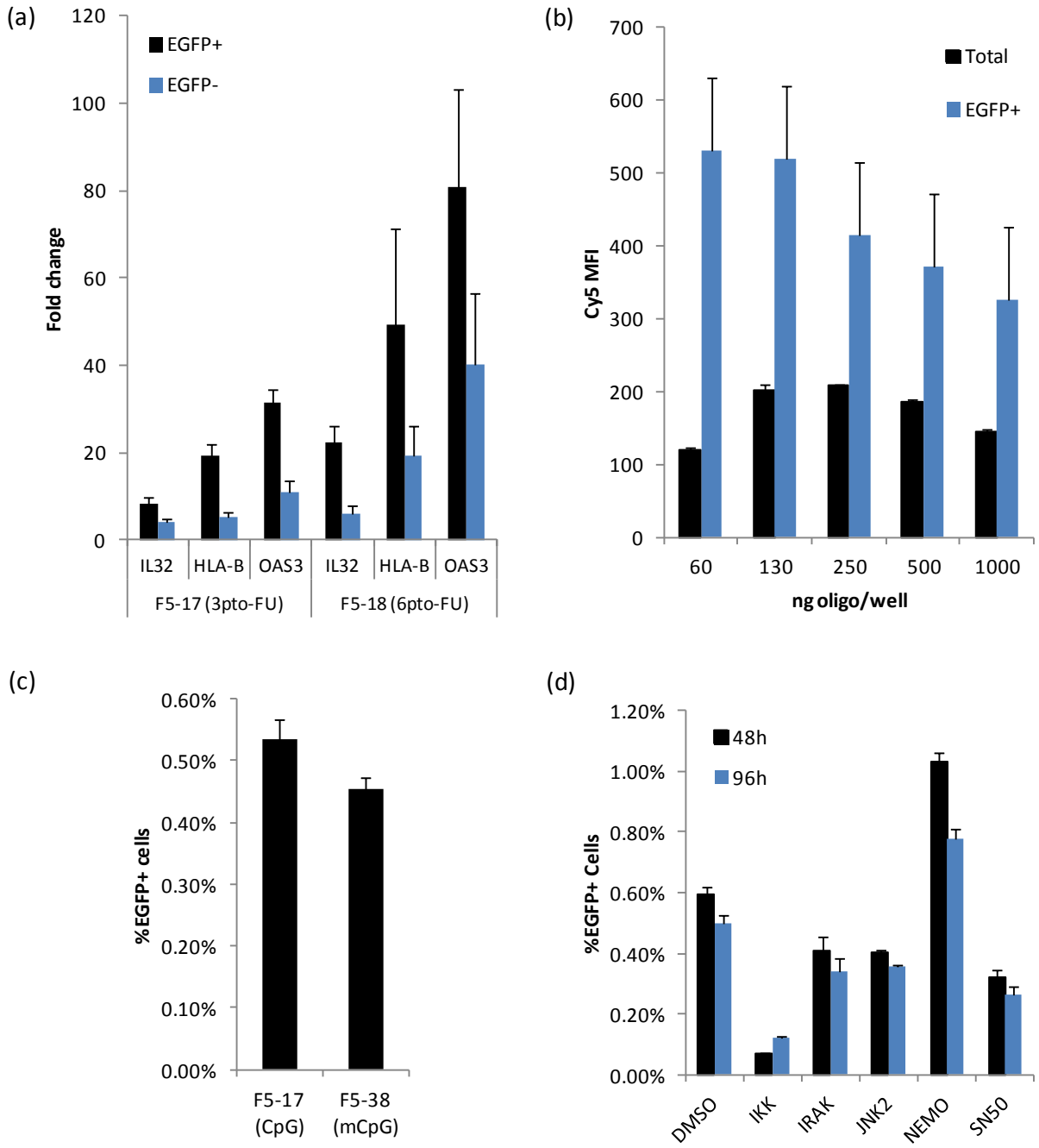


Figure 2.7. Induction of immune-related genes in oligo-transfected cells. (a)RT-qPCR relative quantification was performed for key immune-related genes, normalizing oligo-transfected to untransfected cells. (b) Mean Fluorescence Intensity (MFI) of total and EGFP+cells transfected with both F5-17 and F5-38 (Cy5-labeled) oligo at varying concentrations. Transfections were done in 24-well plates. n=4 (c) Methylation of the CpG sequence present in the targeting oligo has no effect on the %EGFP+ cells. (d) Small-molecule inhibitors against key immune effector proteins, with DMSO as control, added 24 h after F5-17 oligo transfection assayed for EGFP+ cells at 48h and 96h. n=4 technical replicates

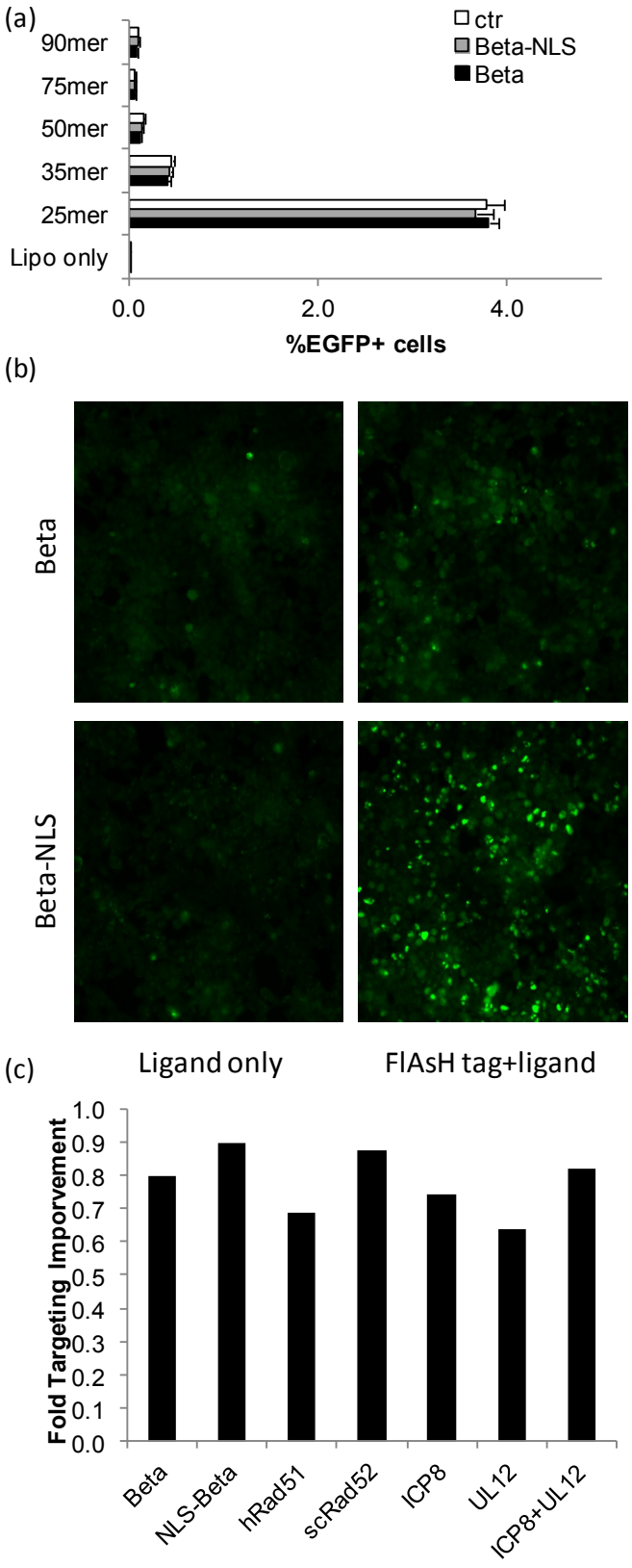
One possible cause for the immune response to the transfected oligos would be the presence of an unmethylated CpG in the oligo sequence(Table 2.1). I tested an oligo with a methyl-C modification (F5-38), but this had no effect on the percentage of EGFP+ cells after oligo transfection (Figure 2.7c). This is not unexpected however, since non-CpG-containing oligos have also been found to activate inflammatory responses(Senn et al. 2005). Additionally, I tested small-molecule inhibitors targeting key immune signaling pathways, out of which NEMO inhibition produced a ~2-fold increase in the percentage of EGFP+ cells (Figure 2.7d). This increase, however, did not translate to higher survival eight days post-transfection(not shown). Further work is needed to determine the specific immune signaling pathway that might be induced in oligo-mediated targeting. Inhibiting such pathway may reduce the toxicity of PTO-modified oligos, leading to further improvements in targeting efficiencies.

Beta and related proteins do not improve oligo recombination in human cells

In *E. coli,* the SSAP Beta increases oligo recombination rates by four orders of magnitude(Ellis et al. 2001). I used the reporter system validated here to check if expressing Beta in human cells would have any effect in oligo-recombination frequencies. Even when trying different oligo sizes and verifying expression and nuclear localization of the protein we were unable to detect any effects in oligo recombination frequencies (Figure 2.8a,b). I expanded the search and tested related proteins, but none was found to significantly affect recombination frequencies (Figure 2.8c).

Figure 2.8. Testing Beta in human cells. (a) Oligo recombination frequencies in cells expressing Beta or Beta-NLS, transfected with targeting oligos of varying length. Ctr=empty plasmid. (b) Protein expression and cellular localization. Cells were transfected with Beta or a Beta with a nuclear localization signal. Beta protein had an added FlAsH N-terminal peptide to detect expression and cellular localization. (c) Fold effects of different overexpressed proteins in oligo recombination frequencies.

Figure 2.8 (Continued)



DISCUSSION

In this work I describe our progress towards oligo-mediated genome engineering in human cells. An efficient version of this technology would be highly desirable since it does not require the site-specific nucleases or involve double-strand breaks, simplifying experimental design and reducing time and costs. Currently, however, oligo-mediated genome engineering in human cells is severely limited by both low oligo incorporation efficiency and the low survival of modified cells. The nature of this limitation is likely multifactorial, with DNA damage being the most studied so far (Olsen et al. 2009). Recent evidence in bacterial and mammalian systems suggests a model of oligo incorporation during DNA replication. Thus, when the oligo carrying the desired genetic change gets incorporated, it can be detected as a DNA replication error by the MMR machinery. The cause of low survival of corrected cells is unknown, but is at least partly caused by the cytotoxicity of the oligos.

To reduce the impact of MMR on oligo-mediated recombinations, I tested if using chemically modified base analogs capable of avoiding MMR recognition in *E. coli* would have a similar effect in human cells. As in *E. coli*, I found that FU and AM outperform thymine and adenine, respectively, in creating T-T or A-A mismatches leading to A/T transversions. By transiently silencing *MSH2* and *MLH1* I found that, at least for FU, the increased targeting efficiency of these modified bases is in part due to avoidance of the MMR machinery. This effect, however, seem to be complex since adding a second mismatch produced variable results, likely due to sequence context effects(Mazurek et al. 2009). Because our reporter assay is limited to restoring the start codon of *mEGFP*, I was only able to study modified bases substituting A and T, but I expect that other modified bases found to work in *E. coli* will have a similar effect.

Further work is required to fully elucidate the mechanism by which modified bases increase targeting frequencies in human cells. These include testing the effects of a complete MMR knockout and the role of other DNA repair pathways such as base excision and nucleotide excision repair on different cell types. An advantage of using modified bases rather than suppressing the endogenous MMR machinery to increase oligo incorporation efficiency is the avoidance of unwanted mutations elsewhere in the genome that MMR suppression is likely to lead to(Dekker et al. 2011), an essential feature for functional genomics applications.

Using FU for MMR-avoidance in an oligo containing 6PTO bonds on each end lead to a substantial number (~3-5%) of EGFP+ cells shortly after transfection, however this EGFP+ population did not proliferate further. This suggests that a process other than MMR-mediated DNA damage signaling may be causing the low survival of corrected cells. Since PTO bonds have been previously implicated in oligo cytotoxicity, I tested varying the number and position of PTO bonds. There was a clear trend where a higher the number of PTO bonds lowers the total number of cells surviving after transfection. Additionally, protecting only the 3' end yielded a higher percentage of corrected cells than a similarly protected and similarly toxic 5' end. Based on this, I designed two new oligos with a reduced number of PTO bonds 3' to the mismatch and used these to examine cell proliferation after transfection in more detail. Using a cell tracking dye I found that the proliferation of corrected cells was consistently lower than that of uncorrected ones, and that the proliferation of corrected cells was increased when the targeting oligo contained a lower number of PTO bonds. Oligo F5-17 oligo allowed isolation of stably modified EGFP+ cells with the correct genotype. This oligo's three PTO bonds are situated immediately 3' to the mismatch base and are therefore internal to the oligo. The oligo would be reduced to only 16 bases long by 3'-to-5' exonucleases, which might make annealing inefficient.

We speculate instead that internal-3' protection might prevent removal of oligos by proofreading polymerases after annealing while still allowing sufficient clearance of excess oligo by autonomous exo- and endonucleases. Supporting this, F5-30, a similar oligo but with the 3 PTO bonds at the 3' terminus resulted in a ~4-fold decrease in survival compared to F5-17.

To try to elucidate the cause of oligo toxicity in relation to PTO bond number I measured the levels of γH2AX phosphorylation. I found that the number of PTO bonds directly correlates to the levels of γH2AX phosphorylation, but EGFP+ cells have consistently lower levels than EGFP- cells. The fact that HeLa cells have abnormal p53 signaling might account for the observation that EGFP+ cells have lower levels of γH2AX phosphorylation. This would be the case, for example, if the observed double-strand breaks are general marker of oligo toxicity inducing apoptosis rather than directly causing double-strand breaks. Future investigation into the mechanisms of PTO toxicity and targeting oligo protection would be desirable.

Cell proliferation experiments showed that modified cells are not an outlier but rather a subset of the population. To get a better understanding of this growth phenotype I did RNA-seq comparing the transcription profiles between corrected and uncorrected cells. This revealed a large number of mildly differentially expressed genes between the two populations, especially genes involved in immune/viral processes, based on GO category enrichment. Interestingly, DNA damage-associated GO categories had a much less significant p-value ($\sim 10^{-2}$ vs. $\sim 10^{-30}$), further arguing against DNA damage causing the low survival of corrected cells. I validated the differential expression of *HLAB*, *OAS3* and *IL32* to use as immune response markers by RT-qPCR, which in addition allowed us to compare expression levels to those of untransfected cells and of cells treated with different targeting oligo designs. This uncovered two important clues: first, both corrected and uncorrected cells showed a much higher fold induction of immune genes

than untransfected cells and second, more oligo PTO modifications led to proportionally higher expression levels. Since cells that get more oligos during transfection are presumably more likely to become EGFP+, that higher oligo concentrations within corrected cells may be inducing a stronger cellular immune response. A similar effect may also explain the toxicity of PTO modifications, since PTO modifications stabilize oligos and will therefore lead to higher sustained oligo concentrations after transfection, which in turn may lead to higher immune signaling. Alternatively, it is possible that cellular DNA sensors might recognize the abnormal PTO backbone itself(Burckstummer et al. 2009). The model of oligo-mediated immune response seems plausible, especially in light of the growing list of cellular immune DNA sensors (Burckstummer et al. 2009; Zhang et al. 2011; Barber 2011) and the role of 3' exonucleases in inflammatory diseases(Coscoy & Raulet 2007). It is possible that the immune response triggered in oligo corrected cells might affect cellular phenotype in such a way that it confounds the effects of the targeted mutation itself. This could be assayed by comparing transcriptional profiles of modified clonal populations with the unmodified 'parental' lines. If necessary and depending on the intended study, mutations of interest could be evaluated in parallel with silent mutations to control for the modification process. Further work is necessary to better understand the roles of DNA double-strand breaks and cellular immunity in oligo-mediated toxicity, including testing these observations in different cell lines.

The wealth of sequencing projects will necessitate novel methods for experimentally testing human genetic variants. For modeling a few high-interest single-loci isogenic changes by targeted manipulation of human cell genomes, the current state-of the-art method of targeted nucleases may be sufficient. However, screening of hundreds of polymorphisms and their potential combinatorial interactions for phenotypic effects is not currently feasible with this

method. An oligo-mediated genome engineering method as used in *E. coli*(Wang et al. 2009) is far better suited for this, where up to ten simultaneous oligo incorporations are possible (Carr et al. 2012) and experimental design is trivial. We estimate our current targeting frequency to be ~0.05% for HeLa cells. This level of targeting efficiency should easily allow the generation of isogenic cells by using a pooled genotyping serial enrichment strategy(Aarts et al. 2009; Dekker et al. 2011). The cellular immune response to transfected oligos described here represents a novel and exciting avenue which might lead to further reduction of toxicity and thus improved targeting efficiencies in multiple cell types. Development of oligo-mediated genome engineering in human induced pluripotent stem cells could enable combinatorial testing of genetic variants in multiple tissues simultaneously, greatly expediting the study of human genetic variation and the genetic basis of human disease.

Here I describe a validated reporter system used to study and improve oligo recombination frequencies in human cells. I used this reporter to test if the *E. coli* SSAP beta would enhance recombination rates. However I found the protein not to be active in the HeLa-F5 cell line, even after verifying expression and nuclear localization. In the next chapter I describe my efforts trying to elucidate how this protein works in *E. coli,* with the goal of eventually adapting its activity to human cells.

MATERIALS AND METHODS

Cell culture and transfections

The HeLa F5 cell line was kindly provided by Dr. Depei Liu. These were cultured in DMEM with GlutaMAX and HEPES, supplemented with 10% HI-FBS, 100 mg/ml streptomycin

and 100 U/ml penicillin (Invitrogen) at 37°C and 5% CO_2 in a humidified incubator. Cells densities were determined using the Countess Automated Cell Counter (Invitrogen). For subculturing, cells were washed with PBS pH 7.4, trypsinized with TrypLE Express(Invitrogen) and neutralized with 10% HI-FBS DMEM.

Oligo transfections

HPLC-purified oligos were ordered from Integrated DNA Technologies, resuspended in distilled water and their concentrations verified by NanoDrop (Thermo Scientific). Cells were pre-seeded at $3x10^6$ cells per 10cm plate two days before transfection. Pre-seeded cells were split to 100,000 cells/well for a 24-well plate or at $3x10^6$ cells for a 10cm plate. The next day, cells were washed with PBS and 10% HI-FBS with no antibiotics was added. Transfections were done with Lipofectamine 2000 (Invitrogen) using a 3:1 ratio of µL lipofectamine to µg oligo. For 24-well plates, 0.5µg of oligo and 1.5µL lipofectamine were diluted in 25µL OptiMEM(Invitrogen) each, incubated at room temperature for 5 mins, combined by pipetting, then incubated again for 20 mins. The oligo complexes were added to cells dropwise and cultured continuously for 36-48hrs. Unless otherwise noted, 20µM thymidine was added to the cells two hours prior transfection.

Flow cytometry analysis

After 36hrs, oligo-transfected cells were washed with PBS, trypsinized, neutralized with 10% HI-FBS DMEM and transferred to 96-well U-bottom plate. Death exclusion was done with 10% propidium iodide (Roche), and total surviving cells were estimated as the total number of living cells counted at a determined volume. The plate was run on a five-laser BD LSRFortessa HTS with FACS Diva 6.1 software. EGFP was detected with the 488nm laser, 530/30 nm filter; CellTrace Violet with the 405nm laser, 450/50 nm filter; propidium iodide with the 561nm,

610/20 nm filter; Cy5 with 640nm laser, 670/30 nm filter. Cell sorting was performed on a BD FACSAria II SORP equipped with 375nm, 405nm, 445 nm, 488nm, 561 nm, and 633nm lasers.

Genotyping and sequencing

We used allele-specific realtime PCR to quantify corrected versus uncorrected *mEGFP* DNA in cells after oligo targeting. A 152bp amplicon was amplified using a forward primer whose 3' terminal base matched either the uncorrected or corrected *mEGFP* target site. Using standard curves for each primer pair, a ratio of corrected to non-corrected template strands could be estimated for each cell population. EGFP+ and EGFP- cells were sorted by FACS and pelleted. To lyse the cells, ~1,500 cells resuspended in 8.9µl water were added to 1µl 10X gold buffer and 0.1µL prepGEM enzyme (ZyGEM) and incubated for 75^{o}C for 5 min then 95^{o}C for 5 min. The reaction was split into 2 x 5µL and each aliquot was mixed with 20 µL PCR mix containing the corrected-specific or non-corrected-specific primer pair. PCR mix: 18.8µL 1.1X Platinum Taq Supermix (Invitrogen), 0.5µL each primer (stock solution 10uM), 0.2µL SYBR green I (stock solution a 1:500 dilution of original tube). Realtime PCR was carried out on an Opticon 2 DNA Engine (MJ Research), under the following cycling conditions: 95^{o}C for 3 min followed by 40 cycles of 95^{o}C 30 sec, 62^{o}C 30 sec, 72^{o}C 30 sec. To sequence the *mEGFP* target site in stable non-corrected and corrected cell clones, a similar PCR protocol was used except with primers N1 and N3 from (Yin et al. 2005) that span the targeted region. PCR products were sequenced by Genewiz, inc. All PCR primer sequences used can be found in Supplementary Table 1.

RNA interference

HeLa F5 cells were seeded 3x10^{6} cells per 10cm plate. The next day, media was changed to 4mL 10% HI-FBS with no antibiotic. Transfection of validated shRNA plasmids (Sigma-

Aldrich) or scramble control (Addgene) was done with FuGENE HD, diluting 15µg of HiSpeed Maxi-Prep(QIAGEN) purified plasmid and 60µL of FuGENE HD(Roche) in 500µL OptiMEM each. The dilutions were mixed by pipetting and incubated for 15mins before adding complexes dropwise to plates. After overnight transfection, cells were washed, trypsinized and re-seeded in 10cm plates with 10% HI-FBS DMEM and 3µg/µL puromycin (Sigma-Aldrich) After two days of puromycin selection, cells were split into 24-well plates 100,000 cells/well for oligo transfections. shRNA target sequences can be found in Supplementary Table 2.

Cell Proliferation assay

HeLa F5 cells seeded at $3x10^6$ cell per 10cm plate were washed and incubated in 5mL PBS with 10µM CellTrace Violet(Invitrogen) for 10 mins at 37^{o}C, washed and grown in 10% HI-FBS DMEM overnight. Cells were split and pre-seeded at $3x10^6$ cell per 10cm plate, then the next day seeded in 24-well plates for oligo transfection. At each timepoint, cells were trypsinzed, centrifuged at 500g for 5 mins and fixed in 4% PFA (Biolegend) for 20 min in the dark at room temperature. After another centrifugation, cells were resuspended in Cell Staining Buffer (Biolegend) and kept at 4^{o}C. After the last timepoint samples were run together on a BD LSRFortessa HTS.

Anti-microtubule drug treatment

Twenty-four hours after cells were transfected with oligos, media was changed and cells were exposed to varying concentrations of either Taxol or Vincristine (EMD) for 24 hours, then washed and assayed with the BD LSRFortessa HTS immediately or two days later (96 hours after oligo transfection).

RNA-seq and qPCR

HeLa F5 cells plated at $3x10^6$ cells on a 10cm plate were washed and 4mL of OptiMEM with 20μM thymidine was added two hours prior transfection with 15μg oligo/45μL lipofectamine. Thirty-six hours after transfection cells were trypsinized, filtered through a 35μm cell strainer (BD Bioscience #352235), counted and centrifuged at 250g for 5 mins. Cells were resuspended at a density of $1x10^7$ cells/mL in 20% HI-FBS DMEM and sorted into EGFP+ and EGFP- directly into Trizol LS(Invitrogen). Extracted RNA was precipitated and concentrated on a RNAeasy Micro kit column (QIAGEN), and the quality of RNA was determined in an Agilent 2100 Bioanalyzer. Each sorting was done from 15 10-cm plates of transfected cells and the RNA of three independent sorting sessions was pooled to ~2μg of total RNA. RNA-seq was performed as described(Christodoulou et al. 2011) with some modifications, i.e. there was no normalization of the library. Reads were aligned using TOPHAT and reads mapping to each gene were analyzed for differential expression with a Bayesian statistic(Audic & Claverie 1997). For RT-qPCR, first-strand synthesis was done with the SuperScript III kit (Invitrogen), and qPCR with the KAPA SYBR Fast Universal 2X mix following manufacturer's protocol. The sequence of the IDT-synthesized qPCR oligos can be found in Supplementary Table 2. The reaction was assayed in a DNA Engine Opticon2 instrument (MJ Research).

Small molecule Immune inhibitors drug treatment

After cells were transfected with oligos for 24 hours, media was changed and cells were exposed 5μM IKK inhibitor VI, 50μM IRAK1/4 inhibitor, 20μM Cell-Permeable NEMO-Binding Domain Binding Peptide, NF-κB SN50, 36μM Cell-Permeable Inhibitor Peptide and 10μM JNK Inhibitor II (EMD) for 24 hours then assayed with the BD LSRFortessa HTS immediately or two days later (96 hours after oligo transfection).

CHAPTER THREE

CHARACTERIZATION OF THE C-TERMINAL DOMAIN OF THE SINGLE-STRAND ANNEALING PROTEIN RED BETA

Xavier Rios[1], Christopher J. Gregg[1], Marc J. Lajoe[1], and George M. Church[1,2]

[1]Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston 02115 USA
[2]Corresponding author

Xavier Rios designed and performed experiments and wrote manuscript. Chris Gregg designed and performed experiments, Marc Lajoe contributed ideas for the generation of the recombinase library, George M Church helped design experiments.

INTRODUCTION

Oligo-mediated recombination as a powerful genome engineering tool

In the previous chapter I showed that it is possible to generate genetically modified cells using ssDNA oligonucleotides. However, the recombination frequencies achieved were low, on the order of ~1/1000 cells. Thus making multiple changes, such as homozygous mutations, would require an ability to detect events happening with a frequency of 10^{-6}. In this chapter, I describe our work on characterizing the single biggest contributing factor for MAGE in *E. coli*: the SSAP Beta, enhancing recombination by four orders of magnitude(Ellis et al. 2001), with the goal of translating its activity to human cells. There are many similarities between oligo recombination in *E. coli* and human cells, suggesting a common recombination pathway. For example, both methods show strong dependence on DNA replication, both methods are inhibited by the endogenous mismatch repair machinery and both methods show strong strand bias.

One of the best example demonstrating the power of oligo-mediated recombination is MAGE, Multiplexed Automatable Genome Engineering (Wang et al. 2009). This powerful technology optimizes lambda Red recombineering in *E. coli* to dramatically improve oligonucleotide-mediated recombination. Recombineering, or lambda Red recombination-mediated genetic engineering, uses the lambda phage's recombination machinery normally expressed from the Red operon during the phage's lytic growth phase (Poteete 2001). The Red machinery promotes recombination between the bacterial chromosome and linear dsDNA molecules, such as PCR products introduced into the cell via electroporation, with far more efficiency than previously used methods for making gene replacements in *E. coli* (Court et al. 2002). The Red operon was discovered from lambda phage recombination defective mutants

isolated from *recA⁻ E. coli* strains that otherwise had WT recombination phenotypes. The mutant genes were initially described as Red α, β and γ. Red α, also known as lambda Exo, is a 5'→3' exonuclease, Red β is a single strand annealing protein (SSAP) and Red γ inhibits the RecBCD nuclease complex. The putative biological functions of the Red operon in the lambda phage may include (1) stimulation of lambda DNA replication, since DNA repair and replication systems are closely intertwined, (2) a DNA repair system against host restriction enzymes, (3) an alternative DNA repair pathway for when the lytic cycle is induced by DNA damage and (4) generating genetic diversity for phage evolution (Poteete 2001).

One of the most attractive features of oligo-mediated recombination is its inherent simplicity, since it is ultimately based on the fundamental property ssDNA has of being both an information carrier and an effector. Oligo recombination appears to be innate in many bacteria even in the absence of phage proteins: *P. syringae* has an native recombination rate of $2400/10^8$, while in *E. coli, S. flexneri* and *S. typhimurion* the native oligo recombination frequency is about 10-fold lower (Swingle, Markel, et al. 2010). Without a SSAP, just 20 nt of homology length nearly achieve maximal recombination frequencies. This is in contrast with Beta, which has been found to have an optimal homology length of 90 nt. In addition, they find that in innate recombination 15-19 bp oligos can be recombined, below the size limit of what Beta can. Further supporting the hypothesis of just ssDNA oligo effectors mediating recombination by annealing, the shorter the oligo is, the more significant its melting temperature becomes, with higher GC content leadings to higher frequencies(Swingle, Markel, et al. 2010).

The single-strand annealing protein Beta

Out of the three lambda Red proteins, the SSAP Beta is the only one required for oligo-mediated recombination, where expressing this protein alone increases recombination

frequencies by four orders of magnitude (Ellis et al. 2001). The Beta protein promotes the renaturation of complementary DNA strands (Kmiec & Holloman 1981). A single monomer binds a minimal ssDNA sequence 28-36bp long, with a K_D of 1.8 μM for a 36-mer. Adding 10 mM Mg^{+2} increased affinity to a ssDNA 36-mer, while 250 mM NaCl abolished it (Mythili et al. 1996). Renaturation of complementary DNA strands was most efficient when ssDNA was pre-annealed with Beta (Karakousis et al. 1998), and linear ssDNA is a more efficient substrate than circular ssDNA, suggesting a preference for annealing ssDNA ends.

Phylogenetic analysis suggests that SSAPs belong to one of three families: RecT/Beta, Rad52 and ERF, which seem to have evolved independently from DNA bacteriophages (Iyer et al. 2002). Rad52 is part of eukaryotic homologous recombination machinery (Mortensen et al. 1996, p.52). The ERF SSAP can also mediate oligo recombination, although with lower efficiency than RecT and Beta (Zhang et al. 2003). The RecT/Beta superfamily can be further subdivided into three groups: the Beta group is present in a large number of bacterial species, while the RecT family is predominantly seen in low-GC Gram-positive bacteria. The third group, EHAP1-like, is more divergent and only present in *E. coli* and *Salmonella*.

Single-strand annealing proteins can mediate strand exchange

RecT is an *E. coli* phage-derived SSAP functionally and structurally related to Beta. The presence of Mg^{2+} reduces RecT binding to both ss- and dsDNA, with dsDNA affected more (Noirot & Kolodner 1998). RecT's affinity to dsDNA is higher than for ssDNA with no Mg^{2+}, but slightly higher for ssDNA in the presence of it. Like Beta, high salt concentrations reduce DNA binding. In addition to strand annealing, RecT is known to catalyze limited strand exchange leading to the formation of D-loops, a reaction inhibited at Mg^{2+} concentrations above 5 mM, high RecT concentrations and high GC-content (Noirot et al. 2003). D-loops are formed

when ssDNA is pre-coated with RecT then added to dsDNA. This reaction is hypothesized to be driven by a stronger binding to the product of the oligo annealing reaction than to either ss- or dsDNA substrates (Hall & Kolodner 1994).

Beta is also capable of limited strand exchange and D-loop formation with low GC-content targets in a reaction inhibited by Mg^{2+} (Li et al. 1998; Rybalchenko et al. 2004). There is polarity to the strand exchange, with 3'-overhangs in dsDNA being more efficient than 5' ones, although the difference disappears with excess amounts of Beta-bound oligo. Similarly to RecT, the order of annealing matters, with ssDNA preannealed with Beta being more efficient (Karakousis et al. 1998). It has been suggested that proteins such as RecT and Beta condense DNA leading to accelerated renaturation (Sikorav & Church 1991). This property, possibly combined with more stable binding of RecT and Beta to the final annealed products, could explain the limited, ATP-independent strand exchange activity observed. However, it is unclear how physiologically relevant the strand displacement activity of these proteins is, especially since it is only efficient with sequences with very low GC-contents. In MAGE, oligos with $\Delta G <$ -12.5 kcal/mol have lower recombination frequencies, which the authors attribute to potential secondary structures (Wang et al. 2009). However, they show that oligo recombinations can take place within broad GC-content range. It is possible that sequences with GC-content low enough to allow strand exchange have abnormally high oligo recombination rates, although this remains to be tested.

Structure-function model for SSAPs

Currently, the three-dimensional atomic structure of Beta is not known. Electron micrographs revealed three quaternary states of the Beta protein (Passy et al. 1999) (Figure 3.1). In the absence of DNA but requiring Mg^{2+}, Beta formed a 12-unit small ring quaternary

structure. In the presence of a ssDNA 30mer and Mg^{2+} the ring structure had 15 units, while longer ssDNA formed 18-unit rings. The small rings were ~145Å in diameter, while the larger ones were 185-210Å . In the presence of dsDNA, long, left-handed helical filaments were seen. Circular dsDNA did not form large rings or filaments, while linear dsDNA did form filaments, a process enhanced with exposed ssDNA overhangs regardless of polarity. The helical filaments had the diameter of the larger rings. In the absence of Mg^{2+}, no rings were seen in the absence of DNA or with 30mers, and with longer 1.2 kb ssDNA the rings observed tended to be incomplete and more heterogeneous structures. Mg^{2+} was not required for dsDNA filament formation, but it enhanced it when denatured complementary ssDNA products were used. Based on these observations, the authors propose a model where the large rings bind to one DNA strand end and initiate its annealing to the complementary strand. Once initiated, the annealing can proceed spontaneously, followed by the formation of a ring-nucleated filament on the duplex DNA. Supporting their hypothesis, under reannealing conditions rings are the first Beta structure to form on the ssDNA overhangs or on the ends of linear dsDNA, followed by filament assembly as the renaturation progresses.

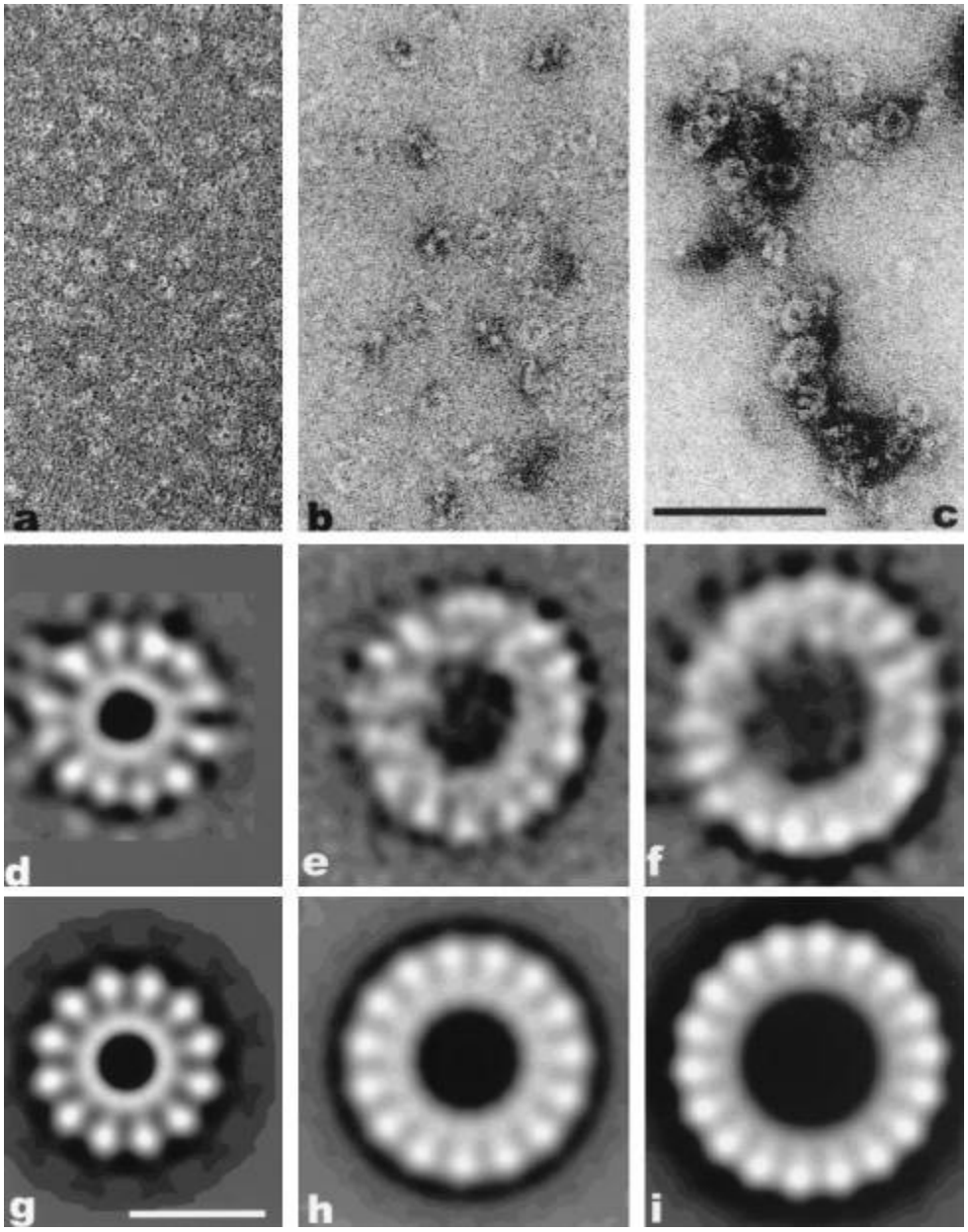


Figure 3.1. Small (left), medium (middle) and large (left) ring structures formed by Beta (Passy et al. 1999). Copyright © 1999, The National Academy of Sciences.

A recent study using atomic force microscopy proposed slightly different Beta quaternary structures(Erler et al. 2009). In this study a single monomer was found to bind ~11bp, and the minimum annealed complex was 16-20 bases consisting of two Beta monomers. In the absence

of DNA, Beta formed an 11-unit gapped ellipses, as opposed to a closed ring, and it seemed that the shallowness of the helix prevented the extension of additional Beta monomers. In the

presence of ssDNA (140mer) Beta formed heterogeneous and disordered complexes in monomeric form, suggesting the binding of ssDNA is driven by an increase in entropy. More specifically, the transition from gapped ellipses to disordered monomers induced by Beta's weak affinity to ssDNA may be explained by the intrinsic instability of a helical structure, as opposed to the ring structure proposed previously (Passy et al. 1999). With annealed dsDNA, Beta formed a left-handed helical filament, similar to the electron microscopy structures. Monomer spacing in the dsDNA helical structure was found to be essentially the same as the gapped ellipses in the absence of DNA. In this model, the left handedness of the helix might compensate for the torsion created by right-handed B-DNA helix after annealing, suggesting that DNA annealing drives the nucleoprotein filament formation (Figure 3.2).

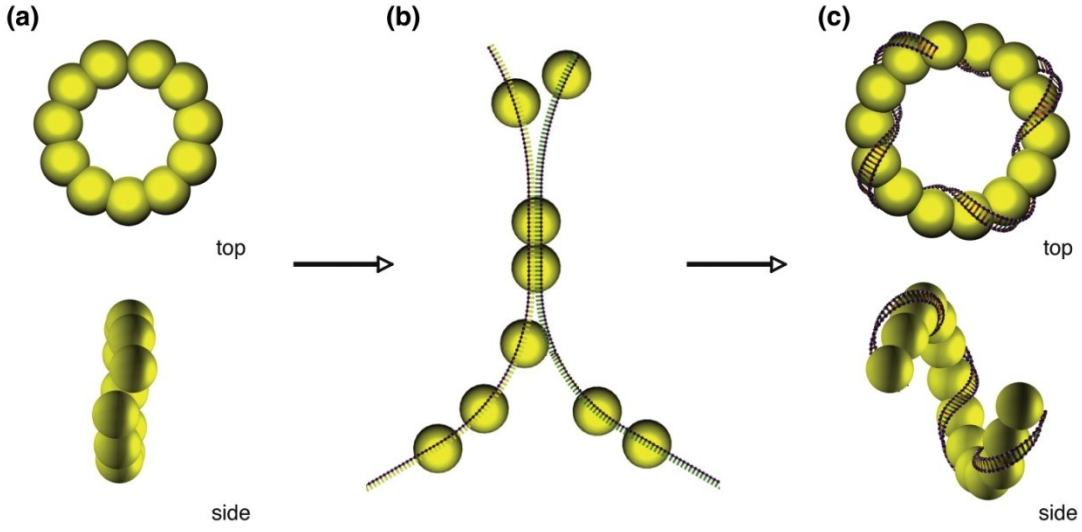


Figure 3.2 Atomic force microscopy-based model for Beta's annealing (Erler et al. 2009).

Both models agree on the final structure formed in the presence of annealed dsDNA being a left-handed helix. In addition, both have similar structures in the absence of any DNA: EM showed rings while AFM revealed gapped ellipses, both of ~11 monomers in length. The main difference between the models is the Beta-ssDNA intermediate. EM images show ring structures larger than the ones seen in the absence of DNA, while AFM shows disassembled monomers coating the DNA. The AFM study required some fixing for the Beta-only elliptical structure, while the rest were done in solution with no fixing. It is possible that the EM images are showing hyper-stabilized ring structures. This difference in structures led to two different mechanism for Beta-mediated annealing. A large ring annealing intermediate suggests a more prominent role for Beta in exposing the DNA bases for homology search, whereas the disordered monomer would support a condensation/crowding effect, although these two are not mutually exclusive. The ring model fails to explain how the helix is formed along annealed dsDNA, while the gapped ellipses model suggests that the helix is a direct extension of the ellipses, with inverted handedness induced by the right hand helix formed by B-DNA as it is annealed.

Electron microscopy data revealed that the presence of Mg^{2+} also stabilizes RecT filaments (Thresher et al. 1995). Combining this with the fact that Mg^{2+} also reduces DNA binding and D-loop formation, it has been proposed that the oligomeric structure might not be as proficient as the free monomeric RecT at DNA binding (Noirot & Kolodner 1998). Interestingly, the RecT/Beta protein family has two highly conserved C-terminal acidic residues, which may be involved in the coordination of Mg^{2+} (Iyer et al. 2002). This implies that the metal ion-dependent conformational switching is likely to be a general feature of the Beta/RecT protein family. Supporting this, the recently described mycobacteria phage SSAP gp61 showed a similar Mg^{+2}-dependent decreased affinity to dsDNA and ssDNA(van Kessel & Hatfull 2007).

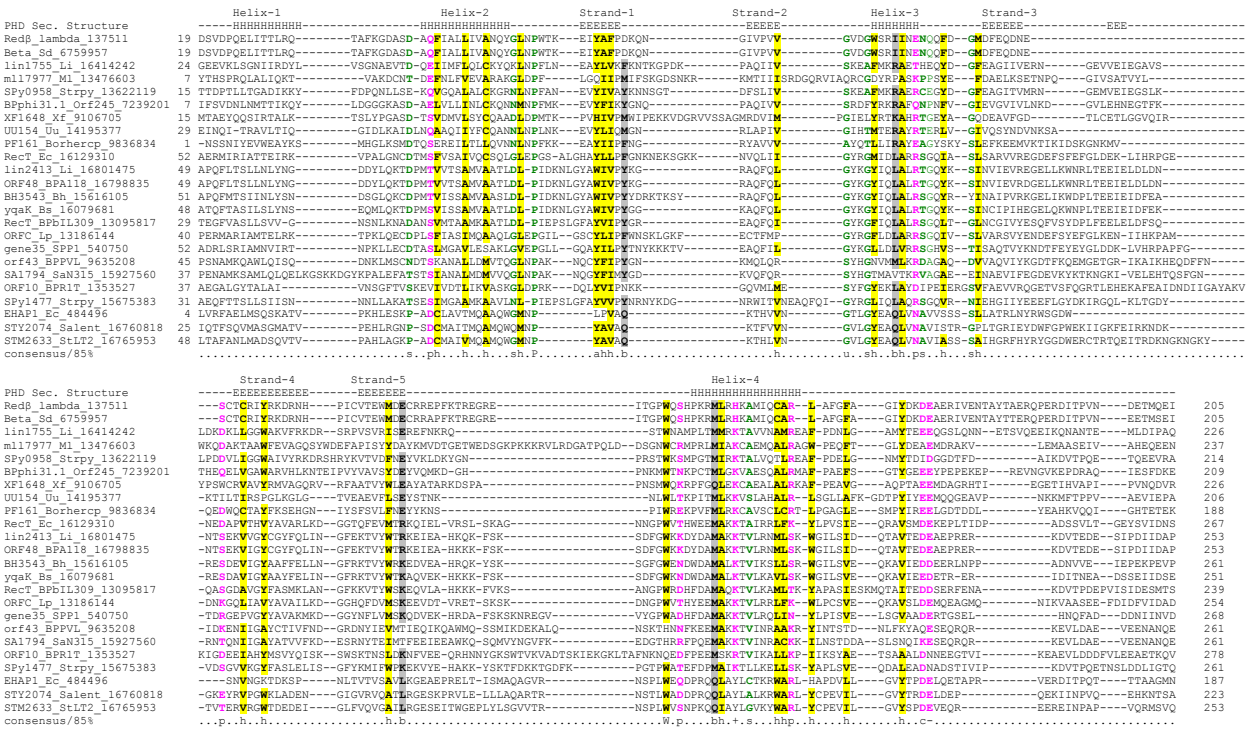


Figure 3.3. Multiple alignment of the RecT/Beta protein family of SSAP. © 2002 Iyer et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL. http://www.biomedcentral.com/1471-2164/3/8

Domain structure of Beta

Complementing the EM and AFM studies, limited proteolysis revealed further details on Beta's structure (Wu et al. 2006). In the presence of DNA, the N-terminus (1-177 aa) is a stable fragment involved in DNA binding, in the absence of DNA a 1-131 fragment is more stable, and with annealed dsDNA 1-177 and 1-230 stable fragments were generated. NHS-biotinylation bonding experiments revealed that many N-terminus lysines were involved in DNA binding, with a K172A mutation abolishing binding completely. The Beta 1-177 truncation was further characterized and based on its elution mass it still had an oligomeric structure of ~17 subunits.

This truncation bound ssDNA more strongly than full-length Beta, demonstrating that the C-terminal domain is not required for DNA binding, although it might be involved in controlling its affinity to ssDNA. In these experiments Mg^{+2} had no effect in the proteolysis of either Beta or Beta+33mer, suggesting it does not cause any major changes in the protein.

Mechanisms of Red-mediated recombination

Beta prefers to bind 3' single-strand overhangs on dsDNA, which is the product of Exo's 5'→3' exonuclease action (Li et al. 1998). This observation led to an elegant mechanism of DNA recombination and double-strand break repair for Beta/Exo mediated single-strand annealing (Figure 3.4). This single-strand annealing double-strand repair mechanism, however, may not explain the high frequencies observed in oligo-mediated recombination. Instead, emerging studies point towards a mechanism of oligo incorporation during DNA replication as a pseudo-Okazaki fragment (Court et al. 2002). One of the earliest pieces of evidence supporting this mechanism is the strand bias observed, where the oligo complementary to the lagging strand incorporates at significantly higher frequencies than the leading strand (Ellis et al. 2001). Additionally, inactivation of the mismatch repair protein MutS enhances oligo recombination up to 100-fold(Costantino & Court 2003). This suggest that at some point during oligo recombination a mismatched dsDNA is generated and recognized by the repair machinery. Similar results are seen in native oligo recombination in the absence of phage-derived SSAPs (Swingle, Markel, et al. 2010). As with Beta-mediated recombination, mismatch repair contributes about a 100-fold effect, and the lagging strand is preferred 10-fold, suggesting these are intrinsic features of the oligo recombination mechanism.

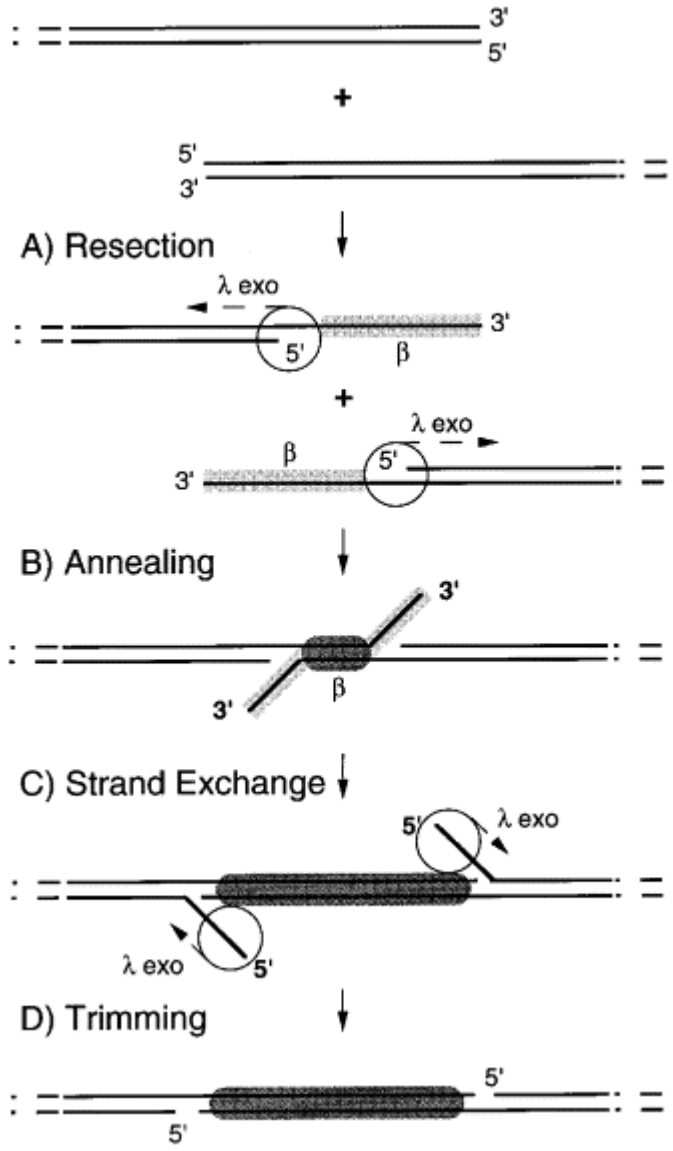


Figure 3.4. Beta/Exo single-strand annealing DNA repair (Li et al. 1998). In this model, two pieces of dsDNA are resected by Exo, leading to the loading of Beta to the exposed 3' overhangs. The Beta-ssDNA nucleoprotein filaments eventually find a complementary region to anneal, followed by strand exchange, displacing 5'- overhangs from the junction. These 5' overhang can then be trimmed by Exo, leaving nicks that can be sealed by DNA ligase. Copyright © 1998, Elsevier.

A replication-repressible plasmid system and oligos with modified dideoxynucleotide ends provided direct evidence suggesting that oligos incorporate(Huen et al. 2006). This study also found that β-clamp mutations enhanced oligo recombination rates. Further supporting a

direct role of DNA replication, mutations in DNA polymerases had significant effects in oligo recombination (Li et al. 2013). Examining mutation incorporation at the ends of the ssDNA oligo revealed that mutations at the 5' end are more likely to be lost. If oligos are incorporated during DNA replication, replicative polymerases with exonuclease activity are likely to affect the mutation loss on the homology arms of the oligos. DNA Pol I has 3'→5' proofreading and 5'→3' nick translation to remove RNA primers on the 5' end of Okazaki fragments. A mutation on DNA Pol I that removed the 3'→5' exonuclease and polymerase activities but kept 5'→3' exonuclease had no effects on overall oligo recombinations or the distribution of mutation loss relative to WT. Mutating the 5'→3' function reduced recombination frequencies 9-fold for lagging and 5-fold for the leading strand. Sequence loss at the 5'end of the lagging strand was completely eliminated, while mutation losses of the leading strand remained unchanged. Deleting the 3'-5' exonuclease activity of DNA Pol III reduced recombination of both strands 40-fold. In this mutant, 5' marker loss in the lagging strand oligo is increased while 3' loss is unaffected. Loss at the 5' end is increased in the leading strand oligo as well, perhaps from an increased role for Pol I. DNA Pol III, the main replicative polymerase, has 3'→5' proofreading and it is released from the lagging strand after it encounters an Okazaki fragment, leaving only a nick. DNA Pol I then removes the RNA primer by nick translation and the resulting nick is fixed by DNA ligase. Due to its low displacing ability, Pol III might be released after encountering an oligo annealed to the lagging strand. Thus, the 5'→3' exonuclease from Pol I is responsible for all mutations lost observed at the 5' end of the lagging strands. Furthermore, a temperature sensitive DNA ligase mutant reduced recombination frequencies 100-fold for the lagging-strand and 260-fold for the leading-strand oligos, implicating that ligase activity is necessary for oligo recombination. This would also explain why mutations at the 5' of the oligo reduce

recombination frequencies, since they would interfere with the ligation of the oligo, while the 3' end mutations can prime off further DNA synthesis before ligation is required.

Further highlighting the importance of lagging strand replication in oligo recombination, mutations in DnaJ primase increased consecutive incorporation of oligo pools (Lajoie et al. 2012). These mutations reduce primase interaction with DNA helicase and increase both the amount of ssDNA available and length of Okazaki fragments. The effect of these mutations in oligo recombination was magnified in a strain with exonucleases ExoI, ExoVII, ExoX, RecJ and lambda Exo were deleted. Deleting the nucleases increased oligo recombinations, but only at low oligo concentrations (Sawitzke et al. 2011). In experiments where multiple oligos are transformed at once, such as the primase experiment from above, each individual oligo concentration gets diluted and thus more likely to be affected by nucleases (Mosberg et al. 2012).

Finally, independent reports describe that knocking out endogenous DNA repair machinery genes such as *recA, recBCD, recORF, or ruvABC* had no effect on oligo recombination, further arguing against alternative recombination models(Huen et al. 2006; Sawitzke et al. 2011). Although there is data associating almost every component of DNA replication to oligo recombination, ultimately *in vitro* replication experiments are needed to conclusively prove oligo incorporation during DNA replication, due to the intrinsic limits of studying replication *in vivo.*

Recombineering beyond E. coli

There is significant interested in developing recombineering methods in other organisms besides *E. coli*. There is evidence that SSAPs from bacteriophages can mediate recombineering

in their bacterial hosts, suggesting a host-specific interaction. One of the most successful examples of this is *Lactobacillus*, where up to 19% oligo recombination frequencies were achieved (Van Pijkeren & Britton 2012). Interestingly, although a SSAP derived from *L. leuteri* achieved these high recombination frequencies, a RecT-like protein from *E. faecalis* had similar activity, while other lactobacilli-derived SSAPs were significantly worse (van Pijkeren et al. 2012). Similarly, although both Beta and RecT are *E. coli* phage-derived SSAPs, Beta is 100-fold better at oligo recombination than RecT, and the *E. faecalis* RecT is almost as good as Beta even though *E. faecalis* is a Gram-positive bacterium (Datta et al. 2008). The mycobacterial RecT/Beta protein gp61 was 10- to 100-fold better than Beta and RecT (van Kessel & Hatfull 2007; van Kessel & Hatfull 2008). Similarly, in *Pseudomonas,* a RecT derived from *P. syringae* produced a 25-fold improvement while Beta had no effect (Swingle, Bao, et al. 2010). In *Corynebacterium*, RecT from *E. coli* was the most efficient SSAP mediating oligo recombination compared to a *Corynebacterium* phage SSAP and to the mycobacteria gp61, while Beta had no effect at all (Binder et al. 2013). Taken together, these results suggest a although a host-specific interaction is likely necessary, it is not sufficient for finding SSAPs that mediate oligo recombination with high frequencies.

Potential SSAP interaction partner

Understanding Beta's species-specific activity is desirable in order to transfer this activity to other organisms. Potential insights on this host-specific interaction can be derived from the eukaryotic SSAP Rad52. In addition to mediating the annealing of complementary DNA strands(Mortensen et al. 1996), Rad52 is able to interact the eukaryotic single-strand binding protein RPA and anneal RPA-coated ssDNA(Sung 1997; New et al. 1998; Shinohara & Ogawa

1998). The interaction of Rad52 with RPA seems to be species specific, yeast Rad52 cannot efficiently anneal DNA pre-coated with human RPA or *E. coli* SSB (Sugiyama et al. 1998; Shinohara et al. 1998). Similarly, in *E. coli* and other bacteria RecO is able to anneal SSB-coated DNA (Kantake et al. 2002), suggesting an analogous species-specific interacting partner for Beta. RecO binds both ss- and dsDNA, and promotes the renaturation of complementary DNA in a reaction stimulated by Mg^{2+} and inhibited at high NaCl concentrations (Luisi-DeLuca & Kolodner 1994). Like Rad52, RecO can only anneal ssDNA coated with its cognate SSB protein and not RPA-coated DNA (Kantake et al. 2002).

The RecO interaction with SSB requires the C-terminal domain of SSB(Hobbs et al. 2007). The crystal structure of RecO showed the SSB C-terminal peptide within a hydrophobic pocket of RecO surrounded by a positively charged surface, with phenylalanine and arginine residues mapped to the region strongly abolishing the interaction when mutated to alanine (Ryzhikov et al. 2011). The last 10 aa of the SSB C-terminal domain are not required for DNA binding but are essential *in vivo* (Curth et al. 1996). This C-terminal peptide may interact with the DNA binding domain of SSB, and upon binding of ssDNA it might become exposed and available to interact with other proteins involved in DNA metabolism (Kozlov et al. 2010; Shereda et al. 2008), including DnaG primase (Naue et al. 2013). Further supporting the existence of species-specific interactions of SSB, chimeric proteins between the *E. coli* and *Mycobacterium tuberculosis* SSB C-terminal domain showed that the C-terminal domain alone could increase the binding of Uracil DNA Glycosylase to a heterologous SSB (Handa et al. 2001). SSB forms a homotetramer independently of the C-terminal domain, and small amounts of otherwise lethal C-terminal domain mutants can functionally complement N-terminal domain mutants, likely forming functional mixed tetramers (Curth et al. 1996).

Interestingly, SSB is frequently part of the Beta/RecT operons, supporting the hypothesis that SSB might be important for their activity *in vivo.* To test the hypothesis that Beta interacts with SSB, I tested if Beta could remove the annealing inhibition of SSB-coated oligos *in vitro.* I further describe this interaction by mutating the C-terminal domain and find that it is required for *in vivo* oligo recombination and for Beta to anneal SSB-coated oligos *in vitro*. In addition, I found that Mg^{2+} has a key role in the Beta-SSB interaction. Based on these observations I propose a model explaining the quaternary structures of Beta and the implications of its interaction with SSB for developing efficient oligo recombinations in other organisms.

RESULTS

I hypothesized that Beta must interact with SSB, an interaction that may explain its species-specific activity. To test this, I added a 6x His tag to the N-terminus of the protein and verified the tag had no effect in its oligo recombination activity in *E. coli* (not shown). Beta and SSB were then expressed and isolated with nearly 100% purity. To test Beta's activity *in vitro*, I developed an oligo annealing assay. In this assay two complementary 90 mer oligos, one with a 3'-Fluorescein and the other with a 5'-Iowa Black FQ dark quencher are mixed and the annealing reaction can be measured on a plate reader as the fluorescein is quenched. (Figure 3.5a). To test the Beta-SSB interaction, I pre-coated the oligos with SSB, which will prevent annealing unless additional factors are included. We find that Beta is capable of overcoming the annealing inhibition of SSB in a somewhat cooperative manner (Figure 3.5b). Interestingly, the ability of Beta to remove the SSB-mediated annealing inhibition is strongly dependent on the

Mg^{2+} concentration; in its absence the inhibition is barely affected (Figure 3.5c). To confirm this Mg^{2+} dependency, we repeated the experiment in the presence of EDTA (Figure 3.5d). To further confirm the Beta-SSB interaction, we tested the effects of deleting the last eight amino acids of the C-terminal domain of SSB, previously shown to be not required for DNA binding but essential for its interaction with other proteins (Curth et al. 1996; Shereda et al. 2008). We find that Beta is no longer able to anneal oligos pre-coated with SSBΔC8 (Figure 3.5d), suggesting that similar to other SSB-interacting proteins, Beta's interaction with SSB requires the C-terminal domain of SSB.

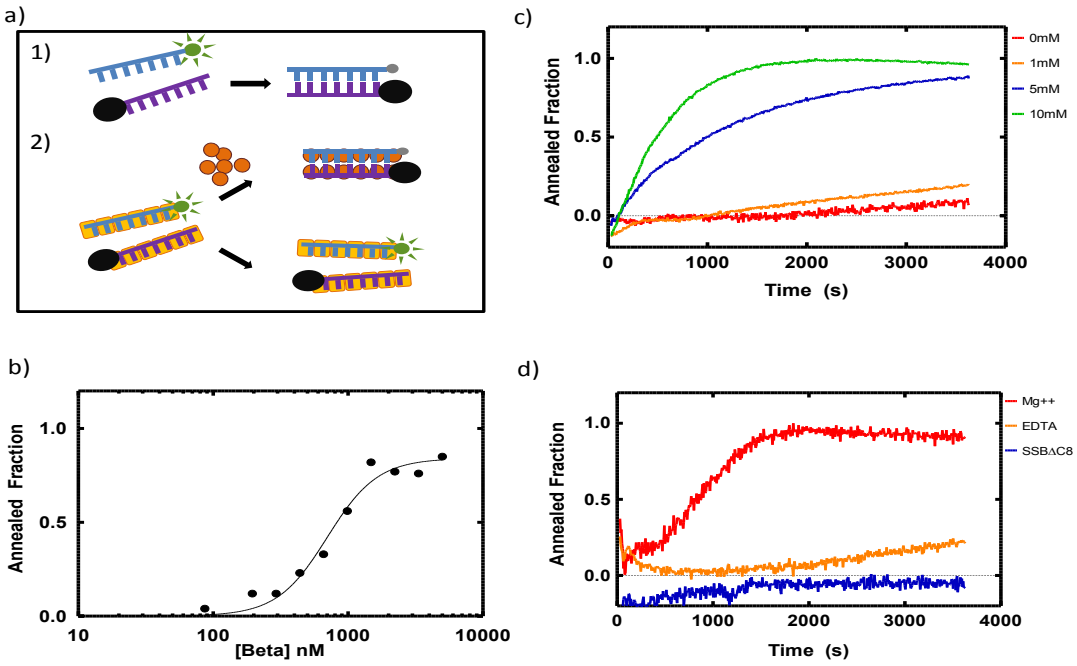


Figure 3.5. Beta interacts with SSB in a Mg^{2+}-dependent reaction. (a) Normally (1), two complementary oligos with compatible FITC fluorophore and quencher anneal, leading to a decay in the fluorescence intensity that can be tracked over time. Thus the fluorescence intensity will be proportional to the amount of starting substrate, while the remaining fraction will be the annealed product. If the oligos are coated with SSB prior mixing (2), they will be prevented from annealing unless additional factors are able to remove the inhibition. (b) Steady state annealed fraction products of SSB-coated oligos at varying amounts of Beta protein. Trend estimates a Hill slope of 2.3, SE 0.46. (c)Effects of Mg^{2+} concentration in Beta-mediated annealing of SSB-coated oligos. (d) Specificity of the Beta-SSB interaction. Neither EDTA or SSBΔC8 allow the annealing reaction.

Our experiments show that Beta can anneal SSB-coated oligos. Next we tested if, similar to analogous proteins, the interaction with SSB is mediated by the C-terminus of Beta. We serially truncated the protein, generating fragments 1-245, 1-228, 1-211, 1-194 and 1-177, the smallest fragment previously found to be sufficient for DNA binding (Wu et al. 2006). Only full-length Beta is capable of achieving oligo recombinations at high frequencies, while deleting just the last 21 amino acids of the protein (truncation 1-245) decreased function at least ~77-fold

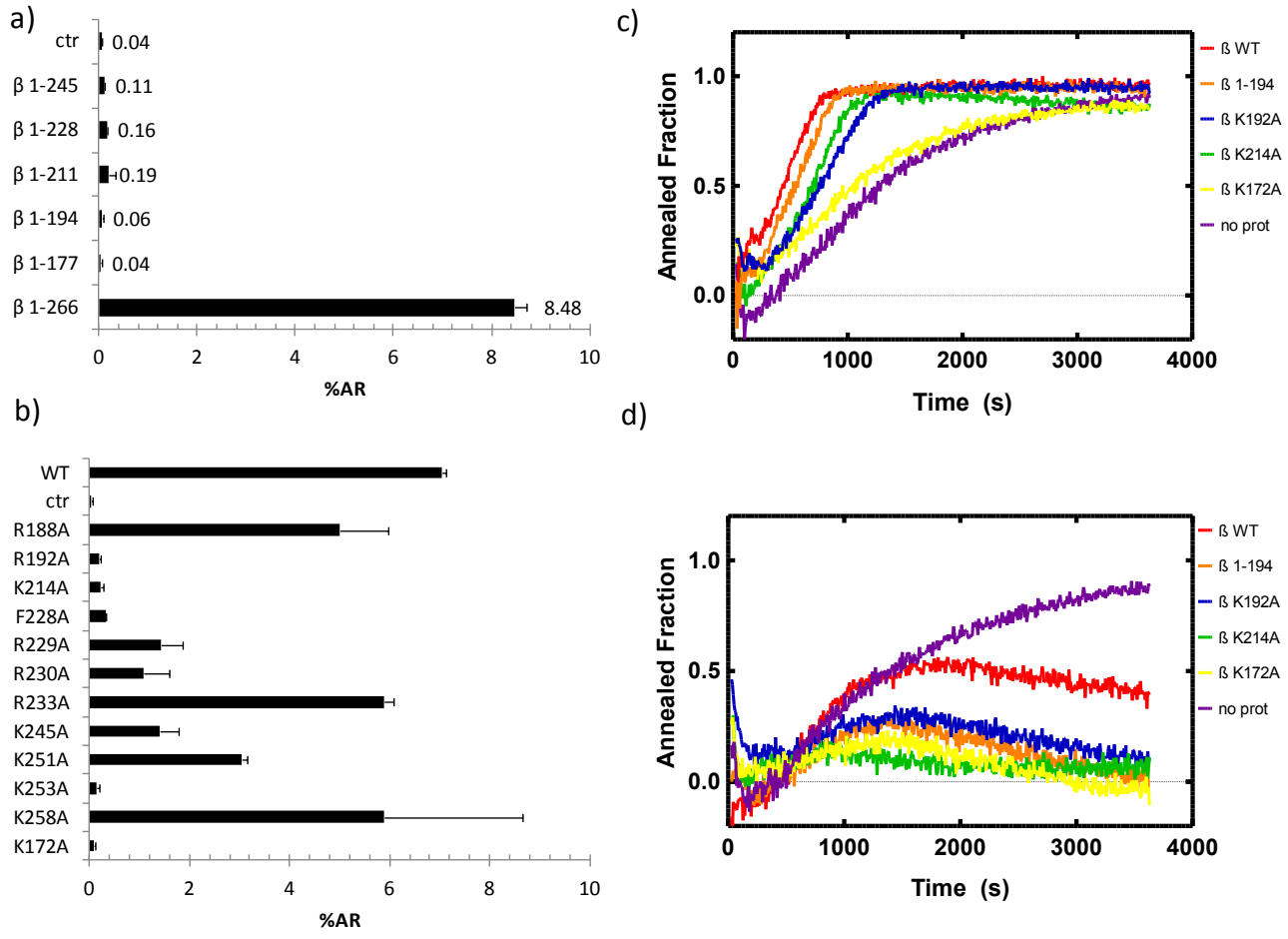


Figure 3.6 Characterizing the C-terminal domain of Beta. Oligo recombination frequencies of Beta C-terminal domain truncations (a) and single amino acid substitutions (b) N=2 biological replicates, ctr: glucose-suppressed WT Beta (c)Oligo annealing assay in the absence of SSB show that Beta and the C-terminal domain mutants accelerate annealing in the absence of Mg^{2+} relative to the no protein control. (d)However, in the presence of Mg^{2+}, all proteins seem to have reduced activity. Error bars: S.D.

(Figure 3.6a). The serial truncations revealed that even though the C-terminal domain is not required for DNA binding, even small deletions significantly reduce its activity *in vivo*. To further evaluate the domain, we mutated the lysine, arginine and phenylalanine amino acids, since these might be essential for interacting with the negatively charged C-terminal domain of SSB, as is the case of RecO (Ryzhikov et al. 2011). Some of the mutations severely reduced recombination frequencies, especially mutations in the 192-228 aa segment (Figure 3.6b). The negative control mutant K172A, previously shown to abolish DNA binding (Wu et al. 2006) also had reduced recombination rates.

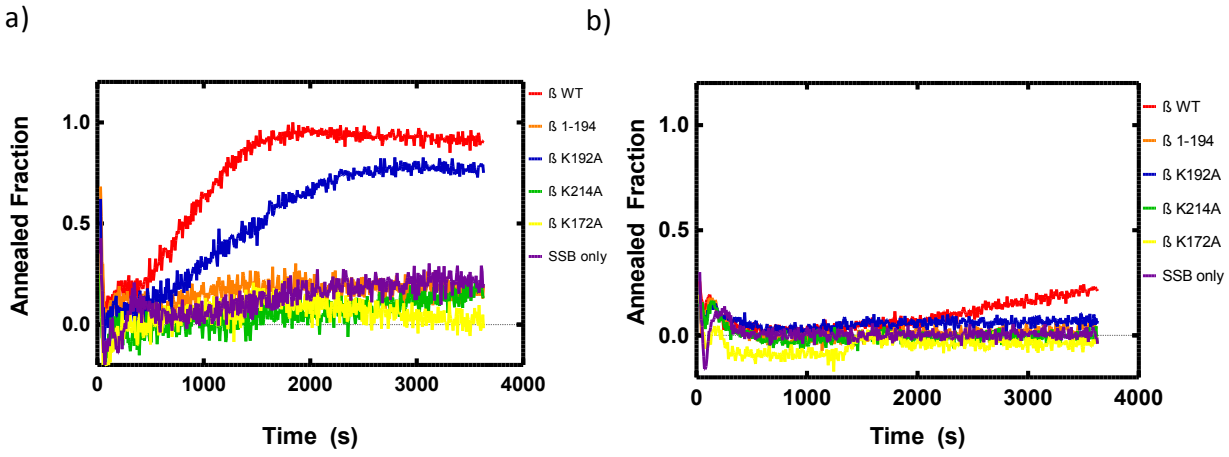


Figure 3.7. The C-terminal domain of Beta is involved in its interaction with SSB. (a) WT Beta and the C-terminal domain mutants were compared in their ability to anneal SSB-coated oligos in the presence of 10mM MgCl2 (b) The same mutants were tested in the presence of EDTA, confirming that Mg2+ is required for the interaction.

The mutation screen revealed that single amino acid substitutions in the C-terminal domain of Beta dramatically reduced oligo recombination frequencies *in vivo.* A subset of these, along with all the truncations and the WT protein were further characterized *in vitro*. Out of the

five truncations tested *in vivo*, only Beta 1-194 was stable under our purification conditions, while there was no problem purifying the point mutants. The purified mutated proteins were compared to the WT in terms of their oligo annealing kinetics, absent SSB (Figure 3.6c-d). In the absence of Mg^{2+}, all mutants behave very similarly to WT Beta in their annealing kinetics, suggesting their strand annealing activity was unaffected. (Figure 3.6c). Unexpectedly, when the assay was repeated in the presence of 10 mM $MgCl_2$, all proteins have reduced annealing (Figure 3.6d). However, the 1-194 truncation and all the point mutants tested show lower annealing than WT Beta. In the absence of Mg^{2+} the K172A mutation had no effect on DNA annealing when compared to the no protein control, however in the presence of Mg^{2+} it inhibited annealing, similar to the 1-177 fragment and the other point mutants. Thus, the C-terminus of Beta is only important for DNA annealing at high Mg^{2+} concentrations.

Mg^{2+} inhibits the strand annealing activity of Beta, and this inhibition is exacerbated by deleting its C-terminal domain or disrupting it with single amino acid substitutions. Since Mg^{2+} is required for Beta to anneal SSB-coated oligos, we hypothesize the C-terminal domain mutations will alter this Beta-SSB interaction. Compared to WT Beta, the 1-194 truncation is unable to break the SSB inhibition, behaving very similarly to the SSB-only control (Figure 3.7a). The K172A mutation is also unable to anneal SSB coated oligos, suggesting that DNA binding might be required for Beta to remove the annealing inhibition of SSB. The K192A mutant had reduced activity, while K214A was more severely affected, being indistinguishable from the SSB only control. In the presence of EDTA instead of Mg^{2+}, none of the Beta mutants is able to anneal SSB coated oligos while WT Beta seems to still have some activity (Figure 3.7b) Thus, the C-terminus of Beta plays a significant role in the interaction with SSB, in a reaction strongly enhanced by Mg^{2+}.

DISCUSSION

The lambda phage SSAP Beta is the single-most important factor enabling multiplex genome engineering in *E. coli,* mediating highly efficient oligo recombination. There have been several attempts to develop similar methods in other bacteria, with mixed results. There seems to be a species-specific interaction, since SSAPs from phages tend to be active in their corresponding bacterial hosts, although there are many exceptions to this. In this work I decided to characterize Beta biochemically, with the goal to elucidate this host-specific interaction and ultimately use this knowledge to export its functions to human cells.

I hypothesized that like the prokaryotic RecO and the eukaryotic SSAP Rad52, Beta would be able to anneal DNA in the presence of SSB. I tested this *in vitro* by tracking the annealing of two complementary oligos coated with SSB, since in the absence of additional factors SSB prevents this annealing. In these experiments I discovered that Beta can be one of these factors and is able to anneal SSB-coated oligos in a somewhat cooperative reaction. This activity is dependent on the concentration of Mg^{2+}, with EDTA strongly inhibiting annealing. I found 10 mM $MgCl_2$ to be the optimal concentration for Beta to interact with SSB, which is also the optimal concentration for RecA activity (Lusetti et al. 2003). This concentration is higher than the free 1-2 mM Mg^{2+} thought to be available intracellularly (Alatossava et al. 1985). It is possible that this deficiency might be compensated *in vivo* by crowding effects (Lavery & Kowalczykowski 1992; Lusetti et al. 2003). Alternatively, the subtle activity detected at 1 mM Mg^{2+} might be sufficient for its *in vivo* activity.

After SSB binds DNA, its negatively charged C-terminal tail becomes available to SSB-interacting proteins (Kozlov et al. 2010; Shereda et al. 2008). As expected, this C-terminal tail is essential *in vivo,* given the fundamental function of SSB. I tested whether Beta requires this C-terminal domain for its ability to anneal SSB-coated ssDNA oligos, and found that deleting the last 8 amino acids of SSB abolishes its interaction with Beta. This suggests that the Beta-SSB interaction is a specific protein-protein interaction and not just competition for DNA binding. Moreover, this protein-protein interaction has been shown to be species-specific. Thus SSB might be the factor that constrains phage SSAPs like Beta to have activity only in bacteria with compatible SSBs.

To further dissect this Beta-SSB interaction I decided to study the C-terminal domain of Beta. First, I tested if this domain was required *in vivo* for oligo recombination even though it has been previously shown not to be required for DNA binding. I found that just deleting the last 24 amino acids reduces oligo recombinations frequencies to levels indistinguishable from background. Further characterization of the domain focused on single-amino acids substitutions targeting residues known to be important for SSB-interacting proteins, specifically lysine, arginine and phenylalanine residues. This revealed several single amino acid substitutions caused dramatically reduced oligo recombination, thus confirming that the C-terminal domain of Beta has a crucial role in oligo recombination.

To expand this observation, I characterized the DNA annealing kinetics of naked ssDNA oligos in the presence of WT Beta and the C-terminal domain mutants. In the absence of Mg^{2+}, the annealing activity of the Beta truncation and the point mutants was largely unaffected. Surprisingly, in the presence of Mg^{2+} the truncation and the point mutations inhibit oligo annealing, behaving like SSB. The fact that the C-terminal domain mutants have this dominant

negative activity in the presence of Mg^{2+} suggests that their ability to interact with magnesium remains. The K172A mutation had no annealing activity relative to the no protein control in the presence of Mg^{2+}, but it also had a dominant negative effect in the presence of it. This residue is adjacent to negatively charged residues strongly conserved in the Beta/RecT class of SSAPs, predicted to be involved in coordinating magnesium binding (Iyer et al. 2002). The dominant negative activity seen suggests that the conformational change induced by magnesium may alter DNA binding in a way that the K172 residue is no longer required.

The amino acid substitutions tested seem to strongly affect DNA annealing in the presence of magnesium. This result is surprising, since the positively charged amino acids tested should not be directly interacting with Mg^{2+}, although they could still be essential for the folding of the domain. Alternatively, the loss of positively charged residues might increase the affinity to Mg^{2+}, perhaps stabilizing a Mg^{2+}-induced conformational intermediate resulting in the dominant negative phenotype. It would be interesting to test other C-terminal domain mutations that showed reduced activity *in vivo*, especially the F228A mutation since it would not affect the charge. In addition, testing the Mg^{2+} responsiveness of shorter truncations might help confirm the presence of a Mg^{2+} binding domain.

The Beta C-terminal domain mutants were also tested on their ability to anneal SSB-coated oligos. In the absence of Mg^{2+}, only WT Beta seems to have some activity. In the presence of Mg^{2+}, all mutants behave fairly similarly to the SSB-only control. The exception is the K192A mutant, which showed some activity, although lower to Beta. The fact that these mutants had defective DNA annealing in the naked-oligo assay means that I cannot tell if their lack of activity in the SSB-coated oligo annealing assay is because these residues are directly interacting with SSB or if they are causing a defect in the Mg^{2+}-dependent conformational

change. The fact that K192A partially works in the presence of SSB but not in its absence suggest that the interaction with SSB is active in this mutant and the SSB interaction induces a different DNA binding mode. This could be mediated by altering SSB binding and forming a Beta-SSB-ssDNA complex, similar to what has been suggested for RecO (Ryzhikov et al. 2011). If this was the case, it may be expected then that K172A would have a similar effect, yet this protein has no activity in the presence of Mg^{2+}, with or without SSB. If there is a magnesium-binding domain in the central part of the protein, mutations closer to it might have a stronger effect, while perhaps the C-terminal domain tail might contain the residues directly interacting with SSB. This might also explain why the longer truncations are slightly more active than the shorter ones. These proteins might still be able to have normal magnesium-induced conformational changes while losing the specific SSB interacting domain, and thus are behaving solely as annealing proteins. Further mutagenesis is required to elucidate this, although it is likely that the magnesium-binding domain and the SSB-interacting domain are interdependent, which may complicate the analysis. Ultimately, crystal structures of all the different conformations might be required to describe the domains with certainty.

The mutations evaluated seem to all have defects that highlight the importance of magnesium concentration for this protein class. Similarly, deletion of RecA C-terminal domain residues significantly alters the effects Mg^{2+} has on this protein's strand exchange activity (Lusetti et al. 2003). Wild-type RecA has optimal activity at Mg^{2+} concentrations of 10 mM, and this dependency decreases progressively with truncations of the C-terminal domain, which is negatively charged. A RecA E343K mutation similarly decreased the Mg^{2+} dependency. Interestingly, both C-terminal domain deletions and high Mg^{2+} enhance RecA's ability to

displace SSB (Eggler et al. 2003). This suggests that magnesium plays a common role in these proteins modulating their interaction with SSB through their C-terminal domain.

In this study I showed that the Lamba phage SSAP Beta is capable of interacting with SSB and override its annealing inhibition, in a reaction greatly enhanced by the presence of Mg^{2+}. This Mg^{2+} requirement for Beta to interact with SSB, combined with the electron microscopy images that revealed Mg^{2+} had a significant role stabilizing the polymeric forms of Beta (Passy et al. 1999) suggests that this form, either rings or ellipsis, is the active conformation required for interacting with SSB and removing its annealing inhibition. We find that amino acid substitutions that remove positive charges near conserved negatively charged residues predicted to coordinate magnesium binding create a dominant negative phenotype when annealing naked oligos in the presence of magnesium. One explanation for it could be that with these mutants the ssDNA oligo is bound by a hyper-stabilized Beta ring or ellipsis, unable to transition to the annealed helical conformation. This might explain why these mutants effectively behave like SSB.

An alternative explanation might be that the ssDNA is binding Mg^{2+}, which in turn decreases the protein's affinity for it. In this case, however, one would expect that the protein mutants would have behaved similar to the no protein control in the naked DNA annealing assay instead of the dominant negative effect seen. It is also possible that the magnesium is activating DNases in the proteins, releasing the fluorophore, but this was verified not to be the case (not shown).

The Beta-SSB interaction described here further solidifies the role of DNA replication in oligo recombinations, supporting the model of oligo incorporation on nascent DNA as pseudo-

Okazaki fragments. During DNA replication, DNA helicase and DNA polymerase closely interact, allowing for continuous synthesis of the leading strand(Langston et al. 2009). The ssDNA generated on the lagging strand is bound by SSB, which protects it from nucleases. Thus Beta interacting with SSB would help explain the strand bias observed in oligo recombination, since the SSB-coated lagging strand would be accessible for annealing by a ssDNA-Beta oligo complex, while the leading strand would be more inaccessible. If the whole chromosome is replicated in the traditional continuous/discontinuous model, strand bias against the leading stand might be expected to be higher. However, recent evidence suggests that this model oversimplifies *in vivo* replication and that both strands can be discontinuous, due to for example DNA lesions or collisions with RNA polymerase(Langston et al. 2009).

After Beta mediates the incorporation of a mutation-carrying oligo, it may form a dsDNA-Beta helical complex that might temporarily protect the oligo from degradation by proofreading polymerases and the mismatch repair system. However, at some point this structure would have to be removed, in order to allow transcription or subsequent DNA replication. At this point a mismatched duplex could be recognized by the mismatch repair machinery, until a second round of replication copies the mutated strand and effectively removes the mismatch.

The Beta-SSB interaction might be generalizable to the rest of the Beta/RecT family of phage-derived SSAPs. This protein family has strongly conserved negatively charged residues predicted to bind magnesium, suggesting they might have a similar response to the divalent cation. Future work comparing different SSAPs biochemically is needed to verify this. It will be interesting to compare RecT with Beta. Since both are derived from *E. coli* phages, it is expected that RecT might also be able to interact with SSB. In fact, our preliminary experiments seem to support this (not shown). However, Beta is about 100-fold better at oligo recombination

than RecT, suggesting there are significant differences between these closely related proteins. Supporting this, electron micrographs revealed that RecT ring formation does not require DNA (Thresher et al. 1995), while ssDNA is essential for Beta to form larger rings. Further comparing Beta to RecT and other SSAPs will help further dissect how Beta mediates oligo recombinations at such high frequencies.

The interaction with SSB might in part explain the variable activity seen in efforts using phage-derived SSAPs to increase oligo recombination frequencies in different bacteria. Closely related species have similar SSB's and thus might be able to interact with the same SSAP. However, the C-terminal domain of these SSAPs tends to be not well conserved, so the protein-protein interaction might be easily changed, or it might not the right kind of interaction. There might need to be a balance, for example if the SSAP is too effective at displacing SSB it might be too toxic to cells, perhaps causing replication fork collapses. Conversely, if the interaction is not strong enough it might not be sufficient to allow the annealing of the mutation-carrying oligo to its target strand. However, different organisms may have specific requirements, be it the magnitude of the interaction with SSB, specific intracellular ionic environments or differences in chromosome structure and the DNA replication machinery. In their phages, these proteins are likely to have the role of DNA repair and recombination machinery, a function that involves an exonuclease and possibly other factors. Thus, it is possible that even a protein as efficient as Beta is not optimized for oligo recombination. Exploring the functional landscape with a library of evolutionarily diverse SSAPs and their mutants may yield improved activities in *E. coli.* This might be a way to discover natural SSAPs with properties specifically suited for organisms of interest. In turn, this could also lead to a deeper understanding of their mechanisms, and may

eventually lead to rationally designed SSAP libraries from which highly active candidates could be selected for any organism of interest.

MATERIALS AND METHODS

Protein purifications

An arabinose inducible expression vector was chosen for expressing the proteins described in this work(Choe et al. 2005). After cloning via isothermal assembly(Gibson et al. 2009) and adding an N-terminus 6x His tag, the arabinose-inducible vector with the constructs of interest were transformed into NiCo21(DE3) competent *E. coli* (NEB). 50 mL LB media with 25 μg/mL chloramphenicol were inoculated with 1:100 overnight confluent cultures ground under glucose repression. The 50 mL cultures were grown for 6 hours at 37 C, then induced with a final arabinose concentration of 0.1%. The cultures were spun down 10 mins 5,000 g, 4 C and the pellets were frozen in a dry ice ethanol bath.

The pellets were lysed with P-BER with Enzymes (Thermo scientific) for 10 mins at room temperature following the manufacturer's instructions. The lysates were mixed 1:1 with binding buffer (40m M Imidazole, 500 nM NaCl, 50 mM Tris pH 7.4), spun down 10 mins 5,000 g, 4 C and the soluble fraction was added to a 20 mL column with 2 mL His GraviTrap (GE Healthcare) pre-equilibrated with binding buffer. After binding, the columns were washed twice with 20 mL of wash buffer (100 mM Imidazole, 500 nM NaCl, 50 mM Tris pH 7.4), then eluted with 4mL of elution buffer (500 mM Imidazole, 500 nM NaCl, 50m M Tris pH 7.4), collecting the flow-through in 1.5mL tubes, ~1mL each.

The protein concentration of the eluted fractions was checked with a Qubit (Invitrogen), and their stability and purity checked by SDS-PAGE. The purest, most concentrated fractions were pooled and buffer exchanged with Zeba desalting columns 7K MWCO (Piercenet) to storage buffer (200 nM NaCl, 50 mM Tris pH 7.4, 1 mM DTT). If required, buffer exchanged protein preps were concentrated with Amicon Ultra-4 10K centrifugal filters (Millipore).

Oligo annealing assay

Fluorophore/quencher complementary oligos were ordered from IDT (5'-AGCAAGCACGCCTTAGTAACCCGGAATTGCGTAAGTCTGCCGCCGATCGTGATGCTGCCTTTGAAAAAATTAATGAAGCGCGCAGTCCA/6-FAM/-3' and 5'-/IABkFQ/TGGACTGCGCGCTTCATTAATTTTTTCAAAGGCAGCATCACGATCGGCGGCAGACTTACGCAATTCCGGGTTACTAAGGCGTGCTTGCT-3'. For the SSB annealing assays, the 10 nM oligo solutions with 20 nM NaCl, 1 mM DTT, 50 mM Tris pH 7.4 were separately incubated with 500 nM SSB or SSBΔC8 for 20 mins at 37 C. The reactions were tracked in a Synergy H4 Hybrid Microplate Reader (Biotek) in half-area, low-bind black 96 well plates. The oligos were serially added to the plate, followed by the SSAP in the same buffer.

Fluorescence quenching-based annealed fraction estimate:

I = Fluorescence intensity at a given time is

$$I = F_f I_f + F_b I_b$$

Where F_f = Free fraction, I_f = Free intensity, F_b= Bound fraction, I_b = Bound fraction. For a DNA annealing assay, F_f is the substrate, and F_b is the product

$$\text{At } t_o \rightarrow F_f = 1, F_b = 0$$

$$F_f + F_b = 1;\ F_f = 1 - F_b \text{ so}$$

$$I = I_f(1 - F_b) + F_b I_b$$

$$F_b = \frac{I - I_f}{I_b - I_f} = 1 - \frac{I - I_b}{I_f - I_b}$$

This calculation is independent of experimental background (B)from the reader:

$$F_b = \frac{(I - B) - (I_f - B)}{(I_b - B) - (I_f - B)} = \frac{I - I_f}{I_b - I_f}$$

I_b was estimated from the minimal steady-state fluorescence of annealed oligos in the presence of protein, while I_f was measured in parallel for each reaction using an unlabeled oligo instead for the quencher. This helped control for the variable background fluorescence of different protein solutions and the fluorescence decay of the FITC fluorophore over the time course measured. The reactions were tracked for an hour, measuring every 7 s. The naked-oligo experiments were done in a similar way, except no SSB was added during the pre-incubation step. Annealing and steady-state graphs were generated using GraphPad Prism 5.

Oligo recombination

The different SSAPs cloned into the arabinose-inducible vectors were transformed into a *ΔmutS* MG1655 *E.coli* strain with a frameshift mutated copy of GFP integrated. Overnight cultures were grown under 0.1% glucose repression to confluence. The next day, these were diluted 1:100 in 2 mL LB 25 μg/mL chloramphenicol and grown at 32 C for two hours. At this point 0.1% arabinose was added and cultures were grown for an additional 30 mins. After induction, 1 mL of cultures were spun down 1min 15,000 g at 4 C, washed twice with 4 C deionized water and resuspended in 50 uL 50 μM oligo in deionized water. This was transferred

to chilled 0.1 mL cuvettes, and the cells were elecroporated on a BioRad Gene Pulser1.78 kV, 25 μF, 200 Ω. After electroporation, cells were recovered overnight in 1 mL LB with 0.1% glucose. For measuring recombination frequencies, overnight cultures were grown 1:100 in LB, then diluted 1:100 in water and analyzed in an LSR Fortessa with a 488 nm laser, 530/30 nm filter.

CHAPTER FOUR

CONCLUDING REMARKS AND FUTURE WORK

Next generation sequencing technologies have revolutionized biomedical research. Current sequencing projects are producing massive amounts of data on human genetic variability, and as cost keep decreasing sequencing whole genomes will become routine (Mardis 2006; von Bubnoff 2008). In order to fully make sense of all this new data, new experimental tools will be required. The goal of this dissertation is to contribute to the development of one of such tools, multiplex genome engineering in mammalian cells. This method, originally developed in *E. coli,* allows oligonucleotide pools to be recombined at very high frequencies (Wang et al. 2009).

The first part of this project, described in Chapter II, focused on studying and validating an oligo recombination reporter system in human cells (Rios et al. 2012). Using this reporter I found that chemically modified base analogs can increase oligo recombination frequencies by avoiding the mismatch repair machinery. In addition, I discovered that part of the toxicity seen in oligo recombination in human cells might be in part from a cellular immune response to the oligo. It is likely that modifications such as PTO bonds increase the half-life of the oligo leading to increased immune signaling and higher toxicity. It might be possible to identify the sensors starting this immune response and knock them out to reduce toxicity and increase recombination frequencies.

By optimizing the oligo design with modified bases and reduced number of PTO bonds I was able to generate stably-modified cells, an accomplishment that eluded the original creators of this reporter cell line (Liu et al. 2009). The frequencies achieved, however, are not nearly good enough for this method to be practically useful, especially compared with methods such as CRISPR/Cas9. The biggest factor contributing to the high oligo recombination frequencies seen in MAGE is the SSAP Beta, capable of increasing recombination rates four orders of magnitude

in *E. coli* (Ellis et al. 2001). I tested whether expressing Beta in human cells had any effect in oligo recombination. After verifying expression, adding a nuclear localization signal tag and testing various oligo lengths, I was unable to detect any activity. Additionally, I tested similar proteins such as RAD52 and ICP8, but these also had no effect in oligo recombination frequencies.

Despite the fact that Beta has such a key role enabling oligo recombination at high frequencies and the numerous attempts of developing MAGE in other organisms, very little is known about how this protein works. This inspired Chapter III, where I study Beta with the goal of eventually translating its activity to human cells. I hypothesized that Beta must be interacting with SSB, a component of the host DNA replication machinery, via a species-specific interaction requiring the C-terminal domain of Beta, analogous to the eukaryotic SSAP RAD52 (Sung 1997; New et al. 1998; Shinohara & Ogawa 1998).

By purifying and characterizing Beta *in vitro*, I was able to prove it is capable of interacting with SSB in a cooperative manner. Surprisingly, this interaction with SSB is strongly dependent on the presence of magnesium, even though magnesium was found to have no effect in its conformational changes (Wu et al. 2006). I further delineate the SSB-interaction domain to the C-terminal domain of Beta. This domain is not required for DNA binding *in vitro*, however I found that truncations and even single amino acid substitutions are capable of dramatically decreasing Beta's oligo recombination activity. These C-terminal domain mutants anneal complementary oligonucleotides similarly to the WT protein in the absence of magnesium. However, in the presence of magnesium these mutants behave more like SSB, preventing the annealing of complementary oligos. In addition, these mutants have reduced or completely abolished capacity of annealing SSB-coated oligos, suggesting that the C-terminal domain is

required for the Beta-SSB interaction. In addition, I show the Beta-SSB interaction requires the negatively charged C-terminal tail of SSB, suggesting this interaction is likely to be a species-specific protein-protein interaction and not just direct competition for DNA binding. This datum suggest that the reason the C-terminal domain Beta mutants are defective at oligo recombination is because they have lost the ability to properly interact with SSB.

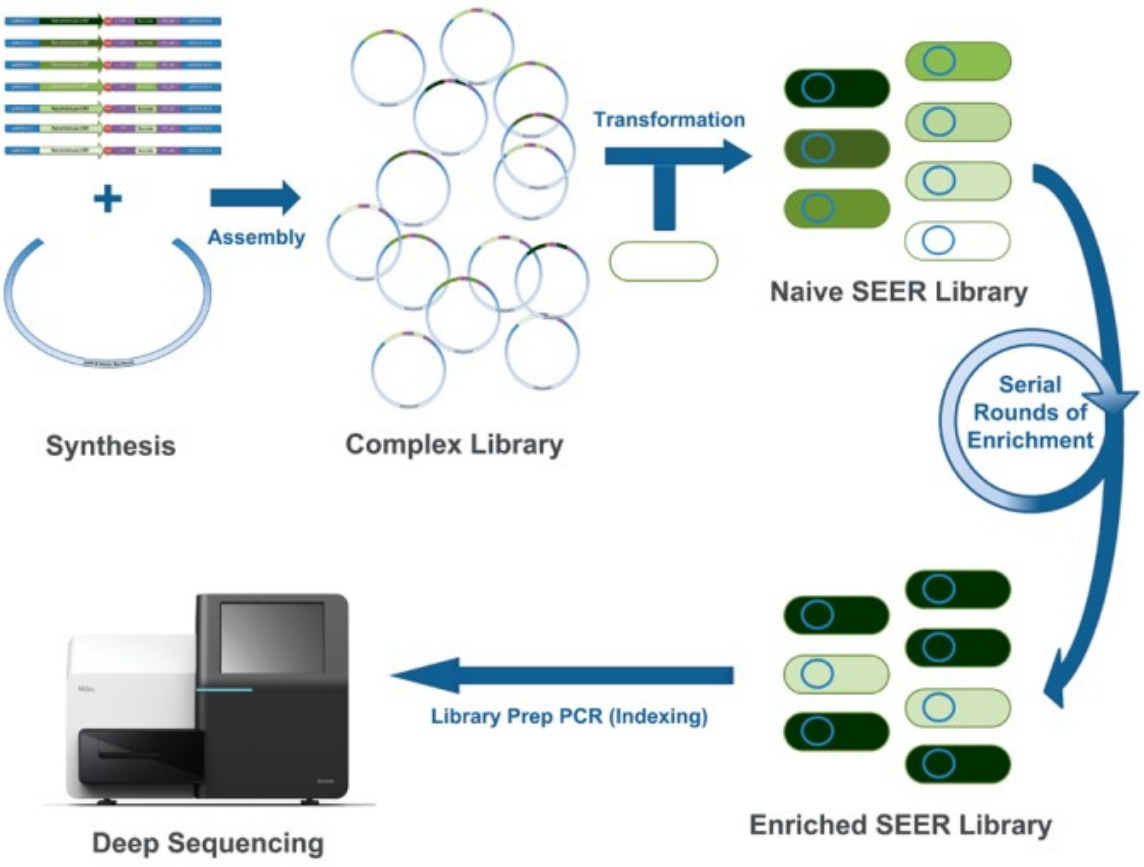


Figure 4.1. Serial Evolutionary Enrichment of Recombinases. Abarcoded library of 87 codon-optimized, naturally occurring SSAP was synthesized and cloned into an arabinose-inducible expression vector. This complex library was transformed into a reporter strain containing a broken antibiotic resistance marker.After a single round of oligo recombination the library is enriched for SSAP activity. This enriched library is mini-prepped, re-transformed and the next round of selection is performed. After several rounds of selection, the enriched libraries are quantified by deep sequencing of the barcodes.

Further biochemical characterization of Beta's C-terminal domain will help determine which residues are directly involved in the Beta-SSB interaction. The mutants studied so far seem to have a defective, dominant-negative phenotype in the presence of magnesium. Thus Beta might have three domains: an N-terminal domain DNA binding domain, a C-terminal SSB-interacting domain and a central magnesium-coordinating domain, although more data is needed to confirm this.

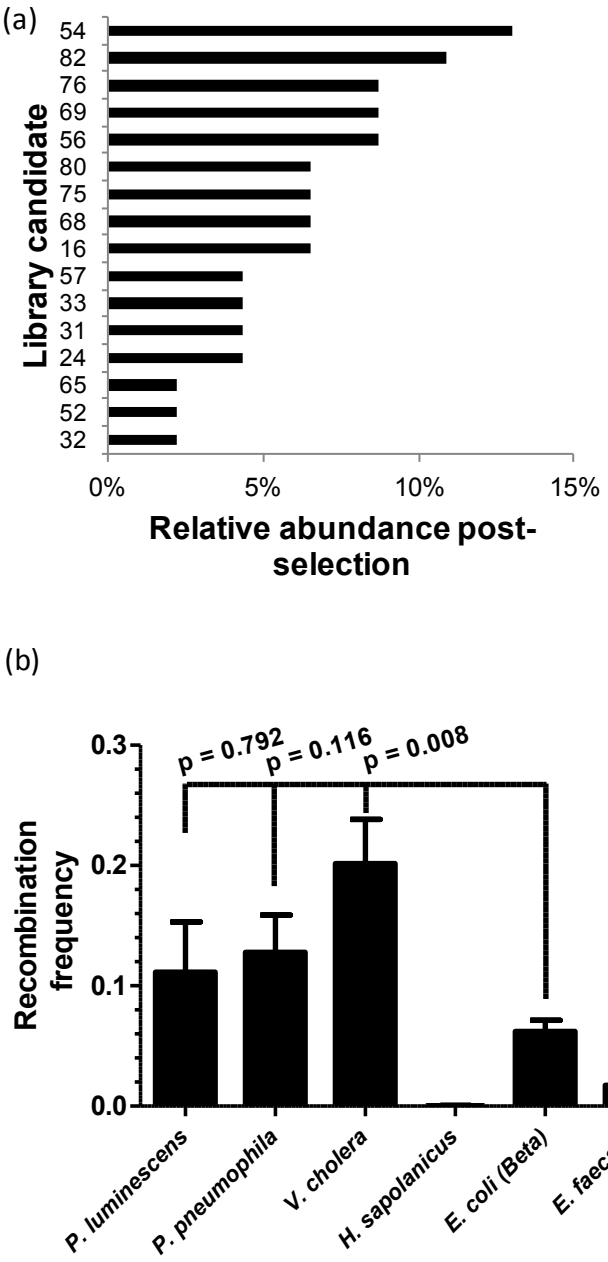


Figure 4.2. Sample results from SSAP library enrichment. (a) After rounds of selection the relative frequencies for each SSAP is determined by sequencing their corresponding barcodes. (b) The top most abundant SSAPs were tested for their oligo recombination activities. Some of the hits seem to have 2-3 fold higher activity than Beta, which is remarkable given Beta's already has frequencies in the 10^{-1} range.

Even though we now know that Beta interacts with SSB, this does not explain why it is such a powerful catalyzer of oligo recombination in *E. coli.* To further elucidate this, I will compare Beta with RecT, another *E. coli* phage-derived SSAP with significantly lower activity than Beta. In addition, I will characterize additional highly active SSAPs. For this, we synthesized a library of 87 barcoded, codon optimized SSAPs, which was submitted to

multiple rounds of selection for oligo recombination activity in *E. coli* (Figure 4.1). Deep sequencing of the barcodes will determine the relative activities of the selected proteins (Figure 4.2a). Preliminary results revealed several candidates with activities similar or higher than Beta (Figure 4.2b). Biochemical characterization of these candidates will provide additional data on what parameters correlate best with *in vivo* activity.

This method of selecting SSAPs from a large library might prove useful for developing MAGE in other bacteria and possibly eukaryotic cells. Alternatively, it might be necessary to combine Beta with RAD52. A library of chimeric Beta-RAD52 proteins could be used to select for a SSAP with the catalytic activity of Beta and the RAD52-RPA interaction. Specifically, the protein motif used by RPA-interacting proteins has been described (Mer et al. 2000). It might be possible to create a library of variations of this motif in Beta's C-terminal domain followed by rounds of selection in human cells.

It is possible that using a SSAP for improving oligo recombination in human cells might not be the best approach, perhaps due to fundamental differences between prokaryotic and eukaryotic genome structures or DNA replication machineries. Overexpressing Rad52p in yeast seems to have no effect in oligo recombination(Liu et al. 2002). However, it is possible that RAD52 is not the right kind of SSAP, just like RecT is not very efficient at oligo recombination in *E. coli*. In addition, RAD52 is functionally analogous to RecO, and overexpressing RecO did not enhance oligo recombination in *E. coli* (not shown), further supporting the idea that only certain kinds of SSAPs have high oligo recombination activities. The biochemical studies developed described in this dissertation may help identify this activity.

An alternative approach for improving oligo recombination in human cells could be based on the homologous recombination proteins RAD51 and RAD54. These proteins can mediate chromatin remodeling and strand invasion(Zhang et al. 2007), potentially a more versatile strategy for genome-wide oligo recombination in eukaryotic cells. Supporting this, overexpressing RAD51 and RAD54 together in yeast enhanced oligo recombination (Liu et al. 2002; DiCarlo, personal communication). In addition, these proteins are well characterized, making them attractive targets for engineering. A good starting point could be testing the I287T and K284E mutations in human RAD51, whose equivalent yeast Rad51p mutations I345T and K342E, respectively, have been shown to increase oligo recombination rates(Liu et al. 2002). Depending on these results, the mutant library screen might be expanded to the rest of the RAD51 L2 loop and adjacent region (327–344 aa in yeast). It would also be interesting to test for any synergy from co-expressing RAD51 and RAD54.

Even then, it is possible that oligo-mediated recombination might not be the optimal pathway for achieving highly efficient genome engineering in eukaryotic cells. An alternative strategy could involve nuclease-mediated oligo recombination. Although the nuclease approach has intrinsic multiplexing limits, this method would still be useful for testing human variants. Double-strand break mediated recombination might benefit from engineered SSAPs, especially one that is able to outcompete the endogenous non-homologous end joining pathway. In addition, improvements in targeted nucleases may translate to targeted nickases(Wang et al. 2012), an alternative method that could be multiplexed but is currently limited by low frequencies.

For this dissertation, it was my goal to develop a genome engineering method for human cells comparable to MAGE. This proved to be quite challenging, but in the process several useful

discoveries were made. I hope in the future these discoveries will contribute to the development of improved methods for engineering human cells and testing human genetic variants. The speed in which this field has progressed has been dazzling. The time is ripe for fundamental advances in our understanding of the genetic basis of human disease leading to significant improvements in healthcare.

REFERENCES

Aarts, M. et al., 2009. Gene modification in embryonic stem cells by single-stranded DNA oligonucleotides. *Methods in molecular biology (Clifton, NJ)*, 530, p.1.

Aarts, M. & te Riele, H., 2010. Progress and prospects: oligonucleotide-directed gene modification in mouse embryonic stem cells: a route to therapeutic application. *Gene Therapy*, 18(3), pp.213–219.

Aarts, Marieke & te Riele, H., 2010. Subtle gene modification in mouse ES cells: evidence for incorporation of unmodified oligonucleotides without induction of DNA damage. *Nucl. Acids Res.*, p.gkq589.

Alatossava, T. et al., 1985. Manipulation of intracellular magnesium content in polymyxin B nonapeptide-sensitized Escherichia coli by ionophore A23187. *Journal of Bacteriology*, 162(1), pp.413–419.

Anon, 2011. Move over ZFNs. *Nat Biotech*, 29(8), pp.681–684.

Audic, S. & Claverie, J.-M., 1997. The Significance of Digital Gene Expression Profiles. *Genome Research*, 7(10), pp.986 –995.

Bamshad, M.J. et al., 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics*, 12(11), pp.745–755.

Barber, G.N., 2011. Innate immune DNA sensing pathways: STING, AIMII and the regulation of interferon production and inflammatory responses. *Current Opinion in Immunology*, 23(1), pp.10–20.

Binder, S. et al., 2013. Recombineering in Corynebacterium glutamicum combined with optical nanosensors: a general strategy for fast producer strain generation. *Nucleic Acids Research*. Available at: http://nar.oxfordjournals.org/content/early/2013/04/28/nar.gkt312 [Accessed April 30, 2013].

Briggs, A.W. et al., 2012. Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Research*. Available at: http://nar.oxfordjournals.org/content/early/2012/06/26/nar.gks624 [Accessed June 29, 2012].

Von Bubnoff, A., 2008. Next-Generation Sequencing: The Race Is On. *Cell*, 132(5), pp.721–723.

Burckstummer, T. et al., 2009. An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome. *Nat Immunol*, 10(3), pp.266–272.

Carr, P.A. et al., 2012. Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Research*, 40(17), pp.e132–e132.

Carr, P.A. & Church, G.M., 2009. Genome engineering. *Nature biotechnology*, 27(12), pp.1151–1162.

Chandrasekharan, S. et al., 2009. Proprietary science, open science and the role of patent disclosure: the case of zinc-finger proteins. *Nature biotechnology*, 27(2), p.140.

Choe, W., Chandrasegaran, S. & Ostermeier, M., 2005. Protein fragment complementation in M.HhaI DNA methyltransferase. *Biochemical and Biophysical Research Communications*, 334(4), pp.1233–1240.

Christodoulou, D.C. et al., 2011. Construction of Normalized RNA-seq Libraries for Next-Generation Sequencing Using the Crab Duplex-Specific Nuclease. *Curr. Protoc. Mol. Biol*, 94, pp.4.12.1–4.12.11.

Cong, L. et al., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, 339(6121), pp.819–823.

Coscoy, L. & Raulet, D.H., 2007. DNA Mismanagement Leads to Immune System Oversight. *Cell*, 131(5), pp.836–838.

Costantino, N. & Court, D.L., 2003. Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants. *Proceedings of the National Academy of Sciences*, 100(26), pp.15748–15753.

Court, D.L., Sawitzke, J.A. & Thomason, L.C., 2002. Genetic Engineering Using Homologous Recombination. *Annual Review of Genetics*, 36(1), pp.361–388.

Curth, U. et al., 1996. In Vitro and in Vivo Function of the C-Terminus of Escherichia Coli Single-Stranded DNA Binding Protein. *Nucleic Acids Research*, 24(14), pp.2706–2711.

Datta, S. et al., 2008. Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proceedings of the National Academy of Sciences*, 105(5), pp.1626 –1631.

Dekker, M. et al., 2011. Transient suppression of MLH1 allows effective single-nucleotide substitution by single-stranded DNA oligonucleotides. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 715(1–2), pp.52–60.

Dekker, M., Brouwers, C. & te Riele, H., 2003. Targeted gene modification in mismatch-repair-deficient embryonic stem cells by single-stranded DNA oligonucleotides. *Nucleic Acids Research*, 31(6), p.e27.

Disterer, P., Simons, J.P. & Owen, J.S., 2009. Validation of oligonucleotide-mediated gene editing. *Gene Therapy*, 16(6), pp.824–826.

Doyon, Y. et al., 2010. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nature Methods*, 8(1), pp.74–79.

Efimova, E. et al., 2003. IG20, a MADD Splice Variant, Increases Cell Susceptibility to γ-Irradiation and Induces Soluble Mediators That Suppress Tumor Cell Growth. *Cancer Research*, 63(24), pp.8768 –8776.

Eggler, A.L., Lusetti, S.L. & Cox, M.M., 2003. The C Terminus of the Escherichia coli RecA Protein Modulates the DNA Binding Competition with Single-stranded DNA-binding Protein. *Journal of Biological Chemistry*, 278(18), pp.16389–16396.

Ellis, H.M., Yu, D. & DiTizio, T., 2001. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proceedings of the National Academy of Sciences*, 98(12), p.6742.

Engstrom, J.U., Suzuki, T. & Kmiec, E.B., 2009. Regulation of targeted gene repair by intrinsic cellular processes. *BioEssays*, 31(2).

Erler, A. et al., 2009. Conformational Adaptability of Red[beta] during DNA Annealing and Implications for Its Structural Relationship with Rad52. *Journal of Molecular Biology*, 391(3), pp.586–598.

Freedman, M.L. et al., 2011. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet*, 43(6), pp.513–518.

Fu, Y. et al., 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, advance online publication. Available at: http://www.nature.com.ezp-prod1.hul.harvard.edu/nbt/journal/vaop/ncurrent/full/nbt.2623.html [Accessed July 17, 2013].

Furukohri, A. et al., 2012. Interaction between Escherichia coli DNA polymerase IV and single-stranded DNA-binding protein is required for DNA synthesis on SSB-coated DNA. *Nucleic Acids Research*, 40(13), pp.6039–6048.

Gibson, D.G. et al., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6, pp.343–345.

Grundberg, E. et al., 2009. Population genomics in a disease targeted primary cell model. *Genome Research*, 19(11), pp.1942–1952.

Hall, S.D. & Kolodner, R.D., 1994. Homologous pairing and strand exchange promoted by the Escherichia coli RecT protein. *Proceedings of the National Academy of Sciences*, 91(8), pp.3205–3209.

Handa, P., Acharya, N. & Varshney, U., 2001. Chimeras Between Single-Stranded DNA-Binding Proteins fromEscherichia Coli and Mycobacterium tuberculosisReveal That

Their C-Terminal Domains Interact with Uracil DNA Glycosylases. *Journal of Biological Chemistry*, 276(20), pp.16992–16997.

Hawkins, E.D. et al., 2007. Measuring lymphocyte proliferation, survival and differentiation using CFSE time-series data. *Nat. Protocols*, 2(9), pp.2057–2067.

Hobbs, M.D., Sakai, A. & Cox, M.M., 2007. SSB Protein Limits RecOR Binding Onto Single-Stranded DNA. *Journal of Biological Chemistry*, 282(15), pp.11058–11067.

Hockemeyer, D. et al., 2011. Genetic engineering of human pluripotent cells using TALE nucleases. *Nature Biotechnology*, 29(8), pp.731–734.

Huen, M.S.Y. et al., 2006. The involvement of replication in single stranded oligonucleotide-mediated gene repair. *Nucl. Acids Res.*, 34(21), pp.6183–6194.

Iyer, L., Koonin, E. & Aravind, L., 2002. Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics*, 3(1), p.8.

Jones, S.J.. et al., 2010. Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. *Genome Biology*, 11(8), p.R82.

Kantake, N. et al., 2002. Escherichia coli RecO protein anneals ssDNA complexed with its cognate ssDNA-binding protein: A common step in genetic recombination. *Proceedings of the National Academy of Sciences*, 99(24), pp.15327 –15332.

Karakousis, G. et al., 1998. The beta protein of phage λ binds preferentially to an intermediate in DNA renaturation. *Journal of Molecular Biology*, 276(4), pp.721–731.

Van Kessel, J.C. & Hatfull, G.F., 2008. Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. *Molecular Microbiology*, 67(5), pp.1094–1107.

Van Kessel, J.C. & Hatfull, G.F., 2007. Recombineering in Mycobacterium tuberculosis. *Nature Methods*, 4(2), pp.147–152.

Kmiec, E. & Holloman, W.K., 1981. Beta protein of bacteriophage lambda promotes renaturation of DNA. *Journal of Biological Chemistry*, 256(24), pp.12636–12639.

Kozlov, A.G., Cox, M.M. & Lohman, T.M., 2010. Regulation of Single-Stranded DNA Binding by the C Termini of Escherichia Coli Single-Stranded DNA-Binding (SSB) Protein. *Journal of Biological Chemistry*, 285(22), pp.17246–17252.

Lajoie, M.J. et al., 2012. Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic Acids Research*, 40(22), pp.e170–e170.

Langston, L.D., Indiani, C. & O'Donnell, M., 2009. Whither the replisome: emerging perspectives on the dynamic nature of the DNA replication machinery. *Cell Cycle (Georgetown, Tex.)*, 8(17), pp.2686–2691.

Lavery, P.E. & Kowalczykowski, S.C., 1992. Enhancement of recA protein-promoted DNA strand exchange activity by volume-occupying agents. *Journal of Biological Chemistry*, 267(13), pp.9307–9314.

Lee, W. et al., 2010. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature*, 465(7297), pp.473–477.

Li, X. et al., 2013. Bacterial DNA polymerases participate in oligonucleotide recombination. *Molecular Microbiology*, 88(5), pp.906–920.

Li, Z. et al., 1998. The beta protein of phage [lambda] promotes strand exchange. *Journal of Molecular Biology*, 276(4), pp.733–744.

Liu, C. et al., 2009. Cell Death Caused by Single-Stranded Oligodeoxynucleotide-Mediated Targeted Genomic Sequence Modification. *Oligonucleotides*, 19(3), pp.281–286.

Liu, L. et al., 2002. Rad51p and Rad54p, but not Rad52p, elevate gene repair in Saccharomyces cerevisiae directed by modified single-stranded oligonucleotide vectors. *Nucl. Acids Res.*, 30(13), pp.2742–2750.

Lu, D. & Keck, J.L., 2008. Structural basis of Escherichia coli single-stranded DNA-binding protein stimulation of exonuclease I. *Proceedings of the National Academy of Sciences*. Available at: http://www.pnas.org/content/early/2008/06/27/0800741105 [Accessed October 10, 2012].

Luisi-DeLuca, C. & Kolodner, R., 1994. Purification and Characterization of the Escherichia coli RecO Protein: Renaturation of Complementary Single-stranded DNA Molecules Catalyzed by the RecO Protein. *Journal of Molecular Biology*, 236(1), pp.124–138.

Lusetti, S.L., Shaw, J.J. & Cox, M.M., 2003. Magnesium Ion-dependent Activation of the RecA Protein Involves the C Terminus. *Journal of Biological Chemistry*, 278(18), pp.16381–16388.

Maeder, M.L. et al., 2009. Oligomerized pool engineering (OPEN): an “open-source” protocol for making customized zinc-finger arrays. *Nat. Protocols*, 4(10), pp.1471–1501.

Mali, P. et al., 2013. RNA-Guided Human Genome Engineering via Cas9. *Science*, 339(6121), pp.823–826.

Manfredi, C. et al., 2010. RecO-mediated DNA homology search and annealing is facilitated by SsbA. *Nucleic Acids Research*, 38(20), pp.6920 –6929.

Mardis, E.R., 2006. Anticipating the $1,000 genome. *Genome Biology*, 7(7), p.112.

Mazurek, A. et al., 2009. Sequence context effect for hMSH2-hMSH6 mismatch-dependent activation. *Proceedings of the National Academy of Sciences*, 106(11), pp.4177–4182.

McClellan, J. & King, M.-C., 2010. Genetic Heterogeneity in Human Disease. *Cell*, 141(2), pp.210–217.

McLean, C.Y. et al., 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotech*, 28(5), pp.495–501.

Melnikov, A. et al., 2012. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nature Biotechnology*, 30(3), pp.271–277.

Mer, G. et al., 2000. Structural Basis for the Recognition of DNA Repair Proteins UNG2, XPA, and RAD52 by Replication Factor RPA. *Cell*, 103(3), pp.449–456.

Metzker, M.L., 2010. Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11(1), pp.31–46.

Meyerson, M., Gabriel, S. & Getz, G., 2010. Advances in understanding cancer genomes through second-generation sequencing. *Nature Reviews Genetics*, 11(10), pp.685–696.

Mortensen, U.H. et al., 1996. DNA strand annealing is promoted by the yeast Rad52 protein. *Proceedings of the National Academy of Sciences*, 93(20), p.10729.

Mosberg, J.A. et al., 2012. Improving Lambda Red Genome Engineering in Escherichia coli via Rational Removal of Endogenous Nucleases. *PLoS ONE*, 7(9), p.e44638.

Mythili, E., Kumar, K.A. & Muniyappa, K., 1996. Characterization of the DNA-binding domain of [beta] protein, a component of phage [lambda] Red-pathway, by UV catalyzed cross-linking. *Gene*, 182(1-2), pp.81–87.

Naue, N. et al., 2013. The helicase-binding domain of Escherichia coli DnaG primase interacts with the highly conserved C-terminal region of single-stranded DNA-binding protein. *Nucleic Acids Research*, 41(8), pp.4507–4517.

New, J.H. et al., 1998. Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A. *Nature*, 391(6665), pp.407–410.

Noirot, P. et al., 2003. Hallmarks of homology recognition by RecA-like recombinases are exhibited by the unrelated Escherichia coli RecT protein. *The EMBO Journal*, 22(2), pp.324–334.

Noirot, P. & Kolodner, R.D., 1998. DNA Strand Invasion Promoted by Escherichia coli RecT Protein. *Journal of Biological Chemistry*, 273(20), pp.12274–12280.

Olsen, P. et al., 2009. Cellular responses to targeted genomic sequence modification using single-stranded oligonucleotides and zinc-finger nucleases. *DNA repair*, 8(3), pp.298–308.

Olsen, P.A., Randøl, M., Luna, L., et al., 2005. Genomic sequence correction by single-stranded DNA oligonucleotides: role of DNA synthesis and chemical modifications of the oligonucleotide ends. *The Journal of Gene Medicine*, 7(12), pp.1534–1544.

Olsen, P.A., Randøl, M. & Krauss, S., 2005. Implications of cell cycle progression on functional sequence correction by short single-stranded DNA oligonucleotides. *Gene Ther*, 12(6), pp.546–551.

Papaioannou, I., Disterer, P. & Owen, J.S., 2009. Use of internally nuclease-protected single-strand DNA oligonucleotides and silencing of the mismatch repair protein, MSH2, enhances the replication of corrected cells following gene editing. *The Journal of Gene Medicine*, 11(3), pp.267–274.

Passy, S.I. et al., 1999. Rings and filaments of beta protein from bacteriophage lambda suggest a superfamily of recombination proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 96(8), pp.4279–4284.

Pattanayak, V. et al., 2011. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nature Methods*, 8(9), pp.765–770.

Patwardhan, R.P. et al., 2012. Massively parallel functional dissection of mammalian enhancers in vivo. *Nature Biotechnology*, 30(3), pp.265–270.

Pickrell, J.K. et al., 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*, 464(7289), pp.768–772.

Van Pijkeren, J.-P. et al., 2012. Exploring optimization parameters to increase ssDNA recombineering in Lactococcus lactis and Lactobacillus reuteri. *Bioengineered*, 3(4), pp.209–217.

Van Pijkeren, J.-P. & Britton, R.A., 2012. High Efficiency Recombineering in Lactic Acid Bacteria. *Nucleic Acids Research*, 40(10), pp.e76–e76.

Poteete, A.R., 2001. What makes the bacteriophage λ Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiology Letters*, 201(1), pp.9–14.

Radecke, S. et al., 2006. Physical incorporation of a single-stranded oligodeoxynucleotide during targeted repair of a human chromosomal locus. *The Journal of Gene Medicine*, 8(2), pp.217–228.

Reinhardt, P. et al., 2013. Genetic Correction of a LRRK2 Mutation in Human iPSCs Links Parkinsonian Neurodegeneration to ERK-Dependent Changes in Gene Expression. *Cell Stem Cell*, 12(3), pp.354–367.

Reyon, D. et al., 2012. FLASH assembly of TALENs for high-throughput genome editing. *Nature Biotechnology*, 30(5), pp.460–465.

Rios, X. et al., 2012. Stable Gene Targeting in Human Cells Using Single-Strand Oligonucleotides with Modified Bases. *PLoS ONE*, 7(5), p.e36697.

Rybalchenko, N. et al., 2004. Strand invasion promoted by recombination protein β of coliphage λ. *Proceedings of the National Academy of Sciences of the United States of America*, 101(49), pp.17056–17060.

Ryzhikov, M. et al., 2011. Mechanism of RecO recruitment to DNA by single-stranded DNA binding protein. *Nucleic Acids Research*, 39(14), pp.6305 –6314.

Sawitzke, J.A. et al., 2011. Probing Cellular Processes with Oligo-Mediated Recombination and Using the Knowledge Gained to Optimize Recombineering. *Journal of Molecular Biology*, 407(1), pp.45–59.

Schaub, M.A. et al., 2012. Linking disease associations with regulatory information in the human genome. *Genome Research*, 22(9), pp.1748–1759.

Senn, J.J., Burel, S. & Henry, S.P., 2005. Non-CpG-Containing Antisense 2′-Methoxyethyl Oligonucleotides Activate a Proinflammatory Response Independent of Toll-Like Receptor 9 or Myeloid Differentiation Factor 88. *Journal of Pharmacology and Experimental Therapeutics*, 314(3), pp.972 –979.

Shereda, R.D. et al., 2008. SSB as an Organizer/Mobilizer of Genome Maintenance Complexes. *Critical Reviews in Biochemistry and Molecular Biology*, 43(5), pp.289–318.

Shinohara, A. et al., 1998. Rad52 forms ring structures and co-operates with RPA in single-strand DNA annealing. *Genes to Cells*, 3(3), pp.145–156.

Shinohara, A. & Ogawa, T., 1998. Stimulation by Rad52 of yeast Rad51- mediated recombination. *Nature*, 391(6665), pp.404–407.

Sikorav, J.-L. & Church, G.M., 1991. Complementary recognition in condensed DNA: Accelerated DNA renaturation. *Journal of Molecular Biology*, 222(4), pp.1085–1108.

Soldner, F. et al., 2011. Generation of Isogenic Pluripotent Stem Cells Differing Exclusively at Two Early Onset Parkinson Point Mutations. *Cell*, 146(2), pp.318–331.

Şöllü, C. et al., 2010. Autonomous Zinc-Finger Nuclease Pairs for Targeted Chromosomal Deletion. *Nucleic Acids Research*. Available at: http://nar.oxfordjournals.org/content/early/2010/08/13/nar.gkq720 [Accessed March 21, 2012].

Sugiyama, T., New, J.H. & Kowalczykowski, S.C., 1998. DNA annealing by Rad52 Protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. *Proceedings of the National Academy of Sciences*, 95(11), pp.6049–6054.

Sung, P., 1997. Function of Yeast Rad52 Protein as a Mediator between Replication Protein A and the Rad51 Recombinase. *Journal of Biological Chemistry*, 272(45), pp.28194–28197.

Sur, I.K. et al., 2012. Mice Lacking a Myc Enhancer That Includes Human SNP rs6983267 Are Resistant to Intestinal Tumors. *Science*, 338(6112), pp.1360–1363.

Swingle, B., Markel, E., et al., 2010. Oligonucleotide recombination in Gram-negative bacteria. *Molecular Microbiology*, 75(1), pp.138–148.

Swingle, B., Bao, Z., et al., 2010. Recombineering Using RecTE from Pseudomonas syringae. *Appl. Environ. Microbiol.*, 76(15), pp.4960–4968.

Thresher, R.J. et al., 1995. Electron Microscopic Visualization of RecT Protein and its Complexes with DNA. *Journal of Molecular Biology*, 254(3), pp.364–371.

Urnov, F.D. et al., 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, 435(7042), pp.646–651.

Vasquez, K.M. et al., 2001. Manipulating the mammalian genome by homologous recombination. *Proceedings of the National Academy of Sciences*, 98(15), pp.8403 – 8410.

Vernot, B. et al., 2012. Personal and population genomics of human regulatory variation. *Genome Research*, 22(9), pp.1689–1697.

Wang, H. et al., 2013. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell*, 153(4), pp.910–918.

Wang, H.H. et al., 2011. Modified Bases Enable High-Efficiency Oligonucleotide-Mediated Allelic Replacement Via Mismatch Repair Evasion. *Nucleic Acids Research*. Available at: http://nar.oxfordjournals.org/content/early/2011/05/23/nar.gkr183 [Accessed March 21, 2012].

Wang, H.H. et al., 2009. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 460(7257), pp.894–898.

Wang, J. et al., 2012. Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Research*, 22(7), pp.1316–1326.

Wang, M. et al., 2000. Species specificity of human RPA in simian virus 40 DNA replication lies in T-antigen-dependent RNA primer synthesis. *Nucl. Acids Res.*, 28(23), pp.4742–4749.

Wu, X.S. et al., 2005. Increased efficiency of oligonucleotide-mediated gene repair through slowing replication fork progression. *Proceedings of the National Academy of Sciences*, 102(7), pp.2508–2513.

Wu, Z. et al., 2006. Domain Structure and DNA Binding Regions of β Protein from Bacteriophage λ. *Journal of Biological Chemistry*, 281(35), pp.25205–25214.

Yin, W.X. et al., 2005. Targeted correction of a chromosomal point mutation by modified single-stranded oligonucleotides in a GFP recovery system. *Biochemical and Biophysical Research Communications*, 334(4), pp.1032–1041.

Zhang, X. et al., 2011. Cutting Edge: Ku70 Is a Novel Cytosolic DNA Sensor That Induces Type III Rather Than Type I IFN. *The Journal of Immunology*, 186(8), pp.4541 –4545.

Zhang, Y. et al., 2003. Phage annealing proteins promote oligonucleotide-directed mutagenesis in Escherichia coli and mouse ES cells. *BMC Molecular Biology*, 4(1), p.1.

Zhang, Z. et al., 2007. Homology-driven chromatin remodeling by human RAD54. *Nature Structural & Molecular Biology*, 14(5), p.397.

APPENDIX ONE

RNAseq top 400 differentially regulated genes

Gene	Chromosome	Description	GFP+ /GFP-
SPINK5	chr5	serine peptidase inhibitor, Kazal type 5	0.178708
RLIM	chrX	ring finger protein, LIM domain interacting	0.224445
EIF5B	chr2	eukaryotic translation initiation factor 5B	0.280111
CR602933	chr5	full-length cDNA clone CS0DC022YF01 of Neuroblastoma Cot 25-normalized of Homo sapiens (human).	0.308464
MT-TS1	chrM	tRNA serine 1	0.311688
TOP2A	chr17	Homo sapiens topoisomerase II alpha-2 (TOP2A) mRNA, partial cds.	0.325499
HSPCA	chr14	SubName: Full=Full-length cDNA clone CS0CAP007YF18 of Thymus of Homo sapiens (human);	0.325829
IRS2	chr13	insulin receptor substrate 2	0.330785
EIF2S2	chr20	eukaryotic translation initiation factor 2 beta	0.361353
RPL17	chr18	ribosomal protein L17	0.362738
HSP90AA1	chr14	heat shock 90kDa protein 1, alpha	0.365069
RPS26P11	chrX	Homo sapiens ribosomal protein S26 pseudogene 11 (RPS26P11), non-coding RNA.	0.366195
MT-TK	chrM	tRNA lysine	0.376643
DDX21	chr10	DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	0.392613
RPL32	chr3	ribosomal protein L32	0.429197
NUCKS1	chr1	nuclear casein kinase and cyclin-dependent	0.430597
RPL26	chr17	ribosomal protein L26	0.462271
FTH1	chr11	ferritin, heavy polypeptide 1	0.497557
RPS24	chr10	Homo sapiens full length insert cDNA clone YB24C12.	0.499453
RPL10	chrX	SubName: Full=Ribosomal protein L10; Flags: Fragment;	0.517146
PPIA	chr7	peptidylprolyl isomerase A	0.558573
RPL35	chr9	ribosomal protein L35	0.577786
OK/SW-cl.46	chr16	RecName: Full=60S ribosomal protein L13;	0.582919
EC45	chr3	RecName: Full=Ribosomal protein L15;	0.585559
RPL13	chr16	ribosomal protein L13	0.589934
RPS14	chr5	ribosomal protein S14	0.592602
AlphaTFEB	chr11	Homo sapiens clone alpha1 mRNA sequence.	0.596378
RPL15	chr3	ribosomal protein L15	0.604601
MT-RNR2	chrM	16S ribosomal RNA	0.609288
RPL11	chr1	SubName: Full=Ribosomal protein L11; Flags: Fragment;	0.617396

RNAseq top 400 differentially regulated genes (Continued)

MT-ND4	chrM	NADH dehydrogenase subunit 4	0.652922
BC018860	chr1	Homo sapiens, clone IMAGE:3141568, mRNA.	0.654636
AB055772	chr17	Homo sapiens mRNA for ribosomal protein S2, partial cds.	0.665313
MT-CO2	chrM	Cytochrome c oxidase subunit II	0.677766
MT-ND6	chrM	NADH dehydrogenase subunit 6	0.698534
MT-CO1	chrM	Cytochrome c oxidase subunit I	0.700384
MT-CYB	chrM	Cytochrome b	0.713237
MT-TF	chrM	tRNA phenylalanine	0.72074
OK/KNS-cl.6	chr16	Homo sapiens OK/KNS-cl.6 mRNA for ribosomal protein S2, complete cds.	0.732352
MT-RNR1	chrM	12S ribosomal RNA	0.733956
ACTG1chr7-	chr7	actin, gamma 1 propeptide	0.736113
RPS2	chr16	ribosomal protein S2	0.736594
ACTB	chr7	SubName: Full=cDNA FLJ32030 fis, clone NTONG2000040, highly similar to Actin, alpha cardiac;	0.746562
MT-ATP6	chrM	ATP synthase F0 subunit 6	0.766903
EEF2	chr19	eukaryotic translation elongation factor 2	1.481708
RPLP1	chr15	ribosomal protein P1	1.53848
DKFZp686L04275	chr12	SubName: Full=Putative uncharacterized protein DKFZp686L04275; Flags: Fragment;	1.586871
TMSB10	chr2	thymosin, beta 10	1.615865
FDPS	chr1	farnesyl diphosphate synthase	1.65258
TUBBchr6_cox_hap2+	chr6_cox_hap2	tubulin, beta	1.662456
UBBchr17+	chr17	ubiquitin B precursor	1.668873
TUBA1B	chr12	tubulin, alpha, ubiquitous	1.672522
C17orf45	chr17	Homo sapiens cDNA FLJ25777 fis, clone TST06567.	1.741873
ACTG1chr17-	chr17	actin, gamma 1 propeptide	1.744692
TUBBchr6+	chr6	tubulin, beta	1.755095
TUBBchr6_dbb_hap3+	chr6_dbb_hap3	tubulin, beta	1.755194
TUBBchr6_mann_hap4+	chr6_mann_hap4	tubulin, beta	1.755194
TUBBchr6_mcf_hap5+	chr6_mcf_hap5	tubulin, beta	1.755194
TUBBchr6_ssto_hap7+	chr6_ssto_hap7	tubulin, beta	1.755194
TUBBchr6_apd_hap1+	chr6_apd_hap1	tubulin, beta	1.755353
TUBBchr6_qbl_hap6+	chr6_qbl_hap6	tubulin, beta	1.755353
A121/SUI1	chr17	Homo sapiens SUI1 isolog mRNA, complete cds.	1.986967
EIF1	chr17	eukaryotic translation initiation factor 1	1.999534
TUBB2Cchr6_cox_hap2+	chr6_cox_hap2	tubulin, beta, 2	2.087218
TUBB2Cchr6+	chr6	tubulin, beta, 2	2.087372
TUBB2Cchr6_dbb_hap3+	chr6_dbb_hap3	tubulin, beta, 2	2.087485

RNAseq top 400 differentially regulated genes (Continued)

TUBB2Cchr6_mann_hap4+	chr6_mann_hap4	tubulin, beta, 2	2.087485
TUBB2Cchr6_mcf_hap5+	chr6_mcf_hap5	tubulin, beta, 2	2.087485
TUBB2Cchr6_ssto_hap7+	chr6_ssto_hap7	tubulin, beta, 2	2.087485
TUBB2Cchr6_apd_hap1+	chr6_apd_hap1	tubulin, beta, 2	2.087752
TUBB2Cchr6_qbl_hap6+	chr6_qbl_hap6	tubulin, beta, 2	2.087752
EEF1A2chr20-	chr20	eukaryotic translation elongation factor 1 alpha	2.137521
EDARADD	chr1	EDAR-associated death domain	2.690573
FLNA	chrX	filamin A, alpha	2.823647
AK307158	chr16	Homo sapiens cDNA, FLJ97106.	2.936142
TAIF	chr16	Homo sapiens cDNA FLJ55649 complete cds, highly similar to Homo sapiens interleukin 32 (IL32), transcript variant 6, mRNA.	2.965964
HLA-Bchr6_qbl_hap6+	chr6_qbl_hap6	SubName: Full=MHC class I antigen; Flags: Fragment;	3.510968
DDTchr22-	chr22	D-dopachrome tautomerase	0.530275
RPS6	chr9	ribosomal protein S6	0.531476
RPS4X	chrX	ribosomal protein S4, X-linked X	0.574014
AF113016	chr11	Homo sapiens clone alpha_est218/52C1 mRNA sequence.	0.558622
RPL21	chr13	SubName: Full=Ribosomal protein L21 variant; Flags: Fragment;	0.450594
HSPA1Achr6_qbl_hap6+	chr6_qbl_hap6	heat shock 70kDa protein 1A	2.603675
AK298056chr6_qbl_hap6-	chr6_qbl_hap6	Homo sapiens cDNA FLJ54370 complete cds, highly similar to Heat shock 70 kDa protein 1.	2.604717
HSPA1Achr6_cox_hap2+	chr6_cox_hap2	heat shock 70kDa protein 1A	2.596007
AK298056chr6_cox_hap2-	chr6_cox_hap2	Homo sapiens cDNA FLJ54370 complete cds, highly similar to Heat shock 70 kDa protein 1.	2.597032
RPS8	chr1	ribosomal protein S8	0.557759
FASN	chr17	fatty acid synthase	2.417843
BAT2D1	chr1	HBxAg transactivated protein 2	0.329105
HLA-Achr6_qbl_hap6+	chr6_qbl_hap6	SubName: Full=MHC class I antigen;	2.184954
ZNF90	chr19	zinc finger protein 90	0.492887
RPS16	chr19	SubName: Full=cDNA FLJ56786, moderately similar to 40S ribosomal protein S16;	0.690873
BAZ1B	chr7	bromodomain adjacent to zinc finger domain, 1B	0.349469
HSPA1Achr6_apd_hap1+	chr6_apd_hap1	heat shock 70kDa protein 1A	2.50454
AK298056chr6_apd_hap1-	chr6_apd_hap1	Homo sapiens cDNA FLJ54370 complete cds, highly similar to Heat shock 70 kDa protein 1.	2.50539
XTP2	chr1	Homo sapiens BAT2-iso mRNA, partial cds.	0.264456
HSPA1Achr6_dbb_hap3+	chr6_dbb_hap3	heat shock 70kDa protein 1A	2.523019
AK298056chr6_dbb_hap3-	chr6_dbb_hap3	Homo sapiens cDNA FLJ54370 complete cds, highly similar to Heat shock 70 kDa protein 1.	2.523943
IL32	chr16	interleukin 32	1.866555
THOC2	chrX	Homo sapiens Tho2 mRNA, complete cds.	0.280104

RNAseq top 400 differentially regulated genes (Continued)

HSPA1Achr6+	chr6	heat shock 70kDa protein 1A	2.506691
AK298056chr6-	chr6	Homo sapiens cDNA FLJ54370 complete cds, highly similar to Heat shock 70 kDa protein 1.	2.507582
TUBA1Cchr12+	chr12	tubulin alpha 6	1.480362
C1orf104	chr1	Homo sapiens cDNA FLJ35976 fis, clone TESTI2013427.	0.483526
DL491750	chr8	DL491750	2.587952
SCD	chr10	stearoyl-CoA desaturase 1	1.841176
MT-ND2	chrM	NADH dehydrogenase subunit 2	0.840077
ENO1	chr1	RecName: Full=Alpha-enolase; EC=4.2.1.11; AltName: Full=2-phospho-D-glycerate hydro-lyase; AltName: Full=Non-neural enolase; Short=NNE; AltName: Full=Enolase 1; AltName: Full=Phosphopyruvate hydratase; AltName: Full=C-myc promoter-binding protein; AltName: Full=MBP-1; AltName: Full=MPB-1; AltName: Full=Plasminogen-binding protein;	1.356919
GOLGA4	chr3	SubName: Full=GOLGA4 protein;	0.204172
BC009321/IMPDH2	chr3	Homo sapiens cDNA clone IMAGE:4123521, **** WARNING: chimeric clone ****.	3.914188
RPL41	chr12	ribosomal protein L41	0.745043
HSPA1Bchr6_qbl_hap6+	chr6_qbl_hap6	heat shock 70kDa protein 1B	1.931597
HSPA1Bchr6_cox_hap2+	chr6_cox_hap2	heat shock 70kDa protein 1B	1.934707
KIAA1532	chr19	SubName: Full=Putative uncharacterized protein ENSP00000371624;	3.021056
MT-ND5	chrM	NADH dehydrogenase subunit 5	0.812256
MAN1A2	chr1	SubName: Full=Putative uncharacterized protein MAN1A2;	0.317346
TUBA1Achr12+	chr12	Synthetic construct Homo sapiens gateway clone IMAGE:100019466 3' read TUBA1A mRNA.	1.450076
CR610533	chr8	SubName: Full=Class IVb beta tubulin;	2.848917
RPLP0chr2+	chr2	SubName: Full=RPLP0 protein;	1.464575
LOC401308	chr11	SubName: Full=Similar to Chain , Heat-Shock Cognate 70kd Protein (44kd Atpase N-Terminal) (E.C.3.6.1.3) Mutant With Asp 206 Replaced By Ser (D206s);	2.002414
MT-ND3	chrM	NADH dehydrogenase subunit 3	1.164973
RPLP2	chr11	ribosomal protein P2	1.440421
RAC1	chr7	ras-related C3 botulinum toxin substrate 1	0.436073
HSPA1Bchr6_apd_hap1+	chr6_apd_hap1	heat shock 70kDa protein 1B	1.918542
DQ597482	chr11	full-length cDNA clone CS0DM002YL19 of Fetal liver of Homo sapiens (human).	0.581733
RPL14	chr3	ribosomal protein L14	0.562145
YWHAB	chr20	tyrosine 3-monooxygenase/tryptophan	0.409504
HSPA1Bchr6_dbb_hap3+	chr6_dbb_hap3	heat shock 70kDa protein 1B	1.921091
CR624826	chr1	Homo sapiens cDNA: FLJ20887 fis, clone ADKA03276.	0.378406

RNAseq top 400 differentially regulated genes (Continued)

EDF1	chr9	endothelial differentiation-related factor 1	1.5108
hCAP-C	chr3	Homo sapiens hCAP-C mRNA for chromosome-associated polypeptide-C, complete cds.	0.260651
HSPA1Bchr6+	chr6	heat shock 70kDa protein 1B	1.910128
SFRS11	chr1	splicing factor, arginine/serine-rich 11	0.298836
DCI	chr16	dodecenoyl-Coenzyme A delta isomerase precursor	0.389571
SSRP1	chr11	structure specific recognition protein 1	0.462444
RPN2	chr20	ribophorin II	0.509197
RBM12	chr20	RNA binding motif protein 12	3.08718
KTN1	chr14	kinectin 1	0.29067
KCP	chr7	Homo sapiens cDNA FLJ33365 fis, clone BRACE2005460, moderately similar to Xenopus laevis mRNA for Kielin.	2.692518
ANXA2chr9+	chr9	annexin A2	1.694536
HMGA1	chr6	SubName: Full=cDNA FLJ54188, moderately similar to High mobility group protein HMG-I/HMG-Y;	1.740566
C15orf63	chr15	chromosome 15 open reading frame 63	1.549347
LOC100129034	chr9	Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 685610.	0.453213
DHCR7	chr11	7-dehydrocholesterol reductase	2.788704
SMC4	chr3	SMC4 structural maintenance of chromosomes	0.344823
EIF4A1	chr17	eukaryotic translation initiation factor 4A	0.585821
ALDOA	chr16	fructose-bisphosphate aldolase A	1.372598
FAM166A	chr9	hypothetical protein LOC401565	0.506966
HSPA8	chr11	heat shock 70kDa protein 8	1.524636
BBX	chr3	HMG-BOX transcription factor BBX	0.358123
MRPS5	chr2	Homo sapiens MRPS5 mRNA for mitochondrial ribosomal protein S5, complete cds.	0.440919
ARF1	chr1	ADP-ribosylation factor 1	2.145166
RPL10A	chr6	ribosomal protein L10a	0.604428
TARG1	chr8	Homo sapiens TARG1 mRNA, 3' untranslated region.	0.274394
MAGED2	chrX	melanoma antigen family D, 2	2.742964
MDH2	chr7	mitochondrial malate dehydrogenase precursor	1.972453
rpl10a	chr6	SubName: Full=HCG1787790; SubName: Full=Putative uncharacterized protein ENSP00000344077;	0.6073
GPI	chr19	glucose phosphate isomerase	1.852699
LASP1	chr17	SubName: Full=cDNA FLJ51834, highly similar to LIM and SH3 domain protein 1;	2.397447
SNORD49B	chr17	Homo sapiens small nucleolar RNA, C/D box 49B (SNORD49B), non-coding RNA.	2.121228
JUP	chr17	junction plakoglobin	2.144249
KIAA0391	chr14	mitochondrial RNase P protein 3 precursor	0.341218
EBP	chrX	emopamil binding protein (sterol isomerase)	1.468221

RNAseq top 400 differentially regulated genes (Continued)

MTDH	chr8	metadherin	0.316266
POLR2A	chr17	DNA-directed RNA polymerase II A	2.373159
SLC25A5	chrX	adenine nucleotide translocator 2	2.074109
CHD4	chr12	chromodomain helicase DNA binding protein 4	0.498823
ACLY	chr17	ATP citrate lyase	1.991025
DDX46	chr5	DEAD (Asp-Glu-Ala-Asp) box polypeptide 46	0.351844
TMSB4X	chrX	thymosin, beta 4	0.726951
FASN variant protein	chr17	Homo sapiens mRNA for FASN variant protein, partial cds, clone: ae00014.	2.234287
ASS1	chr9	argininosuccinate synthetase 1	0.737727
CSRP1	chr1	cysteine and glycine-rich protein 1	1.713545
METAP2	chr12	methionyl aminopeptidase 2	0.440387
VARSchr6_qbl_hap6-	chr6_qbl_hap6	SubName: Full=Valyl-tRNA synthetase; SubName: Full=Valyl-tRNA synthetase	3.845343
GAPD	chr12	RecName: Full=Glyceraldehyde 3-phosphate dehydrogenase; EC=1.2.1.12; Flags: Fragment;	0.822593
EIF4G2	chr11	eukaryotic translation initiation factor 4	1.877482
SNX22	chr15	Homo sapiens cDNA FLJ13952 fis, clone Y79AA1001068.	0.541472
GAPDH	chr12	glyceraldehyde-3-phosphate dehydrogenase	0.823469
GOLGB1	chr3	golgi autoantigen, golgin subfamily b	0.236078
MT-TE	chrM	tRNA glutamic acid	0.478702
COPA	chr1	Synthetic construct DNA, clone: pF1KB9459, Homo sapiens COPA gene for coatomer subunit alpha, complete cds, without stop codon, in Flexi system.	2.170367
SNORD101	chr6	Homo sapiens cDNA FLJ38828 fis, clone MAMGL1000017.	0.350372
MRPS15	chr1	mitochondrial ribosomal protein S15 precursor	0.465634
MAP7D3	chrX	MAP7 domain containing 3	0.255073
HSPA1Lchr6_cox_hap2-	chr6_cox_hap2	heat shock 70kDa protein 1-like	4.734759
HLA-Cchr6_mann_hap4-	chr6_mann_hap4	Homo sapiens HLA-C class I antigen (HLA-C) mRNA, HLA-Cw*05DZ allele, complete cds.	2.107054
HSPA1Lchr6_qbl_hap6-	chr6_qbl_hap6	heat shock 70kDa protein 1-like	4.719672
EPRS	chr1	glutamyl-prolyl tRNA synthetase	0.530624
CENPF	chr1	centromere protein F	0.348649
CCT7	chr2	chaperonin containing TCP1, subunit 7	1.931585
NDUFC2	chr11	NADH dehydrogenase (ubiquinone) 1, subcomplex	2.056072
ARF3	chr12	ADP-ribosylation factor 3	1.956623
hCG_1994130	chr16	SubName: Full=HCG1994130	0.689145
OK/SW-cl.82	chr16	Homo sapiens OK/SW-cl.82 mRNA, complete cds.	0.689441
SURF4	chr9	surfeit 4	1.884385
QM	chrX	Homo sapiens mRNA expressed only in placental villi, clone SMAP26.	2.269846
TM4SF1	chr3	transmembrane 4 superfamily member 1	0.750015

RNAseq top 400 differentially regulated genes (Continued)

CBX3	chr7	SubName: Full=Chromobox homolog 3 (HP1 gamma homolog, Drosophila)	0.638467
AK303306	chr17	SubName: Full=cDNA FLJ51427;	3.142034
RPS15A	chr16	ribosomal protein S15a	0.693197
RPSAchr3+	chr3	ribosomal protein SA	0.601876
M37726	chr1	M37726	0.401813
LYRIC	chr8	Homo sapiens astrocyte elevated gene-1 AEG1 mRNA, complete cds.	0.279476
RPL29	chr3	ribosomal protein L29	0.659075
CPNE1	chr20	Homo sapiens mRNA for copine I variant protein.	1.86948
SLC3A2	chr11	solute carrier family 3, member 2	1.83348
EEF1Gchr11-	chr11	eukaryotic translation elongation factor 1	0.717383
DDB1	chr11	damage-specific DNA binding protein 1	2.231589
PABPC1	chr8	Synthetic construct DNA, clone: pF1KB5695, Homo sapiens PABPC1 gene for polyadenylate-binding protein 1, complete cds, without stop codon, in Flexi system.	0.601032
TAX1BP1	chr7	Tax1 (human T-cell leukemia virus type I)	0.46841
FSCN1	chr7	fascin 1	1.793611
KRT80	chr12	keratin 80	2.548215
NOSIP	chr19	nitric oxide synthase interacting protein	1.911829
VPS72	chr1	transcription factor-like 1	2.614315
CCT4	chr2	RecName: Full=T-complex protein 1, delta subunit;	1.90837
TIMP1	chrX	SubName: Full=cDNA FLJ60276, weakly similar to Metalloproteinase inhibitor 1;	1.559861
KIAA0765	chr20	Homo sapiens mRNA for KIAA0765 protein, partial cds.	1.86527
BAT3chr6-	chr6	HLA-B associated transcript-3	2.476822
EIF4H	chr7	eukaryotic translation initiation factor 4H	1.969434
UBXD2	chr2	SubName: Full=cDNA FLJ38083 fis, clone CTONG2016408, highly similar to UBX domain-containing protein 2;	0.29811
OK/SW-cl.110	chrX	Homo sapiens OK/SW-cl.110 mRNA for phosphoglycerete kinase 1, complete cds.	1.776271
PGK1	chrX	phosphoglycerate kinase 1	1.776271
BAT3chr6_mcf_hap5-	chr6_mcf_hap5	HLA-B associated transcript-3	2.480334
BAT3chr6_dbb_hap3-	chr6_dbb_hap3	HLA-B associated transcript-3	2.478902
BAT3chr6_mann_hap4-	chr6_mann_hap4	HLA-B associated transcript-3	2.478902
CLSPN	chr1	claspin	0.339279
CALU	chr7	calumenin	0.58209
NDUFB4	chr3	SubName: Full=Putative uncharacterized protein ENSP00000244249; Flags: Fragment;	1.734517
SMC3	chr10	structural maintenance of chromosomes 3	0.355898

RNAseq top 400 differentially regulated genes (Continued)

PSMA6	chr14	SubName: Full=cDNA FLJ52022, highly similar to Proteasome subunit alpha type 6 (EC 3.4.25.1); SubName: Full=cDNA, FLJ79122, highly similar to Proteasome subunit alpha type 6 (EC 3.4.25.1);	0.41919
HLA-Cwchr6_mann_hap4-	chr6_mann_hap4	SubName: Full=MHC class I antigen;	2.050767
SFRS2IP	chr12	splicing factor, arginine/serine-rich 2	0.307262
CTSZ	chr20	cathepsin Z preproprotein	0.423975
GHITM	chr10	growth hormone inducible transmembrane protein	1.981593
HLA-Bchr6_mcf_hap5-	chr6_mcf_hap5	SubName: Full=HLA-B protein; Flags: Fragment;	1.700477
DDX27	chr20	DEAD (Asp-Glu-Ala-Asp) box polypeptide 27	0.425882
HLA-Cchr6_mcf_hap5-	chr6_mcf_hap5	Homo sapiens HLA-C class I antigen (HLA-C) mRNA, HLA-Cw*05DZ allele, complete cds.	2.159621
SEC62	chr3	translocation protein 1	0.3955
CTSD	chr11	cathepsin D preproprotein	1.975589
HSP90B1	chr12	SubName: Full=HSP90B1 protein;	0.555335
MT-ND1	chrM	NADH Dehydrogenase subunit 1	0.861913
LARP7	chr4	La ribonucleoprotein domain family, member 7	0.279
PRDX6	chr1	peroxiredoxin 6	1.630754
SND1	chr7	staphylococcal nuclease domain containing 1	1.792087
ATRX	chrX	transcriptional regulator ATRX	0.283264
BRD2chr6_dbb_hap3+	chr6_dbb_hap3	SubName: Full=Bromodomain containing 2; SubName: Full=Bromodomain containing 2	0.448596
CALR	chr19	calreticulin precursor	1.378053
TAF1D	chr11	Homo sapiens cDNA: FLJ23363 fis, clone HEP15507.	0.481044
FUS	chr16	fusion (involved in t(12;16) in malignant	1.994501
BAT2chr6_cox_hap2+	chr6_cox_hap2	HLA-B associated transcript-2	2.079076
DYNLL1	chr12	dynein light chain 1	1.78614
GARS	chr7	glycyl-tRNA synthetase	0.5643
RBM25	chr14	RecName: Full=RNA-binding protein 25; AltName: Full=RNA-binding motif protein 25; AltName: Full=RNA-binding region-containing protein 7; AltName: Full=Arg/Glu/Asp-rich protein of 120 kDa; Short=RED120; AltName: Full=Protein S164;	0.431193
9-Sep	chr17	septin 9	2.284401
FUS/CHOP	chr16	SubName: Full=cDNA FLJ57206, moderately similar to RNA-binding protein FUS;	1.989927
PTP4A2	chr1	protein tyrosine phosphatase type IVA, member 2	0.401035
LRRFIP1	chr2	leucine rich repeat (in FLII) interacting	0.288969
AKAP9	chr7	A-kinase anchor protein 9	0.197231
DNAJC2	chr7	SubName: Full=DNAJC2 protein;	0.349298
SON	chr21	Homo sapiens SON DNA binding protein	0.515389
UBXN4	chr2	UBX domain containing 2	0.353071
RPS26	chr12	ribosomal protein S26	0.605091

RNAseq top 400 differentially regulated genes (Continued)

GPR56	chr16	SubName: Full=G protein-coupled receptor 56;	2.908283
DEK	chr6	DEK oncogene	0.417318
SPOP	chr17	speckle-type POZ protein	0.304696
UBA1	chrX	ubiquitin-activating enzyme E1	1.582153
Y16709	chr9	Homo sapiens mRNA from HIV associated non-Hodgkin's lymphoma (clone hl1-98).	0.435839
RPL27A	chr11	ribosomal protein L27a	0.795919
RSF1	chr11	remodeling and spacing factor 1	0.270706
PRDX5	chr11	peroxiredoxin 5	1.775891
BAT2chr6_dbb_hap3+	chr6_dbb_hap3	HLA-B associated transcript-2	2.068963
BAT2chr6_mann_hap4+	chr6_mann_hap4	HLA-B associated transcript-2	2.079904
CDKN1A	chr6	cyclin-dependent kinase inhibitor 1A	1.890415
RNASEK	chr17	ribonuclease kappa	1.630251
L01117	chr17	L01117	3.367235
PRPF8	chr17	U5 snRNP-specific protein	1.814369
BSG	chr19	Homo sapiens mRNA for CD147, complete cds.	1.478747
hEMMPRIN	chr19	Homo sapiens mRNA for EMMPRIN, complete cds.	1.478107
RPL12	chr9	Homo sapiens mRNA for ribosomal protein L12 variant protein.	1.327549
PKM2	chr15	Homo sapiens Opa-interacting protein OIP3 mRNA, partial cds.	1.293158
PLEC1	chr8	plectin	1.998906
NDUFB9	chr8	NADH dehydrogenase (ubiquinone) 1 beta	1.549053
AB209575	chr19	SubName: Full=Glucose phosphate isomerase variant; Flags: Fragment;	3.185948
C6orf21 variant proteinchr6_ssto_hap7+	chr6_ssto_hap7	Homo sapiens mRNA for C6orf21 variant protein, partial cds, clone: hj06729.	3.998911
GPATCH4	chr1	G patch domain containing 4	0.384625
RIOK1	chr6	RIO kinase 1	0.228556
C6orf21 variant proteinchr6_cox_hap2+	chr6_cox_hap2	Homo sapiens mRNA for C6orf21 variant protein, partial cds, clone: hj06729.	3.972842
C6orf21 variant proteinchr6_qbl_hap6+	chr6_qbl_hap6	Homo sapiens mRNA for C6orf21 variant protein, partial cds, clone: hj06729.	3.972842
RPL13AP20	chr12	SubName: Full=Ribosomal protein L13a variant; Flags: Fragment;	0.558796
TUBA1Cchr12-	chr12	Synthetic construct Homo sapiens gateway clone IMAGE:100017296 3' read TUBA1C mRNA.	1.300839
KIF5B	chr10	kinesin family member 5B	0.380245
HSPA1Lchr6_apd_hap1-	chr6_apd_hap1	heat shock 70kDa protein 1-like	4.323363
HSPA1Lchr6-	chr6	heat shock 70kDa protein 1-like	4.30745
HSPA1Lchr6_dbb_hap3-	chr6_dbb_hap3	heat shock 70kDa protein 1-like	4.30745
POLR2J4	chr7	Homo sapiens cDNA FLJ58900 complete cds, weakly similar to Uroplakin-3B precursor.	2.069547
NDUFS6	chr5	NADH dehydrogenase (ubiquinone) Fe-S protein 6	1.382241

RNAseq top 400 differentially regulated genes (Continued)

C6orf21 variant proteinchr6_mann_hap4+	chr6_mann_hap4	Homo sapiens mRNA for C6orf21 variant protein, partial cds, clone: hj06729.	4.124384
ATP2B1	chr12	plasma membrane calcium ATPase 1	0.340601
ARHGEF18	chr19	Rho/Rac guanine nucleotide exchange factor 18	3.080219
BC015662	chr6	Homo sapiens high mobility group AT-hook 1, mRNA (cDNA clone IMAGE:4844539).	1.562596
RPS19	chr19	ribosomal protein S19	0.731287
BAT2chr6_mcf_hap5+	chr6_mcf_hap5	HLA-B associated transcript-2	2.031376
MESDC2	chr15	mesoderm development candidate 2	0.355542
C6orf21 variant proteinchr6_mcf_hap5+	chr6_mcf_hap5	Homo sapiens mRNA for C6orf21 variant protein, partial cds, clone: hj06729.	4.044016
EEF1A2chr6-	chr6	RecName: Full=Elongation factor 1-alpha 2; Short=EF-1-alpha-2; Short=Elongation factor 1 A-2; AltName: Full=eEF1A-2; AltName: Full=Statin S1;	1.18096
COX7A2L	chr2	cytochrome c oxidase subunit VIIa polypeptide 2	1.929161
DUSP1	chr5	dual specificity phosphatase 1	2.294448
C6orf21 variant proteinchr6_dbb_hap3+	chr6_dbb_hap3	Homo sapiens mRNA for C6orf21 variant protein, partial cds, clone: hj06729.	4.068955
WBP2	chr17	WW domain binding protein 2	1.95577
NFKB2	chr10	nuclear factor of kappa light polypeptide gene	2.126312
BAT3chr6_cox_hap2-	chr6_cox_hap2	HLA-B associated transcript-3	2.112236
BC078172	chr10	Homo sapiens cDNA clone IMAGE:5760022, partial cds.	3.254654
HLA-Cchr6_dbb_hap3-	chr6_dbb_hap3	SubName: Full=MHC class I HLA-C allele; Flags: Precursor; Fragment;	1.799672
COPG	chr3	SubName: Full=Gamma1-COP; SubName: Full=Coatomer protein complex, subunit gamma	1.777813
KIAA1783	chr17	Homo sapiens mRNA for KIAA1783 protein, partial cds.	2.111385
TPR	chr1	SubName: Full=TPR protein; Flags: Fragment;	0.314363
CR596118	chr5	full-length cDNA clone CS0DF026YI08 of Fetal brain of Homo sapiens (human).	0.700465
pp7704	chr1	Homo sapiens HT031 mRNA, complete cds.	3.135332
DKFZp434C119	chr17	SubName: Full=cDNA FLJ50246, highly similar to Mitochondrial carrier protein CGI-69;	2.417396
SOD1	chr21	superoxide dismutase 1, soluble	1.344646
RPS27A	chr2	ubiquitin and ribosomal protein S27a precursor	0.660693
PLSCR3	chr17	phospholipid scramblase 3	1.688149
ATP5G3	chr2	ATP synthase, H+ transporting, mitochondrial F0	1.719891
ARPC4	chr3	actin related protein 2/3 complex subunit 4	1.862117
BC033739	chr16	Homo sapiens cDNA FLJ43590 fis, clone SMINT2001818.	2.785728
ZCRB1	chr12	zinc finger CCHC-type and RNA binding motif 1	0.38508
HLA-Bchr6_dbb_hap3-	chr6_dbb_hap3	SubName: Full=MHC class I antigen; Flags: Fragment;	1.85576
PEA15	chr1	phosphoprotein enriched in astrocytes 15	2.127349

RNAseq top 400 differentially regulated genes (Continued)

SMARCC1	chr3	SWI/SNF-related matrix-associated	0.423784
TTC3	chr21	tetratricopeptide repeat domain 3	0.312403
HIST1H4C	chr6	histone cluster 1, H4c	0.265631
NAP1L1	chr12	nucleosome assembly protein 1-like 1	0.634602
BAT3chr6_ssto_hap7-	chr6_ssto_hap7	HLA-B associated transcript-3	2.070073
BAT3chr6_qbl_hap6-	chr6_qbl_hap6	HLA-B associated transcript-3	2.070565
CD63	chr12	CD63 antigen	0.686107
C6orf21 variant proteinchr6+	chr6	Homo sapiens mRNA for C6orf21 variant protein, partial cds, clone: hj06729.	4.072186
PSMA7	chr20	proteasome alpha 7 subunit	0.61336
ARSDR1	chr14	SubName: Full=cDNA FLJ53708, highly similar to Retinol dehydrogenase 11 (EC 1.1.1.-);	0.435169
VARSchr6_cox_hap2-	chr6_cox_hap2	valyl-tRNA synthetase	1.862007
VARSchr6-	chr6	Homo sapiens cDNA FLJ54383 complete cds, highly similar to Valyl-tRNA synthetase (EC 6.1.1.9).	1.860792
VARSchr6_dbb_hap3-	chr6_dbb_hap3	valyl-tRNA synthetase	1.860792
LUC7L3	chr17	LUC7-like 3	0.396992
C17orf49	chr17	hypothetical protein LOC124944	1.556997
ANKRD11	chr16	ankyrin repeat domain 11	0.395629
TRIM33	chr1	tripartite motif-containing 33 protein	0.339539
VARSchr6_apd_hap1-	chr6_apd_hap1	valyl-tRNA synthetase	1.858637
VARSchr6_mcf_hap5-	chr6_mcf_hap5	valyl-tRNA synthetase	1.858637
DNAJC21	chr5	DnaJ homology subfamily A member 5	0.326358
MSN	chrX	moesin	1.738963
CROP/Luc7A	chr17	SubName: Full=cDNA FLJ59548, highly similar to Cisplatin resistance-associated overexpressed protein;	0.381183
TMEM214	chr2	transmembrane protein 214	2.144868
PYCR1	chr17	pyrroline-5-carboxylate reductase 1	1.93347
SNRPA	chr19	small nuclear ribonucleoprotein polypeptide A	1.962597
MT-ND4L	chrM	NADH dehydrogenase subunit 4L	0.832684
RDX	chr11	radixin	0.399575
KDM5A	chr12	retinoblastoma binding protein 2	0.29715
SDHAchr5-	chr5	Homo sapiens cDNA FLJ58919 complete cds, moderately similar to Succinate dehydrogenase (ubiquinone) flavoprotein subunit, mitochondrial precursor (EC1.3.5.1).	0.327777
TTF1	chr9	transcription termination factor, RNA polymerase	0.319229
RDH11	chr14	retinol dehydrogenase 11	0.44709
PPIG	chr2	peptidylprolyl isomerase G	0.30318
GOLGA2	chr9	SubName: Full=Golgi autoantigen, golgin subfamily a, 2; Flags: Fragment;	0.542955
PHIP	chr6	pleckstrin homology domain interacting protein	0.296799
DYNC1H1	chr14	cytoplasmic dynein 1 heavy chain 1	0.632998

RNAseq top 400 differentially regulated genes (Continued)

TOP1	chr20	DNA topoisomerase I	0.429615
SLTM	chr15	modulator of estrogen induced transcription	0.296135
DKFZp762G052	chr15	SubName: Full=cDNA FLJ30633 fis, clone CTONG2002418, highly similar to Homo sapiens modulator of estrogen induced transcription, transcript variant 1, mRNA; SubName: Full=cDNA FLJ78261;	0.282457
EIF5A	chr17	eukaryotic translation initiation factor 5A	1.29188
TPP1	chr11	tripeptidyl-peptidase I preproprotein	2.997457
SNRNP48	chr6	U11/U12 snRNP 48K	0.253156
GRP94cchr12+	chr12	SubName: Full=Heat shock protein 94c;	0.572111
PSIP1	chr9	PC4 and SFRS1 interacting protein 1	0.310453
EIF3J	chr15	SubName: Full=cDNA FLJ57599, moderately similar to Eukaryotic translation initiation factor 3 subunit 1;	0.514275
ALPP	chr2	placental alkaline phosphatase preproprotein	2.295661
CCDC25	chr8	Homo sapiens cDNA FLJ10853 fis, clone NT2RP4001502.	0.411721
RHOB	chr2	ras homolog gene family, member B precursor	1.843297
RPL7	chr8	ribosomal protein L7	0.666566
AK093299	chr1	Homo sapiens cDNA FLJ35980 fis, clone TESTI2013546.	4.902144
PSMD7	chr16	SubName: Full=cDNA FLJ51373, highly similar to 26S proteasome non-ATPase regulatory subunit 7;	0.485288
C16orf88	chr16	hypothetical protein LOC400506	0.374755
LYAR	chr4	Ly1 antibody reactive homolog	0.301411
TUBB2Achr18+	chr18	SubName: Full=TUBB2A protein; Flags: Fragment;	2.041709
SRPK2	chr7	Synthetic construct DNA, clone: pF1KB3384, Homo sapiens SRPK2 gene for SFRS protein kinase 2, without stop codon, in Flexi system.	0.373769
OAS3	chr12	2'-5'oligoadenylate synthetase 3	3.613097
P2RY11	chr19	purinergic receptor P2Y11	0.302701
RPN1	chr3	ribophorin I precursor	1.622207
PQBP1	chrX	polyglutamine binding protein 1	1.696782
PSMB4	chr1	proteasome beta 4 subunit	1.66446
PRPF40A	chr2	formin binding protein 3	0.364652
NUP188	chr9	nucleoporin 188kDa	2.012372
S QSTM1	chr5	sequestosome 1	1.488766
UPF3B	chrX	UPF3 regulator of nonsense transcripts homolog B	0.256973
HLA-Bchr6-	chr6	SubName: Full=MHC class I antigen;	1.8586
RPL37	chr5	ribosomal protein L37	0.87647
ERGIC1	chr5	RecName: Full=Endoplasmic reticulum-Golgi intermediate compartment protein 1; AltName: Full=ER-Golgi intermediate compartment 32 kDa protein; Short=ERGIC-32;	1.931334
EEF1A1chr6-	chr6	eukaryotic translation elongation factor 1 alpha	1.153656

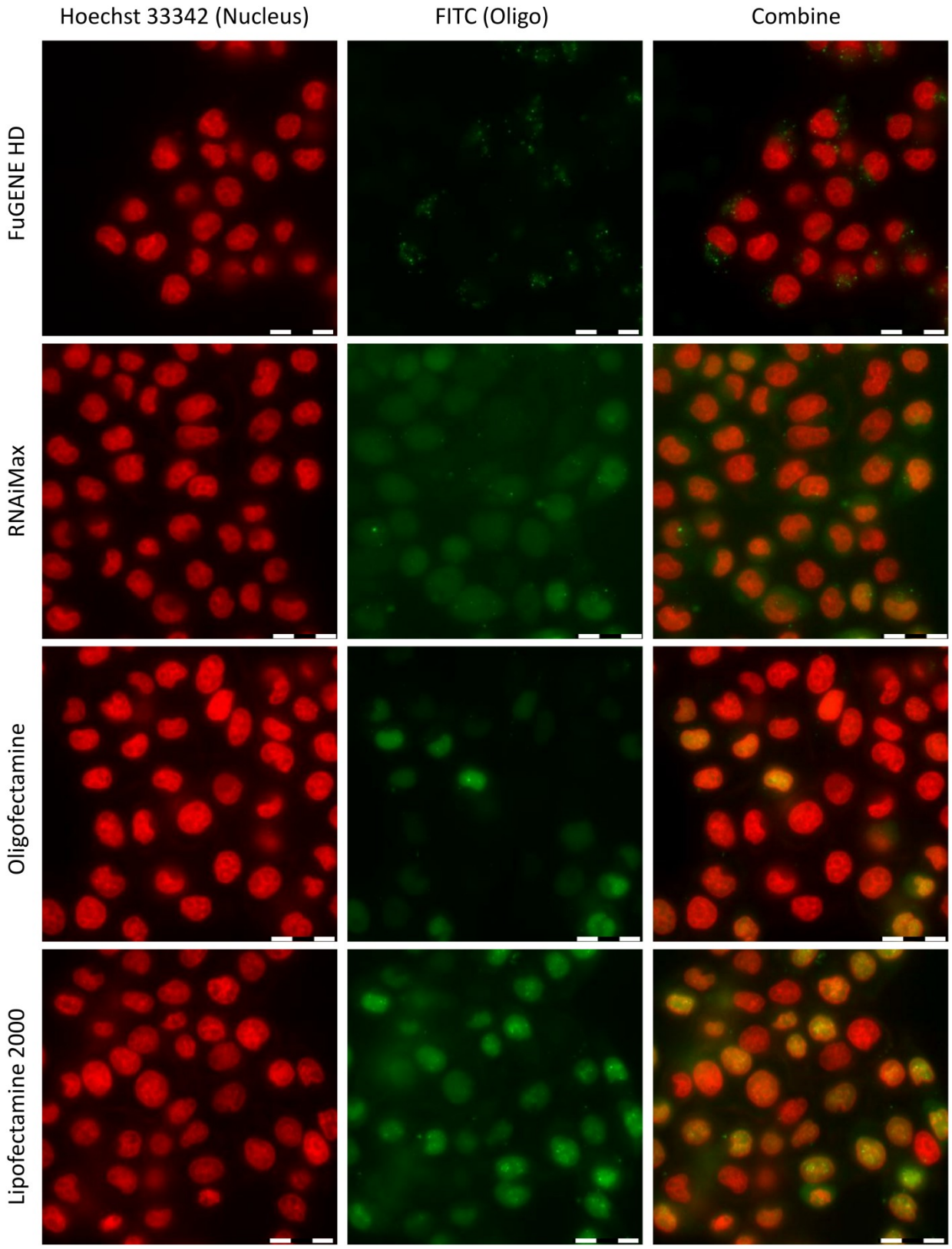
RNAseq top 400 differentially regulated genes (Continued)

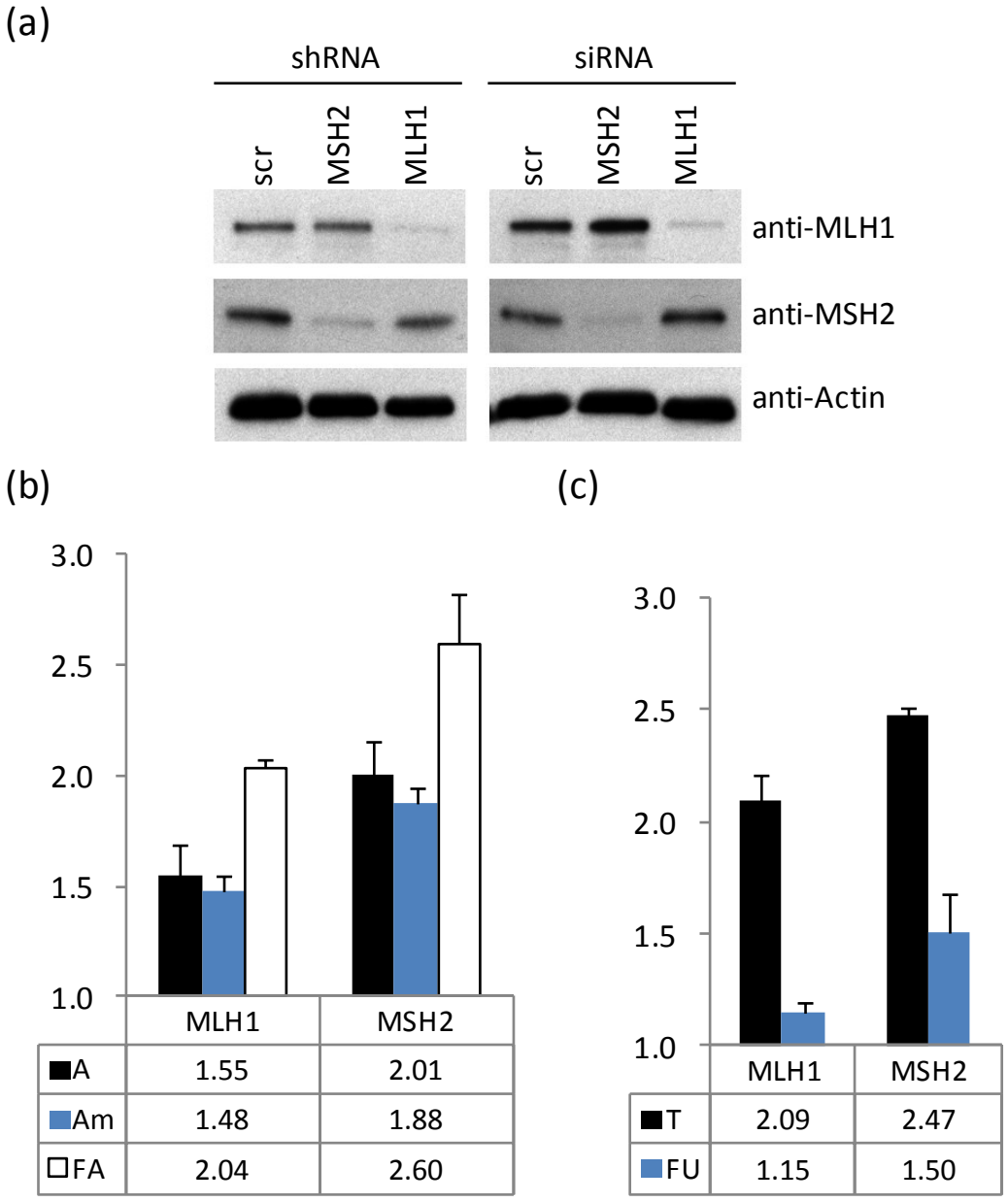
TRAF7	chr16	TNF receptor-associated factor 7	2.000212
PTMS	chr12	parathymosin	1.346629
ATP5O	chr21	Homo sapiens full length insert cDNA clone ZD90F06.	0.594704
PSMB3	chr17	proteasome beta 3 subunit	1.436605

CHAPTER 2 SUPPLEMETARY FIGURES

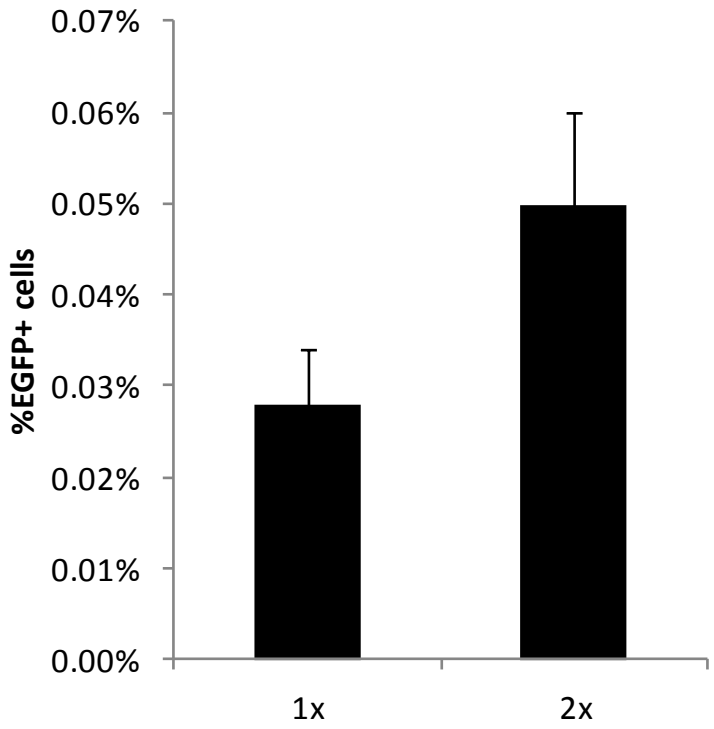
Supplementary Figure 2.1. Nuclear localization of transfected oligos. Cells were grown in glass-bottom 24-well plates and stained with Hoechst 33342 10ug/mL. The next day, cells were transfected with 500ng Fluorescein-labeled oligo using various cationic lipid reagents following manufacturer instructions and imaged 24 hours after transfection. Scale bar=10μm

Supplementary Figure 2.1 (Continued)

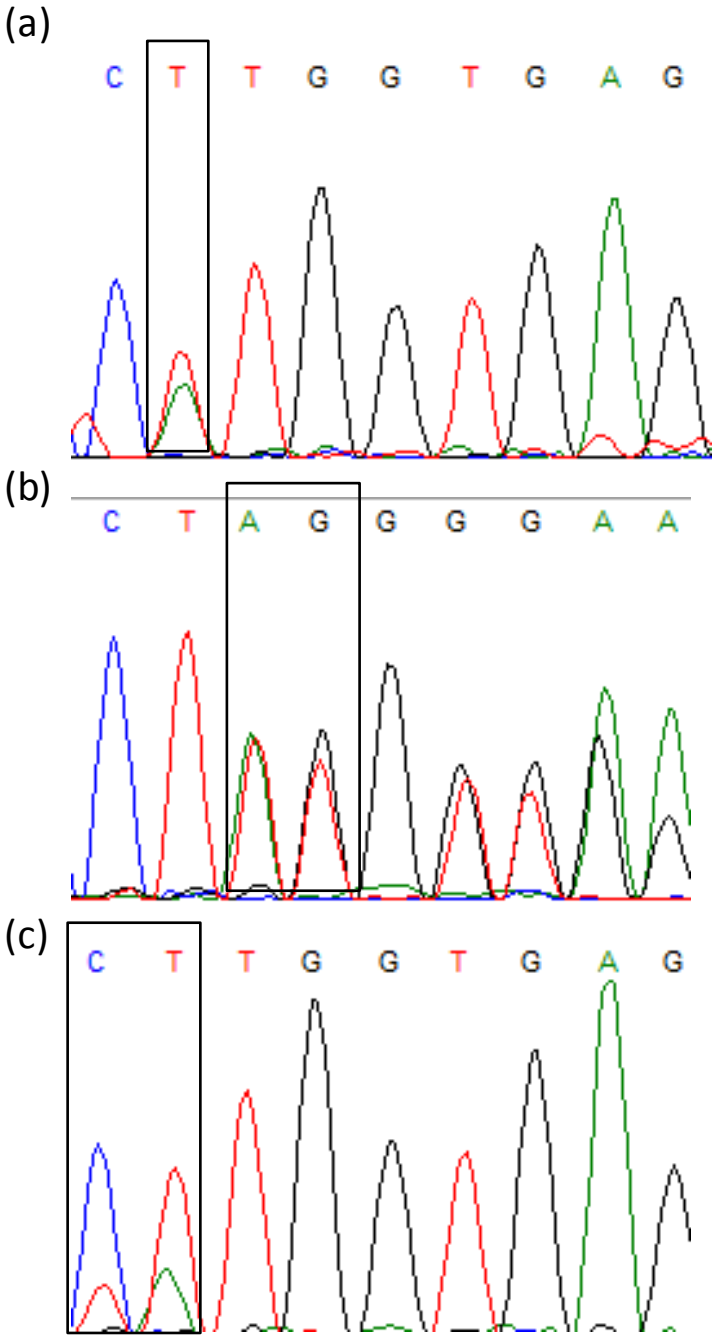




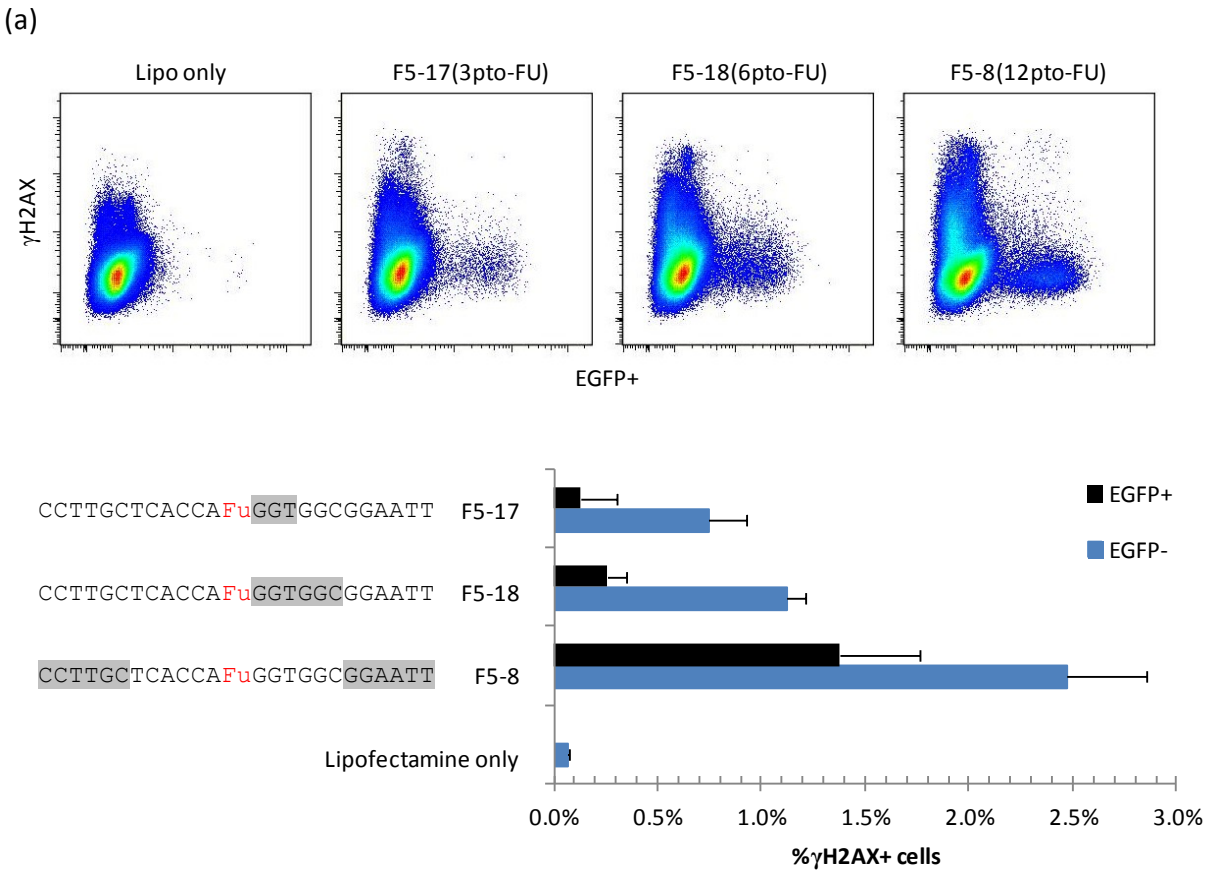
Supplementary Figure 2.2. Validation of RNAi experiments (a) Western Blotting confirming knockdown of MMR components. Primary antibodies anti-MSH2 (ab52266, 1:5,000), anti-MLH1 (ab92312, 1:2,000) and anti-Actin (ab3280, 1:10,000), and secondary antibodies anti-rabbit IgG (ab6721, 1:25,000) and anti-mouse IgG (ab6728 1:25,000) were obtained from Abcam. (b), (c) siRNA targeting mismatch repair components have a lower effect with modified bases. A 24-well plate was treated with 80nM siRNA, 1uL RNAiMax following manufacturer's protocol. siRNA-treated cells were transfected with targeting oligos 72hrs later. N=4 siRNAs sequences can be found in Table 2.2



Supplementary Figure 2.3. Increased transfection reaction leads to higher stable frequencies with a less toxic oligo. HeLa F5 cells transfected with F5-17 oligo, either 1x or 2x DNA:lipofectamine complexes, assayed for %EGFP+ cells eight days after transfection.



Supplementary Figure 2.4. Sequencing corrected EGFP+ clonal cells. Single-cells were sorted into a 96-well plate 8 days post-transfection with (a) F5-17,(b)F5-34, (c)F5-35, grown for two weeks and then sequenced. Sequence spectrum overlap shows one of the two mEGFP copies was modified.



Supplementary Figure 2.5. **γ-H2AX phosphorylation staining.** Cells were plated 500,000/well on a 6-well plate, then transfected with the corresponding oligos. After 36hrs, cells were washed, fixed in 4%PFA, then stained with Alexa Fluor 647 anti-H2A.X-Phosphorylated Antibody (1:25, Biolegend). Cells permeabilized with 10% saponin, shaking for 20 mins at room temperature.