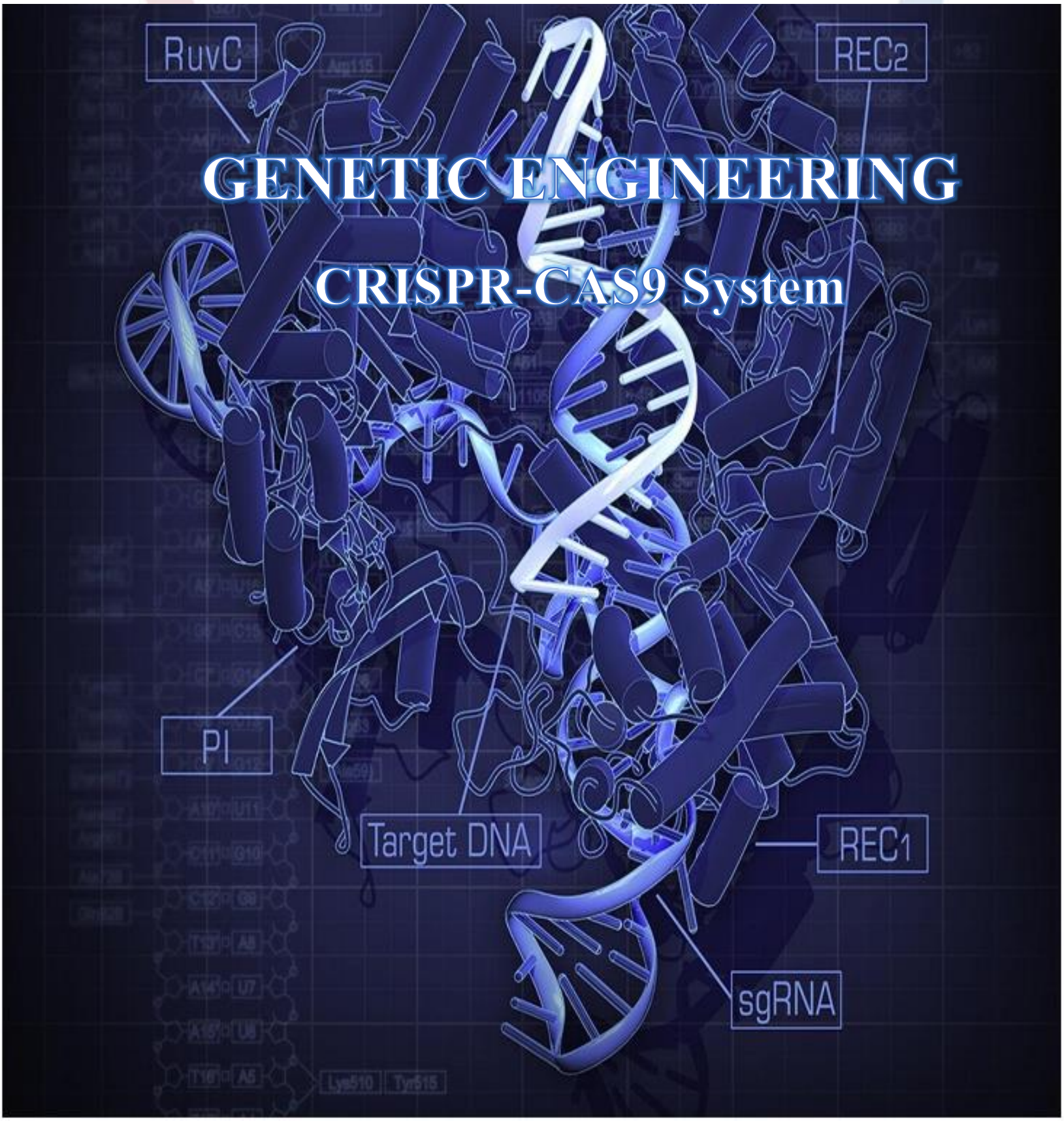
المركز الوطني للمتميزين

NATIONAL CENTER FOR THE DISTINGUISHED

/NCDSyria



GENETIC ENGINEERING

CRISPR-CAS9 System

Syrian Arab Republic.

Ministry of high education.

Organization of distinction and innovation.

National center for the distinguishes.

GENETIC ENGINEERING

Crispr-cas9

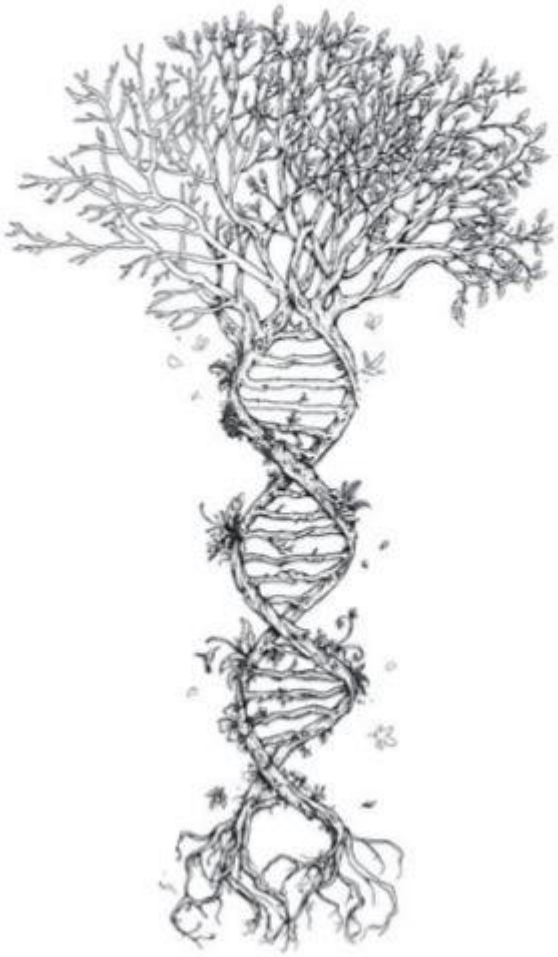
Submitted by: Hasan Mohammad Ammar
Supervisor: M.Nedal Hasan
Date: 1/1/2017

Table of contents:

Table of figures:

1. Abstract:

- Genome engineering is just about its dawn of golden age. New and exciting tools are arising and piling up to its arsenal.
- .After TALENs, Zing-finger a new tool is coming to add up, the newest addition is the CRISPR system, a system that firstly discovered at the Streptococcus thermophilus and its natural role is the sequence-specific adaptive immunity against foreign DNA in bacteria.
- In the present report, CRISPR Cas9, type II system, used in an attempt to knock out the HPRT1 gene expressed in lymphocytes as a proof of concept. HAT medium in combination with Thioguanine used as a selection medium enabling a quick and simple read-out screening method, to verify that a successful editing event took place in the mammalian cells, a detection DNA cleavage kit used in combination with an enzymatic assay.



2. Inquiry of the research:

- In our daily life, we hear a lot about people who died because of diseases they inherit, or kept surfing them for a long period without finding a cure. In the last years, many experiments were done and many tools appeared such as Crisper who managed to modify the viruses' DNA. So, will curing these fatal diseases be possible in the future just by fixing the damaged mutation in the DNA? And can we protect the incoming generation from this disease? And will we build a perfect child one day?

3. List of abbreviation:

- **CRISPR:** Clustered regularly interspaced short palindromic repeats.
- **CAS:** CRISPR-associated (protein).
- **HGPRT:** Hypoxanthine-guanine phosphoribosyl-transferase.
- **DHFR:** Dihydro folate reductase.
- **ZFN S:** Zinc-finger nuclease.
- **TALEN S:** Transcription activator-like effector nuclease.
- **PAM:** Protospacer adjacent motif.
- **Pre-crRNA:** Precursor-crRNA.
- **CrRNA:** CRISPR RNA.
- **SgRNA:** Single guide RNA, also seen as gRNA.
- **TracrRNA:** Trans-activating crRNA.
- **DSB:** Double strand breaks.
- **HR:** Homologous Recombination.
- **NHEJ:** Non-homologous end Joining.
- **NK cells:** Natural Killer cells.
- **HAT:** Hypoxanthine-aminopterin-thymidine medium.
- **TG:** Thioguanine.
- **LNS:** Lesch-Nyhan Syndrome.
- **BP:** Base pair.
- **Knock-out:** introducing a mutation that inactivates completely a gene's function.
- **Knock-in:** introducing a new gene.

4. Introduction:

- Protection of a host genome from invading genetic material is a fundamental challenge faced by all organisms. In prokaryotes, a classical example of a solution to this problem is use of restriction enzymes to recognize and cut specific sequences in invasive DNA, leading to their degradation. Remarkably, an additional solution present in a wide variety of prokaryotes went undetected until very recently: the CRISPR adaptive immune system. Just as the discovery and re-purposing of restriction enzymes ushered in the recombinant DNA revolution in the 1970s, novel methods of engineering the CRISPR system have already resulted in new tools and techniques that have dramatically accelerated our ability to annotate the genome.
- The newest genome-editing tool, CRISPR, is used in a proof of concept study. Primary white mononuclear cells are cultured and treated with CRISPR in order to knockout the HPRT1 gene. Clustered regularly interspaced short palindromic repeats, is abbreviated as CRISPR and is the new breakthrough of genome editing. CRISPR Cas9 type II system is the system that is most often used so far since it has unique advantages compared to the other two types of systems. The HPRT1 gene is responsible for the expression of hypoxanthine-guanine phosphoribosyl-transferase (HGPRT). As its name designates, it is a transferase enzyme used in the purine salvage pathway to degrade DNA and to reintroduce purines back to the synthetic pathway. HGPRTase has an important role in the generation of purine nucleotides. Mutations in the gene can lead to hyperuricemia with the most known disease Lesch-Nyhan syndrome. Lymphocytes are white blood cells expressing the HPRT1 gene in significant amounts. The selection medium HAT is used to grow and select only HGPRT+ cells. HAT medium consists from Hypoxanthine, Aminopterin and Thymidine. The Aminopterin inhibits the enzyme (dihydrofolate reductase, DHFR) that is responsible for the synthesis of nucleic acids. This forces the cells to use salvage pathway as an alternative way of growth, in which a functional HGPRT enzyme is essential. In HAT medium, only the HGPRT+ cells survive and the HGPRT- cells die. HAT is used as a counter selection medium. HGPRT+ cells will be knocked out with the use of CRISPR Cas9 system and the efficiency of this technique is evaluated by using the 6-thioguanine (TG) resistant mutants' analysis. TG is a selection medium for HPRT- negative cells. Therefore, only knock out cells (HPRT-) will survive and the number of the mutant cells points out the efficiency of CRISPR tool. To enhance our observations a rapid award winning biochemical assay is used to monitor the HGPRTase activity and also a genome cleavage detection kit is used to detect the specific locus cleavage of genomic DNA.

5. Introduction to CRISPR:

- Analysis of the Escherichia coli genome in 1987 revealed loci containing repeat sequences with unknown function [2]. When similarly structured loci found in other prokaryotes, the acronym CRISPR (clustered regularly interspaced short palindromic repeats) was coined [3]. Several years later, in silico analysis showed that portions of these sequences mapped to viral and phage genomes, suggesting a role in host genome defense and, by 2007, functional studies had validated this hypothesis [4, 5]. CRISPR loci have identified and characterized in a wide range of both bacteria and archaea, and function as an adaptive immune system.
- As their name suggests, CRISPR loci is organized as repetitive arrays, with variable sequences of approximately 20 nucleotides known as spacers preceding a relatively invariant common sequence. Endogenous CRISPR loci are classified into three groups based on the presence of specific CRISPR-associated (CAS) protein-coding genes located in close proximity to the repetitive arrays [6]. Because both type I and III CRISPR systems require multiple proteins to find and cut DNA, they are not readily amenable to re-purposing. In contrast, the type II CRISPR systems from Streptococcus thermophilus and Streptococcus pyogenes have been extensively studied because only a single protein, Cas9, is required to locate target sequences and perform DNA cleavage; Cas9 of S. pyogenes, in particular, is the most widely used due to its minimal additional sequence requirements [7, 8].
- CAS proteins perform the two distinct functions of CRISPR loci: spacer acquisition and target cutting. Spacer sequences arise from the recognition, processing and insertion of invading genetic material into the CRISPR loci, serving as a molecular memory to protect from subsequent invasion by the same vector. The target-cutting phase begins with transcription of the repetitive array, followed by processing of the spacers into individual CRISPR RNAs (crRNAs). The processed crRNA hybridizes to a separately transcribed trans-activating crRNA (tracrRNA), which together associate with Cas9 and are sufficient for nuclease activity. The crRNA guides Cas9 to complementary DNA, which cleaved by the RuvC and HNH nuclease domains of the Cas9 enzyme [9, 10].
- The three components necessary for target cutting (Cas9 protein, tracrRNA and crRNA) may be programmed to cut a DNA sequence of interest simply by altering the 20 nt variable spacer sequence of the crRNA [10, 11]. The only sequence constraint is the presence of a protospacer adjacent motif (PAM) immediately proximal to the region of crRNA homology on the DNA target; this sequence is NGG for S. pyogenes Cas9, but other Cas9 proteins have alternative PAM requirements [10, 12]. This system may be further simplified by fusion of the tracrRNA and crRNA at the DNA level by inclusion of a linker sequence resulting in transcription of a single guide RNA (sgRNA) encoding the functionalities of both RNAs [10]. Expression of these two components (Cas9 and sgRNA) is both necessary

and sufficient for applying CRISPR technology to a wide variety of biological contexts, including human cells [13, 14] and mouse models [15]. Schematic of interaction between sgRNA and DNA. The crRNA sequence (blue) and tracrRNA sequence (green) are fused together by a short loop (purple) to create a sgRNA. The 20 underlined nucleotides can be programmed to recognize any DNA sequence of interest. For S.pyogenes Cas9, the NGG PAM is required immediately downstream of the target site. The two strands of DNA are cut by the HNH and RuvC nuclease domains of Cas9.

6. Genome editing:

- Genome engineering described as the ability to modify and manipulate precisely DNA sequences in living cells [1]. A new era is emerging fast with new techniques able to engineer the genome with even greater impact. The ability to insert, remove or even edit DNA sequences easily and precisely has attracted the interest of the scientific community in a wide range of biotechnology areas, such as medicine, energy and even environmental studies. From a medical perspective, this fascinating and emerging field in combination with preclinical and clinical trials can potential treat various diseases.
- Targetable nucleases are paving the way for this upcoming arena. Targetable nucleases enable scientists to target and modify theoretically any gene in any organism [1].The nucleases programmed with site-specific DNA binding domains and can have (i) enhanced performance, (ii) accelerated nuclease assembly and (iii) significantly lower cost of genome editing[1]. Zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and engineered homing endonucleases, which or also known as mega nucleases, are the tools used for genome engineering nowadays and could also be labeled as true targeting tools. The newest addition to this list is a bacterial nuclease based on clustered regulatory interspaced short palindromic repeat (CRIPSR) CAS system[1]. The first two above-mentioned tools are using nucleases that bound with modular DNA-binding proteins in order to induce DNA double-strand breaks. Whereas CRISPR Cas9 system uses a nuclease that is guided by a small, 20 nucleotide RNAs through Watson - Crick base pairing to target DNA [1]. The rapid growth of genome-editing field has resulted the availability of various commercially engineered targeted nucleases. However, no method is faultless and each of them has its own pros and cons. In general, the current approaches for genome engineering hampered by (i) low efficiency and (ii) limited number of cell types and organisms that can be targeted [1].
- The ideal genome-editing tool should fulfill the following three criteria: (1) No off-target mutation, (2) Rapid and efficient assembly of the nucleases, (3) High frequency of the desired sequence in the target cell population. Cas9 nuclease, ZFNs and TALENs used for

genome editing by stimulating a double strand break at the target genomic locus [1]. When DNA double strand breaks in eukaryotes, the error-prone non-homologous end-joining pathway (NHEJ) often repairs it.

- The double-strand break facilitates engineering of targeting mutations by serving as a substrate for NHEJ repair mechanism [1]. This creates a band of indels at the cleavage site that can detected by an electrophoresis. NHEJ mutagenesis commonly used as a method for creating targeted knockouts. More than 30 species and 150 human genes and loci have knocked-out by using this method [1]. Moreover, when an exonic sequence is targeted 66 % of indels would expected to result in frame shift mutations [1].

6.1. Existing editing-tools:

- Zinc-finger nucleases are using approximately a 30-aminoacid finger that folds around a zinc ion to form a compact structure that recognizes a 3-base pair of DNA. Consecutive finger repeats are able to recognize and target a wide area of the target DNA. Before the zinc fingers assembled, they optimized to recognize a specific 3-base pair sequence. However not all sites are accessible meaning that the position of the targeted site is not determined by scientists but by DNA accessibility. This disability minimizes the application of zing-finger tool since enzyme active sites and single nucleotide polymorphisms cannot be targeted [1].
- TALENs first discovered in late 2009 when the TALE-DNA-binding code was discovered [1]. For TALENs, two main high-throughput approaches were developed. From which the second one was a clever usage of type II's restriction enzymes called "Golden Gate cloning". Golden Gate cloning is a molecular cloning method, which uses multiple subunits in a designed order. Type IIs restriction enzymes are able to bind in one site and cleave at the adjacent site. This gives the advantage of designing a system and being able to predict where to cut the DNA.
- Quickly, it became obvious that TALENs had a significant advantage over Zing-finger nucleases. Almost all TALENs demonstrate some kind of activity on their chromosomal target site and in comparison, with ZFNs show that it is greatly higher. Both techniques restricted by the accessibility of the chromatin site, implying that not all sites can be approached. However, TALENs has a wider spectrum of sequences that can targeted. In addition, both techniques can lead to potential off target sequence. However, TALENs has fewer chances for off targeting since TALENs designed to recognize 30-36 base pair from the targeted site whereas ZFNs designed to recognize 18-24 base pairs [1]. Mega nucleases are a term used to describe the targetable homing endonucleases that tailored made to a specific target. The target site that these nucleases can recognize is approximately at 24 base pair.

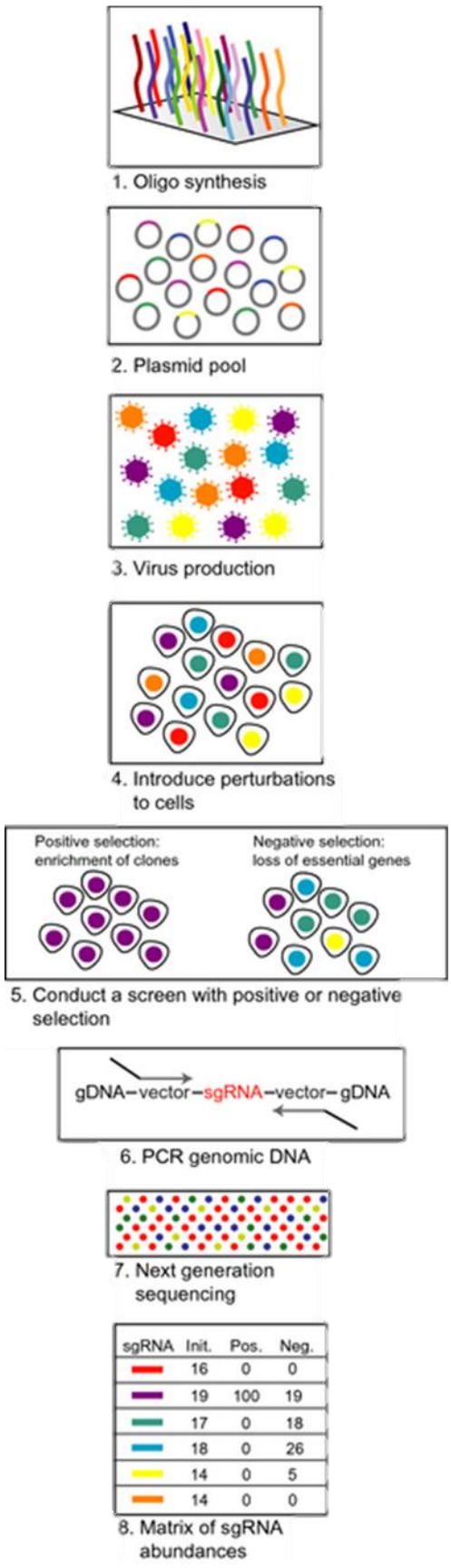
7. Introduction to genetic screens:

- The discovery that CRISPR technology may be easily adapted to work in mammalian cells and delivered by lentiviral vectors led to the possibility of developing CRISPR as a new screening technology. Functional genetic screens have performed using previous technologies for many years, and many of the lessons and challenges from these screens apply to CRISPR, it is therefore instructive briefly to review these previous approaches.
- The discovery of RNA interference (RNAi) a decade ago suggested the possibility of determining loss-of-function phenotypes systematically across many cell types. Here, delivery of a double-stranded RNA of approximately 21 nt triggers a pathway that is endogenous to most eukaryotic cells, in which a protein of the Argonaut family uses one strand of the exogenous RNA as a guide to cleave and thereby degrade complementary mRNAs. Many discoveries have made with RNAi-based screening, using small interfering RNAs directly or DNA-encoded transcriptional units that produce short hairpin RNAs (shRNAs) or long microRNA-like transcripts. The technology has continued to improve, such that effective knockdown is usually easily achieved [16]. While the successes are far too numerous to list here, RNAi also has an inherent weakness that arises when the introduced small RNA enters the microRNA pathway and affects the expression of many mRNAs that have a complementary 'seed' region rather than the full 21 nt of complementarity [17, 18]. While off-target effects have long recognized as a potential confounder in RNAi experiments, their pervasiveness only started to be widely appreciated recently [19, 20]. Given that any single RNAi reagent may function in both on- and off-target pathways, relying on a single RNA sequence as a surrogate for a gene may lead to erroneous conclusions. Proper interpretation of RNAi experiments requires that multiple independent sequences targeting the same gene result in the same phenotype this remains true for CRISPR technology [18, 21].
- Another important aspect of genetic screens, first with RNAi and now with CRISPR, is that the targeting information is contained in a short stretch of nucleotides that may easily been created on a large scale using oligonucleotide synthesis technology. This allows production of large pooled libraries rapidly and less expensively than an arrayed library of individual clones (Figure 1). Pooled libraries comprising many perturbations then introduced into retroviral vectors through traditional cloning methods. Retroviral vectors, including lentivirus, have the important feature of stably incorporating the DNA sequence of the perturbation into the genome of the host cell, resulting in a permanent, heritable genomic modification. Screening all perturbations in a pooled format relies on the ability to physically separate cells with the intended phenotype from the remainder of the population. Overview of pooled screening. Oligonucleotides are synthesized in array format (1), cloned into a

plasmid library (2), and then packaged into a retroviral vector (3). Cells are then infected en masse, typically at low multiplicity of infection, such that most cells receive either zero or one viral insert (4). Selective pressure is then applied to either enrich or deplete particular perturbations (5). Genomic DNA is then harvested and PCR-amplified using primers that add the necessary adaptors (6) for next-generation sequencing (7). The abundance of each sgRNA under each condition (initial abundance, positive selection, negative selection, etc.) is the final output of the screen and the starting point for analysis (8).

- Pooled screening assays fall into two broadly defined types: positive and negative selection screens. The simplest example of a negative selection phenotype is cell viability: perturbations that cause the cell to die are depleted from the population over time. Such an approach has been used to profile genes that are broadly essential to cell viability, as well as to find genes that are essential to specific cell types [22, 23]. This has been especially useful in the context of cancer biology to identify specific vulnerabilities of cancer cells from a certain lineage or carrying a particular oncogene, [24, 25]. A limiting feature of negative selection assays is that, by definition, a perturbation may only be depleted to the extent that it was present in the starting library. For example, in a library of 100 000 perturbations, each perturbation is present at 10 parts per million (p.p.m.) on average. Therefore, the most it may be depleted in a screen is down to 0 p.p.m. This means that negative selection screens may have more signal when the library size is small with a corresponding higher relative abundance of each perturbation. Larger libraries may thus be broken into smaller sub-pools, or a more targeted library may be used as a follow-up to the initial screen. Positive selection screens, however, rely on enrichment of cells, and therefore have a much larger dynamic range: in the same

(Figure 1)



library, a single perturbation may enrich from 10 p.p.m. to 100% of the final population, representing a 100 000-fold enrichment. Because of this, positive selection screens often yield more easily interpreted results than negative selection screens. Examples of positive selection screens include growth assays to rescue from a drug that is toxic to the starting cell population, and flow cytometry assays to purify a population of cells that modulates expression of a sortable marker.

- Regardless of the direction of selection, the identity of the perturbations in the population(s) must be determined following physical separation of cells into two or more populations. Previously, this was often performed using hybridization techniques such as microarrays, but, more recently, massively parallel sequencing has become the method of choice as costs have dramatically fallen. Constant vector sequences are used as priming sites for PCR to amplify the perturbations from viral integration sites in genomic DNA, and the resulting pool of products is sequenced. Here the perturbation of interest (shRNA or sgRNA) acts as its own ‘barcode’ as each sequence is unique and may be mapped to the original gene without any additional sequence information. The PCR primers themselves may also incorporate an additional ‘barcode’ to label samples arising from different conditions, allowing multiple independent cell populations to be pooled into the same sequencing lane to minimize cost. Care must be taken to avoid introducing bias during the amplification and sequencing steps, both in terms of use of optimized protocols and inclusion of the appropriate control populations to correct for these biases when they do occur [26]. Sequencing produces raw read counts of each perturbation, and thus the end of a screen (and the start of data analysis) is a matrix of read counts where the relative abundance of each perturbation is determined for each sample.

8. Genetic screens using CRISPR technology:

- The first reports of genetic screens using CRISPR technology were proof-of-concept studies that largely relied on assays previously screened with RNAi libraries in systems where the underlying biology was already well-understood [27, 28]. Wang et al. generated a library targeting approximately 7000 human genes with a large number of sgRNAs per gene [27]. They used this library to perform a positive selection screen using the purine analog 6-thioguanine, which normally causes cell death in the near-haploid KBM7 cell line, to identify components of the mismatch repair pathway. In a demonstration of the efficacy of CRISPR technology, they observed that all 20 of the most-abundant sgRNAs following selection targeted one of the four expected genes. In a similar screen using the DNA-damaging agent etoposide, TOP2A and CDK6 were identified as mediators of resistance; while the former had already been established as a resistance mechanism by an shRNA screen, the latter represented a new finding [23].
- A contemporaneous report by Shalem et al. used a genome-wide library to examine resistance to vemurafenib, a cancer therapeutic agent that is especially toxic to melanoma cells with the BRAF V600E mutation [28]. Two genes previously reported to result in resistance to vemurafenib were identified, NF1 and MED12, together with several other novel genes, including NF2, CUL3 and members of the STAGA complex [29, 30]. Importantly, the consistency of sgRNAs targeting the same gene was compared to previously published results with shRNA screens, and both papers reported a much higher fraction of perturbations targeting the same gene scoring as hits with CRISPR technology, suggesting a far lower false-positive rate. These two reports also confirmed that CRISPR technology might use for negative selection screening by examining the depletion of essential genes over time, such as components of the ribosome, spliceosome and proteasome. Additional reports have expanded the range of model systems assayed using large-scale CRISPR screens, including resistance to Clostridium α-toxin in mouse ES cells, resistance to diphtheria and anthrax toxins in HeLa cells, and in vivo models of myeloid malignancy[31, 32].
- The use of a modified Cas9 to modulate gene expression recently been demonstrated at the genome scale. Rather than inactivating genes by the introduction of indels following DSBs, nuclease-dead Cas9 (dCas9) tethered to a transcriptional activating domain (CRISPRa) or inhibitory domain (CRISPRi) to alter the transcription of target genes [33, 34]. CRISPRa/i screens performed to identify mediators of resistance to ricin toxin, as well as a cholera/diptheria fusion toxin, in addition to negative selection analysis to identify essential genes. In contrast to previous over-expression technologies such as ORF libraries, which generally use one construct to represent each gene, CRISPRa offers the ability to examine many splice isoforms by modulating endogenous transcript levels rather than relying on

over-expression of a single cDNA-derived transcript. Likewise, by generating knockdown rather than knockout phenotypes, CRISPRi may produce phenotypes that are more similar to those in RNAi studies, and may be more informative when it is desirable or necessary to study phenotypic effects of gene dosage rather than complete loss of function. More recently, it has been show that modification of the tracrRNA-derived portion of the sgRNA to include RNA aptamers may been used as an alternative mechanism to recruit protein factors to DNA to modulate transcription[35, 36].

9. Clustered Regulatory Interspaced Short Palindromic Repeats:

- The research into the defense mechanisms of bacteria brought CRISPR to the scientific community. CRISPR stands for Clustered Regulatory Interspaced Short Palindromic Repeats and Segal et al describes it as a tool of choice for generating site-specific double-strand breaks in DNA[1].

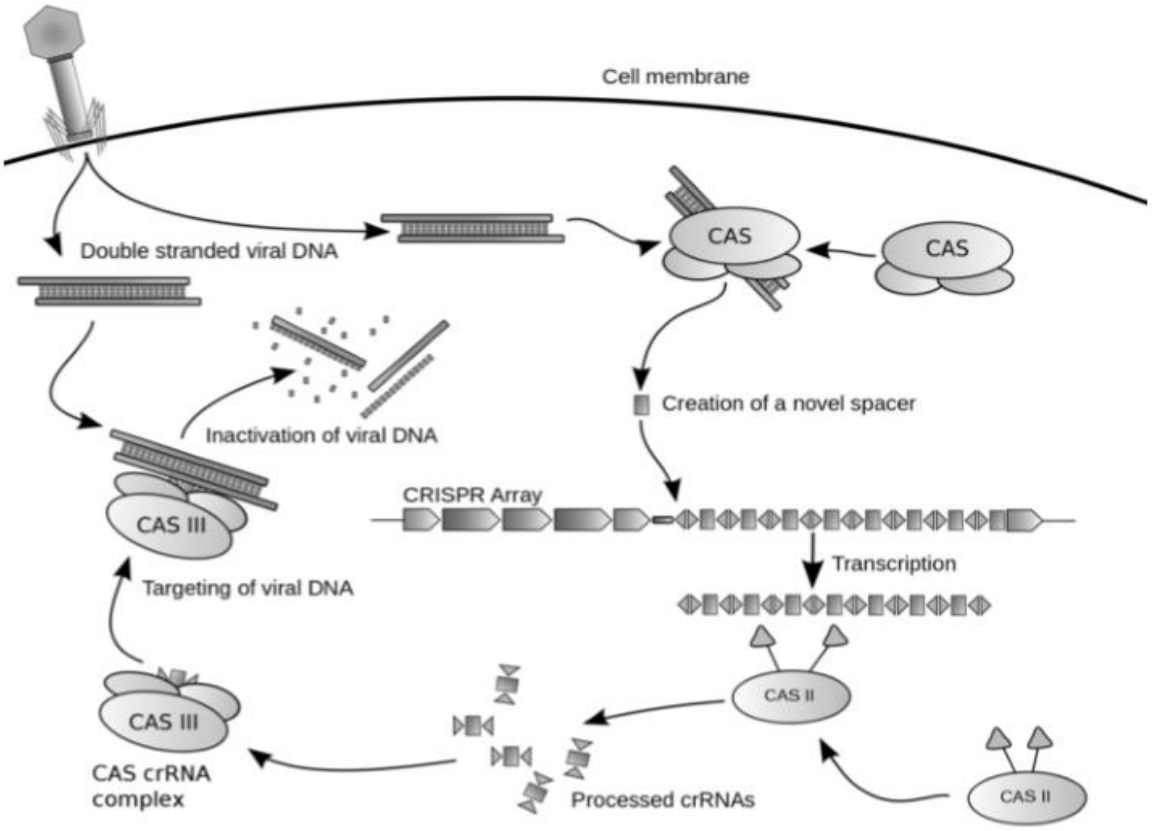


Figure 2: Natural role CRISPR Cas system in bacteria and Archaea.

- In 2007 Danisco team, a Copenhagen-based food ingredient company discovered a way to boost the defenses of the bacterium they were using against phages. S. thermophilus is a bacterium commonly used in dairy industries to ferment milk into yogurt and cheese. Danisco team exposed bacteria to a phage and showed that the bacteria vaccinated against that virus [1]. This established the natural role of CRISPR / CAS system in bacteria and Achaea. The system provides a kind of adaptive immunity from invading nucleic acids by guiding endonucleases to cut a specific non-host sequence [1]. This protects bacteria and Achaea against viruses and plasmids. In short, the immunity based on small RNA molecules that merge into protein complexes and can target specifically viral nucleic acids by base pairing. In general, CRISPR/ CAS defense has three steps (Figure 2).
- In the first one, the injected viral DNA discovered and a part of its DNA inserted to the host CRISPR array as a new spacer. This sequence is usually short and it is about 2-5 nucleotide and called protospacer adjacent motif (PAM). The second step of the response is the transcription of a CRISPR cluster into a long precursor-crRNA (pre-crRNA). The third and final step is the interference reaction. The mature crRNA merged with a larger CAS protein complex and used to identify and destroy the viral genome [1].
- Three different types of CRISPR Cas system (type I, II, III) have been found from the bacterium S. thermophilus (Figure 3) and all of them demonstrate the same architecture design [1].
- The CRISPR cluster can considered as a genomic DNA element whose first part consists of a series of short repeats, usually 24-37 base [5].

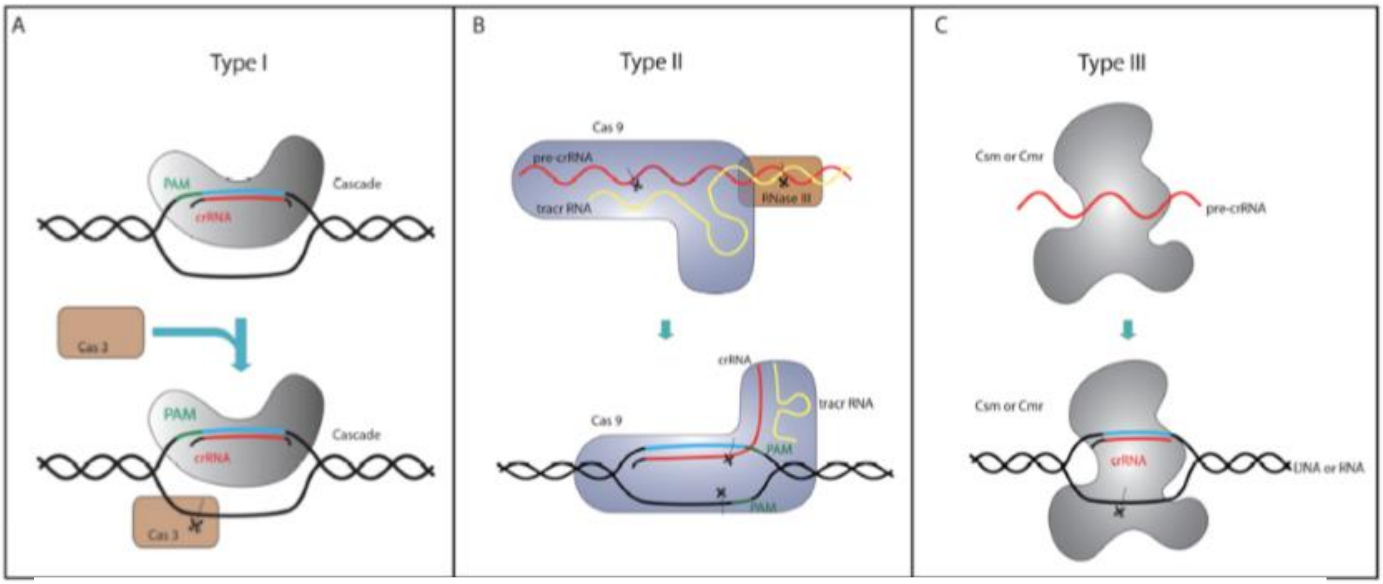


Figure 3: Three different types of CRISPR/ CAS system representing the interference step.

- Pairs that are separate by an exclusive spacer sequence with similar length. Those sequences are the ones that provide bacteria with adaptive immunity. Usually they originate from a viral genome. The second part of the CRISPR/ CAS system is the CAS endonuclease. It differs between the different types, and fulfills the function of providing immunity to bacteria.
- The type I and type III systems use Cas3 and Cas6 endonuclease to cleave pre-crRNA. In type I, the DNA that invades recognized by Cascade: crRNA complex. A PAM motif helps to identify the foreign DNA and nuclease Cas3 used to cleave the target DNA. The type III system uses the nuclease Cas6, on which crRNA bound and recognize the invading DNA or RNA. The type II CRISPR/ CAS system drafts Cas9 endonuclease. The Cas9 enzyme is a nuclease; a protein with the ability to cut DNA strands, and is equipped with two active cutting sites, one site for each strand of the DNA's double helix [1]. A great deal of structural literature is already available and has been important to understand the structural biology of the CRISPR system. It offers a great deal of information about the mechanisms and the evolution of the proteins that are involved [1].
- The type II system uses a completely different mechanism that requires only the use of Cas9 endonuclease to cleave the target sequence. In the type II system, Cas9 expressed with two more RNAs called crRNA (CRISPR RNA) and tracrRNA (transactivating crRNA). Together they form the sequence-specific endonuclease, which cleaves foreign genetic sequences to protect the host cells [1].
- A double strand break induced, cleaving DNA at a site that is complementary to the sequence of the guide RNA. In order for the type II system to be functional Cas9 endonuclease and a small guide RNA sequence are required [1].
- Until now, only type II system is commercially available and widely used for genome editing. There are differences among the different type of systems. One of these is in the interference reaction of both type I and III, relies on multi-protein complexes. This complicates the system and makes it hard to optimize. Another difference is that type III does not need a PAM sequence. That makes it more versatile but less specific [1].
- CRISPR is a new hot tool that has recently been included in the tool kit with the other genome editing techniques. So far, it displays advantages that cannot neglected. (i) Assembly speed, (ii) target efficiency, (iii) potential multi-targeting and (iv) low costs are a few of the many pros [1]. The biggest advantage of all is the (v) simplicity of its method. Until recently genome, engineering required the production of proteins that would have the power to recognize and bind to a specific DNA locus. With this bacterial endonuclease, Cas9, only a small RNA sequence needs to design and can target almost every part of DNA [1].

- Comparing to TALEN and ZNF, a single Cas9 protein can be retargeted by changing the sequence of the single guide RNA (gRNA) making the system easy to use [1]. An ordinary TALEN needs two new 1800 base pairs repeat to be assembled for every new target site whereas CRISPR/ Cas system requires just 20bp. One potential weakness of CRISPR could be that the eight base pairs that are furthest away from the PAM sequence, which is a NGG motif, are tolerant of single base mismatches. This could raise some concerns regarding the specificity of the system [1]. The limiting step of CRISPR technology is the level of RNA expression and assembly into the Cas9 system [1].
- So far, all the reported CRISPR approaches have generated double strands breaks (DSB) to the target sequence. Those breaks can be repaired either by homologous recombination (HR) either by non-homologous end joining (NHEJ). HR repairs fully the break since it uses the wild type allele as a donor template. However, NHEJ is a mechanism that creates mistakes, which can lead to insertion, deletion, frame shift and so on [1].

9.1. CRISPR/ Cas9 system:

- CRISPR refers to system type II that was first discovered at the bacterium S. thermophilus. A short section of nucleotides also referred as protospacer adjacent motif (PAM) identified by the crRNA (Figure 4). With the help of tracrRNA Cas9, nuclease specific double strand breaks caused. Commercially, those two RNA elements are combined into a single chimeric molecule termed guide RNA (gRNA) that enables the simultaneous expression alongside with Cas9 protein [1].

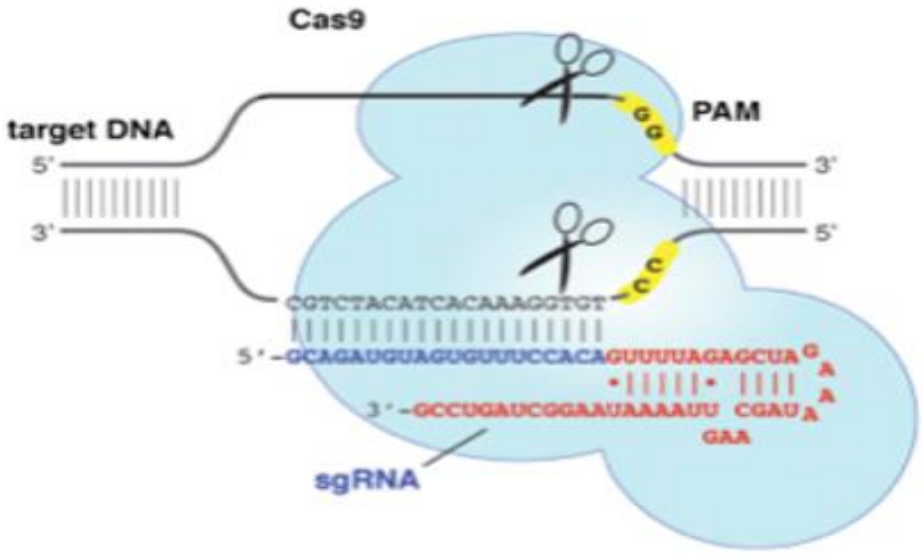


Figure 4: CRISPR/ Cas9 system

- The Cas9 nuclease localize in this way and targets the DNA sequence by a 20 nucleotides guide sequence [1]. This short guide uses common Watson Crick base pairing matching in order to identify the desired genomic locus. Moreover, this guided sequence can tolerate a certain amount of mismatches to the DNA target and this has a drawback since unwanted off-site mutagenesis occur [1]. In this way, the chimeric guide RNA can direct the Cas9 nuclease to almost any genomic locus, which followed by a 5'-NGG PAM motif [1]. The double strand breaks occurred from CRISPR preferably repaired by NHEJ mechanism. As already explained NHEJ is an error prone mechanism that helps introducing insertions, deletion or even frame shifts into mammalian cells. One interesting fact about CRISPR is the ability of multiplexing. The generation of up to five mutations with a single transfection event has observed [1].

10. HPRT1 gene:

- HGPRT [E.C. 2.4.2.8.] is the abbreviation for hypoxanthine-guanine phosphoribosyltransferase which is the enzyme encoded by the HPRT1 gene [1]. The hypoxanthine phosphorybol transferase (HPRT) gene is located on the long arm of the chromosome X of mammalian cells at Xq26-Xq2.7 position and generally is used as a gene model to investigate possible mutations in various mammalian cell lines [1]. HPRT consists of 44 kb of DNA and spread over nine exons, as can see in (Figure 5).
- The gene copied into the mRNA, which is 1,6 kb long. The protein is a tetramer and each

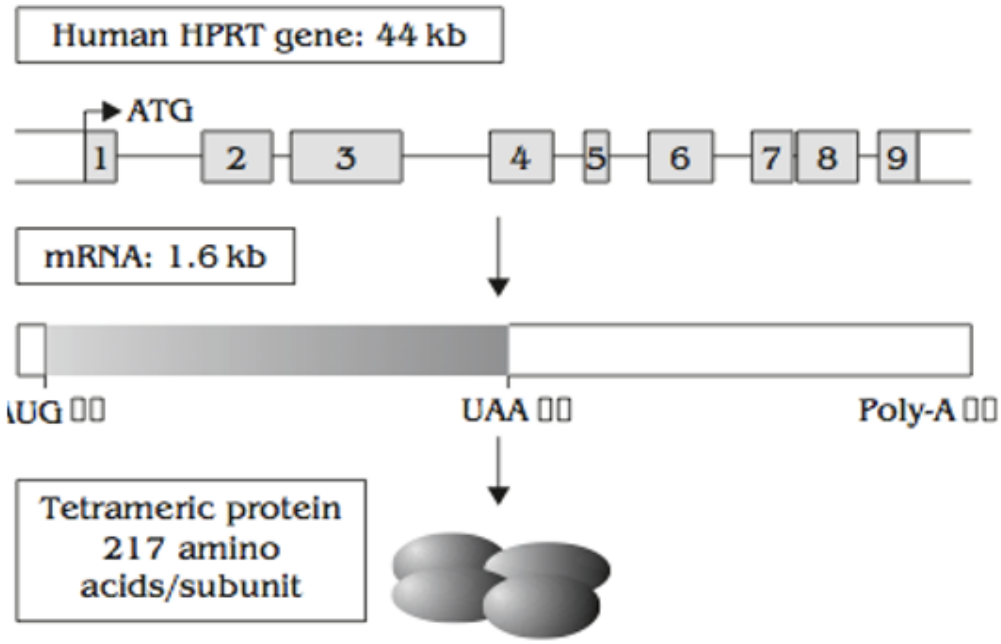


Figure 5: Molecular biology of HGPRTase (Nyhan, 2007)

subunit consists of 217 amino acids [1]. It inherited as an X-linked gene therefore, males generally affected, whereas females are heterozygous carriers and usually are asymptomatic. Up to date more than 300 disease-associate mutations in the HPRT1 gene have been identified [1]. The diagnosis based on clinical findings but also on enzymatic and molecular tests [1].

- The HPRT locus lies on chromosome X, hence only primary male cell lines can conveniently be studied to detect mutagenic effects. As its name suggests, HGPRT is a transferase that catalyzes the conversion of hypoxanthine to inosine mono-phosphate (IMP) and guanine to guanosine monophosphate (GMP), as can be seen in (Figure 6) [1]. HGPRTase actively

expressed in the cytoplasm of every cell of the body with the highest levels to been found in the basal ganglia [1]. HGPRT is a purine salvage enzyme, which transfers a 5-phosphoribosyl group from 5-phosphoribosyl 1-pyrophosphate (PRPP) to the purine. As an example, HGPRT catalyzes the reaction between guanine and phosphoribosyl pyrophosphate (PRPP) to form GMP (Figure 6 & Figure 7). In other words, it converts preformed purine bases to their respective nucleotides [1]. HGPRT has a central role during the generation of purine nucleotides through the purine salvage pathway [1]. Primarily HGPRT deals with purines that come from salvage pathway that occur from the degradation of DNA and purines are re-introduced into the purine synthetic pathways in this way (Figure 6).

Salvage pathway is a metabolic path that helps mammalian cells to obtain precursors for DNA synthesis and repair. More specifically it refers to the pathway from which, purines and pyrimidine are synthesized as intermediated from the degradative pathway of nucleotides. In other words, salvage pathway recovers and reintroduces bases during the degradation of DNA and RNA [1].

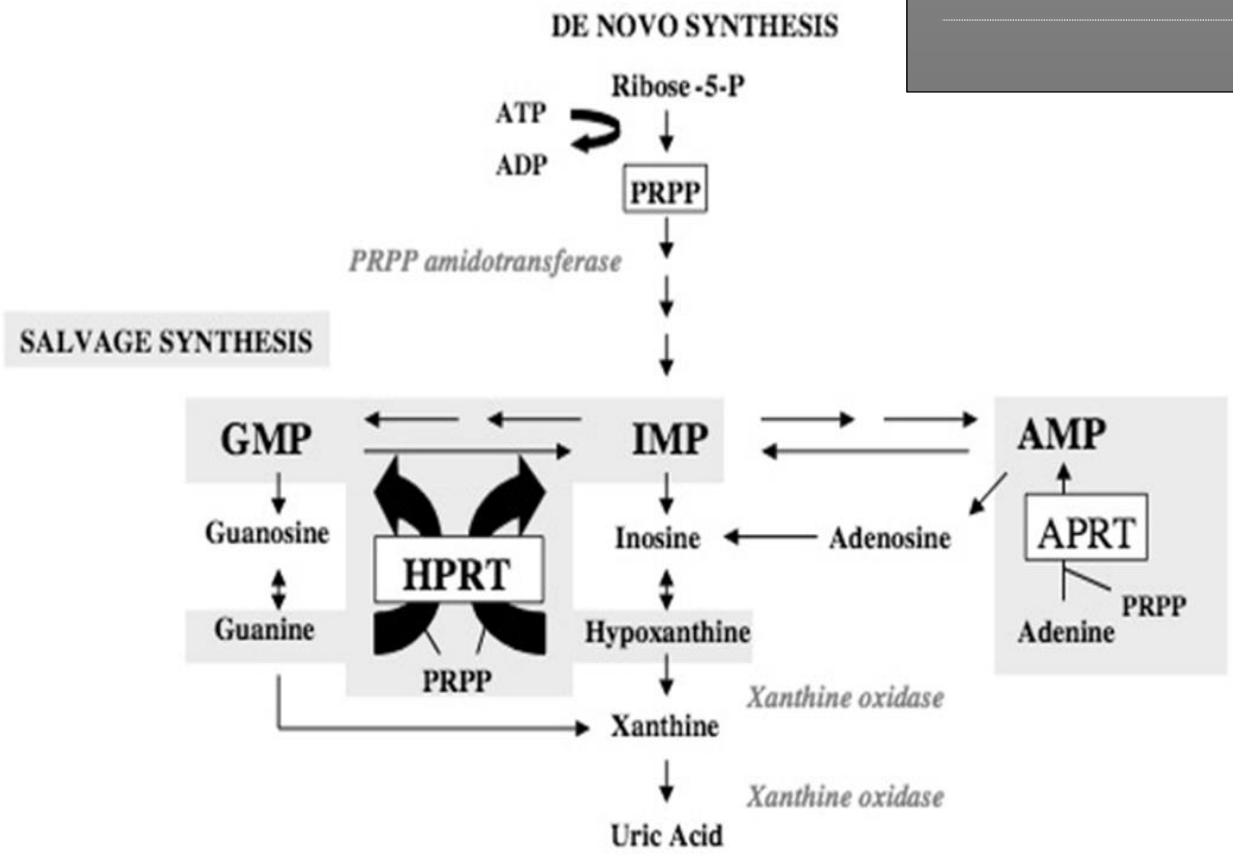


Figure 6: Metabolic scheme of purine metabolism .

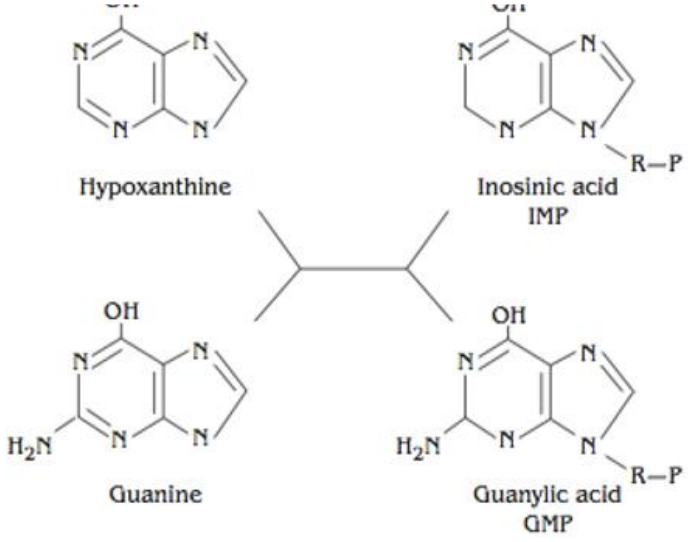


Figure 7: Reactions catalyzed by HPRTase.

- A selective readout assay has developed where since HPRTmutants can see as viable colonies when mutations destroy the functionality of the HPRT gene. This HPRT methodology assists in identifying positive selection, when mutations destroy the functionality of the HPRT gene; since the HPRT-, mutants can actually see as viable colonies [1]. There are three main benefits for making the HPRT gene mutation assay widely used. (i) The target gene encoded on the mammalian X chromosome, something that makes it easy to select for loss of function mutants in cells derived from males. (ii) There are simple and efficient systems for proving the loss of function with cells that survive in presence of 6-thioguanine (6TG). (iii) The HPRT gene conserved among various cell lines and can easily compared between other animal and human cells [1].

10.1. Mutations in the HPRT1 gene lead to hyperuricemia:

- Deficiency of hypoxanthine-guanine phosphoribosyltransferase (HGPRT) activity is generally associated with purine metabolism and can lead to increased excretion of the degradation product uric acid and a wide range of neurological disorders that depends on the level of the enzyme insufficiency [1]. It observed that a portion of the male population with a partial HGPRT deficiency has higher uric acid levels in their blood. Uric acid overproduction triggers the development of gouty arthritis and the formation of uric acid stones in the urinary tract [1]. Patients with partial deficiency at HGRPT enzyme display symptoms in different intensities. HGPRT is associated with two OMIM items, OMIM 300322 & OMIM 300323 caused by mutations that occur at the HPRT locus. Lesch-Nyhan syndrome (LNS, OMIM 300322) has the more severe inefficiency of the enzyme activity and it caused by mutations that occur at the HPRT locus. The other condition linked with HGPRT deficiency named as Kelley-Seegmiller syndrome (OMIM 300323) and it corresponds to a partial deficiency of the enzyme. Patients display some degree of neurological association but not as severe as LNS [1].

10.2. Lesch–Nyhan syndrome:

- Deficiency of the enzyme hypoxanthine-guanine phosphoribosyltransferase (HGPRT) causes a rare inherited disorder in different extend with syndromes like Lesch–Nyhan syndrome (LNS) or Kelley-Seegmiller syndrome [1]. Deficiency in the HGPRT enzyme can occurs because of mutation in the HPRT1 gene take place. The most severe deficiency in HPRT and most common disease of purine metabolism is the Lesch-Nyhan syndrome. When HGPRT partially deactivated, it allows uric acid to be accumulate in all body fluids and results in hyperuricemia and hyperuricosuria (linked with severe kidney and gout difficulties). Neurologically, reduced muscle control and adequate intellectual disability observed to patients [1]. The insufficiency of HGPRT can result in poor utilization of vitamin B12 something that may lead

During cell division, formation of DNA takes place and nucleotides are essential. Adenine and guanine are purine bases and thymidine and cytosine are pyrimidine bases. All of them are bound to deoxyribose and phosphate. Usually nucleotides are synthesized from amino acids. However a small part is recycled from degraded DNA of broken-down cells. This pathway is called salvage pathway. HGPRT is a salvage pathway enzyme for the purines. It channels hypoxanthine and guanine back into DNA synthesis. Deficiency of this enzyme has two main results. Firstly the cell breakdown products cannot be reused, so they are degraded. This increases the level of uric acid; a purine breakdown product. The second result is that the de novo pathway is stimulated due to an excess of PRPP (5-phospho-Dribosyl-1-pyrophosphate or simply phosphoribosyl-pyrophosphate).

to the development of megaloblastic anemia in boys. Most, but not all, patients with this deficiency have severe physical and mental problems through their entire life.

- Lesch-Nyhan syndrome (LNS) is an inborn X-linked inherited disease (Figure 8) or in other words, HPRT gene is a sex-linked locus. It affects infants' development for the first 4 to 6 months [1]. The classical form of the disease is a complete deficiency of the HGPRT enzyme and patients appear to have cognitive impairment, spasticity, dystonia, self-injurious behaviors and also increased concentrations of uric acid in blood and urine that can lead to nephropathy, urinary tract calculi and tophaceous gout. The most distinctive feature of all is the aggressive, self-injury behavior. Patients with this kind of phenotype never learn to stand unassisted or to walk normal [1]. Individuals that predominantly hemizygous males affected while heterozygous females are usually asymptomatic carriers.

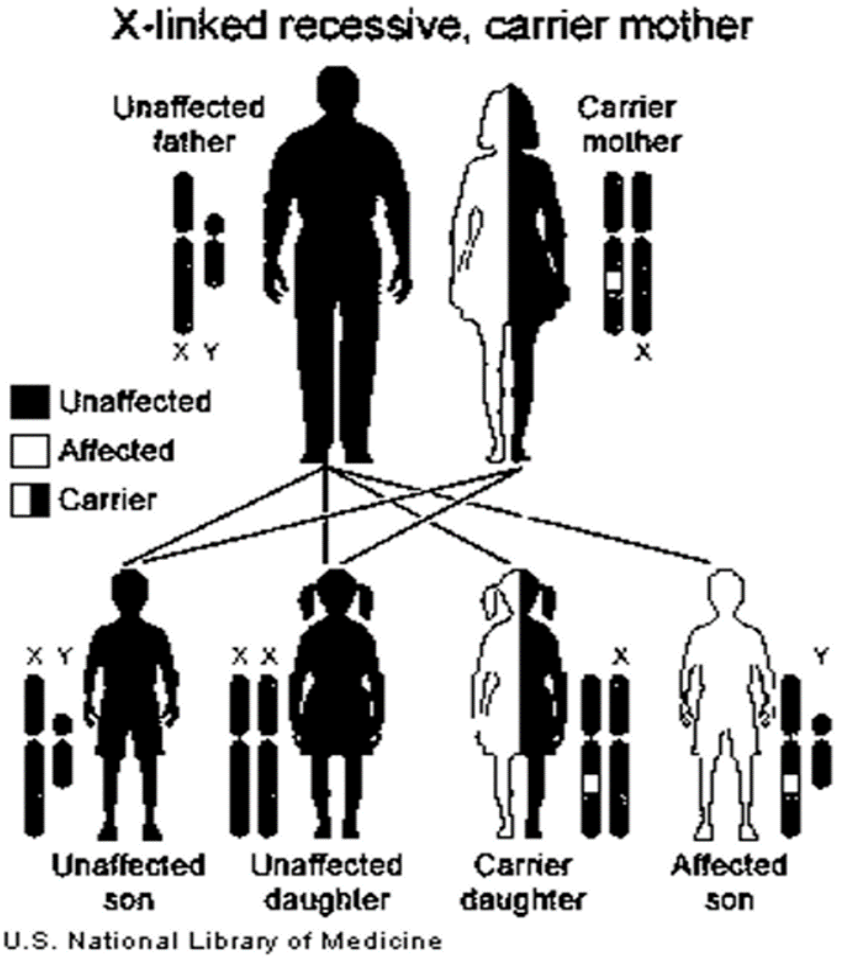


Figure 8: X- linked inherited disease

- These females have an unaffected copy of HPRT that prevents the disease from developing. The gene mutation carried by the mother and passed to her son. The father of an affected male cannot have transmitted the disease and not be the carrier of the mutant allele [1]. However, in one-third of all cases that have examined from new mutations it seems that there are not any family history records. The HPRT1 gene codes the enzyme hypoxanthine-guanine phosphoribosyltransferase (HGPRT, EC 2.4.2.8) [1]. The HGPRT enzyme is involved in the biochemical pathways that the body is using to produce purines, one of the compounds of DNA and RNA. There are a large number of mutations of known HPRT mutations, [1]. Usually mutations that marginally decrease the enzyme functionality by little do not cause severe syndromes like LNS but do produce a milder form of the disease. It seems that there is a hot spot localize in exon 3 where mutations found in 25.7 % of the families. Exons 1, 4 and 9 have implicated in deletion mutations that develop the disease. Amino acid Arginine 51 appears to be a hot spot for mutations in the HPRT1 gene [1]. In addition, reported ten sites of mutation have resulted in the loss of exon 7 from the cDNA.
- The research team has commented that the base sequences flanking exon 7 may have caused the mutation that provoked the splicing errors [1].

Lymphocytes: a group of white blood cells and a vital part of the immune system.

B cells: group of cells belong to lymphocytes group and derive from hematopoietic stem cells (HSC) in the bone marrow.

HAT medium: HAT is a selection medium, ordinarily used for mammalian cell culture and most often used for monoclonal antibody preparation.

Hybridoma is a technology that refers to the production of hybrid cell lines, or else called hybridomas. This is done by merging a specific antibody-producing B-lymphocyte with a cancerous B cell that is selected for its ability to grow in tissue culture and for an absence of antibody chain synthesis. The produced antibodies are all of the same single specificity and therefore are called monoclonal antibodies[1].

10.3. Application:

- B-lymphocytes express the enzyme that enables them to survive when fused to myeloma cells when growing on HAT medium in order to produce monoclonal antibodies. Hybridoma cells produce the monoclonal antibodies. A specific antigen injects into a mammal and procures the antibody production from the mouse spleen. If spleen cells fussed with cancerous immortalized immune cells, then myeloma cells derived. This hybrid cells are then clone to produce identical daughter clones. From those daughter clones, the desired antibody product secreted. In order to select hybridomas HAT medium used. HAT consists of hypoxanthine, aminopterin and

thymidine. The aminopterin inhibits the enzyme dihydrofolate reductase (DHFR), which is necessary for the de novo synthesis of nucleic acids. The cell has no other way to survive but to use the salvage pathway as an alternative.

- The salvage pathway needs functional HGPRT. In HAT medium HGPRT- cell lines will die since they are not able to synthesize nucleic acids from the salvage pathway. Only HGPRT+ cell lines, which are the hybridoma cells and plasma cells, will survive in presence of aminopterin. Plasma cells will eventually die, as they are not immortelle cell lines; however, hybridoma cells that are immortalized will survive. Those hybridoma cells will cloned to produce identical daughter clones secreting the monoclonal antibody product.

11. Optimization of on-target activity:

- When attempting to design sgRNAs to target a gene of interest, CRISPR technology presents an embarrassment of riches, as the number of potential sgRNA sequences scales with the size of the gene. As either the coding or the template strand of the DNA may serve as a target, the S. pyogenes PAM site (NGG) appears on average once every 8 nt. How, then, to choose from the dozens to hundreds of potential sgRNA sequences? Avoidance of off-target activity, discussed in more detail below, may use to eliminate some sgRNA sequences, and target-specific features may be incorporated to choose sgRNAs that are more likely to be effective. For example, to generate loss-of-function alleles of protein coding genes, targeting closer to the N-terminus increases the chance that a frame shift allele will be deleterious, as more of the coding sequence will disrupted. Likewise, for CRISPRa technology, proximity to the transcriptional start site is critical for recruiting appropriate factors [34, 35]. The first genome-wide libraries designed according to these general criteria, but did not take into account any sequence-specific information that may enhance on-target activity.
- The first evidence that sequence-specific rules lead to better on-target activity came from two analyses of a large-scale library [27]. First, pull-down of Flag-tagged Cas9 followed by sequencing of the associated sgRNAs showed that some sequence features led to better association with Cas9. Second, sgRNAs targeting essential genes classified by their ability to deplete from a population of cells as a surrogate for their ability to induce null alleles. These two approaches gave similar results, and showed that the choice of sgRNA sequence may optimized for greater activity. The next study to address this issue utilized a tiling approach to create libraries covering all possible target sites for a selection of mouse and human cell-surface markers [37]. Flow cytometry allowed direct detection and isolation of sgRNAs that led to bi-allelic loss-of-function of the protein of interest. This set of over 1800 sgRNAs of varying efficacy used to characterize sequence features that led to increased activity. Interestingly, an extended PAM sequence beyond the canonical NGG motif was

shown to affect activity, with CGGH being the most-optimal sequence (where H = A, C or T). This quantitative analysis led to a predictive model for designing optimal sgRNA sequences for any target of interest.

- Initial studies have shown that there are sequence features that affect the ability of Cas9 to bind sgRNAs, cleave DNA, and result in a loss-of-function allele. Further characterization of these rules, and incorporation of them into library design, will be critical for successful deployment of genome-wide libraries, and will result in libraries with progressively higher fractions of active sgRNAs. Indeed, a critical analytical step in determining hits from a screen is the requirement for multiple independent sgRNAs targeting the same gene to score; thus, sgRNAs with low activity increase the false-negative rates of a screen. While one approach is simply to increase the size of a library to increase statistical power, this comes with the trade-off of requiring more cells for a screen. Larger libraries increase the cost of a screen, and, for many biological models, such as use of primary cells or in vivo screens, increased cell numbers simply may not be feasible.

12. Minimizing off-target activity:

- While optimizing on-target activity has clear ramifications for use of CRISPR technology, understanding off-target effects is equally important to avoid erroneous interpretation of experimental results. Initial experiments in mammalian cells showed that S. pyogenes Cas9 cleavage activity tolerates a number of mismatches between the sgRNA and the DNA target [38, 39]. In general, mismatches closer to the 5′ end of the RNA tolerated than mismatches close to the PAM. Currently; there are not enough data to create fully predictive models of a sgRNA will lead to appreciable levels of off-target DNA cleavage, as the exact base composition of the mismatch appears to affect activity [38, 40]. Further, non-symmetric pairing between the RNA and DNA, e.g. creation of bulges on either side of the duplex, is also tolerated by Cas9 [41]. These interactions are particularly troubling as current rapid alignment algorithms that scan for potential off-target sites are blind to these interactions, meaning that all current design software ignores this type of potential off-target activity. Given the tolerance of mismatches by Cas9 and the likelihood of finding certain sequence motifs in multiple locations in the genome, it is essential to consider possible off-target locations when designing sgRNA sequences.
- Several strategies have developed to minimize the off-target effects of Cas9. One uses a shortened sgRNA of 17 NT, rather than the standard 20 NT sequence; under the conditions tested, the 17 nt sequence shown to have comparable on-target activity but decreased tolerance to mismatches [42]. To date, no large-scale libraries have created using shorter sgRNAs, and this will be an important test of the generalizability of the efficacy of this

approach. An additional approach borrows conceptually from zinc-finger and transcription activator-like effector nuclease designs, by requiring two sgRNAs to find target sites in close proximity in the genome. One version of this approach involves a 'nickase' version of Cas9 (Cas9n), where one of the two DNA cleavage domains is inactivated, and thus each sgRNA performs a single-strand cut [43]. When two nicks occur physically close to each other on each strand, the result is a DSB. This design minimizes off-target modifications, as single nicks in genomic DNA at off-target sites often repaired without introduction of indels. Similarly, monomers of FokI may fused to dCas9 such that two sgRNA sequences are required to reconstitute active FokI activity and initiate a DSB [44]. Off-target rates have shown to decrease > 1000-fold using these approaches. The need to deliver two sgRNA sequences may limit the utility of this approach for large-scale screens, but this approach is preferred in cases where specificity is especially important (e.g. the creation of edited cell lines).

- A series of genome-wide chromatin immunoprecipitation experiments followed by massively parallel DNA sequencing (ChIP-seq) of dCas9 programmed with various sgRNAs has shown that Cas9 is more likely to associate with DNA in an open chromatin state [45, 46]. Surprisingly, a small seed region of 5–8 nucleotides was shown to be sufficient to allow binding of Cas9 to hundreds of DNA target sites, raising the possibility of rampant off-target effects. Importantly, however, the same sites did not show evidence of cutting when nuclease-active Cas9 used, reaffirming that near-perfect sequence complementarity is required for robust endonuclease activity. In vitro data have also shown that Cas9 first interrogates PAM sites in DNA and then spends progressively more time associated with DNA containing 3′→5′ complementarity to the sgRNA sequence [47]. These data, combined with the specificity shown by CRISPRa and CRISPRi studies discussed above, suggest that the initial step(s) of Cas9 target-site recognition while of a time scale long enough to be detected by ChIP-seq, are largely transient in nature unless there is complementarity throughout the length of the sgRNA.
- Two studies have performed whole-genome sequencing of cells following editing with Cas9 [48, 49]. While this approach is not feasible as a routine test, and the number of unique sgRNA sequences surveyed was limited, the degree of editing at potential off-target sites was no greater than at other unrelated sites in the genome. Importantly, these studies suggest that mutations that arise during clonal expansion of single cells are as much a concern as any off-target effects of Cas9. Analogous to the requirement for multiple independent sgRNAs targeting a gene to score before considering that gene as validated, experiments that rely on edited single cell clones should utilize multiple independent clones to decrease the chances of a random mutation interfering with interpretation.

13. Development and future challenges:

- The amount of progress made with CRISPR technology in the past several years is staggering, and the future offers many opportunities and challenges. For example, Cas9 from S. pyogenes was one of the first Cas9 enzymes studied and continues be extensively used. While a small number of Cas9 proteins from other species have begun to explored, there are thousands of Cas9 proteins endogenous to various prokaryotes that may be better suited to a variety of purposes. S. pyogenes Cas9 is quite large and thus is difficult to efficiently package into many viral vectors. Smaller Cas9 proteins have described but not fully developed for widespread use in screening and routine genome editing. Additional considerations include the tolerance for appending additional functionalities, such as fusions with transcriptional domains, as well as the ability to fold and localize efficiently in mammalian cells. Likewise, the search space of the genome available to Cas9 remains an open question, as the ChIP-seq experiments discussed above suggest the possibility that sequences in closed chromatin may be more refractory to modification. Furthermore, the NGG PAM requirement of S. pyogenes Cas9 limits the sequence space accessible for targeting, which becomes increasingly problematic when the location of editing is of utmost importance, such as when knocking in a reporter construct at the N- or C-terminus of a protein, or modeling a specific SNP. It is worth noting, however, that S. pyogenes Cas9 may soon become fixed in its usage in practice, as any new Cas9 protein will require a great deal of specificity and activity characterization before it may be used appropriately.
- The RNA component of the CRISPR system will also undergo further refinement. Through creation of larger datasets and enhanced modeling techniques, the variable sequence of the sgRNA will continue to undergo design optimization to increase on-target activity and avoid off-target effects. Likewise, the tracrRNA-derived constant region of the sgRNA that associates with Cas9 tolerates mutations in some regions, thus opening the possibility those novel sequence modifications may found that increase the efficacy of this module [35, 50]. Library design and testing lends itself to rapid cycles of improvement and innovation when pooled screens are a possibility. However, assays such as gene expression or image-based profiling require arrayed reagents, i.e. a library of one perturbation per well. Here, achieving full penetrance in the population of cells that receive the perturbation is critical. Further, as the costs of producing such arrayed libraries are an order of magnitude more expensive than for pooled libraries, a careful evaluation of the robustness of CRISPR technology for these approaches is required.
- Despite improvements in creating site-specific dsDNA breaks, first with zinc-finger nucleases, then transcription activator-like effector nucleases and now with Cas9, site-specific incorporation of an exogenously provided DNA sequence remains a highly inefficient process (< 1% efficiency) in many cell types. While strong selection may be used

for certain types of edits, such as knocking in a fluorescent protein and using flow cytometry to enrich for edited cells, many desired gene editing events do not lend themselves to any means of physically separating the small fraction of edited cells from the remainder of the population. For example, it would be highly desirable to model single base mutations discovered in cancer genome sequencing at high throughput [51]. Likewise, the majority of significant associations from genome-wide association studies found in non-protein-coding regions of the genome, and thus gene editing of the endogenous locus is the only way fully understand how a specific variant affects the state of the cell [52]. The current approach of assaying hundreds of single cells isolates to find a successfully edited clone does not scale well to high-throughput assays.

- There are several possibilities for potential improvements. For example, inhibiting the rate of NHEJ by small interfering RNA-based knockdown of critical component(s) may create a transient window during which homologous recombination is the favored repair pathway [53]. Synchronizing cells in S phase, when homologous recombination is endogenously more highly active, is another potential means of increasing the rate of incorporation [54]. Enhanced detection of edited cells is also a possibility; the current standard approach requires cells to lyse for detection of editing by endonuclease- or sequencing-based assays, thus requiring prior expansion and replication of individual cells, usually in a multi-good format, in order to recover viable clone's post-detection. The ability to detect specific sequences in pooled format, such as by using a nuclease-inactive Cas9 coupled to an amplifiable fluorescence signal, would enable flow-cytometry-based enrichment of successfully edited cells [55]. Finally, template-based repair may also improve by bringing the repair template in physical proximity to the repair site, rather than by the current approach of random diffusion in the cell. This may potentially be accomplished by affinity moieties on the DNA that enable association with domains appended onto Cas9, or by modification of the sgRNA module to include the desired repair information. These and other approaches are necessary to enable genome-editing experiments on a larger scale.

14. Conclusion:

- A novel finding as an adaptive immune system in bacteria, CRISPR technology has also proved readily transferrable for use in many organisms, most notably mammalian cells. CRISPR technology may use both to engineer better model systems that more accurately reflect human disease, and to perturb those models to more precisely identity relevant cellular circuitry and potential therapeutic interventions. Furthermore, CRISPR is the first genome-editing technology that may easily performed in any lab through use of inexpensive, easy-to-assemble reagents, democratizing both the process of editing and the work of building new tools from the basic components. Although we may be in the heyday of CRISPR technology, we remain at the early stages of fully understanding the system and expanding its potential.
- CRISPR Cas9, type II system, was used in an attempt to knock out the HPRT1 gene expressing the HGPRTase enzyme in human lymphocytes.
- A possibility was found also to use this technology on animal's cells, therefore, we can conclude that the possibility to modify a primal cell's DNA will allow us in the future to modify an entire human genome.

References:

1. Roidos, P., *Genome editing with the CRISPR Cas9 system.*, in *KTH – ROYAL INSTITUTE OF TECHNOLOGY DEPARTMENT OF BIOTECHNOLOGY.* 2014, Center of Molecular Medicine Karolinska University Hospital.: School of Biotechnology, KTH – Royal Institute of Technology, Stockholm, Sweden. p. 59.
2. Ishino, Y., et al., *Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product.* Journal of bacteriology, 1987. **169**(12): p. 5429-5433.
3. Jansen, R., et al., *Identification of genes that are associated with DNA repeats in prokaryotes.* Molecular microbiology, 2002. **43**(6): p. 1565-1575.
4. Bolotin, A., et al., *Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin.* Microbiology, 2005. **151**(8): p. 2551-2561.
5. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-1712.
6. Makarova, K.S., et al., *Evolution and classification of the CRISPR–Cas systems.* Nature Reviews Microbiology, 2011. **9**(6): p. 467-477.
7. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes.* Science, 2008. **321**(5891): p. 960-964.
8. Gasiunas, G., et al., *Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.* Proceedings of the National Academy of Sciences, 2012. **109**(39): p. E2579-E2586.
9. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-607.
10. Jinek, M., et al., *A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-821.
11. Sapranauskas, R., et al., *The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli.* Nucleic acids research, 2011: p. gkr606.
12. Esvelt, K.M., et al., *Orthogonal Cas9 proteins for RNA-guided gene regulation and editing.* Nature methods, 2013. **10**(11): p. 1116-1121.
13. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-826.
14. Jinek, M., et al., *RNA-programmed genome editing in human cells.* elife, 2013. **2**: p. e00471.
15. Yang, H., et al., *One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering.* Cell, 2013. **154**(6): p. 1370-1379.
16. Knott, S.R., et al., *A computational algorithm to predict shRNA potency.* Molecular cell, 2014. **56**(6): p. 796-807.
17. Jackson, A.L., et al., *Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity.* Rna, 2006. **12**(7): p. 1179-1187.
18. Sigoillot, F.D. and R.W. King, *Vigilance and validation: Keys to success in RNAi screening.* ACS chemical biology, 2010. **6**(1): p. 47-60.
19. Jackson, A.L., et al., *Expression profiling reveals off-target gene regulation by RNAi.* Nature biotechnology, 2003. **21**(6): p. 635-637.
20. Sigoillot, F.D., et al., *A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens.* Nature methods, 2012. **9**(4): p. 363-366.
21. Echeverri, C.J., et al., *Minimizing the risk of reporting false positives in large-scale RNAi screens.* Nature methods, 2006. **3**(10): p. 777-779.
22. Westbrook, T.F., et al., *A genetic screen for candidate tumor suppressors identifies REST.* Cell, 2005. **121**(6): p. 837-848.
23. Luo, B., et al., *Highly parallel identification of essential genes in cancer cells.* Proceedings of the National Academy of Sciences, 2008. **105**(51): p. 20380-20385.
24. Cheung, H.W., et al., *Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer.* Proceedings of the National Academy of Sciences, 2011. **108**(30): p. 12372-12377.
25. Barbie, D.A., et al., *Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1.* Nature, 2009. **462**(7269): p. 108-112.

26. Strezoska, Ž., et al., *Optimized PCR conditions and increased shRNA fold representation improve reproducibility of pooled shRNA screens.* PLoS One, 2012. **7**(8): p. e42341.
27. Wang, T., et al., *Genetic screens in human cells using the CRISPR-Cas9 system.* Science, 2014. **343**(6166): p. 80-84.
28. Shalem, O., et al., *Genome-scale CRISPR-Cas9 knockout screening in human cells.* Science, 2014. **343**(6166): p. 84-87.
29. Whittaker, S.R., et al., *A genome-scale RNA interference screen implicates NF1 loss in resistance to RAF inhibition.* Cancer discovery, 2013. **3**(3): p. 350-362.
30. Huang, S., et al., *MED12 controls the response to multiple cancer drugs through regulation of TGF-β receptor signaling.* Cell, 2012. **151**(5): p. 937-950.
31. Koike-Yusa, H., et al., *Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library.* Nature biotechnology, 2014. **32**(3): p. 267-273.
32. Heckl, D., et al., *Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing.* Nature biotechnology, 2014. **32**(9): p. 941-946.
33. Gilbert, L.A., et al., *CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.* Cell, 2013. **154**(2): p. 442-451.
34. Gilbert, L.A., et al., *Genome-scale CRISPR-mediated control of gene repression and activation.* Cell, 2014. **159**(3): p. 647-661.
35. Konermann, S., et al., *Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex.* Nature, 2014.
36. Zalatan, J.G., et al., *Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds.* Cell, 2015. **160**(1): p. 339-350.
37. Doench, J.G., et al., *Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation.* Nature biotechnology, 2014. **32**(12): p. 1262-1267.
38. Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nature biotechnology, 2013. **31**(9): p. 827-832.
39. Fu, Y., et al., *High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.* Nature biotechnology, 2013. **31**(9): p. 822-826.
40. Fu, B.X.H., et al., *Landscape of target: guide homology effects on Cas9-mediated cleavage.* Nucleic acids research, 2014: p. gku1102.
41. Lin, Y., et al., *CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences.* Nucleic acids research, 2014: p. gku402.
42. Fu, Y., et al., *Improving CRISPR-Cas nuclease specificity using truncated guide RNAs.* Nature biotechnology, 2014. **32**(3): p. 279-284.
43. Ran, F.A., et al., *Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity.* Cell, 2013. **154**(6): p. 1380-1389.
44. Guilinger, J.P., D.B. Thompson, and D.R. Liu, *Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.* Nature biotechnology, 2014. **32**(6): p. 577.
45. Kuscu, C., et al., *Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease.* Nature biotechnology, 2014. **32**(7): p. 677-683.
46. Wu, X., et al., *Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells.* Nature biotechnology, 2014. **32**(7): p. 670-676.
47. Sternberg, S.H., et al., *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.* Nature, 2014. **507**(7490): p. 62-67.
48. Smith, C., et al., *Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs.* Cell stem cell, 2014. **15**(1): p. 12.
49. Hartenian, E. and J.G. Doench, *Genetic screens and functional genomics using CRISPR/Cas9 technology.* FEBS journal, 2015. **282**(8): p. 1383-1393.
50. Nishimasu, H., et al., *Crystal structure of Cas9 in complex with guide RNA and target DNA.* Cell, 2014. **156**(5): p. 935-949.
51. Lawrence, M.S., et al., *Discovery and saturation analysis of cancer genes across 21 tumour types.* Nature, 2014. **505**(7484): p. 495-501.

52. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
53. Shrivastav, M., L.P. De Haro, and J.A. Nickoloff, *Regulation of DNA double-strand break repair pathway choice.* Cell research, 2008. **18**(1): p. 134-147.
54. Lin, S., et al., *Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery.* Elife, 2014. **3**: p. e04766.
55. Tanenbaum, M.E., et al., *A protein-tagging system for signal amplification in gene expression and fluorescence imaging.* Cell, 2014. **159**(3): p. 635-646.