

Thesis for the degree Master of Science

Submitted to the Scientific Council of the Weizmann Institute of Science Rehovot, Israel עבודת גמר (תזה) לתואר מוסמך למדעים

מוגשת למועצה המדעית של מכון ויצמן למדע רחובות, ישראל

By Noa Aharon- Hefetz <sub>מאת</sub> נועה אהרון- חפץ

בחינת הקשר בין ביטוי של רנ"א מעביר לגדילה תאית באדם

Examination of the relation between tRNA expression and human cellular growth

Advisor: Prof. Yitzhak Pilpel מנחה: פרופ' יצחק פלפל

January, 2017

טבת, תשע"ז

## 1. Abstract

Multicellular organisms constantly face a dichotomy between proliferation- differentiation states. On the one hand, specialist cells are necessary for tissue functionality, which requires cells to go through differentiation processes. On the other hand, differentiation restricts the cell's ability to proliferate, which is also an important cellular feature to maintain tissue integrity. Therefore, tight regulation on the proliferation- differentiation decision must take place, to prevent development of cancer and degenerative diseases. To date, there are many known cellular factors that are responsible for the proliferationdifferentiation regulation. These were traditionally dominated by studies which aimed to decipher the transcription and post-transcription regulatory networks that control the mRNA level of key genes in several signaling pathways. Lately, increasing amount of evidences indicates for the relation between translation regulation and cellular state are accumulating. Briefly, cells in different cellular states are using distinct translation regulation programs, that manifest themselves by specific tRNA expression pattern that is complementary to the codon usage of genes expressed at this cellular state. The correlation between the proliferative state of the cell and its tRNA pool raises the question of causality: Does the tRNA pool determine the cellular state, or does the different cellular state dictate distinct expression of tRNA sub-pools? In this project, we aim to address this question. More specifically we asked how changes in the levels of different tRNAs belonging to distinct tRNA groups ("pro-proliferation tRNAs" pool and "pro-differentiation tRNAs" pool, as defined by expression regulation upon changes in cellular growth) influence cell proliferation and viability. To that end, we genomically edited tRNA genes using CRISPR-Cas9 technology in three human cell lines that differ in their proliferation rate, and evaluated the effect of tRNA knock-out on various cellular features. First, we performed competition experiment between HeLa cells carrying sgRNA targeting different tRNAs and measured their fitness. Interestingly, pro- proliferation tRNAs knock-out was more deleterious to the HeLa cell's fitness than prodifferentiation tRNAs knock-out. Next, we investigated proliferation rate and cell cycle progression of cells with deleted tRNA genes. Here, we didn't receive conclusive results as in the competition experiment, although we can note that in general, HeLa cells were more negatively affected by proproliferation tRNAs knock-out compared to less proliferative cells. Our further aim is to explore the role of tRNAs in proliferation- differentiation transition process, like serum starvation and starvation release, senescence induction, reprogramming, ES cells differentiation and trans-differentiation. We hope this project will broaden our knowledge on translation regulation in relation to cellular state, specifically in cancer. Moreover, we intend to study the notion that tRNAs have an active role in cellular state determination, which can uncover a new field in cancer therapy.

# **Table of Contents**

1. Abstract	2
2. Introduction	4
2.1. Proliferation- differentiation decision	4
2.2. tRNA genomic architecture	5
2.3. Changes in tRNA expression levels	6
2.4. tRNA expression and cancer	7
2.5. Genome editing and CRISPR- Cas9 system	8
3. Goals of the study	10
4. Methods	11
4.1. Cell lines	11
4.2. CRISPR plasmids	11
4.3. Lentiviral plasmids	12
4.4. Medium	12
4.5. sgRNA cloning	13
4.6. Cells transfection	16
4.7. Cells transduction	16
4.8. sgRNA competition	16
4.9. Library preparation and deep- sequencing	17
4.10. Deep- sequencing- Data processing	17
4.11. Machine learning techniques for fitness prediction	18
4.12. Generating WI38 fast and slow single clones with tRNA <sup>Ile</sup> AAT knock- out	19
4.13. Proliferation assay	20
4.14. Cell cycle analysis	21
5. Results	22
5.1. Comparing the fitness of tRNA- deleted cells using competition based assay	22
5.2. Creating WI38 fast and slow single clones with characterized tRNA deletion	27
5.3. Measuring proliferation rate of tRNA- deleted cells using MTT based proliferation assay	31
5.4. Exploring cell cycle progression and S phase duration of tRNA- deleted cells using FACS analysis	34
5.5. Hierarchical clustering of all tRNA knock-outs, cell lines and assays	41
6. Discussion	43
7. Literature	49
8. Acknowledgments	51

# 2. Introduction

# 2.1. <u>Proliferation-differentiation decision</u>

One of the most prominent features of multicellular tissues is the proliferation – differentiation decision. Each tissue in a complex organism needs to be able to coordinately function with other tissues, and also within the tissue itself. Moreover, each tissue should have the ability of self-sustain in case of injury or programmed cell death. For that, proper tissue homeostasis requires precise control of replication and differentiation/ quiescence states. Impairment of these processes may result in degenerative or neoplastic diseases<sup>1, 2</sup>.

Throughout the years, a major interest and effort was invested to investigate proliferation- differentiation decision in different aspects of life. For example, Klochendler et al. found that two distinct cell populations, replicating cells and quiescence cells, colonize adult mouse liver adjacent to each other. By preforming Affymetrix expression microarray, they showed strong distinction in the transcriptional program between the two cell populations. Moreover, replicating cells showed high expression of genes associated with cell cycle, mitosis, DNA replication as well as repair and downregulation of liver specific-gene function, while in the quiescent cell population, this transcription pattern indicated the opposite trend<sup>1</sup>.

Beside proliferating populations that reside in quiescent tissue, additional aspect of proliferationdifferentiation transition in nature is the development of malignant transformations. This process can be divided into several stages, each of which is characterized with accumulated genetic alterations<sup>3,4</sup>. Milyavsky et al. described the transcriptional program involved in an *in-vitro* cellular model of malignant transformation. The model consists of WI38 human primary fibroblasts, in which replicative senescence was overcome by expression of hTERT, resulting in sustained proliferation of the cells. Following hTERT induction, cells kept on growing to generate cell line, each of which are characterized with accumulating genetic alteration, some were artificially introduced and others occurred spontaneously. In the end of the transformation process, the cells were characterized with higher proliferation rate, defective contact inhibition checkpoint and sensitivity to H-Ras transformation, in comparison to the primary cells<sup>3</sup>. In this project, we used two cell lines generated from this transformation process. WI38/hTERT<sup>slow</sup> was generated from the beginning of the process and had similar features to the primary cells, and WI38/hTERT<sup>fast</sup>, which was generated towards the end of the process and had similar features to the transformed cells. In their paper from 2005, Milyavsky et al. performed microarray analysis to investigate the transcriptional signature of each cell line throughout the transformation process. In general, they showed that as the transformation process progressed, the expression pattern gradually shifted from differentiation- related genes, cell cycle regulators and apoptotic genes expression to translation- related genes and cancer- associated genes expression<sup>4</sup>.

Another biological process that exemplifies a switch from differentiation to proliferation state is the cellular reprogramming process. Reprogramming is an interesting process of transition from fully differentiated cell back to pluripotent, ESC like, proliferating cell. In 2013, Rais et al. reported on a new reprogramming method that can achieve deterministic and synchronized induced pluripotent stem cells (iPSC) reprogramming with high efficiency within 8 days in mice, following Oct4, Sox2, Klf4 and Myc (OSKM) induction and depletion of Mbd3/NuRD co-repressor complex<sup>5</sup>. RNA-seq revealed sequential activity of genes expressed in the process of converting MEF to naïve iPSC. Those genes represented the transition from the somatic program to the pluripotent one (Zviran et al. in preparation).

The differences between proliferating and differentiated cells that are manifested in transcriptomics and proteomics unique patterns are regulated in a precise manner. Regulation of the proliferation-differentiation transition was intensively investigated throughout the years, when the dominating notion is that specific signaling pathways serve as the main guardians of cell state control<sup>6</sup>. The signaling pathway that promotes proliferation is mainly composed of highly regulated cyclin- depended kinases (Cdk) that activate cascades of events that lead to expression of proliferation- related genes<sup>6</sup>. Later studies revealed that these signaling components, specifically Cdk inhibitors (CKIs), do not only induce cell proliferation, but also have the ability to induce differentiation and maintain this state<sup>2</sup>. Moreover, proliferation and differentiated luteal cells in CDK deficient mice), can show proliferative features such as active DNA synthesis, and cells defined as terminally proliferating (i.e. tumor cells), can restore differentiation phenotype, when cell cycle arrest is enforced<sup>2</sup>.

In recent years, the knowledge of proliferation and differentiation regulation and cell fate determination expanded further more than the traditional signaling pathway control. One such example is the role of miRNA in the cell fate determination during embryonic development. Similar to transcription factors, a single miRNA molecule can regulate the expression of numerous genes. Moreover, their spatial and temporal expression is regulated while being sensitive to changes in the cellular state. These characteristics enable miRNA to be an important candidate for cell fate master regulator<sup>7</sup>.

### 2.2. <u>tRNA genomic architecture</u>

Transfer RNA (tRNA) molecules are a notable component in the cellular translation machinery, in which the tRNA molecules mediate between the mRNA codons and the cognate amino acid by complement anti-codon. Codon redundancy for the 20 amino acids requires tRNAs that encode for the same aminoacid but with different anti-codons (tRNA isoacceptors) that read the different codons for every aminoacid<sup>8, 9</sup>. The isoacceptor families vary in the number of gene members. Interestingly, the number of different tRNA species in the genome is greater than the total number of tRNA isoacceptors- that is, most tRNA isoacceptor can be coded by families of several genes that have the same anticodon, but might slightly differ in the sequence of the tRNA body (referred to as tRNA isodecoders)<sup>8</sup>. An interesting evolutionary trend is in higher eukaryotes, like mammals, that tend to have larger tRNA gene families, meaning, they carry more tRNA isodecoder genes in their genome, compared to unicellular organisms such as yeast. Moreover, the sequence differences between tRNA genes within tRNA isoacceptor family also resides in the internal pol III promotor (referred to as box A box B) that can lead to differential tRNA expression in human tissues and developmental stages. The increase of tRNA gene number and subsequent divergence along evolution is likely the result of genome expansion and may suggest additional role for eukaryotic tRNAs, in protein translation and beyond<sup>8</sup>.

# 2.3. <u>Changes in tRNA expression levels</u>

Translation efficiency is traditionally determinate by the tRNA adaptation index (tAI) that reflects the tRNA abundance according to their gene copy number, which means that the tAI has constant value for all cells from the same organism, as it is determinant by the gene copy number of the tRNA. Translation efficiency can also be measured from the supply to demand ratio, e.g. the tRNA availability (supply) and the codon usage (demand) in specific organism and condition. It was previously shown that the demand component varies between different organism, and even between conditions and cell types within the same organism, according to the tRNA copy number. Interestingly, the changes in the demand do not always positively correlate with the supply availability. For example, in yeast, stress- induced genes are enriched with codons that are translated by rare tRNAs (tRNAs with low copy number), in contrast to house-keeping genes or genes that are highly expressed in normal conditions. To that end, upon stress conditions, the rare tRNAs level must change to compensate for their low availability. This idea may represent a role of supply to demand ratio in shaping the expression pattern of individual genes in different cellular states and condition<sup>10</sup>.

Except of changes in the demand (codon usage), new observations have shown changes in the supply (tRNA availability) between organism, and even between cell types and conditions within the same organism<sup>8, 11, 12, 13</sup>. Currently, direct measurements of tRNA molecules to assess tRNA expression level is

not trivial, especially due to the presence of chemical modifications that are added to the tRNA molecule post-transcriptionally. The common methods for tRNA expression level determination is based on microarrays using complement probes, or genetic and epigenetic features like pol III occupancy of the tRNA genes or histone modification in their vicinity, that marks activation or repression of tRNA transcription<sup>8, 11</sup>. These technologies have contributed to the knowledge that tRNA expression level can change between various cell types that originate from the same species and the same cell type in different conditions. These observations highlight the question of how changes in the tRNA levels affect the cell. Studies that were carried out in yeast cells greatly contribute to the knowledge on the relation between tRNA and cell's fitness. One such study that was carried out in our lab performed a screen of single tRNA deletion library in yeast cells in various environmental conditions and combinatorial tRNA deletions library<sup>14</sup>. They revealed that tRNAs from the same family, with identical sequence, have different expression pattern and affect the cell's fitness differently. This can result from up-stream motifs that contribute to the tRNA transcription by RNA pol III, as well as essentiality of some tRNA in a variety of conditions, compared to less essential tRNA<sup>14</sup>.

# 2.4. <u>tRNA expression and cancer</u>

In recent years, there is growing evidence for the relation between tRNA expression and cancer. In one of these studies, that was performed on cancer cell lines and tumor samples from breast tissue, elevated tRNA levels were detected, relative to non-cancer cell lines/ tissue samples. Moreover, different tRNA families and isoacceptors showed a unique expression pattern in non- cancer samples and cancer samples, which was different between the two cell types (non- cancer vs cancer samples). They also found modest correlation between the codon usage of several codons of cancer- related genes to the elevated tRNA expression in cancer cell lines and samples, suggesting partial adjustment of the tRNA abundance to fit the codon usage of the cell<sup>12</sup>.

Several years ago, our lab has shown an intriguing connection between tRNA expression and the proliferation/ differentiation status of the cell. Briefly, proliferating and differentiated cells are expressing partially distinct sets of tRNA molecules, each compatible with the codon usage of either proproliferating genes or pro-differentiated genes. These findings suggested that different cellular states promote distinct translation regulation programs<sup>13</sup>. In this paper, we defined the two different sets of tRNA molecules as pro- proliferation tRNA pool, which are tRNAs with high expression level in proliferating cells and low expression level in differentiated cells, and pro-differentiation tRNA pool,

which are tRNAs with high expression level in the differentiated cells and low expression level in the proliferating cells.

Gingold et al. pointed out the correlation between tRNA expression and the cell state, but the nature of the relation between the two components wasn't discovered yet. A recent paper published by Goodarzi et al. showed, for the first time, an active role of tRNAs in the metastasis progression in human breast cell line. This paper revealed a causative relation between tRNA expression and cancer development. Moreover, over-expression of the metastasis- promoting tRNAs in non-metastatic breast cell lines resulted in high expression of genes enriched with the corresponding cognate codons. Genes with high content of the cognate codons also showed higher RNA stability relative to total RNA. These results were also true for global tRNA expression pattern and codon usage of highly expressed genes in metastatic breast cell lines and non-metastatic breast cell line- highly expressed genes (measured from ribosomal profiling and Mass-Spec analysis) had codon usage that corresponded to the tRNA types that showed elevated level in those samples<sup>15</sup>.

# 2.5. <u>Genome editing and CRISPR-Cas9 system</u>

These recent discoveries raise the possibility that tRNA expression can affect and perhaps even determine the physiology state of the cell. In my project, we aim to explore this possibility by manipulating the tRNA in human cells with various proliferation capacities. For tRNA perturbation, we performed genomic editing of specific tRNA genes using CRISPR-Cas9 technology. To our knowledge, this is the first time in which this technology is used for manipulating tRNA genes.

Genome editing is one of the recent major developments in the molecular biology field that opens the gate for innovative improvements, from genomic research to targeted therapy. One of the most common techniques for genomic editing is based in the CRISPR-Cas9 method<sup>16, 17</sup>. This system is composed of the Cas9 nuclease that is guided by short sgRNA molecule (~20 nt) that is complementary to the target genomic loci through Watson-Crick base pairing. Cas9 can be guided toward almost any target of interest in immediate vicinity of a protospacer adjacent motif (PAM) sequence by altering the 20-nt guide sequence within the sgRNA<sup>16, 18</sup>. Once the nuclease is bound to the targeted DNA, it creates DNA double- stranded breaks (DSBs), which stimulate the cellular DNA repair machinery. DNA repair can proceed through two alternative pathways; the error-prone nonhomologous end joining (NHEJ) or the high-fidelity homology- directed repair (HDR). Both options can be used to achieve the desired genomic editing. To create a knock-out gene, like we performed in this project, DSBs are re-ligated through the NHEJ process, in which nucleotides are inserted or removed randomly at the site of the brakeage, creating

insertion/deletions (Indel) mutation. In case of protein-coding genes even short Indels are likely to result in gene inactivation, provided that their length is not a multiple of 3 as they disrupt the reading frame. This may lead to knock-out of the gene<sup>16, 17</sup>.

There are several advantages in using CRISPR-Cas9. First, this system enables an easy targeting of any DNA sequence by a simply designing a ~20-nt guide sequence. Second, the cleavage by Cas9 nuclease is precise, creating blunt end cuts between the 17<sup>th</sup> and 18<sup>th</sup> base in the target sequence. Third, Cas9 can be used to target multiple genomic loci simultaneously, by co- delivering a combination of sgRNA to the cell of interest, or by single sgRNA provided that it has multiple targets in the genome<sup>17</sup>. Fourth, CRISPR-Cas9 creates a permanent knock-out of the target gene, which result in no expression of this gene, in comparison with RNAi, that only creates knock-down of the gene expression<sup>19</sup>. The limitations of the CRISPR-Cas9 system is the requirement of specific PAM sequence directly 3' to the target sequence, and the potential off- targets mutagenesis, that should be considered when designing the sgRNA sequence<sup>17</sup>.

In this project, we aim to knock-out tRNA genes which are non-coding gene. There are several challenges in knocking-out non-coding genes. First, small deletion or insertion of nucleotides generated by the CRISPR-Cas9 may not necessarily result in loss of function of the RNA molecule, because there is no open reading frame to disrupt<sup>20</sup>. On the other hand, tRNA is a highly structural molecule with several key positions along the sequence (like the anti-codon nucleotides and the A and B box- promoter sites which reside in the tRNA gene)<sup>9</sup>. Designing sgRNA complementary to those key positions will increase the chances of expressing non-functional tRNA molecule. An additional challenge is designing sgRNA for genes with high copy number in the target genome, like tRNA genes in human. Moreover, only parts of the tRNA genes that belong to the same isoacceptor have identical sequence. Third, evaluation of the tRNA knock-out level is not trivial because of the difficulty in measuring cellular tRNA level in the cell, as was mentioned above.

In this project, we aim to elucidate the causative relation between tRNA level and proliferative state. Despite the challenges, we succeeded to disrupt the tRNA expression pattern in three human cell line that differ in their proliferative status, HeLa, WI38 fast and WI38 slow cells. Perturbation of the tRNA expression was accomplished by delivery of sgRNA that target different tRNA isoacceptor to the cell lines that carry iCas9 gene with inducible expression. Cells with tRNA knock-out (KO) were analyzed for various phenotypic characteristics, such as fitness, proliferation rate and cell cycle progression. In general, we found that perturbation of tRNA levels in human cell line, whether they are part of the pro-

proliferation tRNA class or part of the pro-differentiation tRNA class, affect cell's fitness, proliferation rate and cell cycle progression. Moreover, knock-out of pro-proliferation tRNAs reduce the fitness of HeLa cells more severally than knock-out of pro-differentiation tRNAs. Additionally, HeLa cells were further negatively affected by the reduction of the pro-proliferation tRNAs level than WI38 slow cells. These two cell types represent the two extremes of the proliferation scale. Interestingly, WI38 fast show dual features, some similar to HeLa cells and some to WI38 slow, with a tendency to resemble the WI38 slow cells, in line with their immediate common ancestry. We believe this project will advance the notion that translation regulation has an active role in the proliferation- differentiation decision, including tumor development.

# **3.** Goals of the study

The goals of this study were to reveal the causative relation between tRNA expression pattern and proliferative state in human cells. To address this goal, we planned to:

- Perturbate cellular tRNA level by knock-out of tRNA genes. We planned to achieve this goal by target tRNA genes using CRISPR-Cas9 as genomic editing tool. Each human cell line was planned to carry an iCas9 gene (by stable transfection) and a single sgRNA segment that will target several (typically identical) loci of specific tRNA type.
- Explore the effect of tRNA deletions on the cell's viability and growth. We planned to do so by various phenotypic assays, like fitness measurement following competition between the different cell strains (each carrying sgRNA targeting a different tRNA) and proliferation rate measurement following induced knock-out of distinct tRNA and cell cycle progression.
- Create clonal population with Indel mutation introduced in the tRNA genes. The properties of the Indel mutation are characterized to assess the extent of damage for the tRNA functionality, for example to determine if features such as the anti-codon were abolished.

# 4. Methods

# 4.1. <u>Cell lines</u>

The following cell lines were used in this project:

*HeLa*- Human epithelial cell line originates from cervical cancer tissue. Their karyotype is characterized with aneuploidy (https://www.atcc.org/products/all/CCL-2.aspx). This cell line was kindly given to us by Prof. George Church.

*W138 immortalized cells*- These cells were created in Varda Rotter's lab, more than a decade ago<sup>3</sup>. Briefly, W138 human diploid fibroblast were infected with a recombinant retrovirus encoding for hTERT (marked with puromycin), a telomerase that enables cells to continue to proliferate beyond the replicative senescence checkpoint. Indeed, these cells underwent about 600 population doublings (PDLs). During the 600 PDLs, the cells have accumulated genetic alterations, some of them introduced synthetically and some occurred spontaneously<sup>3</sup>. These genetic alterations changed the cell's molecular and physiological characteristics, such as loss on contact inhibition, reduced doubling time, increased tumorigenicity and a global change in gene expression prolfiles<sup>3,4</sup>. Cells from several time points during the transformation process have been grown to generate cell- lines. In this work, we used the following cell lines: W138/hTERT<sup>slow</sup> (refer to as W138 slow), from the beginning of the transformation process, after ~150 PDLs. These cells are characterized with low proliferation rate, comparable to the primary cells. The second stage is W138/hTERT<sup>fast</sup> (refer to as W138 fast). These cells are characterized with acceleration in proliferation rate ranging between 30 and 52 PDLs/ 50 days, and they appeared after ~520 PDLs<sup>3</sup>.

*WI38 slow-* In addition to hTERT expression, these cells are characterized with high expression of p16<sup>INK4A</sup> and p14<sup>ARF</sup>, resulting from intact INK4A locus. These proteins are important regulators of pRb and p53 pathways, respectively<sup>3</sup>.

*WI38 fast* - In contrast to WI38 slow cells, these cells are characterized with low expression of p16<sup>INK4A</sup> and p14<sup>ARF</sup>, an outcome of methylated INK4A locus. These genetic alterations appeared spontaneously in the cells <sup>3</sup>. These cell lines were kindly given to us by Prof. Moshe Oren.

293T- human epithelial cell- line that originated from embryonic kidney (https://www.atcc.org/Products/All/CRL-1573.aspx). This cell line was used in our work for sgRNA- containing virus production.

# 4.2. <u>CRISPR plasmids</u>

The CRISPR plasmids were kindly sent to us by Prof. George Church's lab.

*pB-Cas9 & pB-support vector* – the pB-Cas9 plasmid contain TRE-rTta system (TRE= Tet response element; rTta= trans-activator for TRE). rTta is constitutively expressed in a non-functional state and in the presence of Doxycycline, it is transformed into its active form, when it binds to TRE-Cas9 and activates its expression. The pB-support vector is based on the PiggyBac transposon system which allows the pB-Cas9 to integrate into the genome. The pB-Cas9 plasmid has a Hygromycin resistance gene as selection marker. For WI38 cells, the same plasmid was used, but with GFP fused to Cas9 with short linker between them, so that upon expression, the linker is broken and the two genes are separated from each other.

*Lenti-sgRNA-puro/lenti-sgRNA-BDS* – this plasmid contains sgRNA sequence consisting of an 18-nt guide sequence and a "scaffold". The sgRNA sequence is under human U6 promotor, and it contains gag element for lentiviral packaging. For HeLa cells, this plasmid has Puromycin resistance gene as selection marker. For both WI38 cells, I had to replace this resistance marker by another Blasticidin, since they already carried the Puromycine resistance from the hTERT introduction.

# 4.3. <u>Lentiviral plasmids</u>

*psPAX2* (Addgene)- A lentiviral packaging plasmid, expressing gag. *pMD2.G* (Addgene)- A lentiviral packaging plasmid, expressing VSV-G envelope.

#### 4.4. <u>Medium</u>

DMEM + NEAA (Life Technologies)- Dulbecco's Modified Eagle Medium with 4.5mg/ml D-Glucose and Non-Essential amino acids. Supplemented with 10% FBS, 1% penicillin/ streptomycin (P/S), 1% L-Glutamine and 1% Sodium Pyruvate. This medium was used as base medium for HeLa cells. For CRISPR plasmids selection, 200µg/ml Hygromycin (iCas9 plasmid) and 2µg/ml Puromycin (sgRNA plasmid) were added to the medium and refreshed every two days. For iCas9 induction, 1µg/ml Doxycycline was added to the medium and refreshed every two days.

*MEM-EAGLE* + *NEAA* (Biological Industries) – Earle's salts base with Non-Essential amino acids. Supplemented with 10% FBS, 1% penicillin/ streptomycin (P/S) and 1% L-Glutamine. This medium was used as a base medium for WI38 slow and fast cells. For CRISPR plasmids selection,  $200\mu$ g/ml Hygromycin (iCas9 plasmid) and  $10\mu$ g/ml Blasticidin (sgRNA plasmid) were added to the medium and refreshed every two days. For iCas9 induction,  $1 \mu$ g/ml Doxycycline was added to the medium and refreshed every two days. *DMEM high glucose* (Biological Industries)- with 4.5mg/ml D-Glucose and 4mM L-Glutamine. Supplemented with 10% FBS, 1% penicillin/ streptomycin (P/S) and 1% L-Glutamine. This medium was used as a base medium for 293T cells.

# 4.5. <u>sgRNA cloning</u>

"Restriction- free" cloning was performed in order to create sgRNA for different tRNA type and controls, and for replacing the antibiotic resistance gene from Puromycin to Blasticidin in lenti-sgRNA plasmid. The "restriction free" cloning method is based on two sequential PCR reactions, the first one is to amplify the target gene by creating Megaprimers, and the second PCR is for introducing the target site into the plasmid. Following cloning, the original plasmid is being degraded using DpnI. For replacement of the antibiotic resistance, a two PCR reactions method was used. For the sgRNA replacement, long primers were ordered and used as megaprimers, therefor only the second PCR reaction was performed. Primers list is provided below (table 1). The first PCR reaction for antibiotics resistance replacement was conducted using iProof master mix (X2), 10µM of each primer and 20ng of lentiCas9- Blast plasmid (template plasmid) in a 50µl total volume reaction, for 32 cycles,  $Tm = 55^{\circ}C$ . The second PCR reaction (common to antibiotics resistance replacement and sgRNA replacement) was conducted using iProof master mix (X2), 1-1.5µl Megaprimers (PCR product from the first PCR- for antibiotic resistance replacement or long primers- for sgRNA replacement) and 50ng of lenti-sgRNA plasmid in a 50µl total volume reaction, for 12 cycles,  $Tm = 60^{\circ}C$ . It is important to mention that the elongation reaction in the second PCR was for 10 minutes. To eliminate the original plasmid, PCR products were incubated with 1µl DpnI enzyme (NEB) for 1 hour at 37°C, and then 20 min at 80°C for inactivation. Following the DpnI treatment, plasmids were transformed into DHa competent bacteria using standard heat shock transformation technique<sup>21</sup>. To find recombinant plasmid, colonies that grow under ampicillin selection were tested by sequencing of the purified plasmid using plasmid miniprep kit (Promega). Primers used for sequencing are:

BDS\_validation- AGGCACCTCGATTAGTTCTC

sgRNA\_validation-TTAGGCAGGGATATTCACCA

For massive plasmid purification, NucleoBond Xtra Midi kit (Macherey- Nagel) was used.

# Table 1- list of primers used for "restriction free" cloning

Sequence for gene amplification – creating megaprimers (first PCR)

Sequence complement to the target plasmid (second PCR)

	Forward primer	Reverse primer
BDS gene (Blasticidin resistance)	CCTCAGACAGTGGTTCAAAGTTTTT TTCTTCCATTTCAGGTGTCGTGACG TACGGCCACCATGGCCAAGCCTTTG TCTC	AATACCAGTCAATCTTTCACAAATT TTGTAATCCAGAGGTTGATTGTCGA CTTAACGCGTTTAGCCCTCCCACAC ATAAC
tRNA <sup>Ile</sup> AAT - sgRNA	CTTTATATATCTTGTGGAAAGGACG AAACACCGGTGGTGCTAATAACGC CAGTTTTAGAGCTAGAAATAGCAA GTTAAAATAAG	CTTATTTTAACTTGCTATTTCTAGCT CTAAAACTGGCGTTATTAGCACCAC CGGTGTTTCGTCCTTTCCACAAGAT ATATAAAG
tRNA <sup>Asn</sup> ATT - sgRNA	CTTTATATATCTTGTGGAAAGGACG AAACACCGGCAATCGGTCAGAGCG TTGTTTTAGAGCTAGAAATAGCAAG TTAAAATAAG	CTTATTTTAACTTGCTATTTCTAGCT CTAAAACAACGCTCTGACCGATTG CCGGTGTTTCGTCCTTTCCACAAGA TATATAAAG
tRNA <sup>Leu</sup> CAA - sgRNA	CTTTATATATCTTGTGGAAAGGACG AAACACCGGGATGGCCGAGTGGTC TAGTTTTAGAGCTAGAAATAGCAA GTTAAAATAAG	CTTATTTTAACTTGCTATTTCTAGCT CTAAAACTAGACCACTCGGCCATC CCGGTGTTTCGTCCTTTCCACAAGA TATATAAAG
tRNA <sup>Val</sup> TAC - sgRNA	CTTTATATATCTTGTGGAAAGGACG AAACACCGGTCTGCTTTACACGCAG AGTTTTAGAGCTAGAAATAGCAAG TTAAAATAAG	CTTATTTTAACTTGCTATTTCTAGCT CTAAAACTCTGCGTGTAAAGCAGA CCGGTGTTTCGTCCTTTCCACAAGA TATATAAAG
tRNA <sup>Leu</sup> TAG - sgRNA	CTTTATATATCTTGTGGAAAGGACG AAACACCGGTCTAAGGCGCTGGAT TTGTTTTAGAGCTAGAAATAGCAAG TTAAAATAAG	CTTATTTTAACTTGCTATTTCTAGCT CTAAAACAAATCCAGCGCCTTAGA CCGGTGTTTCGTCCTTTCCACAAGA TATATAAAG
tRNA <sup>lle</sup> TAT - sgRNA	CTTTATATATCTTGTGGAAAGGACG AAACACCGGTGAGTTCGAGCCTCA CCGTTTTAGAGCTAGAAATAGCAA GTTAAAATAAG	CTTATTTTAACTTGCTATTTCTAGCT CTAAAACGGTGAGGCTCGAACTCA CCGGTGTTTCGTCCTTTCCACAAGA TATATAAAG
tRNA <sup>Arg</sup> <sub>TCG</sub> - sgRNA	CTTTATATATCTTGTGGAAAGGACG AAACACCGGGATAAGGCGTCTGAC TTGTTTTAGAGCTAGAAATAGCAAG TTAAAATAAG	CTTATTTTAACTTGCTATTTCTAGCT CTAAAACAAGTCAGACGCCTTATC CCGGTGTTTCGTCCTTTCCACAAGA TATATAAAG
tRNA <sup>Arg</sup> <sub>TCT</sub> - sgRNA	CTTTATATATCTTGTGGAAAGGACG AAACACCGGCGCAATGGATAGCGC ATGTTTTAGAGCTAGAAATAGCAA GTTAAAATAAG	CTTATTTTAACTTGCTATTTCTAGCT CTAAAACATGCGCTATCCATTGCGC CGGTGTTTCGTCCTTTCCACAAGAT ATATAAAG
tRNA <sup>Glu</sup> TTC	CTTTATATATCTTGTGGAAAGGACG AAACACCGGATTCCTGGTTTTCACC CGTTTTAGAGCTAGAAATAGCAAG TTAAAATAAG	CTTATTTTAACTTGCTATTTCTAGCT CTAAAACGGGTGAAAACCAGGAAT CCGGTGTTTCGTCCTTTCCACAAGA TATATAAAG
tRNA <sup>Lys</sup> TTT - sgRNA	CTTTATATATCTTGTGGAAAGGACG AAACACCGGGGTTCAAGTCCCTGTT CGTTTTAGAGCTAGAAATAGCAAG	CTTATTTTAACTTGCTATTTCTAGCT CTAAAACGAACAGGGACTTGAACC CCGGTGTTTCGTCCTTTCCACAAGA

	TTAAAATAAG	TATATAAAG
	CTTTATATATCTTGTGGAAAGGACG	CTTATTTTAACTTGCTATTTCTAGCT
tRNA <sup>Arg</sup> <sub>ACG</sub> -	AAACACCGGATTCCAGGTTCGACTC	CTAAAACGGAGTCGAACCTGGAAT
sgRNA	CGTTTTAGAGCTAGAAATAGCAAG	CCGGTGTTTCGTCCTTTCCACAAGA
	TTAAAATAAG	TATATAAAG
	CTTTATATATCTTGTGGAAAGGACG	CTTATTTTAACTTGCTATTTCTAGCT
tRNA <sup>Pro</sup> AGG -	AAACACCGGGGGTATGATTCTCGCTT	CTAAAACTAAGCGAGAATCATACC
sgRNA	AGTTTTAGAGCTAGAAATAGCAAG	CCGGTGTTTCGTCCTTTCCACAAGA
C	TTAAAATAAG	TATATAAAG
	CTTTATATATCTTGTGGAAAGGACG	CTTATTTTAACTTGCTATTTCTAGCT
tRNA <sup>Leu</sup> CAG -	AAACACCGGTCTAAGGCGCTGCGT	CTAAAACGAACGCAGCGCCTTAGA
sgRNA	TCGTTTTAGAGCTAGAAATAGCAA	CCGGTGTTTCGTCCTTTCCACAAGA
C	GTTAAAATAAG	TATATAAAG
	CTTTATATATCTTGTGGAAAGGACG	CTTATTTTAACTTGCTATTTCTAGCT
tRNA <sup>Trp</sup> <sub>CCA</sub> -	AAACACCGGCGTGTTCAAATCACG	CTAAAACGACGTGATTTGAACACG
sgRNA	TCGTTTTAGAGCTAGAAATAGCAA	CCGGTGTTTCGTCCTTTCCACAAGA
C	GTTAAAATAAG	TATATAAAG
	CTTTATATATCTTGTGGAAAGGACG	CTTATTTTAACTTGCTATTTCTAGCT
tRNA <sup>Gly</sup> CCC -	AAACACCGGAACCCGGGTCGCAAG	CTAAAACTTCTTGCGACCCGGGTTC
sgRNA	AAGTTTTAGAGCTAGAAATAGCAA	CGGTGTTTCGTCCTTTCCACAAGAT
-	GTTAAAATAAG	ATATAAAG
	CTTTATATATCTTGTGGAAAGGACG	CTTATTTTAACTTGCTATTTCTAGCT
tRNA <sup>Arg</sup> <sub>CCG</sub> -	AAACACCGGGATAAGGCGTCTGAT	CTAAAACGAATCAGACGCCTTATC
sgRNA	TCGTTTTAGAGCTAGAAATAGCAA	CCGGTGTTTCGTCCTTTCCACAAGA
	GTTAAAATAAG	TATATAAAG
	CTTTATATATCTTGTGGAAAGGACG	CTTATTTTAACTTGCTATTTCTAGCT
tRNA <sup>Ser</sup> CGA -	AAACACCGGTTGGACTCGAAATCC	CTAAAACTTGGATTTCGAGTCCAAC
sgRNA	AAGTTTTAGAGCTAGAAATAGCAA	CGGTGTTTCGTCCTTTCCACAAGAT
	GTTAAAATAAG	ATATAAAG
	CTTTATATATCTTGTGGAAAGGACG	CTTATTTTAACTTGCTATTTCTAGCT
tRNA <sup>Pro</sup> CGG -	AAACACCGGGGGTATGATTCTCGCTT	CTAAAACGAAGCGAGAATCATACC
sgRNA	CGTTTTAGAGCTAGAAATAGCAAG	CCGGTGTTTCGTCCTTTCCACAAGA
	TTAAAATAAG	TATATAAAG
_	CTTTATATATCTTGTGGAAAGGACG	CTTATTTTAACTTGCTATTTCTAGCT
tRNA <sup>Thr</sup> CGT -	AAACACCGGACCGACGCCTTACCA	CTAAAACAGTGGTAAGGCGTCGGT
sgRNA	CTGTTTTAGAGCTAGAAATAGCAA	CCGGTGTTTCGTCCTTTCCACAAGA
	GTTAAAATAAG	TATATAAAG
~	CTTTATATATCTTGTGGAAAGGACG	CTTATTTTAACTTGCTATTTCTAGCT
tRNA <sup>Gly</sup> GCC -	AAACACCGGGGGCCGGGAATCGAAC	CTAAAACGGGTTCGATTCCCGGCCC
sgRNA	CCGTTTTAGAGCTAGAAATAGCAA	CGGTGTTTCGTCCTTTCCACAAGAT
	GTTAAAATAAG	ATATAAAG
	CTTTATATATCTTGTGGAAAGGACG	CTTATTTTAACTTGCTATTTCTAGCT
TP53_6a -	AAACACCGGCAGCATCTTATCCGA	CTAAAACTCCACTCGGATAAGATG
sgRNA	GTGGAGTTTTAGAGCTAGAAATAG	CTGCCGGTGTTTCGTCCTTTCCACA
	CAAGTTAAAATAAG	AGATATATAAAG
	CTTTATATATCTTGTGGAAAGGACG	CTTATTTTAACTTGCTATTTCTAGCT
RPS19a -	AAACACCGGTACCCCCAGCTTCCAC	CTAAAACCGCTGTGGAAGCTGGGG
sgRNA	AGCGGTTTTAGAGCTAGAAATAGC	GTACCGGTGTTTCGTCCTTTCCACA
	AAGTTAAAATAAG	AGATATATAAAG

# 4.6. <u>Cells transfections</u>

To generate stable iCas9 cell lines using transfection, WI38 slow and WI38 fast were seeded onto 10cm plates such that cell confluency will be approximately 70% the next day ( $\sim 2*10^6$  cells). 5µg of iCas9 vector (pB-Cas9 & pB-support vector) was transfected to cells with fresh MEM-EAGLE medium (5ml) and 15µl of Poly-jet transfection reagent (SignaGen). After four hours, the medium was replaced with a fresh MEM-EAGLE medium. Five days after the transfection, MEM medium containing 200 µg/ml Hygromycin was added to the transfected cells, refreshed every day for approximately one month.

# 4.7. <u>Cells transduction</u>

Cells transduction is composed of two stages. First stage is to generate a lentivirus carrying the plasmid of interest. The second stage is to transduce the lentivirus to the model cell.

To generate lentivirus containing sgRNA plasmid, 10 cm plates were covered with ~2ml Poly-D-Lysine (Sigma) and then 293T cells were seeded onto the covered 10 cm plates such that cell confluency will be approximately 70% the next day (~1\*10<sup>6</sup> cells). A day after, 2.5µg of PMD2.G (Addgene) and 10.3µg psPAX2 (addgene) packaging vectors were co-transfected with 7.7µg of the appropriate sgRNA plasmid using 40µl of jetPEI (Polyplus) in DMEM medium (5ml). After 48 and 72 hours, virus-containing medium was collected and centrifuged for 15 minutes at 3200g, 4°C. Supernatant was collected to new tube, and 1.25ml PEG solution from PEG virus precipitation kit (BioVision) was added. The virus containing tube was stored in 4°C for at least 12 hours (over-night). Virus- contained tubes were centrifuge for 30 min at 3200g, 4°C. Supernatant was removed and the virus pellet was suspended with 100µl virus resuspension solution from PEG virus precipitation kit (BioVision) and stored in aliquots (20 µl in each aliquot) at -80°C.

To generate stable sgRNA cell- lines using virus transduction, WI38 slow, WI38 fast and HeLa cells were seeded onto 10 cm plates such that cell confluency will be approximately 50% the next day (~ $1.5*10^6$  cells). On the day of transduction (24 hours after cell seeding), cell's medium was replaced with 5µg/ml Poly-Brene (Merck Millipore) contained medium (5ml). 20µl of virus suspension was added to each plate. After 48 hours, fresh medium containing selection antibiotics was added, and replaced every 2 days for approximately 7 days.

# 4.8. <u>sgRNA competition</u>

28 sgRNAs were designed by Idan Frumkin from our lab, to target 10 different pro-proliferation tRNAs, 10 different pro-differentiation tRNAs and 8 different controls (sgRNA targeting p53 gene, ribosomal

protein genes and Lin28 gene). Pooled sgRNAs (all 28 sgRNAs) were packed together in lentiviruses and introduced to HeLa iCas9 cells using lentivirus transduction. Thus, each virion carries one type of sgRNA, and each cell in the population is transduced by one virion. After transduction, cells were selected by adding  $200\mu$ g/ml Hygromycin (Cas9 plasmid) and  $2\mu$ g/ml Puromycin (sgRNA plasmid) to the medium for approximately one week (medium containing antibiotics were refreshed every two days). After selection, 1 µg/ml Doxycycline was added to the cells medium for iCas9 induction. Every two days, half of the cells (from 10 cm plate) were washed with PBS X1, and stored in -80°C, and the other half were reseeded in new 10 cm plate, with fresh medium containing antibiotics and Doxycycline. This procedure continued for 14 days. Eventually, cell samples from five time points (0, 4, 7, 10 and 14 days from the iCas9 induction) were sent for deep-sequencing.

## 4.9. <u>Library preparation and deep-sequencing</u>

Genomic DNA was purified from frozen cells samples of the sgRNA competition experiment using PureLink genomic DNA mini kit (Invitrogene) and used as templates for PCR to amplify specifically the sgRNAs in the population. PCR reaction was conducted using iProof master mix (X2), 10 $\mu$ M of each primer and 20ng of genomic DNA extracted from the samples in a 50 $\mu$ l total volume reaction, for 26 cycles, Tm = 64°C. The primers used to amplify the sgRNA region were:

Forward primer- GCTTACCGTAACTTGAAAGTATTTCGATTTCTTGG

#### Reverse primer- CTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAAC

After PCR clean-up (Promega), samples were run in 2% agarose gel to ensure that the PCR product is compose of a single amplicon in the appropriate size. Next, Hiseq libraries were prepared using the sequencing library module from Blecher-Gonen et al. paper <sup>22</sup>. In summary, blunt ends were repaired, Adenine bases were added to the 3' end of the fragments, barcode adapters containing a T overhang were ligated, and finally the adapted fragments were amplified. The process was repeated for each sample with a different Illumina DNA barcode for multiplexing, and then all samples were pooled in equal amounts and sequenced. We performed a 125bp paired end high output run on HiSeq 2500 PE Cluster Kit v4. Base calling was performed by RTA v. 1.18.64, and de-multiplexing was carried out with Casava v. 1.8.2, outputting results in FASTQ format.

### 4.10. <u>Deep sequencing- Data processing</u>

De-multiplexed data was received in the form of FASTQ files split into samples. First, SeqPrep (https://github.com/jstjohn/SeqPrep) was used to merge paired reads into a single contig, to increase

sequence fidelity over regions of dual coverage. The size of each contig was then compared to the amplicon's length. Next, the forward and reverse primers were found on each contig (allowing for 2 mismatches) and trimmed out. This step was performed for both the forward and reverse complement sequences of the contig, to account for non-directional ligation of the adaptors during library preparation. After the primers were trimmed, the contig was tested again for its length to ensure no Indels had occurred. Contigs were then compared sequentially to all sgRNA, comparing the sequence of each contig to the sequence of each sgRNA. Any contig without a matching sgRNA within two mismatches or less was discarded. Contigs with more than a single matching sgRNA with the same reliability were also discarded due to ambiguity. Each contig that passed these filters was counted in a keyvalue data structure, storing all sgRNA types and their frequency in each sample.

#### 4.11. <u>Machine learning techniques for fitness prediction</u>

For each tRNA that was targeted in the sgRNA competition, we download histone modifications and transcription factors binding scores of HeLa-S3 from the ChIP-seq analysis made by ENCODE/Broad Institute, release 3 (E.Consortium, 2013). For histone modification, signal view was used, which is the density graph of signal enrichment. At each base-pair position, the density is calculated as the number of sequenced tags overlapping a 25 bp window centered at that position. For transcription factor binding, peaks view was used, which is a regions of statistically significant signal enrichment. The score associated with each enriched interval is the mean signal value across the interval. In total, 161 different tRNA features, such as histone modifications, transcription factor binding, gene copy number of the tRNA genes, pol III and nucleosome occupancy, were examined in two machine learning methods, in order to find the tRNA feature that best predict the cell's fitness after tRNA KO, that was measured in the sgRNA competition experiment. The first method was linear regression using Lasso regularization<sup>23</sup>. These methods were performed using Matlab's 'lassoglm' function from the 'Statistics and Machine Learning' toolbox. The  $\lambda$  value was chosen as the value for which the deviance was minimal while keeping the model size below 10 features. The lasso method was run in 19- fold cross validation, meaning in each run one data point of the 20 measurements was left out of the training set and the performance of the model was tested on it. Using the lasso model, two features were chosen, resulting in a Pearson correlation of r = 0.74 between the predicted fitness and the observed fitness. The second method used was "Random Forest". The Random Forest<sup>24</sup> regression model was used using the publicly available R package<sup>25</sup>. Default parameters were used for building 1000 Forests containing 500 trees each. The averaged variable importance was calculated.

# 4.12. <u>Generating WI38 fast and slow single clones with tRNA<sup>lle</sup>AAT knock-out</u>

WI38 fast iCas9 and WI38 slow iCas9 were transduced with tRNA<sup>IIe</sup><sub>AAT</sub> - sgRNA lentivirus. After selection with antibiotic (see cell transduction), these cell lines were grown in 10 cm plate with medium containing antibiotics resistances (200 $\mu$ g/ml Hygromycin and 10 $\mu$ g/ml Blasticidin) and 1  $\mu$ g/ml Doxycycline for iCas9 induction for 7 days (medium containing antibiotics and Doxycycline were refreshed every two days). After 7 days, cells were trypsinized and resuspended with FACS buffer containing PBSX1, 1% FBS and 1mM EDTA pH = 8 (5ml). Suspended cells were sorted to single cells into 96 well plate containing filtered condition medium (medium from ~85% confluent Wild- Type cells plates, enriched with additional 10% FBS), using FACSAria Fusion cell sorter (Becton Dickinson- BD). GFP positive cells were selected using blue laser (excitation 488nm) and a 502nm longpass (LP) followed by a 530/30nm bandpass (BP) emission filters.

After sorting, the condition medium was refreshed every 3 days, until single clones appeared in the wells. Then, antibiotics were added to the wells containing clones, to ensure that only CRISPR-Cas9 clones grow in the wells. When the clone reached ~60 confluence, clones were split to two plates- one for freezing and the other for genomic DNA extraction. In order to find KO clones, genomic DNA was purified in 96 well plate using QuickExtract DNA extraction solution (Epicenter). Next, amplification of specific tRNA<sup>IIe</sup><sub>AAT</sub> genes was performed with PCR reaction using iProof master mix (X2), 10µM of each primer and 4µl of genomic DNA extracted from the samples in a 25µl total volume reaction, for 30 cycles. Primers and annealing temperatures are listed below (table 2). For identification of single amplicons in length of approximately 600 bp, samples were run in 1% gel agarose. Next, samples were cleaned-up using EXO-SAP. EXO-SAP clean-up is composed of two enzymes- Exonuclease I (BioLabs) and Shrimp Alkaline Phosphatase (BioLabs). In order to clean-up the samples, 10µl of the PCR product was incubated with an EXO-SAP mixture containing 0.1µl Exonuclease and its buffer with 0.1µl Phosphatase and its buffer. The incubation time is 30 minutes in 37°C and inactivation in 95°C for 5 minutes. After the PCR clean-up, samples were sequenced by Sanger sequencing.

tRNA locus	Forward primer	Reverse primer	Tm°C
Ile-AAT-5-1	ACAGAAGAGGAGCAACCTGG	TCAAGGACAGTGTTGGGGC	70.9
Ile-AAT-5-2	TCTGCGGTTGGTTCCGTG	ATGCTTGTCTAGCTATTGCGACC	68.6
Ile-AAT-5-3	GTTCCTTTCTGCTGGAGGTA	GGCATCAGACACACATCCAA	65.7
Ile-AAT-9-1	CTGGTGGACTCCCAGAGTTG	CTGAACAGCTTTCTGTCCCCA	69.1
Ile-AAT-8-1	CTGGGCGTACTGCCTTTGAA	CTGGTCTCCAATTCCCGACC	69.1

Table 2- primers to amplify specific loci of tRNA<sup>Ile</sup>AAT

Ile-AAT-2-1	GATGCAACTTCGGATTCGCC	CACAAAAGCCTCCAACCGTA	69.1
Ile-AAT-5-4	CTCACTGCGGAGCTAAGCCT	GCGTGGGATGTGCATCGGTT	71
Ile-AAT-5-5	CGCTGTTGGGTTTTTGCTCT	GTGGATTGGACGGACATCTT	69.1
Ile-AAT-4-1	GCCCTGTAGGTTTCCCAAGG	CCATCGCCTTAACCACTCGG	69.1

The first step in screening single clones with tRNA knock-out is by searching and characterizing Indel mutations in the tRNA genes. The fact that human cells are diploids, constitutes a great challenge, because each allele in each clone could have different type of Indel mutation. Thus, the screening method starts with a search for a reduction in sequencing quality, according to the chromatograms (ab1 Files) that is generated from the Sanger sequencing. Reduction in sequencing quality is when there is an unclear determination of the nucleotide composition in the vicinity of the iCas9 editing site. These clones (with unclear sequence determination near the iCas9 editing site) were further analyzed. The next step was to clarify the type of KO events for the two alleles, using TIDE: Tracking of Indels by DEcomposition<sup>26</sup>. For that, ab1 files of the clone and Wild-Type were uploaded to the software, then parameters were set-up:

'Alignment window'- the area in the sequence where the two sequences should be aligned. It should start from the first nucleotide to  $\sim$ 50 nucleotides prior to the start of the targeted sequence.

'Decomposition window'- the area in sequence where the two sequences should be decomposed. It should start from the beginning of the targeted sequence to the end of the sequence.

'Indel size range'- the number of nucleotide that should be removed or inserted. To detect large Indel mutation, this range was set-up to be  $\sim$ 30 nt.

The TIDE software decomposes the composite sequence (of the single clone) into its two alleles by means of multivariate non-negative linear modeling, with the control sequence (Wild-Type) serving as a template to model the individual Indel components. This decomposition results in an estimate of the relative abundance of every possible Indel within a chosen size range. The software provides the R<sup>2</sup> value as a goodness-of-fit measure, and calculates the statistical significance for each Indel<sup>26</sup>.

To identify the exact nucleotide sequence that was edited (deleted or inserted), the edited sequence was compared to the Wild-Type sequence.

## 4.13. <u>Proliferation assay</u>

To quantify viable cells, MTT based- proliferation assay was used. This assay is based on mitochondrial activity, and involves conversion of MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) to an insoluble formazan by cleavage of the tetrazolium ring by dehydrogenase enzymes. The formazan is

then solubilized and the concentration is determined by absorbance. iCas9 was induced in all three cell lines (HeLa, WI38 fast and WI38 slow) with eight sgRNAs targeting different tRNA genes, and controls (sgRNA targeting p53 gene and ribosomal protein gene, cells carrying iCas9 only) by adding 1µg/ml Doxycycline to antibiotics- containing medium. Every 2 days, medium containing Doxycycline and antibiotics was refreshed through all experiment duration. Induced HeLa and WI38 (fast and slow) cells were seeded in 96 well plates, in three repeats, after 3 and 2 days from the start of the induction, respectively. O.D measurement of HeLa and WI38 (fast and slow) cells started 2 and 1 day after the cells were seeded in 96 well plate, respectively. O.D was measured every day for 3 days. On the day of O.D determination, cell's medium was replaced with fresh medium and 10µl of 12mM MTT solution from Vybrant MTT cell proliferation assay kit (Thermo fisher) was added for a 4 hour incubation in 37°C. Subsequently, 100µl of SDS-HCl solution from Vybrant MTT cell proliferation assay kit (Thermo fisher) was determinate by absorbance at 570 nm. Results were normalized to blank values (wells containing only medium), and O.D ratio was calculated (O.D in the last day divided by O.D in the first day).

# 4.14. <u>Cell cycle analysis</u>

All three cell lines (HeLa, WI38 fast and WI38 slow), that are expressing iCas9 with single sgRNAs were grown in a 10 cm plate with a medium containing antibiotics and Doxycycline induction. HeLa cells were grown in these conditions for 11 days and WI38 fast and slow for 7 days. Following the doxycycline induction, cells were incubated with 5ml medium containing10 $\mu$ M BrdU (Sigma), for one hour. Then, cells were trypsinized, washed with PBS X1 and fixated with 5ml 70% ethanol (in PBS). The fixed cells were stored at 4°C. To detect BrdU incorporation, we use BrdU antibody conjugated to APC (eBioscience). Before adding anti-BrdU-APC to the fixed cells, cells were denaturalized for 30 min incubation in 1ml denaturation solution containing 2N HCl, 0.5 Triton and PBS X1. Next, cells were neutralized by suspension in 1ml neutralization solution containing 0.1M Na<sub>2</sub>B<sub>4</sub>O<sub>7</sub> pH = 8.5 dissolved in DDW (titration with concentrated HCl). After these steps, cells were incubated for 1 hour with 100 $\mu$ l antibody solution containing 1% BSA, 0.5% Triton, 7 $\mu$ l anti-BrdU-APC in PBS X1. After incubation, cells are washed with 1% BSA in PBS and suspend in 500 $\mu$ l PBS containing 50 $\mu$ g/ml Propidium Iodide (Sigma) and 50 $\mu$ g/ml RNAse. Cells were filtered with cell strainer 0.45  $\mu$ m (SPL). Finally, cells were analyzed in LSR II Flow Cytometer (Becton Dickinson- BD). Propidium Iodide was measured using yellow/green laser (excitation 561nm) and a 600nm longpass (LP) followed by a 610/20nm bandpass

(BP) emission filters. Anti-BrdU-APC was measured using red laser (excitation 633nm) and a 670/30nm bandpass (BP) emission filter. Results analysis was performed using the FlowJo software.

# 5. Results

We wish to elucidate how tRNA level perturbation affects human cell's fitness, viability and growth. In a paper published two years ago by Gingold et al., we presented the dichotomy between proliferation and differentiation state of the cells in multicellular organism, and discovered distinct translation regulation programs to the two cellular states, which were based on different expression pattern for tRNA genes, complementary to the codon usage of highly expressed genes for each cellular state<sup>13</sup>. In this current project, we aim to reveal whether the correlation between the tRNA expression pattern and the cellular state is also causative- which element determines the other? In particular, we intend to examine whether manipulation of tRNAs affect cellular growth. Thus, we performed a series of experiments, evaluating the fitness and the proliferative phenotype of cells after the knock-out of different tRNA genes.

## 5.1. <u>Comparing the fitness of tRNA- deleted cells using competition based assay</u>

The first aim in this project was to evaluate the consequence of reducing the levels of different tRNAs on the cell's fitness. To achieve this, we examined the fitness of an heterogenous cell population, in which each cell carries a different sgRNA targeting a different tRNA. The tRNAs belong to two distinct groups-pro-proliferation tRNAs and pro-differentiation tRNAs. Briefly, iCas9 was induced in HeLa iCas9 pooled- sgRNA cells by incubation with doxycycline for 14 days. Every two or three days, sample was taken from the cells plate. From this sample, genomic DNA was extracted and the sgRNA was sequenced using HI-SEQ machine (see Methods).

The deep-sequencing results revealed that 7 out of 10 pro-proliferation tRNA KO in HeLa cells presented a reduction in the read count of their complement sgRNA between one order of magnitude, and up to 5 orders of magnitude (figure 1). In comparison, 5 out of 10 pro- differentiation tRNA KO in HeLa cells displayed a reduction in the read count of their complement sgRNA. Furthermore, the extent of reduction upon knock-out of pro-differentiation tRNAs was much modest, ranging between 1 to 1.5 orders of magnitude (figure 1). 3 out of 10 pro-proliferation tRNA KO demonstrated a slight increase in the read count through the 14 days of iCas9 induction (figure 1), while constant read count is apparent in 5 out of 10 pro-differentiation tRNA KO (figure 1). These results indicate that upon competition, HeLa cells carrying sgRNAs that target pro- proliferation tRNA genes are diluted in heterogenous population that are composed of both sgRNAs targeting pro- proliferation and pro- differentiation tRNAs. In contrast,



population cells carrying sgRNA targeting pro- differentiation tRNA genes are more stable through time of competition.

#### Figure 1 - Fraction of each tRNA KO type in HeLa cells population during 14 days of iCas9 induction.

Pooled sgRNA HeLa- iCas9 cells were doxycycline induced and grown for 14 days. Cell samples from 5 time points during the 14 days of induction were taken for sgRNA deep- sequencing. The fraction of each tRNA KO type in the population was calculated by normalizing the reads number of each sgRNA to the total reads number of the sample (log scale). Each row represents KO of tRNAs (displayed by their anti- codon), in which every two adjacent rows represent two biological repeats of the same tRNA deletion. The first 20 rows represent the pro-proliferation tRNAs and the last 20 rows represent the pro-differentiation tRNAs. Each column represents a time point during the competition assay.

The CRISPR-Cas9 system that was used in this project is highly complex, since each tRNA type has multiple copies in the genome, some with high sequence similarity and even identical, and other with lower sequence similarity. For each tRNA only one sgRNA was constructed, such that the sgRNA will optimally target as many tRNA loci as possible from the same type, with minimum off-targets, especially other tRNA types. Therefore, a notable question that should be tackled is whether the number of predicted loci each sgRNA is supposed to target correlates with the effect of the tRNA KO on the cell's viability. In other words, the concern was that perhaps the pro-proliferation tRNA KOs tend to show a higher effect since the sgRNAs that target them happen to target a larger fraction from their respective tRNA families compared to the pro-differentiation tRNAs. We also calculated and examined the cell's fitness (instead of fraction in the population), to correct read count differences in the first day of sampling (before iCas9

induction started). Indeed, there is a negative correlation between the average fitness of each tRNA KO (calculated from two biological repeats) and the predicted targeting percentage of the corresponding sgRNA (figure 2), suggesting that part of the differences in fitness between different tRNA KO can result from an experimental artifact- an incomplete targeting. Yet, although this confounding factor does affect the results, the interesting observation is that even if it is considered, the pro-proliferation tRNA KOs are still exerting a more sever effect. In particular, the pro-proliferation tRNA KO tend to reduce fitness more than pro-differentiation tRNA KO, even at the same level of targeting percentage, evidenced by the pro-proliferation tRNA (red anticodon) tending to be localized below the trend line, while the pro-differentiation (blue anticodons) above it (figure 2).

tRNA<sup>Glu</sup><sub>TTC</sub>, tRNA<sup>Leu</sup><sub>CAA</sub>, tRNA<sup>Leu</sup><sub>TAG</sub>, and tRNA<sup>Arg</sup><sub>TCG</sub> are the pro-proliferation tRNAs that exhibited the greatest reduction in the population's fraction (figure 1) and the lowest fitness (figure 2). When examining the targeting percentage of the sgRNA targeting these tRNAs, sgRNA targeting tRNA<sup>Arg</sup><sub>TCG</sub> and tRNA<sup>Leu</sup><sub>TAG</sub> are predicted to target 100% of their tRNA loci. Surprisingly, sgRNA targeting tRNA<sup>Leu</sup><sub>CAA</sub> and tRNA<sup>Glu</sup><sub>TTC</sub> are predicted to target less than 70% of their tRNA loci, and yet they exert a very strong effect on the cell's viability. This phenomenon also exists in pro-differentiation tRNA KO. tRNA<sup>Pro</sup><sub>CGG</sub>, tRNA<sup>Gly</sup><sub>GCC</sub>, tRNA<sup>Pro</sup><sub>AGG</sub>, tRNA<sup>Leu</sup><sub>CAG</sub> and tRNA<sup>Ser</sup><sub>CGA</sub> are pro-differentiation tRNA that displayed the greatest reduction in the population's fraction (figure 1) and the lowest fitness from the pro-differentiation tRNA group (figure 2). The sgRNA targeting tRNA<sup>Gly</sup><sub>GCC</sub> has only 35% targeting percentage, while the rest of the sgRNAs with low fitness are predicted to target 100% of their tRNA loci. Thus, although a correlation (r = -0.49) exists between percent targeted loci and fitness, these results suggest that there are additional molecular features, potentially intrinsic to the nature of certain tRNAs, that affect tRNA KO fitness. One such feature appears to be related to the distinction between pro-proliferation and pro-differentiation tRNAs, but others are probably yet to be discovered.





For each tRNA type, average fitness of two biological repeats was measured by dividing the read count of the 14<sup>th</sup> day of iCas9 induction by the read count of the first day- before iCas9 induction (log scale). The average fitness of each tRNA KO type is presented against the targeting percentage of each sgRNA that is calculated by dividing the number of targeted loci for each tRNA type by the total number of loci of each tRNA type. Pro-proliferating tRNAs marked with red anti-codons and pro-differentiation tRNAs marked with blue anti-codons.

In order to find such predictive features, we ran two machine learning methods (linear regression and Random Forest) that examined different tRNA features (genomic and epigenomic features) and chose the features that predicted the tRNA KO cell's fitness best. The tRNA features we examined included different histone modifications, transcription factors binding in the vicinity of the tRNA genes, RNA pol III and nucleosome occupancy, gene copy number of each tRNA type, etc. The two features that best predicted the fitness from the linear regression method were binding of the transcription factor Ezh2 to the tRNA genes and sensitivity to MNase treatment, an indication for nucleosome occupancy (with estimated coefficient of -2.38 and -1.41, p-value of 0.004 and 0.07 and variance explained of 0.35 and 0.12, respectively). The four features that best predicted the fitness from the Random Forest method were binding of Ezh2, GABP, CTCF to tRNA genes and sensitivity to MNase treatment, with importance of 57.5, 25.3, 15.1 and 14.6, respectively. From these two methods, we can conclude that Ezh2 binding to the tRNA genes is the feature that best predicted the cell's fitness (as a result of tRNA KO), and

nucleosomes occupancy (MNase treatment) also greatly influence the fitness of tRNA KO cell. When exploring the relation between the two features, we observed a positive correlation between them, suggesting a similar effect of each feature on the cell's fitness (figure 3).

Interestingly, both features that were found to best explain the tRNA KO fitness, are preserving the inactive state of genes transcription. Ezh2 is a member of the polycomb- group family, and is involved in maintaining the transcriptional repressive state of genes<sup>27</sup>. MNase-seq is a sequencing technique in which genomic DNA with its endogenous histone structure is exposed to MNase treatment, that digests naked DNA (not wrapped around nucleosomes). After MNase treatment, nucleosomes- bound DNA is detached from the nucleosomes and sequenced. DNA occupied by nucleosomes are regions that tend to be transcriptionally repressive<sup>28</sup>. Surprisingly, the pro- proliferation tRNAs (red anti- codons) tend to have higher scores for both features than the pro- differentiation tRNAs (blue anti- codons) (figure 3). This observation, together with the previous studies mentioned, is contradicting to the notion that pro- proliferation tRNAs are more highly expressed in HeLa cells, therefore representing actively transcription state.



Figure 3- Ezh2 binding vs. MNase sensitivity in the vicinity of tRNA genes for each tRNA KO

For each deleted tRNA type, average Ezh2 binding to the tRNA genes and average MNase-seq of the tRNA genes (belong to the same tRNA isoacceptor) were measured. ChIP-seq and MNase- seq data were downloaded from the ENCODE project (see Methods). Pro-proliferating tRNAs marked with red anti-codons and pro-differentiation tRNAs marked with blue anti-codons.

# 5.2. <u>Creating WI38 fast and slow single clones with characterized tRNA deletion</u>

While the above results point on a function of pro-proliferation tRNA in cells growth, the main downside of this experiment is the fact that it was done at the population level, which is likely to contain cells with different genotype in each tRNA locus. Since in our case, repairing DSB generated by CRISPR-Cas9 is based on Non- homologous end joining DNA repair, different tRNA loci targeted by the same sgRNA are likely to contain different Indel mutation. This can result in distinct phenotype of each cell in the heterogenous population, which may create biased results in favor of cells with functional tRNAs. We therefore decided to further examine the effect of pro-proliferation tRNAs KO on cell fitness using a cleaner and better controlled system of single clones carrying the same Indel mutations in genes related to the same tRNA type, that will lead to a knock-out of this tRNA.

Few obstacles exist when attempting to overcome this kind of challenge. As previously mentioned, each tRNA type has several loci in the human genome, and not all sgRNA are predicted to target 100% of the tRNA loci. Second, the induction time needed for maximizing the CRISPR system when targeting multiple gene copies, with minimal off-target effects was not clear. We chose to create a single clone with tRNA<sup>Ile</sup><sub>AAT</sub> KO, on the background of WI38 slow and fast cells. We chose those cells (and not HeLa cells) due to their intact karyotype. We chose to knock- out tRNA<sup>Ile</sup><sub>AAT</sub>, which is part of the proproliferation tRNA pool, following preliminary results of the sgRNA competition that revealed strong reduction in the fitness of HeLa cells lacking expression of this tRNA, and the fact that the sgRNA targeting this tRNA type is predicted to target almost 100% of its loci. Notably, additional sgRNA competition experiments that were carried out at later stages, did not restore the fitness reduction phenotype upon tRNA<sup>Ile</sup><sub>AAT</sub> KO (figure 2).

To create KO cells, we induced the iCas9 for 6 days using Doxycycline and then removed the doxy for 7 days (until the GFP level reduced to basal level), to ensure that no active Cas9 were left after sorting the cells. Then, the cells were sorted to single cells in 96 well plate using FACS sorter. Clones that appeared in the wells (roughly after 1 week of growth) were screened, using PCR on most of the tRNA<sup>Ile</sup><sub>AAT</sub> loci, for identification of tRNA KO clones. Overall, 31 clones were screened, out of which 7 clones had several edited tRNA<sup>Ile</sup><sub>AAT</sub> loci (table 3).

Figure 3 present an example for detection protocol of tRNA gene editing. The first step is to sequence the tRNA gene (specific locus) that was amplified from genomic DNA extracted from the single clone population (for primer list, see table 2). If the gene wasn't edited, the chromatogram (based on the Sanger sequencing) will demonstrate clear and determined sequence of the gene, identical to the Wild- Type sequence. In case of editing, the chromatogram read will become disturbed in the editing area, with no

nucleotide determination, to the end of the amplicon (figure 4A). This is due to heterogeneous editing of the two alleles, that belong to the same tRNA gene, and results in shifts of the sequence by a different number of nucleotides due to insertion and deletions. To characterize the editing pattern in each allele, first the length of the Indel mutations for the two alleles was predicted using the TIDE tool<sup>26</sup> (figure 4B, see Methods). We used this protocol in order to identify tRNA KO in both WI38 fast and slow single clones. WI38 fast iCas9 tRNA<sup>Ile</sup><sub>AAT</sub> KO clone #11, locus Ile-AAT-5-1, was the first clone we succeeded to create and was used as a background for the rest of the WI38 fast clones we further created. Clone #11 was edited in a heterogeneous manner, in which one allele has deletion of 20 nucleotides and the second allele has deletion of 18 nucleotides (table 3, figure 4B). The alleles' sequence, with the identity of the deleted nucleotide is shown in figure 4C.







С

Ile-AAT-5-1-allele #1, 20nc deletion: GCCGGTTAGCTCAGTTGGTTAGAGCGTGGTGCTAATAACGCC AAGGTCGCGGGTTCGATCCCCGTACGGGCCA

Ile-AAT-5-1-allele #2, 18nc deletion: GCCGGTTAGCTCAGTTGGTTAGAGCGTGGTGCTAATAACGCC AAGGTCGCGGGTTCGATCCCCGTACGGGCCA

# tRNA seq sgRNA seq iCas9 editing site deleted nc

#### Figure 4 - Indel mutation detection and characterization protocol of WI38 fast single clone with edited tRNA gene

WI38 fast iCas9 cells with targeted tRNA<sup>lle</sup><sub>AAT</sub> genes were isolated to single cells using FACS sorter. tRNA<sup>lle</sup><sub>AAT</sub> genes were amplified from the genomic DNA of single clones, and sequenced by Sanger sequencing. A| Chromatogram of the Ile-AAT-5-1 locus, illustrates the sequencing quality reduction adjacent to the iCas9 editing site. The low sequencing quality continues to the end of the amplicon. B| Relative abundance of each possible Indel, calculated by TIDE software. C| Ile-AAT-5-1 locus sequence (black), sgRNA sequence (orange), iCas9 editing site (blue) and the deleted nucleotide (enlarged letters) are displayed for each allele.

Other than the WI38 fast clone #11, we succeeded to create additional clones with various deletions and insertions in different loci of tRNA<sup>IIe</sup><sub>AAT</sub> in WI38 slow and WI38 fast that already contain deletion in both alleles in locus IIe-AAT-5-1. These clones are listed below (table 3). It is important to note that while

WI38 slow clones didn't show any change in the proliferation rate (when examined under the microscope), WI38 fast clone #3 and #4 did show small decrease in the doubling time (not shown). WI38 fast clone #16, which had the highest number of edited loci, didn't survived. This may indicate that although KO of this tRNA in the population level didn't show decrease in fitness of HeLa cells, single clones carry only partial KO, like WI38 fast clone #3 and #4, are affected and show reduction in their proliferation rate. Moreover, probably full KO of this tRNA is not viable, as WI38 fast clone #16 with four, partially targeted, loci didn't survive.

Cell line	Clone number	tRNA name	Indel mutation	Anticodon editing
	25	Ile-AAT-5-1	Allele #1 - 7 nc deletion	-
WI38 slow			Allele #2 - no editing	-
W150 510W		Ile-AAT-5-4	Allele #1 - 7 nc deletion	-
			Allele #2 - no editing	-
WI38 slow	1	Ile-AAT-8-1	Allele #1 - 18 nc deletion	+
W 150 SIOW	1		Allele #2 - no editing	-
	27	Ile-AAT-2-1	Allele #1 - 2 nc deletion	-
WI38 slow			Allele #2 - no editing	-
		Ile-AAT-5-4	Allele #1 - 19 nc deletion	-
			Allele #2 - no editing	-
W/120 C	11	Ile-AAT-5-1	Allele #1 - 18 nc deletion	+
W150 last			Allele #2 - 20 nc deletion	+
WI38 fast (clone 11 background)	3	Ile-AAT-5-5	Allele #1 - 9 nc deletion	
WI38 fast (clone 11 background)			Allele #2 - no editing	-
WI38 fast (clone 11 background)	4	Ile-AAT-4-1	Allele #1 - 1 nc insertion	-
w158 fast (clone 11 background)			Allele #2 - no editing	-
	16	Ile-AAT-8-1	Allele #1 - 2 nc deletion	
WI38 fast (clone 11 background)			Allele #2 - no editing	-
		Ile-AAT-2-1	Allele #1 - 14 nc deletion	
			Allele #2 - no editing	-
		Ile-AAT-5-5	Allele #1 - 2 nc deletion	

Table 3- Single clones of WI38 fast and slow iCas9 with edited loci of tRNA<sup>Ile</sup>AAT.

		Allele #2 - no editing	-
	Ile-AAT-4-1	Allele #1 - 2 nc deletion	
		Allele #2 - no editing	-

5.3. <u>Measuring proliferation rate of tRNA- deleted cells using MTT based proliferation assay</u> The sgRNA competition results suggested that targeting pro-proliferation tRNAs in proliferating cells disrupts cells' fitness, while targeting pro-differentiation tRNAs is less deleterious to the cells. We therefore decided to check how the targeting of various tRNAs affects cells proliferation rate, and whether targeting the same tRNA in different cell types (vary in their doubling time) will have different effect.

To assess the effect of tRNA KO on cell's proliferation, we compared between four pro-proliferation tRNA KOs and four pro-differentiation tRNA KOs in three cell lines (HeLa, WI38 fast and WI38 slow). As controls, we used KO of p53, which is expected to result in increased proliferation rate, and a KO of a ribosomal protein, which when targeted is expected to reduce cell proliferation. In addition, to assess whether doxycycline had any effect on the assay, we analyzed each cell line (with iCas9) in the absence or presence of doxycycline.

The proliferation assay results revealed two general observations. First, all tRNA KOs and controls (except ribosomal protein KO in HeLa cells) didn't cause cell death, but rather deceleration or recess in cell's proliferation, as the O.D ratio between the last day and the first day of measurement is higher than 1 (figure 5A-C). Second, in general, pro-differentiation tRNA KO in WI38 fast and slow cells reduces cell's proliferation more severely than in HeLa cells (figure 5A-C).

When focusing on each tRNA KO separately, interesting behaviors are revealed. In the pro-differentiation tRNA group, two out of four tRNAs exhibited similar effect pattern in all three cell lines. tRNA<sup>Pro</sup><sub>CGG</sub> strongly reduced the proliferation rate of HeLa, WI38 fast and WI38 slow, while tRNA<sup>Ser</sup><sub>CGA</sub> had a minor effect on the proliferation rate of these cell lines (figure 5A-C). tRNA<sup>Leu</sup><sub>CAG</sub> showed a moderate effect on the proliferation rate of HeLa cells (figure 5A), but it became significant in WI38 fast and slow cells (figure 5B-C). tRNA<sup>Pro</sup><sub>AGG</sub> showed the opposite pattern from the last mentioned, in which KO of this tRNA in HeLa and WI38 fast dramatically reduced their proliferation rate (figure 5A-B), while in WI38 slow cells this effect declined (figure 5C). In the pro-proliferation tRNA group, tRNA<sup>Arg</sup><sub>TCG</sub> presented medium to weak reduction of the proliferation rate in all three cell lines (figure 5A-C). tRNA<sup>Arg</sup><sub>TCT</sub> and tRNA<sup>Leu</sup><sub>TAG</sub> revealed an interesting pattern, in which they had a great impact on proliferation rate reduced significantly

(figure 5B-C). tRNA<sup>lle</sup><sub>AAT</sub> showed the opposite effect from the last two tRNAs, in which it has minor effect on the proliferation rate of HeLa cells (figure 5A), and in WI38 fast and slow cells this effect become negatively stronger (figure 5B-C). These results indicate that knock- out of different tRNAs from each group (pro- proliferation and pro- differentiation), affect the proliferation rate of all three cell lines in various manners, and not necessarily according to the proliferation- differentiation separation suggested earlier in Gingold et al.<sup>13</sup>.

When comparing the proliferation assay in HeLa and sgRNA competition, also performed in HeLa cell line, we can see that the results correlate to each other. In each tRNA pool (pro-proliferation tRNA pool and pro-differentiation tRNA pool), the tRNA KOs that showed the lowest fitness (figure 2) also reduced the proliferation rate of HeLa cells (figure 5A). However, when examining both tRNA pools, the proliferation assay results don't show clear separation between pro-proliferation and pro-differentiation tRNA, as was found in the sgRNA competition (figure 5A).

An additional interesting observation is that pro- differentiation tRNA KOs negatively affected WI38 cells (fast and slow) more than HeLa cells (figure 5A-C). The shift in the proliferation rate effect from HeLa cells to WI38 cells was also shown in pro-proliferation tRNA KOs, in which two out of four tRNA KOs strongly reduced the proliferation rate in HeLa, but they had a minor effect in this parameter in either WI38 fast and slow cells (figure 5A-C). Interestingly, tRNA<sup>IIe</sup><sub>AAT</sub> (part of the pro-proliferation tRNA pool) had very weak effect on the proliferation rate of HeLa cells, while it became stronger on both types of WI38 cells (figure 5A-C). These results may reflect the nature of the three cell lines, in which HeLa cells are significantly more proliferating and have more cancerous features than WI38 fast and slow cells.

With respect to the control samples, we can observe that the p53 KO and the ribosomal protein (RP) KO showed the expected results in HeLa cells, when the p53 KO didn't affect the proliferation rate (similar to the HeLa iCas9 cells), and the RP KO showed the greatest reduction in proliferation rate, more than the effect of all tRNA KOs (figure 5A). This observation is much weaker in the WI38 fast cells, when the RP KO don't show the most significant reduction in proliferation rate (figure 5B) and in WI38 slow cells, when both p53 KO and RP KO are not showing the expected behavior, i.e. presenting the two ends of the proliferation rate scale (figure 5C). The results of the control samples may point on the complexity in the integrity and verity of the assay in the WI38 fast and slow samples.







Figure 5 - O.D ratio representing cell proliferation for various tRNA KO in three cell lines

Consequence of tRNA deletion on the proliferation rate of A| HeLa cells, B| WI38 fast cells and C| WI38 slow cells. Proliferation rate is illustrated as ratio between the O.D measured in the last day of the experiment to the O.D measured in the first day of the experiment. iCas9 was induced before the first day of O.D measurement, as mentioned in the methods. O.D was determined using MTT- proliferation assay kit. The samples (of all three cell lines) are sorted according to their proliferation rate resulted from the tRNA KOs. Colors marks the type of gene targeted by the sgRNA- controls (green), proproliferation tRNA (red) and pro-differentiation tRNA (blue).

# 5.4. <u>Exploring cell cycle progression and S phase duration of tRNA- deleted cells using FACS</u> analysis

The effect of knocking- out different tRNAs on proliferation rate and fitness, motivated us to explore whether these changes are manifested by variations in the cell cycle progression following tRNA KO. Perhaps knock- out of pro-proliferation tRNAs can result in deceleration of the cell cycle in fast growing cells, or even cell cycle arrest, while in slow growing cells this phenotype will be less significant. We therefore performed detailed analysis of cell cycle progression for selected set of tRNA KO cells (similar to the tRNAs explored in the proliferation assay).

After the iCas9 induction, all three cell lines (HeLa, WI38 fast and WI38 slow) were labeled with BrdU and Propidium Iodide (PI) and analyzed using flow cytometry. BrdU is a thymine analog that

incorporates into newly synthesized DNA, thus marking active DNA synthesis. As such, it can serve as a measure for the combined speed of all replication forks operating in a cell. Propidium iodide binds to DNA molecules and is used to quantitively assess DNA content and hence determine the stage of cell cycle that a cell is at. Using BrdU and PI staining allows to measure very accurately cells in various stages of the cell cycle. Notably, it allows clear separation of cells in G1 from early S phase, or late S phase from G2-M. Generally, only cells in S phase are supposed to have positive values for BrdU incorporation, since only in this phase there is active DNA synthesis. The separation between G0-G1 and G2-M population is by the PI values, when G2-M cells have two times more DNA then G0-G1 cells.

Following FACS measurement the row data was analyzed using FlowJo software. An example of such analysis is illustrated in figure 6, in which WI38 fast iCas9 cells targeted in pro-proliferation tRNA (tRNA<sup>Arg</sup><sub>TCG</sub>) and targeted in pro-differentiation tRNA (tRNA<sup>Ser</sup><sub>CGA</sub>) is analyzed, compared to control (the same cell that is not targeted). When knocking-out tRNAs in WI38 fast cells, S phase population is reduced, and G0-G1 population is increased, compared with induced WI38 fast iCas9 (figure 6). Interestingly, KO of tRNA<sup>Arg</sup><sub>TCG</sub> (pro- proliferation tRNA) reduced the S phase population in WI38 fast cells 10% more than tRNA<sup>Ser</sup><sub>CGA</sub> (pro-differentiation tRNA) KO (figure 6). Additionally, both G0-G1 and G2-M population increased in tRNA<sup>Arg</sup><sub>TCG</sub> KO compared to tRNA<sup>Ser</sup><sub>CGA</sub> KO (figure 6).





WI38 fast iCas9 cells with KO tRNA genes were labeled with BrdU and Propidium Iodide for cell cycle analysis. Using FlowJo software, % cells in each cell cycle phase were determined according to the labeling values. (left panel) % cells in each cell cycle phases for control sample- WI38 fast iCas9 with iCas9 induction. (middle panel) % cells in each cell cycle phases for

WI38 fast iCas9 with KO of tRNA<sup>Arg</sup><sub>CGA</sub>. (right panel) % cells in each cell cycle phases for WI38 fast iCas9 with KO of tRNA<sup>Ser</sup><sub>TCG</sub>.

The change in the distribution of the cell population along the cell cycle phases can suggest that reducing the levels of some tRNAs may affect cells progression through the different cell cycle phases.

When exploring the consequence of different tRNA KOs, that belong to both tRNA groups (proproliferation and pro- differentiation tRNAs) on the cell cycle progression in all three cell lines, interesting behaviors are revealed. tRNA KOs in WI38 fast and slow cells resulted in a wide distribution along the S and G0-G1 population axes, suggesting different outcomes for distinct treatments (KO of control groups, KO of pro-proliferation tRNAs and KO of pro-differentiation tRNAs) (figure 7). In contrast, the same treatments in HeLa cells resulted in more condensed distribution along both S and G0-G1 population axes (figure 7), suggesting smaller effect of gene KO, especially tRNAs KO, in HeLa cells compared to WI38 slow and fast.



Figure 7 – Cell percentage in S and G0-G1 phases for HeLa, WI38 fast and WI38 slow cells with knock-out of different tRNA and control genes

HeLa iCas9 (square), WI38 fast iCas9 (star) and WI38 slow iCas9 (triangle) samples were labeled with BrdU and PI and analyzed for cell cycle progression. Using FlowJo software, cell percentage in each cell cycle phase was determined, according to the labeling values. The samples (of all three cell lines) are labeled differently for controls (green), pro-proliferation tRNA KOs (red) and pro-differentiation tRNA KOs (blue).

When examining each tRNA KO individually, different responses of the cell cycle progression were revealed in all three cell lines. In the pro-proliferation tRNAs group, tRNA<sup>Ile</sup><sub>AAT</sub> KO resulted in minor to modest effect on the S phase population size, relative to the other tRNAs KO in all three cell lines (figure 8A-C). Both tRNA<sup>Arg</sup><sub>TCG</sub> and tRNA<sup>Arg</sup><sub>TCT</sub> strongly reduced the S phase population of HeLa and WI38 fast cells (figure 8A-B). This effect is almost eliminated in WI38 slow cells (figure 8C). tRNA<sup>Leu</sup><sub>TAG</sub> showed the opposite trend, KO of this tRNA slightly reduced the S phase population of HeLa and WI38 fast cells, while in WI38 slow cells the reduction of the same tRNA was more intense (figure 8A-C). For the prodifferentiation tRNAs group, both tRNA<sup>Pro</sup><sub>AGG</sub> and tRNA<sup>Leu</sup><sub>CAG</sub> showed strong decrease in the S phase population of all three cell lines after perturbating their expression level, while low levels of tRNA<sup>Ser</sup><sub>CGA</sub> presented a small decrease in the S phase population in HeLa and WI38 fast, while in WI38 slow cells the sphase population of the same cell lines (figure 8A-C). tRNA<sup>Pro</sup><sub>CGG</sub> KO showed moderate decrease of the S phase population in HeLa and WI38 fast, while in WI38 slow cells this tRNA KO elevated its effect on the S phase population (figure 8A-C). These results reinforced the notion that within each tRNA pool (pro- proliferation and pro- differentiation), knock- out of various tRNAs had a different effect on the proliferation phenotypes of the three cell lines, such as proliferation rate and cell cycle progression.





Figure 8- Percentage of cells in S and G0-G1 phases for HeLa, WI38 fast and WI38 slow cells knocked-out for tRNA and control genes.

Cell percentage in S and G0-G1 phases for A| HeLa icas9 cells B| WI38 fast icas9 cells C| WI38 slow icas9 cells. In each cell line, samples are labeled differently for control (green), pro-proliferation tRNA KOs (red anti-codons) and pro-differentiation tRNA KOs (blue anti-codons).

By looking at the effect of pro-proliferation and pro-differentiation tRNA KOs in different cell types, we found that in WI38 fast cells the separation between pro-proliferation and pro-differentiation tRNAs along the S phase population and G0-G1 population axes was much less dominant than in WI38 slow cells (figure 7). In WI38 fast cell, 7 out of 8 tRNA deletions reduced the S phase population in approximately 20% compared to the WI38 fast iCas9 + Doxy control (figure 8B). In WI38 slow cells, the separation between the two tRNA groups is more apparent, when 3 out of 4 pro-differentiation tRNA deletions showed the strongest effect on the S phase population, and the pro-proliferation tRNA deletions indicated a more modest behavior (figure 8C).

Another informative measurement from the cell cycle FACS analysis is the BrdU incorporation intensity. According to previous studies, BrdU incorporation intensities can reflect replication forks speed during S phase, or in other word, the speed of DNA replication. By treating human cell lines with hydroxyurea (HU), which is a replication fork inhibitor, it was shown that the anti-BrdU intensity is reduced in a dose-depended manner <sup>29</sup>.

In general, there is clear trend of strong reduction in the BrdU incorporation in all three cell lines, while HeLa cells are affected especially from pro-proliferation tRNAs KO, and WI38 slow are affected mainly by pro-differentiation tRNAs KO (figure 9A-C). More specifically, in the pro-proliferation tRNAs group, tRNA<sup>Arg</sup><sub>TCG</sub> KO led to a strong decrease in the BrdU incorporation in all three cell lines, while tRNA<sup>Arg</sup><sub>TCT</sub> and tRNA<sup>Leu</sup><sub>TAG</sub> did not change significantly the BrdU incorporation, relative to other tRNAs KO (figure 9A-C). In the pro-differentiation tRNAs group, tRNA<sup>Ser</sup><sub>CGA</sub> KO decreased the BrdU incorporation in all three cell lines, while tRNA<sup>Pro</sup><sub>AGG</sub> KO almost did not change the BrdU incorporation in HeLa cells, but in WI38 fast and WI38 slow its effect increased.

In addition to tRNA KO samples, we tested control samples (p53 KO, RP KO, iCas9 cells with/without doxycycline incubation). Similar to the proliferation rate assay, the control samples, specifically p53 KO and RP KO don't show the expected consequence on the cell cycle progression and BrdU incorporation, which is a strong reduction in the S phase population and in the BrdU incorporation upon RP KO, and a small to no effect on the S phase population and on the BrdU incorporation upon p53 KO (figure 8 and 9). Together with the proliferation assay, the control samples raise suspicions regarding the validity of the results.







Figure 9 – BrdU incorporation intensity for HeLa, WI38 fast and WI38 slow cells knocked-out for tRNA and control genes.

BrdU incorporation levels were determined following incubation of cell samples with BrdU for 1 hour. BrdU incorporation during S phase for A| HeLa icas9 cells B| WI38 fast icas9 cells. C| WI38 slow icas9 cells. In each cell line, samples are sorted to control (green), pro-proliferation tRNA KOs (red anti-codons) and pro-differentiation tRNA KOs (blue anti-codons).

The cell cycle results strengthen the previous observation, that tRNAs KO exert a negative effect on proliferation-related features such as proliferation rate, cell cycle progression and S phase speed.

# 5.5. Hierarchical clustering of all tRNA knock- outs, cell lines and assays

Finally, we sought to zoom-out, and explore how all 8 tRNA deletions we investigated, and how all different assays and cell lines correlated with each other. To that end, we performed Hierarchical clustering, where one is based on the different tRNAs KO and the other one is based on the different assays and cell lines (figure 10). From the tRNA KO- based clustering we observed almost clear separation between the pro-proliferation tRNA KOs and the pro-differentiation tRNA KOs, which supported our hypothesis that deletion of the two tRNA pools had a different effect on human cells (figure 10). When clustered based on cell line and assay types, we noticed that most of the experiments performed in HeLa cells are clustered together, and WI38 fast and slow are clustered together, but less separable. This may reflect the great difference between the cell lines, especially between HeLa cells and

the WI38 cell lines, when the later originated from the same ancestor cell (figure 10). Interestingly, BrdU incorporation measurement are clustered together, with no relation to cell line (figure 10), suggesting that the different tRNA deletions affects the S phase speed in a consistent manner in all cell lines.





Rows- Hierarchical clustering of the different tRNA deletions, based on all experiments (sgRNA competition, proliferation assay, cell cycle progression and BrdU incorporation). Pro-proliferating tRNAs are labeled in red and pro-differentiation tRNAs are labeled in blue. Columns- Hierarchical clustering of the different assays and cell type, based on all tRNA deletions. Heat- map represent the normalized values of each tRNA deletion in each assay and cell line (Z- transformation normalization).

In general, the results reflect the proliferative nature of the cell lines, while disturbing pro-proliferation tRNA level in HeLa cells showed deleterious effect on the proliferation features of the cells. This phenomenon reduced in its strength when targeting pro-proliferation tRNAs in WI38 cells, especially in WI38 slow cells.

# 6. Discussion

In this project, we aimed to reveal the causative relation between the proliferative state of the cell and the expressed tRNA pool. We addressed this question by evaluating the consequence of tRNA expression perturbation in human cell lines on their viability and growth. Here, we initially report on knock-out of tRNA genes using a genomic editing technique, CRISPR- Cas9.

In this work we study two groups of tRNAs as defined by Gingold et al.: pro- proliferation tRNAs which are tRNAs that correspond to codons that are over-represented among the cell cycle genes, and prodifferentiation tRNAs which are tRNAs that correspond to codons over-represented among the celldifferentiation genes<sup>13</sup>. This means that pro- proliferation tRNAs are more demanded in highly proliferating cells than pro- differentiation tRNA. Nevertheless, it should be noted that prodifferentiation tRNAs may still be necessary and essential for the viability of highly proliferating cells. This is true also for slowly proliferating cells and pro- proliferation tRNAs.

To assess whether deletion of tRNAs in the genomic level affects human cell viability, we performed competition assays between 20 different HeLa strains, each one carries a different sgRNA targeting a unique tRNA type. Half of the sgRNA which were targeting tRNAs that belong to the pro-proliferation tRNA pool, and the other half of the sgRNA which were targeting tRNAs that belong to the pro-differentiation tRNA pool. We found that cells carrying sgRNAs that targeted pro-proliferation tRNAs showed a strong decrease in their fitness over time, while cells carrying sgRNAs that target pro-differentiation tRNAs showed hardly any fitness reduction. This was true for most of the pro-proliferation and pro-differentiation tRNAs types. These findings constitute as proof of concept for our main hypothesis (as was published in *Cell* in 2014)<sup>13</sup>, knocking- out tRNAs that are highly expressed in cancerous cell samples (pro-proliferation tRNAs) is deleterious to these cells and negatively affect the cell's fitness, while knocking-out tRNAs that are lowly expressed in cancerous cell samples (pro-differentiation tRNAs) doesn't present this deleterious effect to the same extent.

We further examined whether we can find any genomic or epigenomic features that can explain the differential fitness obtained upon targeting different tRNA genes in HeLa cells. The first feature we examined was the targeting percentage of each sgRNA. In the human genome, most of the tRNA genes

have several loci, some of them with identical sequence and some are only partly similar. Therefore, each sgRNA we designed is predicted to target only loci with shared sequence that is highly specific for the tRNA of interest, to avoid off targets manipulations. As a result of the high specificity however, in some cases only part of the tRNA loci are predicted to be targeted. We showed that the targeting percentage of the sgRNA is in a negative correlation with the cell's fitness, meaning that gene families that are heavily targeted reduced the fitness more than others. Nonetheless, this alone cannot fully explain the fitness results since even when taking targeting percentage into account, pro- proliferation tRNAs KO reduces the cell's fitness and are more deleterious to the cells than pro- differentiation tRNAs KO. Nevertheless, there were some outliers, in which tRNAs with low targeting percentage strongly reduced the cell's fitness. These outliers suggested that there are other genomic and epigenomic features that may contribute to the fitness determination for each tRNA KO cell. To identify these features, we utilized two machine learning techniques and analyzed tens of genomic and epigenomic feature of tRNA genes in HeLa cells, mainly histone modification and transcription factors bindings in the tRNA genes' vicinity, to find the features that best predict the cell's fitness following tRNA deletion. Two epigenomic features got the highest scores in both methods, Ezh2 binding and MNase- seq (nucleosomes occupancy). The two features negatively correlated with the fitness value, with Ezh2 binding had a higher effect on fitness than MNase- seq. Interestingly, binding of Ezh2 and the presence of nucleosomes in the vicinity of genes have a repressive effect on gene transcription <sup>27</sup>, <sup>28</sup>. Moreover, recent studies have shown that nucleosome presence in the target site can interfere with CRISPR-Cas9 activity <sup>30</sup>. These findings are counter intuitive since deletion of genes that are down-regulated in the cell is less likely to affect cell fitness. Our results, however, show the opposite trend, in which genes with low transcription level (meaning high Ezh2 binding or MNase values) tend to have a stronger effect on the cell's fitness.

The sgRNA competition results indicate that tRNA depletion does indeed cause changes in the cell's viability. More importantly, pro-proliferating tRNA deletion are more deleterious to HeLa cells than prodifferentiation tRNA deletions. The reason for this effect still require further investigation, however we can suggest that different properties like sgRNA targeting ability and binding of suppressive factors to the tRNA genes involve in the negative affect of tRNA deletions.

Next, we aimed to characterize deletion of specific tRNA genes through sets of phenotypic assays. We chose eight tRNA deletions, four belong to the pro-proliferation tRNAs and the other four belong to the pro-differentiation tRNAs. Common feature to these tRNAs is the targeting percentage of their sgRNAs, which is approximately 100%. This way we could compare between the experimental results of these tRNAs knock-out, knowing that they all predicted to be fully targeted. The limitation in choosing 100%

targeted tRNAs is the high probability of having a destructive outcome for the cell's viability, with no relation to specific tRNA group, due to gene essentiality (as mentioned earlier). The phenotypic assays we performed included MTT- based proliferation assay, in which metabolically active cells are quantified and the proliferation and cell's viability is being determined. Additionally, we analyzed cell cycle progression and S phase speed of the tRNA- deleted cells using cell cycle analysis. Beside HeLa cells, we also explored WI38 fast and slow cells (with the same tRNA deletion as HeLa cells), which are cell lines that originated from the same primary WI38 fibroblast cells but differ in their doubling time and few additional genomic characterizations. Together, these three cell lines (HeLa, WI38 fast and WI38 slow) create a model system with a gradually increasing proliferation rate, as WI38 slow cells are in the low end of the proliferation trajectory and HeLa cells are in the high end.

In general, the findings from all phenotypic assays didn't recapitulate the significant separation between pro- proliferation and pro- differentiation tRNA pools, observed in the competition assay. Moreover, these assays emphasized that each tRNA, regardless of its pertinence to the tRNA subsets, contributed in a different manner to the examined phenotype of the cell. For example, tRNA<sup>Ser</sup><sub>CGA</sub> and tRNA<sup>Leu</sup><sub>CAG</sub> are both part of the pro- differentiation tRNA pool. Deletion of these two tRNAs resulted in opposite effects, in which tRNA<sup>Ser</sup><sub>CGA</sub> KO didn't have any effect on the cell cycle progression or the proliferation rate in all three cell lines, while deletion of tRNA<sup>Leu</sup><sub>CAG</sub> led to a strong reduction in the proliferation rate and created G0-G1 arrest in all cell lines (more moderate in HeLa cells). Notably, each tRNA deletion in our system showed a distinct effect and strength on the different proliferation phenotypes we explored, as was mentioned in the Results section. This may suggest that the tRNAs we investigated in this project, that are predicted to be targeted in approximately 100% of their loci, have some degree of essentiality in each of the cell lines we investigated (regardless to their definition as pro-proliferative or pro-differentiation).

Another interesting finding we observed in almost all tRNA deletions is that each cell line reacted differently to the same tRNA KO. Moreover, while HeLa cells and WI38 slow cells behave differently to the same tRNA deletion, WI38 fast cells show common behaviors with both of those cells, depending on the tRNA type. For instance, deletion of tRNA<sup>Arg</sup><sub>TCT</sub> (pro- proliferation tRNA) cause a strong reduction in proliferation rate and G0-G1 arrest in HeLa cells, while in WI38 slow cells this effect almost abolish. In WI38 fast, KO of this tRNA caused G0-G1 arrest, but did not reduce the proliferation rate. This example, and other tRNA deletions, points on the nature of WI38 fast cells, which have an intermediate proliferation rate.

The phenotypic assays revealed an interesting observation that we find hard to interpret. While in some tRNA deletions, we see consistency in the results of the different assays, other tRNA deletion did not

illustrate this behavior, specifically showing opposite trends for the proliferation rate and cell cycle progression results, like deletion of tRNA<sup>Arg</sup><sub>TCG</sub> and tRNA<sup>Leu</sup><sub>TAG</sub> (both belong to the pro- proliferation tRNA pool). These cases are counter intuitive to the logic, that when tRNA KO results in proliferation rate reduction, cell cycle progression should be also disturbed, as deceleration of one of the cell cycle phases or even cell cycle arrest.

I must note though that the controls used in these experiments – KO of the ribosomal protein, and KO of p53 did not yield the expected results, namely reduced and increased growth respectively, in many of the cell lines and assays used. This puzzling result may suggest that some of the assays used might not reliably recapitulate growth rate defects.

From the results of the phenotypic experiments, proliferation assay and cell cycle analysis, we can conclude that:

- Some tRNA KO results reflect the differences between the three cell lines nature. HeLa and WI38 slow, in most cases, reacted differently to the same tRNA deletion. WI38 fast share common behaviors with both HeLa and WI38 slow cells that correspond to their intermediate position on the proliferation rate trajectory of all three cell lines.
- Each tRNA KO that were examined showed a distinct effect on the proliferation features we analyzed, not necessarily complement to its predicted effect, according to its tRNA group definition.
- The results from the proliferation assay and cell cycle analysis don't necessary correlate. For example, KO of specific tRNA cause reduction in proliferation rate, while it doesn't affect the cell cycle progression or speed, and vice versa.
- While the measurements of proliferation rate and cell cycle progression can be very distinct between different cell lines for each tRNA KO, for most cases the measurement of BrdU incorporation is unified for the different cell lines per tRNA deletion.
- Unlike the clear separation between pro-proliferation tRNA and pro-differentiation tRNA observed in the sgRNA competition experiment, the cell cycle and proliferation assay results in less distinct separation. This can be due to several issues that need to be addressed:
  - All three measurements, proliferation rate, cell cycle progression and S phase speed, were measured in the population level. Briefly, the iCas9 was induced for a constant period of time in which the relevant tRNA can be targeted and the surviving population was analyzed. This is certainly complicated by itself, because the population is heterogenous, that is, each cell carries different type of Indel mutation in the tRNA gene and therefor

different tRNA KO. In addition, we cannot rule out the possibility that the population also contains un-targeted cells. This heterogeneity may influence the behavior of the entire population, in which specific cells (for example, cells with partial knock-out) may have a small advantage over other cells, and they will take over the population. To conclude, examining heterogenous population allows fast and high-throughput screening and certainly has its advantages, but can also mask the results.

- As noted earlier, the chosen tRNA for the phenotypic experiment showed deleterious effect on the cell's fitness in the sgRNA competition experiment. This effect correlated with the high targeting percentage of the sgRNAs. This may suggest that full deletion of tRNA genes, regardless of pertinence to specific tRNA pool (pro-proliferation tRNAs or pro-differentiation tRNAs), cause significant reduction in cell viability and growth.
- The results represented in this work represent only one biological repeat for each experiment. Therefore, additional repeats should be performed.

In order to address the heterogeneity issue discussed above we also created clonal cells with KO of specific tRNA. The main challenge we faced originated from the fact that each tRNA type has multiple loci in the human genome. This raises a few difficulties. Firstly, determining the optimal time needed for iCas9 induction so that it will allow cells to accumulate Indel mutations in all loci that are predicted to be targeted by the sgRNA of specific tRNA (in both alleles), without having off-targets. Secondly, applying the right detection method for tRNA- KO clones cells that will be efficient and high-throughput. Nonetheless, we succeeded to create seven WI38 slow and fast clonal cells, carrying various Indel mutations in distinct loci of tRNA<sup>lle</sup><sub>AAT</sub>. All Indel mutation but one (clone #11) resulted in heterogenous DNA repair for the two alleles, in which only one allele underwent editing, while the second allele was identical to the WT allele. This caused only partial knock-out of the locus, or should be referred to as tRNA knock-down. The highest number of edited loci in one clonal population (clone #16) we found was identical to the WT). It's important to note that this clonal population (clone #16) didn't survive. These results may suggest that full knock out of tRNA<sup>lle</sup><sub>AAT</sub> is deleterious to the cells, which may be the reason for not being able to detect this kind of clonal population.

One of the major difficulties in editing non- coding RNA gene is the determination of successful knockout of the gene. While in protein coding genes knock-out by CRISPR is often obtained due to disruption of reading frame, the absence of open reading frame in non- coding gene hardens the possibility of the gene to not be expressed. Nevertheless, because the tRNA molecule is a highly-structured molecule, containing nine key sites (for example, three anticodon positions, three TΨC loop sites, one D stem loop and two paired TΨC stem sites) Indel mutation in these sites may contribute to the knock-out of the tRNA gene functions. Moreover, The TΨC loop and stem sites contain elements from the B box region of the internal promoter. This region is important for the TFIIIC transcription factor binding and tRNA gene transcription by Pol III<sup>9</sup>. We can therefore in future use the information regarding these critical structurefunction positions along the tRNA sequence, to predict whether the editing type in a specific locus is likely to result in dysfunction tRNA molecule. Previous studies have shown that as the number of accumulate mutations increase within specific tRNA molecule, the cell's fitness decrease<sup>9</sup>. Although this study was carried out in yeast cells, we can assume that the negative correlation between number of mutations in the tRNA molecule and the cell's fitness is also true for human cells. Thus, it is likely that deletion of multiple nucleotides, not necessarily including the anti-codon, will disrupt the proper folding of the tRNA which will abolish the functionality of the tRNA molecule and may reduce cell's fitness. This is the case for most Indel mutation present in the single clones we created (see Results).

Our next step in this project is to address the problematic issue of detecting tRNA levels in cells after CRISPR-Cas9 induction. We will achieve this by directly sequencing the tRNA pool in various tRNA KO populations and compare it to the relevant WT cell line using a tRNA deep-seq protocol that is being developed in our lab. Furthermore, we wish to characterize the effect of knocking out the various tRNA genes (both pro-proliferation and pro-differentiation) on the proteome in all 3 cell lines. We believe that the absence of a desired tRNA will influence the protein distribution in the tRNA KO cell, specifically reduce the highly-translated proteins, compared to the Wild Type. We will detect and compare the proteome levels in Wild- Type cells and tRNA- deleted cells using Mass- Spectrometry analysis. In addition, we plan to characterize the Indel mutations generated in the different tRNA loci by deep-sequencing the tRNA genes in the population. Finally, currently our analyses were focused on exploring the effects of changing the tRNA pool in cells with different proliferation rates. Another interesting question is to look on cells undergoing transition in their proliferation/differentiation state. Examples for cell state transition we wish to examine are reprogramming of fibroblasts to iPSs, trans-differentiation, starvation and starvation release and senescence induction.

In conclusion, in this project we report on a new purpose of genomic editing in human cell lines, in which tRNA genes are being edited by CRISPR-Cas9 technique with only a single sgRNA molecule for several genes, all belonging to the same tRNA family. We demonstrated that deletion of tRNA genes affects human cell's fitness, viability and growth. We further presented that three different cell lines, HeLa, WI38 fast and WI38 slow cells that differ in their transformation characterizations and proliferation rate,

react differently to distinct tRNA KO. Moreover, perturbations of some tRNA level were more deleterious to highly proliferating cells like HeLa cell line, while WI38 slow cells showed minimal effect. We believe that the different responses of cell lines, in a distinct proliferation state, to various tRNA KO, can open a new field of cancer investigation and treatment, by targeting tRNAs that are more essential and demanded in cancer cells rather than in normal, differentiated cells, which will lead to cancer tissue damage, while normal cells will not be damaged.

# 7. Literature

- Klochendler, A. *et al.* A Transgenic Mouse Marking Live Replicating Cells Reveals In Vivo Transcriptional Program of Proliferation. *Dev. Cell* 23, 681–690 (2012).
- Zhu, L. & Skoultchi, A. I. Coordinating cell proliferation and differentiation. *Curr. Opin. Genet.* Dev. 11, 91–97 (2001).
- Milyavsky, M. *et al.* Prolonged Culture of Telomerase-Immortalized Human Fibroblasts Leads to a Premalignant Phenotype. *Cancer Res.* 63, 7147–7157 (2003).
- 4. Milyavsky, M. *et al.* Transcriptional programs following genetic alterations in p53, INK4A, and H-Ras genes along defined stages of malignant transformation. *Cancer Res.* **65**, 4530–4543 (2005).
- Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* 502, 65–70 (2013).
- 6. Sherr, C. J. Cancer cell cycles. *Science* **274**, 1672–1677 (1996).
- Ivey, K. N. & Srivastava, D. MicroRNAs as regulators of differentiation and cell fate decisions. Cell Stem Cell 7, 36–41 (2010).
- Goodenbour, J. M. & Pan, T. Diversity of tRNA genes in eukaryotes. *Nucleic Acids Res.* 34, 6137–6146 (2006).
- Li, C., Qian, W., Maclean, C. J. & Zhang, J. The fitness landscape of a tRNA gene. *Science (80-. )*.
  352, 837–840 (2016).
- Gingold, H., Dahan, O. & Pilpel, Y. Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome. *Nucleic Acids Res.* 40, 10053–10063 (2012).
- 11. Barski, A. *et al.* Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes. *Nat. Struct. Mol. Biol.* **17**, 629–634 (2010).
- Pavon-Eternod, M. *et al.* tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Res.* 37, 7268–7280 (2009).
- 13. Gingold, H. et al. A Dual Program for Translation Regulation in Cellular Proliferation and

Differentiation. Cell 158, 1281–1292 (2014).

- 14. Bloom-Ackermann, Z. et al. A Comprehensive tRNA Deletion Library Unravels the Genetic Architecture of the tRNA Pool. *PLoS Genet.* **10**, (2014).
- 15. Goodarzi, H. *et al.* Modulated expression of specific tRNAs drives gene expression and cancer progression. *Cell* **165**, 1416–1427 (2016).
- Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. Science (80-.). 346, 1258096–1258096 (2014).
- Ran, F., Hsu, P., Wright, J. & Agarwala, V. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–308 (2013).
- Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278 (2014).
- 19. Shalem, O. et al. Genome-Scale CRISPR-Cas9 Knockout. Science (80-.). 343, 84–87 (2014).
- 20. Ho, T.-T. *et al.* Targeting non-coding RNAs with the CRISPR/Cas9 system in human cell lines. *Nucleic Acids Res.* **43**, e17 (2015).
- 21. Sambrook, J. & Russell, D. W. Molecular Cloning. (Cold Spring Harbor Laboratory Press, 2001).
- Blecher-Gonen, R. *et al.* High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat. Protoc.* 8, 539–54 (2013).
- 23. Tibshirani, R. Regression Selection and Shrinkage via the Lasso. *Journal of the Royal Statistical Society B* **58**, 267–288 (1996).
- 24. Breiman, L. Randomforest2001. 1-33 (2001). doi:10.1017/CBO9781107415324.004
- 25. Liaw, a & Wiener, M. Classification and Regression by randomForest. *R news* 2, 18–22 (2002).
- 26. Brinkman, E. K., Chen, T., Amendola, M. & Van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, (2014).
- Simon, J. a & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. Nat. Rev. Mol. Cell Biol. 10, 697–708 (2009).
- Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* 10, 161–72 (2009).
- 29. Ge, X. Q. & Blow, J. J. Chk1 inhibits replication factory activation but allows dormant origin firing in existing factories. *J. Cell Biol.* **191**, 1285–1297 (2010).
- Horlbeck, M. A. *et al.* Nucleosomes impede cas9 access to DNA in vivo and in vitro. *Elife* 5, 1–21 (2016).

# 8. Acknowledgments

I would like to thank my mentor, Yitzhak Pilpel, for his support and guidance – You taught me how to think science and how to think outside the box. I wish to thanks Hila Gingold, Idan Frumkin and Danit Finkelshtein who started this project and have been a guiding hand all through my Masters. I wish to thank all my lab members for many fruitful discussions. Special thanks go to Orna Dahan – for helping me on a daily basis, experimentally and scientifically. For Tammy Biniashvily and Dvir Schirman for helping with the computational aspects in my work. I thank Tomer Meir Salame from the Flow cytometry units for helping with the cell cycle experiments and FACS sorting. I thank also Yael Ayalon and Noa Furth from Moshe Oren lab for all the professional support in the tissue culture. Lastly, I would like to acknowledge "The Nancy & Stephen Grand Israel National Center for Personalized Medicine" and Shlomit Gilad who worked on the library preparation and sequencing.