universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Towards a Time- and Cost-Effective Protocol for the Systematic Analysis of Cellular Gene Expression Kinetics in Scale“

verfasst von / submitted by

Lisa Slama, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022/ Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 834
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molekulare Biologie
Betreut von / Supervisor:	Univ.-Prof. Dr. Stefan Ameres

Acknowledgments

Firstly, I would like to thank my supervisor Univ.-Prof. Dr. Stefan Ameres, for letting me work on this exciting project. He gave me lots of input and support, while also challenging me to figure things out on my own.

Next, I want to express my gratitude to Tsimafei Navalayeu, who was my mentor in the lab. Thank you for including me in your PhD project and guiding me. He always knew whenever to push me and when I needed clear explanations. I really appreciate his trust in my work and that I always had the possibility to come to him for any kind of question or confusion.

Finally, I want to thank all my other colleagues from the Ameres lab for all their help, useful feedback, and fun times during breaks. They are a talented group of people, who always push each other to do better, but also to enjoy the time while doing it.

Abstract

Gene expression is an essential and highly regulated process, which involves the production, processing, and decay of RNA. Currently available RNA sequencing protocols provide rapid insights into qualitative and quantitative changes of the transcriptome at steady-state but are blind to the relative kinetics of RNA biogenesis, processing and turnover. This has changed with the recent development of time-resolved RNA sequencing approaches, such as SLAMseq. SLAMseq uses 4-thiouridine (s^4U)-based metabolic RNA labeling in cultured cells followed by thiol-linked alkylation, resulting in specific T-to-C conversions at the site of s^4U-incorporation upon cDNA conversion and sequencing. While this approach provides access to the cellular rate of RNA biogenesis and turnover in the context of steady-state gene expression profiles, its application to massive parallel screening approaches is limited by high sequencing costs associated with conventional global RNA sequencing.

Here, I developed a multiplexed targeted SLAMseq approach to overcome the current limitations in the systematic inspection of gene expression kinetics in scale. To this end, I combined SLAMseq with targeted 3′ mRNA sequencing of candidate transcripts and systematically optimized the parameters and reagents for cDNA library preparation. As a result, I was able to obtain robust kinetic parameters for up to 30 different transcripts with consistent representation despite vastly diverging steady-state expression levels. A further increase in transcript numbers is possible and could be achieved by employing a semi-nested amplification approach. Finally, I propose an advanced workflow for future semi-automated library preparation, which involves in-lysate reverse transcription and barcoding. In summary, I developed robust, scalable, and cost-effective protocols for the systematic profiling of gene expression kinetics that will be useful for the future massive parallel dissection of gene regulatory pathways.

Zusammenfassung

Die Genexpression ist ein wesentlicher und stark regulierter Prozess, der die Produktion, Verarbeitung und den Zerfall von RNA umfasst. Die derzeit verfügbaren RNA-Sequenzierungsprotokolle bieten schnelle Einblicke in die qualitativen und quantitativen Veränderungen des Transkriptoms im Steady-State, erlauben aber keine Einsicht in die relativen zellulären Kinetiken der RNA-Biogenese, -Verarbeitung und -Umsetzung. Dies hat sich mit der Entwicklung von zeitaufgelösten RNA-Sequenzierungsverfahren wie SLAMseq geändert. SLAMseq nutzt die 4-Thiouridin (s^4U)-basierte metabolische RNA-Markierung in kultivierten Zellen, gefolgt von Thiol-verknüpfter Alkylierung, was zu spezifischen T-zu-C-Umwandlungen an der Stelle des s^4U-Einbaus bei der cDNA-Konvertierung und Sequenzierung führt. Während dieser Ansatz Zugang zur zellulären Rate der RNA-Biogenese und des RNA-Umsatzes bietet, ist seine Anwendung bei massiven parallelen Screening-Ansätzen durch die hohen Sequenzierungskosten im Zusammenhang mit der herkömmlichen globalen RNA-Sequenzierung begrenzt.
In dieser Arbeit entwickelte ich einen gezielten Multiplex-SLAMseq-Ansatz, der die systematische Untersuchung der Genexpressionskinetik ermöglicht. Zu diesem Zweck habe ich SLAMseq mit der gezielten 3'-mRNA-Sequenzierung von Kandidaten-Transkripten kombiniert und die Parameter und Reagenzien für das Protokoll zur Vorbereitung der cDNA-Bibliothek systematisch optimiert. Als Ergebnis konnte ich robuste kinetische Parameter für bis zu 30 verschiedene Transkripte mit konsistenter Darstellung erhalten. Eine weitere Erhöhung der Transkript Anzahl ist möglich und kann durch den Einsatz eines verschachtelten Amplifikationsansatzes erreicht werden. Schließlich schlage ich einen minimalisierten Arbeitsablauf für die künftige halbautomatische cDNA Bibliotheksbereitung vor, der die reverse Transkription und das Barcoding im Lysat umfasst. Zusammenfassend lässt sich sagen, dass ich robuste, skalierbare, und kosteneffiziente Protokolle für die systematische Erstellung von Profilen der Genexpressionskinetik entwickelt habe, die für die künftige massive parallele Untersuchung von Genregulationswegen nützlich sein werden.

Table of Contents

ACKNOWLEDGMENTS **3**

ABSTRACT **5**

ZUSAMMENFASSUNG **7**

1. INTRODUCTION **11**

1.1. Gene expression 11
1.2. Transcription 11
1.2.1. Transcriptional Regulation 13
1.2.2. Post-Transcriptional Regulation 14
1.3. mRNAs and Half Lives 16
1.4. Gene Expression Dynamics 17
1.5. SLAMseq 19
1.6. Aim of the Study 24

2. RESULTS **26**

2.1. Reverse Transcription Optimisation 26
2.2. Decreasing RNA input and comparing oligo dT18 vs dT30 28
2.3. Multiplexing with Individual Barcodes 30
2.4. Implementation of the Semi-nested approach 31
2.5. Exhaustive PCR 33
2.6. Titration of Primer concentration 34
2.7. Alkylation and Reverse Transcription in-Lysate 35

3. DISCUSSION **39**

4. MATERIAL AND METHODS **45**

4.1. Cell Line: 45
4.2. Reagents and Buffer: 45
4.3. General Equipment: 46
4.4. Oligos: 47
4.5. Cell Culture 49
4.6. SLAMseq in mESCs 49
4.7. RNA Isolation 49
4.8. Iodoacetamide (IAA) Treatment 50
4.9. TURBO DNA-free Protocol 50
4.10. RNA library preparation 50
4.10.1. Reverse Transcription 50
4.10.2. PCR1 – Second Strand Synthesis 51
4.10.3. PCR2 51
4.10.4. Quality Control 51
4.10.5. Sequencing with MiSeq 52
4.11. Reverse Transcription Optimisation and ExoSAP-IT Addition 52
4.12. RNA Input Decrease and Comparison of dT18 and dT30 53
4.13. Multiplexing with Individual Barcodes 53
4.14. Implementation of the Semi-Nested Approach 53
4.15. Exhaustive PCR 54
4.16. Primer Titration 54
4.17. In-Lysate Protocol 55
4.18. In Vitro Alkylation with DMF 55
4.19. Bioinformatics and Data Analysis 56
4.19.1. Bioinformatics – Targeted SLAMseq Analysis (TSLAM) 57
4.19.2. Data Analysis 58

5. REFERENCES **59**

6. SUPPLEMENTARY MATERIAL **64**

1. Introduction

1.1. Gene expression

Each cell in an organism has the exact same genome, but only a subset of genes is expressed, which establishes unique cell types that carry out specific functions. Hence, the control of gene expression is essential in cellular differentiation and homeostasis and involves all processes by which instructions, stored in the DNA, are converted into purposeful products, including protein synthesis. As a consequence, gene expression is highly controlled at multiple levels, allowing cells to respond to changes in the environment (Guo, 2014). In this thesis, the focus lies on the regulation of gene expression at the level of RNA biogenesis and turnover.

1.2. Transcription

Transcription is the process of producing copies of DNA segments, resulting in RNA transcripts. Transcripts of genes code for proteins and are designated as messenger RNA (mRNA). RNA molecules deriving from other DNA segments, are called non-coding RNAs (ncRNA), these do not code for proteins. Transcription is carried out by the RNA polymerase, which reads a DNA sequence and produces the complementary and antiparallel RNA strand (Dhamija & Menon, 2018).

Transcription takes place in three main steps: Initiation, Elongation and Termination. During initiation, RNA polymerase II binds to the promoter sequence, where it opens the DNA strands to provide a single-stranded template. Through the elongation stage, Polymerase II reads the template strand, adding the complementary nucleotides in the direction from 5′ to 3′. The resulting RNA transcript has the same genetic information as the coding strand. The termination phase occurs when the polyadenylation is completed. Important to point out are the three types of RNA polymerase: RNA Polymerase I for ribosomal RNA (rRNA) synthesis, RNA Polymerase II for mRNA synthesis, and RNA Polymerase III for transfer RNA (tRNA) and 5S rRNA synthesis.

In eukaryotes, transcription occurs in the nucleus and when completed the mRNA gets exported to the cytoplasm. There is additional processing necessary, before

the transcript is called pre-mRNA. RNA modifications in mRNA include 5′ capping, 3′ cleavage and polyadenylation, and splicing (Guo, 2014).

Cellular identity is defined by gene expression profiles; and understanding the dynamics of gene expression regulation is an important part of the answer to many biological questions. Transcription is a highly controlled process that is regulated at the transcriptional and post-transcriptional level (Figure 1). For mRNAs, the transcription and translation rate together predominantly dictate the amount of protein produced; while the regulation of mRNA decay controls the persistence of genetic information. Transcriptional control is influenced by DNA binding proteins and the structure of chromatin, while post-transcriptional regulation includes mRNA processing, RNA binding proteins, mRNA localisation and degradation. Taken together, gene regulation is an essential function of every organism to increase its adaptability and versatility (Mata, Marguerat & Baehler, 2005).

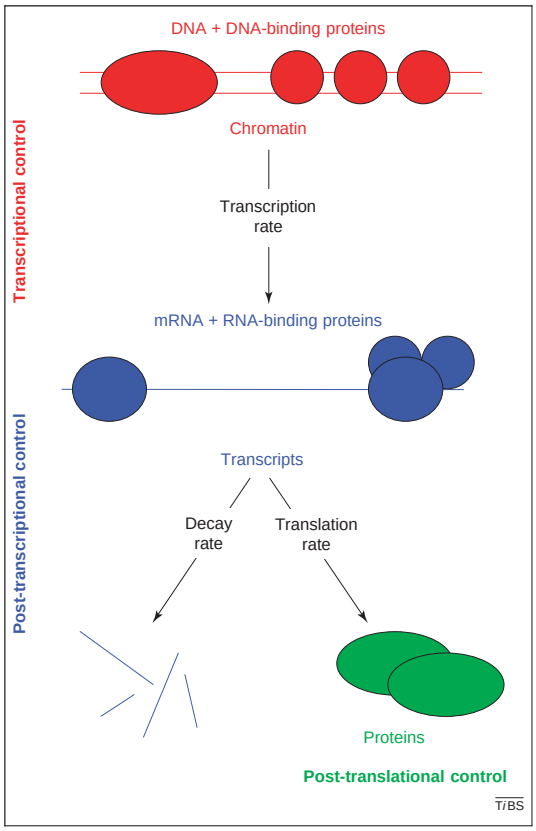


Figure 1:**Overview of steps of gene expression regulation** (Mata, Marguerat & Baehler, 2005) Transcriptional control (red) includes DNA sequence, DNA binding proteins, and the chromatin, which impacts the rate of transcription. In blue the post-transcriptional control is depicted, with mRNA, RNA binding proteins and the transcripts. This part influences the decay and translation rate. Post-translational control of proteins is depicted in green.

1.2.1. Transcriptional Regulation

The ability to adapt to internal stimuli or environmental changes, through activation of some genes and repression of others, is essential to the survival of any cell (Lee & Young, 2013).

One of the main transcription regulators are Histones, which are basic proteins in the nuclei, that can package the DNA into tight nucleosomes. Because of this the genes become inaccessible and are therefore silenced. Is the chromatin in this coiled form it is referred to as heterochromatin. To become accessible for transcription it needs to be opened into euchromatin, which is possible for example due to acetylation, methylation, or phosphorylation. Chromatin structure can be regulated by chromatin re-modellers or histone modifying factors (de Nadal, Ammerer & Posas, 2011).

To influence the rate of transcription at individual gene loci, different DNA binding proteins are in use, that are called Transcription factors (TF). They bind either promoter or enhancer regions of the DNA; depending on the TF, this can lead to an up or down regulation of the nearby gene. TFs can stabilize or hinder the binding of the polymerase, catalyse the reaction of uncoiling or coiling of histones and recruit coactivator or corepressor proteins. The activity of TFs themselves can be regulated, i.e. by ligand binding, phosphorylation, or interaction with coregulatory proteins (Kadonaga, 2004).

Stress is one of the main stimuli affecting cells gene expression, with typical stressors being heat, oxidation, or DNA damage (Russo & Olivias, 2015).
The response to any of these cellular stress events is usually a fast activated signal transduction pathway, which leads to the transcription of different genes to combat the stressor. This synchronously and quick response shows a high gene expression regulation (Shalem et al., 2008). Another trigger that requires precise regulation are metabolic changes. For instance, when nutrient availability is restricted, it induces changes in the transcriptional response. When taking yeast as an example, this would lead to alterations in their growth rate (Slavov & Botstein, 2013). Also, in early development, cues to regulate gene expression are

fundamental, because the cellular fate and function needs to be precisely correct for a living organism to emerge in the end. The correct mRNAs need to be transcribed, also the according amount, and located to their designated destination. Furthermore, all cellular events need to be regulated and quality controlled (Kadyrova et al., 2007; Asaoka-Taguchi et al., 1999).

Quality control in the cytoplasm is carried out mainly by Xrn1, whereas in the nucleus the main player in RNA decay is the exosome. Degradation will be discussed more later (Doma & Parker, 2007).

1.2.2. Post-Transcriptional Regulation

After transcription, pre-mRNA processing takes place before an mRNA is exported from the nucleus to the cytoplasm. The first processing step is the 5′ capping, which involves enzymatic reactions that add 7-methylguanosine to the 5′ end. The initial cap structure gets recognized by the cap binding complex, which together stabilize the mRNA and is an obstruction for the 5′-3′ exonucleases (Proudfoot, Furger & Dye, 2002). Eukaryotic mRNA consists of introns, i.e. non-coding regions, and exons, i.e. coding regions. With the help of the spliceosome, introns are removed and the remaining exons are fused together in a process referred to as splicing. Alternative splicing of exons can diversify the coding sequence of transcripts. Here some introns or exons are either removed or kept, creating unique combinations (Casamassimi & Ciccodicola, 2019). The formation of the 3′ end starts with a cleavage at the site between a conserved AAUAAA hexamer and a downstream sequence element, located within the 3′UTR. The cleavage and polyadenylation specificity factor (CPSF) and the poly-A polymerase (PAP) are required for the cleavage reaction and to direct poly-A addition. Around 200 Adenines are added to the 3′ end. A poly-A binding protein (PABP) binds the poly-A tail, which enhances the processivity of the PAP. The binding of the PABP to the poly-A tail promotes export from the nucleus to the cytoplasm, enhances translation, and protects the RNA from degradation. Poly-A tails of mRNAs in the cytoplasm get gradually shortened, ultimately resulting in mRNA degradation. Multiple factors influence how fast or slow an mRNA gets degraded, including RNA binding proteins and miRNAs (Proudfoot, Furger & Dye, 2002).

As seen in Figure 1, RNA binding proteins (RBP) play a dominant role in post-transcriptional regulation. They bind to mRNAs and form messenger ribonucleoproteins (mRNP), thereby influencing the fate and function of transcripts. RBP regulate multiple processes including mRNA processing, stabilization, localization, and translation. To give some examples, during 3′ end formation CPSF, plays a crucial role in defining the site of mRNA cleavage and polyadenylation. During alternative splicing different RBP recruit spliceosomes to the target site. For the mRNA to leave the nucleus it needs to be transported via a cargo-carrier complex involving different RBP, which are packaged into this complex with proteins, RNA, and other export factors. RNA binding proteins are critical in the localization of mRNAs to their intended target site. The regulation of gene expression relies on spatially regulated protein production, which is especially important during early development (Mata, Marguerat & Baehler, 2005; Moore, 2005).

Beyond transcription, mRNA degradation is a further means by which cells control mRNA levels. Some examples for when mRNAs need to be degraded would be when polyadenylation fails, when there are splicing defects, when there are no introns present in a gene that normally contains introns or when there are double-stranded RNAs (Doma and Parker, 2007). There are different options for mRNA turnover, the two to focus on here are the exonucleolytic and endonucleolytic pathway. Initiation of exonucleolytic decay typically follows deadenylation of mRNAs. Here, the enzyme Xrn1 is responsible for degrading transcripts by first decapping, which refers to the enzymatic removal of their 5′ cap. An alternative exonucleolytic pathway revolves around the exosome complex, which degrades transcripts from their 3′ end. The endonucleolytic pathway on the other hand employs endonucleolytic cleavage, as used by the RNAi machinery.
Decay rates can be controlled by different elements often located within the three-prime untranslated (3′UTR) and recognized by various RBPs (Mata, Marguerat & Baehler, 2005; Houseley & Tollervey 2009). The 3′UTR follows the coding sequence of the mRNA and influences gene expression post transcriptionally. A lot of regulatory elements reside in the 3′UTR of mRNAs. It is a binding site for miRNAs and regulatory proteins. miRNAs are short and non-coding RNAs, their

binding sites are referred to as microRNA response elements (MREs). If a miRNA binds to the MRE at the 3′UTR, it causes translational repression and/or mRNA degradation. Another common segment of the 3′UTR are the AU-rich elements (AREs). AREs consist of many copies of AUUUA sequences in a 50 to 150bp long segment. If an ARE binding protein attaches, it can activate translation, promote mRNA decay, or decrease mRNA stability. PABPs can also affect the translation, decay, stability, and export of mRNAs. A poly-A tail in an mRNA can enhance translation, but its absence or regulated shortening will often lead to mRNA degradation (Weiss & Ito, 2017).

1.3. mRNAs and Half Lives

Biogenesis, processing, and degradation of RNA are the underlying, molecular basis for gene regulation. To comprehend gene expression dynamics, the kinetics of RNA biogenesis and degradation needs to be measured (Herzog et al., 2017). mRNA half-lives can range from minutes to days (Lugowski et al., 2018). The half-lives of mRNAs can be connected to what type of mRNA it is. The most well-known are the housekeeping mRNAs, they are responsible for ensuring basic cellular functions, like metabolic pathways, cell cycle and apoptosis, and can be found in most cell types. The other category are regulatory mRNAs, which usually have shorter half-lives than housekeeping ones, because, as the name suggests, they have regulatory function. Because of that regulatory mRNAs need to be able to react quickly to stimuli, one example would be transcription factors (Morey & Avner, 2004).

The life span of mRNAs can be adjusted in response to environmental changes, stress, metabolic needs, or developmental signals. Getting information about gene expression dynamics requires a robust, cost-effective, and accessible transcriptomics method. Previous approaches were mostly cytotoxic or invasive, where both interfere with gene expression (Mata, Marguerat & Baehler, 2005).

1.4. Gene Expression Dynamics

Gene expression is highly dynamic, reflecting on the organism's ability to react to internal and external stimuli accordingly. This involves the regulation of gene expression on all levels, including RNA synthesis, processing, and decay. But basic RNA sequencing approaches fail to resolve these changes. Traditionally, these steps were seen as individual and isolated from each other (Cole, 2001). Komili and Silver (2008) describe different interplays and couplings of mechanisms of gene expression, showing deep interconnections. For example, at transcriptional regulation, if the DNA is not accessible for the transcription complex, then TFs can also not bind and therefore not activate or repress their targets transcription. So that they can fulfil their function chromatin re-modellers make specifically the DNA regulatory sequences accessible. The post-transcriptionally regulated mRNA processing also has revealed extensive coupling, including connections between the export and polyadenylation, as well as co-transcriptional recruitment of export and splicing factors. This is only a glimpse in the complexity of gene expression and its regulatory mechanisms, but not only is it complex it has also a fast response kinetic and reversibility, which can change the transcription and decay levels within minutes and just as quickly back to the basal state (de Nadal, Ammerer & Posas, 2011).

Basic RNA sequencing only reports the relative abundance of any given transcript at the timepoint of measurement. However, changes in gene expression can be a consequence of alterations in the rate of transcription, turnover, or both. To differentiate between those possibilities, direct changes in gene expression kinetics must be measured before and after perturbations. To experimentally address transcript stability and changes in it, a genetic or pharmacological block in transcription can be performed. A complete inhibition of transcription and measurement of the following degradation of pre-existing RNA is a standard experimental set up. To gather information on the dynamics underlying changes in gene expression, a collection of measurements at predefined time points after a given perturbation is necessary (Mata, Marguerat & Baehler, 2005). One method of perturbation would be to heat inactivate a temperature sensitive RNA polymerase and measure the following decrease in transcript abundance. Another

approach would be to employ cytotoxins to inhibit transcription chemically or use mutants of polymerase II. (Singer & Penman, 1972; Wang et al., 2002) These approaches, together with RNA sequencing, are the prevailing methods to study gene expression and its regulation (Grigull et al., 2004).

Global transcriptional interference methods were often inaccurate as, due to induced cellular stress, alternative pathways were possibly activated, which would interfere with regular transcription and decay (Haimovich et al., 2013). Therefore, metabolic labeling of RNA, which does not interfere with RNA synthesis or stability, was a welcome alternative. Initial protocols used radiolabeled nucleotides with probe-based filter-binding, which was functional, but precise and effective labeling was difficult, and radioactivity is invasive so an influence on the cells was probable (Ross, 1995). Later it was discovered that biologically inert ribonucleotide analogues would drastically optimize these approaches. The most promising one was labeling of transcripts with 4-thiouridine (s^4U) (Schofield et al., 2018).

When s^4U labelled RNAs are coupled with biochemical enrichment methods, i.e. biotin-streptavidin, it is often referred to as 4sU-seq (Boileau et al., 2021). The standard process would be to first label cells with s^4U, extract RNA, conjugate the RNA to biotin and fraction with streptavidin beads, where pre-existing and newly synthesised RNA are separated. Labeled RNA binds to the beads because biotin gets linked to the thiol group of s^4U beforehand. Finally, the nascent and total RNA fractions are either analysed by quantitative PCR, where half-lives can be calculated or individually sequenced (Russo et al., 2017). Schwalb et al. (2016) describe a method that builds on top of 4sU-seq but adds RNA fragmentation before isolating the labeled fragments. The approach is called transient transcriptome sequencing (TT-seq) and here the s^4U labeling of cells only takes place for five minutes, because this time is long enough to capture newly synthesized RNAs, but not long enough to degrade even highly transient RNAs. Due to this narrow time frame the synthesis of most RNAs can be detected without bias.

Biochemical fraction approaches are labour intensive and expensive. Especially the magnetic streptavidin beads need to be purchased at a high price, and the need to sequence several fractions raises the cost again (Boileau et al., 2021).

1.5. SLAMseq

SLAMseq stands for Thiol (SH)-Linked Alkylation for the Metabolic sequencing of RNA. It is a method where newly synthesized and already existing RNA gets measured using metabolic labeling via s^4U. SLAMseq gives insight into RNA synthesis and degradation (Herzog, Fasching & Ameres, 2020). The goal would be to find an approach, that allows RNA labeling, without interfering with any processes in the cell, like a stress or toxin response, which would alter gene expression. A possibility for this would be metabolic RNA labeling of intact cells, with the inert ribonucleotide analog 4-thiouridine (s^4U), which uses the chemical conversion of s^4U into cytosine analogs. RNA prepared from cells with s^4U metabolic labeling can be chemically treated, alkylated, which leads to a T to C conversion in the metabolically labeled transcripts after cDNA preparation and sequencing (Herzog, Fasching & Ameres, 2020). This approach will be described in more detail in the next paragraphs.
Other protocols use oxidative-nucleophilic-aromatic substitution, where s^4U gets converted into cytidine analogues, so a U to C mutation, which can be distinguished when sequencing (Schofield et al., 2018).

SLAMseq can be used for a variety of purposes. Pulse experiments require a short pulse of s^4U, for example for 45 minutes to 1 hour, where the newly synthesized RNA gets labeled. This method has as an output rates of transcription and mRNA half-lives. To this a time course can be added, which would extend the results by dynamic changes of the transcription rate. For pulse chase experiments cells are labeled with s^4U for an increased amount of time, up to 24 hours. This is followed by the chase, which is a washing out of s^4U with unmodified uridine. It leads to the newly synthesised RNA being unlabeled and the existing one being labeled. This method shows the rate of transcription, the rate of decay and mRNA half-lives. Adding drug treatments to an experiment, makes it possible to screen for primary and secondary targets, by observing direct and indirect transcript responses on a transcriptional level. (Herzog et al., 2017)

The concept of how SLAMseq works and what is the readout, is depicted in Figure 2. In SLAMseq cultured cells are grown in 4-thiouridine to allow s^4U nucleotides to

be incorporated into newly synthesized RNA. The labeling duration usually depends on the type of experiment. Cells are sampled at a given timepoints and RNA is extracted. The isolated total RNA contains both existing (unlabeled) and newly synthesized (labeled) RNA for pulse experiments. This is in contrast with the pulse chase experiment, where the s^4U gets washed out with uridine and existing RNA would be labeled and newly synthesized RNA would be unlabeled. Total RNA is then mixed with iodoacetamide (IAA), which modifies the 4-thiol group of s^4U-containing nucleotides and reverse transcribed into cDNA. For labeled RNA transcripts, reverse transcriptase incorporates a G instead of an A at positions where reduced *s^4U-modified nucleotides are encountered. Second-strand synthesis is then performed to generate a double stranded SLAMseq library. Sequencing reads with a T to C mutation distinguish labeled from unlabeled transcripts (Herzog, Fasching & Ameres, 2020).

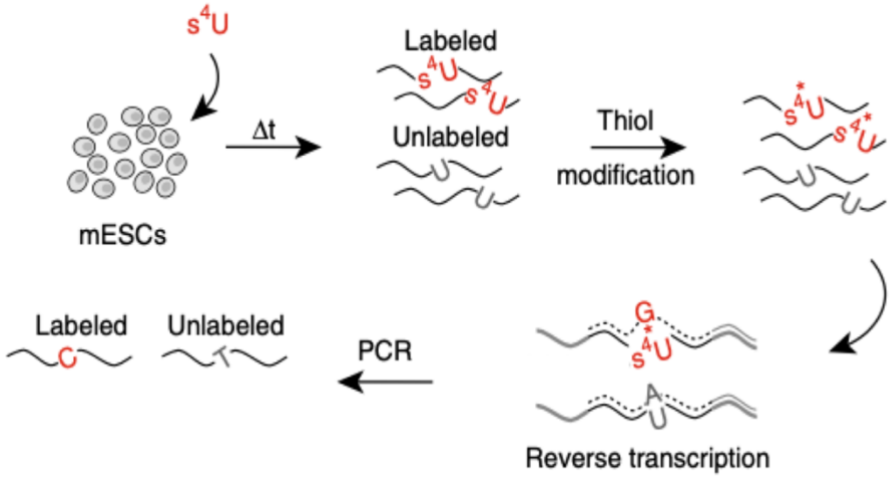


Figure 2: **Schematic overview of SLAMseq** (Herzog et al., 2017)
First, mESCs are labeled with s^4U for a certain period of time, which leads to labeled and unlabeled RNA. After a thiol modification with IAA, the labeled RNA incorporates G instead of A during reverse transcription. After PCR the labeled transcripts are recognizable due to their T to C conversion.

The SLAMseq protocol in combination with any RNA sequencing approach can provide various kinds of information, such as mRNA half-lives. The mRNA 3′ end sequencing is a useful option, because it focuses on one specific region and not the whole genome (Herzog, Fasching & Ameres, 2020). A commonly used commercial protocol is “Lexogen’s QuantSeq 3′ mRNA-Seq Library Prep Kit for Illumina”.

This approach has some clear advantages. For one it reduces the required sequencing depth substantially compared to total RNAseq, because it focuses solely on the 3′ end, therefore less reads are required to get data. It also decreases the bioinformatic workload and its complexity, because there is no need to sort through the whole genome or the normalization of reads to gene length. This targeting has an impact in reducing sequencing and analysis costs (Ma et al., 2019).

The protocol for 3′ end sequencing starts with total RNA as starting material and a following reverse transcription with oligo dT anchored priming. This means that to the oligo dT_{18} a dV (either dG, dA, or dC) and then a dN (dA, dT, dG or dC) are added. Because of this variable sequence, the primer gets anchored to the 5′ end of the Poly A tail of the mRNA, this avoids priming inside of the poly A tail or in other A-rich parts. Then there are two options to proceed. The first one is the untargeted approach, where at second strand synthesis random priming is employed. This means that a random set of hundreds of oligos, usually six nucleotides long, gets mixed into the reaction, which results in the synthesis of many random genes. This method is time and cost intensive, if you have a higher number of samples or transcripts of interest.
In comparison to that, the targeted approach uses gene specific priming, where a specific subset of genes of interest can be targeted with the according primers. The advantage here is, that the sensitivity towards lowly expressed transcripts is amplified. The targeted approach also reduces the needed sequencing depth because it focuses only on the targeted genes of interest. In Figure 3 a schematic overview of the untargeted and targeted mRNA 3′ end sequencing protocol is visualized (Herzog, Fasching & Ameres, 2020).

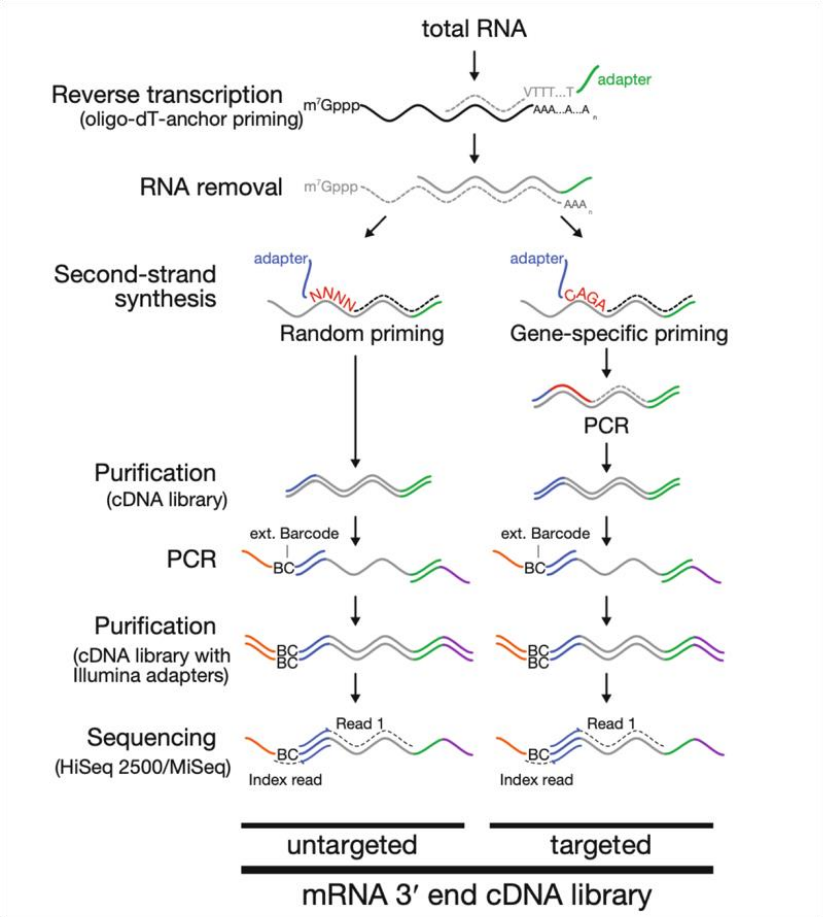


Figure 3: **Schematic overview of library preparation protocol for untargeted and targeted mRNA 3′ end sequencing** (Herzog, Fasching & Ameres, 2020)
For the mRNA 3′ end library preparation RNA gets reverse transcribed using oligo dT anchor priming. After RNA removal the preparation can be split int untargeted and targeted. The first approach uses random priming for second strand synthesis, followed by purification, PCR, another purification, and sequencing. The difference to the targeted approach is, that during second strand synthesis gene specific priming is used, where genes of interest can be targeted, with a following PCR. In the end, both approaches, result in a mRNA 3′ end cDNA library.

The targeted SLAMseq approach has, as previously discussed, many benefits, but there are also some technical limitations, which need solutions. What is important for these experiments is an even representation of transcripts for an equal sequencing depth, so that there is the same amount of information about lowly and highly expressed genes. Because of the different expression rates of potential genes of interest, achieving a balance is challenging. Another limitation of the targeted approach would be that, because of the gene specific priming and the

PCR, shown in Figure 3, the relative abundance of the transcripts is lost. At the same time, the kinetic information given by the relative abundance of T>C conversion containing (i.e newly transcribed) and non-containing (i.e. pre-existing) transcripts is maintained, reporting on possible changes in gene expression dynamics following any given perturbation.

For the optimised, multiplexed, and targeted protocol, a new oligo dT primer setup is introduced, which includes unique molecular identifiers (UMIs). In Figure 4 a scheme of a transcript is displayed, where on the left the sequencing primer and Read 1 are located. On the right there is again the sequencing primer, followed by Read 2, which includes the UMI, the gene specific sample barcode, the oligo dT and the anchor. The UMIs, consisting of 12 random nucleotides, are added during RT with the oligo dT anchored primer, and show how many different molecules were targeted, which gives insight into the diversity of the sample. With this the UMIs can be counted and back calculated to get the relative amplicon abundance.

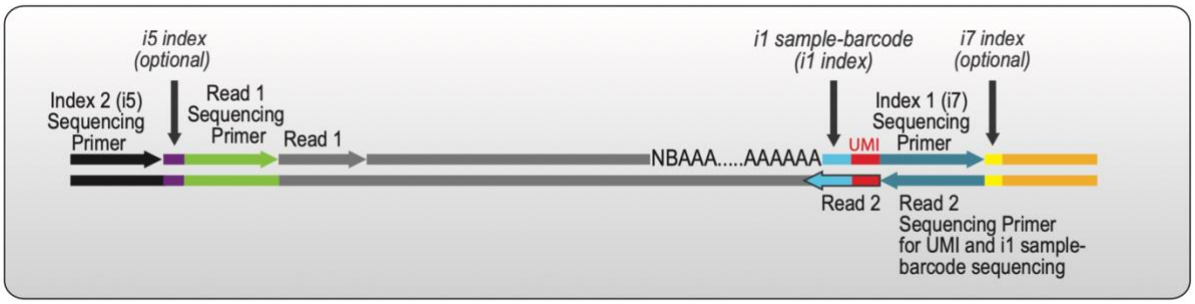


Figure 4: **Scheme of the Transcript**
On the left is the first read (Read 1), which is accompanied by the according sequencing primer and the Illumina index (i5) primer. On the right is the Illumina indexing (i7) primer, followed by the second read (Read 2). Here are also the UMI, the sample barcode and the oligo dT located.

Even though the mRNA 3′ end sequencing protocol is already practical, it is still somewhat time consuming and cost intensive when applied to massive parallel perturbation assays. The targeted approach addresses these issues to a certain extent but evaluating the reaction of genes to many different treatments is still a long and expensive process.

1.6. Aim of the Study

The goal of this thesis is to establish a time- and cost-effective protocol for the systematic analysis of cellular gene expression kinetics in scale. To achieve this, we combined all previous improvements, optimised the protocol even further and added a multiplexing step right after RT.

This means that, after subjecting cells to different treatments, metabolic RNA labelling is performed, followed by RNA extraction, and alkylation. Finally, during the RNA library preparation protocol, barcoded oligo dT primers with unique molecular identifiers (UMIs) will be employed to allow for multiplexed sequencing of a large number of samples. This multiplexing step is a crucial improvement, since it will reduce any number of samples to a single sample that can be prepared with the targeted protocol. After indexing PCR, the sample could even be pooled with more samples, and after sequencing and bioinformatic analysis, every single experimental condition and gene will still be distinguishable from each other (Figure 5).

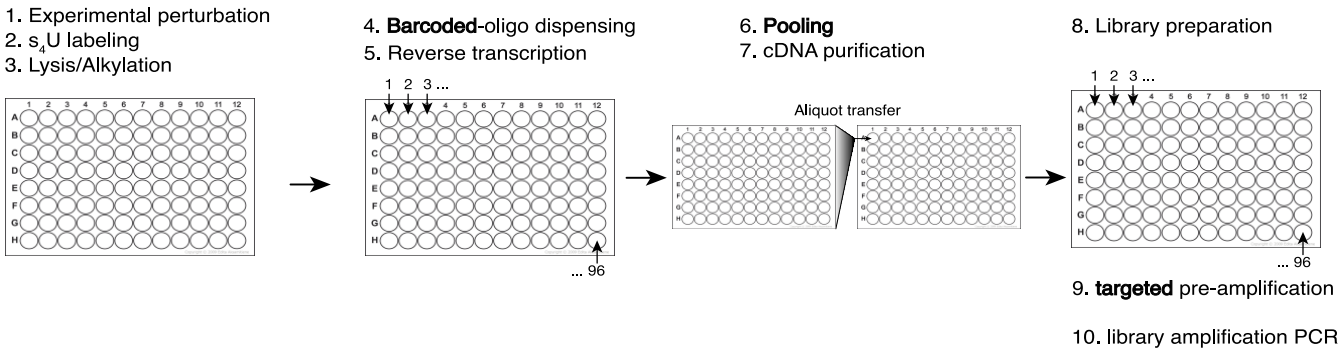


Figure 5: **Schematic overview of a multiplexed and targeted SLAMseq approach**
In a 96 well plate different experimental perturbations can be done. Afterwards s_4U labeling, lysis and alkylation take place, according to the established SLAMseq protocol. The RNA gets transferred to a new plate, where barcoded oligo dT get distributed and RT performed. Afterwards the whole plate gets transferred to one well of a new plate, this can be done 96 times, so until the new plate is full. When the pooling is done, the DNA gets purified, and the libraries prepared. Finally, there is a targeted pre-amplification and the final library amplification PCR before sequencing.

To reach the goal of creating an optimised, high throughput protocol for multiplexed targeted SLAMseq the following steps were taken:
First, we wanted to optimise the individual steps in the protocol, like RT and clean up. And we also worked towards decreasing the input amount to have a robust well working approach. Then we implemented the multiplexing step, as one of the most crucial optimisations. While working on these optimisations it became clear that the primers and their performance were one of the main issues, so we changed the protocol to a semi-nested approach to rescue the primer specificity. With this all set up, we worked towards achieving equal amplicon abundance with exhaustive PCR and primer titration.

2. Results

2.1. Reverse Transcription Optimisation

The experiments were performed with mouse embryonic stem cells (mESC), which were labelled with s^4U for one hour. For the first assessment 38 genes were picked out according to previous results. The criteria for the chosen primers were firstly the stability of the genes, with mRNA half-lives of not more than 3 hours, because if they would be very stable there would not be enough turnover and transcription to incorporate the T>C conversions. Secondly were the bioinformatic criteria with F1 > 0.5 and FN+FP < 100, a more detailed description can be found in the succeeding data analysis section. Following these decisions different reverse transcription (RT) conditions were tested (Figure 6a). Eight different combinations of temperatures and incubation times were used, where 08 was the condition of the protocol that these optimisations were built on. In Figure 6b it shows that all conditions were successful, which resulted in a time reduction from over 50 minutes to condition 03 with 15 minutes at 42°C; with no negative effect on the passed reads. These are reads that contain a useable and detectable primer, and the according RT barcode.

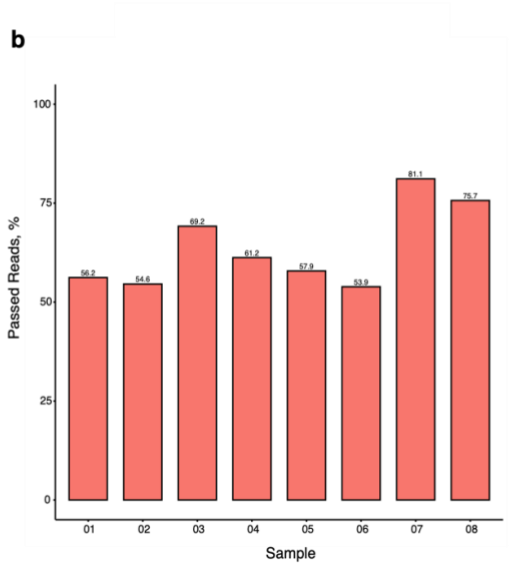
Furthermore, the addition of ExoSAP-IT was tested, which is a reagent, that enzymatically cleans up amplified PCR products, by hydrolysing excessive primers or nucleotides. In the figure it is also visible that the samples that were treated with ExoSAP-IT, sample 01, 03, 05, 07 and 08, had slightly more output, which lead to the incorporation of this reagent in all further experiments.

The targeting of 38 genes at once, proved to be problematic, as the primer performance went below 0.5 of primer ratio for most genes (Figure 6c, Supplementary Figure 1). The primer ratio visualizes the performance of each primer, and it shows the ratio of on and off target reads. The desired result would be that every primer is at 1, which would mean optimal performance.

This first set of experiments showed that the time for the RT could be reduced from almost 60 minutes to 15 minutes and still results in a functional number of sequencing reads. ExoSAP-IT was found to be a helpful reagent to clean up the sample after RT. The interplay of the primers in a high number like 38 resulted in suboptimal primer performance, which is probably due to interaction of the primers with eachother and potentially forming secondary structures and/or binding each other.

a

Sample	RT Conditions	ExoSAP -IT
01	50°C for 15 min	+
02	50°C for 15 min	-
03	42°C for 15 min	+
04	42°C for 15 min	-
05	50°C for 7 min	+
06	50°C for 7 min	-
07	50°C for 50 min	+
08	25°C for 10 min -> 50°C for 50 min	+



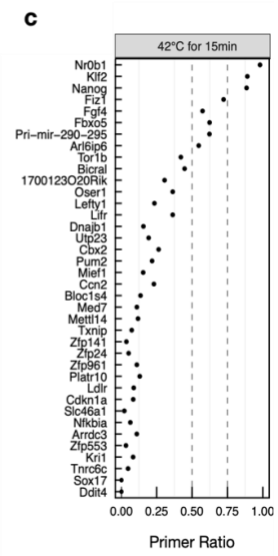


Figure 6: **Reverse Transcription Optimisation**
a) For this experiment 8 different conditions were tested, with different temperatures, time frames and with/without ExoSAP-IT. b) The passed reads contain a useable and detectable primer, and the according RT barcode. They are shown in % for each sample and are the reads that are usable for further bioinformatic analysis. c) The primer ratio is the visualization of the primer performance and shows the fraction of on and off target reads. Here sample 03 is displayed with all genes listed with their according primer ratio.

2.2. Decreasing RNA input and comparing oligo dT18 vs dT30

Due to the problematic primer performance the focus got shifted to optimising the different steps in the protocol first and then specifically working on the primer optimisation. For that reason, Nr0b1 was picked as the control gene, for its high abundance and its performance in the previous experiment. It was tested if decreasing the RNA input is possible and if the length of the T-stretch influences the samples.

Increasing of the T stretch from dT_{18} to dT_{30} is used, for example, by single cell RNA sequencing experiments, because they have very little input material, and a longer T stretch enhances the possibility of a successful aligning with the mRNA (Huang et al., 2018). In Figure 7a it is shown that the RNA input amount does not negatively influence the passed reads and therefore can be reduced. The samples with 50ng showed the best results and were then used for all further experiments. The T_{18} samples performed as expected with no negative influence for any RNA input amount. The T_{30} samples show a drop at 5ng to only 56% passed reads, which is around 25% less than 500ng and 50ng. Nr0b1 had in this experiment an optimal primer ratio of 1 (Figure 7b) and the fraction of T>C reads (Figure 7c) remained the same between different conditions. The reads that are unique have only one distinctive UMI, so the closer the sample is to 1, the more unique it is. The T_{18} samples show at 500ng and 50ng a high deduplication ratio close to 1, whereas 5ng is less unique with only 0.5. The T_{30} samples are regardless of input amount of a similar uniqueness (Figure 7d). The 50ng T_{18} sample had over 80% passed reads and close to 1 as deduplication ratio, so a reduction from 500ng to 50ng was again confirmed, and the standard T_{18} was maintained.

While using 38 genes would serve the purpose to establish a time- and cost-effective protocol for the systematic analysis of cellular gene expression kinetics, due to the unpredictability of the primer mixture picking one gene with high abundance and strong previous performance as a control served these experiments well. The primer performance and T>C conversion rate were as expected, which means the other results are reliable and the decrease of the RNA input amount from 500ng to 50ng, while keeping T_{18}, was reasonable.

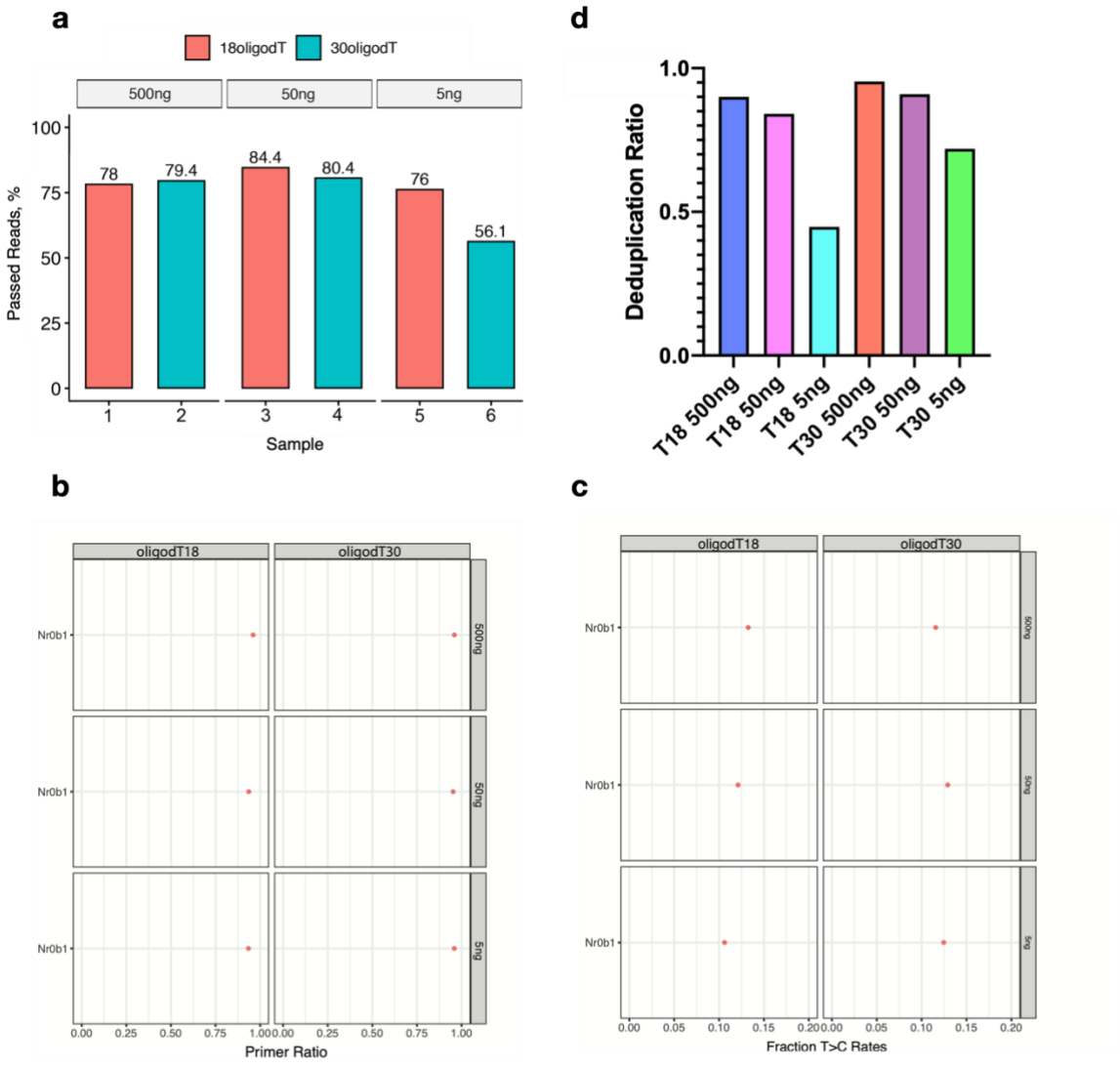


Figure 7: ***Decreasing the RNA input amount***

a) For this experiment three different RNA input amount were used (500ng, 50ng, 5ng) and oligo dT18 compared with oligo dT_{30}. The passed reads contain a useable and detectable primer, and the according RT barcode. They are shown in % for each sample. b) The primer ratio is the visualization of the primer performance and shows the fraction of on and off target reads. Here Nr0b1 was used as a control gene. c) The T>C conversion rate of Nr0b1 over all samples, shows, simplified, how many Ts to Cs were changed. d) The deduplication ratio is where the fraction of deduplicated reads, so unique ones, are compared to all reads using UMIs. The more unique a sample is the less UMIs it has and the closer to 1 it is.

2.3. Multiplexing with Individual Barcodes

One of the most important steps in optimising this protocol is the multiplexing after reverse transcription, where many samples become one. Eight samples, with an individual barcode each, were equally mixed after RT to test if pooling them would cause issues, that could be revealed during bioinformatics analysis. Nr0b1 was used again as the control gene to keep the obtained results reliable. The input RNA amount was for all eight samples the same and therefore the expectation would be, that the passed reads should also be the same. In Figure 8a we see the number of reads split for each of the 8 barcodes with no discrepancy between samples, so every individual barcode performed equally well. So, during the bioinformatic analysis it was possible to extract the samples due to their barcodes, which confirms a successful multiplexing. The same goes for the primer performance, where Nr0b1, as expected, did not fluctuate dependent on which barcode was used (Figure 8b). In Figure 8c it is visible that the UMIs behave the same among each barcode with a deduplication ratio equally close to 1.

Multiplexing samples did not negatively influence the number of passed reads, the primer performance, or the deduplication ratio, so this crucial step in the protocol was successful. This also means that after this step the basic optimisations of the protocol were effectively done, and with the following experiments the problem of the primer performance with multiple genes was confronted.

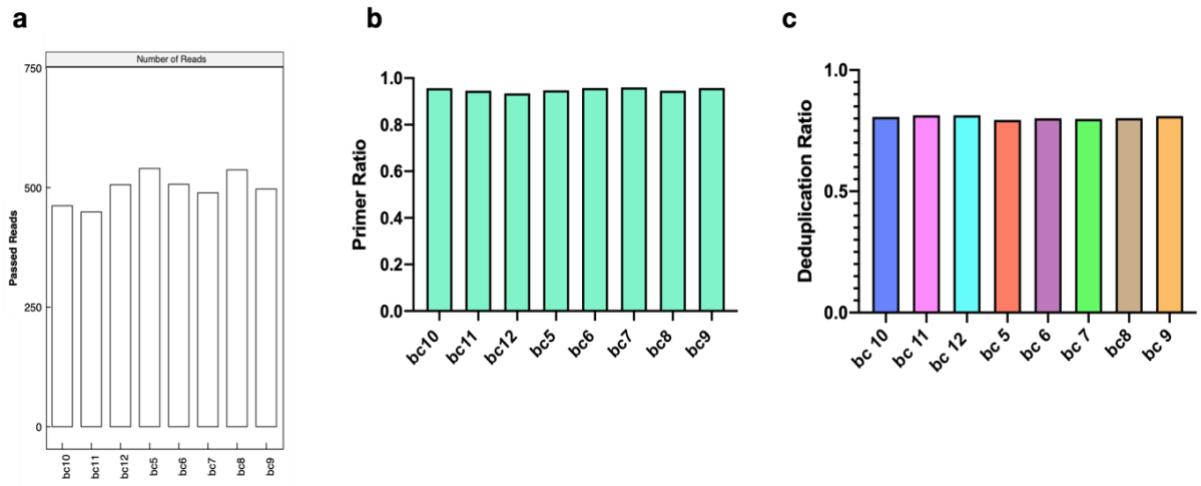


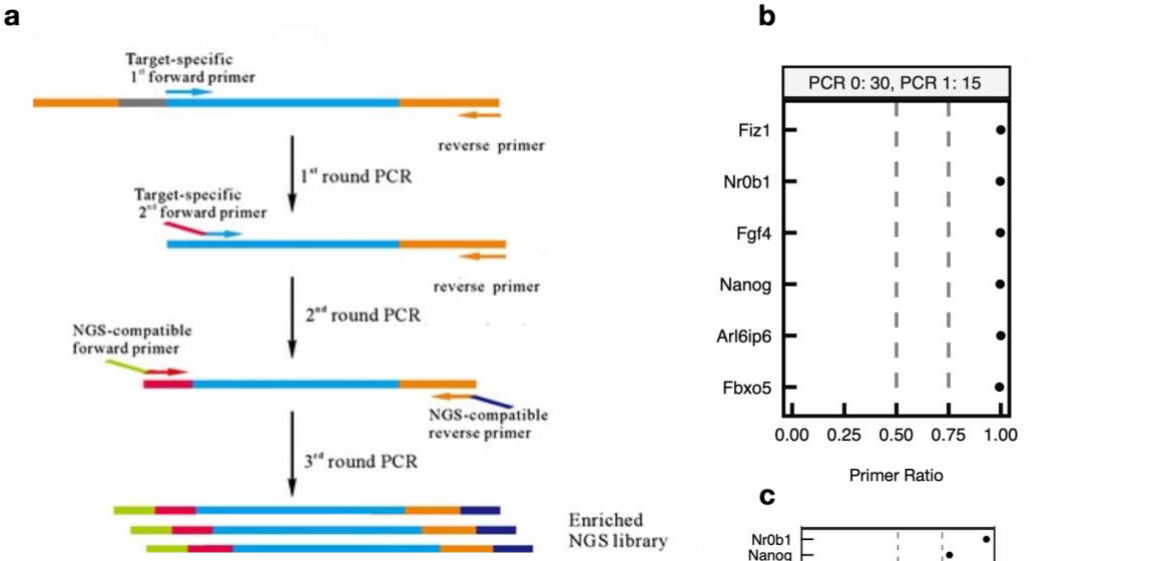
Figure 8: **Multiplexing with Individual Barcodes**
a) Eight samples, with individual barcodes, were multiplexed after RT. The number of passed reads are shown for each barcode. b) The primer ratio is the visualization of the primer performance and shows the fraction of on and off target reads. Nr0b1 was used as the control gene and every barcode is individually depicted. c) The deduplication ratio is where the fraction of deduplicated reads, so unique ones, are compared to all reads, here shown for each individual barcode.

Another condition that was tested, was trying different annealing temperatures during second strand synthesis. The reason behind this was, that different primers have different optimal annealing temperatures, so this was to test if changing the temperature could positively influence the primer performance. It is shown in Supplementary Figure 2, that changing the temperature did not influence the library preparation.

2.4. Implementation of the Semi-nested approach

After successfully optimising and validating the multiplexing of the protocol, the next goal was to improve primer performance. For this, six genes, with the best primer performance were picked out according to criteria, F1 > 0.5 and TP > 50, from previous experiments. To achieve enhanced primer specificity a semi-nested approach was implemented. In Figure 9a a scheme of the setup is visible, here the first PCR is performed with external gene specific forward primers and the standard reverse primer (PCR0). Then another round of PCR with internal gene specific forward primers takes place (PCR1). With this approach the target gene gets amplified with the external primer and then gets again amplified within that transcript with the internal primer, which should increase the overall specificity. Figure 9d lists the 12 samples with their respective PCR0 and PCR1 cycle length. In Figure 9b (Supplementary Figure 3a) we see the primer ratios and as anticipated, primer specificity gets rescued by the semi-nested approach for all conditions. In comparison to Figure 9c, where the same genes were singled out of a previous experiment, which did not apply a semi-nested approach, the difference in primer performance becomes apparent. As mentioned before, we wanted to achieve equal abundance for each amplicon to minimise sequencing depth; in this experiment it became clear that even using longer PCR cycles to exhaust the primers, it did not achieve that (Figure 9e). The fraction T>C rates stayed the same between conditions, which aligns with our expectations (Supplementary Figure 3b). When looking at Supplementary Figure 3c, it is visible that prolonging both PCR steps leads to an anticipated number of unique UMI reads, so although primer specificity is rescued this has no influence on the UMIs.

Implementing the semi-nested approach rescued primer specificity, which should later allow the use of a higher number of primers to be multiplexed and therefore more genes per condition to be analysed. This problem solved, the next issue we faced was the need to balance out amplicon abundance. Otherwise highly abundant genes would result in many reads and lots of information, but data on lesser abundant genes would be very little or be lost altogether.



d

Sample	PCR 0	PCR 1
1	10	8
2	10	13
3	10	20
4	15	8
5	15	13
6	15	18
7	20	8
8	20	13
9	20	15
10	30	8
11	30	13
12	30	15

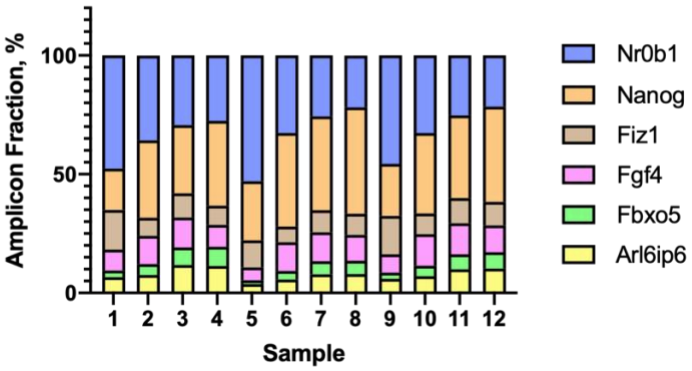


Figure 9: **Implementation of the semi-nested approach**
a) Scheme for the implantation of the semi-nested approach (Chi et al., 2014); three rounds of PCR, with external (PCR0), internal (PCR1) and NGS primers b) The primer ratio is the visualization for the primer performance and shows the fraction of on and off target reads. Here the primer ratio of condition 12 is visible including the 6 used genes. c) This is a plot from previous results, which shows the primer performance of the same 6 genes but prepared with the prior protocol and not the semi-nested approach. d) Table with summary of all samples and their corresponding PCR 0 and PCR 1 cycles. e) Amplicon fraction shows how the 6 genes are present in all 12 samples in relation to each other (in %).

2.5. Exhaustive PCR

Having implemented the semi-nested approach, which rescued primer specificity, the following step was to equalise amplicon abundance. For this an experiment was performed, where PCR1 was run for an extensive number of cycles to exhaust the primers of highly abundant genes fast, and lowly abundant ones would have more time to also use up all their available primers (Supplementary Figure 4a). Achieving equal amplicon abundance would mean that each gene gets an equal number of reads, therefore minimising the required sequencing depth to obtain enough reads to robustly quantify T>C conversion rates for lowly abundant genes. The drawback here is the loss of the information about the relative abundance of each gene in the sample. For this the UMIs were implemented. If there were many duplicates of identical UMIs the same amplicon got amplified repeatedly. But if there were many unique UMIs per transcript, we knew there is a high diversity of amplicons. In Figure 10a we see that increasing the cycle number influenced the uniqueness of UMI reads negatively. This was unfortunate, but not wholly unexpected, because the more cycles the more opportunity for duplicates to amplify. Figure 10b shows that the exhaustive approach did not equalise the amplicon abundance. Although, the sample with PCR1 at 40 cycles did increase Fbxo5's and decreased Nr0b1's fraction, which would be desired. Therefore PCR 1 cycles 15 and 40 were both used for the following experiment. The fraction T>C reads were not influenced by this approach, which proves that these are reliable results (Supplementary Figure 4b).

Using an exhaustive PCR to equalise amplicon abundance would have been the preferred and simple solution. PCR amplification bears multiple risks and for an extensive number of cycles is also unpredictable regarding the primers, what really gets amplified, and the uniqueness of UMIs. The other option that we then explored, was to titrate primer concentration.

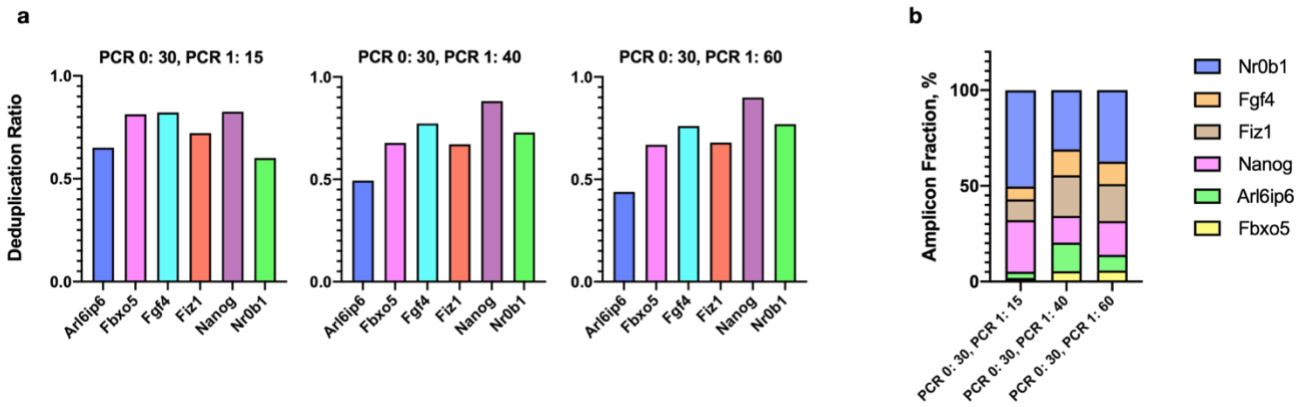


Figure 10: **Exhaustive PCR**
a) The deduplication ratio is where the fraction of deduplicated reads, so unique ones, are compared to all reads. Here shown for three conditions (PCR1 at 15, 40 and 60 cycles) and for all six genes each. b) The amplicon fraction shows how the 6 genes are present in the samples in relation to each other (in %).

2.6. Titration of Primer concentration

Our next approach to equalise amplicon abundance was to titrate the primer concentration. For this the information about amplicon abundance was taken from the previous experiment and primer amounts were rescaled in an inverse ratio for Mix 1. This was to test if amplicon abundance from a 1:1 primer mix could be used for rescaling. Unfortunately, this approach was not successful, therefore the other Mixes were made with arbitrary ratios to get to equalised amplicon abundance with empirical tests (Table 6). As mentioned before for PCR1 the standard amount of 15 cycles and the increased 40 cycles were tested. Sample number six, which is Mix 2 for 30 cycles at PCR0 and 40 cycles at PCR1, showed the most equal amplicon fractions (Figure 11a). This approach showed the way in the right direction, where the changing of primer concentrations and prolonging PCR cycles leads to more equalised amplicon abundance. The fraction of T>C reads were not influenced by this experiment, which demonstrates the reliability of these results (Figure 11b).

To figure out the right primer concentration for each amplicon through empirical experiments seems to be possible, but the positive influence of increased PCR1cycles to exhaust the primers during amplification should not go unnoticed.

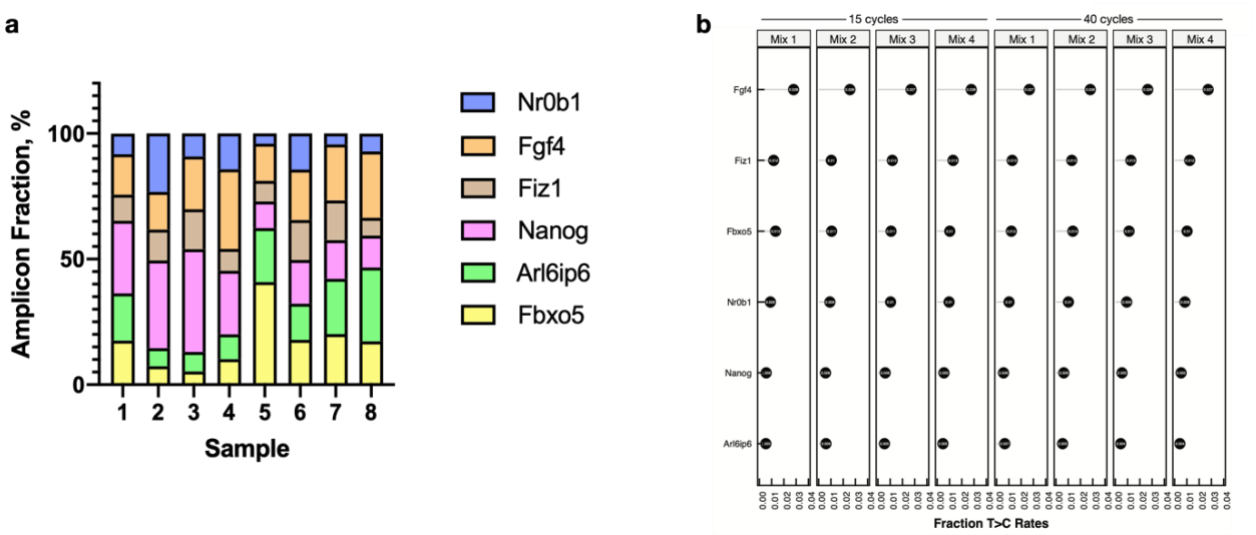


Figure *11*: **Titration of Primer concentration**
a) The amplicon fraction shows how the six genes are present the samples in relation to each other (in %). Sample 1 to 4 ran for 15 cycles at PCR1 and consisted of Mix 1 to 4. Samples 5 to 8 ran for 40 cycles at PCR 1 with mixes 1 to for respectively. b) The T>C conversion rate of the six genes over all samples shows how many Ts to Cs were changed. Samples were run for PCR0 30 cycles and for PCR1 15 or 40 cycles, with four different primer concentration mixes.

2.7. Alkylation and Reverse Transcription in-Lysate

So far, the protocol relied on RNA extraction and alkylation followed by RNA clean-up and reverse transcription. We decided to try and avoid laborious RNA extraction and clean-up by performing IAA treatment and reverse transcription directly in the lysed cells. To put this into perspective, every step of the in vitro protocol was its own reaction that needed to be set up. Compared to the in-lysate, where we would work with the cells, as the name suggests, in lysate, which would get added directly to the alkylation reaction, and to this the RT reagents; therefore, multiple steps and clean ups would become one step, until cDNA is created. Another impactful optimisation would be to alkylate, and reverse transcribe immediately in-lysate, so without RNA extraction. Therefore, a protocol was developed were after cell lysis, immediately IAA alkylation, RT and following the standard protocol would be started (Figure 12a).

One point that needed addressing beforehand was the alkylation step. Routinely DMSO was involved in this reaction, but because in this in-lysate protocol there are no in between clean ups DMSO could negatively affect the RT reaction. As Yasukawa, Konishi & Inouye (2010) described, DMSO in smaller quantities had a positive effect on the RT, but larger quantities inhibited the reaction. In the early protocol development, there would have been 50% of DMSO in the reaction, whereas the paper suggests a maximum of 24%. To circumvent this, we decided to substitute DMSO with DMF and checked that in vitro IAA treatment reaction is unaffected (Figure 12b). All alkylated samples show a similar peak at 297nm, whereas the untreated sample peaks at 333nm, which is expected. With this the safe use of DMF with s^4U in the alkylation reaction was established.

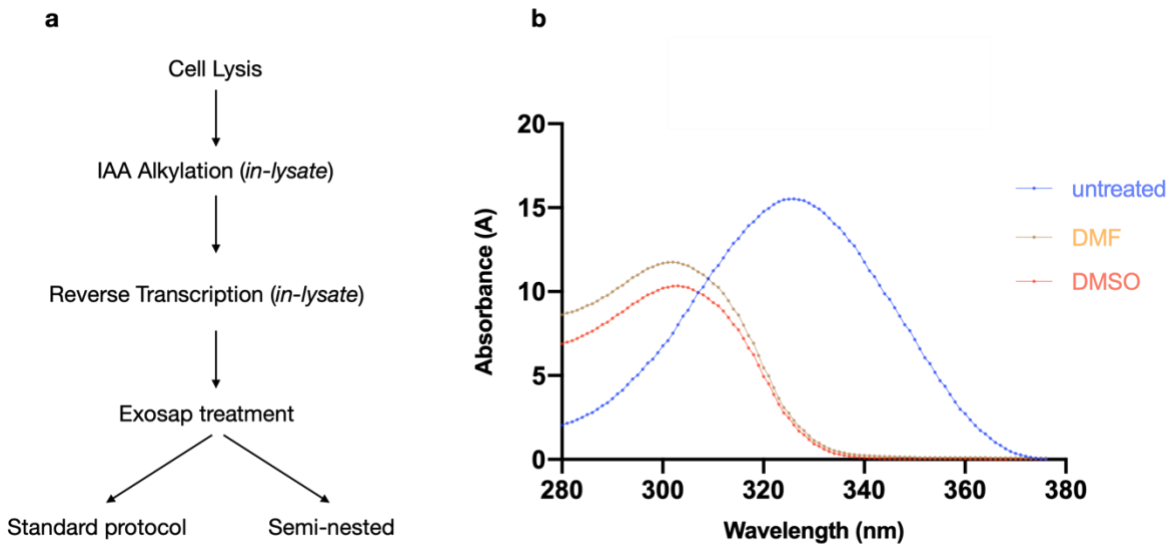


Figure 12: **In-lysate protocol and in vitro alkylation experiment**
a) For these experiments a new protocol was developed to alkylate, and reverse transcribe in the cell lysate. Here is the schematic protocol for this in-lysate approach, visualizing the different steps. b) To test if DMF could replace DMSO an in vitro alkylation, 15 minutes at 50°C, with s^4U was performed. Untreated (blue), DMF (orange) and DMSO (red) visualized with absorbance compared to wavelength. The untreated s^4U should be at around 335nm and the alkylated $*s^4U$ at 297nm.

Next, we viewed the sequencing results of the in-lysate protocol compared to the standard in vitro and the in-lysate semi-nested protocol. In Figure 13a it is shown that the semi-nested approach did not influence the passed reads negatively (Sample 9-12). The primer ratios for the in vitro and in-lysate protocol showed similar patterns to previous results (Figure 13b). The semi-nested in-lysate approach rescued the primer specificity, although two of the lowest abundant genes, Fbox5 and Arl6ip6, did not produce enough reads for useful data. The fraction T>C rates between in vitro and in-lysate approach were similar, which proved the reliability of the results of the new protocol (Figure 13c). Since it was a pilot experiment, we didn't aim to equalise amplicon abundance, which is reflected in Supplementary Figure 5a. Getting useful information from the deduplication ratios (Supplementary Figure 5b) proved to be difficult, because of the low number of reads. This means that for some conditions there is data that allows interpretable results, but some samples had so little reads that if two useful numbers were available this would be the result, which cannot be reliable. The semi-nested approach seemed to have lots of duplications, although in every other aspect it was the best performing. The general variability between the samples of the in-lysate protocol could be explained due to cells as input material instead of a precise nanogram RNA input. In comparison, the standard in vitro approach, including samples 5, 6, 7 and 8, behaved as expected, especially the sample with oligo dT_{18} and the reverse transcriptase SuperScriptIII performed as well as in previous experiments.

To conclude, an in-lysate protocol was created and alkylation with DMF as a substitute for DMSO was successfully established. Looking at the passed reads, primer ratio and fraction T>C rates, the semi-nested in-lysate approach proved to give optimal and reliable results.

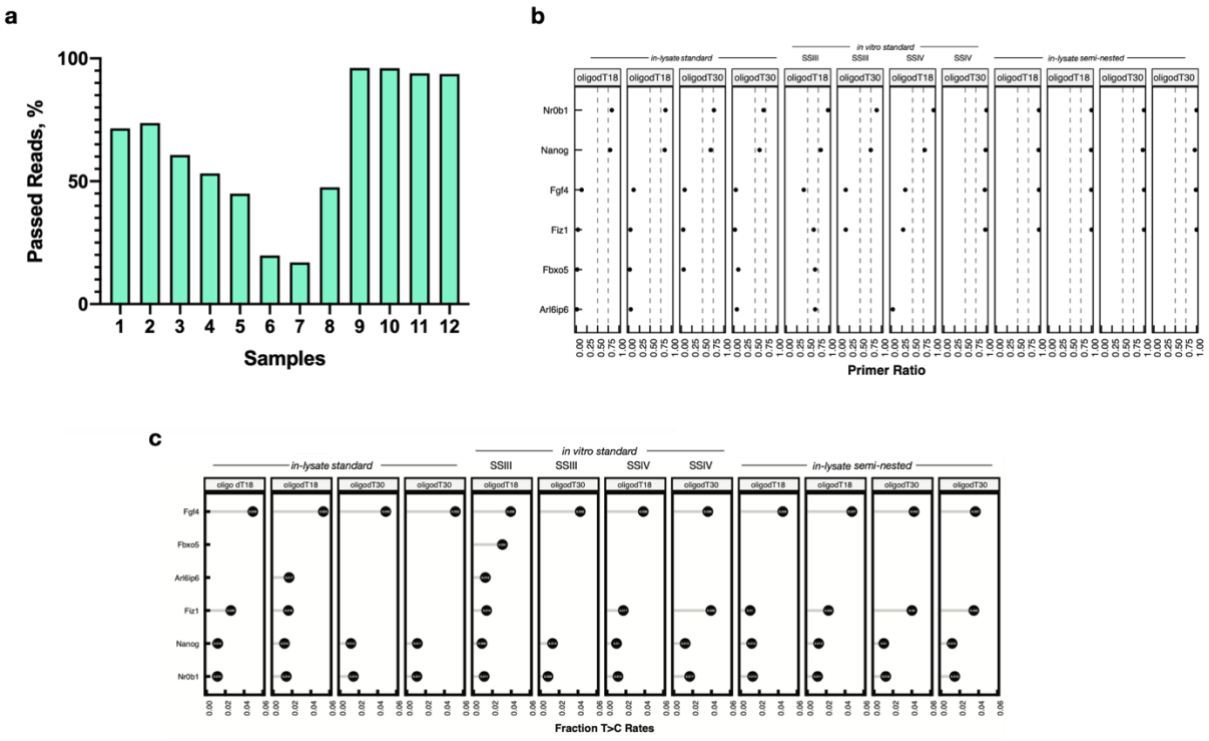


Figure 13: **Alkylation and Reverse Transcription in-lysate**
a) Passed Reads for each sample in %. Sample 1 and 2 are oligo dT_{18} in vivo. Sample 3 and 4 are dT_{30} in vivo. Sample 5 is dT_{18} and SS III in vitro. Sample 6 dT_{30} and SS III. Sample 7 and 8 are SS IV in vitro with dT_{18} and dT_{30}. Sample 9 and 10 are dT_{18} in vivo semi-nested. Sample 11 and 12 the same but dT_{30}. b) The primer ratio is the visualization of the primer performance and shows the fraction of on and off target reads. Here are all samples and with 6 Genes each visible. c) he T>C conversion rate of the six genes over all samples shows how many Ts to Cs were changed. Here, results for all twelve samples are shown.

3. Discussion

The goal of this work was to establish an optimised, hight throughput protocol for targeted multiplexed SLAMseq. To this end, individual steps of the protocol were optimised: the multiplexing with individual barcodes for each sample was established, primer specificity improved, and amplicon abundance balanced.

The optimisation of the reverse transcription, with improved time span and temperature, was the first step. At this point it also became apparent that the multiplexing of many primers would become one of the main problems, because their performance became unpredictable. Even when using primer multiplexing tools (Kechin et al., 2020) the outcome was not as desired (Supplementary Figure 1). The semi-nested approach successfully rescued the primer performance, although this was only tested for a low number of primers (Figure 9). Future experiments with large primer pools should be tested and compared between the standard and semi-nested protocol.

The RNA input amount was successfully reduced from 500ng to 50ng and there could even be follow-up experiments to further reduce the amount. The length of the T stretch did not significantly influence the sequencing results, so the standard, shorter dT_{18} was used. The crucial part of multiplexing after reverse transcription did not run into any issues and could therefore be confidently implemented.

As mentioned, the semi-nested approach rescued primer specificity, which allowed the focus to shift to the equalization of amplicon abundance.
Exhaustion of the primers during PCR alone did not work as hoped (Figure 10), but with titrating the primers according to previous experiments highlighted the right path (Figure 11). Here more elaborate experiments would be needed to get information of all desired primers to then calculate the according ratios, because here it was only done for 6 genes. It generally seemed like, that there is no way to predict primer performance and amplicon abundance beforehand, so the only way is to empirically optimize with the data of multiple sequencing rounds.

An in-lysate protocol was created, for which the use of DMF in the alkylation was effectively tested. As a pilot experiment this gave some interesting insights and proved that an in-lysate approach is possible (Figure 12). Details would need to be worked out more, but the semi-nested in-lysate approach gives hope for an even higher throughput method.

In this protocol a 3′ end sequencing approach was put to use, although with this the general information about the whole genome is lost. Here it is a strength not a weakness, because it reduces the needed sequencing depth, which makes the sequencing costs drop. Furthermore, this approach simplifies the bioinformatic process, because there is no need to normalize the reads to the gene length. This again, saves time and money. Compared to 3′ end sequencing, whole transcript RNAseq distributes reads across the entire length of transcripts with only a low coverage of the 3′ and 5′ end (Ma et al. 2019). 3′ end sequencing does not have this range, but because of the focus on the 3′ end and the minimised sequencing depth, less reads per transcript are required, which in turn allows for a higher degree of multiplexing. This is also the reason why the implementation of the targeted approach, with multiple gene targets, is possible. In this thesis the exhaustive and primer titration approach were used to equalize abundance and therefore get more reads and information also on lesser expressed genes.

A possible expansion of multiplexed targeted 3′ end sequencing would be the application for non-polyadenylated transcripts. The idea would be to first extract RNA, perform in vitro polyadenylation and then follow with the, in this thesis, established protocol, with a gene specific primer against non-coding RNAs. Wu, Schmid & Jensen (2021) describe a 3′ end sequencing method with in vitro polyadenylation by using *E. coli* poly-A polymerase (EPAP) to add an A tail to all RNA 3′ OH ends. Non-coding RNAs without poly-A tails would be for example rRNAs, miRNAs and tRNAs. The in vitro polyadenylation is necessary, because in this protocol at RT an oligo dT primer is used, which relies on the poly-A tail. The EPAP adds adenylated residues to the 3′ends of RNAs using ATP (Yehudai-Resheff & Schuster, 2000).

The amplicon abundance is a complicated topic, which created the issue, that the true steady state abundance is lost due to PCR amplifications. Knowledge of transcript levels and their dynamic changes is necessary to understand transcription, decay, and half-lives of the amplicons. To circumvent this limitation UMIs were introduced to the oligo dT primers. These unique identifiers, consistent of 12 random nucleotides, show how many different molecules were targeted, which gives insight into the diversity of the sample. With this the UMIs can be counted and back calculated to get the relative amplicon abundance. Low abundant amplicons were expected to show less diverse UMIs, which could not be confirmed, see Supplementary Figure 3.
Amplicons, that are more abundant seem to have more duplications than less abundant ones, although that could be due to them having more reads. If a low abundant gene has only two reads that pass the quality control, there cannot be a high discrepancy, so they seem more unique, which is not possible to confirm.

An alternative to the time-consuming RNA extraction with TRIzol would be the King Fisher (KF) Purification System by ThermoFisher, which is available at the MBS facility at the Biocenter. This is a promising possibility to optimise and streamline the RNA extraction. The results shown in Figure 14 were accomplished by a colleague from the Ameres lab, Tsimafei, and demonstrate that the RNA extraction by KF were equally reproducible than the standard extraction with TRIzol. The comparison was done for 1h and 500µM s^4U labelled samples. The T>C conversion rate, obtained after sequencing and analysis, is strongly comparable. Also, in Figure 14d the half-lives in relation to the T>C conversion of KF and TRIzol extraction show a clear negative correlation. When the T>C conversion rate increases, the half-lifes decrease, this means that if there is a fast turnover, the half-life of that gene is shorter.

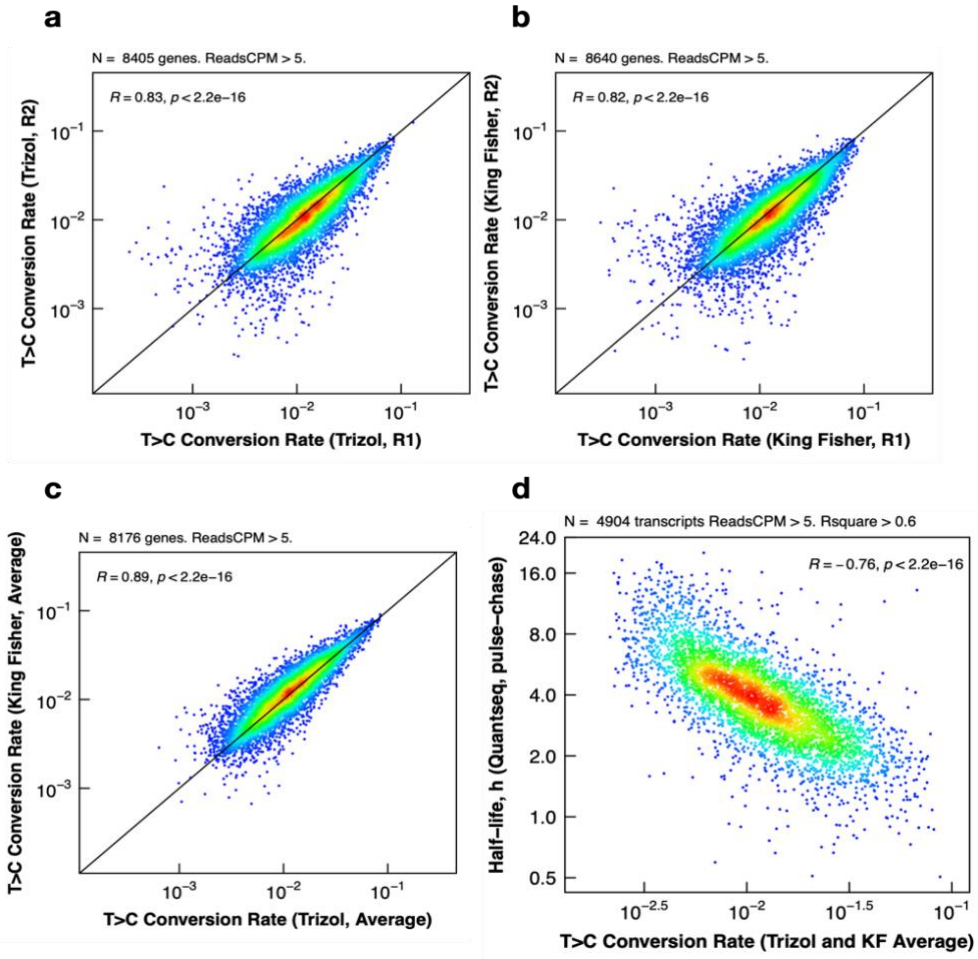


Figure14: **TRIzol extraction compared to King Fisher**

a) To reduce the time for RNA extraction with TRIzol the King Fisher Purification System was tested. a) Here the TRIzol Spearman correlation of T>C conversion rates compared with duplicates are visible. b) These are the King Fisher Spearman correlation of T>C conversion rates in duplicates. c) Here the Spearman correlation of T>C conversion rates of the King Fisher are compared to the TRIzol ones, in average. d) Lastly the Spearman correlation of half-lifes are compared to the T>C conversion rates for King Fisher compared to TRIzol respectively (Average).

In the paper by Ye et al. (2018), they described a similar multiplexing and RT in-lysate approach. This paper was a great source of information, for the preliminary in-lysate protocol set up and confirmed beforehand that these optimizations are possible. They also implemented UMIs to monitor PCR amplification artifacts, used a 3′ end sequencing approach and an early pooling step. In this thesis their concept was proven, but also further steps were taken. On one hand by concentrating on targeted sequencing, compared to the untargeted approach they used, and on the other hand by focusing on metabolic labeling with SLAMseq, to get insight into gene expression dynamics, optimizing multiple other steps in the protocol and decreasing the needed RNA input material.

With all different optimisations and the protocol being streamlined, experiments like compound screening for transcriptional effects in large scales would be possible. If more work gets put into in-lysate RT and alkylation tests, this path could also lead to a strong approach.

Another point to investigate would be to implement robotics. The Opentrons system has already been used for small pools of primers and is also in duty for the NGS facility to multiplex large requests, after QC, for sequencing. This is already helpful but not the large scale that could be possible. For example, bigger, more advanced robots, could be the “Biomek i7” from Beckmann Coulter, are able to do full library preparations, so from RNA to ready libraries, except for the final PCR amplification. In a hypothetical large-scale experiment, where 300 different compounds should be tested in a total of four plates, where each sample is distinguishable through its individual barcode, each plate gets pooled, and these pools are unique due to their Illumina adapters as well. So, an experiment of 4 plates, over 300 samples, could be downscaled to one sequencing sample and all these steps are possible to be done by a robotic system. This would free up an enormous amount of time for the scientists to plan new experiment or analyse data from previous ones (Beckmann Coulter Inc. “https://www.beckman.at/liquid-handlers/biomek-i7”).

If the goal would be to also automate the in-lysate protocol, a different robotic system like the “Echo Liquid Handler”, from Beckmann Coulter, could be the solution. This is a very precise and contamination safe technology, which would be able to work with cells directly. It employs an acoustic droplet ejection (ADE) technology, where sound waves are used to discharge droplets from the starting plate to the destination plate. Because this robot works in droplets, small volumes are possible and this “no touch approach” makes it very safe for cross contamination of cells. The Echo Liquid Handler is also able to work in a 384 well format, which allows experiments in larger scales (Beckmann Coulter Inc. “https://www.beckman.at/liquid-handlers/echo-525” and Ye et al., 2018).

Another interesting topic is single cell sequencing and all its possible approaches and information that it makes available. The paper by Erhard et al. (2019) showed, that SLAMseq can be combined with single-cell analysis and sequencing, to gather information about transcriptional activity for several genes per single cell. This interesting approach would open new possibilities of combinations and improvements. An example would be robotic systems, that can work with cells, which is especially easy when they are resuspended in lysis buffer, like the Echo Liquid Handler, who works with single drops. An automation, maybe even with a direct lysis and alkylation in-lysate would make a noteworthy experiment and the protocol even more high throughput.

With an already optimised and streamlined protocol, these atomisation’s would increase the possibility for even more high throughput experiments. Especially in the search for useful drugs and their influence on the cells a fast, cost efficient and reliable method is crucial. With the results provided and in addition with the discussion and outlook a useful method to further science was implemented.

4. Material and Methods

4.1. Cell Line:

Mouse embryonic stem cells (mESC), clone AN3-12, derived from C57BL/6x129 F1 females, were obtained from IMBA Haplobank (Elling et al., 2017)

4.2. Reagents and Buffer:

Media:

15 % FBS (Gibco), 1x Penicillin-Streptomycin solution (100 U/ml Penicillin, 0.1 mg/ml Streptomycin, Sigma), 2 mM L-Glutamine (Sigma), 1x MEM Non- essential Amino Acid solution (Sigma), 1 mM Sodium Pyruvate (Sigma), 50 µM 2-Mercaptoethanol (Gibco) and 20 ng/ml LIF (in-house produced)

SLAMseq:

4-Thiouridine (Sigma-Aldrich)

1x sterile Phosphate Buffer Saline (PBS)

TRIzol Reagent (Ambion, Life Technologies)

RNA Isolation:

Glycogen (20mg/ml) (Roche)

Chloroform:Isoamyl alcohol 24:1 (AppliChem)

2-Propanol

DTT (1M) (Merck4- Biosciences)

EtOH

H_2O

IAA Treatment:

IAA in 100% EtOH (100mM) (SIGMA)

3 M NaOAc (pH 5.2)

5 M $NaPO_4$

DMSO (SIGMA)

DMF (SIGMA)

TURBO DNA-free Kit (Thermo Fisher)

RNA library preparation:
RNAseOUT Recombinant Ribonuclease Inhibitor
SuperSript III Reverse Transcriptase Kit (ThermoFisher)
SuperSript IV Reverse Transcriptase Kit (ThermoFisher)
10mM dNTPs
MBS Ampure Beads (Beckmann Coulter)
KAPA Hifi Hotstart Readymix
EvaGreen 20x
ExoSAP-IT (ThermoFisher)

Taq MM:
20 mM Tris pH 8.3, 100 mM KCl, 5 mM $MgCl_2$, 0.5 mM each dNTP, 1x Evagreen, 40 µL/ml TAQ (in house produced TAQ from the molecular biology service department)

In-Lysate Protocol:
Lysis Buffer (LB):
0.2% TritonX, 2U/µl RNase Inhibitor, H2O

4.3. General Equipment:

Qubit
Qubit 1x dsDNA HS Assay Kit
Nanodrop
Real-time PCR detection system
Thermal cycler
Centrifuge
Opentrons
Fragment Analyzer
Agligent DNF-474 High Sensitivity NGS Fragment Kit
MiSeq Illumina Machine and reagents (Paired End (PE) 150 cycles Nano)
1M NaOH

200pM PhiX

4.4. Oligos:

Name	Sequence (5′ – 3′)
Bc 1	GTTCAGACGTGTGCTCTTCCGATCTHBNVDNHBNVDNCGGGAAC CCGCATTTTTTTTTTTTTTTTTTVN
Bc 2	GTTCAGACGTGTGCTCTTCCGATCTHBNVDNHBNVDNAAACGTTC ATCCTTTTTTTTTTTTTTTTTVN
Bc 3	GTTCAGACGTGTGCTCTTCCGATCTHBNVDNHBNVDNTTGTCCGA TATGTTTTTTTTTTTTTTTTTVN
Bc 4	GTTCAGACGTGTGCTCTTCCGATCTHBNVDNHBNVDNGTTTAAAG GCAGTTTTTTTTTTTTTTTTTVN
Bc 5	GTTCAGACGTGTGCTCTTCCGATCTHBNVDNHBNVDNCCAAAGAG GGATTTTTTTTTTTTTTTTTTVN
Bc 6	GTTCAGACGTGTGCTCTTCCGATCTHBNVDNHBNVDNTCCTCTCT TCTATTTTTTTTTTTTTTTTTVN
Bc 7	GTTCAGACGTGTGCTCTTCCGATCTHBNVDNHBNVDNGAAGGGT AAAGCTTTTTTTTTTTTTTTTTVN
Bc 8	GTTCAGACGTGTGCTCTTCCGATCTHBNVDNHBNVDNAGTCTCAG CAAATTTTTTTTTTTTTTTTTVN
Bc 10	GTTCAGACGTGTGCTCTTCCGATCTHBNVDNHBNVDNAACCCTGG GAAGTTTTTTTTTTTTTTTTTVN
Bc 11	GTTCAGACGTGTGCTCTTCCGATCTHBNVDNHBNVDNAGGTGGTT CTACTTTTTTTTTTTTTTTTTVN
Bc 12	GTTCAGACGTGTGCTCTTCCGATCTHBNVDNHBNVDNTACGCCCA CGTGTTTTTTTTTTTTTTTTTVN
Bc 89	GTTCAGACGTGTGCTCTTCCGATCTHBNVDNHBNVDNCGGGTAG GGCGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN

Table 1: **Barcoded oligo dT_{18} and dT_{30} (2.5µM)**
Barcodes 1-12 are dT_{18} and Barcode 89 is dT_{30}

Gene Name	Binding Sequence (5′ - 3′)
1700123O20Rik	GGTATGTCACCTGGACTCTCT
Arl6ip6	ATCTCGAGTATTTAGTTGCCAAAG
Arrdc3	GCCACAGTCATGCTCTAGTATT
Bicral	CAGTGTGGCGTGGCTTTATTT
Bloc1s4	CAGTTTTCTGTAGCCTTCTGGG
Cbx2	TTGGAGAAGTGACCCTGTTTG
Ccn2	CACCATAGGTAGGACACGAAG
Cdkn1a	GGCTGATCCTTTCTCAGTGTT
Ddit4	CATTCGGGATCTTCGACACTT
Dnajb1	CAGTCCCTGTCAACTACTGTC
Fbxo5	GGGAGGGTTGTGGTCTTTTTA
Fgf4	TGCAGCTATGTTGCTTCAGG
Fiz1	ACAGTCCTTTGACCACGATTT
Klf2	GAAAGAAGACAGGAGTCTGTGAA
Kri1	TTGAAATGAGGCTACCAGAGC
Ldlr	TGGCTCGGTTTTCATTCTGT
Lefty1	CAATCCCTGTGTGTGCTCTTT
Lifr	ATTGACAGCGGAGACATTCC
Med7	CTGAGCCGATGGATACTGATG
Mettl14	GGCTGGTTTAAACGGTTGAG
Mief1	TACCAGGCTAGGAGGTTACAA
Nanog	TGTGTCATTTTGAGGGGTGA
Nfkbia	AAATCCGGGGAAACAGATCAC
Nr0b1	GATGGAGAAAGCGGTCGTAG
Oser1	TTCAGACCTGCTTACGATGTG
Platr10	AGAGCATTTCTGTGTATCTGTTGAG
Pri-mir-290-295	ACAGGTCGGCTGCTTTCAT
Pum2	ACCCAGTCCTCCTGATACTTTT
Slc46a1	TCAAGCCAGTTATGAGCAACA
Sox17	TTCGGATTTTCAGTCATCTCG
Tnrc6c	TGCTGCCATTTAGCACTGAG
Tor1b	GTGACTTTCAGGCCATAGTCTT
Txnip	GCCTGGGTGACATTCTACATT
Utp23	CCTGAAAGTGTCTTCTGAGCA
Zfp141	TGTATGAACTTGTACCTGAGTAAGA
Zfp24	TGGGTCATATCTGGTGCTTTG
Zfp553	GCTTCAAGTCTCCACAGCTAA
Zfp961	GTTCCAGTTTCAGTCTCCAGT

Table 2: **Forward (Gene specific +i5) Primers (10µM)**
The i5 adaptor sequence consists of “CACGACGCTCTTCCGATCT” and the six genes marked with grey, where the ones picked out for some of the later experiments.

Partial Illumina i7 adaptor	GTTCAGACGTGTGCTCTTCCG

Table 3: **Reverse (partial i7) Primer (10µM)**

Fbxo5	GAAAGTAAAAATAATTATGATGCCTT
Fgf4	CACCTGCCCTGTTCTGATG
Fiz1	CAGACATTCCCTTATGCTGTA
Nanog	TTTTCTTAGAAGCGTGGGTCTT
Nr0b1	GGCATATGTCCACTCAAGT
Arl6ip6	ATGTAGAAATGTTTGAACTTGTTT

Table 4: **External Forward (Gene specific) Primer (10µM)**
Primers used for semi-nested approach

4.5. Cell Culture

Mouse embryonic stem cells were cultured in media, maintained at 37°C with 5% CO_2 and passaged every second day.

4.6. SLAMseq in mESCs

For detailed Protocol see Herzog, Fasching & Ameres (2020).
$5*10^5$ mESC were seeded on a 6-well plate the day before the experiment. On the next day the cells were incubated for 1h with 500µM s^4U at 37°C. Afterwards the cells got washed with 1x PBS. With 500µl TRIzol the cells got lysed, then transferred to a tube and directly processed or stored at -80°C.

4.7. RNA Isolation

For RNA isolation 200µl of chloroform isoamyl alcohol, in a 24:1 per ml of TRIzol ratio, were added. After shaking the tube, it was incubated at RT for 2-3 minutes. Then it was spun down at 16.000g for 15 min. at 4°C. The aqueous phase got transferred into a new tube and after addition of 1µl glycogen, 1/100 volumes of 10mM DTT and 1 volume 2-propanol, vortexed well. This followed a 10 min. at RT incubation and a centrifugation at 16.000g for 20 min. at 4°C. Then the supernatant got removed and the pellet washed with 500µl 75% EtOH and 5µl of 10mM DTT. After vortexing and spinning at 7.500g for 5 min. at RT, the supernatant got removed. The pellet was dried for 5-10 min. and resuspended in 100µl water with 1mM DTT. The extracted RNA was frozen and stored at -80°C.

4.8. Iodoacetamide (IAA) Treatment

For the IAA reaction 5µg of RNA were used, with 5µl of 100mM IAA, 5µl of 500mM $NaPO_4$, 25µl DMSO and, according to the RNA input, water to 50µl. This reaction mix was incubated at 50°C for 15 minutes. Afterwards the reaction got stopped by adding 1µl 1M DTT. After adding 1µl glycogen, 5µl of 3M NaOAc and 125µl of 100% EtOH, the reaction was vortexed and precipitated at -80°C for 30 min. Then it was spun down at 16.000g for 10-15 min on 4°C, the supernatant discarded, and the pellet washed with 1ml of 75% EtOH. After again spinning down at 16.000g for 5 min at 4°C, the supernatant was removed, and the pellet dried for 5-10 min. Finally, it got resuspended in 44µl water.

4.9. TURBO DNA-free Protocol

In addition to the 44µl RNA, the reaction consisted of 5µl 10x DNAse Buffer and 1µl TURBO DNAse. This got incubated at 37°C for 30 min. To stop the reaction 5µl of DNAse Inactivation Reagent got added. After incubation for 5 min at RT, with occasional mixing, the reaction got centrifuged at 10.000g for 1.5 min and 49µl transferred to a fresh tube.

4.10. RNA library preparation

4.10.1. Reverse Transcription

To start RT the following reaction needed to be assembled: 2µl barcoded oligo dT_{18} (2.5µM), 1µl of 10mM dNTPs, 500ng RNA and water up to 14.5µl. After incubating at 65°C for 5 min. and cooling down to 4°C; 4µl of 5x First-Strand buffer, 1µl of 0.1M DTT, 0.25µl RNAse OUT (40 U/µl) and 0.25µl of SuperScript III RT were added. The RT reaction was run at 25°C for 10 minutes, 50°C for 60 min. and 70°C for 15 min. This was followed by MBS magnetic beads clean up. After adding 1.5x of beads to the reaction and incubating for 5 min. at RT, it was placed on a magnet until the solution became clear. Then the supernatant was discarded, and the beads washed twice with 200µl of 80% EtOH. After quickly spinning down the residual EtOH was removed, the beads dried for 1-2 min. and resuspended in 22µl water. The beads were incubated for at

least 1 min, then put on the magnet and 20µl of the supernatant transferred to fresh tubes or plates.

4.10.2. PCR1 – Second Strand Synthesis

The next step in the protocol was the second strand synthesis, where the reaction consisted of: 10µl of cDNA, 12.5µl Taq MM, 1µl gene specific forward primer or primer Mix (10µM), 1µl reverse primer (10µM) and 0.5µl water. The according PCR program was 95°C for 2 minutes, following 12x cycles at 95°C for 30s, 58°C for 30s and 72°C for 30s; and then a final extension with 72°C for 5 min. The PCR was followed with another MBS beads clean up, where everything was as previously done.

4.10.3. PCR2

For the barcode amplification PCR, the reaction consisted of: 25µl KAPA Hifi Hotstart Readymix, 10µl of i5 and i7 barcoded oligo (500nM), 2.5µl Evagreen, 2.5µl ddH2O and 10µl DNA. The PCR program ran at 98°C for 45s, then as recurring cycles at 98°C for 15s, 65°C for 30s and 72°C for 35s. This cycle was repeated until the samples reached almost 5000 RFU, then they were incubated for 1 minute at 72°C.

Afterwards another 1.5x beads clean-up was done, but finally the beads were resuspended in 32µl and 30µl were transferred to fresh tubes or plate.

4.10.4. Quality Control

To confirm sufficient quality of the libraries for sequencing all samples were run on the Fragment Analyzer. On a 96 well plate, each well holds 22µl of HS dilutant marker and 2µl sample. Furthermore, 40ml of DNA Gel with 4µl of dye were prepared. After adding all components into the machine, the DNF-474 High Sensitivity NGS 1-6000bp program was run. With the now known concentrations and sizes of the libraries the samples were handed off to the NGS facility of the VBCF, where a qPCR was run and with these exact values a multiplex of all samples of 2.5nM was created

4.10.5. Sequencing with MiSeq

Before starting a Miseq run the Illumina reagents for the PE 150 Nano kit had to be thawed for 1h. During this time 19µl of the 2.5nM sample were denatured with 1µl of 1M NaOH for 5 minutes. Afterwards 980µl of Hybridisation Buffer were added. From this 63µl were taken and mixed with 5.25µl of 200pM PhiX and 531.75µl of Hybridisation Buffer. This final dilution was filled into the thawed cartridge. The consumables of the Illumina kit, the cartridge and the flowcell were loaded into the machine and the PE 150 cycles program was run.

4.11. RT Optimisation and ExoSAP-IT Addition

To optimise reverse transcription and to test if the addition of ExoSAP-IT to the protocol would positively influence the output, eight different RT reactions, seen in *Table 5*, were used. To the 20µl reaction, 4µl of ExoSAP-IT were added, right after running one of the eight RT programs. Afterwards the samples were incubated at 37°C for 15 min and then at 80°C for 15 min. The samples without ExoSAP-IT were heat inactivated as usual with 70°C for 15 min. Afterwards the protocol progressed to its standard 1.5x bead clean up.

Number	RT Conditions	ExoSAP-IT
1	50°C for 15 min	+
2	50°C for 15 min	-
3	42°C for 15 min	+
4	42°C for 15 min	-
5	50°C for 7 min	+
6	50°C for 7 min	-
7	50°C for 50 min	+
8	25°C for 10 min -> 50°C for 50 min	+

Table 5: **Different conditions for RT optimisation and ExoSAP-IT addition**

4.12. RNA Input Decrease and Comparison of dT18 and dT30

To test the RNA input decrease 500ng, 50ng and 5ng of total RNA input were used for RT. Also, oligo dT_{30} was used as comparison to dT_{18}; it was incorporated in the reaction the same. Furthermore, at this stage of the optimisations, the in between step of the RT was changed from 65°C for 5 min. and then cool down to 4°C; to 65°C for 5 min. and then cool down to 42°C. All bead clean ups were changed from 1.5x to 1x, to narrow down the desired size.

4.13. Multiplexing with Individual Barcodes

For the multiplexing, each sample had an individually assigned barcoded oligo dT_{18} at RT. After completing the newly established and optimised steps, the samples were pooled in a 1:1 ratio, which was for this experiment 2.5µl of each of the eight samples. The final bead clean up after PCR2 was reduced from 1x to 0.9x.

4.14. Implementation of the Semi-Nested Approach

For the semi-nested approach, after RT and clean up, the first PCR changed to PCR0. Here the external Primer was added, therefore the reaction consisted of 10µl of cDNA, 12.5µl of 2x Taq MM, 1µl of external Primer or Mix (10µM), 1µl of the reverse primer (10µM) and 0.5µl water. The PCR program was as before, but for 10, 15, 20 and 30 cycles. After PCR0 5µl of ExoSAP-IT was added to the 25µl reaction. To figure out how much input to use for PCR1, a Qubit 1x dsDNA HS measurement was done. Here 199µl of the Qubit working solution was distributed for each sample in a qubit tube and 1µl of each reaction was added. After two minutes incubation the 1x dsDNA HS measurement at the qubit machine could be done. For the further reactions 1ng of total input was used.

Therefore, the reaction for PCR1 consisted of the input DNA according to Qubit measurement, the rest was the same as for PCR0, except internal instead of external Primers, and 25µl total reaction. The PCR program was the same as PCR0, but for 8 and 13 cycles; and until all primers were exhausted. Afterwards the standard 1x beads clean up and PCR2 was done. For the input material for PCR2 a Qubit measurement was done to not exceed 1ng. The rest of the protocol was as before.

4.15. Exhaustive PCR

In the basic semi-nested protocol, some exhaustive components were already present. The changes to the specific exhaustive protocol were: PCR0 cycled for 30x, between PCR0 and PCR1 no ExoSAP-IT treatment was done, but the Qubit measurement remained. Also, at PCR1 the cycles were 60x, 40x and 15x. The rest of the protocol remained the same.

4.16. Primer Titration

The Primer Titration protocol was also very similar. PCR0 was run for 30 cycles with the standard external Primers. For PCR1 four specific Primer Mixes were created (Table 6) and the program run for 15 and 40 cycles. The rest of the protocol stayed the same as the semi-nested exhaustive PCR.

	Mix1	**Mix2**	**Mix3**	**Mix4**
Fgf4	5.4	4	8	8
Nanog	2.7	2	4	2
Fiz1	2.8	2	4	2
Arl6ip6	7.1	4	8	8
Fbxo5	17	6	12	12
Nr0b1	1	1	1	1

Table 6: **Different Primer Mixes for Primer Titration Experiment**
Primer abundance was set relative to Nr0b1

4.17. In-Lysate Protocol

For the in-lysate protocol 5.000 cells/well were seeded on a 96-well plate the night before, then s^4U labelled for 1h at 37°C as previously described. Instead of RNA isolation, here, the cells were lysed directly. Therefore, they were washed 2x with 1x PBS, 30µl LB was added and they were put on a shaker at 2000rpm for 3 min. at 50°C. Afterwards, 5µl of the cells in lysis buffer, 3µl of IAA (100mM in DMF), 1.5µl $NAPO_4$ (ph8, 500mM), 2µl barcoded oligo dT_{18} or dT_{30} (2.5µM) and 3.5µl PBS were added. This reaction was incubated at 80°C for 2 minutes and then stopped with 1µl DTT (1M). For the RT it got directly added: 0.5µl dNTPs (10mM), 6µl 5x First-Strand Buffer, 0.5µl RNAse OUT (40 U/µl) and 0.5µl of SuperScript III or IV and 6.5µl H_2O. The reaction was incubated at 42°C for 15 minutes, then treated with 6µl ExoSAP-IT and incubated again at 37°C for 15 min. followed by 80°C for 15 min. From here on out the previous protocol could be used again. The only changes were, that for PCR1 5µl cDNA input was used and 35 cycles were run, for the semi-nested approach PCR0 was 30 cycles and PCR1 was 15x.

To compare the in-lysate with the in vitro some modifications were made. For the in vitro approach 50ng RNA, 2µl barcoded oligo dT_{18} or dT_{30} (2.5µM), 0.5µl of 10mM dNTPs, 6µl of 5x First-Strand buffer, 0.5µl RNAse OUT (40 U/µl) and 0.5µl and up to 30µl total volume of water were directly into one reaction combined. Afterwards, the standard RT times and temperatures were used. ExoSAP-IT treatment was performed as usual and the only change for PCR1 was that 10µl cDNA input were used.

4.18. In Vitro Alkylation with DMF

To test if the in vitro alkylation is compatible with DMF, 9.25mg IAA were dissolved in 500µl of DMF, DMSO or 100% EtOH. As negative control DMF was used. Each reaction consisted of 5µl 10mM s^4U, 5µl 100mM IAA (dissolved in the different solutions), 5µl 500mM $NaPO_4$ (pH8) and 32µl H_2O. Every condition was done in triplicates and then incubated at 50°C for 15 min or 80°C for 2 min. Afterwards, the reactions were stopped with addition of 1µl 1M DTT. Then the absorbance was measured on nanodrop, with the negative control (DMF) as blank.

4.19. Bioinformatics and Data Analysis

To understand the bioinformatics and data analysis, the structure of the primers needs to be clear. The primer design for the specific genes consisted of targeting the 3'end with the primers and an amplicon size between 200-400bp. The off priming was checked against a genome and transcriptome database and a Primer BLAST was also done. The setup of the amplicon and all primers can be seen in Figure 15. Read 1 offers information about the gene specific forward primer and the amplicon. Whereas read 2 contains the UMI, for deduplicating, the RT barcode, to distinguish between the multiplexed samples, and the read into the amplicon.

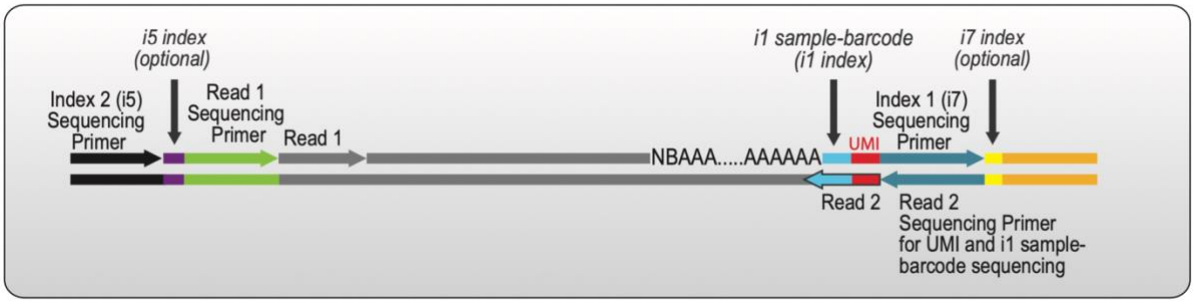


Figure15: **Scheme of the Transcript**
On the left is the first read (Read 1), which is accompanied by the according sequencing primer and the Illumina index (i5) primer. On the right is the Illumina indexing (i7) primer, followed by the second read (Read 2). Here are also the UMI, the sample barcode and the oligo dT located.

The primer design in detail would be that barcoded oligo dT primers have the following structure: 5′–adaptor–UMIs – barcode–oligodT18 (or 30)–VN–3′. The adaptor sequence is GTTCAGACGTGTGCTCTTCCGATCT, the 12nt UMI-sequence is HBNVDNHBNVDN and the 12nt barcodes correspond to i1 barcodes from the Quantseq Pool from Lexogen. The internal primers were designed with NCBI Primer Blast using annotated 3′UTRs (Herzog, Fasching & Ameres, 2020) as input. The primer binding site was selected upstream of the 3'-end of a transcript. The primer pair specificity was checked against "Refseq RNA". The adaptor sequence (CACGACGCTCTTCCGATCT) was added to 5′ end of the internal primer. The external primers were designed with NGS Primer Plex (Kechin et al., 2020) using a list of previously designed internal primers and annotated 3′UTRs. Minimal primer shift between internal and external primers was set to 25nt with maximal external amplicon length 400 nt.

The bioinformatic pipeline used is called TSLAM, which is a nextflow pipeline for the analysis of targeted SLAMseq data. A central step of TSLAM is the extraction of the UMI, barcode and primer sequences from the read pairs. Reads are then mapped to the genome in order to estimate primer performance and off-target rates. Finally, reads are mapped to the amplicons and read pairs, with and without UMI-based PCR deduplication and. All different information get counted and then proceeded to the data analysis.

4.19.1. Bioinformatics – Targeted SLAMseq Analysis (TSLAM)

The sequencing adapters were removed with fastp v0.20.1 and then the read pairs were analysed with a custom python script. This script aligned all configured primers to the 5′ end sequences of the first mate using a global pairwise sequence alignment. Alignment scores were normalized to the primer length and the primer with the highest normalize score was selected unless the score was below a configurable minimum threshold (default: 0.8). The primer sequence was then cleaved off and the primer name was appended to the read pair name.

Then a search for the barcode, UMI and poly-T stretch were conducted on the second mate: First, all barcodes were configured to the 3′ end of the read using a global pairwise sequence. The barcode with the highest score was selected unless the score was below a minimum threshold (default: 0.8). Then the UMI length upstream bases were considered as UMI of the read pair, and it was searched for a poly-T stretch in the downstream part of the read: the goal here was to find a stretch of Ts with configurable maximum length (18 or 30 were used), allowing for a configurable maximum of mismatches due to sequencing errors; two mismatches for dT_{18} and three mismatches for dT_{30} were configured. The read part upstream of the end of the poly-T stretch plus a configurable number of additional bases was then cleaved off and UMI and barcode were appended to the read pair name.

After this, read pairs were filtered if no barcode or primer was found, if the UMI was incomplete or if the poly-T stretch was shorter than a configurable length; for dT_{18} a minimum of 10 and for dT_{30} a minimum of 22 were configured.

Then the remaining read pairs were quality-trimmed, too-short reads (<50bp) removed, and they were written into barcode-specific FASTQ files. Read pairs were then mapped to the targeted amplicons using STAR v2.6.1d. Reads containing T>C conversions and PCR duplicates, based on detected UMI

sequences, were annotated, and read pairs per amplicon with and without PCR duplicates were counted using a custom python script.

4.19.2. Data Analysis

Additionally, read pairs were mapped to the mm10 genome (GRCm38.p6) to calculate primer performance and off-target. All following statistics and calculations were performed with R (Version 1.4.1103).

To evaluate primer performance certain values were needed: True Positives (TP), Fales Negatives (FN) and False Positives (FP). TP are when the target gene can be detected by the bioinformatic analysis in the according read, which means that the primers are working and are specific. FN are when the target gene is in fact present in the read, but the primer can bioinformatically not be fully found, so some primer specificity was lost compared to TP. FP are off priming and miss priming, which means that the primers attached at incorrect locations. The combination of FN and FP are general referred to as off target. To assess primer performance F1 was calculated (F1= TP/(TP+(FN+FP))), this means the higher F1 the better the primers. To pick the best performing primers the conditions F1>0.5 and TP>50 , more than 50 TP reads, were chosen.

Another necessary value were the passed reads. These are all valid reads that contain a useable and detectable primer, and the according RT barcode. The T>C conversion rate was calculated: TC reads/all reads, here TC reads stands for every T that got changed into a C, and they are divided by all reads, so all the Ts together. To filter out errors or spontaneous mutations at least two T>C conversions had to have taken place in a read, so the final calculation was: TC2 reads/all reads
This shows the fraction T>C rate, which can be compared between experiments. Because it was possible to filter out the UMIs, the reads could be deduplicated. That means that only unique UMI reads were counted and this compared to all reads give information about the Amplicons. The closer each value (of the sample or specific amplicon) goes towards one, the more unique (no duplications) the read. The graphs and figures were created using R or Prism 9.

5. References

Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K., & Kobayashi, S. (1999). Maternal Pumilio acts together with Nanos in germline development in Drosophila embryos. *Nature cell biology*, *1*(7), 431–437. https://doi.org/10.1038/15666

Beckmann Coulter Inc. (2022) Echo-525 (https://www.beckman.at/liquid-handlers/echo-525)

Beckmann Coulter Inc. (2022). Biomek-i7 (https://www.beckman.at/liquid-handlers/biomek-i7)

Boileau, E., Altmüller, J., Naarmann-de Vries, I. S., & Dieterich, C. (2021). A comparison of metabolic labeling and statistical methods to infer genome-wide dynamics of RNA turnover. *Briefings in bioinformatics*, *22*(6), bbab219. https://doi.org/10.1093/bib/bbab219

Casamassimi, A., & Ciccodicola, A. (2019). Transcriptional Regulation: Molecules, Involved Mechanisms, and Misregulation. *International journal of molecular sciences*, *20*(6), 1281. https://doi.org/10.3390/ijms20061281

Chi, X., Zhang, Y., Xue, Z., Feng, L., Liu, H., Wang, F., & Qi, X. (2014). Discovery of rare mutations in extensively pooled DNA samples using multiple target enrichment. *Plant biotechnology journal*, *12*(6), 709–717. https://doi.org/10.1111/pbi.12174

Cole C. N. (2001). Choreographing mRNA biogenesis. *Nature genetics*, *29*(1), 6–7. https://doi.org/10.1038/ng0901-6

de Nadal, E., Ammerer, G., & Posas, F. (2011). Controlling gene expression in response to stress. *Nature reviews. Genetics*, *12*(12), 833–845. https://doi.org/10.1038/nrg3055

Dhamija, S., & Menon, M. B. (2018). Non-coding transcript variants of protein-coding genes - what are they good for?. *RNA biology*, *15*(8), 1025–1031. https://doi.org/10.1080/15476286.2018.1511675

Doma, M. K., & Parker, R. (2007). RNA quality control in eukaryotes. *Cell*, *131*(4), 660–668. https://doi.org/10.1016/j.cell.2007.10.041

Elling, U., Wimmer, R. A., Leibbrandt, A., Burkard, T., Michlits, G., Leopoldi, A., Micheler, T., Abdeen, D., Zhuk, S., Aspalter, I. M., Handl, C., Liebergesell, J., Hubmann, M., Husa, A. M., Kinzer, M., Schuller, N., Wetzel, E., van de Loo, N.,

Martinez, J., Estoppey, D., … Penninger, J. M. (2017). A reversible haploid mouse embryonic stem cell biobank resource for functional genomics. *Nature*, *550*(7674), 114–118. https://doi.org/10.1038/nature24027

Erhard, F., Baptista, M., Krammer, T., Hennig, T., Lange, M., Arampatzi, P., Jürges, C. S., Theis, F. J., Saliba, A. E., & Dölken, L. (2019). scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature*, *571*(7765), 419–423. https://doi.org/10.1038/s41586-019-1369-y

Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M. D., & Hughes, T. R. (2004). Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Molecular and cellular biology*, *24*(12), 5534–5547. https://doi.org/10.1128/MCB.24.12.5534-5547.2004

Guo J. (2014). Transcription: the epicenter of gene expression. *Journal of Zhejiang University. Science. B*, *15*(5), 409–411. https://doi.org/10.1631/jzus.B1400113

Haimovich, G., Medina, D. A., Causse, S. Z., Garber, M., Millán-Zambrano, G., Barkai, O., Chávez, S., Pérez-Ortín, J. E., Darzacq, X., & Choder, M. (2013). Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell*, *153*(5), 1000–1011. https://doi.org/10.1016/j.cell.2013.05.012

Herzog, V. A., Fasching, N., & Ameres, S. L. (2020). Determining mRNA Stability by Metabolic RNA Labeling and Chemical Nucleoside Conversion. *Methods in molecular biology (Clifton, N.J.)*, *2062*, 169–189. https://doi.org/10.1007/978-1-4939-9822-7_9

Herzog, V. A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T. R., Wlotzka, W., von Haeseler, A., Zuber, J., & Ameres, S. L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. *Nature methods*, *14*(12), 1198–1204. https://doi.org/10.1038/nmeth.4435

Houseley, J., & Tollervey, D. (2009). The many pathways of RNA degradation. *Cell*, *136*(4), 763–776. https://doi.org/10.1016/j.cell.2009.01.019

Huang, X. T., Li, X., Qin, P. Z., Zhu, Y., Xu, S. N., & Chen, J. P. (2018). Technical Advances in Single-Cell RNA Sequencing and Applications in Normal and Malignant Hematopoiesis. *Frontiers in oncology*, *8*, 582. https://doi.org/10.3389/fonc.2018.00582

Kadonaga J. T. (2004). Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell*, *116*(2), 247–257. https://doi.org/10.1016/s0092-8674(03)01078-x

Kadyrova, L. Y., Habara, Y., Lee, T. H., & Wharton, R. P. (2007). Translational control of maternal Cyclin B mRNA by Nanos in the Drosophila germline. *Development (Cambridge, England)*, *134*(8), 1519–1527. https://doi.org/10.1242/dev.002212

Kechin, A., Borobova, V., Boyarskikh, U., Khrapov, E., Subbotin, S., & Filipenko, M. (2020). NGS-PrimerPlex: High-throughput primer design for multiplex polymerase chain reactions. *PLoS computational biology*, *16*(12), e1008468. https://doi.org/10.1371/journal.pcbi.1008468

Komili, S., & Silver, P. A. (2008). Coupling and coordination in gene expression processes: a systems biology view. *Nature reviews. Genetics*, *9*(1), 38–48. https://doi.org/10.1038/nrg2223

Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell*, *152*(6), 1237–1251. https://doi.org/10.1016/j.cell.2013.02.014

Lugowski, A., Nicholson, B., & Rissland, O. S. (2018). Determining mRNA half-lives on a transcriptome-wide scale. *Methods (San Diego, Calif.)*, *137*, 90–98. https://doi.org/10.1016/j.ymeth.2017.12.006

Ma, F., Fuqua, B. K., Hasin, Y., Yukhtman, C., Vulpe, C. D., Lusis, A. J., & Pellegrini, M. (2019). A comparison between whole transcript and 3' RNA sequencing methods using Kapa and Lexogen library preparation methods. *BMC Genomics*, 20(1), 9. doi:10.1186/s12864-018-5393-3

Mata, J., Marguerat, S., & Bähler, J. (2005). Post-transcriptional control of gene expression: a genome-wide perspective. *Trends in biochemical sciences*, *30*(9), 506–514. https://doi.org/10.1016/j.tibs.2005.07.005

Moore M. J. (2005). From birth to death: the complex lives of eukaryotic mRNAs. *Science (New York, N.Y.)*, *309*(5740), 1514–1518. https://doi.org/10.1126/science.1111443

Morey, C., & Avner, P. (2004). Employment opportunities for non-coding RNAs. *FEBS letters*, *567*(1), 27–34. https://doi.org/10.1016/j.febslet.2004.03.117

Proudfoot, N. J., Furger, A., & Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell*, *108*(4), 501–512. https://doi.org/10.1016/s0092-8674(02)00617-7

Ross J. (1995). mRNA stability in mammalian cells. *Microbiological reviews*, *59*(3), 423–450. https://doi.org/10.1128/mr.59.3.423-450.1995

Russo, J., & Olivas, W. M. (2015). Conditional regulation of Puf1p, Puf4p, and Puf5p activity alters YHB1 mRNA stability for a rapid response to toxic nitric oxide stress in yeast. *Molecular biology of the cell*, *26*(6), 1015–1029. https://doi.org/10.1091/mbc.E14-10-1452

Russo, J., Heck, A. M., Wilusz, J., & Wilusz, C. J. (2017). Metabolic labeling and recovery of nascent RNA to accurately quantify mRNA stability. *Methods (San Diego, Calif.)*, *120*, 39–48. https://doi.org/10.1016/j.ymeth.2017.02.003

Schofield, J. A., Duffy, E. E., Kiefer, L., Sullivan, M. C., & Simon, M. D. (2018). TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nature methods*, *15*(3), 221–225. https://doi.org/10.1038/nmeth.4582

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., & Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science (New York, N.Y.)*, *352*(6290), 1225–1228. https://doi.org/10.1126/science.aad984

Shalem, O., Dahan, O., Levo, M., Martinez, M. R., Furman, I., Segal, E., & Pilpel, Y. (2008). Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Molecular systems biology*, *4*, 223. https://doi.org/10.1038/msb.2008.59

Singer, R. H., & Penman, S. (1972). Stability of HeLa cell mRNA in actinomycin. *Nature*, *240*(5376), 100–102. https://doi.org/10.1038/240100a0

Slavov, N., & Botstein, D. (2013). Decoupling nutrient signaling from growth rate causes aerobic glycolysis and deregulation of cell size and gene expression. *Molecular biology of the cell*, *24*(2), 157–168. https://doi.org/10.1091/mbc.E12-09-0670

Wang, Y., Liu, C. L., Storey, J. D., Tibshirani, R. J., Herschlag, D., & Brown, P. O. (2002). Precision and functional specificity in mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(9), 5860–5865. https://doi.org/10.1073/pnas.092538799

Weiss, C. N., & Ito, K. (2017). A Macro View of MicroRNAs: The Discovery of MicroRNAs and Their Role in Hematopoiesis and Hematologic Disease. *International review of cell and molecular biology*, *334*, 99–175. https://doi.org/10.1016/bs.ircmb.2017.03.007

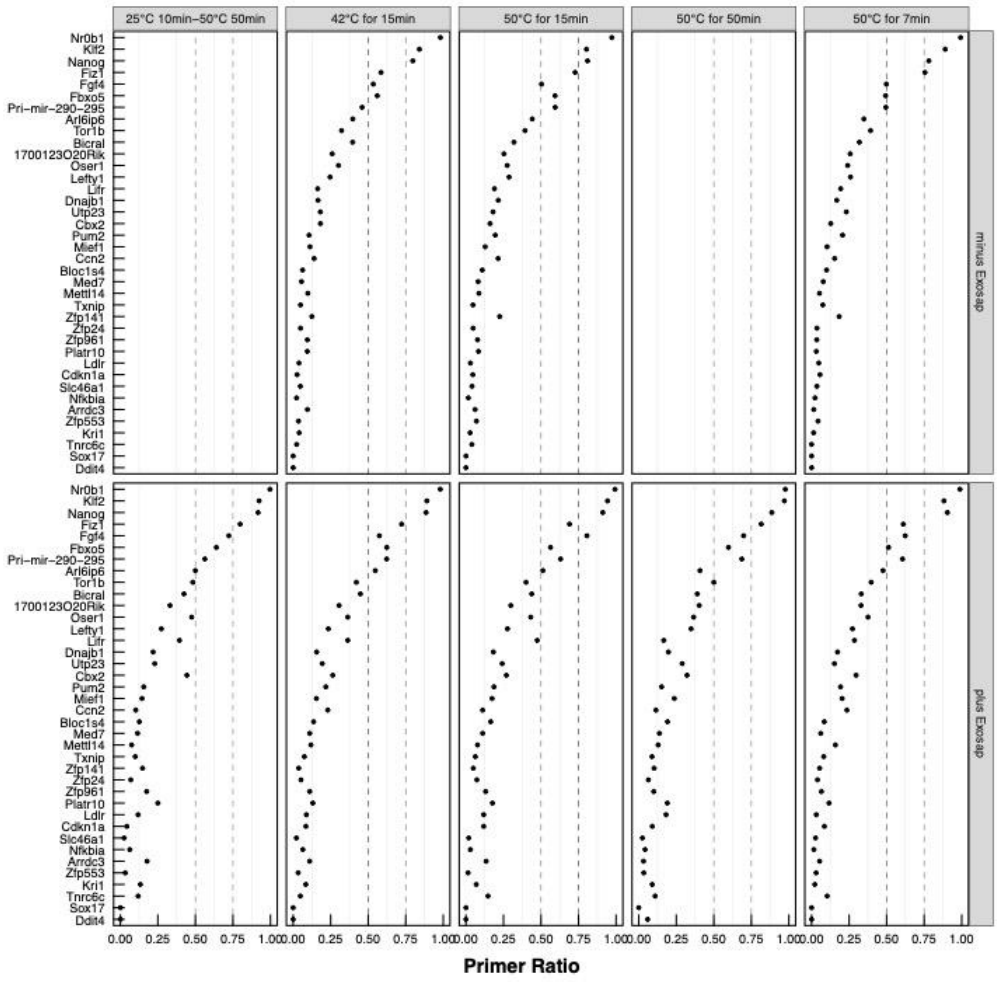
Wu, G., Schmid, M., Heick Jensen, T. (2021). Chapter Seven - 3′ End sequencing of pA+ and pA− RNAs. *Methods in Enzymology*, 655, 139-164. https://doi.org/10.1016/bs.mie.2021.03.020.

Yasukawa, K., Konishi, A., & Inouye, K. (2010). Effects of organic solvents on the reverse transcription reaction catalyzed by reverse transcriptases from avian myeloblastosis virus and Moloney murine leukemia virus. *Bioscience, biotechnology, and biochemistry*, *74*(9), 1925–1930. https://doi.org/10.1271/bbb.100337

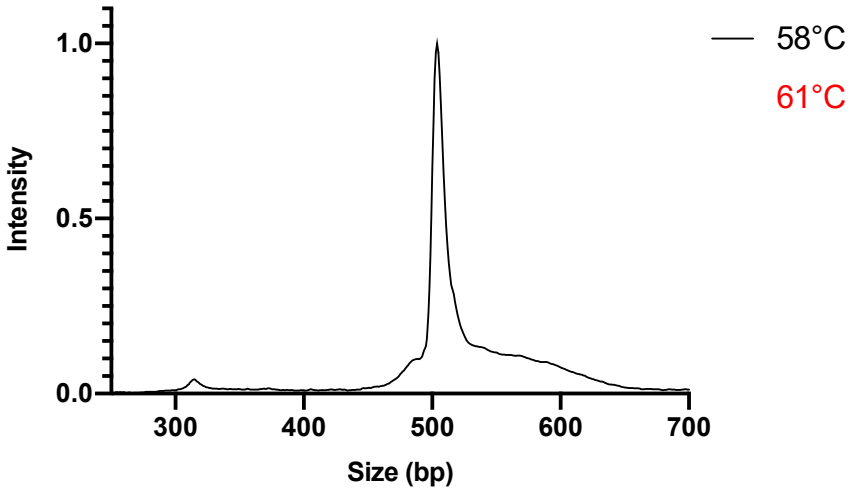
Ye, C., Ho, D. J., Neri, M., Yang, C., Kulkarni, T., Randhawa, R., Henault, M., Mostacci, N., Farmer, P., Renner, S., Ihry, R., Mansur, L., Keller, C. G., McAllister, G., Hild, M., Jenkins, J., & Kaykas, A. (2018). DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. *Nature communications*, *9*(1), 4307. https://doi.org/10.1038/s41467-018-06500-x

Yehudai-Resheff, S., & Schuster, G. (2000). Characterization of the E.coli poly(A) polymerase: nucleotide specificity, RNA-binding affinities and RNA structure dependence. *Nucleic acids research*, *28*(5), 1139-1144. https://doi.org/10.1093/nar/28.5.1139

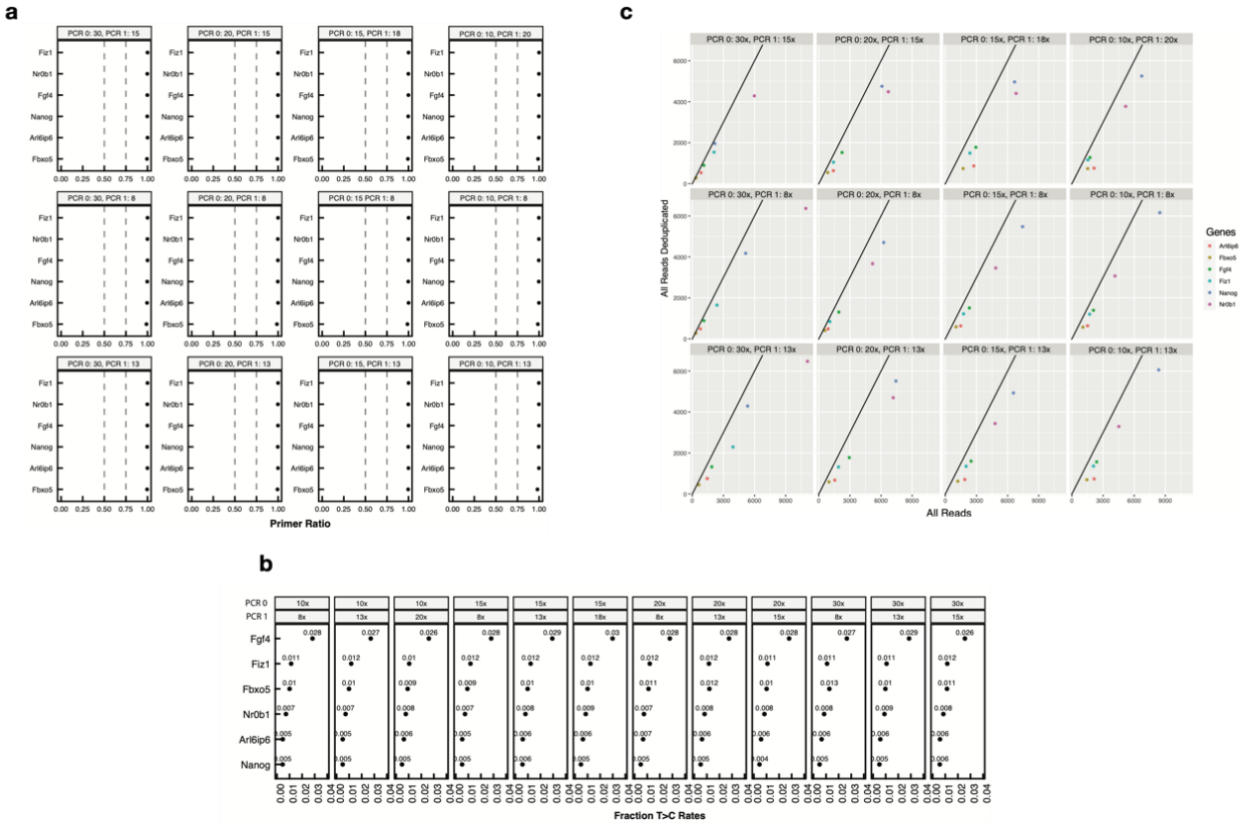
6. Supplementary Material



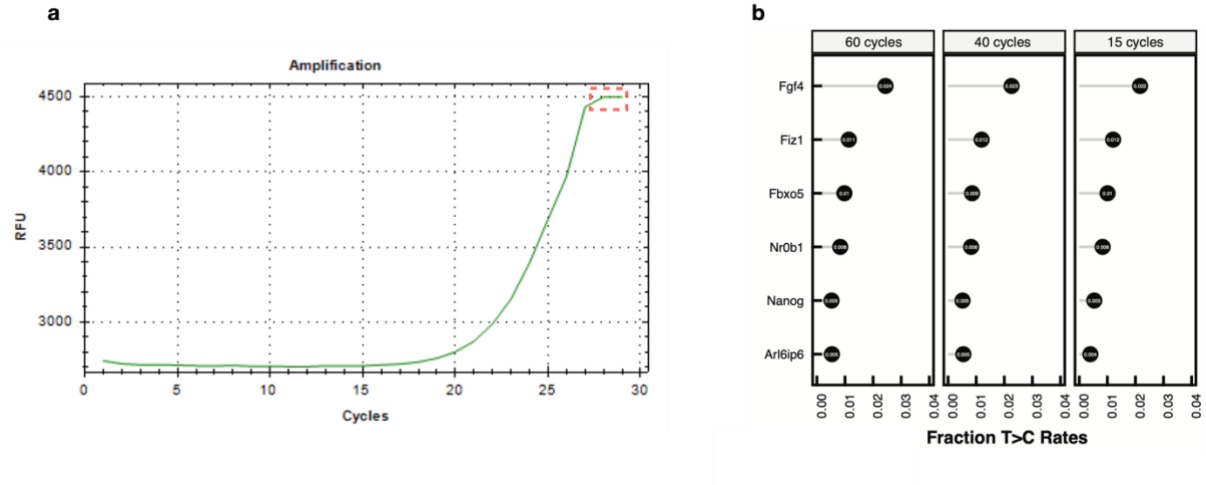
Supplementary Figure 1: **Primer Performance of Reverse Transcription Optimisation**
In this experiment eight different temperature and times were tested to optimise RT. 38 different genes were picked out and samples were cleaned up with or without ExoSap-IT. The primer ratio is the visualization for the primer performance and shows on and off target reads for all genes in each condition.



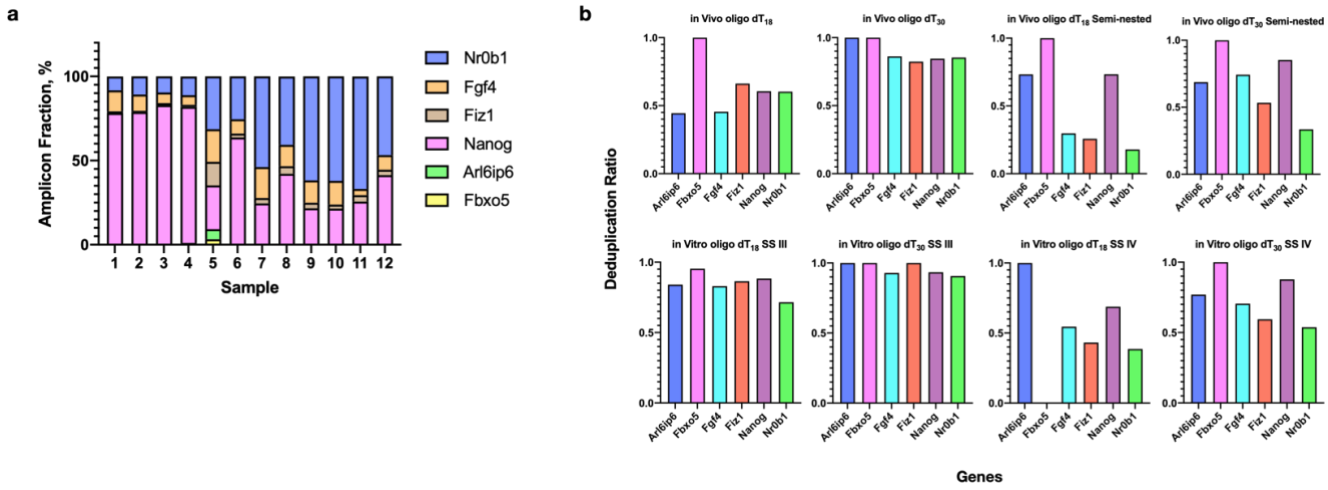
Supplementary Figure 2: **Different Annealing Temperatures for Nr0b1**
For second strand synthesis two different annealing temperatures were tested (58°C in black and 61°C in red), to see if they influence the library preparation. The curves calculated from fragment analyser show the intensity compared to the size in bp.



Supplementary Figure 3: **Additional Results of the Semi-nested Approach**
a) To implement the semi-nested approach 12 different conditions were used with varying cycle lengths for PCR0 and PCR1. The primer ratio is the visualization for the primer performance and shows on and off target reads. Here shown for all conditions and six genes. b) The T>C conversion rate over all samples, shows how many Ts to Cs were changed. c) Here the fraction of deduplicated reads, so unique ones, are compared to all reads. This is shown for all conditions and the genes are marked in colours.



Supplementary Figure 4: **Additional Results of the Exhaustive PCR**
a) Schematical example of reaching Primer exhaustion (red box) during PCR. For exhausting PCR, the plateau needs to be reached, which shows that there is no amplification happening anymore. b) The T>C conversion rate over all samples, shows how many Ts to Cs were changed. Here for six genes and 60, 40 and 15 cycles at PCR1, while PCR0 stays with 30 cycles the same for all.



Supplementary Figure 5: **Additional Results of Alkylation and RT in Vivo**
a) Fractions of all six amplicons per sample in %. Sample 1 and 2 are oligo dT_{18} in vivo. Sample 3 and 4 are dT_{30} in vivo. Sample 5 is dT_{18} and SS III in vitro. Sample 6 dT_{30} and SS III. Sample 7 and 8 are SS IV in vitro with dT_{18} and dT_{30}. Sample 9 and 10 are dT_{18} in vivo semi-nested. Sample 11 and 12 the same but dT_{30}. b) Deduplication ratio is the fraction of deduplicated reads, so unique ones, compared to all reads. This is shown for all conditions and the 6 genes.