From Department of Microbiology, Tumor and Cell Biology

Karolinska Institutet, Stockholm, Sweden

Investigating novel mechanisms for transcriptional memory using genetic engineering

Yerma P. Sanchez

Karolinska Institutet

Stockholm 2023

” Everything starts with a dot.”

- W. Kandinsky

Published by Karolinska Institutet.

Printed by Universitet service US-AB, 2023

ISBN 978-91-8017-072-7

Cover image by Yerma P. Sánchez (inspired by Kandinsky art, idea of F. Righetti)

INVESTIGATING NOVEL MECHANISMS FOR TRANSCRIPTIONAL MEMORY USING GENETIC ENGINEERING

Thesis for Doctoral Degree (Ph.D.)

By Yerma Pareja Sánchez

The thesis will be defended in public at Air&Fire, SciLifeLab, Tomtebodavägen 23A, Solna (Sweden), 22nd of August at 9:30

Principal Supervisor:
Dr. Vicente Jose Pelechano García
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology

Co-supervisor(s):
Prof. Marie Arsenian-Henriksson
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology

Prof. Sebastián Chávez de Diego
Universidad de Sevilla
Department of Genetics

Opponent:
Dr. Anniina Vihervaara
KTH Royal Institute of Technology
Department of Gene Technology

Examination Board:
Dr. Fedrik Wermeling
Karolinska Institutet
Department of Medicine

Prof. Lars-Gunnar Larsson
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology

Prof. Mats Nilsson
Stockholm University
Department of Biochemistry and Biophysics

To my family, friends, and colleagues. Their care and support have made this possible.

Disclaimer: This thesis has been fully generated by natural intelligence

Popular science summary of the thesis

Cells are constantly sensing their environment and their own state. How cells respond to these signals depends on a variety of circumstances. However, it is crucial that they respond efficiently, as it determines their fate. In the case of organisms like us composed by many cells, that means that it will ultimately impact our health. Therefore, it is very important that our cells make the right decisions as fast as possible.

Just as we get better when practicing a task, some of our cells can learn how to react to external signals so they respond more efficiently the next time they are exposed. Even more, they can pass this knowledge to the next generation of cells so the daughters will also be more efficient responding. This is what we here refer to as (transcriptional) memory. Understanding how cells can remember could help us to use it to our benefit. But how do we figure out how cells remember?

Imagine that I wonder how you are able to write your name on a piece of paper. How can I figure out which parts of you are making it possible? That is correct, I can chop pieces of you until you are not able to write anymore. For instance, if I cut your foot, you will have no problems doing it, but if I take away your fingers, you won't be able to write. Thus, I will conclude that you were using your fingers to do it. But of course, you have many, many, many parts that I could remove. Taking one at the time and asking you to write your name for each would take too long. What we can do is to take many people, each of them missing one part, ask them to write their names and then see which parts are missing in those that are not able to. At this point, I would like to remind you that this is just an analogy, scientists are not going around mutilating people…or at least not anymore. Cells don't have feet or hands but also need certain tools to perform their tasks. We can take them away to find out which one was needed for a specific task. That is what I have done here to try to figure out which tools cells need to have memory. However, I am not trying to see how they can do something but how they can do it better. Coming back to the analogy, it is as if I want to know not how you can write your name but how you are able to react and write it faster if I ask you again. This makes everything a little bit more complicated…but we made it. We have found some tools that are important for the cell to learn how to react, the next step will be to see how these tools are used.

Cells keep the information about how to make and how to use these tools in an instruction manual called genome. We have used special technologies to read and modify that manual. A very common way of deleting parts of it is using the genetic scissors called

CRISPR/Cas. With these scissors, we can delete specific parts of the cell's manual so they cannot produce or use a specific tool anymore. However, these scissors are not perfect. Sometimes they don't cut properly or cut in places that they were not intended to. The mistakes are difficult to detect and can confuse the results when we are studying the cell's toolkit. In this thesis we have shown some of these mistakes, how they can sometimes go without being noticed and that they can be important because they may influence how the cell behaves. We also show a how these mistakes can be detected, which can be useful for other researchers.

Finally, we have included a diagnostic test that we developed during the Covid-19 pandemic. It detects the instruction book of the virus if it is present in the sample, by copying it many, many times. When it has been amplified so much, we can label it with fluorescence, or we can visualize it with other methods. It is similar to other tests that were in use, but the ingredients that we use can be produced in a very basic laboratory instead of buying them from a company. That makes it cheaper and independent on the supplies running low, which is very convenient in pandemic times. In addition, as it is designed to detect the instruction book of the virus and we scientists have become very good in figuring out how to do that, it can be adapted to detect other bugs if (or when) another pandemic comes.

.

Abstract

Transcriptional memory is phenomenon that has gained interest due to its potential impact on human health. With our work, we wanted to contribute to the understanding of its regulation. To do so, we performed genome-wide knockout screens in both budding yeast and human cells. For that, we developed experimental platforms selecting an easy-to-assay readout that allowed us to stratify mutants based on phenotype and scrutinize for mutants enriched or depleted in the corresponding strata.

Saccharomyces cerevisiae has been used for many years in genetics mainly because it is a simple unicellular eukaryote that is very easy to manipulate genetically. This made it a very useful model to study fundamental processes of gene expression in eukaryotes, such as transcriptional memory. We have discovered a new layer of complexity to the regulation of this phenomenon in yeast, based on mRNA stability.

Despite the great utility of yeast as a model, the translation of the results to humans is challenging. When studying human cellular models, the biology is closer to the real situation in a human being while keeping it still relatively easy to work with. However, there was still a big challenge: precise and easy genetic manipulation. The discovery of the bacterial CRISPR systems and its application to genetic manipulation of high eukaryotes has paved the path to a whole new era in research. It can be used to knock in or knock out genes or intergenic regions of interest, edition, overexpression, epigenetic modifications, and many more. It has potential not only for fundamental research but also for synthetic biology, gene therapy, diagnostics, personalized medicine, etc. In this thesis, we have used a CRISPR/Cas-generated genome-wide knockout pool to interrogate factors involved in transcriptional memory in human cells. We have identified two putative factors involved in the regulation of transcriptional memory in humans that are interesting candidates for further research.

To validate and further investigate candidates identified in pooled screens, knockout cell lines are frequently produced. A common method for that, is introducing two adjacent DNA double strand breaks (DSBs), using for example CRISPR/Cas with two gRNAs, that often results in the loss of the region in between. We have uncovered the occurrence of unexpected on-target aberrations while using a dual guide CRISPR/Cas system to produce deletion knockouts. We have shown that these events can go unseen and affect the phenotype of the cells. We proposed a workflow for comprehensive analysis of deletion clones.

Finally, in 2020, scientists across the word were urged to help in the SARS-CoV-2 pandemic. As the virus spread quickly, it was crucial to combine the therapeutic efforts with the development of diagnostic tools. The gold standard detection method, RT-qPCR is very sensitive and specific, but is expensive and requires specialized equipment not available in all contexts. In addition, the reagents providers promptly started struggling to meet the global demand. The scientific community responded developing a plethora of alternatives. Some of those methods are based on loop-mediated isothermal amplification (LAMP) of the viral RNA, which can be detected either by fluorescence or change in pH among others. Compared to RT-qPCR, this technic is faster and cheaper, with a simple readout that does not require specific equipment. However, it still required the extraction of the viral RNA, and the supply of related reagents was rapidly affected. Here, we developed a LAMP-based method to detect SARS-CoV-2 without any previous manipulation of the sample, using in-house produced enzymes. Our method performs comparatively to the commercially available options in terms of sensitivity and specificity, and it is compatible with the most commonly used sample carriers. All the plasmids for the production of the enzymes used are publicly available.

List of scientific papers

I. **Differential regulation of mRNA stability modulates transcriptional memory and facilitates environmental adaptation**
Bingnan Li, Patrice Zeis, Yujie Zhang, Alisa Alekseenko, Eliska Fürst, Yerma Pareja Sanchez, Gen Lin, Manu M. Tekkedil, Ilaria Piazza, Lars M. Steinmetz & Vicent Pelechano
Nat Commun, 2023, 14, 910

II. **CRISPR KO genome-wide screen uncovers new factors involved in trancriptional memory**
Yerma P. Sanchez, Xiushan Yin, Marc Friedländer, Vicent Pelechano
Manuscript

III. **Target-enriched nanopore sequencing and *de novo* assembly reveals co-occurrences of complex on-target genomic rearrangements induced by CRISPR-Cas9 in human cells**
Keyi Geng, Lara G. Merino, Linda Wedemann, Aniek Martens, Małgorzata Sobota, Yerma P. Sanchez, Jonas Nørskov Søndergaard, Robert J. White, Claudia Kutter
Genome Research 32.10 (2022): 1876-1891

IV. **Direct detection of SARS-CoV-2 using non-commercial RT-LAMP reagents on heat-inactivated samples**
Alisa Alekseenko*, Donal Barrett*, Yerma Pareja-Sanchez*, Rebecca J Howard, Emilia Strandback, Henry Ampah-Korsah, Urška Rovšnik, Silvia Zuniga-Veliz, Alexander Klenov, Jayshna Malloo, Shenglong Ye, Xiyang Liu, Björn Reinius, Simon J Elsässer, Tomas Nyman, Gustaf Sandh, Xiushan Yin, Vicent Pelechano
*these authors contributed equally to this study
Scientific Reports, 2021 Jan 19;11(1):1820

Scientific papers not included in the thesis

1. **Human prefoldin modulates co-transcriptional pre-mRNA splicing**
   Payán-Bravo L, Fontalva S, Peñate X, Cases I, Guerrero-Martínez JA, Pareja-Sánchez Y, Odriozola-Gil Y, Lara E, Jimeno-González S, Suñé C, Muñoz-Centeno MC, Reyes JC, Chávez S. Nucleic Acids Res. 2021 Jun 21;49(11):6267-6280

2. **Chromatin-sensitive cryptic promoters putatively drive expression of alternative protein isoforms in yeast.**
   Wei W, Hennig BP, Wang J, Zhang Y, Piazza I, Pareja Sanchez Y, Chabbert CD, Adjalley SH, Steinmetz LM, Pelechano V. Genome Res. 2019 Dec;29(12):1974-1984

3. **Tumor suppressor PNRC1 blocks rRNA maturation by recruiting the decapping complex to the nucleolus**
   Gaviraghi M, Vivori C, Pareja Sanchez Y, Invernizzi F, Cattaneo A, Santoliquido BM, Frenquelli M, Segalla S, Bachi A, Doglioni C, Pelechano V, Cittaro D, Tonon G. *EMBO J. 2018 Dec 3;37(23):e99179. doi: 10.15252/embj.201899179*

Contents

List of abbreviations

CRISPR	Clustered regularly interspaced short palindromic repeats
3D	Three dimensional
Cas	CRISPR-associated proteins
ChIP-seq	Chromatin immunoprecipitation coupled with sequencing
COVID-19	Coronavirus disease
CRAC	Crosslink and cDNA analysis
crRNA	Mature clustered, regularly interspaced, short palindromic repeats RNA
Cτ	Cycle threshold
CTD	Carboxyl-terminal domain
DNA	Deoxyribonucleic acid
DSB	Double strand break
eRNA	Enhancer RNA
EXPAR	Exponential amplification reaction
FACS	Fluorescence-activated cell sorting
gRNAs	Guide RNA
GTFs	General Transcription Factors
HLA	Human leukocyte antigen
HR	Homologous Recombination
IFN	Interferon
InDels	Small insertions and deletions
iRNAs	Interference RNAs
LAMP	Loop-mediated isothermal amplification
LRS	Long read sequencing
MACS	Magnetic-activated cell sorting
MERS-CoV	Middle East Respiratory Syndrome Coronavirus
MHC	Major histocompatibility complex
MNase	Micrococcal nuclease
mRNA	Messenger RNA
NASBA	Nucleic acid sequence-based amplification
NHEJ	Non-homologous end joining
NMD	Nonsense-mediated mRNA decay
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
pH	Potential hydrogen
PIC	Pre-initiation Complex
PML	Promyelocytic leukemia
POC	Point Of Care
RBPs	RNA-binding proteins
RCA	Rolling-circle amplification
RNA	Ribonucleic acid
RNA-seq	RNA-sequencing

RNApolII/III	RNA polymerase II/III
RPA	Recombinase polymerase amplification
rRNA	Ribosomal ribonucleic acid
RT	Retrotranscriptase
RT-qPCR	Reverse transcription-quantitative polymerase chain reaction//Real Time-quantitative polymerase chain reaction
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SDA	Strand Displacement Amplification
shRNAs	Short hairpin RNAs
siRNAs	Small (or short) interfering RNA//Synthetic interfering RNA
SLAM-seq	SH-linked alkylation for the metabolic sequencing of RNA
TALENs	Transcription Activator-Like Effector Nucleases
TMA	Transcription Mediated Amplification
tracRNA	Trans-activating CRISPR RNA
tRNA	Transfer RNA
UMI	Unique molecular identifiers
VTM	Viral transport medium
ZFNs	Zinc-finger nucleases

1 Introduction

1.1 Overview

Cells are exposed to a constantly changing environment and their survival greatly depends on their ability to respond to the challenge. Regulation of transcriptional responses is key to achieve a successful response. In any cell, specific sets of genes are expressed in response to internal and external stimuli. In some cases, previous expression events influence subsequent responses of the cell, creating specific expression patterns. The establishment of expression patterns influenced by previous events can be considered, in a broad sense, transcriptional memory. Since this establishment does not rely on genetic variations, it can also be termed epigenetic (transcriptional) memory. The broad term "transcriptional memory" is used within the literature referring to a myriad of phenomena, including transgenerational memory[1] and the maintenance of cell identity[2,3]. The main topic of this thesis is the **Transcriptional Memory** understood as the **transiently inheritable improvement of a transcriptional response due to repeated exposure to an external stimulus**[4,5]. For simplicity, in the context of this thesis, it will be referred to as just transcriptional memory.

Transcriptional memory has been identified in a variety of organisms[6–11], including humans[12–14]. It is well studied in its form of metabolic adaptation[15,16], but also affects both innate[17–19] and adaptive immune responses[20–24]. In addition, it could have a role in the acquisition of reversible tolerance during repeated antitumoral drug treatments[25,26]. Therefore, understanding the molecular bases of transcriptional memory and its regulation could offer great advantages for human health, particularly in immunology and cancer treatment. We intended to shed light into the molecular regulation of this process. To do so, we performed unbiased genome-wide screens in both yeast (Paper I, driven by Bingnan Li)[27] and human cells (Paper II, driven by Yerma P. Sanchez), for which we had to develop the appropriate respective platforms. During the development of screen in humans, we detected the presence of unexpected CRISPR/Cas9-induced on-target events that were locus and cell line independent. We developed a pipeline to uncover the occurrence of these on-target aberrations (Paper III, driven by Keyi Geng)[28]. In addition, we used our expertise in method development to provide a LAMP-based method to detect SARS-CoV-2 (Paper IV, co-driven by Yerma P. Sanchez)[29] in response to the emergency created by the pandemic in 2020. Although this last work is out of the main topic, we include it in the

thesis as I devoted significant amount of time and effort to it, and it greatly impacted my growth as a scientist.

To note, a greatly diverse methodology is used in the papers included in this thesis and it has been partially developed or adapted for the specific propose. Thus, it is described in detail in the corresponding papers and overviewed along the introduction and results sections in this *kappa* rather than in a dedicated section. For example, genetic screens are reviewed in section 1.3 and LAMP amplification in section 1.4.

1.2 Transcriptional Memory

When cells encounter the same stimulus repeatedly, the transcription of some inducible genes can occur faster or stronger the second time, increasing the efficiency of the response. This improvement is passed through mitosis to the daughter cells, that also respond more efficiently to a stimulus that they have never been exposed to (Fig. 1.1). This fascinating phenomenon, denominated "transcriptional memory", has been described in virtually any kind of organism in response to a great variety of stimuli [6–11,30].

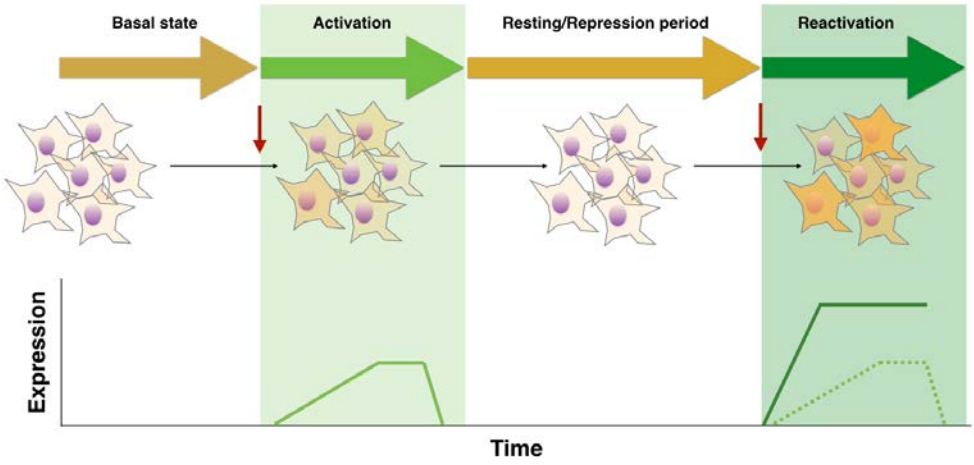

Fig. 1.1: Transcriptional memory as an enhanced response upon repeated stimuli. Cells at basal state (no induced expression) are stimulated (red arrow), which induce a first expression response (light green line). Cells are cultivated until recovery of the basal state and then re-stimulated, inducing an enhanced second response (dark green line, dotted line indicate the first response as a reference) that can be faster and/or stronger than the first.

Due to its impact on adaptation, homeostasis, and potentially in human health, it has been an attractive topic of research in the last decades. Here, I introduce two of the most prominent examples of transcriptional memory, the *GAL* model in *Saccharomyces*

cerevisiae (used in paper I) and the IFNγ/HLA-DR model in human cells (used in paper II). Next, I review some of the molecular mechanisms associated to transcriptional memory and their conservation across species.

1.2.1 Transcriptional memory in yeast: The Gal system

The budding yeast *S. cerevisiae* is an eukaryotic unicellular organism. These particular characteristics made it an extensively used model for studying fundamental processes, such as transcription, as it recapitulates the basis of biological processes of more complex eukaryotes while still being easy to manipulate. In addition, several instances of transcriptional memory have been found in the yeast *S. saccharomyces*, including the enhanced response of *GAL1*[7] and *INO1*[8] genes upon reactivation, or the faster expression of over 1000 genes when cells pre-treated with mild salt concentration are exposed to oxidative stress[31]. Therefore, it is not surprising that the majority of the knowledge in transcriptional memory has been collected from this organism.

1.2.1.1 The Gal regulon

The Gal regulon is a group of genes that encode the structural (*GAL1, GAL2, GAL7, GAL10, MLE*) and regulatory factors (*GAL4, GAL80* and *GAL3*) necessary for the import and metabolism of galactose in the yeast *S. cerevisiae*[32]. The regulation of these genes is the paradigm of gene expression. Gal4p is a transcriptional activator that constitutively binds to UAS_{gal} sequences present in the promoter of the structural Gal genes. In the absence of galactose, the transcriptional inhibitor Gal80p blocks the activation domain of the transcriptional activator Gal4p, preventing the recruitment of the transcription machinery and therefore the expression of the gal genes. Upon galactose addition, the ligand sensor Gal3p interacts with Gal80p inhibiting its action on Gal4p that then recruits the transcription machinery and coactivators, thereby initiating the transcription of the structural Gal genes[33–36] (Fig. 1.2).

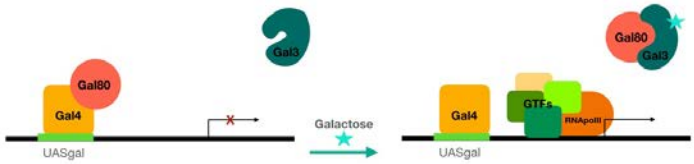

Fig. 1.2: Regulation of the Gal genes in non-inducing and galactose-induced conditions. Gal4 is inhibited by Gal80 in non-induced conditions (left). In the presence of galactose (right), Gal3 sequesters Gal80 abolishing its inhibition and Gal4 recruits the transcription machinery (RNApolII and general transcription factors).

However, *S. cerevisiae* uses glucose as a preferential carbon source. This preference is achieved both by transcriptional regulation of the Gal regulon and direct inhibition of the structural proteins. The presence of glucose inhibits the transcription and the protein galactose transporter Gal2, limiting the galactose available intracellularly. In addition, glucose promotes transcriptional repression of the inducer Gal3p and the transcription factor Gal4p, which results in tight inhibition of the Gal genes expression[37]. This double control system of induction and repression ensures the fine regulation of the yeast metabolism.

Altogether, the Gal system can exist in three different forms, depending on the available carbon source: Repressed (in the presence of glucose), induced (in the presence of galactose) or non-induced (in the presence of other carbon sources such as glycerol or raffinose).

1.2.1.2 Memory of galactose stimulation

When *S. cerevisiae* is cultured in galactose media, structural Gal genes such as *GAL1* start being expressed at mRNA level after ~5 min of galactose addition, with a peak expression at about 50 min[38]. In 2007, Kundu *et al*. observed that this expression levels were reached much faster in a subsequent galactose-mediated induction[7], with maximum expression after only 5 min (Fig. 1.3). The faster reactivation is especially evident when there has been a previous period of glucose repression[39]. Interestingly, this effect is not permanent, although persists for several cell divisions[40], which points to epigenetic regulation.

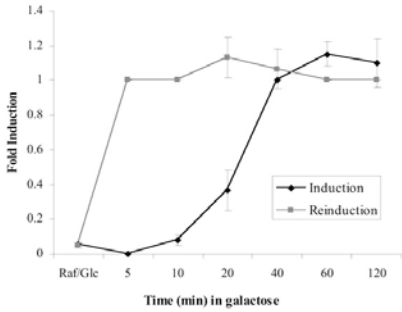

Fig. 1.3: Northern blot quantification (n=3) of induction (black) and re-induction (grey) kinetics of Gal1 upon galactose stimulation. For the first stimulation, cells were previously cultured in raffinose. For the re-induction, the system was repressed by culture in glucose. Figure from Kundu et al., 2007[7].

Although this is not the only example of transcriptional memory in yeast, the great understanding and the complexity of the regulation of the Gal genes, have made this system one of the preferred models to study transcriptional memory at the molecular level. In fact, during the last decades, several overlapping mechanisms have been found to contribute to the memory phenotype, as I will discuss in section 1.2.3.

1.2.2 Transcriptional memory in humans: The IFNγ-HLA-DR system

1.2.2.1 The HLA-DR complex

Human Leukocyte Antigen (HLA) genes encode for proteins of the family of the human major histocompatibility complex (MHC), which are the interface between the immune system and other cells in jawed vertebrates. Genes in the MHC cluster are divided in three categories, class I, class II and class III (Fig. 1.4). Class III genes encode for complement components and cytokines. Class I and II MHC are heterodimeric glycoprotein complexes present in the cell surface responsible of the antigen presentation to the immune system. The genes in this cluster are extremely polymorphic, producing person-to-person variation in the repertoire of MHC complexes of an individual. Antigens presented on MHC molecules are recognized by cells of the immune system and may elicit an immunological response depending on associated co-stimulatory signals, ensuring the elimination of damaged, infected, or exogenous cells as part of immunological surveillance[41]. MHCI molecules are expressed constitutively in almost every cell and is primarily recognized by TCD8+ cells[42]. MHCII molecules are only expressed constitutively in professional antigen-presenting cells such as B cells, macrophages and dendritic cells, that interact mostly with TCD4+ cells[43]. However, class II HLA expression in other cells might be induced in response to certain stimulation[44–46].

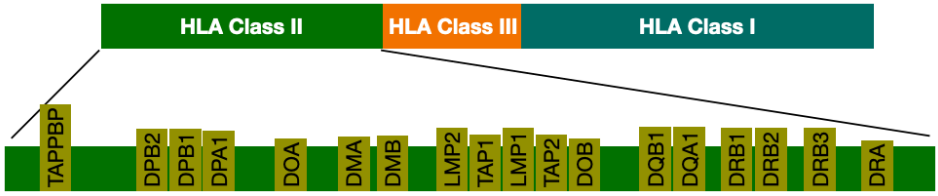

Fig. 1.4: HLA cluster containing Class I, class II and class III MHC genes. Zoom in the genes contained in the MHC class II locus, including HLA-DR genes.

HLA-DR is a MHCII complex (Fig. 1.4). It is composed by a constant alpha subunit that can bind to various beta subunits, encoded by HLA-DRA and HLA-DRB genes[47], respectively. In non-professional antigen presenting cells, these genes can be expressed in response to interferon gamma (IFNγ)[48] and the complex short is lived in the absence of stimulation[49]. IFNγ is a cytokine primarily secreted by T cells and natural killer cells that plays important roles in the regulation of the immune response. It exerts pleiotropic functions including antiviral and antibacterial immunity, enhance antigen presentation, macrophage activation, regulation of Th1/Th2 balance, among many others[50,51]. The IFNγ

receptor is a heterodimeric transmembrane complex composed by an alpha subunit (IFNγR1) that is the main responsible for the ligand binding, and a beta subunit (IFNγR2) that stabilizes the complex and is essential for the signaling transduction[52–54]. These subunits are constitutively associated to tyrosine kinases JAK1 and JAK2 through their cytosolic tail. Upon stimulation, the receptor tetramerizes (two alpha and two beta subunits), the JAK1/2 kinases self- and trans-phosphorylate and recruit STAT1, that is then activated by phosphorylation. Activated STAT1 homodimerizes and translocates to the nucleus, where it activates the transcription of over 200 genes containing the consensus binding sequence GAS (Fig. 1.5). Among those genes, IFNγ reception induces the expression of the transactivator CIITA[48]. CIITA is the master regulator of MHC II genes. Its binding to these genes is required for the expression of these HLA-DR proteins[55].

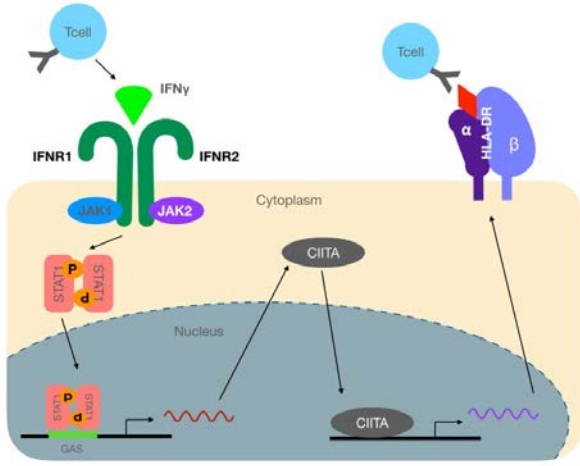

Fig. 1.5: IFNγ-induced expression of HLA-DR. Simplified pathway from reception of IFNγ to surface expression of the HLA-DR complex for antigen presentation.

1.2.2.2 Memory of IFNγ

Memory in human cells has traditionally been associated to the adaptative immune system[21–24,56]. More recently, it was discovered that cells of the innate immune system can be trained to respond faster to pathogens[13,18,19,57–59]. In 2010, Gialitakis *et al*. documented for the first time an example of transcriptional memory in non-immune human cells[14]. They showed that, when repeatedly exposed to IFNγ, HeLa cells induce the expression of HLA-DRA faster and stronger, which has been extensively confirmed afterwards (Fig. 1.6). In addition, HLA-DRA transcriptional memory of IFNγ perdures at least for 14 generations[60].

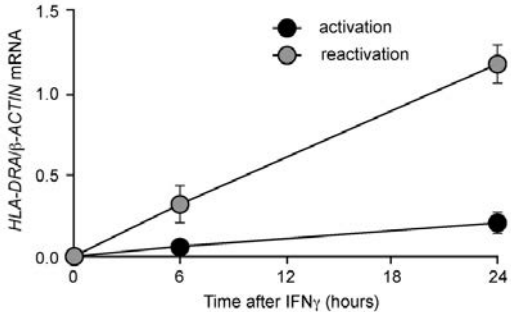

Fig 1.6: Memory in HeLa cells. Expression of HLA-DRA mRNA after induction with IFNγ in naïve (black) and primed cells (grey). Figure from Light et al. (2013)[61]

The similarity of the system to the one in yeast made it a popular model for the study of transcriptional memory at the molecular level. However, memory after repeated treatments with IFNγ is not restricted to MHC-related genes, but it has been also identified for enzyme-coding genes such as GBP4 and GBP5. In addition, other cell types have been shown to keep a memory of the IFNγ response[60], which could indicate that memory has a more generalize role in homeostasis beyond the specialized impact on immunological memory. Perhaps due to the influence of the previous knowledge in transcriptional memory in yeast, the molecular mechanisms of transcriptional memory in humans have not been systematically interrogated but rather assessed in comparison to the features of the models in yeast.

1.2.3 Mechanisms of transcriptional memory in eukaryotes

Despite the great variety of eukaryotic organisms, transcription is a highly conserved process. Thus, it is not surprising that some features of transcriptional memory are commonly found across evolutionary distant species. However, transcriptional memory is a complex process, result of multiple layers of regulation, that partially differs between species and gene models. Here, I review the most prominent molecular mechanisms of transcriptional memory and discuss their conservation across memory models, with particular focus on yeast and humans.

1.2.3.1 Chromatin-based memory mechanisms

Transcription is dramatically affected by the opening state of the chromatin, as it determines the accessibility of transcriptional regulators. Chromatin states are established upon stimulation and can be inherited through mitosis. Consequently, chromatin states have been deeply investigated in relation to transcriptional memory. There are three major

factors determining chromatin states: nucleosome composition, posttranscriptional modifications of histones, and the action of chromatin remodelers.

In transcriptional memory, the presence of the histone variant H2A.Z in the nucleosomes occupying the promoter of memory genes has been proposed as a contributor to their faster transcription. In yeast, it is essential for faster reinduction of both *GAL*1 and *INO1* memory genes[62,63]. The role of this histone variant in memory does not seem to be conserved in human cells nor mice[12,13]. Another histone variant, H3.3, has been related to memory in mouse fibroblasts[13]. This variant is in general associated with the inheritance of active expression patterns through mitosis[64], and it is also incorporated into IFN-inducible genes upon stimulation[65]. However, there are contradictory results whether it is the carrier of memory through mitosis.

Posttranscriptional histone modifications are one of the most accepted hallmarks of transcriptional memory. In particular, di-methylation of histone 3 at lysine 4 (H3K4me2) has been associated with poised promoters and epigenetic inheritance in almost all eukaryotic models tested, and it has been proven essential for transcriptional memory in yeast and humans[12,14,66–69]. Based on the *INO* model in yeast, this modification is predominantly established due to the absence of the Spp1 component in the histone methyl-transferase complex COMPASS specifically after the first induction, then it is recognized and maintained by *SET3* complex[66]. This mechanism has not been confirmed in the *GAL1* memory model. In humans, the homolog to the yeast COMPASS complex, MLL complex, has been proposed as the writer of H2K4me2 during transcriptional memory[66,70], although it has not been investigated whether, as in yeast, it has a memory-specific composition. In addition, how this histone modification is read and maintained during memory in human cells is unknown. Other epigenetic marks, such as H3K4me3 in plants[9] are important for memory, although their role is not fully understood. In addition, mutants in the SAGA histone acetyltransferase complex[71] are specifically affected in the secondary response during memory of the Gal genes in yeast, indicating that this modification could play a role in transcriptional memory.

Ultimately, the level of compaction of the chromatin depends on the action of chromatin remodelers, that can displace and evict nucleosomes leading to open chromatin states. Components of SWI/SNF chromatin remodeling complex have been found essential for the memory of the Gal genes in yeast[7]. However, the role of this complex on memory might be more related to their general function in transcription activation rather than being specific to re-activation[72]. There is not much information about the importance of this mechanism on the transcriptional memory of other genes in other organisms.

In addition to the level of compaction, 3D organization of the chromatin can have a huge impact on gene expression. For example, enhancers located at long distance in the genome can enter in physical contact with the targeted promoter forming a loop. Loops have been reported to support the faster reactivation of *GAL1* and *INO1* genes in *S. cerevisiae*[73], although they do not confer memory on their own[74,75]. Even if it has only been proven in yeast models, the lack of examples of gene loops as mechanism contributing to transcriptional memory might be due to the relatively more difficulty to assess 3D chromatin structure compared to other molecular features.

1.2.3.2 Nucleoporins and nuclear re-localization

Nucleoporins have been related to transcriptional memory in multiple models and organisms. These proteins are components of the nuclear pore complex. However, functions beyond nucleocytoplasmic transport have been attributed to them, including chromatin regulation[76–79]. Examples of nucleoporins with a role in transcriptional memory include Nup100, essential for *INO1* memory in yeast[8,80], and Nup98, necessary for memory of IFNγ in humans and memory of ecdysone in *Drosophyla*[12,81].

In yeast, memory genes are recruited to the nuclear pore after the first induction[82–84] and that leads to enhanced reactivation[62,71]. The localization at the nuclear pore persists for several generations even under repression, conferring enhanced reactivation to the progeny[62]. Multiple factors, including H2A.Z and components of the SAGA histone acetyltransferase complex and Sac3, are critical for the maintenance of these genes at the nuclear periphery and therefore for memory[71]. Nucleoporin-mediated relocation of these genes to the nuclear periphery is, thus, an important requirement for memory. However, while the disruption of the interaction with the nucleoporins completely abolishes memory for *INO1*[8], *GAL1* does not require the relocation to the nuclear periphery[39]. Contrarily, the resistance to oxidative stress acquired by NaCl pre-treatment requires Nup42[85]. In addition, artificial recruitment of the *INO1* gene to the nuclear pore is sufficient for faster reactivation[62], suggesting that it is the localization rather than the nucleoporin what could be the important factor for transcriptional memory. The idea that memory then could be related to a faster export is attractive, but the enhanced transcriptional response is usually measured as total mRNA abundance or even nascent transcription, which would be unaffected by subcellular localization. In addition, depletion of nucleoporins related to nuclear export Nup59 or Nup107 does not affect memory[70,85].This association to the nuclear pore has not been found in any other memory model organism so far. Thus, the

enhancement of transcription at the nuclear pore seems to be due to the associated components rather than faster mRNA export.

Although the relocation of the memory genes to the nuclear periphery seems a particular characteristic of yeast, there are other examples of relocation within the nucleus. After IFNγ treatment, the HLA-DRA locus relocates to PML (promyelocytic leukemia protein) nuclear bodies and remains there for several cell divisions, resulting in faster and stronger reactivation[14]. Despite Nup98 being essential for transcriptional memory of HLA-DRA[12], the role of the nucleoporin in the relocation has not been proven.

Altogether, it seems that transcription during memory is affected by the subcellular localization. In this context, nucleoporins could have a major role bringing together the memory genes and the memory-related transcription machinery but perhaps not a direct role in the faster reactivation itself.

1.2.3.3 Cis and trans promoter regulation

Transcription is regulated by *cis* and *trans* regulators. Specific binding motives at the promoter are *cis* regulatory elements, that can be bound by a corresponding transcription factor that is then acting as a *trans* regulator.

For *INO1* and *GAL1* genes, specific sequences in their promoters have been found to be essential for both, the primary relocation to the nuclear periphery and for the maintenance during memory[8,39,80]. It has been shown that the latest sequences are bound by a specific transcription factor during memory (Slf1 and Tup1 for *INO1* and *GAL1* genes, respectively) and both *cis* and *trans* regulators are essential for the faster reactivation[8,39,80]. Interestingly, the nucleoporins involved in the memory of these genes are required for the binding of the memory transcription factor. Thus, it is possible that the recruitment to the nuclear periphery leads to changes in chromatin accessibility that promotes differential binding affinities for activation or reactivation conditions. Unfortunately, neither these specific motives nor transcription factors have been yet found within the promoters of memory genes in other organisms. Nevertheless, recent studies showed that memory of IFNγ is promoted by a faster recruitment of the transcription factors STAT1 and IRF[86].

1.2.3.4 Transcription machinery

During transcription, RNApolII is recruited to active protein-coding genes together with some general transcription factors (GTFs), forming the pre-initiation complex (PIC). During PIC assembly, the carboxyl-terminal domain (CTD) of the RNApolII is

unphosphorylated. For transcription initiation, RNApolII is phosphorylated at Ser5 of its CTD. As it progresses to elongation, Ser2 becomes increasingly phosphorylated and phosphorylation at Ser5 is progressively removed[87,88] (Fig. 1.7). Thus, there are two time limiting steps for transcription activation: recruitment of the transcriptional machinery to the promoters and the actual transcriptional initiation. Accordingly, the enhanced transcriptional response in transcriptional memory could be the result of the persistence of the transcriptional machinery on recently active promoters, a faster recruitment of the pre-initiation complex (PIC), or faster release from the promoter during the second activation.

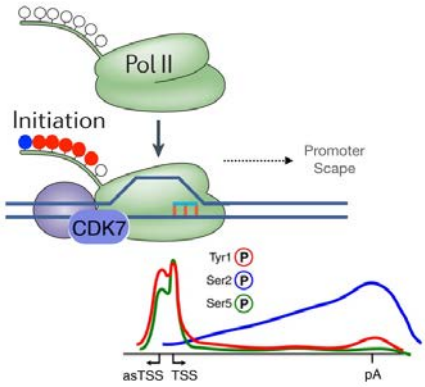

Fig. 1.7: Transcription initiation. Unphosphorylated RNApolII is recruited to the promoter to form the pre-initiation complex, then CDK7-mediated phosphorylation at Ser5 promotes its release from the promoter (top). As transcription progresses, levels of phosphorylation at Ser5 decrease while at ser2 increase (bottom). Modified figure produced combining images from Jeronimo at al. (2016)[89], Kim et al. (2022)[90] and Kuehner et al. (2011)[91]

The retention of RNApolII in recently active promoters can act as a mark for recent transcription[92]. Examples of the RNApolII persistence mechanism in transcriptional memory are found in yeast, plants and humans. In the case of *INO1* and HLA-DRA in *S. cerevisiae* and human cells, components of the PIC remind bound to the promoter in an inactive (unphosphorylated at Ser5) state after the first induction[8,12]. This inactive state is achieved by a differential composition of PIC-Mediator that maintains RNApolII in an unphosphorylated poised state, prevented from transcription but ready to start upon reinduction. In the case of memory genes in *Arabidopsis thaliana* subjected to repeated drought stresses, RNApolII is present at the promoters during the periods between stimuli but phosphorylated at Ser5[30]. Interestingly, in the *GAL1* memory gene promoter in *S. cerevisiae* the PIC is recruited *de novo* before reinduction, as the signal of RNApolII drops immediately after expression (within 20 min) and recovers after 2-4 hours of repression[39]. Thus, the presence of a form of PIC before reactivation contributes to transcriptional memory in different memory models, although the exact mechanism differs.

Alternatively to the presence of PIC before reactivation, faster transcription can be achieved by faster recruitment of the transcription machinery or faster promoter release. Some heat-responsive genes in mouse embryonic fibroblasts showed no accumulation but

faster release of the RNApolII when cells pre-conditioned by heat shock were re-activated[93]. Also, no sign of PIC was found in promoters primed by IFNß for faster reactivation, but there was faster and higher accumulation upon re-exposure[13]. In addition, RNApolII was associated to only a fraction of memory genes in monocytes primed with IFNγ[13]. Similarly, GBP genes in humans showed no accumulation of RNApolII prior to reexposure to IFNγ[60], pointing to a faster recruitment or release as a mechanism. Therefore, the presence of PIC before the second induction is not always a requisite for memory.

Altogether, it seems that the specific mechanism by which the second round of transcription occurs faster in regard to the transcription machinery is promoter-dependent.

1.2.3.5 Cytoplasmic inheritance

Expression patterns are frequently maintained by feed-forward loops. Indeed, a common mechanism to regulate expression cascades is by looped genetic circuits. This is, one of the proteins generated in the pathway is a positive (for amplification) or a negative (for buffering) regulator of upstream factors.

In the context of transcriptional memory, this has been thoughtfully investigated in the Gal system. As discussed in the previous section, the expression of *GAL1* gene depends on the activator Gal3p[94]. However, Gal1p is able to exert the inhibitory functions of Gal3p on Gal80p, leading to transcription[95]. It has been shown that after the first induction, Gal1p accumulates in the cytoplasm, is inherited through mitosis, and leads to faster reactivation in the daughter cells[6,96]. In fact, it has been proposed to be the major contributor to short term transcriptional memory of the *GAL1*[72,97].

The accumulation of transcription factors after the first response has also been proposed in *A. thaliana* as a contributor to increased ability to cope with drought stress[30,98,99]. In humans, this memory mechanism has not been thoughtfully investigated. Early studies showed no differences in the expression of the master regulator of HLA-DR genes CIITA, nor in the abundance and kinetics of STAT1 phosphorylation between primed and naïve cells[14]. However, it is possible that the results were limited by the resolution of the techniques used. In general, the contribution of cytoplasmic inheritance to transcriptional memory might be underestimated due to the focus of most studies on chromatin-related factors.

1.3 Phenotype-based pooled genetic screens

A classical approach in genetics to study a phenotype is to isolate cells (or individuals) with such characteristic and pinpoint which genes are affected in them. This process can be laborious, slow and costly. Alternatively, genes can be specifically targeted to measure their effect on a phenotype. However, this becomes exponentially harder as the number of genes to study increases. Previous knowledge can guide the selection of genes to be target, but sometimes that knowledge might not be available, or the aim of the study might be to discover unexplored gene-phenotype relationships. In phenotype-based pooled genetic screens, multiple individual loci (or their products) are specifically targeted in a way that a traceable record of the modification remains in each variant. Using such approach, big sets of cells carrying different genetic variants can be selected in pool for a particular phenotype, and the perturbations responsible of that phenotype can be easily determined. This approach enables to interrogate thousands of gene products in time and cost-efficient manner without the need of previous knowledge. Recent technological advances that facilitated the performance of pooled screens have resulted in an explosion of this type of studies.

There are three critical steps in the performance of a pooled genetic screen: generation of the cell library with the genetic variants, selection of the cells with the phenotype of interest, and identification of the perturbations leading to the differential phenotypes (Fig. 1.8). Here, I review the most relevant technological advances in each of these steps.

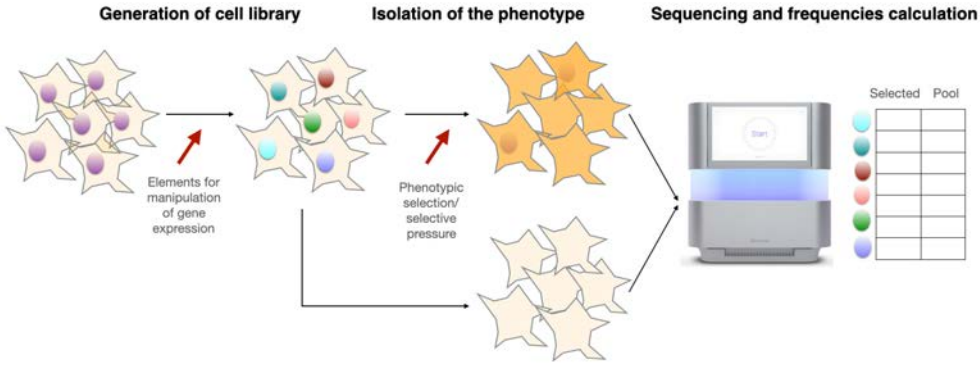

Fig. 1.8: Typical workflow of a pooled genetic screen. First, a cell library of gene expression variants (generally genetic variants) is generated. Then, the cells with the phenotype of interest are selected. The traceable record of the variants is sequenced to calculate and compare their frequencies in the selected subpopulation to the original pool.

1.3.1 Generation of the cell library

Although overexpression and other genetic manipulations are valuable to study gene function, the first approach is usually to investigate the effect of its loss of function. For that, the expression of the gene is impaired, either at the transcriptional or post-transcriptional level in the case of protein coding genes (Fig. 1.9). Usually, the gene is eliminated or replaced, its sequence is altered to produce a premature stop codon, or its translation is inhibited by interference.

1.3.1.1 siRNAs

Small interference RNAs (siRNAs) are short RNA molecules that target mRNA by sequence complementarity and inhibit their translation by cleavage, after being processed and loaded into the RISK complex. Synthetic iRNAs (also siRNAs) can be designed and introduced in cells to target specific mRNAs hijacking the physiological machinery[100]. As they are easy to produce and applicable to a huge range of organisms, they have been extensively used to knockdown protein expression.

This technology has been a revolution and initiated a whole era of massive pooled genetic screens[101]. For pooled genetic screens, siRNAs are frequently delivered through transduction with retroviruses in the form of short hairpin RNAs (shRNAs). Each viral particle is packed with only one shRNA and the cells are infected with a relative low number of them, ensuring that each cell will be infected with a maximum of one virus. Once infected, the shRNA is retrotranscribed and integrated in the cell genome. In this way, it can be expressed to inhibit the synthesis of the targeted gene product, while it serves as a traceable barcode of the perturbation in that cell.

As they are nucleic acids, siRNAs are easy to design and synthetize against virtually any mRNA. In fact, this strategy has been applied to perform genetic screens in various organisms[102–104]. However, the complementarity of the sequence does not directly correlate with the efficiency of the inhibition. The siRNA targeting mechanism is complex and is a developing field in its own[104,105]. Targeting of mRNAs with siRNAs usually results in incomplete depletion of the target protein. Although partial depletion can be useful to study essential genes, it also leads to heterogeneity in the level of interference. In addition, siRNAs are prompt to off target effects[106]. Consequently, siRNA-based pooled screens can be noisy and difficult to interpret. Nevertheless, they have been extensivelly used until the appearance of more precise tools for gene editing that could be applied in large genetic screens.

1.3.1.2 Direct homologous recombination

Alternatively to translation inhibition, genes can be modified at the DNA level. To do so, frequently the own cell machinery to repair DNA double strand breaks (DSBs) is harnessed. There are two major pathways to repair DSBs: homologous recombination (HR)[107] and non-homologous end joining (NHEJ)[108]. While NHEJ religate the ends of the DSBs without error correction, a homologous sequence is used during HR to fill the gap. Thus, HR is error-free, which is desirable for precise gene editing.

While HR is a common pathway in eukaryotes, in very few organisms it can be directly exploded. The budding yeast *S. cerevisiae* undergoes spontaneous homologous recombination efficiently with only 35 nucleotides homology sequences. Typically, a cassette containing a selection marker or auxotrophy is designed to target and replace a specific locus by flanking it with the homologous sequences[109,110]. In addition, these cells can grow in haploid or diploid state, which facilitates the isolation of particular genotypes. Being easy to manipulate, *S. cerevisiae* has been extensively used in genetic studies and there are genome-wide collections of mutants since decades that can be used to perform pooled phenotypic screens[111–113]. In this thesis, we have used one of those collections in paper I[112].

In higher eukaryotes, the efforts have focused on directing the production of DSBs to the specific loci and afterwards either rely on NHEJ to produce the mutation or redirect the repair to the HR pathway.

1.3.1.3 Zinc finger and TALE nucleases

Zinc finger nucleases (ZFNs) are engineered proteins in which a non-specific nuclease (FokI) is fused to a zinc finger domain that recognizes specific DNA sequences[114,115]. When two finger domains recognize the complementary sequences in the targeted DNA locus, the nuclease is activated by dimerization and produces a DSB[116]. In most eukaryotic organisms this will trigger the NHEJ repair pathway, that may resect the ends before religation resulting in small deletions or insertions (InDels)[117]. If the InDels are not multiple of three nucleotides, they will shift the translation reading frame of the resulting mRNA, which will likely produce a premature stop codon. Those faulty mRNAs will be degraded through non-sense mediated decay (NMD)[118].

In ZFNs, the finger domains are usually composed by 6-12 fingers, each of them recognizing three nucleotides. In total, a motive of 18-36 nucleotides can be recognized, which confers high specificity[119,120]. Although these domains can be customized to target virtually any locus, the process is tedious and costly. A further improvement of ZNFs

technology fused a modular DNA binding domain derived from transcription activator-like effectors (TALE) to the nuclease (TALENs)[121]. The modular nature of the DNA binding domain made TALENs a more flexible tool as compared to the ZFNs, being able to freely design the targeting based on a protein-DNA code[122]. However, they are still difficult to upscale to high-throughput screens. Thus, although TALENs made possible to edit genomes previously hard to manipulate and the expected range of applications was growing, the discovery of an even more flexible tool, CRISPR/Cas, relegated them out of the spotlight of gene editing. Nevertheless, the interest in TALENs is resurging due to their high specificity.

1.3.1.4 CRISPR/Cas

Similar to ZFNs and TALENs, CRISPR/Cas consists of a nuclease and a sequence recognition component, and relies on the cell repair machinery to resolve the induced targeted DSB. CRISPR/Cas is a naturally occurring prokaryotic system whose recognition component is purely RNA. Thus, this technology is very easy to customize, scale and parallelize[123].

In bacteria and archaea, CRISPR/Cas functions as an adaptative immune system[124–128]. CRISPR is a genomic array composed by genome-targeting sequences (spacers) in between identical repeats[129–133]. When a cell is infected by a bacteriophage or plasmid, short fragments of the foreign material (protospacers) are integrated in the CRISPR array, keeping a record of the infection. This array is expressed and processed into CRISPR RNAs (crRNAs), which are loaded into the CRISPR-associated nuclease (Cas). Upon reinfection, the ribonucleoprotein is guided towards the specific viral sequence, which is then efficiently cleared[134–136]. This system has been adapted and heavily used for gene editing in previously hard to manipulate genomes.

There are a variety of CRISPR/Cas systems with different properties and processing mechanisms[137]. The first and most commonly used is CRISPR/Cas9[138–141]. Cas9 requires a trans-encoded RNAs (tracRNA) complementary to the repeats to efficiently process pre-crRNA to crRNA[142]. In addition, it requires the presence of a short motive next to the complementary target region (protospacer adjacent motif, PAM)[143]. As the protein and tracRNA sequences are constant, the only element that needs to be adapted for a specific target is the crRNA by sequence complementarity. Several factors, however, affect the efficiency of the crRNA. The most important constrain for the efficiency of Cas9-mediated DSB is the presence of the PAM 3'-NNG in the target region, preferentially preceded by thymines and no cytosines[144,145]. Another consideration is the specificity of the RNA-

guided nuclease activity. Off-target DSBs might occur as a consequence of partial complementarity of the crRNA to the off-target sequence. Further studies uncovered the relevance of the position of the mismatch for the occurrence of the off-target event[146–152]. The increasing knowledge on the molecular bases of CRISPR/Cas efficiency and specificity led to multiple improvements, including Cas9 variants and modifications of the crRNA[144,150–161]. Nowadays, several experimental approaches and predictive algorithms are available for rational crRNA design. Furthermore, there is a great deal of already validated crRNAs, including libraries against whole genomes in several organisms[162]. One of those libraries[163] was used in this thesis for paper II.

Besides the design of the crRNA, other factors are key for the success of CRISPR/Cas-mediated engineering. One important consideration is the delivery of the components. High-throughput experiments are particularly sensitive to low efficiency in the delivery, as it might compromise the coverage of the crRNA library. Conveniently, the RNA components may be provided already combined in a single molecule (gRNA), in the form RNA or a DNA expression cassette. Cas can be delivered as expression cassettes or as a protein. Alternatively, the already assembled ribonucleoprotein complex can be provided[164]. In pooled genetic screens, the gRNA expression cassettes are usually delivered for insertion using retroviruses, as explained for siRNAs. This ensures high efficiency while serving as a tracer of the introduced genetic perturbation. The Cas nuclease can be also integrated, or transiently provided. For high-throughput screens, arrays of validated gRNAs are available, simplifying the production of cell libraries.

As easy to use and readily adaptable tool, CRISPR/Cas has revolutionized the field and has democratized the performance of pooled screens. In addition to point mutations to produce premature stop codons, a dual gRNA-Cas can be directed to produce two adjacent DSBs, which results in the deletion of an entire locus. This approach is useful to study the function of non-coding genes and regulatory regions. We used this strategy in paper III. But the flexibility of the CRISPR/Cas system goes beyond the production of loss of function mutations. Catalytically inactive Cas proteins can be fused to protein domains with other activities, expanding the toolkit to a broad range of gene expression manipulations[165–167]. In addition, as our understanding in the regulation of HR grows, the efficiency of deriving the repair of the DSB towards this pathway is increasing, opening the door to the application of CRISPR/Cas to more precise gene editing[168–170]. Seemingly, the only limit nowadays is our own imagination. However, unexpected events upon repair have been described, limiting current applicability of CRSIPR/Cas in medicine. We discuss some of those limitations in paper III.

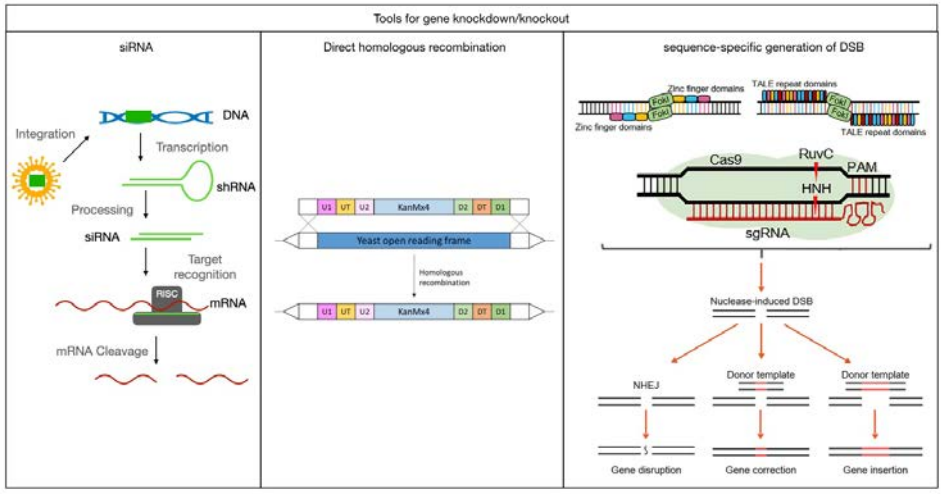

Fig. 1.9: Schematic overview of the most popular strategies for manipulation of the expression of target genes, explained in sections 1.3.1.1-1.3.1.4. Images from Zhang et al. (2019)[171] and Vanderwaeren et al. (2022)[172] are part of this figure.

1.3.2 Phenotypic cell isolation

After generating a cell library of genetic variants, the next step is to identify which of them are phenotypically affected. As in any phenotype-based experiment, choosing an appropriate readout is crucial for a successful pooled genetic screen. Additionally, the subpopulation in the pool with the phenotype of interest must be isolated to determine the genetic variants within it.

The first and most common type of pooled phenotypic-based genetic screen tests viability as a phenotype. This is the simplest strategy, as the cells do not need to be physically separated. However, other more sophisticated screens are performed using endogenous markers or constructed reporters to identify and isolate the cells with the phenotype of interest.

1.3.2.1 Dropout experiments

In the simplest version of dropout experiments, the modified cells are cultured without any other perturbation and the proportions of the variants are measured at consecutive time points. In the course of the experiment, the least fit variants will be outcompeted by the highly proliferative ones. These studies have been very useful to determine gene essentiality in tumoral cells, identifying potential targets for treatments[173–176]. In addition,

this type of screens provided knowledge about the inherent dynamics of the pooled mutants that is applicable to other experimental designs[177,178].

Another layer of information is generated when combined with additional perturbations, such as drug, additional mutations or infection challenge[179–182]. For example, dropout experiments on tumoral cells treated with an antitumor drug, reveal genes responsible of emerging resistances, which is key for the rational design of combinatorial therapies. Applied *in vivo*, this approach can for example be used to discover genes important for homing, tumor invasion or metastasis[183–185].

1.3.2.2 Markers and reporters

In theory, any phenotypic trait can be interrogated in a pooled screen as long as cells can be assessed individually, and the subpopulation of interest can be isolated. Often, molecular phenotypes are traced using a protein or an RNA that correlates with the studied process. This molecule might be endogenously produced or artificially imposed by a reporter system. These molecules are then used to separate the population of interest.

Multiple options exist for separation[186–188]. The great development and increased accessibility of fluorescence-activated cell sorting (FACS) has made it the preferred choice in pooled genetic screens. FACS combines fluidics and optic systems for precise analysis and separation of each individual cell. First, individual cells are englobed in fluid drops and exposed to laser dim of specific wavelength and the diffraction or fluorescence are recorded. Then, the cells with the phenotype of interest (i.e., with specific light scatter and/or fluorescence level) are separated from the stream by an electromagnetic field[189,190]. Thus, FACS can separate cells based on surface or internal molecules, although those must be optically detectable. In reporter systems, this is usually achieved by the synthesis of fluorescent proteins, like GFP and YFP, or adding fluorescent tags to existing proteins. Endogenous markers can be labeled using florescent antibodies or probes. Therefore, reporter systems may simplify downstream sample processing circumventing the need of finding suitable antibodies or other labelling agents. However, it is an additional perturbation to the cell and the possibility of altered regulation compared to the original genomic context must be contemplated.

Despite its supreme utility in pooled screens, FACS separation still requires specialized equipment and expertise. In addition, the flow technology is relatively slow and expensive. Alternatively, cells can be separated without fluorescence labelling using antibodies attached to magnetic beads. Magnetic-activated cell sorting (MACS) can handle higher throughput than FACS, making it cheaper and faster. However, the separation is not as

clean as in FACS and only the positive fraction can be obtained with acceptable purity for a screen. Besides, MACS separation is restricted to separation based on surface proteins, while FACS can be used for internal molecules as well.

1.3.2.3 Non-selective screens for continuous phenotypes

Until recently, pooled screens where mainly reduced to discrete phenotypes, or continuous traits were categorized to allow the phenotype-based separation. For instance, when assessing gene expression usually cutoffs are set, which implies missing the dynamics of the responses. To address this challenge, methods as Perturb-seq[191] have been developed, that allows to assess the transcriptome of each cell and associate it to its genetic variant. This approach enables to relate each perturbation with a phenotype measured in a continuous scale, without the need of threshold-based separation. It is expected that more of these methods expand the applicability of pooled screens.

1.3.3 Identification of the genetic variants

Once the cells with the phenotype of interest have been isolated, the genetic variations responsible of the phenotype must be identified and quantified. For that, the traceable barcodes introduced during the production of the edited pool are sequenced. For instance, in the case of CRISPR/Cas-generated libraries, the integrated guide is sequenced and quantified as a proxy for the abundance of each genetic variant. The great development of sequencing technologies has allowed increased accessibility to this approach. In particular, short read sequencing on Illumina platform has become extremely easy to use as well as cost and time efficient, making it ideal to combine with pooled genetic screens.

1.3.3.1 Target sequencing

To enable simultaneous amplification and library preparation of all the genetic variants, in pooled screens the traceable barcode is introduced in the genome of each cell flanked by common sequences. These sequences are then targeted for amplification by polymerase chain reaction (PCR). As these loci represent a small part of the total DNA of the pool, sequential PCRs are usually performed to gradually enrich the target over the genomic DNA. In addition, during these PCRs grafting sequences needed for sequencing are incorporated (Fig. 1.10).

Illumina platform is based on sequencing by synthesis. This is, the molecules loaded will be used as template for DNA synthesis, while the incorporation of each nucleotide is recorded to decipher the sequence. During each cycle, the incorporation of one new

fluorescently labelled nucleotide is optically detected. The great advance of the Illumina technology is the formation of clusters of each molecule by bridge amplification (Fig. 1.10) before the sequencing cycles, which leads to enhanced signal and therefore dramatically reduced cost. This in turn has made the performance of high-throughput screens much more accessible.

Once the sequences are obtained, they can be mapped to the database of gRNAs or siRNAs introduced, and their proportions in the selected pool can be compared to the original population in order to detect depleted or enriched variants. As CRISPR/Cas screens are currently the most widely used, there are several available tools for the most common analyses[192–194]. Other setups, though, might require customized pipelines.

1.3.3.2 Amplification bias

Targeted sequencing involves heavy processing of the DNA sample, including massive amplification. It is, therefore, susceptible of errors and biases. One of the most common biases occur during the PCR amplification. First, shorter molecules are more efficiently amplified, leading to apparent overrepresentation after sequencing. Second, all the molecules might not be amplified every round. The most abundant are more likely to start being amplified earlier, leading to exponential increase of the abundance differences. In pooled screens, the first is not a problem as the amplicon is usually approximately the same size. However, big differences in the frequency of each variant can lead to the second effect, especially if the complexity of the sequencing library is limited.

One way of minimizing this bias is reducing the number of amplification cycles. In the case of pooled screens this is not easy as the proportion of the amplicon to the rest of the genome is very small. Instead, the material is normally split in several amplification reactions, especially the first PCR, to randomize the initiation bias. In addition, short random nucleotide sequences can be incorporated in the amplicon to serve as a unique molecular identifier (UMI). This sequence will be copied in each round of amplification and can therefore be used to collapse PCR duplicates. Furthermore, the introduction of UMIs in CRISPR/Cas screens has been shown to reduce the amount of cells needed and, consequently, the cost and time of the procedures[195,196].

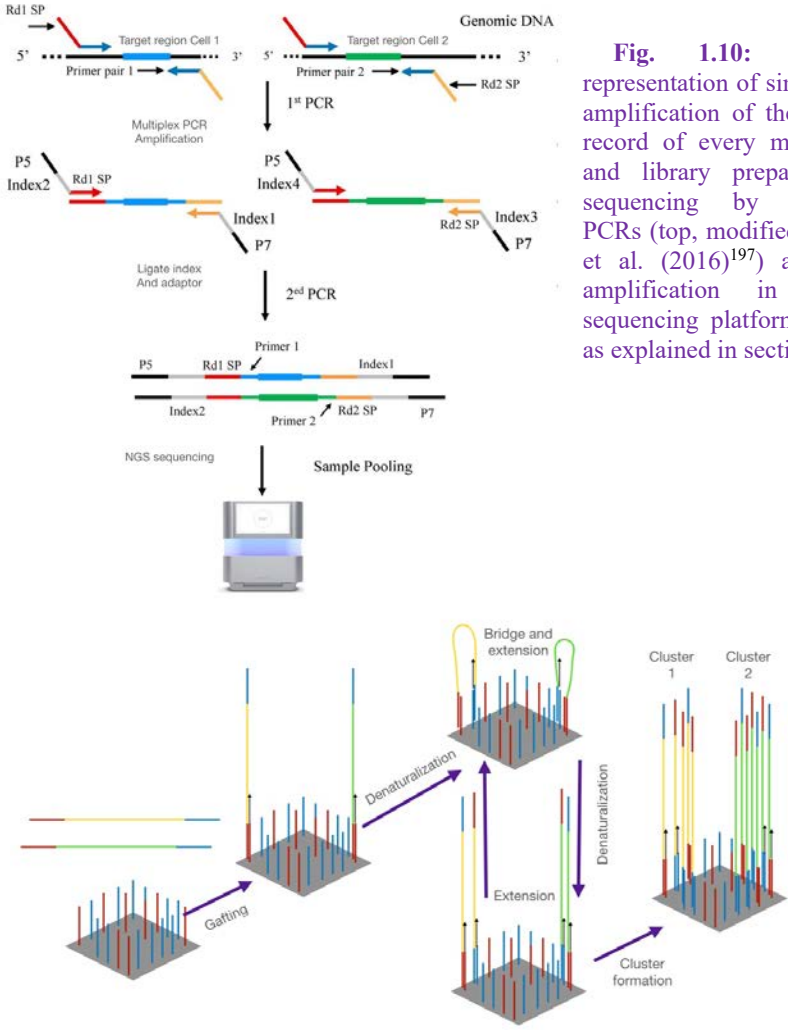

Fig. 1.10: Schematic representation of simultaneous amplification of the traceable record of every modification and library preparation for sequencing by successive PCRs (top, modified from Liu et al. (2016)[197]) and bridge amplification in Illumina sequencing platform (bottom) as explained in section 1.3.3.1

1.4 SARS-CoV-2: The pandemic emergency

In March 2020, the World Health Organization officially declared the pandemic state of the SARS-CoV-2 viral epidemy. The infection started in Wuhan (China) the previous December 2019 by zoonosis, although the animal of origin is still unclear, and quickly spread all around the world leaving a trace of death and economic damage behind.

The effects of the infection rank from completely asymptomatic to death, although most patients develop cold-like symptoms in variate intensities, typically fever, fatigue, respiratory and gastrointestinal issues[198]. Still, the high infectivity of this pathogen overwhelmed the health care systems around the world. In the lack of effective treatment

or vaccines at the beginning of the pandemic, detecting and isolating infected people became crucial. Severe lockdowns were applied in multiple countries, dramatically affecting their economies. Yet, the virus kept rapidly spreading between people and across countries. Up to June 2023, more than 700 millions cases and almost 7 millions deaths had been registered[199], and these sums do not account for undiagnosed cases nor the indirect deaths derived from the collapse of the health care infrastructure.

Luckily, nowadays the alarm rose by SARS-CoV-2 has decreased, as a combination of prophylactic and interventional strategies have succeeded. This success has been possible thanks to an unprecedented joined effort of the scientific community that from the beginning of the pandemic focused on adapting and developing new technologies to detect, prevent and treat SARS-CoV-2 infections.

SARS-CoV-2 is a coronavirus. Coronaviruses are a big family of enveloped viruses that infect birds and mammals[200,201]. Seven coronaviruses are known to infect humans, of which the most virulent are MERS-CoV, SARS-CoV and SARS-CoV-2 [202]. Coronaviruses are positive single strand RNA viruses[203,204]. Thus, their genome can be directly translated by the cell machinery once it is infected, as they resemble the host mRNAs. The coronavirus genome is around 30 kb and contains two overlapping open reading frames (ORF1 and ORF1ab) encoding non-structural proteins, and several smaller ORFs encoding structural proteins and some non-structural proteins. The replication is complex and not fully understood[205,206].

Coronaviruses have a moderate mutational rate and have the ability to recombine their genome upon heterologous co-infection[207,208]. Thus, there is great concern of upcoming coronavirus pandemics. The technologies developed during the SARS-CoV-2 pandemics, as well as the communication networks, emergency protocols and infrastructures established during the pandemic, might be leveraged in the expected case of future viral pandemics. Here, I review some of the most prominent methods used to detect SARS-CoV-2 during the pandemic and discuss their potential applicability for the detection of other pathogens.

1.4.1 Molecular methods for SARS-CoV-2 detection

Viruses are composed by nucleic acids packed into a protein capsid. Thus, either their genomic sequence or proteins can be targeted for detection. During the pandemic, a variety of methods were developed or adapted to detect these sequences and proteins of SARS-CoV-2. Those methods differ in their technical complexity, cost, specificity, sensitivity, and speed. Frequently, there is a tradeoff between these characteristics. Although regular

diagnosis prioritizes sensitivity and specificity, the rapid expansion of the virus pushed the need towards faster and cheaper alternatives that allow massive testing. In fact, massive screening has been demonstrated as a powerful tool to prevent the virus spread[209–211]. In addition, methods that can be simplified in their performance, can be used by non-specialized personal in a no-laboratory environment as point of care (POC) solutions, allowing massive testing without saturating the specialized infrastructures.

Samples are collected normally by nasopharyngeal or oropharyngeal swabs, although sputum and saliva can be also collected for test[212].The sample can be directly tested (in case of POC) or transported to the testing facilities, frequently in a viral transportation media (VTM). VTM are generally composed by salts, a carbon source, antibiotics, serum and a buffering component[212]. Although the use of VTMs has been proven not really necessary[213], they are still broadly used. Thus, their components must be accounted when developing detection methods.

1.4.1.1 Methods based on Protein detection

Methods detecting proteins rely on antibodies (immunoassays). Antibody-based detection is a very well stablished method that has been extensively used for a variety of applications. Antibodies can be produced against virtually any epitope able to trigger humoral response. However, the process of selecting a specific and sensitive antibody can be slow and costly. In addition, mutations in the target proteins can reduce or abolish the target recognition strength and force to redesign the antibodies. The greatest advantage of this strategy, though, is that it can be easily implemented into a lateral flow test. In these tests, the sample is applied in one end of the device and travels by capillarity towards a membrane where the detection antibodies are fixed into a polymeric strip. An additional strip targeting a common antigen in the sample is included as a control. The recognition of the antigens triggers a reaction that can be visualized, usually by a color change. This approach is the base of most home and POC tests. Although less sensitive than laboratory tests, lateral flow tests are easy to use, quick and relatively cheap to produce.

This strategy was promptly adapted to detect SARS-CoV-2 into the popular antigen tests, usually targeting the spike proteins present in the surface of the viral particles. It has allowed high throughput testing without the need of special equipment or expertise, which resulted extremely useful for the rational application of quarantines. During the course of the pandemic several variants of the virus emerged, and all of them had mutations in the spike proteins. Although some tests have reduced sensitivity to some variant, none of the

tests approved by the FDA have completely failed to detect the virus yet[214]. There is no doubt that this approach will be useful in the case of upcoming pandemics.

1.4.1.2 RT-qPCR

Real time quantitative PCR (RT-qPCR) is the gold standard method for nucleic acid-based diagnosis in clinical settings[215]. It measures fluorescence emitted during the amplification through PCR of a DNA molecule. In addition to DNA detection, RT-qPCR can be applied to detect and quantify specific RNAs by adding a previous retrotranscription step. For PCR amplification, the DNA molecule is denatured by high temperature and the single strand sequence of interest is targeted by a primer reverse complementary to the 3' end. For annealing of the primer and further polymerization, the temperature must be reduced. Then, the complementary strand is synthetized which, upon denaturalization, is targeted by a primer with the same sequence than the 5' end than the original sequence of interest (reverse complement of the 3' end in the newly synthetized strand). Thus, in the next cycle, both strands will serve as a template for polymerization. This leads to exponential amplification along the temperature cycles. In RT-qPCR, this amplification is detected in real time due to the incorporation of dyes that bind double-stranded DNA during the reaction, or by using fluorescent probes targeting the specific DNA sequence of interest[216,217]. The fluoresce at each amplification cycle is directly proportional to the amount of DNA in the sample. As the fluorescence curve is exponential, the point where the exponentiality ramps is frequently used as a reference to compare samples, although this threshold can be customized. The number of cycles necessary for the fluorescence to reach that threshold (Cτ value) is inversely proportional to the amount of target originally present in the sample and can be used to quantitatively compare samples, usually test vs control samples[218]. The use of positive and negative controls is essential to ensure reliability and detect potential spurious amplification. This method is extremely sensitive and specific, although multiple factors may result in sample variability[219]. Thus, stringent requirements for sample standardization and calibration are needed for precise quantification across specimens.

As it only requires the design of adequate complementary primers, RT-qPCR has been applied for the detection of a plethora of pathogens for decades. During the pandemic, RT-qPCR has been the standard for diagnosis of SARS-CoV-2, with recommendations and protocols promptly issued by the health organizations. Several kits became available for simultaneous detection of the virus and sample loading controls in one-pot RT-qPCR reactions. Improvements in the pre-processing of the sample were crucial to increase the

applicability of RT-qPCR. The most important, multiple protocols were efficient on detecting the virus on unpurified samples[220–223], bypassing the expensive and time-consuming RNA extraction step. However, quantification of unpurified samples is challenging, as the samples come with a variable presence of other components. Consequently, these tests usually only deliver a qualitative result.

Despite the improvement, RT-qPCR still requires specialized instruments for cycling temperatures and fluorescence detection, as well as certain level of expertise to be performed. In addition, the fast-growing number of samples led to a scarcity of reagents needed for this test. In response, alternative detection protocols were developed based on isothermal amplification.

1.4.1.3 LAMP

Isothermal amplification methods use alternative strategies to amplify DNA without cycling temperatures, which enables their application without the use of thermocyclers. One of the most popular isothermal amplification methods is loop-mediated isothermal amplification (LAMP), which we use in paper IV.

Like PCR, during RT-LAMP the viral RNA is first retrotranscribed and then the DNA sequence of interest is targeted by a primer reverse complementary to the 3' end that is used for the synthesis of the reverse complement (Fig. 1.11). However, this primer carries an overhang sequence reverse complement of a downstream sequence to the priming site (forward internal primer)[224]. Thus, when the complement is synthetized and single stranded, this overhang anneals to the complementary sequence, creating a loop containing the original targeted sequence, and serving as primer for further polymerization. The same occurs by targeting the resulting reverse complement on the other end with an analogous primer with overhang (backward internal primer), creating two species of DNA molecules, one in each orientation. For the formation of the loops and further propagation, the molecules must be single stranded and therefore separated from the template. While in PCR this is achieved by temperature-mediated denaturalization, LAMP uses polymerases with strong stand displacement activity. Once the reverse complement is synthesized from the overhang primer, it is displaced from the template by the synthesis of another molecule from upstream, and the analogous in the opposite end, to displace the overhang primer-mediated complement. Thus, an additional pair of primers must be designed in LAMP against the flanking regions of the targeted sequence. In addition to self-propagation, the looped molecules are still target to the overhang-primers, leading to rapid exponential amplification. Additional primers targeting the loops can be added to speed up the reaction.

The fast amplification confers extremely high potential sensitivity, although it also makes it very sensitive to contamination and other sources of false positives.

The design of LAMP primers is not as straight forward as in PCR, but some tools are available. Although the primers are designed for being very specific, they can self-amplify and thus lead to false positives[224–226]. Thus, meticulous validation is required before applying them for diagnosis. In response to the pandemic, many research groups quickly applied LAMP to detect SARS-CoV-2 and several primer sets have been validated[227–234]. It has been combined with a plethora of sample pre-processing protocols and detection methods.

The greatest advantage of LAMP over PCR is that the reaction takes place at a constant temperature, making it a convenient alternative for diagnosis when thermocyclers are not available. This opens the possibility to adapt it to a POC solution. The change in pH due to the large amount of DNA produced can be detected by an indicator[230], although this requires low or non-buffered reaction conditions that can introduce variability. Alternatively, aggregation of gold particles also produces a color change[235–237]. In addition, efforts have been made to implement the detectable LAMP reaction into a lateral flow device combined with Cas[238] or into an electrochemical device that detects a redox reaction[239].

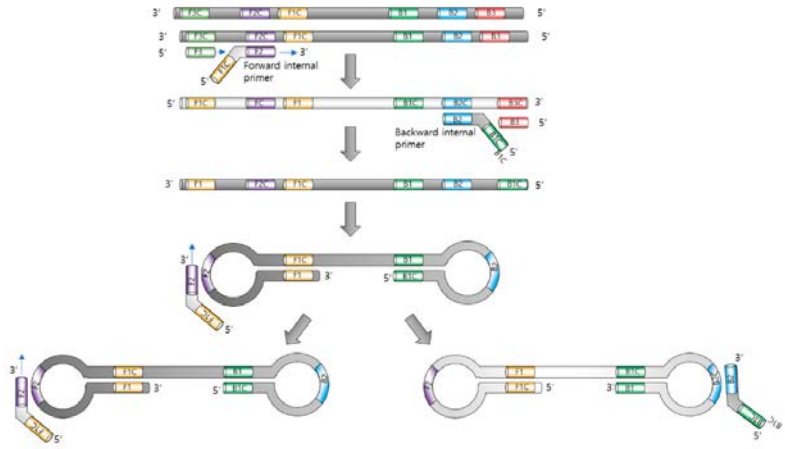

Fig. 1.11: Schematic representation of LAMP amplification, as explained in section 1.4.1.3. Image from Park et al. (2022)[240]

1.4.2 Other isothermal amplification methods

Another popular isothermal amplification approach is rolling circle amplification (RCA). In RCA, the target is circularized using padlock probes and replicated continuously into a very long molecule as the polymerase displaces the previously

synthetized. This method is largely used in molecular diagnostics as it is easily multiplex[241]. The application to detect RNA is not straightforward. Nevertheless, it has been successfully used for RNA target detection and, in particular, for SARS-CoV-2 [242–244].

Other strand displacement methods investigated for SARS-CoV-2 detection include the use of nickases, that cut the newly synthesized strand allowing easy release of short fragments, that then serve as primers for exponential amplification. Examples of this approach are the SDA[245] and EXPAR[246,247] protocols. These methods usually suffer from unspecific amplification. Another approach to amplify nucleic acids without cycling temperatures is using enzymes to unwind the DNA, as in helicase-dependent amplification (HDA)[248] and recombinase polymerase amplification (RPA)[249], both applied for SARS-CoV-2 detection.

Finally, combining a reverse transcriptase with DNA-dependent RNA polymerase and RNase H activity, produce cycles of alternative synthesis of DNA and RNA copies. Some protocols based on this idea have been applied for SARS-CoV-2 detection. The most relevant are transcription-mediated amplification (TMA)[250], that produces linear amplification, and nucleic acid sequence-based amplification (NASBA)[251], that amplifies exponentially. These methods are currently not fully isothermal as they need an initial heating step for primer annealing. However, it has been suggested that the thermal step can by bypassed by a using DNA-binding protein. Both have been used to detect SARS-CoV-2 [250–254].

In summary, isothermal amplification of nucleic acids are flexible, fast, cost-effective and up-scalable approaches that can be applied in the case of future pandemics both in laboratory settings and in POC solutions.

2 Research aims

2.1 Paper I: Differential regulation of mRNA stability modulates transcriptional memory and facilitates environmental adaptation.

- To develop a platform to study transcriptional memory in yeast at genome-wide thelevel.
- To uncover new mechanisms underpinning transcriptional memory
- To decipher the role of mRNA metabolism in transcriptional memory regulation

2.2 Paper II: CRISPR KO genome-wide screen uncovers new factors involved in transcriptional memory.

- To develop a platform to study transcriptional memory in human cells at high throughput.
- To uncover factors involved in transcriptional memory of IFNγ in human cells.

2.3 Paper III: Target-enriched nanopore sequencing and *de novo* assembly reveals co-occurrences of complex on-target genomic rearrangements induced by CRISPR-Cas9 in human cells.

- To establish a data-driven workflow to uncover large on-target rearrangements derived from CRISPR/Cas9-mediated genetic engineering.
- To investigate unexpected on-target effects derived of CRISPR/Cas9 mediated genetic engineering.
- To study the biological impact and factors affecting those on-target unexpected events

2.4 Paper IV: Direct detection of SARS-CoV-2 using non-commercial RT-LAMP reagents on heat-inactivated samples.

- To develop a flexible, cheap, and fast method to detect SARS-CoV-2 on raw samples based on loop-mediated isothermal amplification.
- To benchmark and compare the performance of the protocol with available commercial alternatives on unextracted samples.

3 Results and Discussion

3.1 Paper I: Differential regulation of mRNA stability modulates transcriptional memory and facilitates environmental adaptation

3.1.1 Genome-wide screen of factors involved in transcriptional memory of *GAL1*

As described in section 1.2, the *GAL1* system in *S. cerevisiae* has been extensively used as a model to study transcriptional memory. Although many contributors had been identified, at the time of the conception of this study the system had always interrogated based on previous knowledge. We hypothesized that an unbiased genome-wide genetic screen would uncover new genes modulating transcriptional memory and facilitate a better understanding of this process.

We transformed the genome-wide *S. cerevisiae* deletion collection[255] with a reporter for *GAL1* expression. Next, we exposed the generated cell library to repeated stimulation with galactose as a carbon source. We compared the response of previously exposed cells (primed) with those exposed for the first time (naïve) by measuring the reporter expression using flow cytometry. We separated the positive and negative cells at different time points and sequenced their unique barcode to obtain the distribution of each mutant at each time point (Fig. 3.1.1A). Based on that distribution, we inferred the virtual expression kinetics of each genotype and compare it to the wild-type strain. We found 35 mutants that showed decreased memory, including expected factors such as the chromatin remodeler *ISW2*[256]. Other 37 mutants showed enhanced memory, including *ELP4*, a member of the elongation complex, which was previously reported to have a role in transcriptional memory[257].

Pathway enrichment analysis of the putative modulators of transcriptional memory reported RNA degradation as an enriched term. As nuclear degradation rates have been shown to contribute to the quick reprograming of gene expression after glucose deprivation, we decided to focus on the role of the exosome subunit Rrp6 in transcriptional memory, whose depletion enhanced memory (Fig. 3.1.1B).

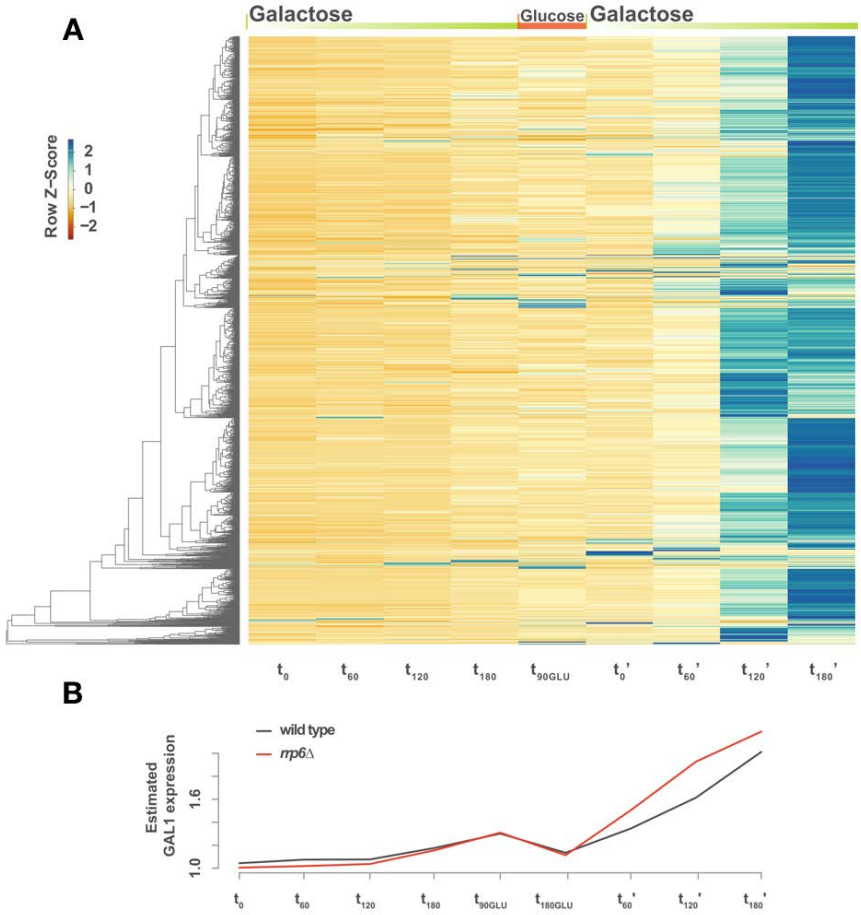

Fig. 3.1.1: Screen for genetic factors controlling transcriptional memory. A) Heatmap depicting strain-specific z-cores for pGAL1 expression at different time points (in minutes). B) Relative expression for rrp6Δ (red line) to wild-type (black) line (see Methods in paper I for details).

3.1.2 The absence of functional nuclear exosome produces changes in transcriptional memory

To validate and further study the effect of Rrp6 on transcriptional memory, we used the strain Δrrp6 and studied its genome-wide transcriptional response to repeated galactose stimulation along with the wild-type strain. Using RNA-seq data of primed vs naïve cells, we found 546 genes in the wild-type strain that showed an enhanced response (Fig. 3.1.2A), including genes related to galactose metabolism. Interestingly, we also found 773 genes that were increasingly repressed upon reexposure, with an overrepresentation of genes related to rRNA processing and ribosome assembly. Compared to the wild-type strain, we found 88 genes with enhanced transcriptional memory. In accordance with the screen results, *GAL1* was among those genes also affected at the mRNA level. We also

found 158 that showed enhanced repression in the mutant (Fig. 3.1.2B). Thus, the lack of functional exosome by depletion of Rrp6 results in enhanced both activation and repressive transcriptional memory.

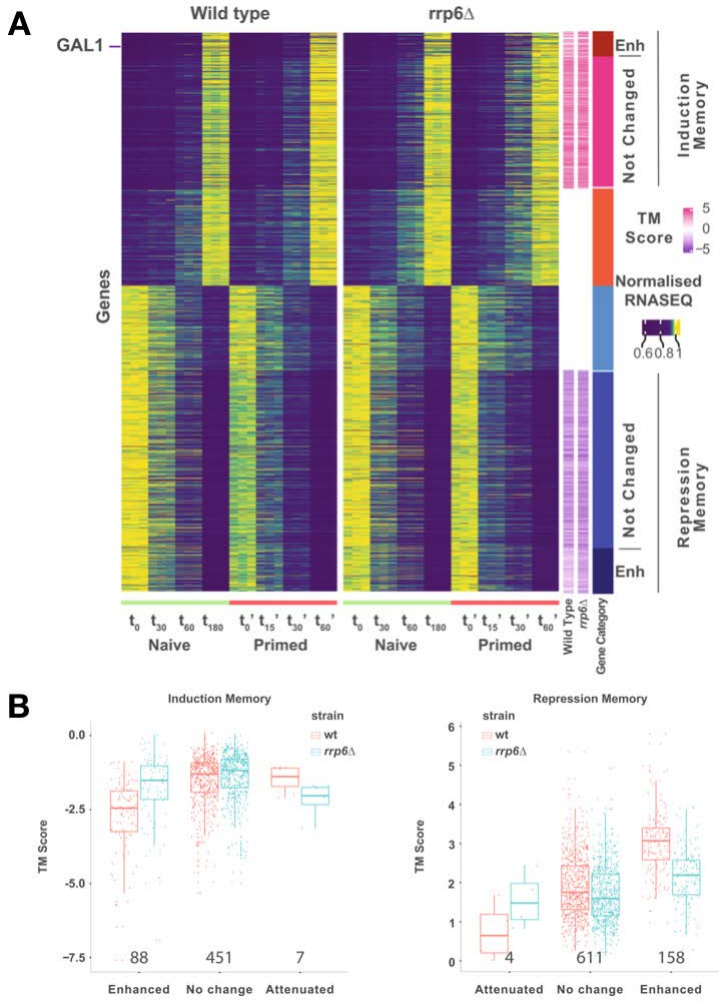

Fig. 3.1.2: Genome-wide identification of genes with transcriptional memory. A) Heatmap of variable genes in RNA-Seq of both wild-type and rrp6Δ strains indicating the relative mRNA abundance of each gene at different time points. Gene-specific transcription memory score is shown in purple to pink for each strain. The rightmost column shows gene category according to Memory Score (TMscore) in rrp6Δ respect to wild type. B) TMscore for genes classified as induction (left) or repression (right) memory, grouped by the differential behavior in the two strains.

3.1.3 The role of exosome subunit Rrp6 on transcriptional memory is not related to non-coding RNA accumulation or chromatin reorganization

One of the functions of the Rrp6-containing exosome complex is the degradation of cryptic unstable transcripts (CUTs)[258,259]. CUTs are small non-coding RNAs produced from inter and intragenic regions of the yeast genome, that are usually quickly degraded. Although whether they have a biological function is not clear, their expression is abundant, which becomes obvious upon depletion of Rrp6. As the depletion of the nuclear exosome in mammal leads to the accumulation of enhancer associated non-coding RNAs (eRNAs)[260], we hypothesized that CUTs accumulation could contribute to transcription activation also in yeast and thus explored the possibility that CUTs could play a role modulating transcriptional memory. However, we did not find differences in the accumulation of these non-coding transcripts between in naïve and primed conditions that could explain the phenotype.

As explained in section 1.2.3, chromatin plays a key role during transcriptional memory. Thus, we also investigated the nucleosome occupancy by MNase treatment on primed and naïve states of wild-type strain and Rrp6 mutant. We only found some subtle difference between wild type and mutant but not between naïve and primed states. We concluded that these differences cannot explain the memory phenotype in the Rrp6 mutant.

3.1.4 Differential abundance of exosome co-factors modulate memory

Recent studies showed that changes in RNA degradation can impact gene expression rewiring in yeast[261,262]. In glucose deprivation conditions, stress responsive genes avoid nuclear decay while genes downregulated in response to glucose withdrawal are targeted more efficiently by nuclear surveillance factors. Thus, it is possible that similar changes in nuclear decay are affecting the memory phenotype. To investigate this hypothesis, we used publicly available CRAC (CRosslink And cDNA analysis) data to search for differences in the intrinsic affinity of transcripts for the nuclear surveillance complexes TRAMP (Trf4/5-Air1/2-Mtr4-polyadenylation) or NNS (Nrd1-Nab3-Sen1) in naïve conditions. Intriguingly, transcripts from induction memory genes are intrinsically more associated to exosome cofactors than are the repression memory transcripts (Fig. 3.1.3A-C). In addition, we observed that those genes that showed enhanced induction memory in the absence of Rrp6 were more intrinsically associated to TRAMP and NNS. Similarly, those genes that showed repression memory in the absence of Rrp6, were associated to TRAMP (Fig. 3.13D-F) and NNS. This suggest that the differential association of the

transcripts of memory genes to co-factors of the nuclear exosome complex could be contributing to the modulation of memory.

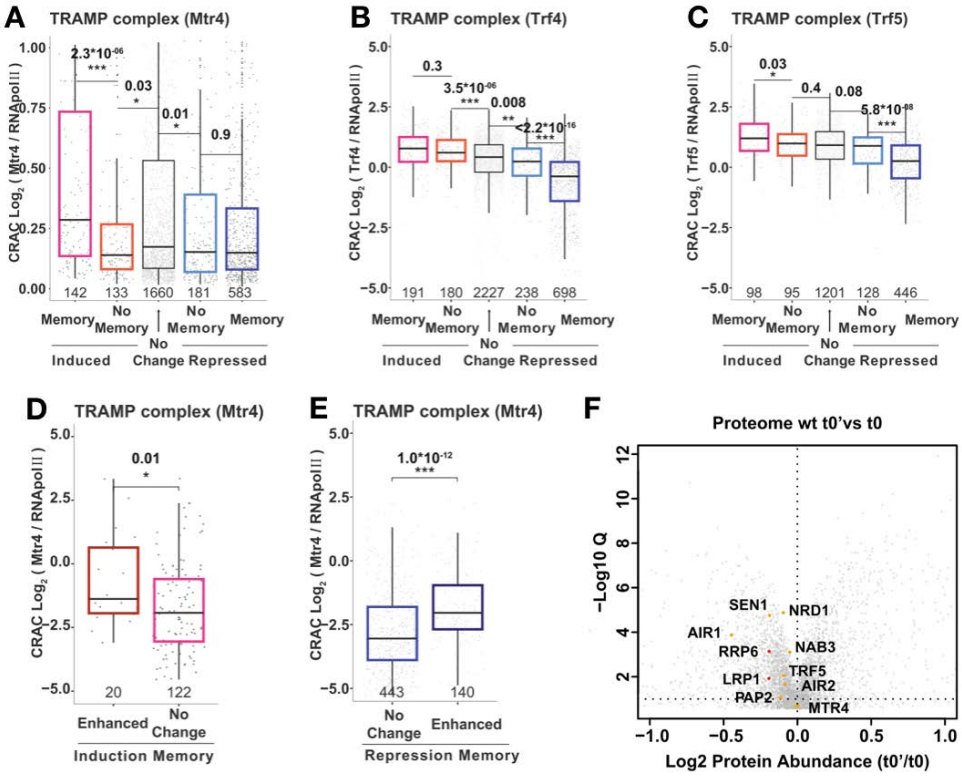

Fig. 3.1.3: Differential association to nuclear exosome co-factors. A-F) Boxplots representing the relative association for the TRAMP complex (e.g., Mtr4) of transcripts grouped by their memory category (A-C) or the differential behavior between rrp6Δ and wild-type strains (D-E). F) Volcano plot showing relative protein abundance changes of primed (t0') and naive (t0) conditions in wild-type cells.

Given their differential association between these RNA-binding proteins (RBPs) and their targets we hypothetized that the difference in mRNA abundance observed in memory genes could be due to differences in RBPs abundances during memory. To test this hypothesis, we performed proteomic analysis of wild-type cells in naïve and primed conditions (Fig. 3.1.3G). We found a significant decrease of nuclear exosome components and its cofactors TRAMP and NNS in primed cells, consistent with the faster accumulation of induction memory genes mRNAs. Interestingly, we also observe difference in abundance of proteins related to cytoplasmic decay, which suggests that the cytoplasmic mRNA degradation may also contribute to transcriptional memory.

3.1.5 Nuclear and cytoplasmic mRNA stability plays a role in transcriptional memory

Next, we investigated how the differential abundance of nuclear and cytoplasmic mRNA degradation machinery and the different mRNA-specific affinities could modulate mRNA stability in naïve and primed conditions.

We performed a genome-wide study of mRNA stability using mRNA metabolic labelling (SLAM-seq) in naïve and primed wild type cells before galactose induction. We observed a generalized stabilization of mRNAs in primed conditions (Fig. 3.1.4A). However, genes that showed induction transcriptional memory were more stabilized in primed respect to naïve conditions than those with repressive memory (Fig. 3.1.4B). We then interrogated the mRNA stability in different mutant strains. As in the wild-type, the *rrp6Δ* strain showed general stabilization of mRNA in primed conditions, although the turnover was generally slower than in the wild-type strain (Fig. 3.1.4C). Similarly, induction memory genes showed increased stability while repression memory genes showed faster turnover in the mutant (Fig. 3.1.4D). We obtained similar results in mRNA stability when depleting the component of the exosome Ski2. Depleting the cytoplasmic exonuclease Xrn1 abolished the global differences between primed than in naïve conditions (Fig. 3.1.4E). However, the specific differences in memory genes remained as in the previous mutants. Altogether, our results indicate that the changes in mRNA stability observed in primed conditions does not depend only on Rrp6 but also on the cytoplasmic decay machinery.

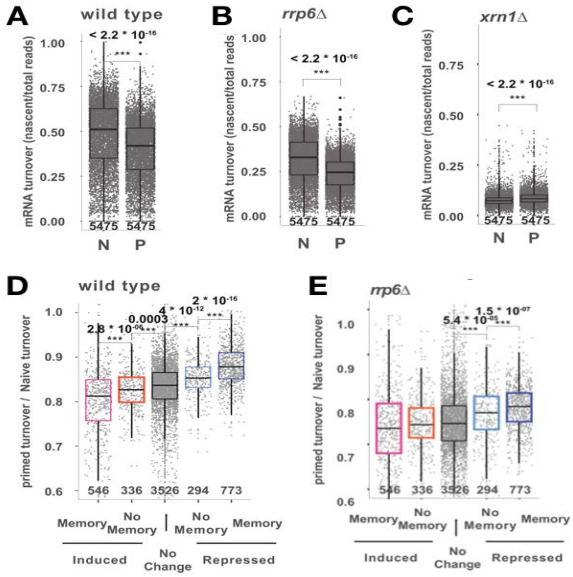

Fig. 3.1.4: Differential mRNA turnover between naïve and primed cells. Boxplots showing the relative mRNA turnover (nascent/total RNA) using SLAM-seq in naïve and primed conditions (A-C) in wild-type (A), *rrp6*Δ (B) or *xrny1*Δ (C) strains. D-E show the change in turnover between naïve and primed conditions for mRNAs of genes categorized by their memory in wild-type (D) or *rrp6*Δ strain (E).

3.1.6 Discussion

In this study, we have performed a genome-wide screen searching for factors that modulate *GAL1* memory in the budding yeast *S. cerevisiae*. Another genome-wide screen exploring memory of the *GAL1* gene was published during the course of this project[263]. Although it decreased the potential impact of our study, both screens can complement each other. We used information from Bheda *et al*. to validate our results and then focused on less explored mechanisms. We uncovered a previously unseen contribution of differential regulation of mRNA stability to transcriptional memory. We have dissected the role of the nuclear degradation machinery and its regulation during memory.

Although studies of gene expression often focus on transcription, both synthesis and decay contribute to the steady state of RNA. Here, we have studied the role of exosome-mediated mRNA stability on memory. We discovered that the memory genes affected by the depletion of Rrp6 had differential intrinsic affinity for nuclear exosome cofactors; the activation memory genes presented very high intrinsic association to the nuclear exosome cofactors NNS and TRAMP, while those with repression memory had very low association. This suggest that activation memory genes are enhanced in the absence of nuclear exosome due to mRNA stabilization. However, the repression memory genes also appear more repressed in the absence of functional exosome, despite the low affinity for its co-factors. This can be explained by the general stabilization of mRNAs upon Rrp6 depletion, that makes these unaffected genes look comparatively less abundant. As activation memory genes accumulate faster upon re-induction, their higher intrinsic affinity for exosome cofactors also points to a differential activity of the exosome during naïve and primed conditions. Supporting this hypothesis, we found lower abundance of proteins related to nuclear RNA degradation in primed cells.

In addition to the nuclear decay, we found evidence that the cytoplasmic degradation machinery contributes to modulate memory. Activation memory genes are more stable in the cytoplasm in primed than in naïve cells, even in the absence of Rrp6. We found that depletion of Xrn1 exonuclease abolished the differences in mRNA stability between naïve and primed conditions, as expected due to the major role of XRN1 in mRNA degradation. Interestingly, we found lower abundance of Xrn1p in primed cells than in naïve, which suggest a similar strategy than for the nuclear decay machinery. As for how the memory genes are particularly affected, we did not find specific changes in degradation rates of these genes between naïve and primed. Plausably, the downregulation of Xrn1 (and subsequent increase in global mRNA stability) in primed conditions could be just exacerbating the effects of upstream memory mechanisms. That is, the combination of a

faster transcription of these genes and lower targeting by the nuclear decay machinery leads to a higher amount of these mRNAs in the cytoplasm where, as mRNA degradation has been generally decreased, quickly accumulates to mount the response more efficiently. Contrarily, the apparent repression memory could be a collateral effect of the overall slower mRNA turnover by which these transcripts would be less affected.

3.2 Paper II: CRISPR KO genome-wide screen uncovers new factors involved in transcriptional memory

3.2.1 Transcriptional memory of IFNγ is reflected on the protein level of HLA-DR in human cells

As described in section 1.2, transcriptional memory has been shown and studied in human cells. Until now, most of the efforts have been mainly focused on investigating if mechanisms of transcriptional memory in yeast are conserved in human cells. Thus, we wanted to perform an unbiased genome-wide screen of factors potentially involved in transcriptional memory, to uncover new mechanisms involved in this process as in our screen in yeast. Similar to what we did in *S. cerevisiae*, we reasoned that transcriptional memory in humans could be studied by investigating changes at the protein level. Thus, differential protein abundance of a gene subjected to transcriptional memory control could be used to measure the response of individual cells and separate them accordingly. In particular, we used the IFNγ-induced expression of HLA-DR as a transcriptional memory model. In this case, we decided to use endogenous HLA-DR as a reporter for transcriptional memory.

At the time of the conception of this study, HLA-DR transcriptional memory of IFNγ had only been shown in Hela cells. To test if memory was reflected at the protein level, we exposed Hela cells to repeated IFNγ treatment and compared HLA-DR abundance in naïve and primed cells by flow cytometry, using a fluorescently labeled anti-HLA-DR antibody (Fig. 3.2.1A). As expected, we found that primed cells expressed HLA-DR protein complex faster than naïve cells (Fig. 3.2.1B).

Using this approach, we tested other human cell lines for memory. The fibroblasts BJ and Hs27, as well as the retina cell line RPE1 showed enhanced primed response (Fig 3.2.1C). Tumoral cell lines are genetically more unstable and tend to accumulate secondary genomic alterations that can result in genetic heterogeneous populations. As such preexisting genetic heterogeneity would introduce additional noise in our intended screen, we decided to use the non-tumoral RPE1 cell line for our screen. In particular, the

telomerase-expressing cell line hTERT-RPE1, as the senescence of the original cell line could impact the feasibility of the screen. We confirmed that HLA-DR in hTERT-RPE1 showed induction memory of IFNγ at the level of protein and mRNA (Fig. 3.2.1D-E).

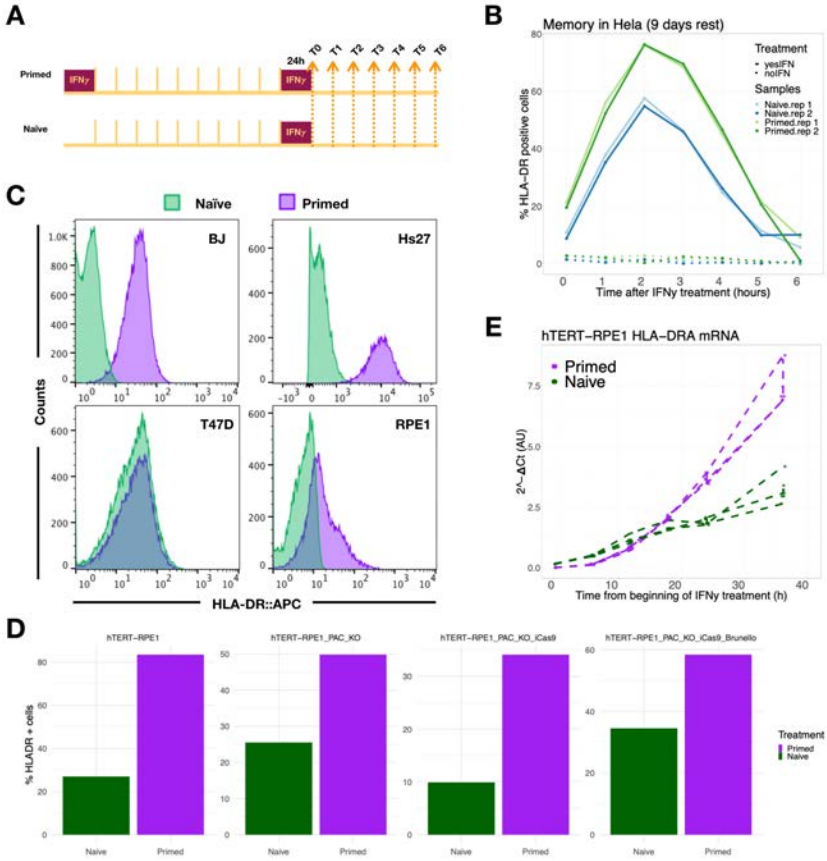

Fig. 3.2.1: Study of transcriptional memory in multiple cell line models. A) Experimental design. B) Percentage of HeLa cells expressing HLA-DRp in naïve or primed conditions. C) Distribution of HLA-DRp expression in naïve and primed cells measured in BJ, Hs27, T47D and RPE. D) RT-qPCR measuring HLA-DRA mRNA expression in naïve and primed conditions of hTERT-RPE1. E) Percentage of HLA-DRp expressing cells in naïve and primed conditions of subsequent modifications of RPE1 cell line measured.

3.2.2 Validation of the screening strategy

To perform a genome-wide screen, we generated a pool of knockout mutant cells using CRISPR/Cas. We first transduced a doxycycline-inducible Cas9 expression cassette into hTERT-RPE1. We selected cells that could properly induce Cas9 expression and transduced with the Brunello gRNA library. This library comprises 77,441 sgRNAs and includes 1000 non-targeting gRNAs that serve as a negative control. We confirmed that

the library was well represented in the pool (Fig. 3.2.2A) and that the pool of mutants still preserved memory (Fig. 3.2.1E).

To confirm that our screen strategy retrieves relevant biological information, we exposed the pool of mutants to IFNγ and calculated the changes in the gRNA frequencies in the HLA-DR positive fraction. First, we separated the positive cells by Fluorescence-activated cell sorting (FACS). We found 4619 guides depleted from the positive fraction respect to the negative control guides, and 39 genes with at least 2 guides among those depleted. We consider those genes as affecting the naïve expression of HLA-DR induced by IFNγ. Amongst the genes affected in the induced HLA-DR expression were the IFNγ receptor subunits IFNGR1 and IFNGR2, the transducers of the signal STAT1, JAK1 and JAK2, the transcriptional activator CTIIA, and the alpha subunit of the HLA-DR complex HLA-DRA (Fig. 3.2.2B). In addition, we performed gene ontology analysis and identified IFNγ and related pathways amongst the most significant (Fig. 3.2.2C). These results indicate that our strategy recapitulates known biology of the expression of HLA-DR induced by IFNγ.

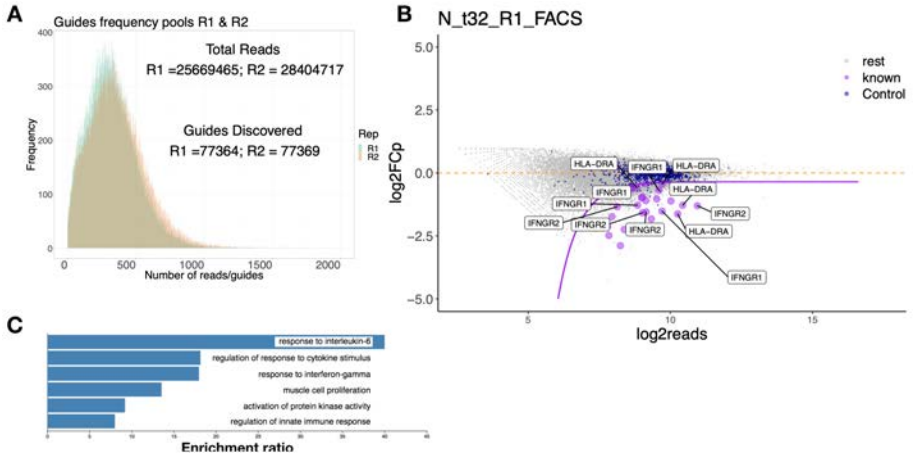

Fig. 3.2.2: Validation of the screening strategy. A) Read/guide distribution and guide discovery of replicate 1 (green) and replicate 2 (orange) of the knockout pools. B) change in probability of being positive for each guide with respect to the control guides (blue). Guides targeting genes known to be involved in the pathway are highlighted (purple dots). D) Gene Ontology terms significantly represented amongst the gene hits called.

Next, we used an alternative method to separate the positive cells after IFNγ exposure, based on magnetic activated cell sorting (MACS). As expected, we identified as depleted genes that are involved in the IFNγ-induced expression of HLA-DR. However, this separation method increased the amount of hits retrieved (9559 guides, 236 genes). This could be a consequence of an enhanced sensitivity due to the higher throughput, but it

could also reflect a higher rate of false positives. As FACS separation is very precise but can handle fewer cells, we decided to combine MACS and FACS separation for the screen to balance the capabilities and limitations of both methods.

3.2.3 KEAP1 and CBFB are novel putative factors modulating memory

To identify new factors involved in memory, we exposed the pool of mutants to repeated IFNγ treatment. Then, we separated the HLA-DR positive naïve or primed cells using MACS or FACS and sequence the integrated guide in each cell. Finally, we calculated the frequency of each guide respect to the control guides in each fraction and compared to the proportions in the original population. To find genes affecting memory, we calculate the join effect of the guides targeting the same gene based on the overall magnitude of the change, the consistency between guides, and the consistency of the hit across samples of the same treatment condition (Zscore). When ranking genes according to their difference in Zscore between naïve and prime conditions, we found CBFB (Core-Binding Factor Subunit Beta) and KEAP1 (Kelch-Like ECH Associated Protein 1 or INrf2) as the top candidates (Fig. 3.2.3A). Further, when we looked at the change of each guide targeting these genes, we found them consistently more depleted from the positive fraction of primed cells than naïve (Fig. 3.2.3B-C). Thus, these genes are putative factors involved in the regulation of memory.

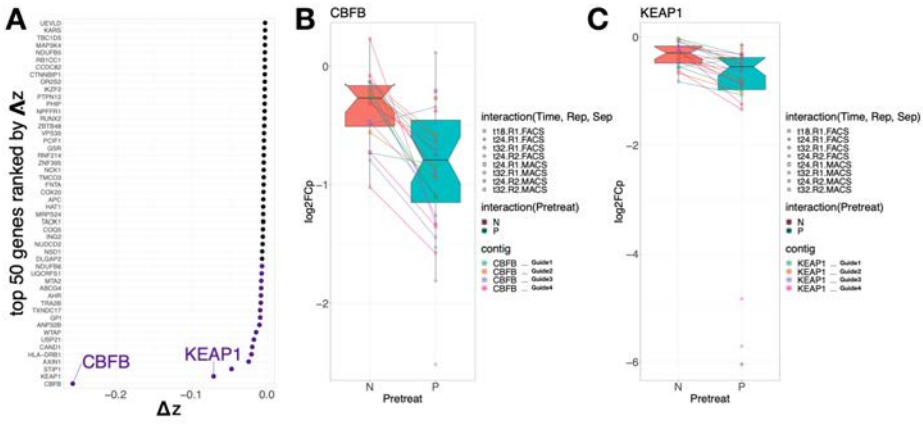

Fig. 3.2.3: Identification of candidates of memory regulators. A) Top 50 candidates ranked according to the difference in Z score (ΔZ) between primed and naïve responses. naïve response. Only top 50 candidates are plotted. B) Change in probability of being positive of each guide respect to negative control guides (log scale) in each sample of the naïve (red) and primed (blue) responses targeting the top candidate CBFB. C) As in B for KEAP1 gene.

3.2.4 Discussion

Here, we have developed a platform to interrogate at genome-wide scale the memory of HLA-DR upon repeated treatments with IFNγ. For that, we have optimized the combination of genome-wide CRISPR/Cas9 knockout pool of mutants with two alternative protein-based separation methods (FACS and MACS) and coupled it with targeted sequencing.

We have demonstrated that HLA-DR memory was reflected at the protein level and therefore protein-based separation methods are adequate to study this phenotype. To do this, we have used the HeLa cells as at the time of this study was the only human example of IFNγ-induce HLA-DR memory. However, tumoral cell lines tend to be genomically unstable, especially HeLa. We suspected that other cell lines could present this memory. We have shown that RPE1, among others, showed enhanced response to repeated IFNγ treatment both at the mRNA and protein level. We considered this cell line an adequate cell model to study IFNγ-induced HLA-DR memory and produced a pool of knockout mutant cells by viral transduction of the CRISPR/Cas9 components.

In a genome-wide screen, the efficiency of the delivery must be extremely high to ensure proper coverage of the entire library. Viral transduction was eventually used in this study as numerous attempts with other delivery strategies were not efficient enough. As IFN type I (IFNα and ß) and IFNs type 2 (IFNγ) pathways may interact, this brought the concern that the exposure to the virus could prime the cells and distortion the phenotype. We inserted the Cas9 expression cassette under inducible promoter in order to temporally separate the transduction and the knockout generation events. Although we have never tested if the exposure to the virus really affects the HLA-DR response, we have demonstrated that the enhanced response in primed conditions is maintained for the pool of mutants when cultured for two weeks between the transduction and the induction of DSBs.

We proceeded to validate our strategy with both, FACS and MACS separation methods, by interrogating the naïve response to IFNγ in the induced knockout pool. Notably, the changes in gRNA distributions in the HLA-DR positive fractions were calculated using reads as contig, which is subjected to amplification biases. The use of UMIs to collapse PCR duplicates have demonstrated to reduce the noise in CRISPR screens. Although our library has UMIs, they were introduced together with the gRNA constructs during the transduction. Since our experimental design involves long culturing periods, each of the original barcoded cells will give rise to hundreds of daughter cells with the same barcode. Thus, collapsing the reads to the UMI counts would not only remove the PCR duplicates

but also real data points, reducing the complexity of the sample. Despite the risk of increasing noise, we were able to clearly identify factors that are essential for the IFNγ-induced expression of HLA-DR in the naïve response, as for example the IFNγ receptor, transductors of the signal, or the HLA-DR alpha subunit, demonstrating that our strategy is able to recapitulate known biology of the IFNγ response. Importantly, we do not suffer from clear proliferation biases as previously reported, despite having extensive expansion of the non-induced pool culture prior the experiment, highlighting the importance of the used inducible Cas9 system.

To find genetic variants affecting memory, we compared the changes in guide distributions in the HLA-DR positive fraction of naïve and primed cells. Initially, we intended to capture a dynamic range of the response by acquiring cells at different time points, that would allow us to construct a virtual expression pattern for each gene, similarly to what we did previously in yeast. Unfortunately, we underestimated the increment in complexity from screens in yeast to human cells. In practice, the difference captured between time points was not enough to establish a kinetics, partially due to the variability between replicates and separation methods. Memory decreases as the cell divides and proliferation rates are affected by the confluency of the culture. RPE1 grows attached to the surface of the culture flasks and needs to be trypsinized for passages, which makes it especially laborious when handling the humungous amount of cells required for this screen. Although this handling was carefully performed, it is possible that heterogeneous confluency resulted in differences in proliferation that in terms led to the observed variability. We proceeded, then, to bin the time points as technical pseudo-replicates, treating memory as a binary phenotype. In retrospective, designing the experiment from the beginning in this way and devoting the sampling effort to increase the number of replicates rather than time points would have been a better approach, making the screen more robust and easier to analyze.

As our experimental design deviates from traditional screens, we developed a customized data analysis strategy. The resulting pipeline was empirically designed to identify genes that showed a more pronounced global effect on the primed than in the naïve response, at the same time that accounting for the consistency of the hit across samples of the same treatment and of the effect of the guides targeting the same gene. Although suboptimal, this approach allowed us to rank the genes according to their effect on the memory phenotype. We identified two very attractive candidates as top hits, CBFB and KEAP1. CBFB (core-binding factor subunit beta) is the regulatory subunit of a heterodimeric transcription factor that enhances the binding capacity of the alpha subunit

to enhancers and promoters. Interestingly, we also find one of its known partners, RUNX2, affected in memory. It is possible that this transcriptional regulator influences the differential kinetics in the expression of HLA-DR by differential binding during the naïve and primed responses, similarly to the Slf1 transcription factor in the memory of *INO* gene in yeast. Another possibility is that alternative alpha subunits of the complex occur during naïve and primed responses leading to differential regulation of the memory genes, similar to what has been described for CDK8-Mediator in HeLa cells. KEAP1 (kelch-like ECH associated protein 1) is a substrate-specific adapter of a BCR (BTB-CUL3-RBX1) E3 ubiquitin ligase complex. Among other functions, it senses and regulate oxidative stress response. It has been reported the KEAP1 can regulate the expression of HLA-DRA in response to IFNγ at the transcriptional level by regulating histone acetylation. Although we capture the effect of KEAP1 depletion in the naïve response, our screen indicates a consistently greater effect on the primed response. As histone modifications are a hallmark of transcriptional memory, it is possible that the enhanced HLA-DR expression in primed cells is modulated by KEAP1-mediated differential histone acetylation. Thus, I think that CBFB and KEAP1 are very interesting candidates for further validation and molecular dissection.

3.3 Paper III: Target-enriched nanopore sequencing and *de novo* assembly reveals co-occurrences of complex on-target genomic rearrangements induced by CRISPR-Cas9 in human cells

3.3.1 Perseverance of the target region in CRISPR/Cas9 deletion clones

As explained in section 1.3, CRISPR/Cas9 can be used to produce genomic deletions when two targeted DBSs are in proximity. Later, deletion clones can be identified by the different band sizes resulting from PCR amplification of the region, which can be further characterized by sequencing. This strategy is particularly important to target non-coding genes, as they do not have an ORF to be disturbed and thus the inductions of InDels will most likely not impair their function. tRNA genes are particularly problematic to target as they are highly similar to each other, making it difficult to target one in particular. To circumvent that, the flanking regions can be used as target for the CRISPR/Cas9-induced DSBs.

Using this strategy, we targeted a locus containing two tRNA genes in the hyperploid hepatocellular carcinoma (HepG2) and the near-haploid chronic myeloid leukemia

(HAP1) by directing Cas9 nuclease to the flanking regions (Fig. 3.3.1A). We selected clones with the expected band size and confirmed the deletion by sanger sequencing. To our surprise, when we performed ChIP-seq on these clones, we found peaks for RNA polIII and H3K4me3 (Fig 3.3.1B), suggesting that the putatively deleted region was still present in their genome. Interestingly, no reads spanned the DSBs points (Fig. 3.3.1C), suggesting that the putatively deleted sequence could be present in another locus.

To confirm this hypothesis, we coupled Xdrop technology to enrich for the target sequence to Oxford Nanopore long read sequencing (LRS) to get insights of the flanking regions. In brief, the DNA molecules are encapsulated in individual droplets together with primers against the target region. Droplets containing the target region are sorted by the positive amplification of the target and the whole content of the droplet is sequenced using Nanopore long read sequencing. Using this approach, we investigated clones derived from the two different cell lines, HAP1 Δt72 and HepG2 Δt15. As expected, the coverage in the flanking regions of the originally target locus was low when mapped to the annotated genome, supporting the hypothesis that the target region was not in its original genomic context.

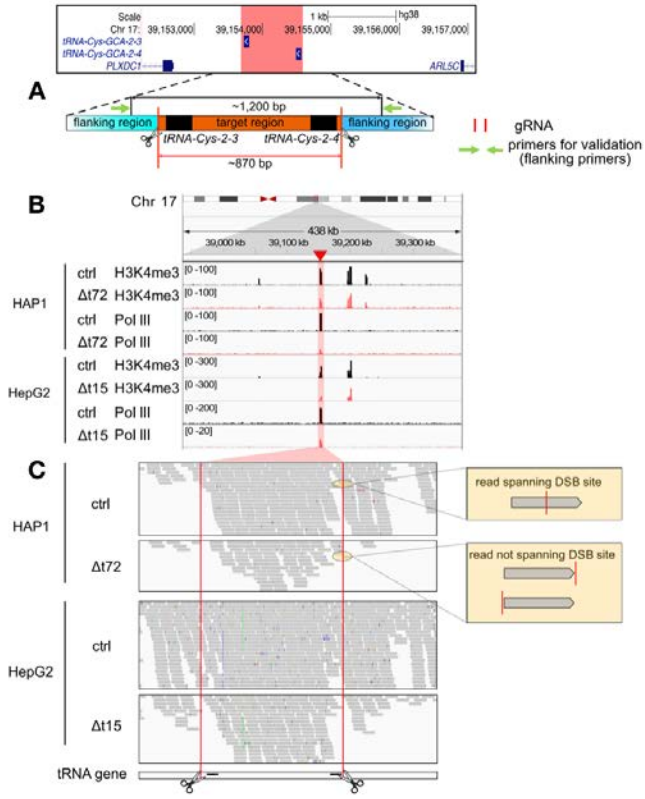

Fig. 3.3.1. The target region remained functional in deletion clones. A) Targeted locus and strategy of Cas9 dual gRNA deletion strategy. B) Normalized ChIP-seq reads for histone H3K4me3 and Pol III covering the target loci (red) in the HAP1 and HepG2 controls (ctrl), and Δt72 and Δt15 clones. C) Alignment tracks show individual reads around the DSBs points (scissors) as in (B).

3.3.2 Unexpected on-target aberrations in clones

To decipher the new genomic context of the target region, we developed a customized de novo assembly pipeline for the long reads of Hap1 Δt72 and HepG2 Δt15 clones, respectively. The obtained sequences from LRS were assembled by recursive alignment to the target region, the novo assembly, and gradual extension of the 5' and 3'. The new alignment revealed three break points in Hap Δt72 respect to the original loci, corresponding to two tandem inverted duplications of the target region (Fig. 3.3.2A). Unclipped reads from our ChIP-seq experiment, as well as combinatorial PCR amplification, supported this configuration. Similar analysis in HepG2 Δt15 revealed an inversion of the target region connected to the 5' flanking region, and a repetition of the target sequence at the 3' end. However, this repetition was followed by a sequence not found in the human genome. Instead, it was identified as a fragment of Escherichia coli genome and a 6000bp of the CRISPR/Cas9 vector used to produce the knockouts (Fig. 3.3.2B). As for the Hap1 clone, we confirmed this configuration with ChIP-seq reads and PCRs. Further, analysis of an additional clone revealed an unseen heterozygosis, with one allele carrying fragments of other chromosomes. We estimated the frequency of these on-target events by PCR of the target region with internal primers as 40% of HAP1 and 47% for HepG2 of the deletion clones, respectively. Thus, unexpected on-targets events are frequent and difficult to detect with traditional validation methods for CRISPR/Cas9 induced mutants.

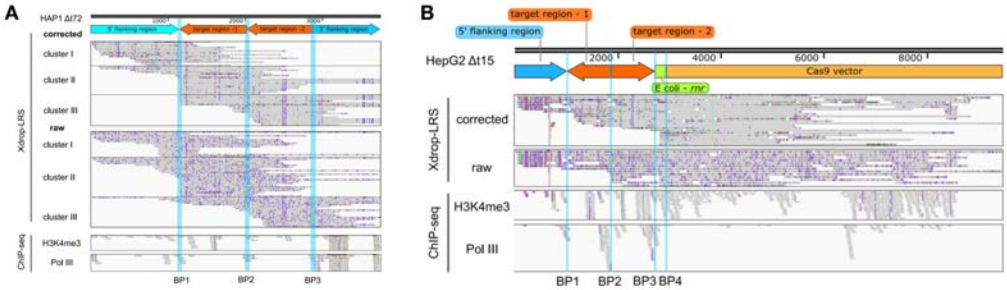

Fig. 3.3.2: On-target rearrangements. Structure of the on-target aberrations deciphered using XDrop-LRS and supported by ChIP-seq reads of Hap Δt72 (A) and HepG2 Δt15 (B) clones.

3.3.3 The unexpected on-target events are not locus specific

We wondered if the similarity between the two tRNAs could be promoting the on-target aberrations. To test this, we deleted the intragenic region between the two tRNA genes, which has a non-repetitive sequence, and investigated the deletion as well as potential on-

target abnormalities using combinatorial external and internal primers for PCR amplification. We detected genomic alterations in 14% of the Hap1 deletion clones and 50% of the HepG2 deletion clones, suggesting that the on-target alterations are not due to the repetitive nature of the targeted sequence.

Next, we reasoned that perhaps using plasmids as delivery system of CRISPR/Cas9 components and the pressure exerted during the drug-based selection of the transfected cells could be enriching for these on-target events. To test this, we generated knockout cell lines using targeted dual DSBs, as described before, but delivering the CRISPR/Cas9 components as an assembled ribonucleoprotein by nucleofection. We used a fluorescently labelled tracRNA to avoid drug selection. In addition, we used an immortalized non-tumoral euploid cell line (hTERT-RPE1) and targeted protein coding genes in different genomic contexts (RNF220 in Chr 1 and SULT1B1 in Chr 4) (Fig. 3.3.3A-B). In both cases, we found deletion clones with on-target abnormalities in about 10% of the deletion clones (Fig. 3.3.3C-D), which is only slightly lower than when deleting the tRNA intergenic region in the HAP1 cells.

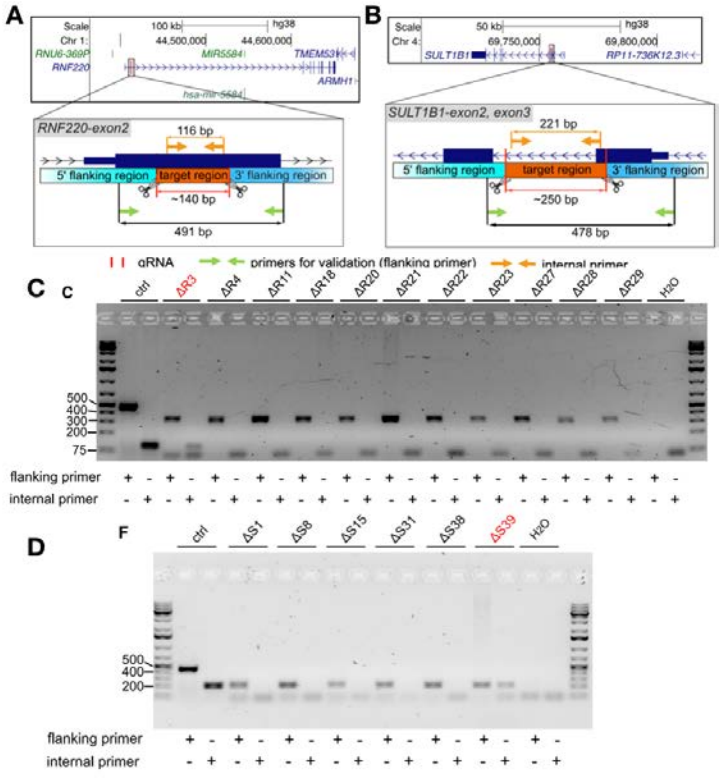

Fig. 3.3.3. Adverse on-target genomic rearrangements occurred using the CRISPR-Cas9-gRNA delivery system. A-B) Genomic locations of RNF220 (A) and SULT1B1 (B) and the dual gRNA design. C-D) Agarose gel of PCR products generated by flanking or the deleted regions in RNF220 (C) and SULT1B1 (D).

3.3.4 Biological consequences of no-target genomic aberrations

We found multiple evidence that the aberrant rearrangements could be transcriptionally active. Firstly, our ChIP-seq data showed H3K4me3 and Pol III ChIP-seq peaks on plasmid-derived sequences in HepG2 Δt15 (Fig. 3.3.4A). In addition, we detected expression of the Cas9 and puromycin resistance gene, as well as of the gRNA-Δt-1 and its scaffold sequence in this clone, and of gRNA-Δt-1 in HepG2 Δt8 (Fig. 3.3.4B). Thus, we wanted to explore whether the detected on-target aberrations could be functional and impact the biology of the cell. As a proxy, we measured proliferation of HAP1 and HepG2 clones carrying on-target genomic alterations, as compared to those with *bonafide* deletions of the intragenic (Δi) or the whole tRNA-containing region (Δt) (Fig. 3.3.4C-D). We found no significant differences between aberrant and *bonafide* modified clones in HepG2 clones, regardless of whether the entire locus or only the intergenic region was deleted. The same was true for HAP1 clones in which the intergenic region (Δi) was targeted. However, we detected higher proliferative rate in HAP1 Δt clones in which the target region was still detectable. Thus, on-target alterations can be transcriptionally active and affect cell behavior, which could ultimately impact the results of experiments that use CRISPR/Cas9 to produce DSBs.

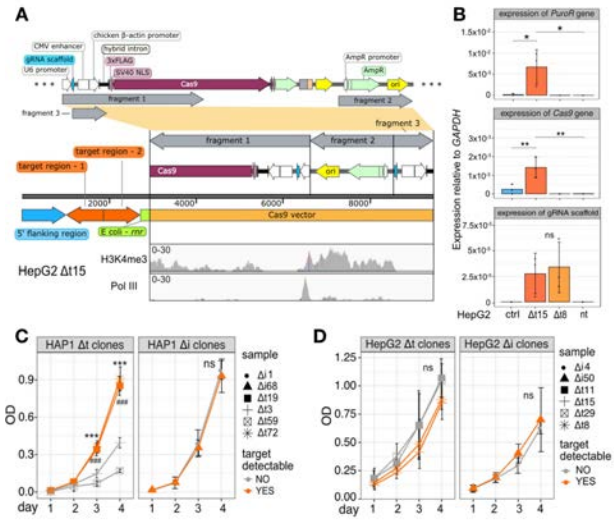

Fig. 3.3.4: Adverse on-target genomic alterations affected cell growth, promoted active transcription, and varied in abundance among deletion clones. (A) H3K4me3 and Pol III occupation at different CRISPR-Cas9 vector sequences integrated in HepG2 Δt15. B) Expression levels of genes that integrated in the HepG2 Δt15 and Δt8. C) Proliferation along 4 days of culture of HAP1 clones with (orange) or without (grey) the presence of target region when deleting the tRNA locus (Δt) or only the intragenic region (Δi). D) As in B for HepG2 clones.

3.3.5 Discussion

In this study, we revealed unprecedently documented co-occurrence of on-target aberrations in multiple cell lines and loci after the induction of CRISPR/Cas9-mediated DSBs. Although the occurrence of genomic alterations after CRISPR/Cas editing have been reported, the full picture of the complex on-target rearrangements remined obscure due to technological limitations. We have developed a workflow coupling XDrop and long read sequencing with a customize the novo assembly pipeline that enables to identify difficult to detect rearrangements such as on-target combinations of inversions and duplications, interchromosomal insertions, or exogenous sequences. Importantly, Nanopore LRS has high error rate. Additionally, sequencing long molecules promotes the appearance of chimeric reads. We validated the obtained assembled contigs by short reads Illumina sequencing, which has high fidelity, and we performed a combinatorial series of PCR to confirm each identified contig.

We applied this technology to decipher the on-target aberrations of multiple clones that underwent dual CRISPR/Cas9 DSBs. We have targeted the same sequence (flanking sequences of the tRNA-containing region) in two different cell lines HAP1 and HepG2, and in both we found aberrant clones, indicating that is not cell line specific. We also found on-target rearrangements when we targeted a different sequence (intragenic region between the two tRNAs) in the same cell lines, suggesting that these events are not due to some specific characteristic of the targeted sequence. Nevertheless, both flanking and intergenic regions are in the same locus for which we cannot discard the influence of the genomic context. In addition, other factor such as the use of plasmids to deliver CRISPR/Cas9 components, drugs selection, and the big distance between the two induced DSBs could be promoting the aberrant rearrangements. However, we have produced deletion clones by dual gRNA in another cell line, hTERT-RPE1 in which all these aforementioned elements differed from the previous strategy and still detected clones with major aberration. We estimated the frequency of aberrant clones amongst the deletion clones that we obtained. Although in some cases this frequency was lower (i.e., in hTERT-RPE1), the total number of clones were too low to give a reliable indication. In addition, the relative contribution of each of these factors cannot be separated. Thus, although based on our data the on-target aberrations can occur at relatively high frequency, this would need to be determined in a dedicated experiment.

Although the integration of exogenous sequences is not surprising when present during the induction of DSBs, foreign sequences tend to be heterochromatinzed. To our surprise, we found expression of on-target integration of Cas9, guides and puromycin resistance

cassettes. In addition, we showed that some clones that retained the target sequence in aberrant on-target configuration gained a proliferative advantage. This could lead to additional confounding in subsequent experiments and should be therefore detected during clone selection after CRSIPR/Cas9 DSBs induction. Importantly, these on-target aberrations are allele specific, which means that they can go unseen during the regular detection of the deletion by PCR fragment size (i.e., if the allele that carries the aberration does not amplify). Thus, we strongly recommend to always test for the presence of target sequence by amplifying the internal region of the fragment to be deleted and perform further investigation if necessary. Here we propose a method combining XDrop-LRS and de novo genome assembly to characterize those aberrant on-target structures with high precision.

This and other studies rising awareness on undesired effects of using CRISPR tools are extremely relevant, as these methods are widely used sometimes without the adequate precautions. Although in many cases aberrant clones can be discarded when detected, the method presented here could be useful when the sample is precious, and the number of clones is limited.

3.4 Paper IV: Direct detection of SARS-CoV-2 using non-commercial RT-LAMP reagents on heat-inactivated samples

3.4.1 Production and optimization of in-house LAMP enzymes and primers

As explained in section 1.4, the fast spreading of the SARS-CoV-2 challenged the testing capability world-wide. The overwhelming testing demand stimulated the development of faster and easier tests to complement RT-qPCR, and the supply of reagents for the most common testing methods were running short. Thus, we focused on the optimization of a LAMP-based detection on unextracted samples, using in-house produced retrotranscriptase (RT) and thermophilic polymerase with strand displacement activity.

First, we expressed and purified Bst LF (*Geobacillus stearothermophilu*) polymerase and two chimeric version of this protein and Klantaq (Thermus aquaticus), v5.9 and v7.16, and tested their ability to amplify synthetic fragments of the SARS-CoV-2 sequence in a LAMP reaction by measuring the incorporation of a fluorescent dye. For this step, we used previously validated LAMP primers and a commercial retrotranscriptase. For each polymerase, we optimized the buffer composition and enzyme concentration to minimize the reaction time while maximizing the separation of true positives from spurious

amplification of negative controls (Fig. 3.4.1A-B). We found that the chimeras v7.16 and v5.9 outperformed Bst LF. Of those, v7.16 showed the best performance, being able to detect 10^4 copies of the viral sequence in little more than 10 min with no non-specific amplification (Fig. 3.4.1C-D).

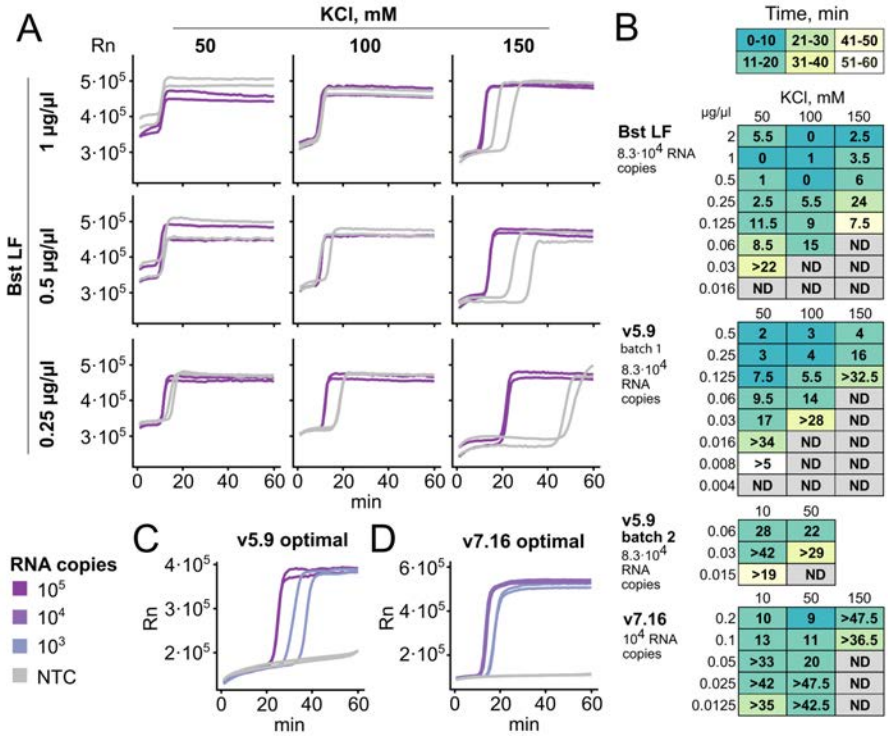

Fig 3.4.1. Optimization of enzyme amount and buffer composition for RT-LAMP. A) Example of RT-LAMP optimization varying Bst LF amount and the KCl concentration. B) Summary of optimization of enzyme amount and KCl concentration for Bst LF, v5.9, and v.7.16. Numbers indicate the time difference between the slowest-amplifying positive control replicate and the fastest-amplifying negative control replicate. Color indicates the required time to detect the slowest-amplifying positive control. C–D) Optimal conditions determined for v5.9 (C) and v7.16 (D).

Next, we explored alternative RT enzymes in combination with the chimeras v7.16 and v5.9. We tested RTX, with or without proofreading domain, and MashUp-RT at different concentrations in the optimal buffer conditions for each chimera polymerase. While both RTX versions were seemingly incompatible with our LAMP reactions, MashUp-RT efficiently amplified the synthetic SARS-CoV-2 sequence in combination with both v7.16 and v5.9 (Fig. 3.4.2A). In fact, the results were comparable to those obtained using the

commercial RT. Thus, we concluded that MashUp-RT and v5.9 or MashUp-RT and v7.16 were suitable enzyme combinations for fast and sensitive detection of SARS-CoV-2.

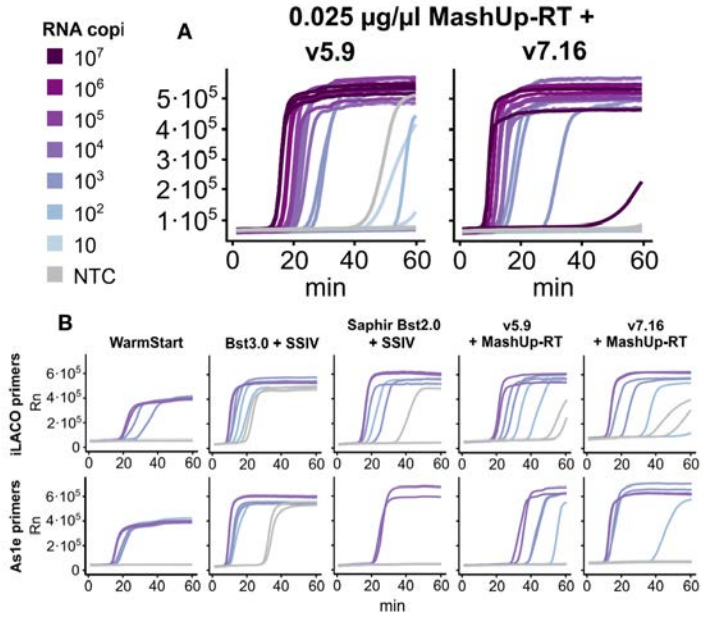

Figure 3.4.2: Benchmarking of in-house produced enzymes against commercial alternatives. A) Optimal detection of synthetic SARS-CoV-2 RNA using v5.9 or v7.16 supplemented with MashUp-RT. B) Synthetic RNA templates amplified with iLACO or As1e primers and either and the indicated combinations of enzymes.

3.4.2 Benchmarking of in-house LAMP with commercial alternatives

As our aim is to provide an alternative to commercially available options, we compared the performance of our in-house reaction with other solutions on the market (WarmStart Colorimetric master mix from NEB, Saphir Bst2.0 Turbo from Jena Biosciences, and Bst3.0 from NEB). In addition, we explore multiple validated primer sets for LAMP and selected iLACO and As1e for the benchmark experiment. All the tested combinations resulted in positive detection of the synthetic SARS-CoV-2 sequence. However, we found dramatic differences in the sensitivity and background signal (Fig. 3.4.2B). The performance of the in-house combination of v7.17-MashUpRT with As1e primers was comparable to the best commercial alternative Bst3.0, having similar detection time and no amplification in the negative control. Thus, we selected this combination for further application.

3.4.3 Application of the in-house LAMP reaction to nasopharyngeal samples

As most samples are collected in nasopharyngeal swabs, we wanted to explore the ability of our in-house LAMP reaction to detect in this samples. To go one step further, we intended to apply it on unextracted heat-inactivated samples. However, we first tested the compatibility of the reaction with some virus transport media (VTM) commonly used for the collection of these samples (Virocult (MWE), Sigma Transwab (MWE), eSwab (Copan) and Beaver (BEAVER biomedical)) using the synthetic SARS-CoV-2 sequence (Fig. 3.2.3A). Three of them were compatible with our in-house reaction adding up to 10% of the reaction volume. Beaver VTM needed to be diluted three more times, and still then only the chimera v5.6 was able to amplify the viral sequence efficiently.

Knowing that, we obtained 184 nasopharyngeal patient samples collected in Virocult, TransSwab and eSwab VTM to test our LAMP reaction. The viral load of those samples was previously analyzed using GeneXpert SARS-CoV-2; 142 classified as positive ($C\tau <$ 42) and 42 as negative. We perform LAMP reactions on the unextracted samples with v7.17-MashUpRT with As1e (2 replicates) or iLACO primers. For comparison, we included the popular commercial solution WarmStart. We used the previously determined $C\boldsymbol{\tau}$ values as reference to compare to the time of detection (Fig. 3.4.3B) and to calculate the sensitivity (Fig. 3.4.3C) and specificity (Fig. 3.4.3D) of the LAMP reactions. Although the detection capacity of LAMP dramatically drops on samples with high $C\tau$, we found a high and comparable specificity and sensitivity across all the LAMP reactions. Thus, we demonstrated that our in-house LAMP mix offers an efficient alternative to the commercially available products, being able to detect SARS-CoV-2 in unextracted samples with medium to high viral load for time and cost efficiency.

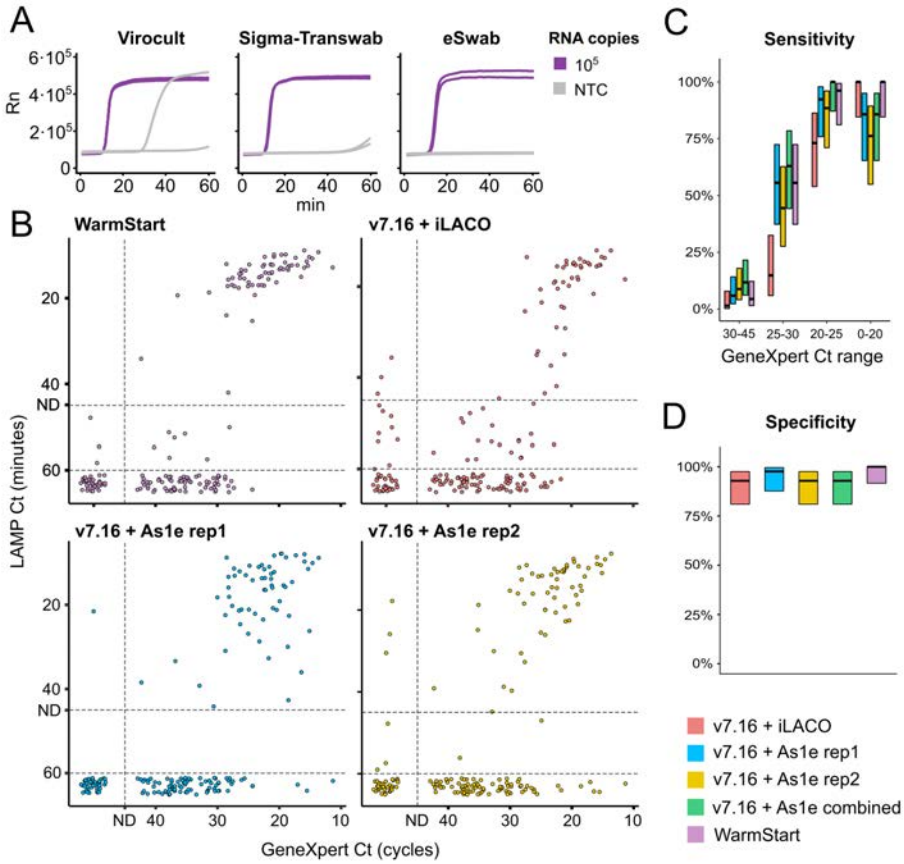

Fig. 3.4.3: Applicability of RT-LAMP to unextracted nasopharyngeal samples. (A) Effect of common VTM on RT-LAMP amplification with v7.16 and MashUp-RT. B) Comparisons of RT-LAMP Cτ (minutes) and GeneXpert RT-qPCR Cτ (cycles) for 184 clinical samples. v7.16 + RT-MashUP reaction mix was tested with iLACO (red) and As1e primer sets (blue/yellow). As1e primer set was also tested with WarmStart Colorimetric master mix (violet). ND designates the thresholds for calling positives. C) Reaction sensitivity according to SARS-CoV-2 abundance as determined by GeneXpert. D) Reaction specificity, as % of samples considered negatives by RT-qPCR that were also negative by RT-LAMP.

3.4.4 Discussion

Here, we have optimized a LAMP reaction to detect SARS-CoV-2 in unextracted patient swab samples using in-house reagents.

We have designed and optimized an in-house LAMP mix with comparable performance to the commercial alternatives. Our mix of enzymes can be easily produced with a simple expression and extraction protocol. Although access to certain equipment is necessary, large amounts of the enzyme is obtained in each batch, enough to perform hundreds of thousands of tests. Thus, it is time and cost efficient. Having alternatives to the

commercially available reagents was important as the increasing testing demand was challenging the reagents supply. We optimized the buffer composition and enzymes concentration, although we recommend calibrating the amount of enzyme for every batch as the efficiency might vary.

LAMP reaction has been previously proven efficient to detect the virus in purified samples. Purification removes potential inhibitors of the reaction and concentrates the RNA, leading to more sensitive and robust detection, but it is an expensive and time-consuming step that requires specialized equipment and personnel. We have successfully applied our reaction to unextracted swab samples in some of the most commonly used VTMs. Although with lower sensitivity than RT-qPCR, it was still able to identify positive samples with medium to high viral load. Thus, it would be useful in the context of massive or frequent testing. Although there is certain correlation between the viral load and the time of detection, we do not recommend using it as a quantitative test, as differences in the sample composition affect the efficiency of the reaction.

We followed continuous measurement of fluorescent dye incorporation during the optimizations to be able to track the reaction in real time. However, once established, the positive amplification can be determined by measuring end point fluorescence or other methods to detect amplified DNA. To note, although we have optimized our reaction to minimize spurious amplification, longer reaction times increase the probability of false positives.

Altogether, I think this method could be useful for continuous testing or in low resources circumstances. It can be also adapted to offer POC solutions or home tests if combined with an easy readout. In addition, as it relies on primer sequence complementarity, it can be easily adapted to detect other pathogens.

4 Conclusions and perspectives

In conclusion, in this thesis we performed two genome wide screens to identify genes affecting memory, one in the *GAL1* system in the budding yeast *S. cerevisiae* and one in the IFNγ-induced HLA-DR in the epithelial human cell line RPE1. In yeast, mRNA decay was found to modulate memory, while in humans CBFB and KEAP1 were identified as putative contributors. In addition, we present a method to uncover undesired on-target events that can appear as a consequence of CRISPR-mediated DSBs, and a LAMP-based rapid test for SARS-CoV-2 using in-house produced enzymes. Thus, this thesis is a combination of biology research and technological development.

As for transcriptional memory, we uncovered the role of the mRNA stability in the memory of the *GAL1* system in *S. cerevisiae*, which brings yet another layer of regulation to the already heavily studied *GAL1* system. This is valuable information to understand how cells adapt to their environment, and what are the basis of the regulation of gene expression. Besides the academic interest, deep understanding of metabolic adaptation mechanisms could allow fine manipulation of gene expression responses, which could be beneficial for the industrial production of metabolites. In addition, it could help to understand similar phenomenon that have an impact on human health, such as in the memory of hyperglycemia[264–266]. Nevertheless, although yeast is very often used to generate knowledge in gene regulation that can be applied to higher eukaryotes, it is becoming more and more evident that the mechanisms of transcriptional memory are not widely conserved. As genetic manipulation of human cells is becoming increasingly accessible, it is more practical to perform the studies directly in human models or more closely related species. The model chosen in this thesis for studying memory in humans is highly relevant as HLA-DR complex is the interface with the immune system. It also became obvious very quickly that we should select cells that are naturally present in the human body. Cells in physiological conditions are frequently exposed to IFNγ and, therefore, a deep understanding of how they respond to it is fundamental. The idea of non-immune cells aiding the immune system to mount an efficient response is not new[267,268]. We believe that the IFNγ-mediated memory of HLA-DR in non-immune cells could influence immune responses such as anti-tumor immunity[267–269] and untoimmunity[270–272]. However, this hypothesis must be further investigated. The discovery of targets for HLA-DR memory modulation could be used to investigate in these directions and perhaps for a future application in drug design. From the screen in RPE1, CBFB and KEAP1 appear as putative factors affecting memory. The clear next step is to produce knockout cell lines to

validate and decipher the role of these proteins in memory. Additionally, those mutants could be used to study differential interaction with the immune system.

CRISPR/Cas genome-wide screens are a powerful strategy to uncover new biology in human cells, but are technically challenging, partially because handling such large cultures is greatly demanding. Thus, the data generated from the human screen is a valuable resource and can be object of further analysis. Although we and others still perform screens manually, it is evident that automation is the future of this approach. Currently there are labs, facilities and companies with optimized platforms that are making this technology more accessible and adaptable to more complex phenotypes.

Another challenge when modifying cells using CRISPR/Cas is off- and on-targets effects. A good research effort has been devoted to the study of off-target events and their predictions. We rose awareness of the on complex on-target aberrations that might go unseen in the regular pipelines for selection of deletion clones. We propose a cost-efficient method that could be used to deconvolute the complex combinatorial sequences that might occur after Cas-mediated DSBs for high precision characterizations of the clones. This could be relevant in the clinics, as cells modified for gene therapy require of this high level of confidence. In addition, a direct application could be to study the factors that promote or prevent this type of events, as well as their frequency, including new variants of Cas9 or other nucleases.

Finally, we developed a rapid and economic test able to detect SARS-CoV-2 in samples with medium to high viral load, without the need specialized equipment or personal. As any other primer-based method, ours could be adapted to the detection of new pandemic pathogens. To facilitate its application, however, it is important to keep working on the read-out method. Ideally, to be implemented in a lateral flow device or similar that could be used as a home test or POC solution. The combination of isothermal amplification with Cas-mediated detection seems strongly promising[273–276]. Further, the fast development of Cas tools for diagnosis might make the pre-amplification step unnecessary[277–279]. Although there has been a myriad of diagnosis methods for the SARS-CoV-2 detection, we feel very recomforted to have contributed with our little part during the pandemic crisis. We have been contacted from countries with less resources that were interested in implementing the method, and we have sent enzymes and primers all around the world. In addition, it has been a great example of teamwork and interdisciplinary collaboration in an unbelievably fast pace. Altogether, it might not have changed the world, but it has dramatically impacted me as a scientist and as a person.

5 Acknowledgements

I cannot possibly name everyone that has gotten my back along this tough journey. If you don't find your name, know that is due to lack of space, not lack of gratitude. I'm very grateful for all the love and support that I have received, but I would like to mention some people in particular.

First of all, my Supervisor, **Vicent** (AKA *The Boss*). You are just the best. Always available, always trying to make our life easier, and always improving yourself. It is fascinating how you are evolving. Thank you for truly engaging my developing process, for letting me learn from my mistakes, and for being unconditionally there, with a comforting smile. You know I haven't had it all easy. Your support has been crucial to keep me going. My co-supervisor **Marie**. I know you think you didn't do much, but every meeting with you was an injection of positivity and hope. Your care means a lot to me. And of course, **Sebastián**. No olvido que fuiste el primero que le diste una oportunidad a una loquilla con gorro de oso que quería jugar con pipetas. Gracias por hacer possible que esté aquí hoy. Even though they are not my official supervisors, **Claudia** and **Marc** have given me valuable advice and feedback. Thank you, **Claudia,** for deriving me to the help I needed. **Marc**, thank you for sharing your little torpedo with me some days.

I would like to thank my opponent, **Anniina Vihervaara,** and the members of the examination board **Fredrik Wermeling, Lars-Gunnar Larsson and Mats Nilson**. I really appreciate the time and effort that you are putting in evaluating my PhD work.

To the **Kutländchanos!** All past and present members, I feel lucky to be part of it. Thank you for creating an open collaborative environment, full of good feedback and even some fun. I will miss working with you. **Keyi**! You are so adorable. Your combination of humbleness and intelligence is astonishing. It has been a pleasure to work with you. **Christian**, you are amazing. Thank you for the moments shared in the lab… and in the pub, in the council, board games, climbing... a bit everywhere, aren't you? An so are you, **Marcel**! Thank you for being so patient when my snacks wrappers overflowed to your desk. **Bingnan**, thank you for all those *memories*, your optimism, and for welcoming me in the Pelechano family. **Yujie <3**, together from the beginning to the end. Thank you for always making fun of everything and for the Chinese lessons (shieshie). Good luck with your defense! **Sus** (AKA *Vaggagita*), you are an example of maturity and love for science. Thank you for always remembering my birthday! **Donal**, your passion and curiosity are admirable. Those months working hand in hand were *the bests of my PhD*. I will forever

remember you and your frenetic talking. My sweet Lilit, I know you are doing all big stuff but…I miss you. You have a unique charm. Thank you for your hugs. **Eva**, thank you for your effort in team building and for answering all the random questions that I threw across the cupboard. To my smoking buddies, **Guido**, **Lefteris**, **Morteza**, and the new acquisition, **Santo**. Thank you for giving me a shoulder to cry on and for all those politically incorrect conversations. I really wish that if we quit (which we should), we will keep those running. **Cristina**, such a good student! Sharp, focused, with no fear of making questions nor trying new things. I really enjoyed your time here.

You deserve your own paragraph, **Dr. Alisa**. The pioneer PhD. I'm not going to flatter you mentioning your intelligence and your outstanding analytical skills, you already know that. I'm just going to tell you that since that hug in 2016 you are deep in my heart. I will be always here for you, just as you are for me. I love you and I feel so proud of being your friend. And thank you for bringing **Maria** and **Smara** to my life. What four! Each worse than the other up there! I have great time with you, girls.

Thank you to the **pub crew**. I really enjoyed making people happy once a week. **Gus**! I admire you. Good luck to you and **Hazal** in your new live! To the **climbing group**, thank you for pushing me! A mi gente de **Españoleando** ☺ Qué buenas comilonas nos hemos dado…y que sean muchas más! And to my **F1 team.** You understand that this thesis is like a Ferrari, had potential but… at the end surprised that it finished xD

My chosen family: **Fran, Francesco** (AKA *Monito*), **Roberto** and **Serena**. Thank you for all your support, for our boardgame evenings, for the nice food and for endless love. I hope we can keep growing together…and end up living closer!

David. No tengo palabras para explicarte lo mucho que eres y serás para mi. Gracias pore star a mi lado, pase lo que pase. Gracias por achucharme cuando flojeaba y por animarme cuando decía que no llegaba, por tu comprension y tolerancia, tu amor y cariño.

Mami! Gracias por haber creído siempre en mi. Si tengo algo de luchadora, lo saqué de ti. Espero que estés orgullosa de tu enanita del bosque. Ten clara una cosa: nadie es perfecto, pero hay gente buena y gente no tan Buena, y tú eres de las *extra-ordinaria* ☺

Last but not least, **Mikael**, mi *pollito*. You have arrived at the end of this journey, when I was the deepest in the ditch. I know it has not been easy to deal with me. Thank you for your patience and efforts to make my days beautiful. But overall, thank you for bringing some *lagomness* to this new self that I'm working on. Looking forward to building our life together.

6 References

1. Bonasio, R. The expanding epigenetic landscape of non-model organisms. *Journal of Experimental Biology* vol. 218 114–122 Preprint at https://doi.org/10.1242/jeb.110809 (2015).
2. Balint, B. L. *et al.* Arginine Methylation Provides Epigenetic Transcription Memory for Retinoid-Induced Differentiation in Myeloid Cells. *Mol Cell Biol* **25**, 5648–5663 (2005).
3. Bellec, M. *et al.* The control of transcriptional memory by stable mitotic bookmarking. *Nat Commun* **13**, (2022).
4. D'Urso, A. & Brickner, J. H. Epigenetic transcriptional memory. *Current Genetics* vol. 63 435–439 Preprint at https://doi.org/10.1007/s00294-016-0661-8 (2017).
5. Bheda, P., Kirmizis, A. & Schneider, R. The past determines the future: sugar source history and transcriptional memory. *Current Genetics* vol. 66 1029–1035 Preprint at https://doi.org/10.1007/s00294-020-01094-8 (2020).
6. Zacharioudakis, I., Gligoris, T. & Tzamarias, D. A Yeast Catabolic Enzyme Controls Transcriptional Memory. *Current Biology* **17**, 2041–2046 (2007).
7. Kundu, S., Horn, P. J. & Peterson, C. L. SWI/SNF is required for transcriptional memory at the yeast GAL gene cluster. *Genes Dev* **21**, 997–1004 (2007).
8. Light, W. H., Brickner, D. G., Brand, V. R. & Brickner, J. H. Interaction of a DNA zip code with the nuclear pore complex promotes H2A.Z incorporation and INO1 transcriptional memory. *Mol Cell* **40**, 112–125 (2010).
9. Lämke, J., Brzezinka, K., Altmann, S. & Bäurle, I. A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. *EMBO J* **35**, 162–175 (2016).
10. Ng, R. K. & Gurdon, J. B. *Epigenetic memory of active gene transcription is inherited through somatic cell nuclear transfer*. www.pnas.orgcgidoi10.1073pnas.0409813102 (2005).
11. Pascual-Garcia, P. *et al.* Metazoan Nuclear Pores Provide a Scaffold for Poised Genes and Mediate Induced Enhancer-Promoter Contacts. *Mol Cell* **66**, 63-76.e6 (2017).
12. Light, W. H. *et al.* A Conserved Role for Human Nup98 in Altering Chromatin Structure and Promoting Epigenetic Transcriptional Memory. *PLoS Biol* **11**, (2013).
13. Kamada, R. *et al.* Interferon stimulation creates chromatin marks and establishes transcriptional memory. *Proc Natl Acad Sci U S A* **115**, E9162–E9171 (2018).

14. Gialitakis, M., Arampatzi, P., Makatounakis, T. & Papamatheakis, J. Gamma Interferon-Dependent Transcriptional Memory via Relocalization of a Gene Locus to PML Nuclear Bodies. *Mol Cell Biol* **30**, 2046–2056 (2010).
15. Bheda, P. Metabolic transcriptional memory. *Molecular Metabolism* vol. 38 Preprint at https://doi.org/10.1016/j.molmet.2020.01.019 (2020).
16. Zapata, R. C. *et al.* Adipocytes control food intake and weight regain via Vacuolar-type H+ ATPase. *Nat Commun* **13**, (2022).
17. Netea, M. G. *et al.* Trained immunity: A program of innate immune memory in health and disease. *Science* vol. 352 427 Preprint at https://doi.org/10.1126/science.aaf1098 (2016).
18. Maheshwari, A. Innate Immune Memory in Macrophages. *Newborn* **2**, 60–79 (2023).
19. Chen, L. Innate Immune Memory in Hematopoietic Stem/Progenitor Cells: Myeloid-Biased Differentiation and the Role of Interferon. *Frontiers in Immunology* vol. 12 Preprint at https://doi.org/10.3389/fimmu.2021.621333 (2021).
20. Zhao, Z. *et al.* Sustained tnf-α stimulation leads to transcriptional memory that greatly enhances signal sensitivity and robustness. *Elife* **9**, 1–27 (2020).
21. Dunn, J., McCuaig, R., Tu, W. J., Hardy, K. & Rao, S. Multi-layered epigenetic mechanisms contribute to transcriptional memory in T lymphocytes. *BMC Immunol* **16**, (2015).
22. Barski, A. *et al.* Rapid Recall Ability of Memory T cells is Encoded in their Epigenome. *Sci Rep* **7**, (2017).
23. Tu, W. J. *et al.* Priming of transcriptional memory responses via the chromatin accessibility landscape in T cells. *Sci Rep* **7**, (2017).
24. Chong, M. M. W. *et al.* Epigenetic propagation of CD4 expression is established by the Cd4 proximal enhancer in helper T cells. *Genes Dev* **24**, 659–669 (2010).
25. May, W. A. *et al.* Characterization and drug resistance patterns of ewing's sarcoma family tumor cell lines. *PLoS One* **8**, (2013).
26. Shaffer, S. M. *et al.* Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).
27. Li, B. *et al.* Differential regulation of mRNA stability modulates transcriptional memory and facilitates environmental adaptation. *Nat Commun* **14**, (2023).

28. Geng, K. *et al.* Target-enriched nanopore sequencing and de novo assembly reveals co-occurrences of complex on-target genomic rearrangements induced by CRISPR-Cas9 in human cells. *Genome Res* **32**, 1876–1891 (2022).
29. Alekseenko, A. *et al.* Direct detection of SARS-CoV-2 using non-commercial RT-LAMP reagents on heat-inactivated samples. *Sci Rep* **11**, (2021).
30. Ding, Y., Fromm, M. & Avramova, Z. Multiple exposures to drought 'train' transcriptional responses in Arabidopsis. *Nat Commun* **3**, (2012).
31. Guan, Q., Haroon, S., Bravo, D. G., Will, J. L. & Gasch, A. P. Cellular memory of acquired stress resistance in Saccharomyces cerevisiae. *Genetics* **192**, 495–505 (2012).
32. Bassel, J. & Mortimer, R. *Genetic Order of the Galactose Structural Genes in Saccharomyces cerevisiae*. *JOURNAL OF BACTERIOLOGY* (1971).
33. Platt, A. & Reece, R. J. *The yeast galactose genetic switch is mediated by the formation of a Gal4p-Gal80p-Gal3p complex*. *The EMBO Journal* vol. 17 (1998).
34. Blank, T. E., Woods, M. P., Lebo, C. M., Xin, P. & Hopper, J. E. *Novel Gal3 Proteins Showing Altered Gal80p Binding Cause Constitutive Transcription of Gal4p-Activated Genes in Saccharomyces cerevisiae*. *MOLECULAR AND CELLULAR BIOLOGY* vol. 17 (1997).
35. *Galactose induction in yeast involves association of Gal80p with Gal1p or Gal3p*.
36. Yano, K.-I. & Fukasawa, T. *Galactose-dependent reversible interaction of Gal3p with Gal80p in the induction pathway of Gal4p-activated genes of Saccharomyces cerevisiae*. *Biochemistry* vol. 94 www.pnas.org. (1997).
37. Johnston, M., Flick, J. S. & Pextont, T. *Multiple Mechanisms Provide Rapid and Stringent Glucose Repression of GAL Gene Expression in Saccharomyces cerevisiae*. *MOLECULAR AND CELLULAR BIOLOGY* vol. 14 (1994).
38. Yarger, J. G., Halvorson, H. O. & Hopper, J. E. *Regulation of galactokinase (GALl) enzyme accumulation in Saccharomyces cerevisiae*. *Molecular and Cellular Biochemistry* vol. 61 (1984).
39. Sood, V., Cajigas, I., D'Urso, A., Light, W. H. & Brickner, J. H. Epigenetic transcriptional memory of GAL genes depends on growth in glucose and the tup1 transcription factor in saccharomyces cerevisiae. *Genetics* **206**, 1895–1907 (2017).
40. Sood, V. & Brickner, J. H. Genetic and Epigenetic Strategies Potentiate Gal4 Activation to Enhance Fitness in Recently Diverged Yeast Species. *Current Biology* **27**, 3591-3602.e3 (2017).

41. Cruz-Tapias, P., Castiblanco, J. & Anaya, J.-M. Autoimmunity: From Bench to Bedside. in *Autoimmunity: From Bench to Bedside* (eds. Anaya, J., Shoenfeld, Y., Rojas-Villarraga, A. & et al.) (El Rosario University Press, 2013).
42. Sun, J., Leahy, D. & Kavathas, P. B. *Interaction Between CD8 and Major Histocompatibility Complex (MHC) Class I Mediated by Multiple Contact Surfaces that Include the oL2 and oL3 Domains of MHC Class I*.
43. Functional interaction between human T-cell protein CD4 and the major histocompatibility complex HLA-DR antigen.
44. Aub, J., Niederwieser, ck D., Romani, N. & Huber, C. *Human Interferon-}, Induces Expression of HLA-DR on Keratinocytes and Melanocytes*. *Arch Dermatol Res (I985)* vol. 277.
45. Maenaka, A. *et al.* Interferon-γ-induced HLA Class II expression on endothelial cells is decreased by inhibition of mTOR and HMG-CoA reductase. *FEBS Open Bio* **10**, 927–936 (2020).
46. Reinisch, W. *et al. Donor dependent, interferon-γ γ γ γ induced HLA-DR expression on human neutrophils in vivo*. *Clin Exp Immunol* vol. 133 (2003).
47. Doxiadis, G. G. M., Hoof, I., De Groot, N. & Bontrop, R. E. Evolution of HLA-DRB genes. in *Molecular Biology and Evolution* vol. 29 3843–3853 (2012).
48. Muhlethaler-Mottet, A., Otten, L. A., Steimle, V. & Mach, B. Expression of MHC class II molecules in different cellular and functional compartments is controlled by differential usage of multiple promoters of the transactivator CIITA expression. *EMBO J* **16**, 2851–2860 (1997).
49. Keskinen, P., Ronni, T., Matikainen, S., Lehtonen, A. & Julkunen, I. *Regulation of HLA class I and II expression by interferons and influenza A virus in human peripheral blood mononuclear cells*. *Immunology* vol. 91 (1997).
50. Billiau, A. *Interferon-y: Biology and Role in Pathogenesis*. (1996).
51. Boehm, U., Klamp, T., Groot, M. & Howard, J. C. *Cellular responses to interferon-gamma*. *Annu. Rev. Immunol* vol. 8 http://www.annurev.org/sup/material.htm. (1997).
52. Kotenko, S. V. *et al.* Interaction between the components of the interferon γ receptor complex. *Journal of Biological Chemistry* **270**, 20915–20921 (1995).
53. Hemmi, S., Böhni, R., Stark, G., Marco, F. DI & Aguet, M. A novel member of the interferon receptor family complements functionality of the murine interferon γ receptor in human cells. *Cell* **76**, 803–810 (1994).
54. Biologic functions of the IFN-γ receptors.

55. Steimle, V., Otten, L. A., Zufferey, M. & Mach, B. *Complementation Cloning of an MHC Class II Transactivator Mutated in Hereditary MHC Class II Deficiency (or Bare Lymphocyte Syndrome)*. *Cell* vol. 75 (1993).
56. Li, J. *et al.* Nuclear PKC-θ facilitates rapid transcriptional responses in human memory CD4+T cells through p65 and H2B phosphorylation. *J Cell Sci* **129**, 2448–2461 (2016).
57. Netea, M. G. *et al.* Trained immunity: A program of innate immune memory in health and disease. *Science* vol. 352 427 Preprint at https://doi.org/10.1126/science.aaf1098 (2016).
58. Martinez-Gonzalez, I. *et al.* Allergen-Experienced Group 2 Innate Lymphoid Cells Acquire Memory-like Properties and Enhance Allergic Lung Inflammation. *Immunity* **45**, 198–208 (2016).
59. Netea, M. G., Latz, E., Mills, K. H. G. & O'Neill, L. A. J. Innate immune memory: A paradigm shift in understanding host defense. in *Nature Immunology* vol. 16 675–679 (Nature Publishing Group, 2015).
60. Siwek, W., Tehrani, S. S. H., Mata, J. F. & Jansen, L. E. T. Activation of Clustered IFNγ Target Genes Drives Cohesin-Controlled Transcriptional Memory. *Mol Cell* **80**, 396-409.e6 (2020).
61. Light, W. H. & Brickner, J. H. Nuclear pore proteins regulate chromatin structure and transcriptional memory by a conserved mechanism. *Nucleus (United States)* **4**, 357–360 (2013).
62. Brickner, D. G. *et al.* H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS Biol* **5**, 704–716 (2007).
63. Zovkic, I. B. & Walters, B. J. H2A.Z helps genes remember their history so we can remember ours. *BioEssays* **37**, 596–601 (2015).
64. Ng, R. K. & Gurdon, J. B. Epigenetic inheritance of cell differentiation status. *Cell Cycle* vol. 7 1173–1177 Preprint at https://doi.org/10.4161/cc.7.9.5791 (2008).
65. Tamura, T. *et al.* Inducible deposition of the histone variant H3.3 in interferon-stimulated genes. *Journal of Biological Chemistry* **284**, 12217–12225 (2009).
66. D'urso, A. *et al.* Set1/COMPASS and Mediator are repurposed to promote epigenetic transcriptional memory. (2016) doi:10.7554/eLife.16691.001.
67. Bevington, S. L. *et al.* Inducible chromatin priming is associated with the establishment of immunological memory in T cells. *EMBO J* **35**, 515–535 (2016).

68. Schaner, C. E., Deshpande, G., Schedl, P. D. & Kelly, W. G. *A Conserved Chromatin Architecture Marks and Maintains the Restricted Germ Cell Lineage in Worms and Flies*.

69. Singh, P. *et al.* Environmental history modulates Arabidopsis pattern-triggered immunity in a histone acetyltransferase1-dependent manner. *Plant Cell* **26**, 2676–2688 (2014).

70. Light, W. H. & Brickner, J. H. Nuclear pore proteins regulate chromatin structure and transcriptional memory by a conserved mechanism. *Nucleus (United States)* **4**, 357–360 (2013).

71. Cabal, G. G. *et al.* SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature* **441**, 770–773 (2006).

72. Kundu, S. & Peterson, C. L. Dominant Role for Signal Transduction in the Transcriptional Memory of Yeast GAL Genes . *Mol Cell Biol* **30**, 2330–2340 (2010).

73. Tan-Wong, S. M., Wijayatilake, H. D. & Proudfoot, N. J. Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev* **23**, 2610–2624 (2009).

74. Brickner, J. H. Transcriptional Memory: Staying in the Loop. *Current Biology* vol. 20 Preprint at https://doi.org/10.1016/j.cub.2009.11.013 (2010).

75. Cloutier, S. C. *et al.* Regulated Formation of lncRNA-DNA Hybrids Enables Faster Transcriptional Induction and Environmental Adaptation. *Mol Cell* **61**, 393–404 (2016).

76. Raices, M. & D'Angelo, M. A. Nuclear pore complexes and regulation of gene expression. *Current Opinion in Cell Biology* vol. 46 26–32 Preprint at https://doi.org/10.1016/j.ceb.2016.12.006 (2017).

77. Pascual-Garcia, P. & Capelson, M. Nuclear pores as versatile platforms for gene regulation. *Current Opinion in Genetics and Development* vol. 25 110–117 Preprint at https://doi.org/10.1016/j.gde.2013.12.009 (2014).

78. Arib, G. & Akhtar, A. Multiple facets of nuclear periphery in gene expression control. *Curr Opin Cell Biol* **23**, 346–353 (2011).

79. Tamura, K. Nuclear pore complex-mediated gene expression in Arabidopsis thaliana. *J Plant Res* **133**, 449–455 (2020).

80. Brickner, D. G., Coukos, R. & Brickner, J. H. INO1 transcriptional memory leads to DNA zip code-dependent interchromosomal clustering. *Microbial Cell* **2**, 481–490 (2015).

81. Pascual-Garcia, P., Little, S. C. & Capelson, M. Nup98-dependent transcriptional memory is established independently of transcription. *Elife* **11**, (2022).
82. Casolari, J. M. *et al. Genome-Wide Localization of the Nuclear Transport Machinery Couples Transcriptional Status and Nuclear Organization*. *Cell* vol. 117 http://www.cell.com/ (2004).
83. Brickner, J. H. Transcriptional memory at the nuclear periphery. *Current Opinion in Cell Biology* vol. 21 127–133 Preprint at https://doi.org/10.1016/j.ceb.2009.01.007 (2009).
84. Brickner, J. H. & Walter, P. Gene recruitment of the activated INO1 locus to the nuclear membrane. *PLoS Biol* **2**, (2004).
85. Guan, Q., Haroon, S., Bravo, D. G., Will, J. L. & Gasch, A. P. Cellular memory of acquired stress resistance in Saccharomyces cerevisiae. *Genetics* **192**, 495–505 (2012).
86. Tehrani, S. S. *et al.* <scp>STAT1</scp> is required to establish but not maintain interferon-γ-induced transcriptional memory. *EMBO J* (2023) doi:10.15252/embj.2022112259.
87. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology* vol. 16 129–143 Preprint at https://doi.org/10.1038/nrm3952 (2015).
88. Hsin, J. P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes and Development* vol. 26 2119–2137 Preprint at https://doi.org/10.1101/gad.200303.112 (2012).
89. Jeronimo, C., Collin, P. & Robert, F. The RNA Polymerase II CTD: The Increasing Complexity of a Low-Complexity Protein Domain. *Journal of Molecular Biology* vol. 428 2607–2622 Preprint at https://doi.org/10.1016/j.jmb.2016.02.006 (2016).
90. Kim, H. S. *et al.* Mammalian Ssu72 phosphatase preferentially considers tissue-specific actively transcribed gene expression by regulating RNA Pol II transcription. *Theranostics* **27**, 186–206 (2022).
91. Kuehner, J. N., Pearson, E. L. & Moore, C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol* **12**, 283–294 (2011).
92. Hui Ng, H., ois Robert, F., Young, R. A. & Struhl, K. *Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity*. *Orphanides and Reinberg* vol. 11 (2003).

93. Vihervaara, A. *et al.* Stress-induced transcriptional memory accelerates promoter-proximal pause release and decelerates termination over mitotic divisions. *Mol Cell* **81**, 1715-1731.e6 (2021).
94. Nogi, Y. *GAL3 Gene Product Is Required for Maintenance of the Induced State of the GAL Cluster Genes in Saccharomyces cerevisiae*. vol. 165 (1986).
95. Meyer, J., Walker-jonah, A. & Hollenberg, C. P. *Galactokinase Encoded by GAL] Is a Bifunctional Protein Required for Induction of the GAL Genes in Kluyveromyces lactis and Is Able To Suppress the gal3 Phenotype in Saccharomyces cerevisiae*. vol. 11 (1991).
96. Zacharioudakis, I. & Tzamarias, D. Bimodal expression of yeast GAL genes is controlled by a long non-coding RNA and a bifunctional galactokinase. *Biochem Biophys Res Commun* **486**, 63–69 (2017).
97. Stockwell, S. R., Landry, C. R. & Rifkin, S. A. The yeast galactose network as a quantitative model for cellular memory. *Molecular BioSystems* vol. 11 28–37 Preprint at https://doi.org/10.1039/c4mb00448e (2015).
98. Forestan, C. *et al.* Epigenetic signatures of stress adaptation and flowering regulation in response to extended drought and recovery in Zea mays. *Plant Cell Environ* **43**, 55–75 (2020).
99. Virlouvet, L., Ding, Y., Fujii, H., Avramova, Z. & Fromm, M. ABA signaling is necessary but not sufficient for RD29B transcriptional memory during successive dehydration stresses in Arabidopsis thaliana. *Plant Journal* **79**, 150–161 (2014).
100. Alshaer, W. *et al.* siRNA: Mechanism of action, challenges, and therapeutic approaches. *Eur J Pharmacol* **905**, (2021).
101. Sims, D. *et al.* High-throughput RNA interference screening using pooled shRNA libraries and next generation sequencing. *Genome Biol* **12**, (2011).
102. Okamura, Y., Makita, N., Hizukuri, Y. & Hayashi, Y. Genome-wide siRNA screening in mouse bone marrow-derived macrophages revealed that knockdown of ribosomal proteins suppresses IL-10 and enhances TNF-α production. *Journal of Clinical and Experimental Hematopathology* **58**, 87–94 (2018).
103. López-Varea, A. *et al.* Genome-wide phenotypic RNAi screen in the Drosophila wing: Global parameters. *G3: Genes, Genomes, Genetics* **11**, (2021).
104. Goh, W. S. S. *et al.* A genome-wide RNAi screen identifies factors required for distinct stages of C. elegans piRNA biogenesis. *Genes Dev* **28**, 797–807 (2014).
105. Safari, F., Barouji, S. R. & Tamaddon, A. M. Strategies for improving siRNA-induced gene silencing efficiency. *Adv Pharm Bull* **7**, 603–609 (2017).

106. Seok, H., Lee, H., Jang, E. S. & Chi, S. W. Evaluation and control of miRNA-like off-target repression for RNA interference. *Cellular and Molecular Life Sciences* vol. 75 797–814 Preprint at https://doi.org/10.1007/s00018-017-2656-0 (2018).
107. Wright, W. D., Shah, S. S. & Heyer, W. D. Homologous recombination and the repair of DNA double-strand breaks. *Journal of Biological Chemistry* vol. 293 10524–10535 Preprint at https://doi.org/10.1074/jbc.TM118.000372 (2018).
108. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology* vol. 18 495–506 Preprint at https://doi.org/10.1038/nrm.2017.48 (2017).
109. Bernardi, B. & Wendland, J. Homologous Recombination: A GRAS Yeast Genome Editing Tool. *Fermentation* vol. 6 Preprint at https://doi.org/10.3390/FERMENTATION6020057 (2020).
110. Wach, A., Brachat, A., Pöhlmann, R. & Philippsen, P. New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. *Yeast* **10**, 1793–1808 (1994).
111. Pierce, S. E., Davis, R. W., Nislow, C. & Giaever, G. Genome-wide analysis of barcoded saccharomyces cerevisiae gene-deletion mutants in pooled cultures. *Nat Protoc* **2**, 2958–2974 (2007).
112. Giaever, G. & Nislow, C. The yeast deletion collection: A decade of functional genomics. *Genetics* **197**, 451–465 (2014).
113. Pagé, N. *et al. A Saccharomyces cerevisiae Genome-Wide Mutant Screen for Altered Sensitivity to K1 Killer Toxin*. www.resgen.com/products/YEASTD.php3;
114. Kim, Y.-G., Cha, J. & Chandrasegaran, S. *Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain (Flavobacterium okeanokoites/chimeric restriction endonuclease/protein engineering/recognition and cleavage domains)*. vol. 93 (1996).
115. Klug, A. & Rhodes, D. *Reviews 'Zinc fingers': a novel protein motif for nucleic acid recognition*. *Btochem Soc q'rans-act* vol. 12 (1987).
116. Doyon, Y. *et al.* Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods* **8**, 74–79 (2011).
117. Pannunzio, N. R., Watanabe, G. & Lieber, M. R. Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *Journal of Biological Chemistry* vol. 293 10512–10523 Preprint at https://doi.org/10.1074/jbc.TM117.000374 (2018).

118. Maquat, L. E. *When cells stop making sense: Effects of nonsense codons on RNA metabolism in vertebrate cells*. *RNA* vol. 1 (1995).
119. Liu, Q., Segal, D. J., Ghiara, J. B. & Iii, C. F. B. *Design of polydactyl zinc-finger proteins for unique addressing within complex genomes*. vol. 94 www.pnas.org. (1997).
120. Segal, D. J., Crotty, J. W., Bhakta, M. S., Barbas, C. F. & Horton, N. C. Structure of Aart, a Designed Six-finger Zinc Finger Peptide, Bound to DNA. *J Mol Biol* **363**, 405–421 (2006).
121. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**, 143–150 (2011).
122. Boch, J. *et al.* Breaking the Code of DNA BindingSpecificity of TAL-Type III Effectors. *Science (1979)* **326**, 1509–1512 (2009).
123. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science (1979)* **339**, 819–823 (2013).
124. Terns, M. P. & Terns, R. M. CRISPR-based adaptive immune systems. *Current Opinion in Microbiology* vol. 14 321–327 Preprint at https://doi.org/10.1016/j.mib.2011.03.005 (2011).
125. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Geophys Res Lett* **31**, (2004).
126. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (N Y)* **155**, 733–740 (2009).
127. Brouns, S. J. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science (1979)* **321**, 960–964 (2008).
128. Garneau, J. E. *et al.* The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
129. Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (N Y)* **151**, 2551–2561 (2005).
130. Bolotin, A. *et al.* Complete sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus. *Nat Biotechnol* **22**, 1554–1558 (2004).
131. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**, 174–182 (2005).

132. Jansen, R., Embden, J., Gaastra, W. & Schouls, L. *Identification of genes that are associated with DNA repeats in prokaryotes*. (2002).
133. Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (N Y)* **151**, 2551–2561 (2005).
134. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Reviews Genetics* vol. 11 181–190 Preprint at https://doi.org/10.1038/nrg2749 (2010).
135. van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M. & Brouns, S. J. J. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends in Biochemical Sciences* vol. 34 401–407 Preprint at https://doi.org/10.1016/j.tibs.2009.05.002 (2009).
136. Sorek, R., Kunin, V. & Hugenholtz, P. *CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea*. http://crispr.u-psud.fr/crispr/ (2008).
137. Makarova, K. S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology* vol. 9 467–477 Preprint at https://doi.org/10.1038/nrmicro2577 (2011).
138. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
139. CHARPENTIER, E. & DOUDNA, J. Rewriting a genome. *Nature* **495**, (2013).
140. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2013**, (2013).
141. Charpentier, E. & Marraffini, L. A. Harnessing CRISPR-Cas9 immunity for genetic engineering. *Current Opinion in Microbiology* vol. 19 114–119 Preprint at https://doi.org/10.1016/j.mib.2014.07.001 (2014).
142. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
143. Cencic, R. *et al.* Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. *PLoS One* **9**, (2014).
144. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32**, 1262–1267 (2014).
145. Wong, N., Liu, W. & Wang, X. WU-CRISPR: Characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol* **16**, (2015).

146. Josephs, E. A. *et al.* Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res* **43**, 8924–8941 (2015).
147. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**, 839–843 (2013).
148. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* **32**, 677–683 (2014).
149. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822–826 (2013).
150. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**, 279–284 (2014).
151. Radzisheuskaya, A., Shlyueva, D., Müller, I. & Helin, K. Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. *Nucleic Acids Res* **44**, (2016).
152. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827–832 (2013).
153. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
154. O'Geen, H., Henry, I. M., Bhakta, M. S., Meckler, J. F. & Segal, D. J. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res* **43**, 3389–3404 (2015).
155. Ran, F. A. *et al.* XDouble nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell* **154**, (2013).
156. Wright, A. V. *et al.* Rational design of a split-Cas9 enzyme complex. *Proc Natl Acad Sci U S A* **112**, 2984–2989 (2015).
157. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science (1979)* **351**, 84–88 (2016).
158. Tsai, S. Q. & Joung, J. K. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat Rev Genet* **17**, 300–312 (2016).
159. Tycko, J., Myer, V. E. & Hsu, P. D. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Molecular Cell* vol. 63 355–370 Preprint at https://doi.org/10.1016/j.molcel.2016.07.004 (2016).

160. Naeem, M. & Alkhnbashi, O. S. Current Bioinformatics Tools to Optimize CRISPR/Cas9 Experiments to Reduce Off-Target Effects. *Int J Mol Sci* **24**, (2023).
161. Feng, S. *et al.* Strategies for High-Efficiency Mutation Using the CRISPR/Cas System. *Frontiers in Cell and Developmental Biology* vol. 9 Preprint at https://doi.org/10.3389/fcell.2021.803252 (2022).
162. Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* **9**, (2018).
163. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184–191 (2016).
164. He, Z. Y. *et al.* Non-viral and viral delivery systems for CRISPR-Cas9 technology in the biomedical field. *Science China Life Sciences* vol. 60 458–467 Preprint at https://doi.org/10.1007/s11427-017-9033-0 (2017).
165. Kanafi, M. M. & Tavallaei, M. Overview of advances in CRISPR/deadCas9 technology and its applications in human diseases. *Gene* vol. 830 Preprint at https://doi.org/10.1016/j.gene.2022.146518 (2022).
166. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology* vol. 38 824–844 Preprint at https://doi.org/10.1038/s41587-020-0561-9 (2020).
167. Brezgin, S., Kostyusheva, A., Kostyushev, D. & Chulanov, V. Dead cas systems: Types, principles, and applications. *International Journal of Molecular Sciences* vol. 20 Preprint at https://doi.org/10.3390/ijms20236041 (2019).
168. Liu, M. *et al.* Methodologies for improving HDR efficiency. *Frontiers in Genetics* vol. 10 Preprint at https://doi.org/10.3389/fgene.2018.00691 (2019).
169. Chen, E. *et al.* Decorating chromatin for enhanced genome editing using CRISPR-Cas9. *Proc Natl Acad Sci U S A* **119**, (2022).
170. Chen, H., Neubauer, M. & Wang, J. P. Enhancing HR Frequency for Precise Genome Editing in Plants. *Front Plant Sci* **13**, (2022).
171. Zhang, H. X., Zhang, Y. & Yin, H. Genome Editing with mRNA Encoding ZFN, TALEN, and Cas9. *Molecular Therapy* vol. 27 735–746 Preprint at https://doi.org/10.1016/j.ymthe.2019.01.014 (2019).
172. Zhang, H. X., Zhang, Y. & Yin, H. Genome Editing with mRNA Encoding ZFN, TALEN, and Cas9. *Molecular Therapy* vol. 27 735–746 Preprint at https://doi.org/10.1016/j.ymthe.2019.01.014 (2019).

173. Gao, S. *et al.* CRISPR screens identify cholesterol biosynthesis as a therapeutic target on stemness and drug resistance of colon cancer. *Oncogene* **40**, 6601–6613 (2021).
174. Tzelepis, K. *et al.* A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep* **17**, 1193–1205 (2016).
175. Kiessling, M. K. *et al.* Identification of oncogenic driver mutations by genome-wide CRISPR-Cas9 dropout screening. *BMC Genomics* **17**, (2016).
176. Li, F. *et al.* Identification of UBE2I as a Novel Biomarker in ccRCC Based on a Large-Scale CRISPR-Cas9 Screening Database and Immunohistochemistry. *Front Mol Biosci* **9**, (2022).
177. Jiang, L. *et al.* CRISPR/Cas9-Induced DNA Damage Enriches for Mutations in a p53-Linked Interactome: Implications for CRISPR-Based Therapies. *Cancer Res* **82**, 36–45 (2021).
178. Munoz, D. M. *et al.* CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov* **6**, 900–913 (2016).
179. Coelho, R. *et al.* Overlapping gene dependencies for PARP inhibitors and carboplatin response identified by functional CRISPR-Cas9 screening in ovarian cancer. *Cell Death Dis* **13**, (2022).
180. Yeung, A. T. Y. *et al.* A Genome-Wide Knockout Screen in Human Macrophages Identified Host Factors Modulating Salmonella Infection. (2019) doi:10.1128/mBio.
181. Cenik, B. K. *et al.* A synthetic lethality screen reveals ING5 as a genetic dependency of catalytically dead Set1A/COMPASS in mouse embryonic stem cells. (2022) doi:10.1073/pnas.
182. Cai, W. *et al.* A Genome-Wide Screen Identifies PDPK1 as a Target to Enhance the Efficacy of MEK1/2 Inhibitors in NRAS Mutant Melanoma. *Cancer Res* **82**, 2625–2639 (2022).
183. Hong, Y. *et al.* ST3GAL1 and βII-spectrin pathways control CAR T cell migration to target tumors. *Nat Immunol* (2023) doi:10.1038/s41590-023-01498-x.
184. Sangaré, L. O. *et al.* In Vivo CRISPR Screen Identifies TgWIP as a Toxoplasma Modulator of Dendritic Cell Migration. *Cell Host Microbe* **26**, 478-492.e8 (2019).
185. Jin, M., Gong, Y., Peng, J., Hu, X. & Shao, Z. In Vivo CRISPR Screens Identify RhoV as a Pro-Metastasis Factor of Triple-Negative Breast Cancer . *Cancer Sci* (2023) doi:10.1111/cas.15783.

186. Tomlinson, M. J., Tomlinson, S., Yang, X. B. & Kirkham, J. Cell separation: Terminology and practical considerations. *Journal of Tissue Engineering* vol. 4 1–14 Preprint at https://doi.org/10.1177/2041731412472690 (2013).
187. Amos, P. J., Cagavi Bozkulak, E. & Qyang, Y. Methods of cell purification: A critical juncture for laboratory research and translational science. *Cells Tissues Organs* **195**, 26–40 (2011).
188. Bacon, K., Lavoie, A., Rao, B. M., Daniele, M. & Menegatti, S. Past, Present, and Future of Affinity-based Cell Separation Technologies. *Acta Biomaterialia* vol. 112 29–51 Preprint at https://doi.org/10.1016/j.actbio.2020.05.004 (2020).
189. McKinnon, K. M. Flow cytometry: An overview. *Curr Protoc Immunol* **2018**, 5.1.1-5.1.11 (2018).
190. Alice L. Givan. Flow cytometry: an introduction. in *Flow Cytometry protocols. Methods in Molecular Biology* (eds. Hawley, T. & Hawley, R.) vol. 699 (Humana Press, 2011).
191. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866.e17 (2016).
192. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **12**, (2011).
193. Spahn, P. N. *et al.* PinAPL-Py: A comprehensive web-application for the analysis of CRISPR/Cas9 screens. *Sci Rep* **7**, (2017).
194. Hanna, R. E. & Doench, J. G. Design and analysis of CRISPR–Cas experiments. *Nature Biotechnology* vol. 38 813–823 Preprint at https://doi.org/10.1038/s41587-020-0490-7 (2020).
195. Michlits, G. *et al.* CRISPR-UMI: Single-cell lineage tracing of pooled CRISPR-Cas9 screens. *Nat Methods* **14**, 1191–1197 (2017).
196. Schmierer, B. *et al.* CRISPR/Cas9 screening using unique molecular identifiers. *Mol Syst Biol* **13**, 945 (2017).
197. Liu, Y. *et al.* A new genetic diagnostic for enlarged vestibular aqueduct based on next-generation sequencing. *PLoS One* **11**, (2016).
198. Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J. & Hsueh, P. R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents* vol. 55 Preprint at https://doi.org/10.1016/j.ijantimicag.2020.105924 (2020).
199. WHO Coronavirus (COVID-19) Dashboard. *https://covid19.who.int* (2023).

200. King, A., Lefkowitz, E., Adams, M. & Carstens, E. *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses*.
201. Robson, F. *et al.* Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. *Molecular Cell* vol. 79 710–727 Preprint at https://doi.org/10.1016/j.molcel.2020.07.027 (2020).
202. U.S. Department of Health & Human Services. Human Coronavirus Types. *https://www.cdc.gov/coronavirus/types.html* (2020).
203. KV Holmes, M. L. *Fields virology*. (Lippincott-Raven, 1996).
204. F Fenner. *The viruses*. (Intervirology, 1976).
205. Snijder, E. J. *et al.* Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol* **331**, 991–1004 (2003).
206. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology* vol. 19 155–170 Preprint at https://doi.org/10.1038/s41579-020-00468-6 (2021).
207. Focosi, D. & Maggi, F. Recombination in Coronaviruses, with a Focus on SARS-CoV-2. *Viruses* vol. 14 Preprint at https://doi.org/10.3390/v14061239 (2022).
208. Sun, Y., Wang, M., Lin, W., Dong, W. & Xu, J. Massive-scale genomic analysis reveals SARS-CoV-2 mutation characteristics and evolutionary trends. *mLife* **1**, 311–322 (2022).
209. Larremore, D. B. *et al. Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening*. (2021).
210. Mina, M. J., Parker, R. & Larremore, D. B. Rethinking Covid-19 Test Sensitivity — A Strategy for Containment. *New England Journal of Medicine* **383**, e120 (2020).
211. Feng, Z. *et al.* Mass screening is a key component to fight against SARS-CoV-2 and return to normalcy. *Medical Review* **2**, 197–212 (2022).
212. Vindeirinho, J. M., Pinho, E., Azevedo, N. F. & Almeida, C. SARS-CoV-2 Diagnostics Based on Nucleic Acids Amplification: From Fundamental Concepts to Applications and Beyond. *Frontiers in Cellular and Infection Microbiology* vol. 12 Preprint at https://doi.org/10.3389/fcimb.2022.799678 (2022).
213. Rodino, K. G. *et al.* Evaluation of saline, phosphate-buffered saline, and minimum essential medium as potential alternatives to viral transport media for SARS-CoV-2 testing. *Journal of Clinical Microbiology* vol. 58 Preprint at https://doi.org/10.1128/JCM.00590-20 (2020).

214. FDA. SARS-CoV-2 Viral Mutations: Impact on COVID-19 Tests. *https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/sars-cov-2-viral-mutations-impact-covid-19-tests#:~:text=The%20presence%20of%20mutations%20in,design%20differences%20of%20each%20test.* (2023).
215. Dramé, M. *et al.* Should RT-PCR be considered a gold standard in the diagnosis of COVID-19? *Journal of Medical Virology* vol. 92 2312–2313 Preprint at https://doi.org/10.1002/jmv.25996 (2020).
216. Deepak, S. A. *et al. Real-Time PCR: Revolutionizing Detection and Expression Analysis of Genes*. *Current Genomics* vol. 8 (2007).
217. Sy, R., Rothman, R. E. & Yang, S. *PCR-based diagnostics PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings*. *THE LANCET Infectious Diseases* vol. 4 http://infection.thelancet.com (2004).
218. VanGuilder, H. D., Vrana, K. E. & Freeman, W. M. Twenty-five years of quantitative PCR for gene expression analysis. *BioTechniques* vol. 44 619–626 Preprint at https://doi.org/10.2144/000112776 (2008).
219. Porter, W. T., Kelley, E. J., Bowers, J. R. & Engelthaler, D. M. Normalization of SARS-CoV-2 viral load via RT-qPCR provides higher-resolution data for comparison across time and between patients. *Virus Res* **306**, (2021).
220. Panpradist, N. *et al.* Simpler and faster Covid-19 testing: Strategies to streamline SARS-CoV-2 molecular assays. *EBioMedicine* **64**, (2021).
221. Bruce, E. A. *et al.* Direct RT-qPCR detection of SARS-CoV-2 RNA from patient nasopharyngeal swabs without an RNA extraction step. *PLoS Biol* **18**, (2020).
222. Smyrlaki, I. *et al.* Massive and rapid COVID-19 testing is feasible by extraction-free SARS-CoV-2 RT-PCR. *Nat Commun* **11**, (2020).
223. Israeli, O. *et al.* Evaluating the efficacy of RT-qPCR SARS-CoV-2 direct approaches in comparison to RNA extraction. *International Journal of Infectious Diseases* **99**, 352–354 (2020).
224. Notomi, T. *et al. Loop-mediated isothermal amplification of DNA*. *Nucleic Acids Research* vol. 28 (2000).
225. Meagher, R. J., Priye, A., Light, Y. K., Huang, C. & Wang, E. Impact of primer dimers and self-amplifying hairpins on reverse transcription loop-mediated isothermal amplification detection of viral RNA. *Analyst* **143**, 1924–1933 (2018).

226. Rolando, J. C., Jue, E., Barlow, J. T. & Ismagilov, R. F. Real-time kinetics and high-resolution melt curves in single-molecule digital LAMP to differentiate and study specific and non-specific amplification. *Nucleic Acids Res* **48**, (2021).
227. Zhu, X. *et al.* Multiplex reverse transcription loop-mediated isothermal amplification combined with nanoparticle-based lateral flow biosensor for the diagnosis of COVID-19. *Biosens Bioelectron* **166**, (2020).
228. Vindeirinho, J. M., Pinho, E., Azevedo, N. F. & Almeida, C. SARS-CoV-2 Diagnostics Based on Nucleic Acids Amplification: From Fundamental Concepts to Applications and Beyond. *Frontiers in Cellular and Infection Microbiology* vol. 12 Preprint at https://doi.org/10.3389/fcimb.2022.799678 (2022).
229. Rabe, B. A. & Cepko, C. SARS-CoV-2 detection using isothermal amplification and a rapid, inexpensive protocol for sample inactivation and purification. **117**, 24450–24458 (2020).
230. Yu, L. *et al. Rapid detection of COVID-19 coronavirus using a reverse transcriptional loop-mediated isothermal amplification (RT-LAMP) diagnostic platform*. http://primerexplorer.jp/lampv5e/index.html (2020).
231. Gärtner, K. *et al.* A fast extraction-free isothermal LAMP assay for detection of SARS-CoV-2 with potential use in resource-limited settings. *Virol J* **19**, (2022).
232. Schmidt, J. *et al.* A semi-automated, isolation-free, high-throughput SARS-CoV-2 reverse transcriptase (RT) loop-mediated isothermal amplification (LAMP) test. *Sci Rep* **11**, (2021).
233. Alekseenko, A. *et al.* Direct detection of SARS-CoV-2 using non-commercial RT-LAMP reagents on heat-inactivated samples. *Sci Rep* **11**, (2021).
234. Tamanaha, E., Zhang, Y. & Tanner, N. A. Profiling RT-LAMP tolerance of sequence variation for SARS-CoV-2 RNA detection. *PLoS One* **17**, (2022).
235. Chang, C. C. *et al.* Gold nanoparticle-based colorimetric strategies for chemical and biological sensing applications. *Nanomaterials* vol. 9 Preprint at https://doi.org/10.3390/nano9060861 (2019).
236. Sivakumar, R., Dinh, V. P. & Lee, N. Y. Ultraviolet-inducedin situgold nanoparticles for point-of-care testing of infectious diseases in loop-mediated isothermal amplification. *Lab Chip* **21**, 700–709 (2021).
237. Grammatikos, S., Svoliantopoulos, I. & Gizeli, E. Naked-Eye Detection of LAMP-Produced Nucleic Acids in Saliva using Chitosan-capped AuNPs in a Single-Tube Assay. doi:10.1101/2023.06.09.23291198.

238. Nguyen, L. T., Smith, B. M. & Jain, P. K. Enhancement of trans-cleavage activity of Cas12a with engineered crRNA enables amplified nucleic acid detection. *Nat Commun* **11**, (2020).

239. Chen, B. J., Mani, V., Huang, S. T., Hu, Y. C. & Shan, H. C. P. Bisintercalating DNA redox reporters for real-time electrochemical qLAMP. *Biosens Bioelectron* **129**, 277–283 (2019).

240. Park, J. W. Principles and Applications of Loop-Mediated Isothermal Amplification to Point-of-Care Tests. *Biosensors* vol. 12 Preprint at https://doi.org/10.3390/bios12100857 (2022).

241. Soares, R. R. G., Madaboosi, N. & Nilsson, M. Rolling Circle Amplification in Integrated Microsystems: An Uncut Gem toward Massively Multiplexed Pathogen Diagnostics and Genotyping. *Acc Chem Res* **54**, 3979–3990 (2021).

242. Kim, H. soo, Abbas, N. & Shin, S. A rapid diagnosis of SARS-CoV-2 using DNA hydrogel formation on microfluidic pores. *Biosens Bioelectron* **177**, (2021).

243. Chaibun, T. *et al.* Rapid electrochemical detection of coronavirus SARS-CoV-2. *Nat Commun* **12**, (2021).

244. Lee, H. N. *et al.* A Lateral Flow Assay for Nucleic Acid Detection Based on Rolling Circle Amplification Using Capture Ligand-Modified Oligonucleotides. *Biochip J* **16**, 441–450 (2022).

245. Walker, G. T. *et al. Strand displacement amplification-an isothermal, in vitro DNA amplification technique*. *Nucleic Acids Research* vol. 20.

246. Carter, J. G. *et al.* Ultrarapid detection of SARS-CoV-2 RNA using a reverse transcription-free exponential amplification reaction, RTF-EXPAR. **118**, (2021).

247. Ness, J. Van, Ness, L. K. Van & Galas, D. J. *Isothermal reactions for the amplification of oligonucleotides*. *PNAS April* vol. 15 www.pnas.orgcgidoi10.1073pnas.0730811100 (2003).

248. Zhou, M., Fan, C., Wang, L., Xu, T. & Zhang, X. Enhanced Isothermal Amplification for Ultrafast Sensing of SARS-CoV-2 in Microdroplets. *Analytical Chemistry* vol. 94 4135–4140 Preprint at https://doi.org/10.1021/acs.analchem.2c00008 (2022).

249. Yu, Y. *et al.* Reverse transcription recombinase polymerase amplification coupled with CRISPR-Cas12a for facile and highly sensitive colorimetric SARS-CoV-2 detection. *Anal Chem* **93**, 4126–4133 (2021).

250. Pham, J. *et al. Performance Characteristics of a High-Throughput Automated Transcription-Mediated Amplification Test for SARS-CoV-2 Detection*. https://doi.org/10 (2020).
251. Wu, Q. *et al. INSIGHT: A population-scale COVID-19 testing strategy combining point-of-care diagnosis with centralized high-throughput sequencing*. *Sci. Adv* vol. 7 (2021).
252. Ju, Y. *et al.* Rapid and accurate clinical testing for COVID-19 by nicking and extension chain reaction system-based amplification (NESBA). *Biosens Bioelectron* **196**, (2022).
253. Xing, W. *et al.* A High-Throughput, Multi-Index Isothermal Amplification Platform for Rapid Detection of 19 Types of Common Respiratory Viruses Including SARS-CoV-2. *Engineering* vol. 6 1130–1140 Preprint at https://doi.org/10.1016/j.eng.2020.07.015 (2020).
254. Gorzalski, A. J. *et al.* High-Throughput Transcription-mediated amplification on the Hologic Panther is a highly sensitive method of detection for SARS-CoV-2. *Journal of Clinical Virology* **129**, (2020).
255. Giaever, G. & Nislow, C. The yeast deletion collection: A decade of functional genomics. *Genetics* **197**, 451–465 (2014).
256. Kundu, S., Horn, P. J. & Peterson, C. L. SWI/SNF is required for transcriptional memory at the yeast GAL gene cluster. *Genes Dev* **21**, 997–1004 (2007).
257. Bheda, P. *et al.* Single-Cell Tracing Dissects Regulation of Maintenance and Inheritance of Transcriptional Reinduction Memory. *Mol Cell* **78**, 915-925.e7 (2020).
258. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033–1037 (2009).
259. Ntini, E. *et al.* Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol* **20**, 923–928 (2013).
260. Pefanis, E. *et al.* RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* **161**, 774–789 (2015).
261. Bresson, S., Tuck, A., Staneva, D. & Tollervey, D. Nuclear RNA Decay Pathways Aid Rapid Remodeling of Gene Expression in Yeast. *Mol Cell* **65**, 787-800.e5 (2017).
262. Van Nues, R. *et al.* Kinetic CRAC uncovers a role for Nab3 in determining gene expression profiles during stress. *Nat Commun* **8**, (2017).

263. Bheda, P. *et al.* Single-Cell Tracing Dissects Regulation of Maintenance and Inheritance of Transcriptional Reinduction Memory. *Mol Cell* **78**, 915-925.e7 (2020).
264. Bheda, P. Metabolic transcriptional memory. *Molecular Metabolism* vol. 38 Preprint at https://doi.org/10.1016/j.molmet.2020.01.019 (2020).
265. Drzewoski, J., Kasznicki, J., Trojanowski, Z. & Chorób Wewnętrznych Oddziałem Diabeto logii, K. *The role of 'meta bolic memory' in the natural history of diabetes mellitus*. *Pol Arch Med Wewn* vol. 119 (2009).
266. The "Metabolic Memory" Theory and the Early Treatment of Hyperglycemia in Prevention of Diabetic Complications.
267. Mao, X. *et al.* Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. *Molecular Cancer* vol. 20 Preprint at https://doi.org/10.1186/s12943-021-01428-1 (2021).
268. Petitprez, F., Meylan, M., de Reyniès, A., Sautès-Fridman, C. & Fridman, W. H. The Tumor Microenvironment in the Response to Immune Checkpoint Blockade Therapies. *Frontiers in Immunology* vol. 11 Preprint at https://doi.org/10.3389/fimmu.2020.00784 (2020).
269. Gajewski, T. F., Schreiber, H. & Fu, Y. X. Innate and adaptive immune cells in the tumor microenvironment. *Nature Immunology* vol. 14 1014–1022 Preprint at https://doi.org/10.1038/ni.2703 (2013).
270. Barnabei, L., Laplantine, E., Mbongo, W., Rieux-Laucat, F. & Weil, R. NF-κB: At the Borders of Autoimmunity and Inflammation. *Frontiers in Immunology* vol. 12 Preprint at https://doi.org/10.3389/fimmu.2021.716469 (2021).
271. Kayama, H., Okumura, R. & Takeda, K. Interaction Between the Microbiota, Epithelia, and Immune Cells in the Intestine. *IY38CH02_Takeda ARjats.cls April* **5**, (2020).
272. Zhang, X. M., Liu, C. Y., Shao, Z. H. & Guo, L. S. Advances in the role of helper T cells in autoimmune diseases. *Chinese Medical Journal* vol. 133 968–974 Preprint at https://doi.org/10.1097/CM9.0000000000000748 (2020).
273. Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science (1979)* **356**, 438–442 (2017).
274. Gootenberg, J. S. *et al.* Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a and Csm6. *Science (1979)* **360**, 439–444 (2018).
275. López-Valls, M. *et al.* CASCADE: Naked eye-detection of SARS-CoV-2 using Cas13a and gold nanoparticles. *Anal Chim Acta* **1205**, (2022).

276. Chen, J. S. *et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity One Sentence Summary*. vol. 360 (2018).
277. Nguyen, L. T., Smith, B. M. & Jain, P. K. Enhancement of trans-cleavage activity of Cas12a with engineered crRNA enables amplified nucleic acid detection. *Nat Commun* **11**, (2020).
278. Santiago-Frangos, A. *et al.* CRISPR-Cas, Argonaute proteins and the emerging landscape of amplification-free diagnostics. *Methods* **205**, 1–10 (2022).
279. Shi, K. *et al. A CRISPR-Cas autocatalysis-driven feedback amplification network for supersensitive DNA diagnostics*. *Sci. Adv* vol. 7 (2021).