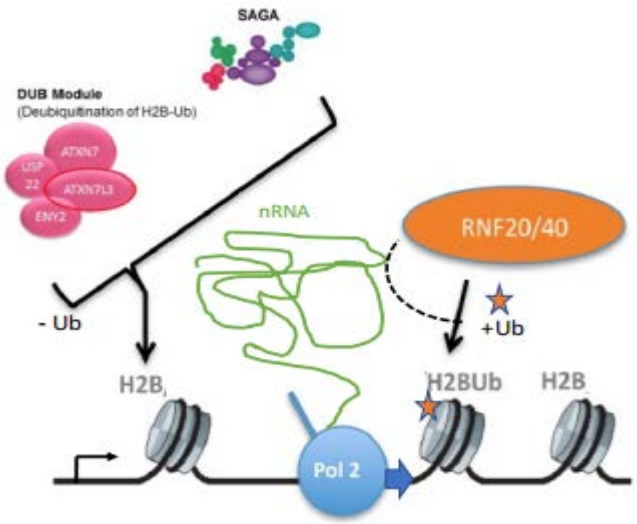
ΙΔΡΥΜΑ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΕΡΕΥΝΑΣ

Master thesis

Title: ***Designing tools for discovering RNF20/40 mode of action in transcription regulation***



Thesis Committee:

Lavigne Matthieu (Thesis Supervisor, Researcher IMBB-FORTH)

Giakountis Antonis (Assistant Professor Biochemistry and Biotechnology, University of Thessaly)

Pavlopoulos Anastasios (Researcher IMBB-FORTH)

Vaios Theodosiou **DECEMBER 2022**

ABSTRACT

Post-translational modifications (PTMs) of histones play a central role in regulating all nuclear processes requiring access to DNA in all eukaryotic cells and are extremely important in keeping human cells healthy. Recent studies suggest that the balanced activity of RNF20/40 (E3-Ubiquitin ligase) and the de-ubiquitination (DUB) module of SAGA determines the structural properties of histone H2B and controls transcription elongation. The evidence that RNF20/40 potentially binds nascent RNA directly raised our interest to determine if nRNA shape and size can influence H2Bub-associated functions. Furthermore, findings that loss of RNF20/40 and H2B monoubiquitylation (H2Bub) is observed in several cancers and is linked to an aggressive phenotype, and is also an indicator of poor prognosis. In this study, we aimed to create important molecular and cellular tools that will be important to map RNF20 DNA and RNA binding sites. Preliminary loss-of-functions or overexpression results also suggest that RNF20 functions are essential in Huh7 cells and determine the growth of many cancer cell lines. This thesis sets the stage for the characterization of how RNF20 is recruited and operates in the genome.

ΠΕΡΙΛΗΨΗ

Οι μετα-μεταφραστικές τροποποιήσεις (PTMs) των ιστονών διαδραματίζουν κεντρικό ρόλο στη ρύθμιση όλων των πυρηνικών διεργασιών που απαιτούν πρόσβαση στο DNA σε όλα τα ευκαρυωτικά κύτταρα και είναι εξαιρετικά σημαντικές για τη διατήρηση της υγείας των ανθρώπινων κυττάρων. Πρόσφατες μελέτες υποδεικνύουν ότι η ισορροπημένη δραστηριότητα του RNF20/40 (λιγάση της ουβικιτίνης) και του συμπλόκου της αποουβικουιτίνωσης (DUB) του SAGA καθορίζει τις δομικές ιδιότητες της ιστόνης H2B και ελέγχει την επιμήκυνση της μεταγραφής. Οι ενδείξεις ότι το RNF20/40 δεσμεύει δυνητικά το εκκολαπτόμενο RNA προκάλεσε άμεσα το ενδιαφέρον μας να προσδιορίσουμε εάν το σχήμα και το μέγεθος του nRNA μπορούν να επηρεάσουν τις λειτουργίες που σχετίζονται με το H2Bub. Επιπλέον, τα ευρήματα ότι η απώλεια του RNF20/40 και της μονοουβικιτίνωσης της H2B (H2Bub), παρατηρείται σε αρκετούς καρκίνους και συνδέεται με επιθετικό φαινότυπο και είναι επίσης δείκτης κακής πρόγνωσης. Σε αυτή τη μελέτη, στοχεύσαμε να δημιουργήσουμε σημαντικά μοριακά και κυτταρικά εργαλεία που θα είναι σημαντικά για τη χαρτογράφηση των θέσεων δέσμευσης DNA και RNA του RNF20. Τα προκαταρκτικά αποτελέσματα απώλειας λειτουργίας ή υπερέκφρασης υποδηλώνουν επίσης ότι οι λειτουργίες του RNF20 είναι απαραίτητες στα κύτταρα Huh7 και καθορίζουν την ανάπτυξη πολλών καρκινικών κυτταρικών σειρών. Αυτή η διατριβή θέτει τη βάση για τον χαρακτηρισμό του τρόπου με τον οποίο ο RNF20 στρατολογείται και δρά στο γονιδίωμα.

INTRODUCTION

Composition and structure of chromatin

All organisms, prokaryotic and eukaryotic, face the challenge of packaging the disproportionately long DNA molecule within the confined space of the cell. The genome of eukaryotic organisms varies in size from 10^8 bp (C. Elegans) to $5{,}4 \times 10^9$ (Z. mays), while the haploid genome of a human cell consists of 3×10^9 bp. Despite this size, which would correspond to a length of about one meter if unpacked, the human genome is packed inside the cell's nucleus in a space equivalent to only about 10 micrometers in diameter. The spatial organization of the genetic material is based on a mechanism of gradual (step-wise and reversible through cell cycle) condensation of DNA and various proteins (distinguished between histones and non-histones), which together constitute chromatin . 146 bp of DNA wraps around complexes (octamers) of histone proteins to form nucleosomes [1] (Figure1), which bundle into so called "beads on a string" structure [2] (Figure 2). Further condensation of the chromatin is achieved towards a helical, secondary structure (chromatin fibril 30nm), and further chromatin packing occurs to form mitotic chromosomes.

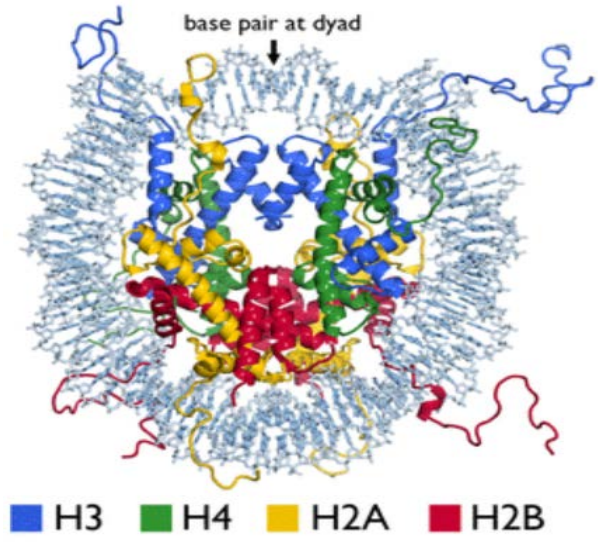


Figure 1: Nucleosome core particle structure. Histones and DNA are depicted in cartoon and stick representations, respectively, and colored as indicated. McGinty and Tan, 2015

Eukaryotic nucleosomes organize chromatin by wrapping 147 bp of DNA around a histone core particle comprising two molecules each of histone H2A, H2B, H3 and H4. The DNA entering and exiting the particle may be bound by the linker histone H1. Whereas deposition of bulk histones is confined to S-phase, paralogs of the common histones, known as histone variants, are available to carry out functions throughout the cell cycle and accumulate in post-mitotic cells[3]. Histone variants confer different structural properties on nucleosomes by wrapping more or less DNA or by altering nucleosome stability. They carry out specialized functions in DNA repair, chromosome segregation and regulation of transcription initiation, or perform tissue-specific roles.

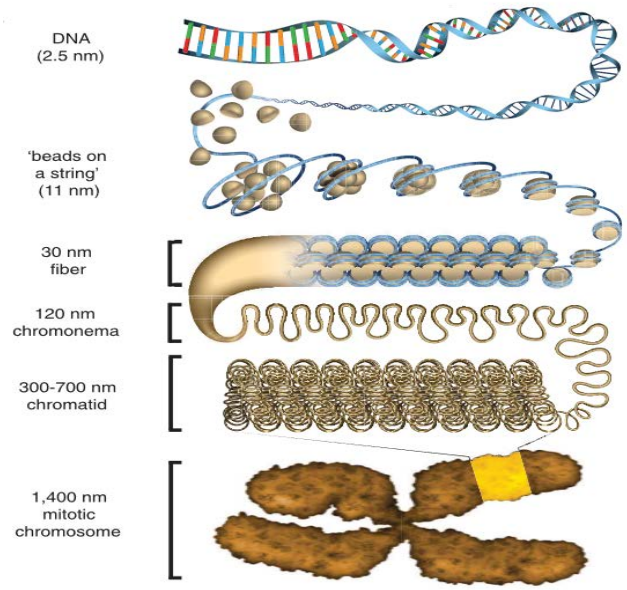


Figure 2: Hierarchical chromatin-folding model. Hd, et. al. 2017

Histones are the basic proteins that enable the first level of folding of chromatin. Their great functional and structural importance is attested by their highest degree of evolutionary conservation. Histones were identified in 1884 and characterized by Albrecht Kossel. All members of the histone family are relatively small with a molecular weight of 10-12 kDa and with neutral pH on average. They are rich in positively charged amino acids lysine, arginine and histidine. The core histones are H2A, H2B, H3 and H4 and can be interchanged in particular contexts

with numerous variant histones (see above), which often confer different structural properties on nucleosomes and often have distinct functions in cell division, transcription, DNA repair, differentiation and chromatin remodeling[3]. The linking histones H1 and H5 (H5 found only in birds) are important structural components for chromatin higher-level packaging[4]. H3-H4 forms a heterodimer and then a tetramer (H3-H4)2 on which the two H2A-H2B dimers bind. A nucleosome is defined as ~146 bp of DNA wrapped 1.6 times around histone octamers (Figure 3) but depending on the type of DNA (B, z), a different number of base pairs are included in each turn[5].

Comment [ml1]: Either you put a number or the name… (if numbers the refs should be listed by number at the ref section)

To better understand how chromatin structure is regulated, it is important to look at how DNA wraps around histones. Each histone is packed into the nucleosome through its carboxyl-terminus which has a characteristic conformation known as the "histone fold", while its amino-terminus protrudes from the nucleosome and is very often subject to post-translational modifications[6]. These modifications play a particularly important role in shaping the chromatin structure, and these physico-chemical parameters directly impact gene expression regulation. There are three main ways by which chromatin structure changes at the nucleosome level: 1) nucleosome remodeling, 2) nucleosome removal (mainly done by ATP-consuming enzymes known as ATP-dependent chromatin remodeling complexes) and 3) histone replacement (carried out by histone chaperones that replace normal histones with histone variants).

Nucleosomes are found on average every 200 base pairs throughout the genome, but this distance is variable. There are regions of heterochromatin, which appear mostly as transcriptionally inactive, where distribution of nucleosomes tends to be uniform and dense. Other regions of so-called euchromatin, which are characterized as transcriptionally active, display irregular and sparser spacing [7]. (Figure 3)

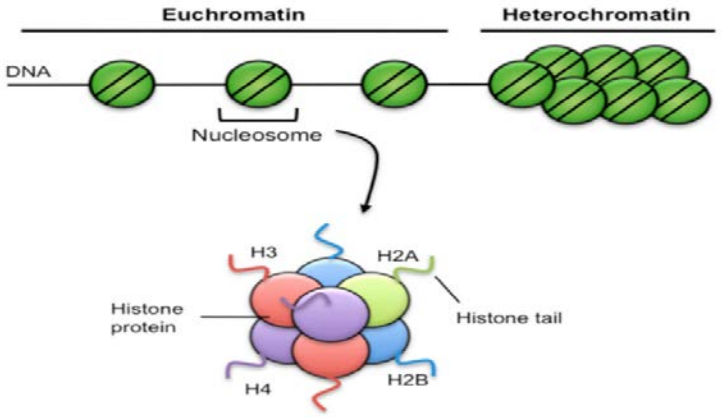


Figure 3: Chromatin structure. Chromatin has two broad structures. The first is euchromatin, characterized by sparse nucleosome density and is generally associated with active gene transcriptional activity. Heterochromatin is characterized by high nucleosome density, is very

compacted and is generally associated with repression of gene transcription. Nucleosomes consist of 147 bp of DNA wound 1.6 turns around a complex of histone proteins, comprising two of each of the H2A, H2B, H3, and H4 histones. Each histone has a soluble amino terminal tail that can be covalently modified by specific epigenetic marks such as acetylation, methylation, and phosphorylation. Huisinga et. al. 2006

Regulation levels of gene expression

Robust Spatial-temporal patterning of gene expression across the population of cells is essential during differentiation processes and/or during response to extracellular signals. Mechanisms of regulation of gene expression are multi-layered and complex. Two main levels of control are found in the process of expressing the information encoded in a gene: first transcription into mRNA and then translation of mRNA into protein. In the case of eukaryotic genes, two more levels are distinguished. Specifically, because the most eukaryotic genes are discontinuous or interrupted (contain introns), between transcription and translation there is another level, the so-called maturation of mRNA, during which introns are excised and exons are spliced. Next, in eukaryotic cells proteins undergo additional modifications after their synthesis (post-translational) in order to enable full/specific functions.

One can summarize the levels of gene regulation as follows:

1. **Transcriptional control**. It determines the amount, frequency and rate of transcription of each gene. It is the stage in which the synthesis and interpretation of the extracellular stimuli that will determine the pattern of gene expression. The correct transcriptional control is the most common and most diverse way of regulating gene expression, as it acts on the first stage of expression of genetic information and prevents waste of materials and energy.
2. **Post-transcriptional control.** It first controls the manner and rate of processing of synthesized mRNAs and then the rate of their export from the nucleus and their destination in the cytoplasm. More specifically, it is determined which of the possible alternative products of maturation (alternative splicing) will occur and how quickly this will occur. Then, the mature mRNA is controlled by specific protein complexes that allow or deny its exit from the nucleus and direct it to its site of translation, in a specific compartment of the cytoplasm. At the level of post-transcriptional control, also acts the RNA interference pathway.
3. **Translation control.** It chooses which mRNA will be translated, by how many ribosomes, and how many times. Essentially, at this level the lifetime of each mRNA is controlled on the one hand and therefore the duration during which it is available for translation and on the other hand the frequency of formation of the preinitiation complex and the rate of elongation of the newly synthesized protein.
4. **Post-translational control.** It controls all modifications (eg, amino acid truncation, phosphorylation, ubiquitination, glycosylation, acetylation or

addition of other molecules) after protein synthesis, which will lead to the functional maturation and proper targeting of the protein within the cell.

The expression of most genes is primarily regulated at the transcriptional level. The most important factors of transcriptional regulation are the cis and trans regulatory elements of a gene, i.e., the DNA regulatory sequences (promoter, enhancer) and the proteins (transcription factors, co-activators, repressors, mediators) that recognize and bind specifically to these sequences, respectively. The packaging of the genome of higher eukaryotic organisms in the form of chromatin adds another level of regulation of a gene's transcriptional activity. In fact, the degree of chromatin compaction (euchromatin/heterochromatin) and the local chromatin architecture (position and composition of nucleosomes) are dynamic features that can be modified by the action of specific macromolecular complexes and enzymes (complexes chromatin remodeling, histone acetyltransferases, methylases, etc.), determining if and when the respective gene will be expressed.

Transcription consists of multiple regulated steps:

1. In many cases transcription begins when a pioneer factor (can directly bind condensed chromatin) binds to nucleosome covered DNA and increases its accessibility by recruiting nucleosome remodelers and histone acetyltransferases, that can free up nucleosome-depleted regions[8].
2. As chromatin opens, DNA elements become accessible to additional transcription factors (TFs). TFs binding can drive spatio-temporal specific binding of the preinitiation complex (PIC) in underlying regulatory regions[9]. This assembly of approximately 100 proteins works together with RNA polymerases to initiate transcription of protein-coding and non-coding RNA genes.
3. After PIC is formed, TFIIH functions as a helicase to unwind the double-stranded DNA and allows pol 2 to initiate transcription. (In eukaryotes, polymerase I and III synthesizes ribosomal RNAs and small structural RNAs, respectively, whereas polymerase II produces protein-coding mRNAs, long non-coding RNAs, primary microRNAs and enhancer RNAs).
4. The pol II initially transcribes 20-60 nucleotides before promoter-proximal pausing. This is a regulatory checkpoint for execution of transcription programs and a major rate-limiting step in gene expression. This step is essential for nascent RNA protection by 5' capping.
5. Pausing of Pol II is stabilized by NELF and DSIF[10]. Release from the pause stage occurs after P-TEFb activation and binding. This complex phosphorylates NELF-E, DSIF and Ser2 residues of Pol II CTD (the C-terminal domain is the largest subunit of the RNA pol II), thanks to cyclin dependent kinase 9 (cdk9).
6. During productive elongation, various elongation factors promote the process of pol II and also connect the nascent transcription to processes such as co-transcriptional splicing[11].
7. Finally, the transcript is cleaved when Pol 2 reaches the 3' end (transcription termination site (TTS) and polyadenylation site) signaling for pre-mRNA polyadenylation. The pre-mRNA is released from Pol II after cleavage and

short nascent RNA synthesized after TTS are exposed to XRN2 digestion, which destabilizes the Pol II via the torpedo mechanism. Once the terminated Pol II dissociates from the DNA it can be recycled to a new round of transcription on the same or other neighboring promoters.

Histone modifications

Histone post-translational modifications (PTMs) are covalent additions of chemical groups to the amino-terminal of histones, although there are modifications that take place in the folding structures of histones, as well as at the C-terminal[12]. The known histone modifications are: 1) methylation, 2) acetylation, 3) ubiquitination, 4) phosphorylation, 5) biotinylation, 6) sumoylation, 7) isomerization, 8) deamination, 9) ADP-ribosylation. These modifications can be carried out at specific amino acid residues of each histone and their addition (also known as "writing") or removal ("erasing") is very dynamic and highly regulated as a result of the action of specialized enzymes[13] (Figure 4). Covalent modifications of histones can be linked to transcriptional activity of neighbor genes (e.g., hyperacetylation of histones H3 and H4[14] as it facilitates the access of the transcriptional machinery, or with gene repression, which depends on specific lysine residues that are methylated e.g, H3K27me3[15].

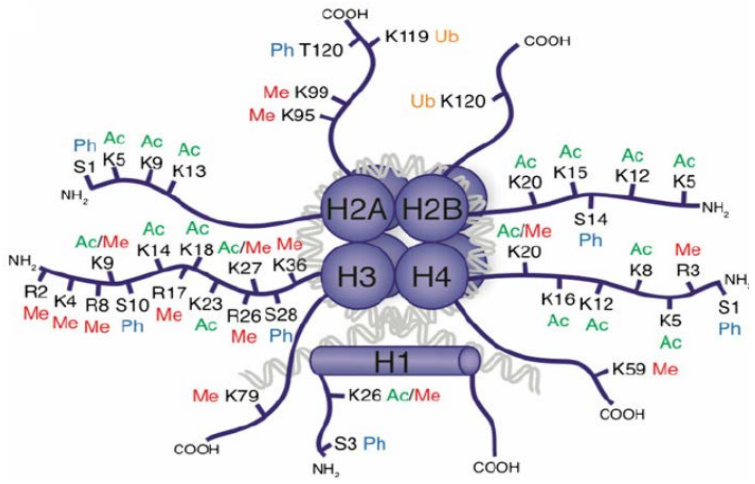


Figure 4: Main covalent histone post- translational modifications PTMs are highlighted on the N- and C- terminal tails of each histone.

Me=methylation, Ac= acetylation, Ub=ubiquitination, Ph=phosphorylation

An important modification addressed in this thesis is the monoubiquitination of the histone H2B at lysine 120. First, histone H2B is a core histone protein that determines the structure of the nucleosomes in eukaryotic cells. Like all core histones, H2B consists of two domains: a basic N-terminal domain and a histone-fold C-terminal domain[16,17]. The histone-fold domain has two defined characteristics: a) it heterodimerizes with a second histone (H2B with H2A), b) two heterodimers then associates to (H3-H4)2 tetramer to assemble into an octamer, around which wraps DNA to form the nucleosome. The basic N-terminal 'tail' domains lie at the periphery of the nucleosome and they do not have any defined structure. H2B sequence consists of 126 amino acids, many of these have a positive charge at cellular pH (7.0-7.4)[18]. This positive charge allows them to interact with the negatively charged DNA. Finally, 13 somatic histone H2B isoforms have been identified by various biochemical and mass spectrometric (MS) approaches (Figure 5) and also H2B isoforms are expressed at varying levels in different cells, suggesting isoform-specific, and possibly cell-type-specific, H2B gene regulation[19].

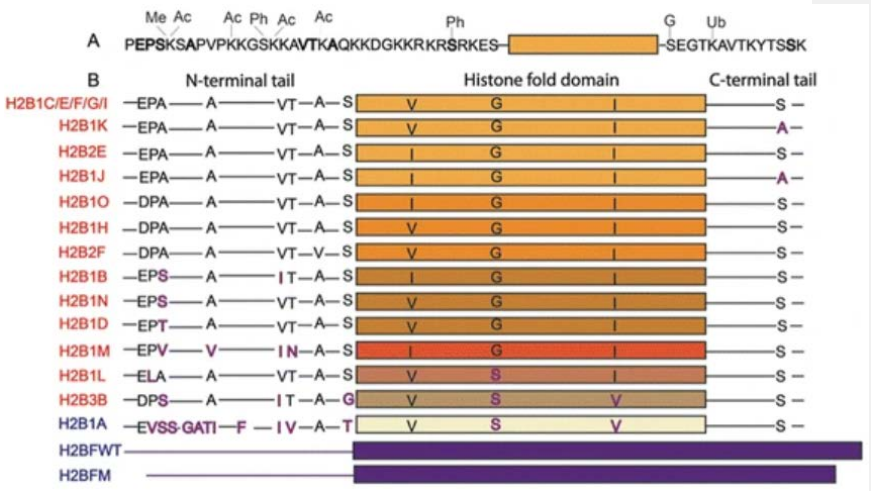


Figure 5: Points of sequence variation and modifications on human H2B isoforms. ***(A)*** *Modification sites on the H2B N-terminal tail and C-terminal tail. Sites of sequence variation are in bold.* ***(B)*** *Sequence alignment of H2B isoforms. Points of sequence variation are highlighted in purple. Somatic isoforms are listed in red, and testis-specific variants are in blue. Variants H2BFM and H2BFWT differ significantly from other H2B isoforms across their entire sequences. Me = methylation, Ac = acetylation, Ph = phosphorylation, G = glycosylation, Ub = ubiquitination. RC Molden et. al. 2015*

Ubiquitination of histone H2B in mammalian cells was identified over three decades. Ubiquitin (ub) is a highly conserved protein of 76 amino acids that plays a critical role in cellular homeostasis. Ub (alone or conjugated to another protein) accounts for approximately 0,1-5% of total proteins within cells[20,21]. In mammalian cells, ub has a half-life of about 28-31 hours. Ubiquitination is catalyzed in three steps by three types of enzymes: E1, E2, and E3. E1 is a ubiquitin-activating enzyme that loads

ubiquitin in an ATP-dependent manner. E1 transfers ubiquitin to E2, which is a ubiquitin-conjugating enzyme[22]. Attachment of ub to target proteins is dictated in a specific manner by E3 thanks to one of the more than 600 E3 ligases known to date. Specifically, E3 ub ligase transfers the ub from the E2 enzyme onto a lysine residue on the target protein, via an isopeptide bond. This bond can be then cleaved by DUB enzymes (Figure 6). The DUB module of the SAGA (Spt-Ada-Gcn5 acetyltransferase) transcriptional coactivator complex is the primary complex responsible for removing ub from histone H2Bub in eukaryotes[23]. DUBs can liberate ub from target proteins, they can impact the abundance of free ub in the cell, which can then be re-used to target other proteins[24]. Histone proteins are major targets for ubiquitination, with about 5–15% of histone H2A and 1–2% of histone H2B being monoubiquitinated. Monoubiquitination (ub1) of histone H2B at lysine 120 in metazoan cells plays an important role in cellular homeostasis[25]. The addition of a protein like ub may be a point of access to other factors that lead to differential regulation of chromatin and can impact on gene expression levels. So far, we know that the monoubiquitination of H2B affects the assembly/disassembly of nucleosomes and therefore affects chromatin relaxation/compaction[26]. It has also been shown that H2Bub regulates transcription based on its genomic position. Within genes H2Bub appears to stimulate transcriptional elongation while its presence at promoters is suggested to inhibit Pol II recruitment[27,28]. In parallel, the presence of H2Bub inhibits transcription in low-expressed genes, while activating it in highly expressed genes[29].

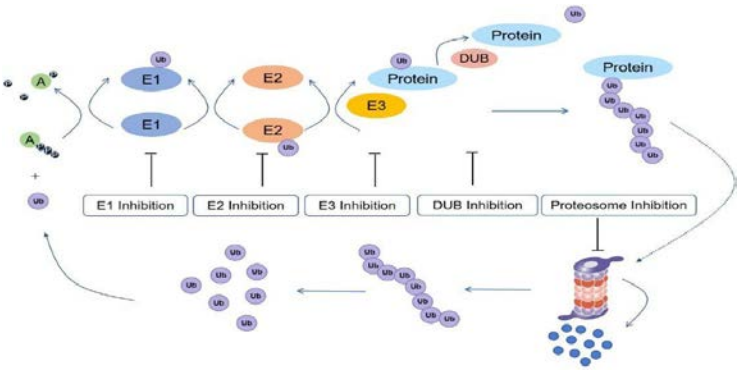


Figure 6: Ubiquitination process and potential drug inhibition targets (E1: ubiquitin-activating enzyme; E2: ubiquitin conjugating enzyme; E3: ubiquitin protein ligase; Protein: substrate/target protein; DUB: deubiquitinating enzyme). S.Q Song, 2021

Experiments that have been done in the past have shown that presence of H2Bub leads to disruption of chromatin compaction (in vitro reconstituted nucleosome arrays)[26], while other studies have found that it stabilizes nucleosomes (MNase sensitivity assays)[30]. These conflicting results make us question why it is difficult to accurately describe the role of H2Bub in different experiments and also when analyzing individual genes in a snapshot of time. However, we can definitely conclude that the local effect of this modification on the compression/relaxation of the chromatin is dynamic and complicated and probably depends on spatio-

temporal physico-chemical conditions imposed at given genomic loci. It is worth mentioning that monoubiquitination of H2B K120 is involved in histone cross-talk with H3K4 and H3K79 methylation, which are both hallmarks of actively transcribed genes. Recent advances in molecular and genomics techniques have, however, revealed that the trans-tail crosstalk is linked to a more diverse cascade of histone modifications and has various functions in co-transcriptional processes. A relatively recent paper shows a comprehensive mechanism of cross-talk between histone ub and methylation. The methyltransferase DOT1 interacts with the tail of histone H4 and this leads to the change in conformation in the globular core of H3, so K79 is exposed for methylation [31](Worden et al, 2019). Finally, H2Bub regulates FACT genic distribution, histone binding, and chaperone activity[32]. FACT (facilitates chromatin transcription) is an essential histone chaperone and as its name suggests, it plays a role in the elongation and removal of nucleosomes[33].

Nascent RNA role on transcription regulation

The transcription machinery transcribes much more RNA, in terms of sequence, than is needed to code for a protein. Taking under consideration that energy is needed to carry out cellular processes, such as transcription, it would be surprising if the cell spent energy without any significant reason. Eukaryotic genomes are known to transcribe several classes of RNA (as discussed above), and beyond regulatory small and long-non coding RNAs it is suggested that unprocessed nascent RNA from protein-coding genes, could also regulate gene expression. This suggests that RNAs are both products and regulators of gene regulation. RNA molecules contain sequences that allow them to interact with DNA, RNA or proteins in a specific or nonspecific manner. Interestingly, several DNA-binding proteins are also capable of directly binding RNA[34]. RNA-binding proteins (RBPs) are key players in the regulation of gene expression in all organisms and play important roles in post-transcriptional processes in all eukaryotes, such as regulation of splicing, mRNA transport, and regulation of mRNA translation and decay. Nascent RNA are suggested to regulate gene expression through RBPs binding and this hypothesis is further strengthened by the fact that more and more Pol 2 associated elongation factors (EF) have been shown to first bind to nRNA before being assembled into the functional Pol II elongation complex[35]. ChIP-seq experiments showed that several RBPs located in important regions of chromatin (in terms of gene expression), participate directly in transcriptional control[36]. A new perspective on the role of nascent RNA posits that it may govern the global recruitment or eviction of regulatory factors to /from chromatin. A recent study suggests that nascent RNA, besides recruiting RNA processing factors, antagonizes interactions of a set of transcriptional regulators with chromatin, perhaps functioning as an ‘‘RNA control center’’ to regulate factor recruitment and activity during the transcription process[37] (Figure 7). One more recent study suggests how nascent RNA-rich Transcription Loops (TLs) are forming while transcribing Pol2 convoys travel on gene bodies and can modify the shape of the harboring genomic loci, because large nascent

ribonucleoprotein particles (RNPs) make their space in the 3D structure of the nucleus[38] (Figure 8). Despite these exciting advances, the question of what is the impact of such nRNA on gene regulation remains largely open, but the development and progress of techniques which capture nascent RNA is promising for the immediate future.

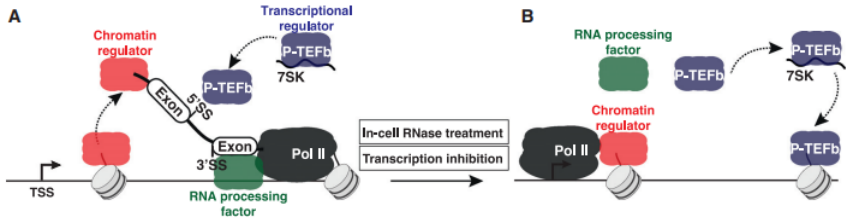


Figure 7: (A) The studies by Skalska et al. (2021) suggest: (1) chromatin regulators are relocated from chromatin to nascent RNA during transcription, (2) P-TEFb is transferred from 7SK to 50 SS, and (3) RNA processing factors are tethered to Pol II through interactions with pre-mRNA. (B) Upon in-cell RNase treatment or transcription inhibition: (1) chromatin regulators gain interactions with chromatin, (2) P-TEFb is released from nascent RNA and associates with chromatin in a 7SK-dependent manner, and (3) RNA processing factors are evicted from chromatin.

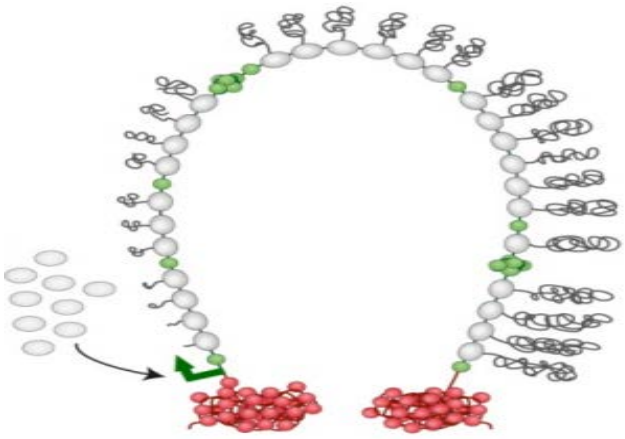


Figure 8: Transcription loops (TLs) are decorated by elongating RNAPIIs moving along the gene axis and carrying nRNAs that undergo co-transcriptional splicing. *Leidescher et al. 2022*

RNF20/40 complex

In mammalian cells, E3 ubiquitin ligases can be divided into two major families based on the presence of either a really interesting new gene (RING) finger domain or a homologous to the E6-AP carboxyl-terminus domain. Mammalian Bre1 complexes (BRE1A/B (RNF20/40) in humans) function similarly to their yeast homolog Bre1 as ubiquitin ligases in monoubiquitylation of histone H2B. RNF20 genes encode for a RING protein, which consists of 975 amino acids with a predicted mass of 113662 Da and is located in chromosome 9 in humans. RNF20 protein contains three coiled-coil domains and a C-terminal RING finger domain, the latter of which is implicated in protein–DNA (protein-RNA?) and protein–protein interactions. RNF20 heterodimerizes with RNF40, which consists of 1001 amino acids (113678 Da, chromosome 16), which contains one more coiled-coil domain than RNF20[39]. RNF20/40 activity depends on RAD6 (UBE2A/UBE2B) E2 ub conjugating enzyme and interacts with WAC to facilitate H2Bub in the vicinity of elongating Pol 2[40]. RNF20/40 complexes and PAF1 are recruited and stabilized at active gene promoters by direct binary interactions and cooperate to regulate P-TEFb[28].

Remarkably, recent studies have shown that the substantial changes in specific exon inclusion (alternative splicing) are not associated with significant change in relative H2Bub levels within the affected exons or their surrounding introns. Therefore, exonic H2Bub patterns are not dictated by the selective activity of the splicing machinery[41]. On the other hand, RNF20/40 is a binding partner of p53 and plays an important role in tumor suppression by regulating the transcription and mRNA splicing of p53 target genes[42]. Along these lines, it has been reported that PRPF8 (Pre-mRNA-splicing factor 18) has a ubiquitin-binding motif (UBM) at the C-terminus[43] and H2Bub is important for the recruitment of PRPF8 for p21 and PUMA mRNA splicing.

Other functions of RNF20 include its role in DNA damage repair (DDR), a corrective process by which a dedicated cellular signaling network acts. Ten years ago, RNF20/40 was identified as a novel element of the DDR that acts through modifications of H2Bub[44]. During the abduction of double strand breaks (DSB), the ATM phosphorylates ser172 and ser553 residues of RNF20 and ser114 of RNF40. Phosphorylated RNF20/40 is then recruited to sites of DSB, where it creates H2Bub modifications associated with DNA damage, thereby facilitating chromatin opening and increasing accessibility for DNA repair proteins (BRCA1, BRCA2, RAD51)[45,46].

In addition to H2B, other non-histone substrates have been reported for the RNF20/40 complex. For instance, RNF20/40 monoubiquitinated Eg5, a key player in the assembly of the spindle during mitosis[47], and Eef1BδL, a heat shock transcription factor[48]. The balance between ubiquitination and deubiquitination is essential for normal cellular function and alteration in this balance often leads to pathogenesis (i.e. carcinogenesis). The reduction of H2Bub and RNF20/40 levels observed in some advanced cancers, suggest that these enzymes may be cancer markers and new targets for cancer therapy, but the exact molecular role of RNF20 seems to vary in different types of cancers. There are situations like breast cancer, where in RNF20/40 was found to be highly expressed compared with normal human breast epithelial cell lines. On the other hand, it was found that depletion of RNF20 increased breast cancer cell proliferation and migration potential[49](Figure 9).

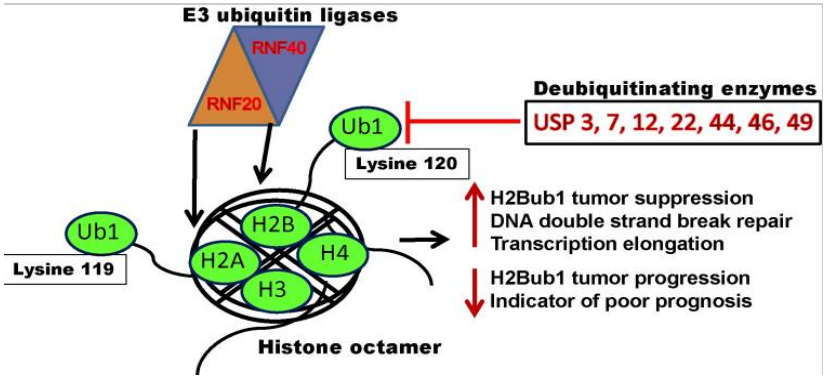


Figure 9: Reversible post-translational modification (PTM) ubiquitination of H2A (K119) and H2B (K120) on specific amino acid residues on core histone tails regulates various biological processes, including transcription elongation, inflammation, DNA replication, and DNA DSB repair processes, and is highly dynamic in nature. G. Sethi et al. 2018

Last examples of the difficulty to assign a beneficial or detrimental role for RNF20/40 are that, RNF20 also represses transcription elongation by inhibiting TFIIS binding to chromatin. TFIIS depletion selectively suppresses the activation of proto-oncogenes. Excessive cancer activity can promote TFIIS. In contrast, recent publications claim that the expression of RNF20 contributes to the development of breast cancer, through the epigenetic suppression of the expression of E-cadherin.

Experimental hypothesis

Nascent transcripts (RNA molecules still tethered to RNA polymerase) are overlooked chromatin components in respect to their potential regulatory role. The possible impact of nascent RNA molecules on health has thus recently gained attention with phenomena such as inappropriate retention of introns, alternative splicing or even widespread premature intronic polyadenylation demonstrated to inactivate tumor suppressor genes in cancers[50–52]. These facts combined with the rapid progress of genome- and proteome-wide techniques for capture of nascent RNA and associated DNA and proteins could be a good opportunity to shed more light on the exact role of nascent RNA as a regulatory factor in transcription. As mentioned in the introduction, the mammalian complex RNF20/40 acts as a ubiquitin ligase and is responsible for monoubiquitylation of histone H2B (H2Bub). This step is transient (reversible) for H2Bub and facilitates subsequent methylation of histone H3 at K4 and K79 of the same nucleosome via histone crosstalk, a process that facilitates Pol 2 elongation in transcribed genes.

The long-term goal of this project is to understand what regulates histone H2Bub localization/ concentration in the genome. Data from previous experiments of the laboratory and from other groups demonstrate that transcription elongation rate depends on a well-balanced activity of ub writing by RNF20/40 and deubiquitylation (DUB) by the USP22-module of SAGA. Together this turnover modifies the structural properties of histone H2B-H2A dimer, and therefore modulates chromatin remodeling activity in its vicinity, especially when elongating RNA polymerase traverses the nucleosome. A strong argument for formulating our hypothesis is preliminary evidence that RNF20 can interact with nascent RNA. Indeed, a study listing potential proteins interacting with EU-labeled/purified nascent RNA listed RNF 40 as a strong candidate[53]. Moreover, recent results from our group (Fanourgakis et al, in revision) established how H2Bub levels dropH2Bub abruptly after each intron-exon junction along gene bodies (Figure 10). It would therefore be important to decipher the exact molecular mechanism that rules the writing/erasing of this PTM, and how the involved regulators determine proper gene expression in homeostasis or disease. This line of investigation might provide new therapeutic opportunities in some cancer types or other transcription-related diseases, where these processes are impaired.

During my MSc thesis, I developed essential molecular tools that will allow us to test the above hypothesis and demonstrate how RNF20/40 can control transcription output. In this context, I first proceeded to engineer plasmid constructs that allow to knock out (KO) RNF20 or RNF 40 using the CRISPR/Cas9-nickase technology. Plasmids were then transfected in human hepatocarcinoma cells Huh7 to test for loss-of-function effects. In parallel, I have purified and transfected plasmids designed to overexpress (OE) exogenously RNF20 and RNF40 proteins tagged with FLAG and HA sequences, respectively. This strategy is important to resolve inherent difficulties to ChIP endogenous proteins due to lack of efficient antibodies against RNF20 and RNF40 on the market for Chip-seq or RIP-seq experiments. Overall, the generation of

these KO and OE stable cell lines is essential for future genome-wide characterization of RNF20/40 functions.

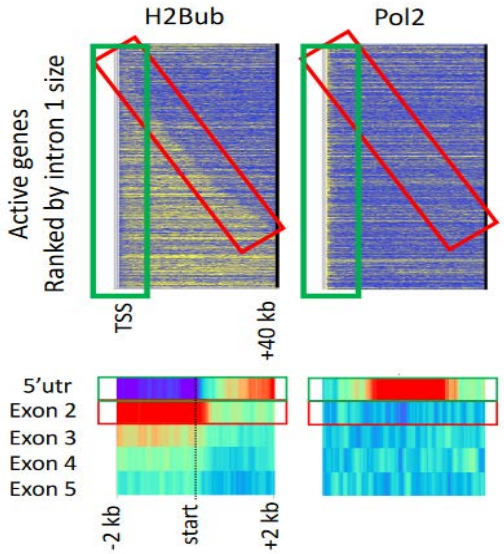


Figure 10: Preliminary chip-seq data showing the distribution of H2Bub (left) and pol2 (right) along active genes longer than 40kb sorted to intron 1 size (top) and across consecutive intron-exon junctions (e2-e5).

Material and methods

CRISPR/Cas9 technology

CRISPR/Cas9 genome editing technology was adapted from the natural genome editing system that bacteria use as an immune system to edit genomes of complex animals and has helped scientists understand the role and function of a growing number of target sequences. When this system is introduced into cells, the guide RNA recognizes the intended DNA sequence and the Cas9 enzyme cuts the DNA at the targeted location. For Cas9 to cut, it must recognize a so-called PAM sequence at the targeted loci. Therefore, this PAM sequence serves as a binding signal for Cas9 after a specific gRNA directs it towards this site in particular, but the exact sequence to target as a PAM depends on which Cas protein you use.

How do nickases work?

The cleavage activity of the S.Pyogenes Cas9 endonuclease is mediated by the coordinated functions of 2 catalytic domains, RuvC and HNH. The RuvC domain cleaves the non-targeting strand, while the HNH domain cleaves the targeting strand, which is complementary to the guide RNA. A Cas9 nickase variant can be generated by alanine substitution at key catalytic residues within these domains. The RuvC mutant D10A produces a nick on the targeting strand, while the HNH mutant H840A generates a nick on the non-targeting strand. DSBs are known to be essential for efficient genome editing. In eukaryotic cells, neither single nor dual gRNAs targeting the same strand led to efficient genome editing, because of a high-fidelity, base excision repair pathway. However, the combined use of 1 of the 2 Cas9 nickases with a pair of gRNAs targeting opposite DNA strands in close proximity generates a staggered DSB with overhangs. (Figure 11)

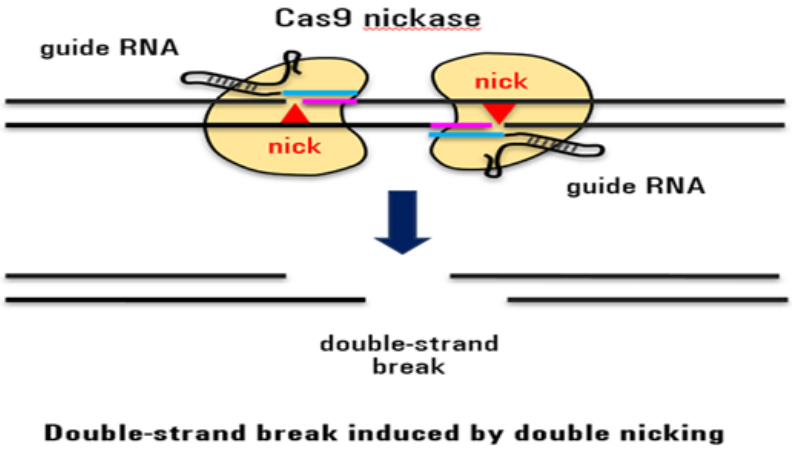


Figure 11: Double strand break induced by double nicking in close proximity.

Experimental procedure for disputing a gene with Cas9 nickase (D10A)

1. Design gRNA for knock out/knockdown (Benchling)
2. Cloning of the gRNA pairs in expression plasmid.
3. Transfection of gRNA and Cas9-nickase components in Huh7
4. Selection transfected cells (3-7day selection with antibiotic)
5. Screening and validation of KO clones. (Western blot)

Design guide RNA for knock out/knockdown of RNF20 and RNF40 genes

There are several parameters to consider when designing sgRNA for CRISPR experiments:

- GC content of the sgRNA sequence is important, as higher GC content will make it more stable - it should be 40-80%.
- The length of the sgRNA should be between 17-24 nucleotides, depending on the specific Cas nuclease you're using. Shorter sequences can minimize off-target effects, however, if the sequence is too short, the opposite effect can also occur.
- Mismatches between gRNA and target site can lead to off-target effects, depending on the number of mismatches and their position/s.
- It may be necessary to design multiple sgRNAs for each gene of interest, due to the fact that activity and specificity can be unpredictable.

We used the tool Benchling to design the following pairs of guides RNA, which will be inserted into the AIO-puro plasmid (plasmid #74630 addgene). The selection of the specific pairs of gRNA was made to target exons 3 and 4, as they are among the largest initial exons of the RNF genes and they also have high off target score. The off-target score tells you the inverse probability of Cas9 off-target binding. A higher score means the sequence has less chance to bind to sequences in the rest of the genome. The on-target score represents the cleavage efficiency of Cas9. Off-target scores are calculated based on Hsu et al. 2013.

Gene	Oligonucleotide
RNF20 G1 (pair 1) Exon 3	TGTCAACCGATACTGGAGTC
RNF20 G2 (pair 1) Exon 3	TCGGTTGACAATCAATAGTG
RNF20 G1 (pair 2) Exon4	CGTAAAGATGACCGAGAGAG
RNF20 G2 (pair 2) Exon4	TCCTGATTGCTATCAGAGTC
RNF20 G1 (pair 3) Exon3	TATTGATTGTCAACCGATAC
RNF20 G2 (pair 3) Exon3	CAATAGTGAGGCATCATCAG
RNF40 G1 Exon3	TCCTCATCGTCAATCGCTAC
RNF40 G2 Exon3	CCAACTTCTCAATTCGTTCT

Table1: Oligonucleotide sequences for CRISPR-Cas9 nickase (D10A)

RNF20 AIO-puro Constructology

We use the the All-in-One plasmid (AIO, Addgene # 74630) encoding dual U6 promoter-driven sgRNAs and Cas9-D10A nickase linked via 2A peptide with **puromycin** resistant marker (figure 2). The expression of the latter enables us to select for cells having received the plasmid when treating them with Puromycin

antibiotic and therefore enables us to maximize genome editing in the population of cells. After designing and receiving the oligonucleotides (synthesized by Macrogen Europe) (table 2), we proceeded to phosphorylation of the 5'end of the single stranded oligos.

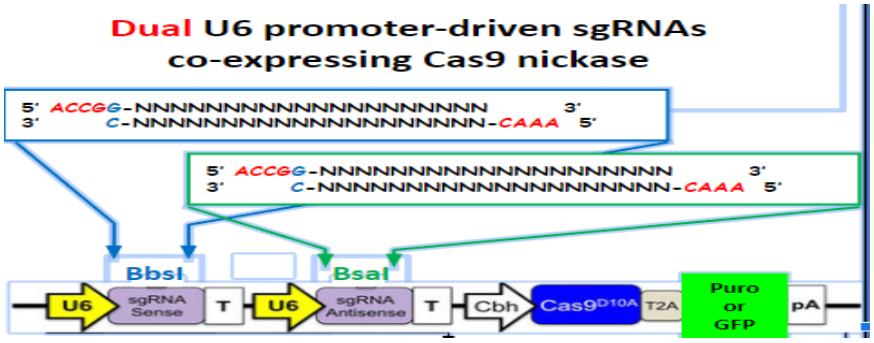


Figure 12: Co-expression systems for Cas9 nickase and gRNAs

	Coding sequence	Reverse complementary sequence
RNF20 pair 1 G1	ACCG-TGTCAACCGATACTGGAGTC	AAAC-GACTCCAGTATCGGTTGACA
RNF20 pair 1 G2	ACCG-TCGGTTGACAATCAATAGTG	AAAC-CACTATTGATTGTCAACCGA
RNF20 pair 2 G1	ACCG-CGTAAAGATGACCGAGAGAG	AAAC-CTCTCTCGGTCATCTTTACG
RNF20 pair 2 G2	ACCG-TCCTGATTGCTATCAGAGTC	AAAC-GACTCTGATAGCAATCAGGA
RNF20 pair 3 G1	ACCG-TATTGATTGTCAACCGATAC	AAAC-GTATCGGTTGACAATCAATA
RNF20 pair 3 G2	ACCG-CAATAGTGAGGCATCATCAG	AAAC-CTGATGATGCCTCACTATTG
RNF40 pair 1 G1	ACCG-TCCTCATCGTCAATCGCTA	AAAC-GTAGCGATTGACGATGAGGA
RNF40 pair 1 G2	ACCG-G-CCAACTTCTCAATTCGTTCT	AAAC-AGAACGAATTGAGAAGTTGG-C

Table 2: Oligonucleotides synthesized by Macrogen Europe.

So, we used T4 PolyNucleotide Kinase (PNK, NEB #M0201), a recombinant enzyme that catalyzes the transfer (or exchange) of P_i from the γ position of ATP to the 5´ -hydroxyl terminus of polynucleotides (double-and single-stranded DNA and RNA). We next hybridizing the complementary strands before cloning the dsDNA into AIO-puro plasmid, according to the following experimental procedure:

1. Set up AIO Puro digest reaction with BbsI, as described below.

AIO DIGEST REACTION (final Volume: 30µl) **With BbsI**	
VECTOR 2µg	8µl from 250ng/µl
10x Buffer BbsI (NEB 2)*	3µl
BbsI	2µl
H_2O	17µl

- Digest reactions 37° for ~2 hours and 50 min
- Run on 1% agarose gel
- Gel purification (total 40µl elution per reaction – 2 elutions of 20µl per column)
- Measure nanodrop

2. Dissolve primers in water for injection filtered. Prepare 100 µM dilutions.

3. Phosphorylate and anneal the oligos separately as described below in <u>PCR tubes</u>.

Oligo Phosphorylation Reaction (final Volume: 20µl)	
RNF20 pair 1 coding G1 for AIO vector (100µM)	2µl
RNF20 pair 1 reverse complementary G1 (100µM)	2µl
10x T4 Ligation Buffer	2µl
T4 PNK	1µl

H_2O	13µl

Oligo Phosphorylation Reaction (final Volume: 20µl)	
RNF20 pair 2 coding G1 for AIO vector (100µM)	2µl
RNF20 pair 2 reverse complementary G1 (100µM)	2µl
10x T4 Ligation Buffer	2µl
T4 PNK	1µl
H_2O	13µl

Oligo Phosphorylation Reaction (final Volume: 20µl)	
RNF20 pair 3 coding for AIO vector G1 (100µM)	2µl
RNF20 pair 3 reverse complementary G1 (100µM)	2µl
10x T4 Ligation Buffer	2µl
T4 PNK	1µl
H_2O	13µl

Oligo Phosphorylation Reaction (final Volume: 20µl)	
RNF40 pair 1 coding G1 for AIO vector (100µM)	2µl
RNF40 pair 1 reverse complementary G1 (100µM)	2µl
10x T4 Ligation Buffer	2µl
T4 PNK	1µl
H_2O	13µl

We follow exactly the same procedure for G2 coding and reverse complementary sequences as in the tables above. Each pair of sets targets a given region. This is necessary as CAS9 Nickase requires that DNA is nicked twice in a short distance to simulate a double strand break (DSB). The DNA is then repaired via NHEJ, a process that alters the underlying sequence and is likely to introduce mutations such as stop codons, which infer the desired KO effect.

Phosphorylation-annealing reaction conditions

Temperature	Time
37°C	30min
95°C	5min
90°C	1min
85°C	1min
80°C	1min
75°C	1min
70°C	1min
65°C	1min
60°C	1min
55°C	1min
50°C	1min
45°C	1min
40°C	1min
35°C	1min
30°C	1min
25°C	1min
4°C	For ever

- Transfer each 20µl reaction into epp tube
- Add to each tube 180µl H_2O
- Add 200µl phenol/chl/iaa (25:24:1)
- Vortex very well (20sec)
- Centrifuge 14000rpm/10min/RT
- Transfer upper phase in clean tube
- Add 200µl chl/iaa (24:1)
- Centrifuge 14000rpm/10min/4°C
- Transfer upper phase in clean tube
- Add 20µl 3M Na-acetate pH5.2 and 550µl EtOH
- Precipitation at -80°C for 1 hour or at -20°C overnight (preferable)
- Centrifuge 14000rpm / 4°C / 30min
- Wash with cold 80% EtOH
- Centrifuge 14000rpm / 4°C / 10min
- Dry pellet
- Resuspend in 22µl H_2O
- Measure 2µl in nanodrop (we need a certain concentration of oligos for ligation step)

4. **Set Up the ligation reaction as follows and incubate overnight at 16°C (this can be done in a PCR tube)**

Ligation Reaction F.V. = 20µl	
Gel purified AIO-Puro BbsI digested vector	50ng
Annealed oligos (dsDNA) for Guide A-(AIO) construct	170-190ng
10x Ligation Buffer (#B0202S, NEB)	2µl
T4 DNA Ligase (#M0202S, NEB)	1µl
H_2O	Up to 20µl

5. DH5a Transformation and plasmid amplification/purification

LB preparation: LB recipe 1L -> Bacto Tryptone: 10gr, 1%, Bacto yeast extract: 5gr, 0.5%, NaCl: 10gr, 1%, split into conical flasks and sterilize by autoclaving for 20'.
LB agar preparation: LB agar recipe 1L: To the above ingredients add 15gr agar after volume adjustments and sterilize by autoclaving. Addition of antibiotics should be done after the medium is cooled.

- Add 5ng from the ligation reaction into 200µl of competent DH5a bacteria cells.

- Heat shock: Put the tube in a water bath at 42 for exactly 45 sec.
- Let the cells on ice for 2 min.
- Add 1 ml LB and incubate at 37 for 1h
- Plate everything after centrifuge at 1000 rpm for 5 min at RT
- Alkaline lysis miniprep plasmid isolation:
    1. Set up 4-5 ml liquid LB cultures and incubate overnight (16-18 hours) at 37 with shaking
    2. Pour 1,5 ml of culture and spin culture at 13.000 rpm for 30 sec.
    3. Resuspend pellet in 100 µl GTE buffer (solution 1), mix/vortex
    4. Add 200 µl freshly prepared 0,2M NaOH/1% SDS (solution 2), invert tubes 5 times and incubate no more than 5 minutes
    5. Add 150 µl 5M potassium acetate (solution 3), invert tubes 5 times and incubate no more than 5 minutes
    6. Add 150 µl chloroform: IAA (24:1) and vortex
    7. Spin 13.000 rpm/5 min and take 400 µl from the up phase into a new tube

8. Add 1 ml ice cold EtOH 100% and mix well by inverting the tubes
9. Spin at 13.000 rpm/5 min and discard the liquid
10. Rinse pellet with 350 µl EtOH 70% and spin 13.000 rpm/2 min
11. Remove all alcohol and resuspend pellets in dH2O with RNase A (10mg/ml)

Solution 1 (GTE buffer): 25mM Tris, 10mM EDTA, 50mM glucose

Solution 2: 0,2M sodium hydroxide, + 1% SDS

Solution 3: 5M K acetate

6. Check construct with PCR

Plasmid DNA:2 ng
2 ng plasmid AIO-Puro (Without insert) for control

PCR REACTION (20µl)	**MASTER MIX FOR 12 Reactions**
2 ng plasmid DNA	
2µl 10x Taq Buffer (minotech)	24µl 10x taq Buffer
0,4µl Fw. Primer (25µM)	4,8µl Fw primer (25µM)
0,4µl Re. Primer (25 µM)	4,8µl Re primer (25µM)
0,4 µl dNTPs (10mM	4,8 µl dNTPs (10mM)
0,1-0,2µl(5U./µl) Taq Polymerase	1,2-2,4µl(5U./µl) Taq Polymerase
H20 (F.V. =20µl)	H20(F.V.=240µl)

94◦C for 2min
94◦C for 30sec
55◦C for 30sec
PCR CONDITIONS
72◦C for 13sec
Go to step 2 for 20 cycles
4◦C for ∞

	Forward primer	Reverse primer
Construct 1 (pair 1 RNF20 a)	ACCGGTGTCAACCGATACTGGAGTC	CTGCCCGACCTTTCCCTTTA
Construct 2 (pair 2 RNF20 a)	ACCGGCGTAAAGATGACCGAGAGAG	CTGCCCGACCTTTCCCTTTA
Construct 3 (pair 3 RNF20 a)	ACCGGTATTGATTGTCAACCGATAC	CTGCCCGACCTTTCCCTTTA
Construct 4 (pair 1 RNF40 a)	ACCGGTCCTCATCGTCAATCGCTAC	CTGCCCGACCTTTCCCTTTA

Table 3: Forward and reverse primers used for successful ligation of gRNA1 check with PCR.

The reverse primer CTGCCCGACCTTTCCCTTTA constitutes a sequence of AIO-puro plasmid, is located 225 bp downstream of BbsI cut site and designed to check the insertion of guide RNA 1.

7. Set up AIO Puro digest reaction with BsaI, as described below.

AIO-Guide A CONSTRUCTS DIGEST REACTION (final Volume: 10µl) With BsaI	
Guide A AIO-Vector	3000ng
10x Buffer BsaI (NEB 3.1)	1µl
BsaI	1µl(10units/µl)
H_2O	Up to 10µl

- Incubate at 37^{oC} for ~3hours

- Run on 1% agarose gel
- Gel purification (total 40µl elution per reaction – 2 elutions of 20µl per column). Gel purification was carried out by MACHEREY-NAGEL's clean up kits to quantitative high recovery DNA purification with this kit are: a**) Excise DNA fragment / Solubilize gel slice**, using a specific buffer, which often contains a pH indicator, solubilize the gel-encased DNA. Usually, the buffer and gel slice are heated until all of the gel is dissolved. The pH indicator is used to ensure that the buffer maintains the optimal pH for DNA binding. Acidic pH usually enhances DNA adsorption to the membrane. b) **Bind DNA to column**, virtually all DNA gel extraction kits rely on silica membrane spin columns. Silica membranes bind DNA molecules in the presence of high ionic-salt buffers that drive hydrogen bond formation between silica and DNA. DNA binds the silica membrane as the sample is passed through the column by centrifugation. c) **Wash silica membrane**, while the DNA remains bound to the silica membrane, contaminants such as nucleotides, proteins and other impurities are removed by alcohol-based washes. During several washes, salts are also removed, which sets the stage for elution of DNA. d) **Dry silica membrane**, e) **Elute DNA**, DNA is released from the silica membrane by eluting with a low-ionic solution, such as TE (Tris-EDTA) or water. Low salt solutions disrupt the hydrogen bonds that hold DNA on the membrane. Elution is most efficient under basic conditions, between the pH of 8 and 9. Moreover, it is recommended to heat up the elution buffer and let it sit on the membrane for up to five minutes to release more DNA molecules.

8. Set Up the ligation reaction as follows and incubate overnight at 16°C (this can be done in a PCR tube)

Ligation Reaction Final Volume = 20µl	
Gel purified AIO-Puro G1 BsaI digested vectors	~50ng
Annealed oligo for Guide B-(AIO) constructs	~170-190ng
10x Ligation Buffer	2µl
T4 DNA Ligase	1µl
H_2O	Up to 20µl

The optimal insert to vector DNA ratio is usually between 2:1 and 10:1. A larger amount of insert increases your chances of a successful cloning reaction. Higher concentrations of DNA reaction components will result in a higher rate of reaction. Ligation efficiency is also contingent on the integrity of the cohesive ends of the fragments being ligated.

9. DH5a Transformation and plasmid amplification/purification as described in step 4.

10. COLONY PCR

We relied on colony PCR for rapid screening of positive (inserted gRNA sequence in vector) clones from bacteria that have grown up on selective media following a transformation step, to verify that the desired genetic construct is present. To do so we designed and synthesized appropriate primers (show their sequence + make the diagram to explain what to expect from the PCR reaction for positive v negative clones) for positive and negative controls.

PCR CONDITIONS
94°C for 2min
94°C for 30sec
55°C for 30sec
72°C for 30sec
Go to step 2 for 20 cycles
4°C for ∞

	Forward primer	Reverse primer
Construct 1 (pair 1 RNF20 b)	ACCGGTGTCAACCGATACTGGAGTC	AAACCACTATTGATTGTCAACCGAC

Construct 2 (pair 2 RNF20 b)	ACCGGCGTAAAGATGACCGAGAG AG	AAACGACTCTGATAGCAATCAGGAC
Construct 3 (pair 3 RNF20 b)	ACCGGTATTGATTGTCAACCGATAC	AAACCTGATGATGCCTCACTATTGC
Construct 4 (pair 1 RNF40 b)	ACCGGTCCTCATCGTCAATCGCTAC	AAACAGAACGAATTGAGAAGTTGGC

Table 4: Forward and reverse primers used for gRNA1 and gRNA 2 for successful ligation check with PCR.

PCR REACTION (20µl)
plasmid DNA (unknown concentration)
2µl 10x Taq Buffer (minotech)
0,4µl Fw. primer (25µM)
0,4µl Re. primer (25 µM)
0,4 µl dNTPs (10mM
0,1-0,2µl(5U./µl) Taq Polymerase
H20 (F.V. =20µl)

Midipreps

The purification of all DNA plasmids (FLAG-RNF20-pcDNA3, HA-RNF40-PIRES-puro2, AIO-puro with the appropriate guide RNA for KO), was done at mini prep scale (Alkaline lysis miniprep plasmid isolation) with homemade buffers for diagnostic digestions (1,5ml bacteria culture).

While for the purification of larger amounts of pDNA stocks, we transferred 1ml of exponentially growing bacteria from mini-prep primary culture in 50 ml LB + antibiotic (in our case ampicillin 0,01 mg/ml). The culture was incubated for 16-18 hours at 37 C and the cells were pelleted at 3500 rpm. Supernatant was discarded and the pellet was processed according to the ZymoPURE II Plasmid Midiprep kit protocol. Utilizing a modified alkaline lysis in conjunction with a patented binding system, the ZymoPURE II Midiprep kit can process up to 50 ml of bacteria culture. The plasmid DNA is rapidly bound onto a column with either a vacuum or a centrifuge instead of a slow gravity flow column. The chaotropic salt binding buffer allows the highest DNA binding of any column method. Powerful wash buffers

remove all traces of protein and salt. DNA is eluted in a low-salt buffer to allow for pH stabilization of the DNA in storage. Additionally, there is no alcohol precipitation step required and the elution is performed using a microcentrifuge. FLAG-RNF20 pcDNA3 and HA-RNF40 PIRES-puro2 plasmids were kindly given by Dr J. Kim, South Korea. We confirmed them with diagnostic digestions with the appropriate enzymes and sequencing.

Human Cell culture

We used the human cell line HuH-7 established in 1982 from a well differentiated hepatocyte derived cellular carcinoma cell line that was originally taken from a liver tumor in a 57-year-old Japanese male. Huh-7 is an immortal cell line composed of epithelial-like, tumorigenic cells. The majority of Huh-7 cells show a chromosome number between 55 and 63 (mode 60) and are highly heterogeneous. These cells are adherent to the surface of flasks/plates and typically grow as 2D monolayers. Although containing many mutations, it is worthy to note the Huh7 cells have a point mutation in the p53 gene. The reason we chose this particular cell line is that it has 2 copies of chromosomes 9 and 16 where are the genetic locus of RNF20 and RNF40 respectively (Figure 13) and the cells grow nicely, and are suitable for microscopy analyses. This will help us to more easily achieve a homozygous situation via CRISPR/Cas9 genome editing, in contrast to other cell lines that have many more copies of these genetic loci.

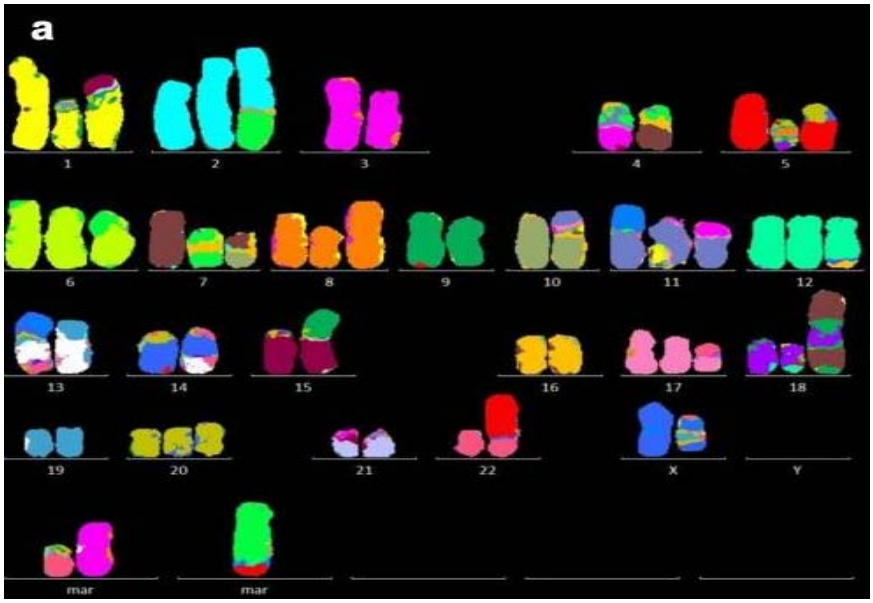


Figure 13: Example of M-FISH karyogram from α major clone in the HuH-7 cell line,

F Kasai, 2018.

We grow the cells incomplete nutrient medium D 10 [DMEM (Dulbecos Modified Eagles Medium (Thermo Scientific # 11965, enriched with 10 % FBS (fetal bovine serum #. Gibco 10437-028) and 50µg/ml Gentamicin, #. Applichem, 1405-41-0). The cells are kept at 37 o C with 5% CO2, they are adherent cells. Passages were performed after PBS wash, by applying 0.25% trypsin dissolved in PBS (stock: 2,5% trypsin-EDTA
10x, Gibco, 15400-054) for 5 mins at 37 C and by diluting the cells 3 to 5x with D10 according to our needs, or by counting the cells and diluting the desired number in D10 to seed them in a new plate/well.

Transient transfection protocol for adherent cells

JetPEI® transfection reagent is a linear polyethyleneimine (PEI) derivative, free of components of animal origin, providing highly effective and reproducible gene delivery to **adherent and suspension cells**. It is a cationic polymer composed of PEI that efficiently encapsulates and protects nucleic acids by forming less than 100 nm nanoparticles. Cationic polymers can be defined as macromolecules that bear positive charges. Plasmids then enter the nucleus either upon the mitotic disassembly of the nuclear envelope or through nuclear pore complexes in the absence of cell division, using a different set of proteins (Figure 14).

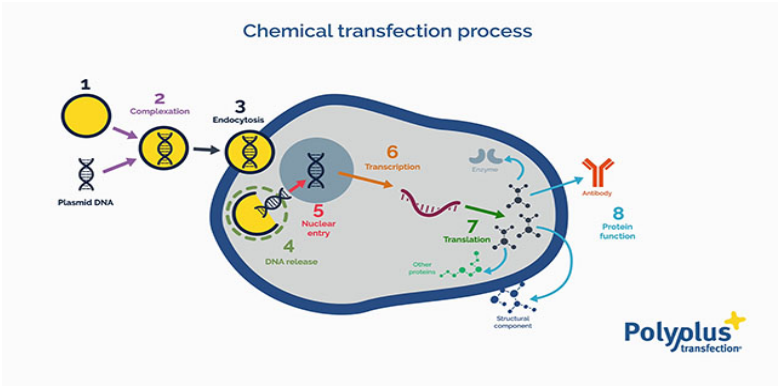


Figure 14: Chemical transfection process via JetPEI® transfection reagent.

Transfection procedure in a 12-well plate for jetpei DNA transfection kit (Polyplus, ref. number 101000053):

1. For each transfection reaction, dilute 2 µg of DNA in 150 mM NaCl to a final volume of 50 µL. Vortex gently and spin down briefly.
2. Vortex jetPEI® reagent for 5 sec and spin down before use.
3. Dilute 4 µL of jetPEI® reagent in 150 mM NaCl to a final volume of 50 µL. Vortex gently and spin down briefly.
4. Add the 50 µL jetPEI® solution into the 50 µL DNA solution all at once. Please note that mixing the solutions in the reverse order may reduce transfection efficiency.
5. Vortex the solution immediately and spin down briefly.
6. Incubate for 30 minutes at room temperature.
7. For optimal transfection conditions with jetPEI® we recommend using cells 50-70% confluent on the day of transfection.
8. Per well, add the 100 µL jetPEI® /DNA mix drop-wise to the cells in 1 mL of serum-containing medium and homogenize by gently swirling the plate.
9. Return the plates to the cell culture incubator.
10. Repeat step 8 the next day
11. Perform reporter gene assay 24h after second transfection.

Cell culture antibiotic selection

Antibiotics are commonly used in cell culture to prevent contamination, maintain aseptic conditions, or select for cells containing genetic modifications. The use of antibiotics can minimize loss of cell lines and cell cultures, saving time and conserving precious resources. Proper antibiotic selection depends on both cell type and which contaminants you are trying to prevent. Selective antibiotics for

generating stable cell lines or other recombinant cultures should be chosen based on the antibiotic resistance gene or selectable marker.
In our case, we used puromycin for selection of the cells that have received AIO-puro plasmid (contains guide RNA for RNF20 CRISPR KO, Cas9 nickase and resistance gene for puromycin) and G418 for selection of cells that received FLAG-RNF20 pcDNA3 plasmid. Puromycin is an antibiotic protein synthesis inhibitor which causes premature chain termination during translation and G418 blocks polypeptide synthesis by inhibiting the elongation step. Importantly, we tested Puromycin survival by testing different concentrations to determine optimal antibiotic concentration for selection. We established what concentration of puromycin kills 100% of untransfected cells by adding different concentrations of puromycin (0.5, 1, 1,5, 2, 3, 4, 5 μg/ml) for 5 days. Regarding G418, we relied on previous data so we proceeded directly with the selection (1 mg/ml) in transfected cells with FLAG-RNF20 pcDNA3 plasmid for 5 days.

Protein extraction from Cells

- Collect cells from each 12-well plate in a 1.5ml tube (with scraper or trypsin, both cases were tested)
- Resuspend the cells in the tubes with 80 μl RIPA solution (which allows cell lysis and whole cell protein extraction-WCE)
- Vortex for 1 minute and incubation for 5 minutes on ice
- Pipetting or vortex again for homogenous lysis of all the cells, followed by another 15 minutes incubation on ice.
- Centrifuge for 1 minute at 13000 rpm
- Transfer the supernatant to a new tube
- Add to it 80 μl of 2X Laemmli buffer (loading solution which contains β-mercaptoethanol).
- Mercaptoethanol reduces disulfide bonds, which develop between two -SH groups of cysteines of the same or two different polypeptide chains.
- Boil the sample for 10 minutes at 100 o C
- Centrifuge at 13000 rpm for 1 minute
- Transfer the supernatant to a new tube. The supernatant must be loaded immediately on ice and is ready to be used for western immunoblotting experiments.

Laemmli buffer	**RIPA buffer**
250 mM pH: 6.8 Tris-HCl	50 MM Tris-HCL, pH 8
30% glycerol	150 mM NaCl
10% SDS	5 mM EDTA pH 8
0.02% blue bromophenol	1% NP-40
5% β-mercaptoethanol	0,5% NaDOC
	1 mM PMSF

	1X Proteinase inhibitors cocktail
	ddH_2O

Western Blot

Denaturing Polyacrylamide gel electrophoresis is the most popular method of quantification and separation of proteins. The polyacrylamide gel is a three-dimensional mesh of long aliphatic polyacrylamide chains joined together by N-N methylene-bis-acrylamide (MBA) molecules. Polyacrylamide gels (PAGE) are the most suitable for electrophoresis because they are composed of chemical neutral compounds and are easily formed. Also, the size of the pores can be adjusted by choosing different concentrations of acrylamide and MBA. Smaller proteins move more easily through the pores of the gel, while the larger ones are delayed. The polymerization of the acrylamide and MBA, takes place at room temperature (RT) and requires two polymerizing agents: APS and TEMED as the latter catalyzes the formation of free radicals from APS. Lower polyacrylamide concentration (4-8 %) generates large pores and therefore proteins of high molecular weight are better resolved. On the contrary higher concentration (8-15%) enables them to resolve smaller proteins. The most common electrophoresis method is the one performed in destructive conditions where proteins are separated only according to their molecular weight in the presence of **Sodium dodecyl sulfate** (**SDS**) an anionic detergent (SDS-PAGE). SDS denatures proteins and protein-SDS complexes have a negative charge. Under these conditions the proteins are separated based on their molecular mass resulting in the smallest proteins to move faster towards the positive electrode and the larger more slowly. The experimental process of western blotting begins with the analysis of the protein samples on a denaturing gel polyacrylamide-SDS.

We first cast a Separating gel:

Separating gel 10 % (5mL)
1,67 mL Acrylamide 30 % $^{w}/_{v}$
2,05 mL ddH_2O
1,23 mL 1,5 M Tris-HCl pH 8,8
25 µL 20 % SDS
20 µL 10 % APS
10 µL TEMED

Once it sets, we pour on the top the stacking gel, which has a large pore size to stack the proteins in a narrow area:

Stacking gel 4 % (2,5 mL)
425 µL Acrylamide 30 % w/v (Sigma, A3574)
1,428 mL ddH2O
625 µL 0,5 M Tris-HCl pH 6,8
12,5 µL 20 % SDS
10 µL 10 % APS
5 µL TEMED (AppliChem)

We load the samples and run them in an electrophoresis device with EL buffer, until the proteins are well separated. When the protein molecules reach the separating gel, the migration of those molecules is slowed down because the separating gel is a high concentration gel with a small pore size that can act as a considerable barrier for the movement of the protein molecules. This slow down allows the other proteins are migrating slowly to catch up, resulting in a narrow, concentrated band in between the two gels. After electrophoresis, the separated proteins are transferred onto a nitrocellulose membrane, presence of transfer buffer at 200 mA for 90 minutes. After transfer, the membrane must be blocked to prevent non-specific binding of the antibody to the membrane surface, with 5% milk/PBST (PBS, 0,1 % Tween 20).

10X EL (1L)	**Transfer Buffer**
30,3 g Tris base	200 ml methanol
144 g glycine	100 ml EL buffer
10 g SDS	700 ml ddH_2O
ddH_2O	

Western blot quantification by Image J

1. Save original western blot image in JPEG, PNG or TIF on your computer.
2. Open the Image J in the Application folder on your computer. If you do not have this software, you may download from NIH website at: https://imagej.nih.gov/ij/download.html.
Make sure you download the one that fits your computer.
3. Open the image from ImageJ.

4. Select the rectangle icon and define your quantification area for the biggest band.
5. Use this to quantify all bands and backgrounds nearby for each band using “Command M”. The integrated density (IntDen) will show with the Area.
6. Copy and paste the numbers to Excel sheet and label the lanes with ID and treatment. Subtract the background value for each band to get Corrected IntDen.
7. Copy paste the image as a record.
8. Do step 1-6 for the loading control (GAPDH).
9. Divide the Corrected IntDen for your protein of interest by the corrected IntDen for the loading control to obtain Normalized IntDen.
10. Calculate the mean value of your control group in your gel image.
11. Divide all the values by this number to obtain fold change.

RESULTS

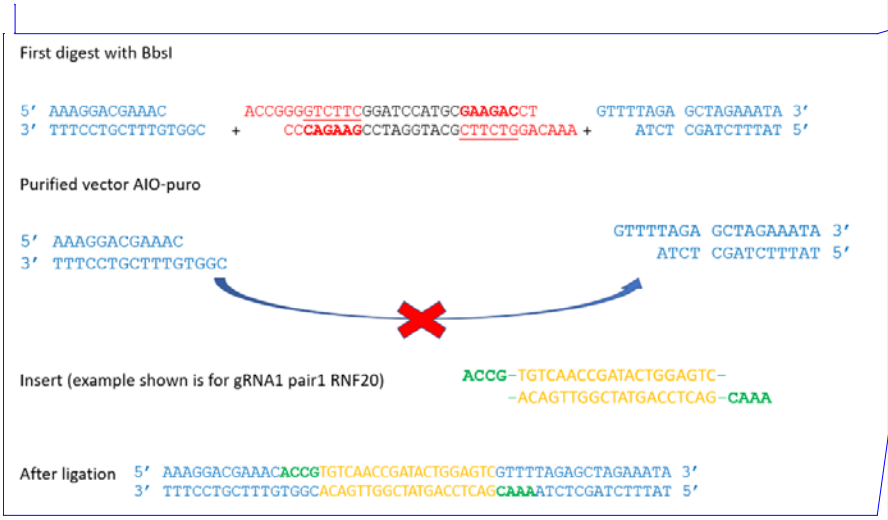
Initially, we designed the guide RNAs and found that pairs of gRNAs target exon 3 and 4 for RNF20 and RNF40 with better off target effect, they are what we described in table 1. In the table 5 mention again the sequences of gRNAs including on target and off target scores.

Guide sequence	On Target score	Off target score (pair)	Off target score (single)
TGTCAACCGATACTGGAGTC TCGGTTGACAATCAATAGTG RNF20 pair 1	53.4 69.7	18.4	45.8 40.2
CGTAAAGATGACCGAGAGAG TCCTGATTGCTATCAGAGTC RNF20 pair 2	53.3 68.4	18.8	47.2 42
TATTGATTGTCAACCGATAC CAATAGTGAGGCATCATCAG RNF20 pair 3	49.5 69.7	19.7	49.0 40.2
TCCTCATCGTCAATCGCTAC CCAACTTCTCAATTCGTTCT RNF40 pair 1	45.7 66.4	18.1	48.7 37.2

Table 5: Sequences of gRNAs we designed, including on target and off target scores.

We choose pairs that had the highest score in relation to other suggested pairs of gRNAs. Paired scores should be compared relative to other paired scores, while the single guide scores should be compared to other single guide scores. It is a bit unintuitive that the score is lower for the paired compared to single, considering that the nickase pair will have a lower off-target cut probability.

Subsequently, to insert the sequences of gRNAs into AIO-puro plasmid, we used two restriction enzymes type IIS, BbsI and BsaI. Type IIS restriction enzymes comprise a specific group of enzymes which recognize asymmetric DNA sequences and cleave at a defined distance outside of their recognition sequence, usually within 1 to 20 nucleotides. In this case, BbsI and BsaI recognize and cut twice the AIO-puro plasmid in close proximity, thus creating ends with different overhangs and providing us the direction we will introduce the sequence by designing the oligonucleotides with the appropriate overhangs. Plasmid self-ligation is improbable without an insertion. (Figure 15)



Comment [ml2]: The arrow is confusing draw a simple line that starts fro the 5′ shown on the outside.
Now it looks you ligate the two bits together…

Comment [ml3]: After ligation don't right the dashes « - «

Figure 15: Schematic representation of digestion with BbsI restriction enzyme and ligation with DsDNA sequence for gRNA 1. BbsI restriction enzyme recognizes GAAGAC(2/6)^ sites. By the same reasoning, digestion with BsaI and insertion of dsDNA sequence for gRNA 2. BsaI restriction enzyme recognizes GGTCTC(1/5)^ sites.

We ligated gRNA encoding dsDNA into AIO-puro plasmid, which expresses Cas9-nickase. The cloning process was carried out in 2 steps. In the first cloning step, the insertion of the first gRNA , after digesting the plasmid with an BbsI enzyme checked with PCR. Using as a forward primer the sequence of guide RNA1 for each pair of guides and as a reverse primer a sequence of AIO puro plasmid we expect to see a product at 213 bp after successful ligation (Figure 16).

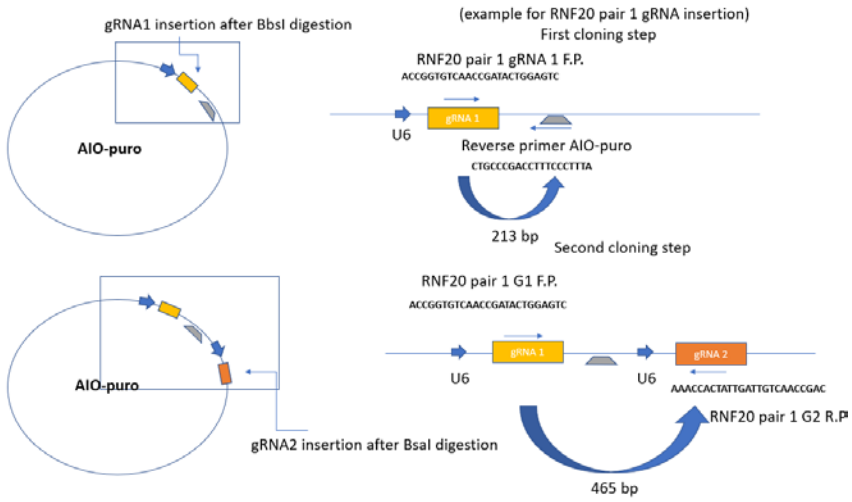
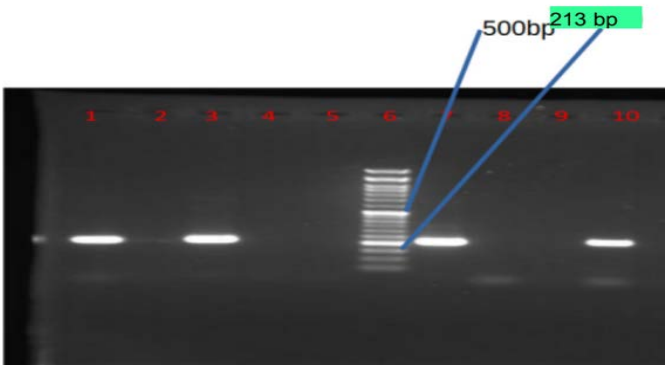


Figure 16: Schematic representation for plasmid screening after first and second cloning step with PCR.

The percentage of positive clones after minipreps was 5/5 for all constructs (1a,2a,3a,4a) (data not shown). Construct a contains only gRNA1 from each pair of gRNAs. In figure 17, PCR was repeated once for each construct with the appropriate primers.

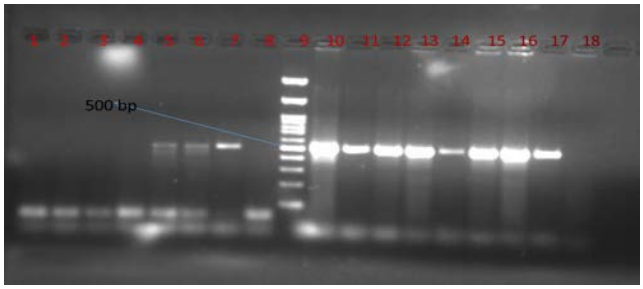


Lane	DNA	primers
1	Construct 1a	pair 1 RNF20 a (table 3)
2	Construct 1a	pair 1 RNF20 b (table 4)-negative control
3	Construct 2a	pair 2 RNF20 a (table 3)
4	Construct 2a	pair 1 RNF20 b (table 4)-negative control
5	H20	pair 1 RNF20 b (table 4)
6	Ladder 1kb	
7	Construct 3a	pair 3 RNF20 a (table3)
8	Construct 3a	pair 1 RNF20 b (table 4)-negative control
9	Construct 4a	pair 1 RNF20 b (table 4)-negative control
10	Construct 4a	pair 1 RNF40 a (table 3)

Figure 17: Agarose gel after PCR for gRNA 1 insertion check. Positive clones contain gRNA1 for RNF20 KO pair 1, pair 2, pair 3 and RNF40 pair 1 respectively. Expected product 213bp. Constructs a includes only the sequence of the first gRNA after digestion with BbsI.

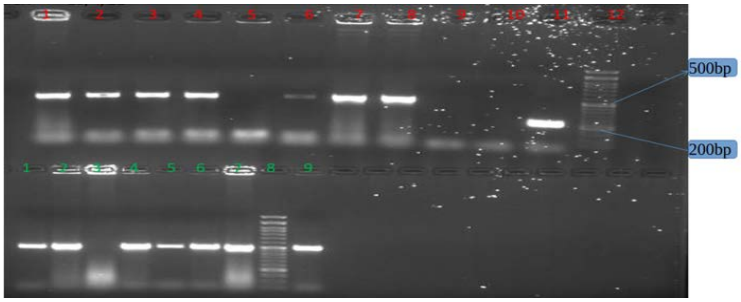
In the second cloning step involving digestion of the selected positive clones with BsaI, we purified the vector and ligated it with dsDNA sequences containing the guide RNA 2 (see methods and Figure 18). We then proceeded to check if insertion of guides RNA 2 was successful by colony PCR. Using for primers the 2 sequences of each guide RNA pair (table 4), we expected to see a PCR product at 465bp, in the appropriate temporal conditions for each pair of primers (Figure 16). Briefly, we made colony PCR of 8 different colonies for each construct, which contains different pairs of gRNAs. For RNF20 pair 1 KO we had a 3/8 success rate, 7/ 8 for RNF20 KO pair 3 and RNF40 pair 1 KO and 8/8 for RNF20 pair 2 KO (Figure 18,19). Since we carried out PCR using as forward primer and reverse primer sequences of gRNAs are ligated into AIO-puro (table 4), the absence of one of the two gRNAs is enough to not get a band after PCR reaction.

Comment [ml4]: I added this to clarify what you do between the two PCRs,,,



Lane	DNA	primers
1-8	Construct 1 b	pair 1 RNF20 b (table 4)
9	Ladder 1kb	
10-17	Construct 2 b	pair 2 RNF20 b (table 4)
18	H20	pair 1 RNF20 b (table 4)

Figure 18: Agarose gel after colony PCR. Expected product 465bp. Constructs b includes both sequences of gRNAs after digestion with BsaI.



Lane	DNA	primers
1-8	Construct 3 b	pair 3 RNF20 b (table 4)
9	H20	pair 3 RNF20 b (table 4)
10	Construct 3 b	pair 1 RNF20 b (table 4)- negative control
11	Construct 4 b	pair 1 RNF20 b (table 4)-negative control
12	Ladder 1kb	
1-7 +9	Construct 4 b	pair 1 RNF40 b (table 4)
8	Ladder 1kb	

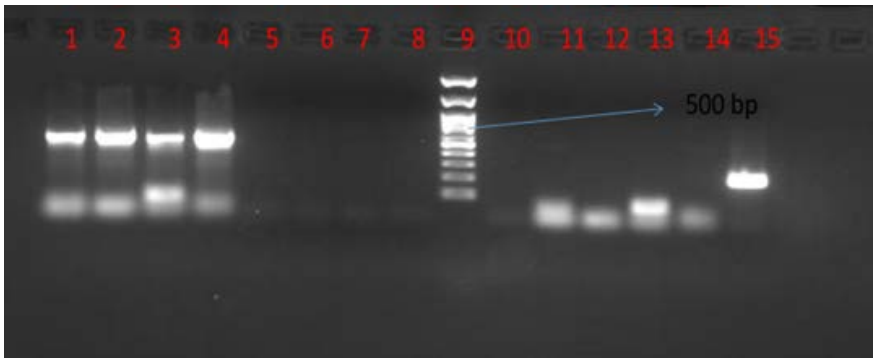
Figure 19: Agarose gel after colony PCR. Expected product 465bp. Constructs b includes both sequences of gRNAs after digestion with BsaI.

Then, knowing which samples from colony PCR contain the desired sequences, we repeated the same procedure for one positive control from each construct. The PCR was repeated this time using negative controls. Different incompatible primers for each construct were used to confirm the specificity of the product. Finally, we made 4 different AIO-puro constructs with different pairs of guide RNAs, 3 constructs for RNF20 and 1 for RNF40, after confirmation with PCR (Figure 20).

Comment [ml5]: This is really confusing, what primers were used… if in methods refer to their names.

Comment [ml6]: You mean new clones? No.. you want to say we tested our positive clones with primers that should not give product: which ones…. Could be worth having a figure with schematic too here

Comment [v7]:



Lane	DNA	primers
1	Construct 1b	pair 1 RNF20 b (table 4)
2	Construct 2b	pair 2 RNF20 b (table 4)
3	Construct 3b	pair 3 RNF20 b (table 4)
4	Construct 4b	pair 1 RNF40 b (table 4)
5	Construct 1a	pair 1 RNF20 b (table 4)-negative control
6	Construct 2a	pair 2 RNF20 b (table 4)-negative control
7	Construct 3a	pair 3 RNF20 b (table 4)-negative control
8	Construct 4a	pair 1 RNF40 b (table 4)-negative control
9	Ladder	
10	H20	pair 1 RNF20 b (table 4)-negative control
11	Construct 1b	pair 2 RNF20 b (table 4)-negative control
12	Construct 2b	pair 1 RNF20 b (table 4)-negative control
13	Construct 3b	pair 1 RNF40 b (table 4)-negative control
14	Construct 4b	pair 3 RNF20 b (table 4)-negative control
15	Construct 1b	pair 1 RNF20 a (table 3)-positive control

Figure 20: Repeat PCR with positive clones from colony PCR. Constructs a includes only the sequence of the first gRNA after digestion with BbsI. Constructs b includes both sequences of gRNAs after digestion with BsaI. Expected product 465 bp for positive clones and 213 for positive control.

For the implementation of transfection, we had available and checked the efficiency of two different reagents, Escort IV (Sigma Aldrich, Product No. L3287) and Jetpei DNA transfection kit (Polyplus, ref. number 101000053). We first checked transfection efficiency by introducing a plasmid expressing GFP (px461, addgene # 48140). efficiency of transfection was quantified based on the number of cells with detectable green fluorescence signal under microscopy two days post transfection. Although both reagents did not cause major toxicity to cells (no cell death the next days was observed), we could see in figure 22 how the Jetpei DNA transfection kit was more likely to give successful gene transfer than Escort IV (Figure 21). The experiment was repeated and the efficiency between those reagents did not change.

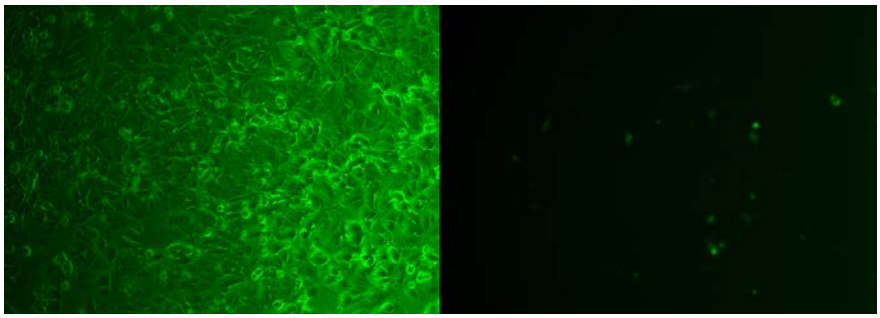


Figure 21: Huh7 cells transfected with px461-GFP plasmid via Escort IV.

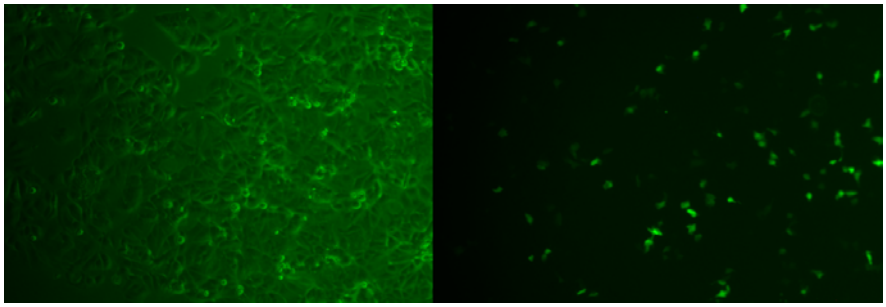


Figure 22: Huh7 cells transfected with px461-GFP plasmid via Jetpei DNA transfection kit.

Provided that, we ended up to jetpei as the most appropriate reagent for transfection from what we have at our disposal, we proceeded to import the necessary components to induce gene silencing of RNF20 gene via CRISPR/Cas9 nickase technology. We decided to do double transfection to get more KO by maximizing the number of cells receiving a plasmid. Two days after the double transfection with 2 different AIO-puro constructs (for exon 3 and exon 4), we collected the cells and proceeded to protein extraction and quantification. Equal amounts of protein samples were loaded on a 12% SDS-PAGE gel and analyzed by

Comment [ml8]: Something missing? How from cloning you go to transfection????

Western Blot to check the success of transfection. anti-RNF20 and anti-GAPDH antibodies were used to reveal the amounts of RNF20 (normalized to GAPDH) detected in control (cells transfected with AIO-puro plasmid w/o the appropriate guide RNAs) and KD cells (cells transfected with CRISPR/Cas9 components).

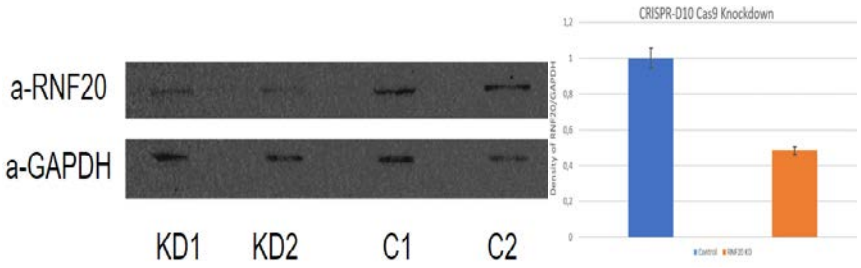


Figure 23: Western blot experiment for knockdown of RNF20 gene via CRISPR/CAS9 nickase technology. Quantification of Western blot bands by densitometry, graph derived from the average of the signal obtained for the two lanes of KD and WT above and another two replicates after repeating the experiment.

The figure 23 shows a clear reduction of RNF20 protein levels compared with control cells two days post transfection (2,057 folds after image J quantification). These results thus confirm that our gRNA plasmids are functional for KO of RNF20. We then initiated a procedure for selection of clones that have successfully received the AIO-puro plasmid with puromycin to enrich for cells with no RNF20/40 expression. After puromycin treatment on control cells, we found that a concentration of 1.5 µg/ml is sufficient to kill all cells in 5 days (w/o resistance gene to puromycin). Applying this regimen of antibiotics for 5 days, we enriched our cultures with cells that survive and have most likely received the plasmids with the CRISPR/Csa9 components. However, we also noticed that they could not grow and after a few days cells were dying. This indicates that the KO of RNF20 may be lethal for the cells. Our data are in agreement with published experiments showing that, knockdown of Bre1a or Bre1b with lentivirus-based shRNA constructs targeting either of the two subunits of the complex, significantly impaired cell growth in different mouse cell types including RIF-1, C3H 10T1/2 and MES cells[54] (Chernikova et al. 2012). We also took advantage of a new functional tool, the Dependency Map (DepMap), summarizing data from CRISPR experiments with random guide RNAs in a variety of cancer cell lines (data from 1812 human cancer cell lines). The DepMap project utilizes CRISPR gene knockout as the primary method to map genomic dependencies in cancer cell lines.

In our case, there is available data for RNF20 CRISPR KO in Huh7 cells. As we can see from gene effect (Chronos) score, the score is very close to -1 which means that the

absence of RNF20 is important for the growth of this cancer cell line, a result compatible with our observation that cells do not survive after constitutive silencing of RNF20 gene (Figure 24).

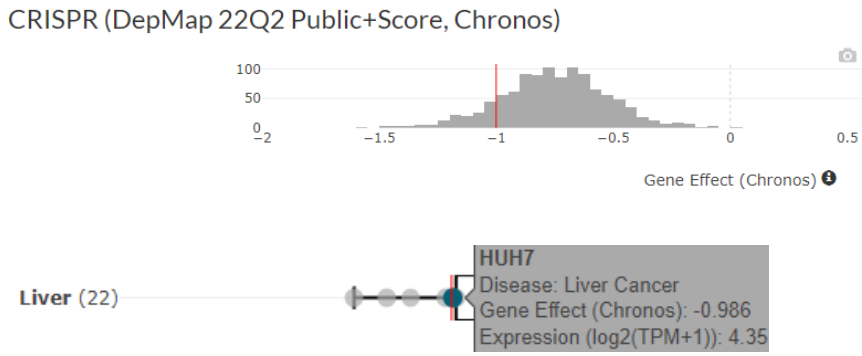


Figure 24: Effect of the depletion of RNF20 on liver cell lines, a lower Chronos score indicates a higher likelihood that the gene is essential. A value of -1 (red line) represents the median score of all pan-essential genes.

With the initial goal of mapping the binding sites of RNF20 on genome and nascent RNA, we thought of proceeding with the generation of a cell line expressing RNF20 and RNF40 with a tag, as there is no well-characterized ChIP antibody on the market. Dr J. Kim kindly sent us two plasmids containing the coding sequence of RNF20 and RNF40 tagged with FLAG and HA sequences, respectively. We received the plasmids and confirmed their content by diagnostic digestions with restriction enzymes and sequencing.

What we observed after sequencing is that the last of the 8 amino acids of the FLAG tag is different from the one recognized by the commercial a-FLAG antibodies on the market. Afterward, we performed transfection with FLAG-RNF20 pcDNA3 plasmid into Huh7 cells. Two days post transfection collected the cells and proceeded to protein extraction and quantification. Equal amounts of protein samples were loaded on a 12% SDS-PAGE gel and analyzed by Western Blot to check the success of transfection. anti-RNF20 and anti-GAPDH antibodies were used to reveal the amounts of RNF20 (normalized to GAPDH) detected in control and OE cells.

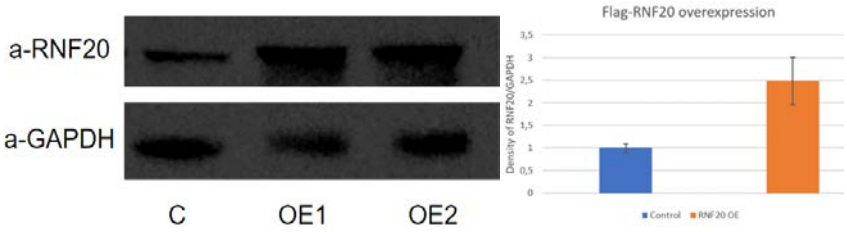


Figure 25: Western blot experiment for overexpression of RNF20 gene via transfection with FLAG-RNF20 pcDNA3 construct. Quantification of Western blot bands by densitometry, graph derived from the average of 4 different transfection and western blot experiments.

As we also see in figure 25, there is a significant increase in protein levels of RNF20 in cells (2,412 folds after image J quantification) that have received the plasmid containing FLAG-RNF20. We quantified the bands with ImageJ and indicate the ratio of the signal (OE/control). Seven days after selection with G418 0,5mg/ml, we observed here as well difficulty to growth of Huh7 cells in relation to cells that have received GFP plasmid. Recent data, which are in agreement with this result, shows that overexpression of RNF20 reduces SREBP1 expression. Finally, several W.B. experiments were conducted with a-FLAG antibody in protein extracts, where RNF20 was overexpressed but could not reveal bands colocalizing with RNF20, by testing 2 different a-FLAG antibodies (Cell signaling, DYKDDDDK Tag (9A3) Mouse mAb, #8146 and DYKDDDDK Tag (D6W5B) Rabbit mAb #14793).

Discussion

In the framework of this master thesis, we prepared basic molecular tools for our future experiments which will aim at assessing the loss of function effects of RNF20/40 in chromatin organization and gene expression, as well as mapping RNF20/40 interaction sites in the genome (by ChIP-seq) and to nascent RNA (by RIP-seq) to confirm interactions detected in data from EU-MS experiments[53]. While trying to procedure stable cell lines KO/KD and overexpression for RNF20 we were also interested in the role of RNF20 in the development and evolution of cancer. Preliminary results show that in cases of liver cancer the levels of RNF20 increase (Figure 26). Our results after KO experiments in Huh7 cells in combination with data from Depmap shows that RNF20 must be a necessary gene for this specific cell line but also for others.

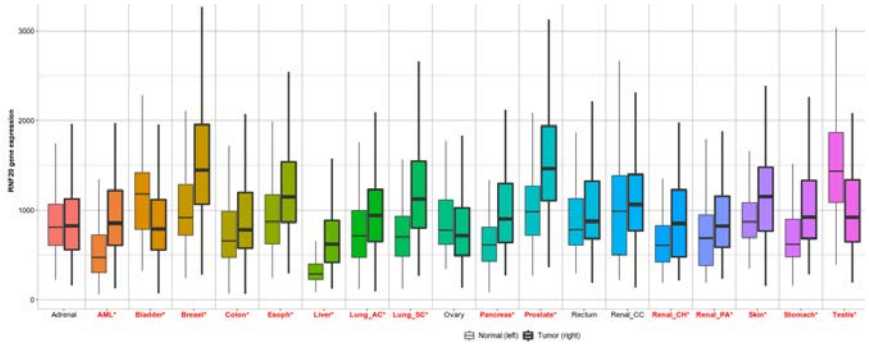


Figure 26: Differential gene expression analysis of RNF20 in tumor and normal cells from TNMplot.

Difficulties in establishing stable constitutive RNF20 KO cell lines led us to consider alternative technologies for acute protein depletion such as the dTAG system. We have already designed primers for CRISPR-mediated specific locus knock in (for RNF20 and RNF40) with dTag system. This approach constitutes a robust method to probe the role of proteins in living cells, based on the rapidity of induced protein depletion and its reversibility (Figure 27). Briefly, CRISPR-mediated genome engineering is used to tag the endogenous protein of interest at the carboxy-terminus with a double HA-epitope tag and F36V mutant form of FKBP12 [FKBP(F36V)-2HA]. This process involves electroporation of cells to deliver gene-specific targeting vectors, as well as ribonucleoprotein complexes of Cas9 and a target gene-specific sgRNA. After enrichment and validation, factor degradation is triggered by addition of a bifunctional small molecule degrader—such as dTAG-13 (Nabet et al., 2018) or dTAG-47 (Guarnaccia et al., 2021), that couples the 2HA-FKBP(F36V)-tagged protein to the cereblon (CRBN) ubiquitin protein ligase, triggering its rapid ubiquitin-mediated proteolysis[55,56].

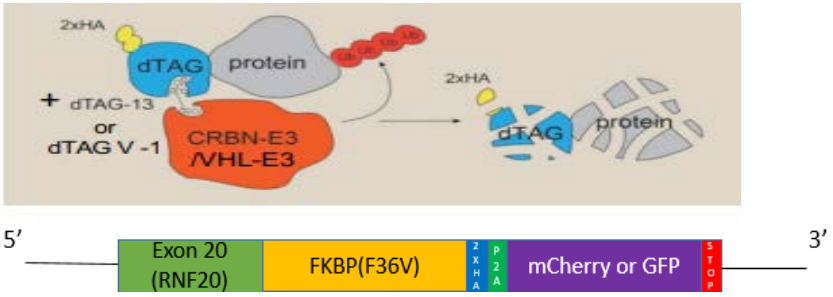


Figure 27: Strategy to generate modified cell lines by dTAG KI acute degradation of RNF20 protein.

The same strategy will also allow us to improve RNF20 overexpression, maybe with a different tag (2xHA tag in C term or N term of inserted sequence depending on the plasmid we will use, the protein will therefore be expressed with a tag constitutively). After transfection with FLAG-RNF20 pcDNA3 plasmid, we observed an increase in the levels of RNF20, but we were unable to capture the protein with a-FLAG antibody. This is a big problem for us because we need such a tagged-RNF20/40 cell line to be able to pull down RNF20 for ChIP or RIP experiments. Although in the original paper describing this plasmid, they were able to perform anti-FLAG IP with the generated cell lines, our sequencing results of the plasmid indicate that there is a mutation in 1 amino-acid (say which one (indicate sequence observed versus theoretical). The difference in the last amino acid, after sequencing of FLAG-RNF20 pcDNA plasmid, which recognizes the a-FLAG antibody might therefore be responsible for our inability to detect OE RNF20 with anti-FLAG antibodies by W.B. We also argue that a plasmid that expresses the RNF20 protein tagged with 3X FLAG would facilitate the detect of ion FLAG sequence compared with the 1X FLAG plasmid we have right now. Despite these problems with the FLAG-tag, we noticed that overexpression of RNF20 such as KO/KD in Huh 7 cells may have some effect on the progress and growth of these cancer cells. From what it seems, the disruption of the RNF20 levels down and up could cause a problem in the proliferation of cancer cells (data not shown). Furthermore, it was found that reducing the expression of SREBP1 leads to a decrease in the expression of genes involved in lipid metabolism in liver and kidney cancer cells and suppresses the proliferation of cancer cells[57] (Figure 28).

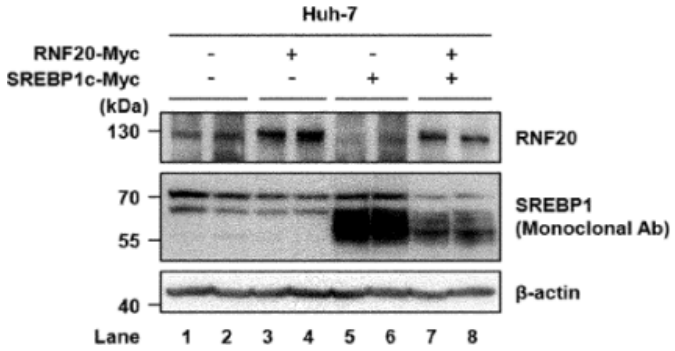


Figure 28: RNF20 overexpression reduces SREBP1 in Huh7 cells.

Certainly, the experiments regarding the overexpression of RNF20 should be repeated. We should introduce into the cells a new construct that will express the protein with a tag, which we will be able to detect after protein extraction and Western blot with a-Tag antibody. This will allow us to re-observe the effects of RNF20 overexpression in HuH7 cancer cells but also to proceed with future experiments related to the mapping of the protein.

A fine balance in activity between RNF20/40 (Ubiquitin ligase) and the de-ubiquitination (DUB) module of SAGA is known to be primordial for smooth elongation of Pol II through nucleosomes. Based on genome wide (re)distribution maps of H2Bub obtained with or without inhibition of transcription, in normal cells or cells that lack DUB activity, we suspect that the potential recruitment of the writer enzyme to given genomic loci depends on nascent RNA concentration around the transcribing polymerase (Fanourgakis et al, in revision). Our future experiments about this project will include:

1.Producing RNF-Tag and RNF inducible-degradation KI in human hepatocytes (HuH7 cells) via CRISPR technology.

2. Generating NGS chromatin and nascent RNA profiles in WT and mutant cells perturbed conditions (e.g., transcription and splicing inhibition) by performing ChIP-seq and RIP-seq of RNF-tag and nascent RNA-seq.

3.Verify nRNA particles behavior by intron specific RNA-FISH and CRISPR-Cas9 based locus specific labeling/pull-down of nascent RNA or H2Bub.

About the 3rd bullet, the volume of nRNA tethered to a traveling Polymerase 2 could influence the level of RNF20/40 recruitment at genomic loci and therefore we want to check if nascent RNA volume is locally changing at intron-exon junctions. This hypothesis implies that nRNA particles grow from Pol2 and stay condensed above elongating Pol2 according to Leidescher et al 2022.

Together these results will clarify the mechanism ruling H2Bub chromatin modification during transcription elongation and determine how nRNA feedback is

probably instrumental in this process. Overall, our preliminary efforts described in this thesis are important to generate the necessary tools to understand the effects of overexpression or absence of RNF20/40 on the proliferation of cancer cells. Deciphering the exact molecular mechanism that regulates chromatin modification by RNF20/40 and explaining how this process participates in gene expression might provide opportunities for new therapeutic approaches in certain types of cancer or other transcription-related diseases.

BIBLIOGRAPHY

1. McGinty, R. K. & Tan, S. Nucleosome Structure and Function. *Chem. Rev.* **115**, 2255–2273 (2015).
2. Hd, O. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, eaag0025–eaag0025 (2017).
3. Talbert, P. B. & Henikoff, S. Histone variants at a glance. *J. Cell Sci.* **134**, jcs244749 (2021).
4. Fyodorov, D. V., Zhou, B.-R., Skoultchi, A. I. & Bai, Y. Emerging roles of linker histones in regulating chromatin structure and function. *Nat. Rev. Mol. Cell Biol.* **19**, 192–206 (2018).
5. Potaman, V. N. & Sinden, R. R. *DNA: Alternative Conformations and Biology*. *Madame Curie Bioscience Database [Internet]* (Landes Bioscience, 2013).
6. Mariño-Ramírez, L., Kann, M. G., Shoemaker, B. A. & Landsman, D. Histone structure and nucleosome stability. *Expert Rev. Proteomics* **2**, 719–729 (2005).
7. Huisinga, K. L., Brower-Toland, B. & Elgin, S. C. R. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma* **115**, 110–122 (2006).
8. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
9. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549–561 (1989).

10. Wu, C.-H. *et al.* NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes Dev.* **17**, 1402–1414 (2003).
11. Neugebauer, K. M. Nascent RNA and the Coordination of Splicing with Transcription. *Cold Spring Harb. Perspect. Biol.* **11**, a032227 (2019).
12. Bowman, G. D. & Poirier, M. G. Post-translational modifications of histones that influence nucleosome dynamics. *Chem. Rev.* **115**, 2274–2295 (2015).
13. Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447**, 407–412 (2007).
14. Suganuma, T. & Workman, J. L. Crosstalk among Histone Modifications. *Cell* **135**, 604–607 (2008).
15. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
16. Parra, M. A., Kerr, D., Fahy, D., Pouchnik, D. J. & Wyrick, J. J. Deciphering the Roles of the Histone H2B N-Terminal Domain in Genome-Wide Transcription. *Mol. Cell. Biol.* **26**, 3842–3852 (2006).
17. Gnesutta, N., Nardini, M. & Mantovani, R. The H2A/H2B-like histone-fold domain proteins at the crossroad between chromatin and different DNA metabolisms. *Transcription* **4**, 114–119 (2013).
18. Moreland, R. B., Langevin, G. L., Singer, R. H., Garcea, R. L. & Hereford, L. M. Amino acid sequences that determine the nuclear localization of yeast histone 2B. *Mol. Cell. Biol.* **7**, 4048–4057 (1987).
19. Molden, R. C., Bhanu, N. V., LeRoy, G., Arnaudo, A. M. & Garcia, B. A. Multi-faceted quantitative proteomics analysis of histone H2B isoforms and their modifications. *Epigenetics Chromatin* **8**, 15 (2015).
20. Takada, K., Hibi, N., Tsukada, Y., Shibasaki, T. & Ohkawa, K. Ability of ubiquitin radioimmunoassay to discriminate between monoubiquitin and multi-ubiquitin chains. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1290**, 282–288 (1996).

21. Ryu, K.-Y., Baker, R. T. & Kopito, R. R. Ubiquitin-specific protease 2 as a tool for quantification of total ubiquitin levels in biological specimens. *Anal. Biochem.* **1**, 153–155 (2006).
22. Scheffner, M., Nuber, U. & Huibregtse, J. M. Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* **373**, 81–83 (1995).
23. Structure of the human SAGA coactivator complex | Nature Structural & Molecular Biology. https://www.nature.com/articles/s41594-021-00682-7.
24. Love, K. R., Catic, A., Schlieker, C. & Ploegh, H. L. Mechanisms, biology and inhibitors of deubiquitinating enzymes - Nature Chemical Biology. *Nat. Chem. Biol.* **3**, 697–705 (2007).
25. Chen, S., Li, J., Wang, D.-L. & Sun, F.-L. Histone H2B lysine 120 monoubiquitination is required for embryonic stem cell differentiation. *Cell Res.* **22**, 1402–1405 (2012).
26. Fierz, B. *et al.* Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nat. Chem. Biol.* **7**, 113–119 (2011).
27. Tanny, J. C., Erdjument-Bromage, H., Tempst, P. & Allis, C. D. Ubiquitylation of histone H2B controls RNA polymerase II transcription elongation independently of histone H3 methylation. *Genes Dev.* **21**, 835–847 (2007).
28. Wu, L., Li, L., Zhou, B., Qin, Z. & Dou, Y. H2B Ubiquitylation Promotes RNA Pol II Processivity via PAF1 and pTEFb. *Mol. Cell* **54**, 920–931 (2014).
29. Batta, K., Zhang, Z., Yen, K., Goffman, D. B. & Pugh, B. F. Genome-wide function of H2B ubiquitylation in promoter and genic regions. *Genes Dev.* **25**, 2254–2265 (2011).
30. Chandrasekharan, M. B., Huang, F. & Sun, Z.-W. Ubiquitination of histone H2B regulates chromatin dynamics by enhancing nucleosome stability. *Proc. Natl. Acad. Sci.* **106**, 16686–16691 (2009).

31. Worden, E. J., Hoffmann, N. A., Hicks, C. W. & Wolberger, C. Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. *Cell* **176**, 1490-1501.e12 (2019).
32. Murawska, M. *et al.* The Chaperone FACT and Histone H2B Ubiquitination Maintain S. pombe Genome Architecture through Genic and Subtelomeric Functions. *Mol. Cell* **77**, 501-513.e7 (2020).
33. Formosa, T. The role of FACT in making and breaking nucleosomes. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1819**, 247–255 (2012).
34. Hudson, W. H. & Ortlund, E. A. The structure, function and evolution of proteins that bind DNA and RNA. *Nat. Rev. Mol. Cell Biol.* **15**, 749–760 (2014).
35. Battaglia, S. *et al.* RNA-dependent chromatin association of transcription elongation factors and Pol II CTD kinases. *eLife* **6**, e25637.
36. Xiao, R. *et al.* Pervasive Chromatin-RNA Binding Protein Interactions Enable RNA-Based Regulation of Transcription. *Cell* **178**, 107-121.e18 (2019).
37. Skalska, L. *et al.* Nascent RNA antagonizes the interaction of a set of regulatory proteins with chromatin. *Mol. Cell* **81**, 2944-2959.e10 (2021).
38. Spatial organization of transcribed eukaryotic genes | Nature Cell Biology. https://www.nature.com/articles/s41556-022-00847-6.
39. Structure and Function of the RING Domains of RNF20 and RNF40, Dimeric E3 Ligases that Monoubiquitylate Histone H2B - PubMed. https://pubmed.ncbi.nlm.nih.gov/27569044/.
40. Zhang, F. & Yu, X. WAC, a Functional Partner of RNF20/40, Regulates Histone H2B Ubiquitination and Gene Transcription. *Mol. Cell* **41**, 384–397 (2011).
41. Fuchs, G. & Oren, M. Writing and reading H2B monoubiquitylation. *Biochim. Biophys. Acta* **1839**, 694–701 (2014).

42. Wu, C. *et al.* The RNF20/40 complex regulates p53-dependent gene transcription and mRNA splicing. *J. Mol. Cell Biol.* **12**, 113–124 (2020).
43. Bellare, P., Kutach, A. K., Rines, A. K., Guthrie, C. & Sontheimer, E. J. Ubiquitin binding by a variant Jab1/MPN domain in the essential pre-mRNA splicing factor Prp8p. *RNA* **12**, 292–302 (2006).
44. Cao, J. & Yan, Q. Histone Ubiquitination and Deubiquitination in Transcription, DNA Damage Response, and Cancer. *Front. Oncol.* **2**, (2012).
45. Kao, C.-F. *et al.* Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B. *Genes Dev.* **18**, 184–195 (2004).
46. Qian, M.-X. *et al.* Acetylation-Mediated Proteasomal Degradation of Core Histones during DNA Repair and Spermatogenesis. *Cell* **153**, 1012–1024 (2013).
47. Ubiquitin ligase RNF20/40 facilitates spindle assembly and promotes breast carcinogenesis through stabilizing motor protein Eg5 | Nature Communications. https://www.nature.com/articles/ncomms12648.
48. In, S., Kim, Y.-I., Lee, J. E. & Kim, J. RNF20/40-mediated eEF1BδL monoubiquitylation stimulates transcription of heat shock-responsive genes. *Nucleic Acids Res.* **47**, 2840–2855 (2019).
49. Sethi, G., Shanmugam, M. K., Arfuso, F. & Kumar, A. P. Role of RNF20 in cancer development and progression – a comprehensive review. *Biosci. Rep.* **38**, BSR20171287 (2018).
50. Jacob, A. G. & Smith, C. W. J. Intron retention as a component of regulated gene expression programs. *Hum. Genet.* **136**, 1043–1057 (2017).
51. Kitamura, K. & Nimura, K. Regulation of RNA Splicing: Aberrant Splicing Regulation and Therapeutic Targets in Cancer. *Cells* **10**, 923 (2021).
52. Levesque, L., Salazar, N. & Roy, S. W. Distinct Minor Splicing Patterns across Cancers. *Genes* **13**, 387 (2022).

53. Bao, X. *et al.* Capturing the interactome of newly transcribed RNA. *Nat. Methods* **15**, 213–220 (2018).
54. Chernikova, S. B. *et al.* Deficiency in mammalian histone H2B ubiquitin ligase Bre1 (Rnf20/Rnf40) leads to replication stress and chromosomal instability. *Cancer Res.* **72**, 2111–2119 (2012).
55. Nabet, B. *et al.* The dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.* **14**, 431–441 (2018).
56. Guarnaccia, A. D., Weissmiller, A. M. & Tansey, W. P. Gene-specific quantification of nascent transcription following targeted degradation of endogenous proteins in cultured cells. *STAR Protoc.* **2**, 101000 (2021).
57. Kim, J. B. & Lee, J. H. Use of Rnf20 for Diagnosis and Treatment of Kidney or Liver Cancer and Screening for Therapeutic Agents for Kidney or Liver Cancer. (2018).